# Contents

1	Introduction to quantum mechanics			
2	The hydrogen atom         2.1       Spectrum of $\mathcal{L}^2$	<b>6</b> 7		
3	Spin         3.1       Heisenberg Uncertainty Principle         3.2       Quantum numbers	<b>13</b> 19 21		
4	Combining two spin- $\frac{1}{2}$ particles 4.1 The hydrogen molecule ion $H_2^+$	<b>23</b> 26		
<b>5</b>	Slater Determinant	27		
6	Hartree–Fock Theory         6.1       One-body part         6.2       Two-body part         6.3       Spin symmetries (Fokotume)	28 28 28 33		
7	Second Quantization7.1The Fermionic Fock space7.2The Jordan-Wigner transformation7.3Hamiltonian in second quantization	<b>35</b> 35 35 37		
8	Contractions         8.1       Wick's contraction theorem         8.2       Particle-Hole formalism	<b>38</b> 39 40		
9	Post Hartree–Fock Methods         9.1       Configuration interaction	44 46 48 49 52 53 56		
	<ul> <li>9.4.1 Size-extensivity of the exact wave function</li></ul>	56 57 57 59 60		
10	Molecular Properties and Response Theory         10.1 Geometric derivatives         10.2 Electric properties         10.2.1 Magnetic properties         10.3 Computation of properties         10.3.1 Sum-over-state expression for exact wave functions	<b>61</b> 61 63 63 64		

	1	0.3.2 General expressions for approximate methods													64
	1	0.3.3 First-order energy derivative													65
	1	0.3.4 Second-order energy derivative													65
	1	0.3.5 Linear response equations													66
	1	0.3.6 Lagrangian formalism for non-variational param	met	ters	5.										66
-	10.4 E	Example of linear-response equations for Hartree-Fock	Ξ.												66
	1	0.4.1 Exponential parametrization													66
	1	0.4.2 Electronic gradient and Hessian			•	• •	•	•	•	•	•	•	•	•	67
11 \$	Specia	al topics													68
-	11.1 T	Cailored coupled cluster													68
-	11.2 G	Green's functions													68

### **1** Introduction to quantum mechanics

The theory of quantum mechanics explores the behavior of matter and energy at atomic scales. At these scales, the classical rules of physics we are familiar with no longer apply. This introduction is designed for graduate students in mathematics with little to no background in physics. We will use the famous double-slit experiment as a gateway to uncover key ideas in quantum mechanics, culminating in the electronic Schrödinger equation, a fundamental tool for studying particles at the atomic level. We will come to the conclusion that electrons behave like both particles and waves which is know as the particle-wave duality. As most readers will be – at least at the intuitive level – familiar with classical mechanics, we will begin our exposition with classical physics and then transition into the quantum theory of particles.

#### Point particles vs classical waves

In classical physics, waves and point particles are two very different models for physical systems, each with an exceptionally large range of applications.

Point particles which are idealized particles with no spatial extension are very commonly used in physics to describe the behavior of (uniform) rigid body dynamics, such as the "projectile motion" or "billiard mechanics". The notion of (point) particles is governed by Newtonian mechanics going back to the  $17^{th}$  century, which is manifested in three Newtonian Law's of motion

- 1. A body remains at rest, or in motion at a constant speed in a straight line, except insofar as it is acted upon by a force.
- 2. At any instant of time, the net force on a body is equal to the rate at which the body's momentum is changing with time.

$$F = \frac{dp}{dt} = m\frac{dv}{dt} = ma \tag{1}$$

3. If two bodies exert forces on each other, these forces have the same magnitude but opposite directions.

The describing quantities of (point) particles are (1) their mass, (2) their position, and (3) their forces.

Classical waves are propagating dynamic disturbances, i.e., changes from an equilibrium state. Periodic waves oscillate repeatedly about an equilibrium (resting) value at some frequency. When the entire waveform moves in one direction, it is said to be a traveling wave; by contrast, a pair of superimposed periodic waves traveling in opposite directions makes a standing wave. In a standing wave, the amplitude of vibration has nulls at some positions where the wave amplitude appears smaller or even zero. Commonly, a wave is described by a scalar field

$$u: \mathbb{R}^3 \times \mathbb{R}_{\geq 0} \to \mathbb{R} \; ; \; (\mathbf{r}, t) \mapsto u(\mathbf{r}, t)$$
(2)

assigning at any point in time  $t \in \mathbb{R}$  and any point in space  $\mathbf{r} \in \mathbb{R}^3$  a magnitude  $u(\mathbf{r}, t)$ . Classical waves are governed by the wave equation, i.e.,

$$\frac{\partial^2 u}{\partial t^2} = c^2 \Delta_{\mathbf{r}} u,\tag{3}$$

where  $c \in \mathbb{R}_{>0}$ . Unlike point particles, waves have continuous values at many points in space that vary with time. The spatial extent of waves can vary with time due to diffraction or wave interferences, see Fig. 1. These are characteristic observations of waves.





Figure 1: (a) Left panel shows diffraction of water wave on a single list. (b) Right panel shows interference pattern of concentric waves.

Unlike waves, point particles do not exhibit diffraction or interference.

#### The Double-Slit Experiment

We will transition to atomic scales; imagine an apparatus where particles at the atomic scale, such as electrons, are fired onto a detector, see Fig. 2

The particles can be individually detected and appear to behave like point particles in classical physics. Now, imagine a setup where the particles are sent toward a barrier with two small openings, called slits, see 3. Beyond the slits is a screen that detects where the particles land.

We now consider two scenarios:



Figure 2: electron beam



Figure 3: electron beam

- 1. Blocking one slit: If we close one slit, the particles form a simple pattern on the screen directly behind the open slit, see ??. This result again aligns with our expectation from the point particle picture.
- 2. **Opening both slits**: With both slits open, we expect the result to be the sum of the patterns from each slit. However, what we observe is far more surprising; a series of bright and dark bands, called an interference pattern, similar to what we see if waves (not particles) were passing through the slits, see 3.

This experiment shows that particles such as electrons behave like waves under certain conditions. More surprisingly, if we place detectors at the slits to observe which slit the individual electron passes through, see ??, the interference pattern disappears, and the electrons behave like particles again. This puzzling behavior is at the heart of quantum mechanics.

But, how do we interpret this result and what does it mean for our understanding of physics? The double-slit experiment suggests that particles have a dual nature: they can behave like both, particles and waves. This is called the *wave-particle duality* challenging our classical picture of physics in which these concepts are fundamentally different. The quantum mechanical framework allows us to describe the interference pattern arising in the dual slit experiment. Quantum mechanically, particles are characterized by a *wave function*, which determines the probability of finding the particle in a particular location. As the name suggests, the propagation of the particle's wave function is governed by "some type of wave equation". In particular, when no measurement is made, the particle's wave function spreads out like a wave producing the wave-like interference patterns observed in the double-slit experiment. However, when we measure which slit the particle goes through, the wave function "collapses" into a specific location, and the interference pattern vanishes – the electron now behaves like

a classical particle.

Mathematically, the wave function is a function in space and time, denoted by

$$\psi: \mathbb{R}^m \times \mathbb{R}_{\geq 0} \to \mathbb{C} \; ; \; (\mathbf{x}, t) \mapsto \psi(\mathbf{x}, t) \tag{4}$$

where  $m \in \mathbb{N}$ . The wave function is the central object in quantum mechanics that encapsulates all the information about the system's state. Its squared modulus is the probability density of finding the particles at positions  $\mathbf{x} \in \mathbb{R}^m$  at time  $t \in \mathbb{R}_{\geq 0}$ , hence,  $\psi \in L^2(\mathbb{R}^m)$  for all  $t \in \mathbb{R}_{\geq 0}$ . The governing equation for the wave function is the Schrödinger equation, proposed by Erwin Schrödinger in 1925 [?]. In it's time-dependent form, the Schrödinger equation reads

$$i\frac{\partial\psi(\mathbf{x},t)}{\partial t} = H\psi(\mathbf{x},t),\tag{5}$$

where *i* is the imaginary unit and *H* is the Hamiltonian operator, representing the total energy of the system of interest. For a particle moving in a potential  $V(\mathbf{x})$ , the Hamiltonian is

$$H = -\frac{1}{2}\nabla^2 + V(\mathbf{x}),\tag{6}$$

where  $-\frac{1}{2}\nabla^2$  describes the kinetic energy, and  $V(\mathbf{x})$  describes the potential energy. In many cases, however, we are interested in systems that do not change in time, such as electrons in a stable atom or molecule. Here, the Schrödinger equation simplifies to

$$H\psi(x) = E\psi(x),\tag{7}$$

where E describes the system's energy. Narrowing down on the application to atoms and molecules, we can specify the potential further. Consider a molecule with  $N_{\text{nuc}}$  nuclei and N electrons, the wave function depends on both nuclear positions  $\mathbf{R} = (\mathbf{R}_1, ..., \mathbf{R}_{N_{\text{nuc}}}) \in \mathbb{R}^{3N_{\text{nuc}}}$  with  $\mathbf{R}_i \mathbb{R}^3$  and the electronic positions  $\mathbf{r} = (\mathbf{r}_1, ..., \mathbf{r}_N) \in \mathbb{R}^{3N}$  with  $\mathbf{r}_i \in \mathbb{R}^3$ 

$$H\Psi(\mathbf{r},\mathbf{R}) = E\Psi(\mathbf{r},\mathbf{R}) \tag{8}$$

where

$$H = -\sum_{i \in \llbracket N \rrbracket} \frac{1}{2} \Delta_{\mathbf{r}_{i}} - \sum_{i \in \llbracket N \rrbracket} \sum_{j \in \llbracket N_{\text{nuc}} \rrbracket} \frac{Z_{j}}{\|\mathbf{r}_{i} - \mathbf{R}_{j}\|} + \frac{1}{2} \sum_{i \in \llbracket N \rrbracket} \sum_{i \neq j \in \llbracket N \rrbracket} \frac{1}{\|\mathbf{r}_{i} - \mathbf{r}_{j}\|} - \sum_{i \in \llbracket N_{\text{nuc}} \rrbracket} \frac{1}{2M_{i}} \Delta_{\mathbf{R}_{i}} + \frac{1}{2} \sum_{i \in \llbracket N_{\text{nuc}} \rrbracket} \sum_{i \neq j \in \llbracket N_{\text{nuc}} \rrbracket} \frac{Z_{i}Z_{j}}{\|\mathbf{R}_{i} - \mathbf{R}_{j}\|}$$
(9)

where  $M_i$  is the nuclear mass relative to the electronic mass, and  $Z_i$  are the atomic numbers found on the periodic table. A common simplification is the Born-Oppenheimer approximation, which assumes that nuclei move much more slowly than electrons, meaning that the kinetic potential energy coming exclusively from nuclei simply enter as a constant. This reduces the problem to the electronic Schrödinger equation:

$$H(\mathbf{R})\Psi(\mathbf{r}) = E\Psi(\mathbf{r}) \tag{10}$$

where

$$H(\mathbf{R}) = -\sum_{i \in \llbracket N \rrbracket} \frac{1}{2} \Delta_{\mathbf{r}_i} - \sum_{i \in \llbracket N \rrbracket} \sum_{j \in \llbracket N_{\text{nuc}} \rrbracket} \frac{Z_j}{\|\mathbf{r}_i - \mathbf{R}_j\|} + \frac{1}{2} \sum_{i \in \llbracket N \rrbracket} \sum_{i \neq j \in \llbracket N \rrbracket} \frac{1}{\|\mathbf{r}_i - \mathbf{r}_j\|}.$$
 (11)

Subject of this topic course will be to solve this innocent looking eigenvalue problem.

# 2 The hydrogen atom

For the hydrogen atom, the (stationary) Schrödinger equation takes the simple form

$$\left(-\frac{1}{2}\Delta - \frac{1}{\|\mathbf{R} - \mathbf{r}\|}\right)\Psi(\mathbf{r}) = E\Psi(\mathbf{r}),\tag{12}$$

where  $\mathbf{R} \in \mathbb{R}^3$  denotes the position of the hydrogen nucleus. Without loss of generality, we can define our coordinate system such that the nucleus defines the coordinate origin simplifying the equation to

$$\left(-\frac{1}{2}\Delta - \frac{1}{\|\mathbf{r}\|}\right)\Psi(\mathbf{r}) = E\Psi(\mathbf{r}).$$
(13)

We note that this is a spherically symmetric problem. This can be seen as follows: Consider the rotation matrix R, then

$$\mathbf{r} \mapsto R\mathbf{r} = \mathbf{r}' \tag{14}$$

describes a rotation of the frame. Since  ${\cal R}$  is unitary and the 2-norm is unitarily invariant, we have

$$\|\mathbf{r}'\| = \|R\mathbf{r}\| = \|\mathbf{r}\|.$$
(15)

Moreover, in the rotated frame the partial derivative is given by

$$\partial_{x'_i} = \sum_j R_{i,j} \partial_{x_j} \qquad \Rightarrow \qquad \partial^2_{x'_i} = \sum_{k,j} R_{i,k} R_{i,j} \partial_{x_j} \partial_{x_k}. \tag{16}$$

Then

$$\Delta' f = \sum_{i} \sum_{k,j} R_{i,k} R_{i,j} \partial_{x_j} \partial_{x_k} f = \sum_{k,j} \left( \sum_{i} R_{i,k} R_{i,j} \right) \partial_{x_j} \partial_{x_k} f = \sum_{k} \partial_{x_k}^2 f = \Delta f.$$
(17)

This suggests that we investigate this problem in spherical coordinates. Recall the spherical coordinate transformation

$$s: [0,\pi] \times [0,2\pi) \times \mathbb{R}_{\geq 0} \to \mathbb{R}^3; \quad \begin{bmatrix} \theta \\ \phi \\ r \end{bmatrix} \mapsto \begin{bmatrix} r\sin(\theta)\cos(\phi) \\ r\sin(\theta)\sin(\phi) \\ r\cos(\theta) \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$
(18)

and

$$s^{-1}: \mathbb{R}^3 \to [0,\pi] \times [0,2\pi) \times \mathbb{R}_{\geq 0} ; \quad \begin{bmatrix} x\\ y\\ z \end{bmatrix} \mapsto \begin{bmatrix} \tan^{-1}(y/x)\\ \cos^{-1}(z/\sqrt{x^2+y^2+z^2})\\ \sqrt{x^2+y^2+z^2} \end{bmatrix} = \begin{bmatrix} \phi\\ \theta\\ r \end{bmatrix}$$
(19)

The Laplacian in spherical coordinates is given by

$$\Delta = \frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \phi^2}$$
$$=: \frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial}{\partial r} \right) - \frac{1}{r^2} \mathcal{L}^2$$
(20)

where we defined

$$\mathcal{L}^{2} = -\frac{1}{\sin\theta} \frac{\partial}{\partial\theta} \left( \sin\theta \frac{\partial}{\partial\theta} \right) - \frac{1}{\sin^{2}\theta} \frac{\partial^{2}}{\partial\phi^{2}}.$$
 (21)

The separation of the radial and angular part in Eq. (20) suggests to use the separation of variables ansatz, i.e.,

$$\Psi(r,\theta,\phi) = R(r)Y(\theta,\phi).$$
(22)

Inserting this ansatz into the Schrödinger equation (12), this yields

$$\begin{pmatrix} -\frac{1}{2}\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{1}{2}\frac{1}{r^2}\mathcal{L}^2 - \frac{1}{|r|}\right)R(r)Y(\theta,\phi) = ER(r)Y(\theta,\phi) \\ \Leftrightarrow \qquad \left(-\frac{1}{2}\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) - \frac{1}{|r|} - E\right)R(r) + \frac{R(r)}{2r^2}\frac{\mathcal{L}^2Y(\theta,\phi)}{Y(\theta,\phi)} = 0$$

$$(23)$$

We see that solving this equation is subject to characterizing the spectrum of  $\mathcal{L}^2$ .

# 2.1 Spectrum of $\mathcal{L}^2$

We seek to solve the eigenvalue problem

$$\mathcal{L}^2 Y(\theta, \phi) = k Y(\theta, \phi). \tag{24}$$

Given the structure of the operator  $\mathcal{L}^2$ , see Eq. (21), we again propose the separation of variables ansatz

$$Y(\theta,\phi) = \Theta(\theta)\Phi(\phi), \tag{25}$$

which yields

$$\begin{pmatrix} -\frac{1}{\sin\theta} \frac{\partial}{\partial\theta} \left( \sin\theta \frac{\partial}{\partial\theta} \right) - \frac{1}{\sin^2\theta} \frac{\partial^2}{\partial\phi^2} \right) \Theta(\theta) \Phi(\phi) = k\Theta(\theta) \Phi(\phi)$$

$$\Leftrightarrow \qquad \frac{\sin^2\theta}{\Theta(\theta)} \left( -\frac{1}{\sin\theta} \frac{\partial}{\partial\theta} \left( \sin\theta \frac{\partial}{\partial\theta} \right) - k \right) \Theta(\theta) - \frac{1}{\Phi(\phi)} \frac{\partial^2}{\partial\phi^2} \Phi(\phi) = 0$$

$$(26)$$

We first separate  $\Phi(\phi)$  which yields

$$\begin{cases} -\frac{\partial^2}{\partial\phi^2}\Phi(\phi) = m^2\Phi(\phi) \\ \Phi(0) = \Phi(2\pi) \end{cases}$$
(27)

with separation constant  $m^2$ , and periodic boundary condition because of the spherical symmetry of the problem. The solution to Eq. (27) is given by

$$\Phi(\phi) = A e^{im\phi},\tag{28}$$

with  $m \in \mathbb{Z}$  to fulfill the azimuthal periodic boundary condition.

For  $\Theta(\theta)$  we then obtain

$$\left(-\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{m^2}{\sin^2\theta}\right)\Theta(\theta) = k\Theta(\theta)$$
(29)

Performing the coordinate transformation

$$\zeta := \cos(\theta) \quad \text{and} \quad \xi(\cos(\theta)) := \Theta(\theta)$$
(30)

yields

$$\frac{d}{d\zeta} \left( (1-\zeta)^2 \frac{d\xi}{d\zeta} \right) + \left( k - \frac{m^2}{1-\zeta^2} \right) \xi = 0$$
(31)

with  $|\xi(1)|, |\xi(-1)| < \infty$ . This differential equation is a (well-)known differential equation, namely, the generalized Legendre differential equation [?]. Recall the conventional Legendre equation, i.e.,

$$(1-x^2)\frac{d^2y}{dx^2} - 2x\frac{dy}{dx} + \ell(\ell+1)y = 0,$$
(32)

where  $\ell$  is the degree. This equation describes solutions to Laplace's equation for azimuthally symmetric boundary conditions, i.e., when the solution has no dependence on the azimuthal angle  $\phi$ . However, to account for non-zero azimuthal dependence, particularly in gravitational and electromagnetic potentials, the equation was extended to the generalized Legendre differential equation, i.e.,

$$(1-x^2)\frac{d^2y}{dx^2} - 2x\frac{dy}{dx} + \left[\ell(\ell+1) - \frac{m^2}{1-x^2}\right]y = 0,$$
(33)

introducing the order m, which corresponds to the projection of angular momentum along the z-axis. The solutions to this equation are the associated Legendre functions  $P_{\ell}^{m}(x)$ , which are derived from the Legendre polynomials  $P_{\ell}(x)$  using the relation

$$P_{\ell}^{m}(x) = (1 - x^{2})^{\frac{m}{2}} \frac{d^{m}}{dx^{m}} P_{\ell}(x).$$
(34)

These functions generalize  $P_{\ell}(x)$  by incorporating azimuthal dependence and are crucial in describing solutions to Laplace's equation in spherical coordinates for problems with full rotational symmetry. This formalism, developed to handle problems involving angular momentum and wave-like behavior, became essential in physics, geophysics, and engineering.

The Eq. (31) is a Legendre differential equation on [-1, 1], which yields

$$k = \ell(\ell+1), \qquad \ell \in \mathbb{Z}_{\geq 0} \tag{35}$$

Each eigenvalue  $\ell(\ell+1)$  corresponds to  $\ell+1$  degenerate, orthogonal eigenfunctions, denoted

$$\xi(\zeta) = P_{\ell}^{m}(\zeta) := (-1)^{m} (1 - \zeta^{2})^{\frac{m}{2}} \frac{d^{m}}{d\zeta^{m}} P_{\ell}(\zeta), \qquad m \in \{0, ..., \ell\}$$
(36)

where  $P_{\ell}^m$  are the associated Legendre polynomials and

$$P_{\ell}(\zeta) = \frac{(-1)^{\ell}}{2^{\ell}\ell!} \frac{d^{\ell}}{d\zeta^{\ell}} (1-\zeta^2)^{\ell}$$
(37)

are the Legendre polynomials [?]. We also define the *negative associated Legendre polyno*mials as

$$P_{\ell}^{-m}(\zeta) = (-1)^m \frac{(\ell-m)!}{(\ell+m)!} P_{\ell}^m(\zeta)$$
(38)

Example 2.0.1. The first six Legendre polynomials are given by

 $P_0(\zeta) = 1 \qquad P_1(\zeta) = \zeta \qquad P_2(\zeta) = \frac{1}{2}(3\zeta^2 - 1)$  $P_3(\zeta) = \frac{1}{2}(5\zeta^3 - 3\zeta) \qquad P_4(\zeta) = \frac{1}{8}(35\zeta^4 - 30\zeta^2 + 3) \qquad P_5(\zeta) = \frac{1}{8}(63\zeta^5 - 70\zeta^3 + 15\zeta)$ 

The first few associated Legendre polynomials are give by

$$P_0^0(x) = 1$$

$$P_1^{-1}(x) = \frac{(1-x^2)^{1/2}}{2} \qquad P_1^0(x) = x \qquad P_1^1(x) = -(1-x^2)^{1/2}$$

$$P_2^{-2}(x) = \frac{3(1-x^2)}{24} \qquad P_2^{-1}(x) = \frac{3x(1-x^2)^{1/2}}{6} \qquad P_2^0(x) = \frac{(3x^2-1)}{2} \qquad P_2^1(x) = -3x(1-x^2)^{1/2} \qquad P_2^2(x) = 3(1-x^2)$$

Changing the variables back to  $\theta$  and  $\Theta(\theta)$ , this yields that

$$\Theta(\theta) = P_{\ell}^{m}(\cos(\theta)) \tag{39}$$

Hence

$$Y_{\ell,m}(\theta,\phi) = C_{\ell,m} P_{\ell}^m(\cos(\theta)) e^{im\phi}$$
(40)

where  $C_{\ell,m}$  a normalization constant which is chosen such that

$$\int_{[0,\phi]\times[0,2\pi)} Y_{\ell,m}(\theta,\phi) Y_{\ell',m'}(\theta,\phi) \sin^2(\theta) \ d\theta d\phi = \delta_{l,l'} \delta_{m,m'}$$

$$\Leftrightarrow \qquad C_{\ell,m} = (-1)^m \sqrt{\frac{(2\ell+1)}{4\pi} \frac{(\ell-m)!}{(\ell+m)!}}$$

$$\tag{41}$$

The functions  $Y_{\ell,m}(\theta,\phi)$  are the spherical harmonics fulfilling

$$\mathcal{L}^2 Y_{\ell,m} = \ell(\ell+1) Y_{\ell,m} \tag{42}$$

where the eigenvalue had geometric multiplicity  $2\ell + 1$ .

**Example 2.0.2.** Let's consider  $\ell = 1$  and m = 0, then

$$Y_{1,0}(\theta,\phi) = \sqrt{\frac{3}{4\pi}}\cos(\theta).$$
(43)

We see that  $Y_{1,0}$  only depends on the polar angle  $\theta$  and not on the azimuthal angle  $\phi$ . A possible way to visualize this function is via the map

$$f_{\ell}^{m}: (\theta, \phi) \mapsto P_{\ell}^{m}(\cos(\theta))e^{im\phi} \begin{bmatrix} \sin(\theta)\sin(\phi)\\ \sin(\theta)\cos(\phi)\\ \cos(\theta) \end{bmatrix}$$
(44)

Using this procedure, we can visualize  $Y_{1,0}$  as shown in Fig. 4. This procedure can similarly be applied to the first few spherical harmonics in Fig. 5.



Figure 4: Construction of a visualization of  $Y_{1,0}$ .



Figure 5: Graphical representation of the first few spherical harmonics, using the function  $f_{\ell}^m$  in Eq. (44). Blue portions represent regions where the function is positive, and yellow portions represent where it is negative. The rows correspond to  $\ell = 0$  (s),  $\ell = 1$  (p),  $\ell = 2$  (d),  $\ell = 3$  (f); the columns correspond to  $m \in \{-\ell, ..., \ell\}$ .

We shall now return to Eq. (23). Having found the spectrum of  $\mathcal{L}^2$ , Eq. (23) simplifies to

$$-\frac{1}{2}\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial R(r)}{\partial r}\right) - \frac{1}{r}R(r) + \frac{R(r)}{2r^2}\ell(\ell+1) = ER(r), \qquad r > 0, \tag{45}$$

which is the radial equation. Note that

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial R(r)}{\partial r}\right) = \frac{1}{r^2}\left(2r\frac{\partial R(r)}{\partial r} + r^2\frac{\partial^2 R(r)}{\partial r^2}\right) = \left(\frac{2}{r}\frac{\partial R(r)}{\partial r} + \frac{\partial^2 R(r)}{\partial r^2}\right) \\ = \frac{1}{r}\frac{\partial}{\partial r}\left(R(r) + r\frac{\partial R(r)}{\partial r}\right) = \frac{1}{r}\frac{\partial^2}{\partial r^2}rR(r)$$
(46)

Substituting

$$u(r) = rR(r) \tag{47}$$

yields

$$-\frac{1}{2}\frac{\partial^2}{\partial r^2}rR(r) - R(r) + \frac{R(r)}{2r}\ell(\ell+1) = ErR(r)$$

$$\Rightarrow \quad -\frac{1}{2}\frac{\partial^2}{\partial r^2}u(r) + \tilde{V}(r)u(r) = Eu(r)$$

$$(48)$$

where

$$\tilde{V}(r) = \frac{\ell(\ell+1)}{2r^2} - \frac{1}{r}.$$
(49)

We will now investigate the solution to Eq. (48) in two limiting cases. First, in the far field. We note that

$$\tilde{V}(r) \xrightarrow{r \to \infty} 0 \tag{50}$$

which simplifies Eq. (48) to

$$-\frac{1}{2}\frac{\partial^2}{\partial r^2}u(r) = Eu(r).$$
(51)

For E > 0 the solution to Eq. (51) yields

$$u(r) = c_1 e^{i\sqrt{E2}r} + c_2 e^{-i\sqrt{E2}r},$$
(52)

which cannot be square integrable, hence, any value E > 0 cannot be an isolated eigenvalue. For  $E \leq 0$  the solution is given by

$$u(r) = c_1 e^{\sqrt{|E|^2}r} + c_2 e^{-\sqrt{|E|^2}r}.$$
(53)

Since this solution must remain finite as  $r \to \infty$ ,  $c_1$  must be zero, hence,

$$u(r) = ce^{-\sqrt{|E|^2r}}.$$
 (54)

Second, in the near field. We note that

$$\tilde{V}(r) \xrightarrow{r \to 0} -\frac{\ell(\ell+1)}{2r^2} \quad \text{and} \quad u(r) \xrightarrow{r \to 0} 0$$
(55)

which yields

$$-\frac{1}{2}\frac{\partial^2}{\partial r^2}u(r) - \frac{\ell(\ell+1)}{r^2}u(r) = 0.$$
 (56)

Equation (56) has the solution

$$u(r) = c_1 r^{\ell+1} + c_2 r^{-\ell} \tag{57}$$

but since  $r^{-\ell} \to \infty$  for  $r \to 0$ ,  $c_2$  must be zero, hence,

$$u(r) = cr^{\ell+1}.\tag{58}$$

Combining the near field and far field solution then yields the general solution

$$u(r) = Cr^{\ell+1} e^{-\sqrt{\frac{|E|}{2}}r} G(r)$$
(59)

where

$$G(r) = A_0 + A_1 r + A_2 r^2 + \dots$$
(60)

"stitches" the near field and far field solution together. Inserting this ansatz into Eq.  $\left(48\right)$  yields the recursion

$$A_k = -2A_{k-1} \frac{1 - (\ell + k)\sqrt{-E2}}{(\ell + k)(\ell + k + 1) - \ell(\ell + 1)}.$$
(61)

For this series to terminate we require

$$0 = 1 - (\ell + k)\sqrt{-2E} \qquad \Leftrightarrow \qquad E_{k,\ell} = -\frac{1}{2(\ell + k)^2} \tag{62}$$

# 3 Spin

We now want to transition to multi-electron systems. To that end, we need to address a fundamental quantity that is – surprisingly – not explicitly incorporated into the electronic Schrödinger equation, namely, the spin. Understanding the spin of an electron is a cornerstone of quantum mechanics, yet it can be an abstract concept for learners. The Stern-Gerlach experiment provides an intuitive and visualizable way to grasp the key ideas surrounding spin. In the early 20th century, physicists were intrigued by the idea that electrons and other particles possess intrinsic angular momentum beyond what classical physics could explain. This intrinsic property, later named *spin*, was suspected but not clearly understood. The Stern-Gerlach experiment, conducted in 1922 by Otto Stern and Walther Gerlach, aimed to probe the nature of this intrinsic property by observing the deflection of particles in a magnetic field.

The core idea of the experiment is deceptively simple. Here's a breakdown of the essential components, see Fig. 6:

- Source of Silver Atoms: The experiment initially used silver atoms which contain a single valence electron in an unpaired state (Ag has 47 electrons) as a convenient proxy for electrons.
- Collimated Beam: The silver atoms are heated and directed into a narrow beam.
- Non-Uniform Magnetic Field: The atoms pass through a region of non-uniform magnetic field, which exerts a different force depending on the magnetic moment of each atom.
- Detection Screen: After passing through the magnetic field, the atoms strike a detection screen, revealing their distribution.



Figure 6: Setup of the Stern-Gerlach experiment

Hot silver atoms emerge from an oven with an initial velocity directed along the *y*-axis. These atoms then pass through a region containing an inhomogeneous magnetic field, created by combining a uniform magnetic field oriented along the *z*-axis with a small perturbation that introduces a gradient. As the silver atoms traverse this magnetic field, their final positions along the *z*-axis are recorded on a detection screen placed to the right. Inside the oven, individual silver atoms are produced carrying one unpaired valence electron that imparts a net magnetic moment. This magnetic moment, referred to as the electron's spin, is represented by a vector  $\mu \in \mathbb{R}^3$ . The direction of this vector determines whether the atom's trajectory bends upwards or downwards within the magnetic field gradient.

Before the experiment was conducted, classical physics predicted that the magnetic moments of the atoms would orient randomly. As a result, a continuous smear of detection points was expected on the screen, see Fig. 7.

However, quantum mechanics made a strikingly different prediction: due to the quantized nature of spin, the beam should split into discrete parts, see Fig. 7.



Figure 7: (a) The prediction of the result of the Stern–Gerlach experiment from classical theory and (b) the experimental result.

By blocking one particle beam, we may use the Stern-Gerlach experiment to generate filters, see Fig. 8. Embracing the Dirac notation, we denote the two states that pass through a Stern-Gerlach apparatus with magnetic field in z-direction (SG<sub>z</sub> apparatus) by  $|+_z\rangle$  and  $|-_z\rangle$ , respectively, see Fig. 8.



Figure 8: (a) The Stern–Gerlach apparatus along the z-direction and (b) a filtering apparatus.

Note that the spatial orientation of the magnetic field does not have any fundamental effect on the outcome. More precisely, we may rotate the apparatus in a way that the magnetic field is oriented along the x-axis instead, and the outcome would be the same as above, however, rotated around the y-axis. The resulting apparatus would be a Stern-Gerlach apparatus with a magnetic field in x-direction (SG<sub>x</sub> apparatus) and the out-coming states would be  $|+_x\rangle$  and  $|-_x\rangle$ , respectively.

Having the ability to combine different filters may allow us to perform a number of experiments. First, we would combine two  $SG_z$  apparatuses – or equivalently two  $SG_x$  apparatuses. Once a spin is filtered out, we do not observe a beam split if we filter twice or n times in fact, see Fig. 9. On the other hand, filtering one spin out of the beam does not affect the other spin, i.e., once an x-spin of filtered out, we still observe a beam split when transitioning through a  $SZ_z$  apparatus.



Figure 9: (a) The Stern–Gerlach apparatus along the z-direction and (b) a filtering apparatus.

An interesting observation happens when concatenating three SG filters,  $SZ_z-SZ_x-SZ_z$ . Intuitively, the first filter would filter out a spin component in z-direction, the second filter would filter out a spin component in x-direction, and we should no observe a beam split when the beam transitions the third SZ filter. However, the experiment shows a different outcome, see Fig. 10.



Figure 10: Sequence of three SG filters  $SZ_z-SZ_x-SZ_z$ .

The solution to this mysterious outcome can be obtained using only  $2 \times 2$  matrices and will lead to the Heisenberg uncertainty principle. The theory of a single spin- $\frac{1}{2}$  particle can described using a two-dimensional vector space  $\mathcal{H} \simeq \mathbb{C}^2$  (isometric isomorphic) called the state vector space. The states  $|\pm_z\rangle$  form a basis of  $\mathcal{H}$ , i.e., for any  $|\psi\rangle \in \mathcal{H}$  there exist  $ci_1, c_2 \in \mathbb{C}$ , s.t.

$$|\psi\rangle = c_1|+_z\rangle + c_2|-_z\rangle. \tag{63}$$

In particular,  $|\pm_x\rangle$  are also states in  $\mathcal{H}$  and can be extended using  $|\pm_z\rangle$ . Moreover,  $\mathcal{H}$  is a Hilbert space, and  $|\pm_z\rangle$  are orthonormal w.r.t. the inner product structure. In Dirac notation, this means

$$\langle +_z | +_z \rangle = \langle -_z | -_z \rangle = 1$$
 and  $\langle +_z | -_z \rangle = 0.$  (64)

Now, since the orientation of the Stern-Gerlach experiment is somewhat arbitrary, we may change  $|\pm_z\rangle$  to  $|\pm_x\rangle$  in the above discussion. This in turn means that  $|\pm_x\rangle$  are linearly independent and therewith form a basis of  $\mathcal{H}$ .

In quantum mechanics, a measurement corresponds to a linear operator that is self-adjoint on  $\mathcal{H}$ , and the possible outcomes correspond to the operator's spectrum. The Stern-Gerlach apparatus SG<sub>z</sub> corresponds to a measurement that measures the spin in z-direction. According to the above postulate, this means that it corresponds to a linear and self-adjoint operator on  $\mathcal{H}$ , we shall denote it  $S_z$ , and we know the eigenstates of  $S_z$  are  $|\pm_z\rangle$ . Similarly, we may define  $S_x$  and  $S_y$  together with their spectra  $|\pm_x\rangle$  and  $|\pm_y\rangle$ , respectively. Since  $|\pm_x\rangle \in \mathcal{H}$ and  $|\pm_y\rangle \in \mathcal{H}$  we may wonder about their expansion using the basis  $|\pm_z\rangle$ . The Stern-Gerlach experiment provides some insight into this expansion. We first note that when  $|+_x\rangle$  passes through SG<sub>z</sub> the outcome is  $|\pm_x\rangle$  with equal probability, i.e., the outcome suggests a bimodal symmetric distribution. Sine we moreover have a phase that can be introduced, we find

$$|+_{x}\rangle = \frac{1}{\sqrt{2}} \left( |+_{z}\rangle + e^{i\alpha} |-_{z}\rangle \right) \tag{65}$$

and we find

$$|-_x\rangle = \frac{1}{\sqrt{2}} \left( |+_z\rangle - e^{i\alpha} |-_z\rangle \right) \tag{66}$$

Similarly we find

$$|+_{y}\rangle = \frac{1}{\sqrt{2}} \left( |+_{z}\rangle + e^{i\beta}|-_{z}\rangle \right) \tag{67}$$

and

$$|-_{y}\rangle = \frac{1}{\sqrt{2}} \left( |+_{z}\rangle - e^{i\beta}|-_{z}\rangle \right) \tag{68}$$

In order to relate  $\alpha$  and  $\beta$  we consider passing  $|\pm_x\rangle$  through SG<sub>y</sub>; which yields a bimodal distribution! Hence

$$|\langle \pm_y | \pm_x \rangle|^2 = \frac{1}{2} \tag{69}$$

Since

$$\langle \pm_y | \pm_x \rangle = \frac{1}{2} \left( \langle +_z | \pm e^{-i\beta} \langle -_z | \right) \left( | +_z \rangle \pm e^{i\alpha} | -_z \rangle \right)$$

$$= \frac{1}{2} \left( \langle +_z | +_z \rangle \pm e^{-i\beta} \langle -_z | +_z \rangle \pm e^{i\alpha} \langle +_z | -_z \rangle + e^{-i(\beta - \alpha)} \langle -_z | -_z \rangle \right)$$

$$= \frac{1}{2} \left( 1 + e^{-i(\beta - \alpha)} \right)$$

$$(70)$$

this yields

$$\frac{1}{2} = |\langle \pm_y | \pm_x \rangle|^2 = \frac{1}{2} \left( 1 + \cos(\beta - \alpha) \right)$$
(71)

and therewith

$$\alpha - \beta = \pm \frac{\pi}{2} + 2\pi n, \qquad n \in \mathbb{Z}.$$
(72)

We introduce the convention that  $\alpha=0$  and  $\beta=\pi/2$  which yields that

$$|\pm_x\rangle = \frac{1}{\sqrt{2}}(|+_z\rangle \pm |-_z\rangle) \quad \text{and} \quad |\pm_y\rangle = \frac{1}{\sqrt{2}}(|+_z\rangle \pm i|-_z\rangle)$$
(73)

Since  $\mathcal{H} \simeq \mathbb{C}^2$  we may perform the identification

$$|+_{z}\rangle = \begin{bmatrix} 1\\0 \end{bmatrix}$$
 and  $|-_{z}\rangle = \begin{bmatrix} 0\\1 \end{bmatrix}$  (74)

which yields

$$|+_x\rangle = \frac{1}{\sqrt{2}} \begin{bmatrix} 1\\1 \end{bmatrix}$$
 and  $|-_x\rangle = \frac{1}{\sqrt{2}} \begin{bmatrix} 1\\-1 \end{bmatrix}$  (75)

and

$$|+_{y}\rangle = \frac{1}{\sqrt{2}} \begin{bmatrix} 1\\ i \end{bmatrix}$$
 and  $|-_{y}\rangle = \frac{1}{\sqrt{2}} \begin{bmatrix} 1\\ -i \end{bmatrix}$  (76)

This choice of functions allows us to formulate  $S_x$ ,  $S_z$ , and  $S_z$  in matrix form. To that end, we use the spectral theorem which yields

$$S_{z} = \frac{1}{2} \left( |+_{z} \times +_{z}| - |-_{z} \times -_{z}| \right) = \frac{1}{2} \begin{bmatrix} 1 & 0\\ 0 & -1 \end{bmatrix}$$
(77)

as  $|\pm_z\rangle$  are eigenvectors corresponding to eigenvalues  $\pm \frac{1}{2}$ . The choice of these eigenvalues Similarly we find

$$S_{x} = \frac{1}{2} (|+_{x} \times +_{x}| - |-_{x} \times -_{x}|) = \frac{1}{2} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$S_{y} = \frac{1}{2} (|+_{y} \times +_{y}| - |-_{y} \times -_{y}|) = \frac{1}{2} \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}$$
(78)

Following the convention that  $S = (S_x, S_y, S_z) = \frac{1}{2}(\sigma_x, \sigma_y, \sigma_z)$  we define the Pauli matrices as

$$\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad \sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \qquad \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$
(79)

One immediate observation is that linear operators for different observables do not necessarily commute, e.g.,

$$S_{x}S_{z} = \frac{1}{4} \begin{bmatrix} 0 & 1\\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0\\ 0 & -1 \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 0 & -1\\ 1 & 0 \end{bmatrix}$$
  
$$S_{z}S_{x} = \frac{1}{4} \begin{bmatrix} 1 & 0\\ 0 & -1 \end{bmatrix} \begin{bmatrix} 0 & 1\\ 1 & 0 \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 0 & 1\\ -1 & 0 \end{bmatrix}$$
(80)

which yields

$$[S_z, S_x] = S_z S_x - S_x S_z = \frac{1}{2} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = i S_y$$
(81)

This means that  $S_x$  and  $S_z$  are not simultaneously diagonalizable and we refer to them as incompatible in a quantum mechanical context. Similarly, we find that

$$[S_x, S_y] = iS_z \quad \text{and} \quad [S_y, S_z] = iS_x \tag{82}$$

Hence,  $S_x, S_y$  and  $S_z$  are mutually incompatible. On the other hand

$$S^2 = S_x^2 + S_y^2 + S_z^2 \tag{83}$$

is compatible with all spin operators along any individual direction.

Compatibility has an important physical consequence. Recall that the quantum mechanics postulation that the final state from any measurement leads to an eigenstate of the operator corresponding to a physical observable. Then if two operators can be simultaneously diagonalized using the same set of eigenstates, it means that one can simultaneously measure the corresponding quantities, e.g., the spin magnitude  $S^2$  and the individual spin operators. The compatibility condition is sufficient and necessary. In other words, if the two operators are incompatible, then one cannot always simultaneously measure the values of the two physical observables. This is the core concept behind the *uncertainty principle*, which can be formulated in terms of an inequality for the fluctuation of the measurements for two operators Aand B. For a given operator A and quantum state  $\Psi$ , we define an operator

$$\Delta A = A - \langle A \rangle I := A - \langle \Psi | A | \Psi \rangle I \tag{84}$$

Thus, the expectation value of  $\Delta A$  is given by

$$\langle \Delta A \rangle = \langle A \rangle - \langle A \rangle = 0 \tag{85}$$

and its variance is

$$\left\langle \Delta A^2 \right\rangle = \left\langle (A - \left\langle \Psi | A | \Psi \right\rangle I)^2 \right\rangle = \left\langle A^2 \right\rangle - \left\langle A \right\rangle^2 \tag{86}$$

Now, if two operators A and B are compatible, and  $\Psi$  is one of their common eigenvectors, then

$$\left\langle \Delta A^2 \right\rangle = \left\langle \Delta B^2 \right\rangle = 0. \tag{87}$$

This means that there is no uncertainty in measuring the values of both A and B simultaneously. Since A is hermitian,  $\Delta A$  is also hermitian; together with the Cauchy–Schwarz inequality yields

$$\begin{split} |\langle \Delta A \Delta B \rangle|^2 &= |\langle \langle \psi | \Delta A \rangle (\Delta B | \psi \rangle)|^2 \\ &\leq \langle \psi | \Delta A (\Delta A)^* | \psi \rangle \langle \psi | (\Delta B)^* \Delta B | \psi \rangle \\ &= \langle \Delta A^2 \rangle \langle \Delta B^2 \rangle \end{split}$$
(88)

and note that

$$\Delta A \Delta B = \frac{1}{2} \left( \Delta A \Delta B + \Delta B \Delta A \right) + \frac{1}{2} \left( \Delta A \Delta B - \Delta B \Delta A \right) = \frac{1}{2} \{ \Delta A, \Delta B \} + \frac{1}{2} [\Delta A, \Delta B]$$
(89)

Since  $\{\Delta A, \Delta B\}$  is hermitian,  $\langle \{\Delta A, \Delta B\} \rangle \in \mathbb{R}$ . Similarly, we note that  $i[\Delta A, \Delta B]$  is hermitian, hence,  $\langle [\Delta A, \Delta B] \rangle$  is purely imaginary. This yields

$$\begin{split} |\langle \Delta A \Delta B \rangle|^2 &= \frac{1}{4} |\langle \{\Delta A, \Delta B\} \rangle + \langle [\Delta A, \Delta B] \rangle|^2 \\ &\geq \frac{1}{4} |\langle [\Delta A, \Delta B] \rangle|^2 = \frac{1}{4} |\langle [A, B] \rangle|^2 \end{split}$$
(90)

where we used that

$$\begin{aligned} [\Delta A, \Delta B] &= [A - \langle A \rangle I, B - \langle B \rangle I] \\ &= (A - \langle A \rangle I)(B - \langle B \rangle I) - (B - \langle B \rangle I)(A - \langle A \rangle I) \\ &= AB - \langle B \rangle A - \langle A \rangle B + \langle A \rangle \langle B \rangle - BA + \langle A \rangle B + \langle B \rangle A - \langle A \rangle \langle B \rangle \tag{91} \\ &= AB - BA \\ &= [A, B]. \end{aligned}$$

Hence

$$\langle \Delta A^2 \rangle \langle \Delta B^2 \rangle \ge \frac{1}{4} |\langle [A, B] \rangle|^2$$
 (92)

which is known as the uncertainty principle. It states that there is a lower bound for the product of the uncertainty of two operators  $\langle \Delta A^2 \rangle$  and  $\langle \Delta B^2 \rangle$  given by the expectation value of the commutator. Due to the uncertainty principle, one cannot obtain simultaneously precise measurements of, e.g.,  $S_x$  and  $S_z$ .

#### 3.1 Heisenberg Uncertainty Principle

The Heisenberg uncertainty principle is defined as the uncertainty of the position operator and momentum operator. First introduced in 1927 by German physicist Werner Heisenberg, the formal inequality relating the standard deviation of position and the standard deviation of momentum was derived by Earle Hesse Kennard later that year and by Hermann Weyl in 1928.

The explanation for "why" the position and momentum operators are of the presented form requires more physical background and motivation which we will not provide here.

In real space the position operator is defined as

$$|x\psi\rangle = |x\psi\rangle \quad \Leftrightarrow \quad (x\psi)(x) = x\psi(x)$$
(93)

for any  $|\psi\rangle \in \mathcal{H}$ . Note that for an eigenstate of the position operator, i.e.,

$$x|\psi\rangle = x_0|\psi\rangle,\tag{94}$$

this means that  $\Psi(x) = 0$  if  $x \neq x_0$ . This however contradicts that  $|\psi\rangle$  is normalizable. In fact the position operator does not have an eigenstate that is square integrable. Instead the eigen decomposition of the position operator is given by the Dirac  $\delta$ -distribution loosely defined as

$$\delta(x - x_0) = \begin{cases} \infty, & \text{if } x = x_0 \\ 0, & \text{else} \end{cases}$$
(95)

with

$$\int \delta(x - x_0) dx = 1. \tag{96}$$

Note that to fully grasp this, we require the theory of distributions. The momentum operator determines the momentum of a quantum mechanical particle and is defined as

$$p = -i\frac{d}{dx} \tag{97}$$

hence, when applied to a wavefunction  $\psi(x)\psi(x)$  this yields

$$p\psi(x) = -i\psi'(x). \tag{98}$$

We may now compute the commutator of the position operator and the momentum operator

$$[x, p]\psi(x) = (xp - px)\psi(x) = x(-i\psi'(x)) + i\psi(x) + ix\psi'(x) = i\psi(x)$$
(99)

hence

$$[x,p] = i \tag{100}$$

which is known as the canonical commutation relation. In particular, position and momentum cannot be simultaneously determined. A more quantitative version of this statement is given by the uncertainty principle

$$\sqrt{\langle \Delta x^2 \rangle} \sqrt{\langle \Delta p^2 \rangle} = \frac{1}{2} \tag{101}$$

#### 3.2 Quantum numbers

Recall from the hydrogen atom computation that we found the states  $\psi_{n,\ell,m}$ , where  $n \ge 1$ ,  $0 \le \ell \le n-1$ , and  $-\ell \le m_\ell \le \ell$ . We have now learned that an electron also has a spin value  $m_s = \{\pm \frac{1}{2}\}$ . This means that every energetic state of the electron in the hydrogen is described by these four numbers:

principal quantum number	$n \ge 1$
azimuthal quantum number	$0\leqslant \ell\leqslant n-1$
magnetic quantum number	$-\ell \leqslant m_\ell \leqslant \ell$
spin quantum number	$m_s = \{\pm \frac{1}{2}\}$

In fact, this characterization of electronic states in an atom generalizes to multi-electron systems atoms. However, there is a catch:

Pauli-exclusion principle [?]:

"In a poly-electron atom it is impossible for any two electrons to have the same two values of all four of their quantum numbers, which are: n, the principal quantum number;  $\ell$ , the azimuthal quantum number;  $m_{\ell}$ , the magnetic quantum number; and  $m_s$ , the spin quantum number."

Consequently, in multi-electron atoms the electrons occupy different electronic states. The "filling" of electronic states follows the Aufbau principle see Fig. 11.



Figure 11: (a) Visualization of the Aufbau principle. and (b) the energies of the different states.

The Aufbau principle relates to the first three quantum numbers, which are sufficient for the hydrogen atom. In multi-electron atoms the spin quantum number needs to be carefully considered. Here Hund's rule applies, which says that the orbitals of the subshell are each occupied singly with electrons of parallel spin before double occupation occurs. **Example 3.0.1.** Consider the molecule chlorine which has 17 electrons, following the Aufbau principle and Hund's rule we find the following configuration, see Table. 1.

Table 1: Orbital occupation for chlorine,

Therefore the electronic configuration is

$$1s^2 2s^2 2p^6 3s^2 3p^5 = [\text{Ne}] 3s^2 3p^5 \tag{102}$$

**Remark 3.0.1.** The electronic configuration of an atom can be determined quicker by looking at the periodic table, see Fig. 12.



Figure 12: Orbital blocks in periodic table

We then "walk from left to right and top to bottom" to determine the electronic configuration.

# 4 Combining two spin- $\frac{1}{2}$ particles

When we investigated a single spin particle, we could use the  $|\pm_z\rangle$  as a basis. Now what is the correct basis for two spin- $\frac{1}{2}$  particles? A natural extension to the  $|\pm_z\rangle$  basis is the tensor product basis, i.e.,

$$|+_{z}\rangle \otimes |+_{z}\rangle = |+_{z}, +_{z}\rangle \qquad |+_{z}\rangle \otimes |-_{z}\rangle = |+_{z}, -_{z}\rangle |-_{z}\rangle \otimes |+_{z}\rangle = |-_{z}, +_{z}\rangle \qquad |-_{z}\rangle \otimes |-_{z}\rangle = |-_{z}, -_{z}\rangle$$
(103)

in which the first element describes the state of particle one and the second element describes the state of particle two. More formally, we may consider two Hilbert spaces  $\mathcal{H}_A$  and  $\mathcal{H}_B$  of dimension  $N_A$  and  $N_B$  describing the individual particles, respectively, and denote their bases  $\{|\varphi_i^A\rangle\}_{i=1}^{N_A}$  and  $\{|\varphi_i^B\rangle\}_{i=1}^{N_B}$ . The tensor product space is then defined as

$$\mathcal{H}_A \otimes \mathcal{H}_B = \operatorname{Span}\left\{ |\varphi_i^A, \varphi_j^B \rangle \mid i \in \llbracket N_A \rrbracket, j \in \llbracket N_B \rrbracket \right\},$$
(104)

where

Here,  $\{|\varphi_i^A, \varphi_j^B\rangle\}$  form a new basis. Recall the tensor product of vectors

$$|+_{z}\rangle \otimes |-_{z}\rangle = \begin{bmatrix} 1\\0 \end{bmatrix} \otimes \begin{bmatrix} 0\\1 \end{bmatrix} = \begin{bmatrix} 0\\1\\0\\0 \end{bmatrix}$$
(105)

Since  $\mathcal{H}_A$  and  $\mathcal{H}_B$  are Hilbert spaces, their tensor product space inherits an inner product structure

$$\langle \varphi_i^A, \varphi_j^B | \varphi_{i'}^A, \varphi_{j'}^B \rangle_{A \otimes B} = \langle \varphi_i^A | \varphi_{i'}^A \rangle_A \langle \varphi_j^B | \varphi_{j'}^B \rangle_B \tag{106}$$

Therefore, if  $\{|\varphi_i^A\rangle\}_{i=1}^{N_A}$  and  $\{|\varphi_i^B\rangle\}_{i=1}^{N_B}$  are orthonormal, the tensor product basis inhertis the orthonormality, i.e.,

$$\left\langle \varphi_{i}^{A}, \varphi_{j}^{B} | \varphi_{i'}^{A}, \varphi_{j'}^{B} \right\rangle = \left\langle \varphi_{i}^{A} | \varphi_{i'}^{A} \right\rangle \left\langle \varphi_{j}^{B} | \varphi_{j'}^{B} \right\rangle = \delta_{i,i'} \delta_{j,j'} \tag{107}$$

Moreover,  $\mathcal{H}_A \otimes \mathcal{H}_B$  is a Hilbert space itself of dimension  $N_A \times N_B$ . Similarly, we may extend operators to tensor product spaces. Let A and B be two operators defined on  $\mathcal{H}_A$  and  $\mathcal{H}_B$ , respectively. Then

$$(A \otimes B)|\varphi_{i'}^A, \varphi_{j'}^B\rangle := |A\varphi_{i'}^A, B\varphi_{j'}^B\rangle.$$
(108)

Returning to our example of two spin- $\frac{1}{2}$  particles, we can then define the operators

$$S_z^{(1)} = S_z \otimes I$$
 and  $S_z^{(2)} = I \otimes S_z$  (109)

and the total spin operator in z-direction

$$S_z^{\text{tot}} = S_z^{(1)} + S_z^{(2)} \tag{110}$$

Similarly, we can define  $S_x^{\text{tot}}$  and  $S_y^{\text{tot}}$ , as the corresponding spin magnitude operator

$$(S^{\text{tot}})^2 = (S_x^{\text{tot}})^2 + (S_y^{\text{tot}})^2 + (S_z^{\text{tot}})^2$$
(111)

We observe that

$$S_{z}^{\text{tot}}|+_{z},+_{z}\rangle = S_{z} \otimes I|+_{z},+_{z}\rangle + I \otimes S_{z}|+_{z},+_{z}\rangle = \frac{1}{2}|+_{z},+_{z}\rangle + \frac{1}{2}|+_{z},+_{z}\rangle = |+_{z},+_{z}\rangle$$

$$S_{z}^{\text{tot}}|-_{z},-_{z}\rangle = S_{z} \otimes I|-_{z},-_{z}\rangle + I \otimes S_{z}|-_{z},-_{z}\rangle = -\frac{1}{2}|-_{z},-_{z}\rangle - \frac{1}{2}|-_{z},-_{z}\rangle = -|-_{z},-_{z}\rangle$$

$$S_{z}^{\text{tot}}|+_{z},-_{z}\rangle = S_{z} \otimes I|+_{z},-_{z}\rangle + I \otimes S_{z}|+_{z},-_{z}\rangle = \frac{1}{2}|+_{z},-_{z}\rangle - \frac{1}{2}|+_{z},-_{z}\rangle = |0\rangle$$

$$S_{z}^{\text{tot}}|-_{z},+_{z}\rangle = S_{z} \otimes I|-_{z},+_{z}\rangle + I \otimes S_{z}|-_{z},+_{z}\rangle = -\frac{1}{2}|-_{z},+_{z}\rangle + \frac{1}{2}|-_{z},+_{z}\rangle = |0\rangle$$

$$(112)$$

(112) Hence all basis vectors of the tensor product basis are eigenstates of  $S_z^{\text{tot}}$ . We now want to investigate if  $S_z^{\text{tot}}$  and the corresponding total spin square operator commute. To that end, we first note that

$$(S_z^{\text{tot}})^2 = (S_z \otimes I)^2 + (S_z \otimes I) (I \otimes S_z) + (I \otimes S_z) (S_z \otimes I) + (I \otimes S_z)^2$$
  
=  $\frac{1}{2}I \otimes I + 2S_z \otimes S_z$  (113)

which yields

$$(S^{\text{tot}})^2 = \frac{3}{2}I \otimes I + 2S_x \otimes S_x + 2S_y \otimes S_y + 2S_z \otimes S_z$$
(114)

and therefore

$$[(S^{\text{tot}})^2, S_z^{\text{tot}}] = \frac{3}{2} [I \otimes I, S_z^{\text{tot}}] + 2[S_x \otimes S_x, S_z^{\text{tot}}] + 2[S_y \otimes S_y, S_z^{\text{tot}}] + 2[S_z \otimes S_z, S_z^{\text{tot}}]$$
  
= 2[S<sub>x</sub> \otimes S<sub>x</sub>, S<sub>z</sub>^{\text{tot}}] + 2[S\_y \otimes S\_y, S\_z^{\text{tot}}] + 2[S\_z \otimes S\_z, S\_z^{\text{tot}}](115)

We then note that

$$[S_{\alpha} \otimes S_{\alpha}, S_{z}^{\text{tot}}] = [S_{\alpha} \otimes S_{\alpha}, S_{z}^{(1)}] + [S_{\alpha} \otimes S_{\alpha}, S_{z}^{(2)}]$$
  
= 
$$[S_{\alpha}, S_{z}] \otimes S_{\alpha} + S_{\alpha} \otimes [S_{\alpha}, S_{z}]$$
 (116)

Since the spin operators for single particles are cyclic we find

$$[S_x \otimes S_x, S_z^{\text{tot}}] = -iS_y \otimes S_x - iS_x \otimes S_y$$
  

$$[S_y \otimes S_y, S_z^{\text{tot}}] = iS_x \otimes S_y + iS_y \otimes S_x$$
  

$$[S_z \otimes S_z, S_z^{\text{tot}}] = 0$$
(117)

which yields that

$$[(S^{\text{tot}})^2, S_z^{\text{tot}}] = 0.$$
(118)

Therefore  $(S^{\text{tot}})^2$  and  $S_z^{\text{tot}}$  can be simultaneously diagonialized. More specifically,

$$(S^{\text{tot}})^2|+_z,+_z\rangle = 2|+_z,+_z\rangle$$
 and  $(S^{\text{tot}})^2|-_z,-_z\rangle = 2|-_z,-_z\rangle$  (119)

On the other hand, we note that

$$(S^{\text{tot}})^{2}|+_{z}, -_{z}\rangle = \frac{3}{2}I \otimes I|+_{z}, -_{z}\rangle + 2S_{x} \otimes S_{x}|+_{z}, -_{z}\rangle + 2S_{y} \otimes S_{y}|+_{z}, -_{z}\rangle + 2S_{z} \otimes S_{z}|+_{z}, -_{z}\rangle = \frac{3}{2}|+_{z}, -_{z}\rangle + \frac{1}{2}|-_{z}, +_{z}\rangle + \frac{1}{2}|-_{z}, +_{z}\rangle - \frac{1}{2}|+_{z}, -_{z}\rangle = |+_{z}, -_{z}\rangle + |-_{z}, +_{z}\rangle$$
(120)

and similar

$$(S^{\text{tot}})^2|_{-z}, +_z\rangle = |-_z, +_z\rangle + |+_z, -_z\rangle$$
(121)

Hence,  $|\pm_z, \mp_z\rangle$  are not eigenstates of  $(S^{\text{tot}})^2$ . Instead,

$$(S^{\text{tot}})^{2} \left(\frac{1}{\sqrt{2}}|+_{z},-_{z}\rangle + \frac{1}{\sqrt{2}}|-_{z},+_{z}\rangle\right)$$

$$= \frac{1}{\sqrt{2}}(S^{\text{tot}})^{2}|+_{z},-_{z}\rangle + \frac{1}{\sqrt{2}}(S^{\text{tot}})^{2}|-_{z},+_{z}\rangle$$

$$= \frac{1}{\sqrt{2}}(|+_{z},-_{z}\rangle + |-_{z},+_{z}\rangle) + \frac{1}{\sqrt{2}}(|-_{z},+_{z}\rangle + |+_{z},-_{z}\rangle)$$

$$= 2\left(\frac{1}{\sqrt{2}}|+_{z},-_{z}\rangle + \frac{1}{\sqrt{2}}|-_{z},+_{z}\rangle\right)$$
(122)

and similar

$$(S^{\text{tot}})^2 \left( \frac{1}{\sqrt{2}} |+_z, -_z\rangle - \frac{1}{\sqrt{2}} |-_z, +_z\rangle \right) = 0.$$
(123)

Thus, the operator  $(S^{\text{tot}})^2$  can be used to distinguish the 0-eigenspace of  $S_z^{\text{tot}}$ . So we can summarize the eigenstates as

$$\begin{array}{c|c} \text{State} & \text{Type} & (S^{\text{tot}})^2 & S_z^{\text{tot}} \\ \hline \frac{1}{\sqrt{2}} \left( |+_z, -_z\rangle - |-_z, +_z\rangle \right) & \text{singlet} & 0 & 0 \\ \hline |+_z, +_z\rangle & & 1 \\ \frac{1}{\sqrt{2}} \left( |+_z, -_z\rangle + |-_z, +_z\rangle \right) & \text{triplet} & 2 & 0 \\ |-_z, -_z\rangle & & -1 \end{array}$$

## 4.1 The hydrogen molecule ion $H_2^+$

We now consider the  ${\rm H}_2^+$  molecule consisting of two hydrogen atoms, but with one electron removed. The Hamiltonian then reads

$$H = \frac{1}{2}\Delta_{\mathbf{r}} - \frac{1}{\|\mathbf{r}\|} - \frac{1}{\|\mathbf{r} - \mathbf{R}\|}.$$
(124)

A crude approximation to the solution would be a linear combination of two 1s orbitals centered at 0 and  $\mathbf{R}$  respectively, i.e.,

$$\psi(\mathbf{r}) = c_1 \psi_{1,0,0}(\mathbf{r}) + c_2 \psi_{1,0,0}(\mathbf{r} - \mathbf{R}).$$
(125)

In this framework, we assume the Galerkin space

$$\mathcal{H} = \text{Span}\{\psi_{1,0,0}(\cdot), \psi_{1,0,0}(\cdot - \mathbf{R})\}$$
(126)

which yields the matrix eigenvalue problem

$$\begin{bmatrix} \varepsilon & -t \\ -t & \varepsilon \end{bmatrix} \mathbf{c} = E \begin{bmatrix} 1 & s \\ s & 1 \end{bmatrix} \mathbf{c}$$
(127)

where

$$\varepsilon = \int_{\mathbb{R}^3} \psi_{1,0,0}(\mathbf{r}) (H\psi_{1,0,0})(\mathbf{r}) d\mathbf{r}$$
  
$$-t = \int_{\mathbb{R}^3} \psi_{1,0,0}(\mathbf{r}) (H\psi_{1,0,0})(\mathbf{r} - \mathbf{R}) d\mathbf{r}$$
  
$$s = \int_{\mathbb{R}^3} \psi_{1,0,0}(\mathbf{r}) \psi_{1,0,0}(\mathbf{r} - \mathbf{R}) d\mathbf{r}$$
 (128)

Given that s, t > 0, the state eigenvalue and eigenfunctions are

$$E_g = \frac{\varepsilon - t}{1 + s}, \qquad \mathbf{c}_g = \frac{1}{\sqrt{2(1 + s)}} \begin{bmatrix} 1\\1 \end{bmatrix}$$
(129)

and

$$E_e = \frac{\varepsilon + t}{1 - s}, \qquad \mathbf{c}_g = \frac{1}{\sqrt{2(1 - s)}} \begin{bmatrix} 1\\ -1 \end{bmatrix}$$
(130)

# 5 Slater Determinant

## 6 Hartree–Fock Theory

$$\min_{|\psi\rangle \in \mathcal{H}_N} \frac{\langle \psi | H | \psi \rangle}{\langle \psi | \psi \rangle} \leq \min_{|\psi\rangle \in \mathcal{S}_N} \frac{\langle \psi | H | \psi \rangle}{\langle \psi | \psi \rangle}$$
(131)

where

$$\mathcal{S}_N = \left\{ \psi \in \mathcal{H}_N \mid \exists (\phi_1, ..., \phi_N) \in \mathcal{C}_N ; \psi = \frac{1}{\sqrt{N!}} \bigwedge_{i=1}^N \phi_i \right\},\$$

is the set of Slater determinants and the set of N-frames

$$\mathcal{C}_N = \left\{ (\phi_1, ..., \phi_N) \mid \phi_i \in H^1 \text{ and } \langle \phi_i, \phi_j \rangle_{L^2} = \delta_{i,j} \text{ for } 1 \leq j, i \leq N \right\}$$

is the *Stiefel manifold*. Note that the functional in Eq. (131) over  $S_N$  is invariant with respect to unitary transformations. To remedy this non-uniqueness we introduce the Grassmannian as suitable quotient space of  $C_N$ , i.e.,

$$\mathcal{G}_N = \mathcal{C}_N \big|_{\sim},\tag{132}$$

where the equivalence classes are given by

$$[\Phi]_{\sim} = \left\{ \tilde{\Phi} \in \mathcal{C}_N \mid \exists U \in \mathcal{U}(N) \ s.t. \ \Phi = U\tilde{\Phi} \right\}.$$
(133)

In other words, we identify all orthonormal N-frames spanning the same subspace. This yields the following minimization problem

$$E_{\rm HF} = \inf_{\Psi \in \mathcal{S}_N} \mathcal{J}(\Psi) = \inf_{\Phi \in \mathcal{G}_N} \varepsilon(\Phi), \qquad (134)$$

#### 6.1 One-body part

$$\langle \Psi | h | \Psi \rangle = \dots = \sum_{i=1}^{N} \int_{\mathbb{R}^{3} \times \left\{ \pm \frac{1}{2} \right\}} \frac{1}{2} |\nabla_{r} \phi_{i}(\mathbf{x})|^{2} + V_{\text{ext}}(\mathbf{r}) |\phi_{i}(\mathbf{x})|^{2} d\mathbf{x}$$
(135)

#### 6.2 Two-body part

Recall that the two-body interaction term is given by

$$H_I := \sum_{i < j} g(i, j) := \sum_{i < j} \frac{1}{\|\mathbf{r}_i - \mathbf{r}_j\|}$$
(136)

Again, let  $\Psi = \Phi[j_1, ..., j_N]$  be a Slater determinant. Then

$$\langle \Psi | H_I | \Psi \rangle = \sum_{i < j} \left\langle \Psi \left| \frac{1}{\|\mathbf{r}_i - \mathbf{r}_j\|} \right| \Psi \right\rangle$$

$$= \binom{N}{2} \left\langle \Psi \left| \frac{1}{\|\mathbf{r}_1 - \mathbf{r}_2\|} \right| \Psi \right\rangle$$

$$(137)$$

We note that

$$\begin{split} \left\langle \Psi \middle| \frac{1}{\|\mathbf{r}_{1} - \mathbf{r}_{2}\|} \middle| \Psi \right\rangle &= \int_{\left(\mathbb{R}^{3} \times \{\pm \frac{1}{2}\}\right)^{\otimes N}} \Psi^{*}(\mathbf{x}_{1}, ..., \mathbf{x}_{N}) \frac{1}{\|r_{1} - r_{2}\|} \Psi(\mathbf{x}_{1}, ..., \mathbf{x}_{N}) d\mathbf{x}_{1} ... d\mathbf{x}_{N} \\ &= \frac{1}{N!} \sum_{\pi, \pi' \in S_{N}} \operatorname{sgn}(\pi) \operatorname{sgn}(\pi') \int_{\left(\mathbb{R}^{3} \times \{\pm \frac{1}{2}\}\right)^{\otimes 2}} \phi^{*}_{\pi(1)}(\mathbf{x}_{1}) \phi^{*}_{\pi(2)}(\mathbf{x}_{2}) \frac{1}{\|r_{1} - r_{2}\|} \phi_{\pi'(1)}(\mathbf{x}_{1}) \phi_{\pi'(2)}(\mathbf{x}_{2}) \\ &\times \prod_{i=3}^{N} \int_{\mathbb{R}^{3} \times \{\pm \frac{1}{2}\}} \phi^{*}_{\pi(i)}(\mathbf{x}) \phi_{\pi'(i)}(\mathbf{x}) d\mathbf{x} \end{split}$$

Note that

$$\prod_{i=3}^{N} \int_{\mathbb{R}^{3} \times \{\pm \frac{1}{2}\}} \phi_{\pi(i)}^{*}(\mathbf{x}) \phi_{\pi'(i)}(\mathbf{x}) d\mathbf{x} = \prod_{i=3}^{N} \delta_{\pi(i),\pi'(i)}$$
(138)

which implies that either

 $\pi(1) = \pi'(1) = i_1$  and  $\pi(2) = \pi'(2) = i_2$  (139)

or

$$\pi(1) = \pi'(2) = i_1$$
 and  $\pi(2) = \pi'(1) = i_2$  (140)

where  $i_1, i_2 \in \{1, ..., N\}$ . In case of Eq. (139) we have that

$$\operatorname{sgn}(\pi) = \operatorname{sgn}(\pi') \implies \operatorname{sgn}(\pi)\operatorname{sgn}(\pi') = 1$$

Whereas in case of Eq. (140) we have that

$$\operatorname{sgn}(\pi) = -\operatorname{sgn}(\pi') \implies \operatorname{sgn}(\pi)\operatorname{sgn}(\pi') = -1$$

Thus

$$\left\langle \Psi \middle| \frac{1}{\|r_1 - r_2\|} \middle| \Psi \right\rangle = \frac{(N-2)!}{N!} \sum_{i=1}^N \sum_{i \neq j=1}^N \left( \int_{\left(\mathbb{R}^3 \times \left\{ \pm \frac{1}{2} \right\} \right)^{\otimes 2}} \phi_i^*(\mathbf{x}_1) \phi_j^*(\mathbf{x}_2) \frac{1}{\|r_1 - r_2\|} \phi_i(\mathbf{x}_1) \phi_j(\mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 \right) \\ - \int_{\left(\mathbb{R}^3 \times \left\{ \pm \frac{1}{2} \right\} \right)^{\otimes 2}} \phi_i^*(\mathbf{x}_1) \phi_j^*(\mathbf{x}_2) \frac{1}{\|r_1 - r_2\|} \phi_j(\mathbf{x}_1) \phi_i(\mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 \right) \\ := \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{i \neq j=1}^N \langle ij||ij\rangle - \langle ij||ji\rangle$$

Hence

$$\langle \Psi | H_I | \Psi \rangle = \frac{1}{2} \sum_{i,j=1}^{N} \langle ij | | ij \rangle - \langle ij | | ji \rangle$$
(141)

Putting it all together

$$\begin{aligned} \mathcal{E}_{\mathrm{HF}}(\{\phi_i\}_{i=1}^N) &= \sum_{i=1}^N \int_{\mathbb{R}^3 \times \{\pm \frac{1}{2}\}} \frac{1}{2} |\nabla_r \phi_i(x)|^2 + V_{\mathrm{ext}}(r) |\phi_i(x)|^2 dx \\ &+ \frac{1}{2} \sum_{i,j} \int_{\left(\mathbb{R}^3 \times \{\pm \frac{1}{2}\}\right)^{\otimes 2}} \frac{|\phi_i(x)|^2 |\phi_j(x')|^2}{\|r - r'\|} dx dx' \\ &- \frac{1}{2} \sum_{i,j} \int_{\left(\mathbb{R}^3 \times \{\pm \frac{1}{2}\}\right)^{\otimes 2}} \frac{\phi_i^*(x) \phi_j^*(x') \phi_j(x) \phi_i(x')}{\|r - r'\|} dx dx' \\ &= \sum_{i=1}^N \langle i||i\rangle + \frac{1}{2} \sum_{i,j} \langle ij||ij\rangle - \langle ij||ji\rangle \end{aligned}$$

How do we optimize this energy function?

We assume a given set of spin-orbitals  $\{\chi_i\}_{i=1}^K$  where  $N \leq K$ . Note that in this context,  $\chi_i$  may be inspired by atomic orbital functions are are not necessarily orthogonal, i.e.,

$$S_{i,j} := \langle \chi_i | \chi_j \rangle \tag{142}$$

and  $\mathbf{S} \neq \mathbf{I}_{K}$ . We then make the ansatz

$$\phi_i = \sum_{j=1}^K \mathbf{C}_{j,i} \chi_j \tag{143}$$

defining  $\mathbf{C} \in \mathbb{C}^{K \times N}$ . Applying this to the kinetic energy part we find

$$\frac{1}{2} \sum_{i=1}^{N} \int_{\mathbb{R}^{3} \times \{\pm \frac{1}{2}\}} (\nabla_{r} \phi_{i}(\mathbf{x}))^{*} \nabla_{r} \phi_{i}(\mathbf{x}) d\mathbf{x}$$

$$= \frac{1}{2} \sum_{i=1}^{N} \sum_{j,k=1}^{K} \mathbf{C}_{j,i}^{*} \mathbf{C}_{k,i} \int_{\mathbb{R}^{3} \times \{\pm \frac{1}{2}\}} (\nabla_{r} \chi_{j}(\mathbf{x}))^{*} \nabla_{r} \chi_{k}(\mathbf{x}) d\mathbf{x}$$

$$= \sum_{i=1}^{N} \frac{1}{2} \sum_{j,k=1}^{K} \mathbf{C}_{j,i}^{*} \mathbf{C}_{k,i} \langle \nabla \chi_{j}, \nabla \chi_{k} \rangle$$
(144)

Overall, this yields

$$\mathcal{E}_{\mathrm{HF}}(C) = \sum_{i=1}^{N} \frac{1}{2} \sum_{j,k=1}^{K} \mathbf{C}_{j,i}^{*} \mathbf{C}_{k,i} \langle \nabla \chi_{j}, \nabla \chi_{k} \rangle$$
$$- \sum_{i=1}^{N} \sum_{j,k=1}^{K} \mathbf{C}_{j,i}^{*} \mathbf{C}_{k,i} \langle \chi_{j}, V_{\mathrm{ext}} \chi_{k} \rangle$$
$$+ \frac{1}{2} \sum_{i,j=1}^{N} \sum_{k,\ell,m,n=1}^{K} \mathbf{C}_{k,i}^{*} \mathbf{C}_{\ell,j}^{*} \mathbf{C}_{m,i} \mathbf{C}_{n,j} \langle \chi_{k} \chi_{\ell} || \chi_{m} \chi_{n} \rangle$$
$$- \frac{1}{2} \sum_{i,j=1}^{N} \sum_{k,\ell,m,n=1}^{K} \mathbf{C}_{k,i}^{*} \mathbf{C}_{\ell,j}^{*} \mathbf{C}_{m,j} \mathbf{C}_{n,i} \langle \chi_{k} \chi_{\ell} || \chi_{m} \chi_{n} \rangle$$

Introducing the tensors

$$\begin{split} \mathbf{T}_{j,k} &= \frac{1}{2} \langle \nabla \phi_j | \nabla \phi_k \rangle \\ \mathbf{V}_{j,k} &= \langle \phi_j | V | \phi_k \rangle \\ \mathbf{h} &= \mathbf{T} - \mathbf{V} \\ \mathbf{V}_{i,j,k,l} &= \langle \chi_i \chi_j | | \chi_k \chi_l \rangle \end{split}$$

we find

$$\begin{aligned} \mathcal{E}_{\mathrm{HF}}(\mathbf{C}) &= \mathrm{Tr}\left(\mathbf{C}^{\dagger}\mathbf{h}\mathbf{C}\right) \\ &+ \frac{1}{2}\sum_{i,j=1}^{N}\sum_{k,\ell,m,n=1}^{K}\mathbf{C}_{k,i}^{*}\mathbf{C}_{\ell,j}^{*}\mathbf{C}_{m,i}\mathbf{C}_{n,j}\mathbf{V}_{k,\ell,m,n} \\ &- \frac{1}{2}\sum_{i,j=1}^{N}\sum_{k,\ell,m,n=1}^{K}\mathbf{C}_{k,i}^{*}\mathbf{C}_{\ell,j}^{*}\mathbf{C}_{m,j}\mathbf{C}_{n,i}\mathbf{V}_{k,\ell,m,n} \end{aligned}$$

We define the one-particle reduced density matrix (1-RDM)  $\mathbf{D} \in \mathbb{C}^{K \times K}$  element-wise as

$$\mathbf{D}_{i,j} = [\mathbf{C}\mathbf{C}^{\dagger}]_{i,j} = \sum_{k=1}^{N} \mathbf{C}_{i,k}\mathbf{C}_{j,k}^{*}.$$
(145)

which yields

$$\mathcal{E}_{\mathrm{HF}}(\mathbf{D}) = \mathrm{Tr} (\mathbf{h} \mathbf{D}) + \frac{1}{2} \sum_{k,\ell,m,n=1}^{K} \mathbf{D}_{m,k} \mathbf{D}_{n,\ell} \mathbf{V}_{k,\ell,m,n} - \frac{1}{2} \sum_{k,\ell,m,n=1}^{K} \mathbf{D}_{n,k} \mathbf{D}_{m,\ell} \mathbf{V}_{k,\ell,m,n} =: \mathrm{Tr} (\mathbf{h} \mathbf{D}) + \frac{1}{2} \mathrm{Tr} (\mathbf{J}(\mathbf{D}) \mathbf{D}) - \frac{1}{2} \mathrm{Tr} (\mathbf{K}(\mathbf{D}) \mathbf{D})$$
(146)

where

$$\mathbf{J}(\mathbf{D})_{k,m} = \sum_{\ell,n=1}^{K} \mathbf{D}_{n,\ell} \mathbf{V}_{k,\ell,m,n}$$
(147)

is the Hartree or Coulomb operator leading to

$$\operatorname{Tr}\left(\mathbf{J}(\mathbf{D})\mathbf{D}\right) = \sum_{k,m=1}^{K} \mathbf{J}(\mathbf{D})_{k,m} \mathbf{D}_{m,k} = \sum_{k,\ell,m,n=1}^{K} \mathbf{D}_{n,\ell} \mathbf{V}_{k,\ell,m,n} \mathbf{D}_{m,k}$$
(148)

and

$$\mathbf{K}(\mathbf{D})_{k,n} = \sum_{\ell,m=1}^{K} \mathbf{D}_{m,\ell} \mathbf{V}_{k,\ell,m,n}$$
(149)

is the Fock or exchange operator leading to

$$\operatorname{Tr}\left(\mathbf{K}(\mathbf{D})\mathbf{D}\right) = \sum_{k,n=1}^{K} \mathbf{K}(\mathbf{D})_{k,n} \mathbf{D}_{n,k} = \sum_{k,\ell,m,n=1}^{K} \mathbf{D}_{m,\ell} \mathbf{V}_{k,\ell,m,n} \mathbf{D}_{n,k}$$
(150)

Given the compact form of the energy functional, we will now take a closer look at C or equivalently D. We first note that we have used that

$$\delta_{i,j} = \langle \phi_i | \phi_j \rangle = \sum_{k,\ell} C_{k,i}^* \langle \chi_k | \chi_\ell \rangle C_{\ell,j}$$
(151)

which yields the condition that

$$\mathbf{C}^{\dagger}\mathbf{S}\mathbf{C} = \mathbf{I}_N \tag{152}$$

Hence we can write the Hartree–Fock method as a constraint minimization, namely,

$$\min_{\mathbf{C}\in\mathbb{C}^{K\times N}} \quad \mathcal{E}_{\mathrm{HF}}(\mathbf{C})$$
subject to  $\mathbf{C}^{\dagger}\mathbf{S}\mathbf{C} = \mathbf{I}_{N}$ 
(153)

This yields the Lagrangian

$$\mathcal{L}[\mathbf{C}, \Lambda] = \mathcal{E}_{\mathrm{HF}}(\mathbf{C}) - \mathrm{Tr}[\mathbf{\Lambda}(\mathbf{I} - \mathbf{C}^{\dagger}\mathbf{S}\mathbf{C})]$$
(154)

We find that

$$0 \stackrel{(!)}{=} \nabla_{\mathbf{C}} \mathcal{L}[\mathbf{C}, \Lambda]$$
  
=  $2\mathbf{h}\mathbf{C} + \frac{1}{2}4\mathbf{J}(\mathbf{C}\mathbf{C}^{\dagger})\mathbf{C} - \frac{1}{2}4\mathbf{K}(\mathbf{C}\mathbf{C}^{\dagger})\mathbf{C} + 2\mathbf{S}\mathbf{C}\Lambda$  (155)  
 $\Leftrightarrow \mathbf{F}(\mathbf{C})\mathbf{C} := (\mathbf{h} + \mathbf{J}(\mathbf{C}\mathbf{C}^{\dagger}) - \mathbf{K}(\mathbf{C}\mathbf{C}^{\dagger}))\mathbf{C} = \mathbf{S}\mathbf{C}\Lambda$ 

where  $F(\mathbf{C})$  is the Fock matrix constructed from the  $\mathbf{C} \in \mathbb{C}^{K \times N}$ . Diagonalizing  $\Lambda$ , i.e.,

$$\mathbf{\Lambda} = \mathbf{U}\mathcal{E}\mathbf{U}^* \tag{156}$$

yields the non-linear generalized eigenvalue problem (GEVP)

$$F(CU)CU = SCU\mathcal{E}$$
  

$$\Leftrightarrow F(C)C = SC\mathcal{E}$$
(157)

This is known as the Roothan Hall equations, and the self-consistent field (SCF) iteration method solves this problem self-consistently, i.e.,

$$\mathbf{C}^{(k)} \to \mathbf{F}(\mathbf{C}^{(k)}) \to \mathbf{C}^{(k+1)} \tag{158}$$

until convergence is reached. Note that we here used a certain notational ambiguity, namely, solving the GEVP yields a matrix  $\mathbf{C} \in \mathbb{C}^{K \times K}$ . These are *all* molecular orbitals. Those corresponding the lowest N eigenvalues  $\varepsilon$  are called occupied molecular orbitals and are used to generate the Fock matrix for the next iteration which is consistent with the Aufbau principle. The remaining orbitals are the virtual molecular orbitals.

Having the 1-RDM as the central object, we make the following observations: First, we see that  $\mathbf{D}$  is hermitian since

$$\mathbf{D}^{\dagger} = (\mathbf{C}\mathbf{C}^{\dagger})^{\dagger} = \mathbf{C}\mathbf{C}^{\dagger} = \mathbf{D}$$
(159)

Second, we note that

$$\operatorname{Tr}[\mathbf{SD}] = \operatorname{Tr}[\mathbf{C}^{\dagger}\mathbf{SC}] = \operatorname{Tr}[\mathbf{I}_N] = N$$
(160)

Third, we find

$$\mathbf{DSD} = \mathbf{CC}^{\dagger}\mathbf{SCC}^{\dagger} = \mathbf{CC}^{\dagger} = \mathbf{D}$$
(161)

$$\begin{array}{ll} \min_{\mathbf{D} \in \mathbb{C}^{K \times K}} & \mathcal{E}_{\mathrm{HF}}(\mathbf{D}) \\ \text{subject to} & \mathbf{D} \in \mathcal{D} \end{array} \tag{162}$$

where

$$\mathcal{D} = \{ \mathbf{D} \in \mathbb{C}^{K \times K} \mid \mathbf{DSD} = \mathbf{D}, \ \mathrm{Tr}[\mathbf{SD}] = N, \ \mathbf{D}^{\dagger} = \mathbf{D} \}$$
(163)

Note that  $\mathbf{0} < \mathbf{S}$ , i.e.,  $\mathbf{S}$  is positive definite and when the orbitals  $\{\chi\}_{i=1}^{K}$  are orthonormalized, e.g.,

$$\tilde{\chi}_i = \sum_{j=1}^{K} [\mathbf{S}^{-\frac{1}{2}}]_{i,j} \chi_j$$
(164)

then

$$\langle \tilde{\chi}_{i} | \tilde{\chi}_{j} \rangle = \sum_{k,\ell=1}^{K} [\mathbf{S}^{-\frac{1}{2}}]_{i,k} \langle \chi_{k} | \chi_{\ell} \rangle [\mathbf{S}^{-\frac{1}{2}}]_{j,\ell} = [\mathbf{S}^{-\frac{1}{2}} \mathbf{S} \mathbf{S}^{-\frac{1}{2}}]_{i,j} = \delta_{i,j},$$
(165)

we have

$$\mathcal{D} = \{ \mathbf{D} \in \mathbb{C}^{K \times K} \mid \mathbf{D}^2 = \mathbf{D}, \mathrm{Tr}[\mathbf{D}] = N, \mathbf{D}^{\dagger} = \mathbf{D} \}$$
(166)

This set is equivalent to the Grasssmann manifold. Therefore we may also directly minimize the energy functional  $\mathcal{E}_{\mathrm{HF}}(\mathbf{D})$  over the Grassmann manifold. Note that this argument may also be made for  $\mathcal{E}_{\mathrm{HF}}(\mathbf{C})$  via  $\mathbf{C} \in \mathbb{C}^{K \times N}$ . For obvious reasons, methods that follow this optimization procedures are called direct minimization approaches to the Hartree–Fock problem.

#### 6.3 Spin symmetries (Fokotume)

Up to this point, we assumed that  $\{\chi\}_{i=1}^K \subseteq H^1(\mathbb{R}^3 \times \{\pm \frac{1}{2}\})$  is a very general set of functions. However, this approach is more general than it needs to be. Recall that the Hamiltonian does not explicitly depend on the spin degree of freedom. In particular the ansatz

$$\chi_i(\mathbf{x}) = \xi_i(\mathbf{r})m_i(s) \tag{167}$$

where  $\xi$  are the spatial orbitals and m are the spinors, is justified. The different spin symmetry approaches then corresponds to different ansätze in the LCAO formalism, namely,

Spin unrestricted HF

$$\phi_{i,s} = \sum_{p=1}^{K/2} C_{p,i}^{(s)} \xi_i \otimes m_s$$
(168)

here  $\mathbf{C} \in \mathbb{C}^{K \times N}$  is block diagonal, with not necessarily equal blocks.

Spin restricted HF

$$\phi_{i,s} = \sum_{p=1}^{K/2} C_{p,i} \xi_i \otimes m_s \tag{169}$$

here  $\mathbf{C} \in \mathbb{C}^{K \times N}$  is block diagonal, with equal blocks. Spin generalized HF

$$\phi_i = \sum_{p=1}^{K/2} \sum_{s=1}^2 C_{p,i} \xi_i \otimes m_s$$
(170)

here  $\mathbf{C} \in \mathbb{C}^{K \times N}$  is block diagonal, with not necessarily equal blocks.

Note that these symmetries have effect the optimization functional. In case of spin restricted Hartree–Fock we find that the Fock matrix is given by

$$F(\mathbf{C}) = \mathbf{h} + \mathbf{J}(\mathbf{C}\mathbf{C}^{\dagger}) - \frac{1}{2}\mathbf{K}(\mathbf{C}\mathbf{C}^{\dagger})$$
(171)

where  $\mathbf{C} \in \mathbb{C}^{k/2 \times N/2}$  is only one of the diagonal blocks, and the ERI using in the Coulomb and exchange term are coming from K/2 spatial orbitals. The corresponding RHF energy is then given by

$$E_{RHF} = \frac{1}{2} \operatorname{Tr}[\mathbf{D}(\mathbf{h} + \mathbf{F}(\mathbf{C}))]$$
(172)

where  $\mathbf{D} = \mathbf{C}\mathbf{C}^{\dagger}$ .

### 7 Second Quantization

#### 7.1 The Fermionic Fock space

Starting point are the atomic spin orbitals  $\{\phi_i\}$  which are sufficiently smooth functions defined on  $\mathbb{R}^3 \times \{\pm 1/2\}$ . The set of atomic orbitals is denoted  $\mathcal{B}$  and spans the finite-dimensional space h. In our case,  $|\mathcal{B}| = n \gg d$ . In physics and chemistry parlance, h is commonly referred to as the single-particle Hilbert space.

The *M*-particle (Hilbert) space, denoted  $\mathcal{H}^{(M)}$ , is then the *M*-th exterior power of *h* equipped with the natural orthonormal basis, denoted  $\mathfrak{B}^{(M)}$ , coming from wedge-products of elements of  $\mathcal{B}$  (these are the Slater determinants). We define the fermionic Fock space as the Grassmann algebra on  $h = \mathcal{H}^{(1)}$ :

$$\mathcal{F} = \bigoplus_{M=0}^{n} \mathcal{H}^{(M)}.$$
(173)

All vector spaces defined above are real vector spaces. Recall that by Pauli's exclusion principle, atomic spin orbitals can either be occupied or unoccupied, hence, any element in  $\mathcal{F}$  can be expressed by means of an occupation vector, i.e., a K-tuple of zeros and ones.

Example 7.0.1. We may write an occupation vector using Dirac notation: Consider

$$|s_1, \dots, s_n\rangle = \frac{1}{\sqrt{M!}}\phi_1^{s_1} \wedge \phi_2^{s_2} \wedge \dots \wedge \phi_n^{s_n}$$
(174)

where  $M = \sum_{i} s_i \leq n$  and  $s_i \in \{0, 1\}$  for all i = 1, ..., n. A general element in  $\mathcal{F}$ , is then given as

$$|\Psi\rangle = \sum_{s_1,...,s_n \in \{0,1\}} \Psi(s_1,...,s_n) |s_1,...,s_n\rangle$$
(175)

where  $\Psi(s_1, ..., s_n) \in \mathbb{C}$ .

The central objects in the second quantization are the fermionic creation and annihilation operators, i.e.,

$$a_{p}^{\dagger}: \mathcal{F} \to \mathcal{F} \; ; \; |s_{1}, ..., s_{n}\rangle \mapsto (-1)^{\sigma(p)}(1-s_{p})|s_{1}, ...s_{p-1}, 1-s_{p}, s_{p+1}, ..., s_{n}\rangle$$

$$a_{p}: \mathcal{F} \to \mathcal{F} \; ; \; |s_{1}, ..., s_{n}\rangle \mapsto (-1)^{\sigma(p)}s_{p}|s_{1}, ...s_{p-1}, 1-s_{p}, s_{p+1}, ..., s_{n}\rangle$$
(176)

where  $\sigma(p) = \sum_{q=1}^{p-1} s_q$ . As the name suggests, these operators create or annihilate a "particle" in the *pth* atomic spin-orbital.

#### 7.2 The Jordan-Wigner transformation

One particularly useful way of thinking of occupation vectors is by means of tensor products. To that end, we denote

$$\begin{pmatrix} 1\\ 0 \end{pmatrix} \equiv \text{unoccupied} \quad \text{and} \quad \begin{pmatrix} 0\\ 1 \end{pmatrix} \equiv \text{occupied}$$

Note that this is an arbitrary choice, but this is the convention used in the community. Then, we may identify

$$|s_1, \dots, s_n\rangle = \binom{1-s_1}{s_1} \otimes \dots \otimes \binom{1-s_n}{s_n}.$$
(177)

Example 7.0.2. Let n = 2, then

$$|01\rangle = \begin{pmatrix} 1\\0 \end{pmatrix} \otimes \begin{pmatrix} 0\\1 \end{pmatrix} = \begin{pmatrix} 0\\1\\0\\0 \end{pmatrix}$$
(178)

This perspective is particularly useful since it allows us to derive a discretization of the Hamiltonian in terms of exponentially large but extremely sparse matrices. We begin by defining the matrix

$$a = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

and note that

$$\begin{pmatrix} 1\\0 \end{pmatrix} = a \begin{pmatrix} 0\\1 \end{pmatrix}$$
 and  $\begin{pmatrix} 0\\1 \end{pmatrix} = a^{\dagger} \begin{pmatrix} 1\\0 \end{pmatrix}$ 

In other words, a annihilates and  $a^{\dagger}$  creates a particle. Moreover, we define the matrix

$$\sigma_z = \begin{pmatrix} 1 & 0\\ 0 & -1 \end{pmatrix}$$

which is also known as the Pauli-z matrix. Note that

$$(-1)\begin{pmatrix} 0\\1 \end{pmatrix} = \sigma_z \begin{pmatrix} 0\\1 \end{pmatrix} \text{ and } \begin{pmatrix} 1\\0 \end{pmatrix} = \sigma_z \begin{pmatrix} 1\\0 \end{pmatrix}$$

In other words, the Pauli-z matrix multiplies by the factor -1 if the orbital is occupied, and by 1 if the state is unoccupied. We can therefore use the Pauli-z matrix to get the parity in Eq. (176) correct.

Then, the fermionic creation and annihilation operators are sparse matrices of the form

$$a_p^{\dagger} = \underbrace{\sigma_z \otimes \dots \otimes \sigma_z}_{p-1 \text{ times}} \otimes a^{\dagger} \otimes \underbrace{I \otimes \dots \otimes I}_{n-p-1 \text{ times}} \quad \text{and} \quad a_p = \underbrace{\sigma_z \otimes \dots \otimes \sigma_z}_{p-1 \text{ times}} \otimes a \otimes \underbrace{I \otimes \dots \otimes I}_{n-p-1 \text{ times}} \tag{179}$$

where

$$I = \begin{pmatrix} 1 & 0\\ 0 & 1 \end{pmatrix} \tag{180}$$

**Example 7.0.3.** *Let* n = 3*, then* 

Clearly, one does not want to store the creation and annihilation matrices explicitly, as they scale as  $\mathcal{O}(2^n)$ , yet their very structured sparsity pattern can be leveraged.

Before venturing further into the Hamiltonian description, we want to highlight the most important property of the fermionic creation and annihilation matrices. **Theorem 7.1.** the fermionic creation and annihilation operators obey the canonical anticommunication relation (CAR):

$$[a_p, a_q]_+ = [a_p^{\dagger}, a_q^{\dagger}]_+ = 0 \quad and \quad [a_p, a_q^{\dagger}]_+ = \delta_{p,q}.$$
(182)

*Proof.* Let p < q. Then

$$a_{p}a_{q} = I \otimes \dots \otimes I \otimes \underbrace{a\sigma_{z}}_{=-a} \otimes \sigma_{z} \otimes \dots \otimes \sigma_{z} \otimes a \otimes I \otimes \dots \otimes I$$
$$= -I \otimes \dots \otimes I \otimes \underbrace{\sigma_{z}a}_{=a} \otimes \sigma_{z} \otimes \dots \otimes \sigma_{z} \otimes a \otimes I \otimes \dots \otimes I$$
$$= -a_{q}a_{p}$$
(183)

For p = q we find  $a^2 = 0$  tensored with identities. Similarly, we find that  $[a_p^{\dagger}, a_q^{\dagger}]_+ = 0$ .

For Slater determinants there holds

$$|s_1,...,s_n\rangle = \prod_{i=1^n} (a_i^{\dagger})^{s_i} |-\rangle = a_k^{\dagger}...a_1^{\dagger} |-\rangle$$

#### 7.3 Hamiltonian in second quantization

Given the fermionic creation and annihilation operators, the Coulomb Hamiltonian can be reformulated in second quantized form it reads

$$H = \sum_{p,q=1}^{n} h_{p,q} a_p^{\dagger} a_q + \frac{1}{2} \sum_{p,q,r,s=1}^{n} v_{p,q,r,s} a_p^{\dagger} a_r^{\dagger} a_s a_q,$$
(184)

where

$$h_{p,q} = \int_X \phi_p^*(x_1) \left( -\frac{\Delta}{2} - \sum_j \frac{Z_j}{|r_1 - R_j|} \right) \phi_q(x_1) dx_1$$
(185)

and

$$v_{p,q,r,s} = \int_{X \times X} \frac{\phi_p^*(x_1)\phi_q(x_1)\phi_r^*(x_2)\phi_s(x_2)}{|r_1 - r_2|} dx_1 dx_2,$$
(186)

An important observation is that although the matrix  $H \in \mathbb{C}^{2^n \times 2^n}$ , the number of coefficients scales much lower, namely,  $h = [h_{p,q}] \in \mathbb{C}^{n \times n}$  and  $v = [v_{p,q,r,s}] \in \mathbb{C}^{n \times n \times n \times n}$ . Moreover, we note that

 $h = h^{\dagger}$ 

or in the case of real-valued atomic spin orbitals

 $h = h^{\top}$ 

and that

$$v_{p,q,r,s} = v_{r,s,p,q} = v_{q,p,s,r}^* = v_{s,r,q,p}^*$$

or in the case of real-valued atomic spin orbitals

$$v_{p,q,r,s} = v_{r,s,p,q} = v_{q,p,s,r} = v_{s,r,q,p} = v_{q,p,r,s} = v_{s,r,p,q} = v_{p,q,s,r} = v_{r,s,q,p}$$

# 8 Contractions

In quantum mechanics, the eigenstate is a nice quantity to have, but of interest are  $expectation\ values$ 

$$\langle \Psi, \mathcal{O}\Psi \rangle$$
 (187)

since they correspond to measurable quantities, called observables. In the case of the Hamiltonian, the observable is the system's energy. Remember that

$$[a_p, a_q]_+ = [a_p^{\dagger}, a_q^{\dagger}]_+ = 0 \text{ and } [a_p^{\dagger}, a_q]_+ = \delta_{p,q}.$$
 (188)

Then

$$\langle ij|\sum_{p,q}h_{p,q}a_p^{\dagger}a_q|kl\rangle = \sum_{p,q}h_{p,q}\langle -|a_ja_ia_p^{\dagger}a_qa_l^{\dagger}a_k^{\dagger}|-\rangle$$
(189)

and

$$\langle -|a_{j}\overline{a}_{i}a_{p}^{\dagger}a_{q}a_{l}^{\dagger}a_{k}^{\dagger}|-\rangle = (-1)\langle -|\overline{a}_{j}a_{p}^{\dagger}a_{i}a_{q}a_{l}^{\dagger}a_{k}^{\dagger}|-\rangle + \delta_{p,i}\langle -|a_{j}a_{q}a_{q}^{\dagger}a_{k}^{\dagger}|-\rangle$$

$$= (-1)^{2} \langle -|a_{p}^{\dagger}a_{j}a_{q}a_{q}a_{l}^{\dagger}a_{k}^{\dagger}|-\rangle + (-1)\delta_{p,j}\langle -|a_{i}a_{q}a_{q}^{\dagger}a_{k}^{\dagger}|-\rangle$$

$$+ (-1)\delta_{p,i}\langle -|\overline{a}_{j}a_{l}^{\dagger}a_{q}a_{k}^{\dagger}|-\rangle + \delta_{p,i}\delta_{l,q}\langle -|\overline{a}_{j}a_{k}^{\dagger}|-\rangle$$

$$= (-1)^{2}\delta_{p,j}\langle -|\overline{a}_{i}a_{l}^{\dagger}a_{q}a_{k}^{\dagger}|-\rangle + (-1)\delta_{p,i}\delta_{l,j}\langle -|\overline{a}_{i}a_{k}^{\dagger}|-\rangle$$

$$+ (-1)^{2}\delta_{p,i}\langle -|a_{k}^{\dagger}a_{j}a_{q}a_{k}^{\dagger}|-\rangle + (-1)\delta_{p,i}\delta_{l,j}\langle -|\overline{a}_{q}a_{k}^{\dagger}|-\rangle$$

$$+ (-1)\delta_{p,i}\delta_{l,q}\langle -|a_{k}^{\dagger}a_{j}|-\rangle + (-1)^{2}\delta_{p,j}\delta_{i,l}\langle -|\overline{a}_{q}a_{k}^{\dagger}|-\rangle$$

$$+ (-1)^{2}\delta_{p,j}\langle -|a_{k}^{\dagger}a_{j}a_{q}a_{k}^{\dagger}|-\rangle + (-1)^{2}\delta_{p,j}\delta_{i,l}\langle -|\overline{a}_{q}a_{k}^{\dagger}|-\rangle$$

$$+ (-1)^{2}\delta_{p,j}\delta_{q,l}\langle -|a_{k}^{\dagger}a_{j}|-\rangle + (-1)\delta_{p,j}\delta_{q,l}\delta_{i,k}\langle -|I|-\rangle$$

$$+ (-1)^{2}\delta_{p,j}\delta_{l,j}\langle -|a_{k}^{\dagger}a_{q}|-\rangle + (-1)\delta_{p,i}\delta_{l,j}\delta_{q,k}\langle -|I|-\rangle$$

$$+ \delta_{p,i}\delta_{l,q}\delta_{k,j}$$

$$= (-1)^{3}\delta_{p,j}\delta_{i,l}\langle -|a_{k}^{\dagger}a_{q}|-\rangle + (-1)^{2}\delta_{p,j}\delta_{i,l}\delta_{q,k}\langle -|I|-\rangle$$

$$+ \delta_{p,i}\delta_{l,q}\delta_{k,j}$$

$$= (-1)^{3}\delta_{p,j}\delta_{i,l}\langle -|a_{k}^{\dagger}a_{q}|-\rangle + (-1)^{2}\delta_{p,j}\delta_{i,l}\delta_{q,k}\langle -|I|-\rangle$$

$$+ (-1)\delta_{p,j}\delta_{q,l}\delta_{i,k} + (-1)\delta_{p,i}\delta_{l,j}\delta_{q,k} + \delta_{p,i}\delta_{l,q}\delta_{k,j}$$

$$= \delta_{p,j}\delta_{i,l}\delta_{q,k} - \delta_{p,j}\delta_{q,l}\delta_{i,k} - \delta_{p,i}\delta_{l,j}\delta_{q,k} + \delta_{p,i}\delta_{l,q}\delta_{k,j}$$

Hence

$$\begin{aligned} \langle ij|\sum_{p,q}h_{p,q}a_p^{\dagger}a_q|kl\rangle &= \sum_{p,q}h_{p,q}(\delta_{p,j}\delta_{i,l}\delta_{q,k} - \delta_{p,j}\delta_{q,l}\delta_{i,k} - \delta_{p,i}\delta_{l,j}\delta_{qk} + \delta_{p,i}\delta_{l,q}\delta_{k,j}) \\ &= h_{j,k}\delta_{i,l} - h_{j,l}\delta_{i,k} - h_{i,k}\delta_{l,j} + h_{i,l}\delta_{k,j} \end{aligned}$$

Which can be evaluated in  $\mathcal{O}(1)$ . For comparison, a naïve VecMatVec evaluations in first quantization costs  $\mathcal{O}\left(\binom{n}{d}^{3}\right)$ . To reach the final formula of the matrix elements following the procedure in Eq. (190) is very tedious. A shorter path is Wick's theorem.

#### 8.1 Wick's contraction theorem

At the core of Wick's theorem stands the normal ordering of an operator string O which we shall denote by  $\{O\}$ . The normal ordering of an operator string means that all creation operators are to the left of the string and all annihilation operators are to the right. The procedure is to move all creation operators are to the left as if they anti-commuted.

**Example 8.0.1.** Let's investigate the normal ordering of operator strings of length two:

$$\{a_p a_q\} = a_p a_q \qquad \{a_p^{\dagger} a_q^{\dagger}\} = a_p^{\dagger} a_q^{\dagger} \qquad \{a_p^{\dagger} a_q\} = a_p^{\dagger} a_q \qquad \{a_p a_q^{\dagger}\} = -a_q^{\dagger} a_p \qquad (191)$$

Example 8.0.2. Let's look at a second, slightly more complicated example

$$\{a_s a_r a_p^{\dagger} a_q a_t^{\dagger} a_u^{\dagger}\} = \{a_p^{\dagger} a_s a_r a_q a_t^{\dagger} a_u^{\dagger}\} = -\{a_p^{\dagger} a_t^{\dagger} a_s a_r a_q a_u^{\dagger}\} = a_p^{\dagger} a_t^{\dagger} a_u^{\dagger} a_s a_r a_q$$
(192)

A key observation is that the vacuum expectation value of a normal ordered string of operators is zero, i.e.,

$$\langle -|\{O\}|-\rangle = 0. \tag{193}$$

Note that this is an immediate consequence of

$$\mathbf{0} = a_i |-\rangle = \left(\langle -|a_i^{\dagger}\right)^{\dagger}.$$
(194)

We now define contractions of two operators as

$$\prod_{xy}^{\square} = xy - \{xy\} \tag{195}$$

For the creation and annihilation operators there are now four options in total, namely

$$\begin{aligned}
\overline{a_p^{\dagger}a_q^{\dagger}} &= a_p^{\dagger}a_q^{\dagger} - \{a_p^{\dagger}a_q^{\dagger}\} = a_p^{\dagger}a_q^{\dagger} - a_p^{\dagger}a_q^{\dagger} = 0\\
\overline{a_p^{\dagger}a_q} &= a_pa_q - \{a_pa_q\} = a_pa_q - a_pa_q = 0\\
\overline{a_p^{\dagger}a_q} &= a_p^{\dagger}a_q - \{a_p^{\dagger}a_q\} = a_p^{\dagger}a_q - a_p^{\dagger}a_q = 0\\
\overline{a_p^{\dagger}a_q^{\dagger}} &= a_pa_q^{\dagger} - \{a_pa_q^{\dagger}\} = a_pa_q^{\dagger} + a_q^{\dagger}a_p = \delta_{p,q}
\end{aligned}$$
(196)

We see that the only non-zero contribution occurs when an annihilation operator appears to the left of a creation operator.

We are now able to formulate Wick's theorem:

**Theorem 8.1.** An operator string can be written as a linear combination of normal-ordered strings, *i.e.*,

$$ABC...XYZ = \{ABC...XYZ\} + \sum_{\text{singles}} \{\overrightarrow{ABC}...XYZ\} + \sum_{\text{doubles}} \{\overrightarrow{ABC}...XYZ\} + \sum_{\text{doubles}} \{\overrightarrow{ABC}...XYZ\} + ...$$

$$(197)$$

**Corollary 8.1.1.** Only fully contracted terms contribute to vacuum expectation values.

With these results, we may now return to the problem of evaluating vacuum expectation values. Recall that by direct – yet tedious – application of the CAR we found

$$\langle -|a_j a_i a_p^{\dagger} a_q a_l^{\dagger} a_k^{\dagger}| - \rangle = \delta_{p,j} \delta_{i,l} \delta_{q,k} - \delta_{p,j} \delta_{q,l} \delta_{i,k} - \delta_{p,i} \delta_{l,j} \delta_{qk} + \delta_{p,i} \delta_{l,q} \delta_{k,j}$$
(198)

Alternatively to the bruit force derivation, we may apply Wick's theorem to the expectation value. Noting that only fully contracted terms contribute to the vacuum expectation value we find

$$\langle -|a_{j}a_{i}a_{p}^{\dagger}a_{q}a_{l}^{\dagger}a_{k}^{\dagger}|-\rangle = \langle -|a_{j}a_{i}a_{p}^{\dagger}a_{q}a_{l}^{\dagger}a_{k}^{\dagger}|-\rangle$$

$$+ \langle -|a_{j}a_{i}a_{p}^{\dagger}a_{q}a_{l}^{\dagger}a_{k}^{\dagger}|-\rangle$$

$$+ \langle -|a_{j}a_{i}a_{p}^{\dagger}a_{q}a_{l}^{\dagger}a_{k}^{\dagger}|-\rangle$$

$$+ \langle -|a_{j}a_{i}a_{p}^{\dagger}a_{q}a_{l}^{\dagger}a_{k}^{\dagger}|-\rangle$$

$$+ \langle -|a_{j}a_{i}a_{p}^{\dagger}a_{q}a_{l}^{\dagger}a_{k}^{\dagger}|-\rangle$$

$$= \delta_{p,j}\delta_{i,l}\delta_{q,k} - \delta_{p,j}\delta_{q,l}\delta_{i,k} - \delta_{p,i}\delta_{l,j}\delta_{qk} + \delta_{p,i}\delta_{l,q}\delta_{k,j}$$

$$(199)$$

Clearly, this is much more tractable than the brut force derivation. However, it is still very tedious when we are considering a general N-particle state. Consider the expectation value

$$\langle i_1 i_2 \dots i_N | a_p^{\dagger} a_q^{\dagger} a_r a_s | j_1 j_2 \dots j_N \rangle = \langle -|a_{i_N} \dots a_{i_1} a_p^{\dagger} a_q^{\dagger} a_r a_s a_{j_N}^{\dagger} \dots a_{j_1}^{\dagger}| - \rangle$$
(200)

which rapidly yields an intractable number of Wick contractions.

#### 8.2 Particle-Hole formalism

Fortunately, Wick's theorem extends to the Fermi-vaccuum, that is, a single reference Slater determinant. To that end, we consider the Slater determinant

$$|\phi_0\rangle = |\mathbf{0}\rangle = a_{j_N}^{\dagger} \dots a_{j_1}^{\dagger}|-\rangle, \qquad (201)$$

and without loss of generality we assume  $j_i = i$ . As before, we denote the orbitals  $i \in [[N]]$  as occupied and  $a \in [[K]] \setminus [[N]]$  as virtual. Before proceeding, we make the observation that the vacuum state  $|-\rangle$  is the only occupation number vector for which

$$a_p|-\rangle = \mathbf{0} \qquad \forall p \in \llbracket K \rrbracket$$
 (202)

holds. Interestingly, we observe that

$$a_a |\phi_0\rangle = a_i^{\dagger} |\phi_0\rangle = \mathbf{0} \qquad \forall i \in \llbracket N \rrbracket \text{ and } a \in \llbracket K \rrbracket \setminus \llbracket N \rrbracket$$

$$(203)$$

This suggests that we can interpret the reference state  $|\phi_0\rangle$  as "vacuum" state in a quasiparticle picture, called the Fermi-vacuum. We can then define the ladder operators with respect to this quasi-particle picture, i.e.,

$$c_p := \begin{cases} a_p^{\dagger}, & \text{if } p \in \llbracket N \rrbracket \\ a_p, & \text{else.} \end{cases}$$
(204)

The physical interpretation of this is that  $c_a^{\dagger} = a_a^{\dagger}$  creates a particle (electron) and  $c_i^{\dagger} = a_i$  creates a hole (vacancy). We can then extend Wick's theorem to this new framework where we are seeking normal ordering with respect to the Fermi-vacuum, denoted by  $\{\cdot\}_0$ . For the one-body Hamiltonian we find

$$\sum_{p,q=1}^{K} h_{pq} a_p^{\dagger} a_q = \sum_{p,q=1}^{K} h_{pq} \left( \{ a_p^{\dagger} a_q \}_0 + \{ a_p^{\dagger} a_q \}_0 \right)$$
(205)

Recall that the only non-zero contribution appears when an annihilation operator appears to the left of a creation operator, i.e.,  $c_p c_q^{\dagger}$ . Given the definition of the particle-hole operators this only occurs if  $p, q \in [N]$ , leading to

$$\{\overline{a}_{p}^{\dagger}\overline{a}_{q}\}_{0} = \delta_{p,q}\mathbb{1}_{[\![N]\!]}(p).$$

$$(206)$$

Hence,

$$\sum_{p,q=1}^{K} h_{pq} a_p^{\dagger} a_q = \sum_{p,q=1}^{K} h_{pq} \{ a_p^{\dagger} a_q \}_0 + \sum_{i=1}^{N} h_{ii}$$
(207)

For the two-electron part, we note that the non-zero contractions are

$$a_{p}^{\dagger}a_{r}^{\dagger}a_{s}a_{q} = \{a_{p}^{\dagger}a_{r}^{\dagger}a_{s}a_{q}\}_{0} + \{\overline{a_{p}^{\dagger}a_{r}^{\dagger}a_{s}a_{q}}\}_{0} + \{\overline{a_{p}^{\dagger}a_{r}^{\dagger}a_{s}a_{q}}\}_{0} + \{\overline{a_{p}^{\dagger}a_{r}^{\dagger}a_{s}a_{q}}\}_{0} + \{\overline{a_{p}^{\dagger}a_{r}^{\dagger}a_{s}a_{q}}\}_{0} + \{\overline{a_{p}^{\dagger}a_{r}^{\dagger}a_{s}a_{q}}\}_{0}$$

$$(208)$$

$$+ \{\overline{a_{p}^{\dagger}a_{r}^{\dagger}a_{s}a_{q}}\}_{0} + \{\overline{a_{p}^{\dagger}a_{r}^{\dagger}a_{s}a_{q}}\}_{0}$$

Evaluating the contraction for the double contractions leads to

$$\{a_{p}^{\dagger}a_{r}^{\dagger}a_{s}a_{q}\}_{0} = -\delta_{p,s}\mathbb{1}_{\llbracket N \rrbracket}(p)\delta_{r,q}\mathbb{1}_{\llbracket N \rrbracket}(q)$$

$$\{a_{p}^{\dagger}a_{r}^{\dagger}a_{s}a_{q}\}_{0} = \delta_{p,q}\mathbb{1}_{\llbracket N \rrbracket}(p)\delta_{r,s}\mathbb{1}_{\llbracket N \rrbracket}(r)$$
(209)

which results in

$$\frac{1}{2} \sum_{pqrs=1}^{K} v_{p,q,r,s} \left( \{ a_{p}^{\dagger} a_{r}^{\dagger} a_{s} a_{q} \}_{0} + \{ a_{p}^{\dagger} a_{r}^{\dagger} a_{s} a_{q} \}_{0} \right)$$

$$= \frac{1}{2} \sum_{pqrs=1}^{K} v_{p,q,r,s} \left( \delta_{p,q} \mathbb{1}_{[N]}(p) \delta_{r,s} \mathbb{1}_{[N]}(r) - \delta_{p,s} \mathbb{1}_{[N]}(p) \delta_{r,q} \mathbb{1}_{[N]}(q) \right) \quad (210)$$

$$= \frac{1}{2} \sum_{p,r=1}^{N} \left( v_{p,p,r,r} - v_{p,r,r,p} \right)$$

Similarly, for the singles we find

$$\{ \overline{a_{p}^{\dagger}a_{r}^{\dagger}a_{s}}a_{q} \}_{0} = -\{\overline{a_{p}^{\dagger}a_{s}}a_{r}^{\dagger}a_{q} \}_{0} = -\delta_{p,s} \mathbb{1}_{[N]}(p)\{\overline{a_{r}^{\dagger}a_{q}}\}_{0} 
\{ \overline{a_{p}^{\dagger}a_{r}^{\dagger}a_{s}}a_{q} \}_{0} = \{\overline{a_{p}^{\dagger}a_{q}}a_{r}^{\dagger}a_{s} \}_{0} = \delta_{p,q} \mathbb{1}_{[N]}(p)\{\overline{a_{r}^{\dagger}a_{s}}\}_{0} 
\{ \overline{a_{p}^{\dagger}a_{r}^{\dagger}a_{s}}a_{q} \}_{0} = \delta_{r,s} \mathbb{1}_{[N]}(r)\{\overline{a_{p}^{\dagger}a_{q}}\}_{0} 
\{ \overline{a_{p}^{\dagger}a_{r}^{\dagger}a_{s}}a_{q} \}_{0} = -\{\overline{a_{p}^{\dagger}a_{r}^{\dagger}a_{q}}a_{s} \}_{0} = -\delta_{r,q} \mathbb{1}_{[N]}(r)\{\overline{a_{p}^{\dagger}a_{s}}\}_{0}$$

$$(211)$$

which results in

$$\frac{1}{2} \sum_{pqrs=1}^{K} v_{p,q,r,s} \left( \{ \overline{a}_{p}^{\dagger} \overline{a}_{r}^{\dagger} \overline{a}_{s} a_{q} \}_{0} + \{ \overline{a}_{p}^{\dagger} \overline{a}_{r}^{\dagger} \overline{a}_{s} a_{q} \}_{0} + \{ \overline{a}_{p}^{\dagger} \overline{a}_{r}^{\dagger} \overline{a}_{s} a_{q} \}_{0} + \{ \overline{a}_{p}^{\dagger} \overline{a}_{r}^{\dagger} \overline{a}_{s} a_{q} \}_{0} \right) \\
= \frac{1}{2} \sum_{pqrs=1}^{K} v_{p,q,r,s} \left( \delta_{p,q} \mathbb{1}_{[[N]]}(p) \{ \overline{a}_{r}^{\dagger} \overline{a}_{s} \}_{0} - \delta_{p,s} \mathbb{1}_{[[N]]}(p) \{ \overline{a}_{r}^{\dagger} \overline{a}_{q} \}_{0} \\
+ \delta_{r,s} \mathbb{1}_{[[N]]}(r) \{ \overline{a}_{p}^{\dagger} \overline{a}_{q} \}_{0} - \delta_{r,q} \mathbb{1}_{[[N]]}(r) \{ \overline{a}_{p}^{\dagger} \overline{a}_{s} \}_{0} \right) \\
= \frac{1}{2} \left( \sum_{rs=1}^{K} \sum_{i=1}^{N} v_{i,i,r,s} \{ \overline{a}_{r}^{\dagger} \overline{a}_{s} \}_{0} - \sum_{qr=1}^{K} \sum_{i=1}^{N} v_{i,q,r,i} \{ \overline{a}_{r}^{\dagger} \overline{a}_{q} \}_{0} \\
+ \sum_{pq=1}^{K} \sum_{i=1}^{N} v_{p,q,i,i} \{ \overline{a}_{p}^{\dagger} \overline{a}_{q} \}_{0} - \sum_{ps=1}^{K} \sum_{i=1}^{N} v_{p,i,i,s} \{ \overline{a}_{p}^{\dagger} \overline{a}_{s} \}_{0} \right) \\
= \frac{1}{2} \sum_{pq=1}^{K} \sum_{i=1}^{N} \left( v_{i,i,p,q} \{ \overline{a}_{p}^{\dagger} \overline{a}_{q} \}_{0} - v_{i,q,p,i} \{ \overline{a}_{p}^{\dagger} \overline{a}_{q} \}_{0} \\
+ v_{p,q,i,i} \{ \overline{a}_{p}^{\dagger} \overline{a}_{q} \}_{0} - v_{p,i,i,q} \{ \overline{a}_{p}^{\dagger} \overline{a}_{q} \}_{0} \right) \\
= \sum_{pq=1}^{K} \sum_{i=1}^{N} \left( v_{p,q,i,i} - v_{p,i,i,q} \} \{ \overline{a}_{p}^{\dagger} \overline{a}_{q} \}_{0} \right) \\$$

where used that

$$v_{p,q,r,s} = v_{r,s,p,q}.$$
 (213)

Therefore, the final form of the electronic structure Hamiltonian is given by

$$H = E^{\rm HF} + \sum_{p,q=1}^{K} \left( h_{pq} + \sum_{i=1}^{N} \left( v_{p,q,i,i} - v_{p,i,i,q} \right) \right) \{ a_p^{\dagger} a_q \}_0 + \frac{1}{2} \sum_{pqrs} v_{pqrs} \{ a_p^{\dagger} a_r^{\dagger} a_s a_q \}_0$$
(214)

where the Hartree–Fock energy appears naturally

$$E^{(\mathrm{HF})} = \sum_{i=1}^{N} h_{ii} + \frac{1}{2} \sum_{p,r=1}^{N} \left( v_{p,p,r,r} - v_{p,r,r,p} \right) = \langle \phi_0 | H | \phi_0 \rangle.$$
(215)

The normal ordered Hamiltonian is then given by

$$H_N := H - \langle \phi_0 | H | \phi_0 \rangle = \sum_{p,q=1}^K f_{p,q} \{ a_p^{\dagger} a_q \}_0 + \frac{1}{2} \sum_{pqrs} v_{pqrs} \{ a_p^{\dagger} a_r^{\dagger} a_s a_q \}_0$$
(216)

where

$$f_{p,q} := h_{pq} + \sum_{i=1}^{N} \left( v_{p,q,i,i} - v_{p,i,i,q} \right).$$
(217)

### 9 Post Hartree–Fock Methods

We will now explore post-Hartree–Fock methods, i.e., methods that aim to improve upon the Hartree–Fock solution. The energetic correction to the Hartree Fock energy will be referred to as *correlation energy*.

We being this chapter with a short exposition of excitation matrices which are a common and useful tool to characterize post-Hartree–Fock methods. The *N*-particle Hartree–Fock solu tion, often also referred to as reference state, takes the following simple form assuming molecular orbitals:

$$|\Psi_0\rangle = |1, ..., 1, 0, ...0\rangle = \underbrace{\begin{pmatrix} 0\\1 \end{pmatrix} \otimes \cdots \otimes \begin{pmatrix} 0\\1 \end{pmatrix}}_{N-\text{times}} \otimes \underbrace{\begin{pmatrix} 1\\0 \end{pmatrix} \otimes \cdots \otimes \begin{pmatrix} 1\\0 \end{pmatrix}}_{N_B-N-\text{times}} \in \mathcal{H}^N,$$
(218)

where the first N entries in the occupation vector are set to one, and the remaining entries are zero. As common for post-Hartree–Fock theory and outlined before, define  $\mathcal{V}_{\text{occ}} = [\![N]\!] = \{1, ..., N\}$  and  $\mathcal{V}_{\text{virt}} = [\![N_B]\!] \setminus [\![N]\!] = \{N + 1, ..., N_B\}$  as occupied and virtual orbitals, respectively. Assume  $a_1, ..., a_k \in \mathcal{V}_{\text{virt}}$ , and  $i_1, ..., i_k \in \mathcal{V}_{\text{occ}}$ . Then,

$$X_{\binom{a_1,\dots,a_k}{i_1,\dots,i_k}} = a_{a_k}^{\dagger} \dots a_{a_1}^{\dagger} a_{i_1} \dots a_{i_k}$$
(219)

defines an excitation matrix, and the set of all excitation matrices on  $\mathcal{H}^N$ , i.e., the *N*-particle sector in  $\mathcal{F}$ , is given by

$$\mathfrak{E}(\mathcal{H}^{(N)}) = \left\{ X_{\mu} \mid \mu = \begin{pmatrix} a_1, \dots, a_k \\ i_1, \dots, i_k \end{pmatrix}, a_j \in \mathcal{V}_{\text{virt}}, i_j \in \mathcal{V}_{\text{occ}}, k \leq N \right\}.$$
(220)

The excitation indices  $\mu$  that excite from the occupied into the virtual orbitals define the multi-index set

$$\mathcal{I} = \left\{ \mu \mid \mu = \begin{pmatrix} a_1, \dots, a_k \\ i_1, \dots, i_k \end{pmatrix}, a_j \in \mathcal{V}_{\text{virt}}, i_j \in \mathcal{V}_{\text{occ}}, 1 \le k \le N \right\}.$$
(221)

Note that the above construction of the excitation matrices yields that excitation matrices are particle number preserving, i.e.,

$$[X_{\mu}, N] = 0 \quad \forall \mu \in \mathcal{I}, \tag{222}$$

where

$$N = \sum_{p=1}^{N_B} a_p^{\dagger} a_p \tag{223}$$

describes the number operator. Note that this notation is ambiguous since N also describes the particle number, i.e., an element in  $\mathbb{N}$ . However, it will be clear from context whether N denotes the number operator or the particle number.

Since this set of excitations corresponds to simply replacing indices in the string [1, ..., N] with indices in the string  $[N + 1, ..., N_B]$  (plus some additional permutation), we deduce that there is a one-to-one relation between excitation operators and Slater determinants except for the reference Slater determinant  $|\Psi_0\rangle$ . In other words, the excitation operators map the reference Slater determinant  $|\Psi_0\rangle$  to all other Slater determinants.

**Proposition 9.0.1.** There exists a one-to-one relation between the N-particle basis functions  $\mathfrak{B}^{(N)}$  and  $\mathfrak{E}(\mathcal{H}^{(N)}) \cup \{I\}$ .

*Proof.* Note that this is true by construction! However, we may formalize the construction a bit further to clearly see why this is true. Since excitation matrices are defined w.r.t. the reference determinant  $|\Psi_0\rangle$  it follows immediately that  $|\Psi_0\rangle = I|\Psi_0\rangle$ . Next consider

$$|\Psi_P\rangle = \xi_{p_1} \wedge \dots \wedge \xi_{p_N} \in \mathcal{H}^{(N)},\tag{224}$$

where  $P = \{p_1, ..., p_N\}$ . Comparing  $\{1, ..., N\}$  to  $\{p_1, ..., p_N\}$  we can identify a multi-index  $\mu$  describing the indices that have to be changed in  $\{1, ..., N\}$  to obtain  $\{p_1, ..., p_N\}$ . More precisely,  $\mu$  describes an excitation from  $\mathcal{V}_{\text{occ}} \setminus P$  to  $P \cap \mathcal{V}_{\text{virt}}$ . Due to the canonical ordering, this multi-index  $\mu$  is unique. Then, by definition we obtain  $|\Psi_P\rangle = \text{sign}(\mu)X_{\mu}|\Psi_0\rangle$ , which shows the claim.

The above result is the fundamental result that allows us to express any target wave function  $|\Psi\rangle \in \mathcal{H}^{(N)}$  through a sequence of excitation matrices applied to the reference determinant instead of an expansion through basis vectors, i.e.,

$$|\Psi\rangle = \Omega |\Psi_0\rangle := \left(c_0 I + \sum_{\mu} c_{\mu} X_{\mu}\right) |\Psi_0\rangle.$$
(225)

Before proceeding to the different post-Hartree–Fock methods, we highlight a few properties of the excitation matrices. The first property we consider is the commutativity of the excitation matrices.

**Proposition 9.0.2.** Let  $X_{\mu}, X_{\nu} \in \mathfrak{E}(\mathcal{H}^{(N)})$ . Then  $[X_{\mu}, X_{\nu}] = 0$ .

*Proof.* Let

$$X_{\mu} = X_{\binom{a_1, \dots, a_k}{i_1, \dots, i_k}} = a_{a_k}^{\dagger} \dots a_{a_1}^{\dagger} a_{i_1} \dots a_{i_k} \quad \text{and} \quad X_{\nu} = X_{\binom{b_1, \dots, b_\ell}{j_1, \dots, j_\ell}} = a_{b_\ell}^{\dagger} \dots a_{b_1}^{\dagger} a_{j_1} \dots a_{j_\ell}.$$

The proof is conducted in two steps:

First, we seek to permute all creation operators in the commutator to the left using the CAR. We begin with the following product and note that when permuting  $a_{b_{\ell}}^{\dagger}$  to the right of  $a_{a_1}^{\dagger}$  we merely pick up a sign, since  $b_{\ell} \notin \mathcal{V}_{\text{occ}}$ , i.e.,

$$a_{a_k}^{\dagger} \dots a_{a_1}^{\dagger} a_{i_1}^{\dagger} \dots a_{i_k}^{\dagger} a_{b_{\ell}}^{\dagger} \dots a_{b_1}^{\dagger} a_{j_1} \dots a_{j_{\ell}} = (-1)^k a_{a_k}^{\dagger} \dots a_{a_1}^{\dagger} a_{b_{\ell}}^{\dagger} a_{i_1} \dots a_{i_k} a_{b_{\ell-1}}^{\dagger} \dots a_{b_1}^{\dagger} a_{j_1} \dots a_{j_{\ell}}.$$

This furthermore yields

$$a_{a_k}^{\dagger} \dots a_{a_1}^{\dagger} a_{i_1} \dots a_{i_k} a_{b_\ell}^{\dagger} \dots a_{b_1}^{\dagger} a_{j_1} \dots a_{j_\ell} = (-1)^{\ell \cdot k} a_{a_k}^{\dagger} \dots a_{a_1}^{\dagger} a_{b_\ell}^{\dagger} \dots a_{b_1}^{\dagger} a_{i_1} \dots a_{i_k} a_{j_1} \dots a_{j_\ell}$$

and similar

$$a_{b_{\ell}}^{\dagger} \dots a_{b_{1}}^{\dagger} a_{j_{1}} \dots a_{j_{\ell}} a_{a_{k}}^{\dagger} \dots a_{a_{1}}^{\dagger} a_{i_{1}} \dots a_{i_{k}} = (-1)^{\ell \cdot k} a_{b_{\ell}}^{\dagger} \dots a_{b_{1}}^{\dagger} a_{a_{k}}^{\dagger} \dots a_{a_{1}}^{\dagger} a_{j_{1}} \dots a_{j_{\ell}} a_{i_{1}} \dots a_{i_{k}}.$$

Second, we wish to unify the index sequence of the creation and annihilation operators in the two summands of the commutator. Applying the CAR again, we find

$$a_{b_{\ell}}^{\dagger} \dots a_{b_{1}}^{\dagger} a_{a_{k}}^{\dagger} \dots a_{a_{1}}^{\dagger} a_{j_{1}} \dots a_{j_{\ell}} a_{i_{1}} \dots a_{i_{k}} = (-1)^{\ell} a_{a_{k}}^{\dagger} a_{b_{\ell}}^{\dagger} \dots a_{b_{1}}^{\dagger} a_{a_{k-1}}^{\dagger} \dots a_{a_{1}}^{\dagger} a_{j_{1}} \dots a_{j_{\ell}} a_{i_{1}} \dots a_{i_{k}},$$

which yields

$$a_{b_{\ell}}^{\dagger} \dots a_{b_{1}}^{\dagger} a_{a_{k}}^{\dagger} \dots a_{a_{1}}^{\dagger} a_{j_{1}} \dots a_{j_{\ell}} a_{i_{1}} \dots a_{i_{k}} = (-1)^{2*\ell \cdot k} a_{a_{k}}^{\dagger} \dots a_{a_{1}}^{\dagger} a_{b_{\ell}}^{\dagger} \dots a_{b_{1}}^{\dagger} a_{i_{1}} \dots a_{i_{k}} a_{j_{1}} \dots a_{j_{\ell}}.$$

Note that we have here assumed that  $\mu \cap \nu = \emptyset$ , otherwise the expression is trivially zero due to the nilpotency of the creation and annihilation operators. Overall this yields

$$\begin{split} [X_{\mu}, X_{\nu}] &= \left[ X_{\begin{pmatrix} a_{1}, \dots, a_{k} \\ i_{1}, \dots, i_{k} \end{pmatrix}}, X_{\begin{pmatrix} b_{1}, \dots, b_{\ell} \\ j_{1}, \dots, j_{\ell} \end{pmatrix}} \right] \\ &= a_{a_{k}}^{\dagger} \dots a_{a_{1}}^{\dagger} a_{i_{1}} \dots a_{i_{k}} a_{b_{\ell}}^{\dagger} \dots a_{b_{1}}^{\dagger} a_{j_{1}} \dots a_{j_{\ell}} - a_{b_{\ell}}^{\dagger} \dots a_{b_{1}}^{\dagger} a_{j_{1}} \dots a_{j_{\ell}} a_{a_{k}}^{\dagger} \dots a_{a_{1}}^{\dagger} a_{i_{1}} \dots a_{i_{\ell}} \\ &= (-1)^{\ell \cdot k} a_{a_{k}}^{\dagger} \dots a_{b_{1}}^{\dagger} a_{i_{1}} \dots a_{i_{k}} a_{j_{1}} \dots a_{j_{\ell}} - (-1)^{\ell \cdot k} a_{b_{\ell}}^{\dagger} \dots a_{b_{1}}^{\dagger} a_{i_{1}} \dots a_{i_{k}} a_{j_{1}} \dots a_{j_{\ell}} \\ &= (-1)^{\ell \cdot k} a_{a_{k}}^{\dagger} \dots a_{a_{1}}^{\dagger} a_{b_{\ell}}^{\dagger} \dots a_{b_{1}}^{\dagger} a_{i_{1}} \dots a_{i_{k}} a_{j_{1}} \dots a_{j_{\ell}} - (-1)^{3 \cdot \ell \cdot k} a_{a_{k}}^{\dagger} \dots a_{a_{1}}^{\dagger} a_{b_{\ell}}^{\dagger} \dots a_{i_{k}}^{\dagger} a_{j_{1}} \dots a_{i_{k}} a_{j_{1}} \dots a_{j_{\ell}} \\ &= 0. \end{split}$$

Another important property is that the excitation matrices inherited the nilpotency from the fermionic creation and annihilation matrices.

**Proposition 9.0.3.** Let  $X_{\mu} \in \mathfrak{E}(\mathcal{H}^{(N)})$ . Then  $X_{\mu}^2 = 0$ .

*Proof.* Recall that  $(a_p^{\dagger})^2 = (a_p)^2 = 0$  by construction (see Eq. (179)). Let

$$X_{\mu} = X_{\binom{a_1, \dots, a_k}{i_1, \dots, i_k}} = a_{a_k}^{\dagger} \dots a_{a_1}^{\dagger} a_{i_k} \dots a_{i_1}.$$

Then

$$X_{\mu}^{2} = a_{a_{k}}^{\dagger} \dots a_{a_{1}}^{\dagger} a_{i_{k}} \dots a_{i_{1}}^{\dagger} a_{a_{k}}^{\dagger} \dots a_{a_{1}}^{\dagger} a_{i_{k}} \dots a_{i_{1}}^{\dagger} \\ = - \underbrace{a_{a_{k}}^{\dagger} a_{a_{k}}^{\dagger}}_{=0} a_{a_{k}-1}^{\dagger} \dots a_{a_{1}}^{\dagger} a_{i_{k}} \dots a_{i_{1}} a_{a_{k}-1}^{\dagger} \dots a_{a_{1}}^{\dagger} a_{i_{k}} \dots a_{i_{1}} \\ = 0.$$
(227)

#### 9.1 Configuration interaction

In the CI method, the electronic wave function is constructed as a linear combination of Slater determinants, i.e.,

$$|\Psi_{CI}\rangle = \Omega_{\rm CI}|\Psi_0\rangle := \left(c_0I + \sum_{\mu} c_{\mu}X_{\mu}\right)|\Psi_0\rangle.$$
(228)

where we defined the CI wave operator  $\Omega_{\text{CI}}$ . The linear coefficients **c** are then determined by a variational optimization of the expectation value of the electronic energy, i.e.,

$$\mathbf{0} = \nabla_{\mathbf{c}} \frac{\langle \Psi_{CI} | H | \Psi_{CI} \rangle}{\langle \Psi_{CI} | \Psi_{CI} \rangle} \quad \Leftrightarrow \quad H\mathbf{c} = E\mathbf{c}, \tag{229}$$

where H is denoting the Hamiltonian in both the Fock space and in corresponding particle sector, respectively. Note that the right hand side corresponds to a standard hermitian eigenvalue problem. When no restriction on the coefficient vectors  $\mathbf{c}$  is made, we refer to this method as full configuration interaction (FCI) method. For FCI wave functions, the number of Slater determinants – therewith the size of the matrix H – increases very rapidly with the number of electrons and with the number of orbitals. This behavior is illustrated in Table 2, where we have listed the number of determinants with spin projection zero obtained by distributing an even number of 2k electrons among 2k orbitals for  $1 \leq k \leq 10$ .

2k	$N_{ m det}$
2	4
4	36
6	400
8	4 900
10	63  504
12	$853\ 776$
14	$11\ 778\ 624$
16	$165 \ 636 \ 900$
18	$2 \ 363 \ 904 \ 400$
20	$34 \ 134 \ 779 \ 536$

Table 2: The number of Slater determinants  $N_{det}$  with spin projection zero obtained by distributing k electrons among 2k orbitals

In Table 2, the number of Slater determinants  $N_{det}$  is the product of the number of determinants for the spin up and spin down electrons separately. Thus, for a system with n orbitals containing k up electrons and k down electrons, the number of determinants is given by

$$N_{\rm det} = \binom{n}{k}^2 \tag{230}$$

Thus, the FCI expansion is numerically intractable for any but the smallest electronic systems. Indeed, the usefulness of the FCI model is mostly that, for small systems, it may provide benchmarks for other determinantal wave-function models. Therefore, for most systems, one has to truncate the (full) CI expansion. This can either be done at a given level of excitations for manageable calculations referred to as *truncated* CI, e.g., CISD, CISDT, etc., or by restricting the expansion to a selected number of determinants referred to as *selected* CI.

The truncated CI method has a serious shortcoming. Consider the *size consistency* property that the total energy of a system composed of two non-interacting fragments A and B must be the sum of the total energies of the separate fragments, i.e.,

$$E(A \cdots B) = E(A) + E(B). \tag{231}$$

This property is particularly important in chemistry since it is often concerned with systems composed of fragments (atoms, molecules). It is of course satisfied for the exact total energy, but not necessarily with approximate methods. A method which gives total energies satisfying this property is said to be *size-consistent*. For example, the unrestricted HF method is size consistent but the restricted HF method is generally not. The FCI method is size-consistent, but the truncated CI method has the important drawback of being generally not size-consistent.

#### 9.2 Perturbation Theory

#### 9.2.1 General Rayleigh–Schrödinger perturbation theory

The basic idea of quantum-mechanical perturbation theory is to partition the Hamiltonian operator into two parts

$$H = H_0 + U \tag{232}$$

where  $H_0$  is some zero-order Hamiltonian

$$U = H - H_0 \tag{233}$$

is the perturbation. The orthonormal eigensolutions to the zero-order Hamiltonian

$$H_0 |\Psi_i^{(0)}\rangle = E_i^{(0)} |\Psi_i^{(0)}\rangle$$
 (234)

are then used to expand the eigenfunction  $|\Psi_i^{(0)}\rangle$  of the exact Hamiltonian in Eq. (232). To that end, we perform a perturbation theory argument, let  $\lambda$  be a dimensionless parameter that can take on value ranging continuously from 0 (no perturbation) to 1 (the full perturbation). The perturbed Hamiltonian is:

$$H_{\lambda} = H_0 + \lambda U \tag{235}$$

The energy levels and eigenstates of the perturbed Hamiltonian are again given by

$$H_{\lambda}|\Psi_{i}^{(\lambda)}\rangle = E_{i}^{(\lambda)}|\Psi_{i}^{(\lambda)}\rangle$$
(236)

The objective now is to express  $E_i^{(\lambda)}$  and  $|\Psi_i^{(\lambda)}\rangle$ . If the perturbation is sufficiently weak, they can be written as a Maclaurin power series in  $\lambda$ , i.e.,

$$E_{i}^{(\lambda)} = E_{i}^{[0]} + \lambda E_{i}^{[1]} + \lambda^{2} E_{i}^{[2]} + \dots$$

$$|\Psi_{i}^{(\lambda)}\rangle = |\Psi_{i}^{[0]}\rangle + \lambda |\Psi_{i}^{[1]}\rangle + \lambda^{2} |\Psi_{i}^{[2]}\rangle + \dots$$
(237)

where

$$E_{i}^{[k]} = \frac{1}{k!} \frac{d^{k} E_{i}^{(\lambda)}}{d\lambda^{k}} \Big|_{\lambda=0}$$

$$|\Psi_{i}^{[k]}\rangle = \frac{1}{k!} \frac{d^{k} |\Psi_{i}^{(\lambda)}\rangle}{d\lambda^{k}} \Big|_{\lambda=0}.$$

$$(238)$$

We will subsequently focus only on the ground state considerations. Note that the zerothorder wave function is just the solution to the zero-order Hamiltonian, i.e.,  $|\Psi_0^{[0]}\rangle = |\Psi_0^{(0)}\rangle$ . We are moreover free to choose the normalization of  $|\Psi_0^{(\lambda)}\rangle$ . A convenient choice is the intermediate normalization, i.e.  $\langle \Psi_0^{(0)} | \Psi_0^{(\lambda)} \rangle = 1$  for all  $\lambda$ . Since the zeroth-order wave function is normalized as  $\langle \Psi_0^{(0)} | \Psi_0 \rangle = 1$ , it implies that  $\langle \Psi_0^{(0)} | \Psi_0^{[k]} \rangle = 0$  for all  $k \ge 1$ , i.e. the wavefunction correction at each order is orthogonal to the zeroth-order wave function. Inserting Maclaurin power series in to the Schrödinger equation yields

$$(H_0 + \lambda U)(|\Psi_0^{[0]}\rangle + \lambda |\Psi_0^{[1]}\rangle + \lambda^2 |\Psi_0^{[2]}\rangle + ...) = (E_0^{[0]} + \lambda E_0^{[1]} + \lambda^2 E_0^{[2]} + ...)(|\Psi_0^{[0]}\rangle + \lambda |\Psi_0^{[1]}\rangle + \lambda^2 |\Psi_0^{[2]}\rangle + ...)$$
(239)

Looking at this equation order by order in  $\lambda$ , we obtain at zeroth order

$$H_0 |\Psi_0^{(0)}\rangle = E_0^{(0)} |\Psi_0^{(0)}\rangle \tag{240}$$

which is just eq. (234). At first order, we obtain

$$H_0|\Psi_0^{[1]}\rangle + U|\Psi_0^{(0)}\rangle = E_0^{(0)}|\Psi_0^{[1]}\rangle + E_0^{[1]}|\Psi_0^{(0)}\rangle.$$
(241)

Projecting onto  $|\Psi_{0}^{(0)}\rangle$  yields

$$\langle \Psi_{0}^{(0)} | H_{0} | \Psi_{0}^{[1]} \rangle + \langle \Psi_{0}^{(0)} | U | \Psi_{0}^{(0)} \rangle = E_{0}^{(0)} \langle \Psi_{0}^{(0)} | \Psi_{0}^{[1]} \rangle + E_{0}^{[1]} \langle \Psi_{0}^{(0)} | \Psi_{0}^{(0)} \rangle$$

$$\Leftrightarrow \quad E_{0}^{(0)} \langle \Psi_{0}^{(0)} | \Psi_{0}^{[1]} \rangle + \langle \Psi_{0}^{(0)} | U | \Psi_{0}^{(0)} \rangle = E_{0}^{(0)} \langle \Psi_{0}^{(0)} | \Psi_{0}^{[1]} \rangle + E_{0}^{[1]} \langle \Psi_{0}^{(0)} | \Psi_{0}^{(0)} \rangle$$

$$\Leftrightarrow \quad \langle \Psi_{0}^{(0)} | U | \Psi_{0}^{(0)} \rangle = E_{0}^{[1]}$$

$$(242)$$

whereas projecting onto  $|\Psi_i^{(0)}\rangle$ . yields

$$\langle \Psi_{i}^{(0)} | H_{0} | \Psi_{0}^{[1]} \rangle + \langle \Psi_{i}^{(0)} | U | \Psi_{0}^{(0)} \rangle = E_{0}^{(0)} \langle \Psi_{i}^{(0)} | \Psi_{0}^{[1]} \rangle + E_{0}^{[1]} \langle \Psi_{i}^{(0)} | \Psi_{0}^{(0)} \rangle$$

$$\Leftrightarrow \qquad E_{i}^{(0)} \langle \Psi_{i}^{(0)} | \Psi_{0}^{[1]} \rangle + \langle \Psi_{i}^{(0)} | U | \Psi_{0}^{(0)} \rangle = E_{0}^{(0)} \langle \Psi_{i}^{(0)} | \Psi_{0}^{[1]} \rangle + E_{0}^{[1]} \langle \Psi_{i}^{(0)} | \Psi_{0}^{(0)} \rangle$$

$$\Leftrightarrow \qquad \langle \Psi_{i}^{(0)} | \Psi_{0}^{[1]} \rangle = -\frac{\langle \Psi_{i}^{(0)} | U | \Psi_{0}^{(0)} \rangle}{E_{i}^{(0)} - E_{0}^{(0)}}.$$

$$(243)$$

Since  $\langle \Psi_0^{(0)} | \Psi_0^{[1]} \rangle = 0$ , this leads to the first-order wave-function correction

$$|\Psi_{0}^{[1]}\rangle = -\sum_{i\neq 0} \frac{\langle \Psi_{i}^{(0)} | U | \Psi_{0}^{(0)} \rangle}{E_{i}^{(0)} - E_{0}^{(0)}} | \Psi_{i}^{(0)} \rangle$$
(244)

Similarly, at second order perturbation we get

$$H_0|\Psi_0^{[2]}\rangle + U|\Psi_0^{[1]}\rangle = E_0^{(0)}|\Psi_0^{[2]}\rangle + E_0^{[1]}|\Psi_0^{[1]}\rangle + E_0^{[2]}|\Psi_0^{(0)}\rangle$$
(245)

which yields

$$E_0^{[2]} = \langle \Psi_0^{(0)} | U | \Psi_0^{[1]} \rangle = -\sum_{i \neq 0} \frac{|\langle \Psi_i^{(0)} | U | \Psi_0^{(0)} \rangle|^2}{E_i^{(0)} - E_0^{(0)}}$$
(246)

#### 9.2.2 Møller-Plesset perturbation theory – spin-unrestricted theory

Møller-Plesset (MP) perturbation theory is a particular case of Rayleigh–Schrödinger perturbation theory for which the zeroth-order Hamiltonian is chosen to be the (many-electron) Hartree–Fock (sometimes also simply called Fock) Hamiltonian

$$H_0 = F \tag{247}$$

and the corresponding perturbation operator W is thus the difference between the electronelectron Coulomb interaction and the effective HF potential  $V_{\text{eff}}$  also called the *fluctuation potential*. In MO formulation this reads

$$W = V - V_{\text{eff}} = \frac{1}{2} \sum_{p,q,r,s} v_{p,q,r,s} a_p^{\dagger} a_r^{\dagger} a_s a_q - \sum_{p,q} \sum_{i \in [\![N]\!]} (v_{p,q,i,i} a_p^{\dagger} a_q - v_{p,i,i,q} a_p^{\dagger} a_q)$$
(248)

**Zeroth order:** The zeroth-order ground-state wave function  $|\Psi_0^{[0]}\rangle = |\Psi_0^{(0)}\rangle$  is the HF single determinant  $|\Psi_0\rangle$ , and the zeroth-order excited-state wave functions are the singly, doubly, triply, etc., excited determinants, i.e.,  $X_{\binom{a}{i}}|\Psi_0\rangle$ ,  $X_{\binom{ab}{ij}}|\Psi_0\rangle$ ,  $X_{\binom{abc}{ijk}}|\Psi_0\rangle$ , etc. The zeroth-order ground-state energy  $E_0^{[0]}$  is given by the sum of occupied orbital energies, i.e.,

$$E_0^{[0]} = \sum_{i \in [\![N]\!]} \varepsilon_i = E_0^{(0)}.$$
(249)

We moreover note that for any given excitation index  $\mu \in \mathcal{I}$ , i.e.,

$$\mu = \begin{pmatrix} a_1, \dots, a_k\\ i_1, \dots, i_k \end{pmatrix},\tag{250}$$

the function  $|\Psi_{\mu}\rangle = X_{\mu}|\Psi_{0}\rangle = a^{\dagger}_{m_{N}}...a^{\dagger}_{m_{1}}|\Psi_{0}\rangle$  is an eigenfunction of the Fock matrix, with

$$F|\Psi_{\mu}\rangle = \sum_{i=1}^{N} \varepsilon_{m_i} |\Psi_{\mu}\rangle.$$
(251)

Hence, the corresponding zeroth-order energy is given by

$$E_{\mu}^{(0)} = \sum_{i=1}^{N} \varepsilon_{m_i} = E_0^{(0)} + \varepsilon_{a_1} + \dots + \varepsilon_{a_k} - \varepsilon_{i_1} + \dots + \varepsilon_{i_k}.$$
 (252)

**First order (MP1):** Following the general power series expansion, we find that the firstorder energy correction is the expectation value of the HF determinant over the fluctuation potential, which is

$$E_{0}^{[1]} = \langle \Psi_{0}^{(0)} | W | \Psi_{0}^{(0)} \rangle$$

$$= \langle \Psi_{0} | W | \Psi_{0} \rangle$$

$$= \langle \Psi_{0} | V | \Psi_{0} \rangle - \langle \Psi_{0} | V_{\text{eff}} | \Psi_{0} \rangle$$

$$= \frac{1}{2} \sum_{i,j \in \llbracket N \rrbracket} \langle ij | | ij \rangle - \sum_{i \in \llbracket N \rrbracket} \langle \phi_{i} | V_{\text{eff}} | \phi_{i} \rangle$$

$$= -\frac{1}{2} \sum_{i,j \in \llbracket N \rrbracket} \langle ij | | ij \rangle$$
(253)

where we used the short hand notation

$$\langle ij||ij\rangle = v_{i,i,j,j} - v_{i,j,i,j}.$$
(254)

Therefore, the sum of the zeroth-order energy and first-order energy correction is simply the HF energy

$$E_{0}^{[0]} + E_{0}^{[1]} = \sum_{i \in \llbracket N \rrbracket} \varepsilon_{i} - \frac{1}{2} \sum_{i,j \in \llbracket N \rrbracket} \langle ij || ij \rangle$$

$$= \sum_{i \in \llbracket N \rrbracket} \left( h_{ii} + \sum_{j \in \llbracket N \rrbracket} \langle ij || ij \rangle \right) - \frac{1}{2} \sum_{i,j \in \llbracket N \rrbracket} \langle ij || ij \rangle$$

$$= \sum_{i \in \llbracket N \rrbracket} h_{ii} + \frac{1}{2} \sum_{i,j \in \llbracket N \rrbracket} \langle ij || ij \rangle$$

$$= E^{(HF)}$$

$$(255)$$

Second order (MP2): Since first order Møller-Plesset perturbation theory basically yields Hartree–Fock, the first "relevant" and, in practice, mostly used Møller-Plesset variant is second-order Møller-Plesset (MP2), i.e., the second order correction. For the energy correction, we find

$$E_0^{[2]} = E_{\rm corr}^{(\rm MP2)} = -\sum_{\mu \in \mathcal{I}} \frac{|\langle \Psi_0 | W | \Psi_\mu \rangle|^2}{E_\mu^{(0)} - E_0^{(0)}}$$
(256)

We observe that the numerator can be further simplified. The general rules that apply here are the *Slater–Condon* rules, however, we can use a direct computation. For a singly excited Slater determinant, i.e.,  $|\Psi_{\mu}\rangle = X_{\binom{a}{i}}|\Psi_{0}\rangle = |\Psi_{i}^{a}\rangle$ 

$$\langle \Psi_{0} | W | \Psi_{i}^{a} \rangle = \sum_{j \in \llbracket N \rrbracket} \langle ij | | aj \rangle - \langle i | V_{\text{eff}} | a \rangle$$

$$= \sum_{j \in \llbracket N \rrbracket} \langle ij | | aj \rangle - \sum_{j \in \llbracket N \rrbracket} \langle ij | | aj \rangle$$

$$= 0$$

$$(257)$$

For a doubly excited Slater determinant, we find

$$\langle \Psi_0 | W | \Psi_{ij}^{ab} \rangle = \langle ij | | ab \rangle \tag{258}$$

For any higher-order excited Slater determinant, the numerator will be zero. Hence,

$$E_0^{[2]} = E_{\rm corr}^{\rm (MP2)} = -\sum_{\substack{\mu \in \mathcal{I} \\ |\mu|=2}} \frac{|\langle \Psi_0 | W | \Psi_\mu \rangle|^2}{E_\mu^{(0)} - E_0^{(0)}} = -\sum_{i < j \in \mathcal{V}_{\rm occ}} \sum_{a < b \in \mathcal{V}_{\rm virt}} \frac{|\langle ij||ab \rangle|^2}{\varepsilon_a + \varepsilon_b - \varepsilon_i - \varepsilon_j}$$
(259)

where we used that the denominator can be simplified using the zeroth-order energy expression in eq. (252), namely,

$$E_{\mu}^{(0)} - E_{0}^{(0)} = \varepsilon_{a_{1}} + \dots + \varepsilon_{a_{k}} - \varepsilon_{i_{1}} + \dots + \varepsilon_{i_{k}}.$$
 (260)

We moreover see that by using the antisymmetry property of the integral expression, i.e.  $\langle ij||ab\rangle = -\langle ij||ba\rangle = -\langle ji||ab\rangle$ , and the fact that  $\langle ij||ab\rangle = 0$  if i = j or a = b, the MP2 correlation energy can also be written without constrained sum, i.e.,

$$E_{\rm corr}^{\rm (MP2)} = -\frac{1}{4} \sum_{i,j \in \mathcal{V}_{\rm occ}} \sum_{a,b \in \mathcal{V}_{\rm virt}} \frac{|\langle ij||ab\rangle|^2}{\varepsilon_a + \varepsilon_b - \varepsilon_i - \varepsilon_j}$$
(261)

Third order (MP3): Similarly, we can continue computing third-order Møller Plesset. After some tedious but straight forward work we find

$$E_{0}^{[3]} = \frac{1}{8} \sum_{a,b,c,d \in \mathcal{V}_{occ}} \sum_{rs \in \mathcal{V}_{virt}} \frac{\langle ab || rs \rangle \langle rs || cd \rangle \langle cd || ab \rangle}{(\varepsilon_{r} + \varepsilon_{s} - \varepsilon_{a} - \varepsilon_{b})(\varepsilon_{r} + \varepsilon_{s} - \varepsilon_{c} - \varepsilon_{d})} + \frac{1}{8} \sum_{a,b \in \mathcal{V}_{occ}} \sum_{rstu \in \mathcal{V}_{virt}} \frac{\langle ab || rs \rangle \langle rs || tu \rangle \langle tu || ab \rangle}{(\varepsilon_{r} + \varepsilon_{s} - \varepsilon_{a} - \varepsilon_{b})(\varepsilon_{t} + \varepsilon_{u} - \varepsilon_{a} - \varepsilon_{b})} + \sum_{a,b,c \in \mathcal{V}_{occ}} \sum_{rst \in \mathcal{V}_{virt}} \frac{\langle ab || rs \rangle \langle cs || tb \rangle \langle rt || ac \rangle}{(\varepsilon_{r} + \varepsilon_{s} - \varepsilon_{a} - \varepsilon_{b})(\varepsilon_{r} + \varepsilon_{t} - \varepsilon_{a} - \varepsilon_{c})}$$
(262)

We note that the calculation of the third- or higher-order Møller-Plessert corrections is often considered as not worthwhile in comparison with alternatives such as *coupled-cluster* methods.

We note that we starting from an unrestricted HF calculation, MP perturbation theory is correctly size consistent at each order. This is a consequence of the fact that the energy correction at each order cannot be factorized in uncoupled sums. This coupling between *all* orbital indices is known as the *linked-cluster theorem*.

#### 9.2.3 Spin-restricted MP2

For closed-shell systems, with spin-singlet symmetry, the MP2 correlation energy expression can be simplified by summing over the spin coordinates. This yields

$$E_{\rm corr}^{(\rm MP2)} = -\frac{1}{4} \sum_{i,j \in \mathcal{V}_{\rm occ}} \sum_{a,b \in \mathcal{V}_{\rm virt}} \frac{|\langle ij|ab \rangle - \langle ij|ba \rangle|^2}{\varepsilon_a + \varepsilon_b - \varepsilon_i - \varepsilon_j}$$
  
$$= -\frac{1}{4} \sum_{i,j \in \mathcal{V}_{\rm occ}} \sum_{a,b \in \mathcal{V}_{\rm virt}} \frac{(\langle ij|ab \rangle - \langle ij|ba \rangle)(\langle ab|ij \rangle - \langle ba|ij \rangle)}{\varepsilon_a + \varepsilon_b - \varepsilon_i - \varepsilon_j}$$
  
$$= -\frac{1}{2} \sum_{i,j \in \mathcal{V}_{\rm occ}} \sum_{a,b \in \mathcal{V}_{\rm virt}} \frac{\langle ij|ab \rangle \langle ab|ij \rangle - \langle ij|ab \rangle \langle ab|ji \rangle}{\varepsilon_a + \varepsilon_b - \varepsilon_i - \varepsilon_j}$$
(263)

where the last line was obtained by expanding, using the permutation symmetry property of the integrals such as  $\langle ab|ij \rangle = \langle ba|ji \rangle$  and relabeled the dummy indices. We can now perform the summations over the spin coordinates. This yields

$$E_{\rm corr}^{\rm (MP2)} = -\sum_{i,j\in\mathcal{V}_{\rm occ}^{\rm spa}} \sum_{a,b\in\mathcal{V}_{\rm virt}^{\rm spa}} \frac{2\langle ij|ab\rangle\langle ab|ij\rangle - \langle ij|ab\rangle\langle ab|ji\rangle}{\varepsilon_a + \varepsilon_b - \varepsilon_i - \varepsilon_j}$$
(264)

where  $\mathcal{V}_{occ}^{spa}$  and  $\mathcal{V}_{occ}^{spa}$  denotes the spatial occupied and virtual orbitals, respectively.

#### 9.3 Coupled cluster theory

As has become apparent throughout this course, at the core of many electronic structure methods stands the idea of seeking approximations to the wave function via a non-linear ansatz. Coupled cluster methods are here no exception.

In coupled cluster theory, we embrace the ansatz that

$$|\Psi\rangle = \Omega_{\rm CC} |\Psi_0\rangle = e^{T(t)} |\Psi_0\rangle \tag{265}$$

where

$$T(t) = \sum_{\mu \in \mathcal{I}} t_{\mu} X_{\mu} \tag{266}$$

is the *cluster matrix/operator* and its expansion coefficients are called *cluster amplitudes*. For sake of readability, we will subsequently not explicitly write the t dependence of T(t), but use simply T. In order to justify this ansatz mathematically, we need to understand the space of cluster matrices better. To that end, we define the  $\mathbb{C}$ -vector space

$$\mathfrak{b} = \left\{ T = \sum_{\mu} t_{\mu} X_{\mu} \mid \mu \in \mathcal{I} \right\} = \operatorname{span} \left( \mathfrak{E}(\mathcal{H}^{(N)}) \right),$$
(267)

Utilizing the propositions discussed earlier, we will demonstrate that this vector space possesses a highly structured nature. Our next step is to introduce the concept of the exponential of cluster matrices, which forms a key mathematical bridge between cluster matrices and wave operators. This involves drawing a connection between the Lie algebra, as embodied by the cluster matrices, and the Lie group comprising wave operators, thereby establishing an essential theoretical foundation of our modern understanding of coupled cluster theory. To begin this exploration, we first assert that  $\mathfrak{b}$  constitutes some form of Lie algebra. As it turns out, this assertion holds true.

#### **Theorem 9.1.** Elements in b commute and are nilpotent.

*Proof.* Recall that by Proposition 9.0.2 excitation matrices commute. Therefore, for  $T_1, T_2 \in \mathfrak{b}$  we find

$$[T_1, T_2] = \sum_{\mu} \sum_{\nu} t_{\mu} t_{\nu} [X_{\mu}, X_{\nu}] = \sum_{\mu} \sum_{\nu} t_{\mu} t_{\nu} [X_{\nu}, X_{\mu}] = [T_2, T_1],$$

hence, cluster matrices also commute.

To show nilpotency, let  $T \in \mathfrak{b}$  and we expand  $T^{N+1}$  which yields

$$T^{N+1} = \sum_{\substack{k_1+k_2+\dots+k_m=N+1\\k_1,k_2,\dots,k_m \ge 0}} \binom{N+1}{k_1,k_2,\dots,k_m} \prod_{j=1}^m (t_{\mu_j} X_{\mu_j})^{k_j},$$
(268)

where

$$\binom{N+1}{k_1, k_2, \dots, k_m} = \frac{N+1!}{k_1! k_2! \cdots k_m!}$$

is a multinomial coefficient and  $m = |\mathcal{I}|$ . Note that  $k_i < 2$  due to Proposition 9.0.3, meaning that  $(k_1, ..., k_m)$  is a binary string containing N + 1 ones. However, since  $|\mathcal{V}_{occ}| = N$ , there exists one  $i \in \mathcal{V}_{occ}$  that appears at least twice in each matrix  $\prod_{j=1}^{m} (t_{\mu_j} X_{\mu_j})^{k_j}$  and since  $a_i^2 = 0$  this yields that  $T^{N+1} = 0$ , which shows the claim.

**Remark 9.1.1.** Theorem 9.1 illustrates that  $\mathfrak{b}$  equipped with the standard matrix commutator  $[\cdot, \cdot]$  forms a nilpotent Abelian Lie algebra.

We may now prove that coupled cluster theory is a meaningful ansatz, in the sense that it enable to parameterize a sufficiently large portion of the N-particle Hilbert space.

**Theorem 9.2.** Let  $|\Psi\rangle \in \mathcal{H}^{(N)}$  with  $\langle \Psi_0 | \Psi \rangle = 1$ . Then there exists a unique  $T \in \text{span}(\mathfrak{E}(\mathcal{H}^{(N)}))$  such that

$$|\Psi\rangle = e^{T(t)}|\Psi_0\rangle. \tag{269}$$

*Proof.* Recall that in the full configuration interaction ansatz the wave function is expressed as

$$|\Psi\rangle = \left(c_0 I + \sum_{\mu \in \mathcal{I}} c_\mu X_\mu\right) |\Psi_0\rangle.$$
(270)

Since  $\langle \Psi | \Psi_0 \rangle = 1$ , we know that  $c_0 = 1$ , and define

$$C := \sum_{\mu \in \mathcal{I}} c_{\mu} X_{\mu} \in \mathfrak{b}.$$
 (271)

Next, we define

$$T := \log(I+C) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n+1} C^{n+1} = \sum_{n=0}^{N} \frac{(-1)^n}{n+1} C^{n+1}$$
(272)

where we used that  $C \in \mathfrak{b}$ , and therefore  $C^{N+1} = 0$ . As a polynomial of matrices, this expression is well defined, which concludes the proof.

**Remark 9.2.1.** We call the condition that  $\langle \Psi_0, \Psi \rangle = 1$  intermediately normalized. In particular, the exponential ansatz in Eq. (265) is not normalized.

**Remark 9.2.2.** Due to the intermediate normalization, the coupled cluster ansatz can only parameterize states that have overlap equal to one with the chosen reference  $|\Psi_0\rangle$ . However, we may renormalize any wavefunction with non-zero overlap to the reference as

$$|\Psi\rangle = \frac{1}{\langle\Psi_0|\tilde{\Psi}\rangle}|\tilde{\Psi}\rangle \tag{273}$$

which is now intermediately normalized. Put differently, the coupled cluster ansatz can parameterize any wave function that is not orthogonal to the chosen reference state  $|\Psi_0\rangle$ .

An important question to ask is how this ansatz affects our ability to express solutions to the Schrödinger equation. We recall the Schrödinger equation

$$H|\Psi\rangle = E|\Psi\rangle. \tag{274}$$

Assuming that  $\langle \Psi_0 | \tilde{\Psi} \rangle \neq 0$ , we define

$$|\Psi\rangle = \frac{1}{\langle \Psi_0 | \tilde{\Psi} \rangle} | \tilde{\Psi} \rangle = e^T | \Psi_0 \rangle \tag{275}$$

noting that  $|\Psi\rangle$  still solves the Schrödinger equation as eigenspaces are linear spaces. Hence

$$He^{T}|\Psi_{0}\rangle = Ee^{T}|\Psi_{0}\rangle.$$
(276)

**Theorem 9.3.** The linked coupled cluster equations are equivalent to solving the Schrödinger equation, *i.e.*,

$$He^{T}|\Psi_{0}\rangle = Ee^{T}|\Psi_{0}\rangle \quad \Leftrightarrow \quad \begin{cases} \langle\Psi_{0}|e^{-T}He^{T}|\Psi_{0}\rangle = E, \\ \langle\Psi_{\mu}|e^{-T}He^{T}|\Psi_{0}\rangle = 0, \quad \forall \mu \in \mathcal{I}. \end{cases}$$
(277)

*Proof.* Let's first assume that  $|\Psi\rangle$  solves the Schrödinger equation. Since the matrix exponential is invertible we may write Eq. (276) as

$$e^{-T}He^{T}|\Psi_{0}\rangle = E|\Psi_{0}\rangle \tag{278}$$

When we project onto  $|\Psi_0\rangle$  we recover the energy since  $||\Psi_0|| = 1$ , i.e.,

$$\langle \Psi_0 | e^{-T} H e^T | \Psi_0 \rangle = E | \Psi_0 \rangle.$$
(279)

Moreover, since  $0 = \langle \Psi_{\mu} | \Psi_0 \rangle = \langle \Psi_0 | X_{\mu} | \Psi_0 \rangle$  for all  $\mu \in \mathcal{I}$ , we get

$$\langle \Psi_{\mu} | e^{-T} H e^{T} | \Psi_{0} \rangle = 0.$$
(280)

We now consider that the linked coupled cluster equations are fulfilled. Note that the Slater determinants form a basis of  $\mathcal{H}^{(N)}$ , in particular,

$$I = |\Psi_0 \times \Psi_0| + \sum_{\mu \in \mathcal{I}} |\Psi_\mu \times \Psi_\mu|$$
(281)

Let  $|\Phi\rangle \in \mathcal{H}^{(N)}$  be arbitrary, we find

$$\langle \Phi | (H-E)e^{T} | \Psi_{0} \rangle = \langle \Phi | e^{T}e^{-T}(H-E)e^{T} | \Psi_{0} \rangle$$

$$= \langle \Phi | e^{T} | \Psi_{0} \rangle \langle \Psi_{0} | e^{-T}(H-E)e^{T} | \Psi_{0} \rangle$$

$$+ \sum_{\mu \in \mathcal{I}} \langle \Phi | e^{T} | \Psi_{\mu} \rangle \langle \Psi_{\mu} | e^{-T}(H-E)e^{T} | \Psi_{0} \rangle$$

$$= \langle \Phi | e^{T} | \Psi_{0} \rangle \underbrace{\langle \Psi_{0} | e^{-T}He^{T} | \Psi_{0} \rangle}_{=E} - E \langle \Phi | e^{T} | \Psi_{\mu} \rangle \underbrace{\langle \Psi_{\mu} | e^{-T}He^{T} | \Psi_{0} \rangle}_{=0} - E \sum_{\mu \in \mathcal{I}} \langle \Phi | e^{T} | \Psi_{\mu} \rangle \underbrace{\langle \Psi_{\mu} | \Psi_{0} \rangle}_{=0}$$

$$(282)$$

Hence,

$$\langle \Phi | (H-E)e^T | \Psi_0 \rangle = 0 \tag{283}$$

for all  $|\Phi\rangle \in \mathcal{H}^{(N)},$  which characterizes an eigenpair.

Remark 9.3.1. The trick to write

$$I = |\Psi_0 \rangle \langle \Psi_0| + \sum_{\mu \in \mathcal{I}} |\Psi_\mu \rangle \langle \Psi_\mu|$$
(284)

is known as resolution of identity.

#### 9.4 Size-extensivity

In quantum chemistry, we often need to compare energies of different systems. For example, the atomization energy is obtained by subtracting the energy of the molecule from those of its atoms. If we are to obtain reliable results, the energies of the molecule and its fragments must be calculated with comparable accuracy. More generally, our methods should be such that we may apply them to large and small systems alike and expect to obtain results of the same accuracy. Obviously, terms such as "comparable accuracy" are vague. Fortunately, there are situations where we may give a very precise meaning to this term, namely, for systems comprising two or more non-interacting subsystems. In such cases, our methods should generate the same energy for the system, irrespective of whether we have carried out the calculations for each subsystem separately or for all subsystems simultaneously. We summarize this requirement by stating that the calculations and, more generally, the computational methods should be *size-extensive*.

#### 9.4.1 Size-extensivity of the exact wave function

Consider a system of two non-interacting molecular fragments A and B. For such a system, the Hamiltonian operator may be written in the form

$$H^{(A\dots B)} = H^{(A)} + H^{(B)}$$
(285)

where  $H^{(A)}$  is associated with fragment A and  $H^{(B)}$  with fragment B. Note that the ladder operators of  $H^{(A)}$  anticommute with those of  $H^{(B)}$  since the associated orbitals are orthogonal, i.e.,

$$\left[\left(a_{p}^{(A)}\right)^{\dagger}, a_{q}^{(B)}\right]_{+} = 0 \tag{286}$$

**Remark 9.3.2.** The Hamiltonian in Eq. (285) is an idealized operator that cannot be realized in practice, but – in the limit of an infinite separation between the fragments – it provides an exact representation of the true operator.

The exact solution for the combined system represented by the Hamiltonian in Eq. (285) satisfies the Schrödinger equation:

$$H^{(A\cdots B)}|\Psi^{(A\cdots B)}\rangle = E^{(A\cdots B)}|\Psi^{(A\cdots B)}\rangle$$
(287)

We shall consider how this solution is related to the exact solutions for the fragments

$$H^{(A)}|\Psi^{(A)}\rangle = E^{(A)}|\Psi^{(A)}\rangle$$

$$H^{(B)}|\Psi^{(B)}\rangle = E^{(B)}|\Psi^{(B)}\rangle$$
(288)

Recall that

$$|\Psi^{(A)}\rangle = \left(c_0 I + \sum_{\mu \in \mathcal{I}^{(A)}} c_{\mu}^{(A)} X_{\mu}\right) |\Psi_0\rangle = \Omega_{\text{FCI}}^{(A)} |\Psi_0\rangle$$

$$|\Psi^{(B)}\rangle = \left(c_0 I + \sum_{\mu \in \mathcal{I}^{(B)}} c_{\mu}^{(B)} X_{\mu}\right) |\Psi_0\rangle = \Omega_{\text{FCI}}^{(B)} |\Psi_0\rangle$$

$$(289)$$

and note that

$$\Omega^{(A)}\Omega^{(B)} = \Omega^{(B)}\Omega^{(A)} \tag{290}$$

Then

$$H^{(A\cdots B)}\Omega^{(A)}\Omega^{(B)}|\Psi_{0}\rangle = (H^{(A)} + H^{(B)})\Omega^{(A)}\Omega^{(B)}|\Psi_{0}\rangle$$
  
$$= H^{(A)}\Omega^{(A)}\Omega^{(B)}|\Psi_{0}\rangle + H^{(B)}\Omega^{(A)}\Omega^{(B)}|\Psi_{0}\rangle$$
  
$$= \Omega^{(B)}H^{(A)}\Omega^{(A)}|\Psi_{0}\rangle + \Omega^{(A)}H^{(B)}\Omega^{(B)}|\Psi_{0}\rangle$$
  
$$= (E^{(A)} + E^{(B)})\Omega^{(B)}\Omega^{(A)}|\Psi_{0}\rangle$$
(291)

Note that we have here used that

$$\Omega_{\rm FCI}^{(A)}\Omega_{\rm FCI}^{(B)} = \Omega_{\rm FCI}^{(A\cdots B)}$$
(292)

is again a wave operator that is within the theoretical framework of the FCI ansatz. It is important to point out that this is not true for truncated CI. To see this, consider a system with N > 2 electrons, then

$$\Omega_{\text{CISD}}^{(A)} \Omega_{\text{CISD}}^{(B)} \neq \Omega_{\text{CISD}}^{(A \cdots B)}$$
(293)

cannot be generally true as the left-hand side contains four-fold excitations – since we are taking products of two-fold excitations. Hence, a sufficient requirement for size-extensivity is that the corresponding wave operator factorizes.

#### 9.4.2 Size-extensivity of the coupled-cluster ansatz

A key property of (truncated) coupled cluster ansatz as opposed to truncated configuration interaction ansatz, is that it is size consistent. This is an immediate consequence of cluster matrices being commutative, i.e.,

$$\Omega_{\rm CC}^{(A)}\Omega_{\rm CC}^{(B)} = e^{T^{(A)}}e^{T^{(B)}} = e^{T^{(A)}+T^{(B)}} = \Omega_{\rm CC}^{(AB)}$$
(294)

where  $T^{(A)} + T^{(B)}$  is not only a valid cluster operator, but it also maintains the potential truncated excitation rank imposed by the respective level of theory (vide infra).

#### 9.5 The coupled cluster working equations

We have seen that the coupled cluster parametrization is not a true exponential because the cluster matrices are nilpotent. Therefore, the coupled cluster working equations, i.e.,

$$0 = \langle \Psi_{\mu} | e^{-T} H e^{T} | \Psi_{0} \rangle \qquad \forall \mu \in \mathcal{I}$$
(295)

form a system of polynomial equations that we are trying to solve. Since cluster matrices are nilpotent to order N, the working equations Eq. (295) are polynomials at most of order 2N. However this simple estimation does not take further structure of the Hamiltonian into account. In particular, since the electronic structure Hamiltonian is a two-body operator, matrix elements corresponding to states that are more than two-fold excitations apart will be zero. This can be verified by direct computation via Wick's theorem and is more generally known as Slater–Condon rules:

For a one-body operator in an N-particle system, i.e.,

$$h = \sum_{pq} h_{pq} a_p^{\dagger} a_q \tag{296}$$

the matrix elements are given by

$$\langle \Psi | h | \Psi \rangle = \sum_{i=1}^{N} \langle \phi_i | h | \phi_i \rangle = \sum_{i=1}^{N} h_{ii}$$

$$\langle \Psi | h | \Psi_m^p \rangle = \langle \phi_m | h | \phi_p \rangle = h_{mp}$$

$$\langle \Psi | h | \Psi_{mn}^{pq} \rangle = 0.$$

$$(297)$$

For a two-body operator in an N-particle system, i.e.,

$$V = \sum_{pqrq} v_{pqrs} a_p^{\dagger} a_r^{\dagger} a_q a_s \tag{298}$$

the matrix elements are given by

$$\langle \Psi | V | \Psi \rangle = \frac{1}{2} \sum_{i,j=1}^{N} \langle ij | | ij \rangle$$

$$\langle \Psi | V | \Psi_m^p \rangle = \sum_{i=1}^{N} \left( \langle \phi_m \phi_i | \phi_p \phi_i \rangle - \langle \phi_m \phi_i | \phi_i \phi_p \rangle \right)$$

$$\langle \Psi | V | \Psi_{mn}^{pq} \rangle = \langle \phi_m \phi_n | \phi_p \phi_q \rangle - \langle \phi_m \phi_n | \phi_q \phi_p \rangle.$$
(299)

We may now apply these rules to the coupled cluster equations. We first note that the similarity transformed Hamiltonian can be expressed using a series expansion known as the Hadamard lemme, which is a consequence of the Baker–Campbell–Hausdorff (BCH) formula. More precisely,

$$e^{-T}He^{T} = \sum_{n \ge 0} \frac{1}{n!} [H, T]_{n}$$
(300)

where

$$[H,T]_n = [[...[[H,T],T],....],T]$$
(301)

are n-fold nested commutators. Applying the Slater-Condon rules, we find that the only non-zero contributions in Eq. (300) are given by

$$\langle \Psi_{\mu} | e^{-T} H e^{T} | \Psi_{0} \rangle = \sum_{n=0}^{4} \langle \Psi_{\mu} | [H, T]_{n} | \Psi_{0} \rangle.$$
(302)

More explicitly, we derived that

$$e^{-T}He^{T} = H + [H,T] + \frac{1}{2}[[H,T],T] + \frac{1}{6}[[[H,T],T],T] + \frac{1}{24}[[[H,T],T],T],T]$$
(303)

where

$$[H, T] = HT - TH,$$

$$[[H, T], T] = HT^{2} - 2THT + T^{2}H,$$

$$[[[H, T], T], T] = HT^{3} - 3THT^{2} + 3T^{2}HT - T^{3}H,$$

$$[[[[H, T], T], T], T] = HT^{4} - 4THT^{3} + 6T^{2}HT^{2} - 4T^{3}HT + T^{4}H.$$
(304)

Hence, the coupled cluster working equations are a system of polynomial equations at most of degree four.

#### 9.5.1 Truncated coupled cluster methods

Up to this point, we introduced a non-linear parametrization of the wave function that is of equal size as the corresponding linear system. From a numerical perspective, this is a bad idea since the complexity of solving non-linear systems is higher. However, when truncating the coupled cluster ansatz, this procedure becomes computationally more feasible while providing high-accuracy results which can be arguably accounted to the non-linear parametrization.

Truncations in the coupled cluster theory correspond to sparsity enforcement in the cluster amplitude vector, i.e.,

$$T(\mathbf{t}) = \sum_{\mu \in \mathcal{I}} t_{\mu} X_{\mu} \approx \sum_{\mu \in \tilde{\mathcal{I}}} t_{\mu} X_{\mu} = T((\tilde{\mathbf{t}}))$$
(305)

where  $\tilde{\mathcal{I}} \subset \mathcal{I}$ . In general, there are no limitations on how we pick the multi-index set  $\tilde{\mathcal{I}}$ , however, in practice, the most commonly used ones are

$$\mathcal{I}_{SD} = \left\{ \mu \mid |\mu| \leq 2 \right\} \quad \text{and} \quad \mathcal{I}_{D} = \left\{ \mu \mid |\mu| = 2 \right\}$$
(306)

which correspond to the coupled cluster with singles and doubles (CCSD) and coupled cluster with doubles (CCD) variants, respectively. With this restriction of the cluster amplitudes, the working equations become over-determined. We, therefore, always restrict the projective equations to the same set of excitations that are involved in the cluster expansion, i.e., we solve

$$0 = \langle \Psi_{\mu} | e^{-T(\mathbf{t})} H | e^{T(\mathbf{t})} | \Psi_0 \rangle \qquad \forall \mu \in \tilde{\mathcal{I}}.$$
(307)

The CCD working equations: We consider the cluster expansion

$$T_D = \sum_{i,j \in \mathcal{V}_{\text{occ}}} \sum_{a,b \in \mathcal{V}_{\text{virt}}} t_{ij}^{ab} X_{ij}^{ab}$$
(308)

A vanilla version of the working equations that are implemented in quantum-chemistry software packages can now be derived via direct application of Wick's theorem and the Slater-Condon rules. We illustrate this at hand of a few selected terms: We note that the zeroth order term is directly given by the Slater-Condon rule, namely,

$$\langle \Psi_{kl}^{cd} | H | \Psi_0 \rangle = \langle \phi_k \phi_l | \phi_c \phi_d \rangle - \langle \phi_k \phi_l | \phi_d \phi_c \rangle \tag{309}$$

Before proceeding to the remaining terms, we first note that in this approximation only few terms in the BCH expansion contribute:

$$\langle \Psi_{kl}^{cd} | [H, T_D] | \Psi_0 \rangle = \langle \Psi_{kl}^{cd} | HT_D - T_D H | \Psi_0 \rangle,$$
  
$$\langle \Psi_{kl}^{cd} | [[H, T_D], T_D] | \Psi_0 \rangle = \langle \Psi_{kl}^{cd} | HT_D^2 - 2T_D H T_D | \Psi_0 \rangle.$$
 (310)

We may now take a closer look at the first order term. We observe

$$\langle \Psi_{kl}^{cd} | HT | \Psi_0 \rangle - \langle \Psi_{kl}^{cd} | TH | \Psi_0 \rangle = \sum_{i,j} \sum_{a,b} t_{ij}^{ab} \langle \Psi_{kl}^{cd} | HX_{ij}^{ab} | \Psi_0 \rangle - \sum_{i,j} \sum_{a,b} t_{ij}^{ab} \langle \Psi_{kl}^{cd} | X_{ij}^{ab} H | \Psi_0 \rangle$$

$$= \sum_{i,j} \sum_{a,b} t_{ij}^{ab} \langle \Psi_{kl}^{cd} | H | \Psi_{ij}^{ab} \rangle - t_{kl}^{cd} \langle \Psi_0 | H | \Psi_0 \rangle$$

$$(311)$$

where now have to compute

$$\langle \Psi_{kl}^{cd} | H | \Psi_{ij}^{ab} \rangle = \sum_{pq} h_{pq} \langle \Psi_{kl}^{cd} | a_p^{\dagger} a_q | \Psi_{ij}^{ab} \rangle + \frac{1}{2} \sum_{pqra} v_{pqrs} \langle \Psi_{kl}^{cd} | a_p^{\dagger} a_r^{\dagger} a_q a_s | \Psi_{ij}^{ab} \rangle$$

$$= \sum_{pq} h_{pq} \langle \Psi_0 | a_l^{\dagger} a_k^{\dagger} a_c a_d a_p^{\dagger} a_q a_b^{\dagger} a_a^{\dagger} a_i a_j | \Psi_0 \rangle$$

$$+ \frac{1}{2} \sum_{pqra} v_{pqrs} \langle \Psi_0 | a_l^{\dagger} a_k^{\dagger} a_c a_d a_p^{\dagger} a_r^{\dagger} a_q a_s a_b^{\dagger} a_a^{\dagger} a_i a_j | \Psi_0 \rangle$$

$$(312)$$

Recall that by Corollary 8.1.1, only fully contracted terms contribute to the vacuum expectation values. Hence

$$\langle \Psi_{0} | a_{l}^{\dagger} a_{k}^{\dagger} a_{c} a_{d} a_{p}^{\dagger} a_{q} a_{b}^{\dagger} a_{a}^{\dagger} a_{i} a_{j} | \Psi_{0} \rangle = \langle \Psi_{0} | \{ a_{l}^{\dagger} a_{k}^{\dagger} a_{c} a_{d} a_{p}^{\dagger} a_{q} a_{b}^{\dagger} a_{a}^{\dagger} a_{i} a_{j} \}_{0} | \Psi_{0} \rangle$$

$$+ \langle \Psi_{0} | \{ a_{l}^{\dagger} a_{k}^{\dagger} a_{c} a_{d} a_{p}^{\dagger} a_{q} a_{b}^{\dagger} a_{a}^{\dagger} a_{i} a_{j} \}_{0} | \Psi_{0} \rangle$$

$$+ \langle \Psi_{0} | \{ a_{l}^{\dagger} a_{k}^{\dagger} a_{c} a_{d} a_{p}^{\dagger} a_{q} a_{b}^{\dagger} a_{a}^{\dagger} a_{i} a_{j} \}_{0} | \Psi_{0} \rangle$$

$$+ \langle \Psi_{0} | \{ a_{l}^{\dagger} a_{k}^{\dagger} a_{c} a_{d} a_{p}^{\dagger} a_{q} a_{b}^{\dagger} a_{a}^{\dagger} a_{i} a_{j} \}_{0} | \Psi_{0} \rangle$$

$$= \dots$$

$$(313)$$

# 9.6 CCSD(T) – The Gold Standard of Quantum Chemistry

### 10 Molecular Properties and Response Theory

Consider an isolated molecular system governed by a Hamiltonian  $H_0$ . We introduce a static perturbation V(x) that depends on a parameter x, which represents the strength of the perturbation. By definition, the perturbation vanishes when x = 0. Consequently, the perturbed Hamiltonian is given by

$$H(x) = H_0 + V(x)$$
(314)

and the total energy can be expanded with respect to x as

$$\mathcal{E}(x) = \frac{\langle \Psi(x) | H(x) | \Psi(x) \rangle}{\langle \Psi(x) | \Psi(x) \rangle} = \mathcal{E}^{(0)} + \mathcal{E}^{(1)}x + \frac{1}{2}\mathcal{E}^{(2)}x^2 + \dots$$
(315)

where  $\Psi(x)$  is the wave function of the state considered, which is usually the ground state. In this expansion,  $\mathcal{E}^{(0)} = E_0$  is the unperturbed energy associated to the Hamiltonian  $H_0$ , and the energy derivatives

$$\mathcal{E}^{(1)} = \frac{d\mathcal{E}}{dx}\Big|_{x=0} ; \quad \mathcal{E}^{(2)} = \frac{d^2\mathcal{E}}{dx^2}\Big|_{x=0} ; \quad \mathcal{E}^{(3)} = \frac{d^3\mathcal{E}}{dx^3}\Big|_{x=0} ; \quad \text{etc.}$$
(316)

are called time-independent molecular properties. They are characteristic of the molecule and its quantum state, and contain important information about the response of the system to the perturbation.

#### 10.1 Geometric derivatives

One of the most significant examples of a static perturbation, within the Born–Oppenheimer approximation, is a deformation of the molecular geometry. In this context, the parameter  $\delta \mathbf{R}$  describes changes in the positions of the nuclei:

$$\mathcal{E}(\mathbf{R}) = \mathcal{E}^{(0)} + \sum_{i} \mathcal{E}_{i}^{(1)} \delta \mathbf{R}_{i} + \frac{1}{2} \sum_{i,j} \delta \mathbf{R}_{i} \mathcal{E}_{i,j}^{(2)} \delta \mathbf{R}_{j} + \dots$$
(317)

we call

$$\mathcal{E}^{(1)} = \nabla_{\mathbf{R}} \mathcal{E} \tag{318}$$

the molecular gradient, and

$$\mathcal{E}_{i,j}^{(2)} = \frac{d^2 \mathcal{E}}{d\mathbf{R}_i \mathbf{R}_j} \tag{319}$$

is called the molecular Hessian. These quantities are used in geometry optimization for locating and characterizing critical points on the molecular potential energy surface (energy minimum, saddle point at a transition state). There are also used for calculating spectroscopic constants such as harmonic vibrational frequencies.

#### **10.2** Electric properties

In the context of electronic properties, a central object is the *charge density*. It is the measure of the probability of an electron being present at an infinitesimal element of space surrounding any given point. Given the wavefunction  $|\Psi\rangle \in \mathcal{H}^{(N)}$  we may define this via

$$n(\mathbf{r}) = \langle \Psi | \hat{\rho}(\mathbf{r}) | \Psi \rangle \tag{320}$$

where

$$\hat{\rho}(\mathbf{r}) = \sum_{i=1}^{N} \delta(\mathbf{r} - \mathbf{r}_i).$$
(321)

This can be simplified straight-forwardly to

$$n(\mathbf{r}) = N \sum_{s_1} \dots \sum_{s_N} \int_{\mathbb{R}^3} \dots \int_{\mathbb{R}^3} |\Psi(\mathbf{r}, s_1; \mathbf{r}_2, s_2; \dots; \mathbf{r}_N, s_N)|^2 \, \mathrm{d}\mathbf{r}_2 \dots \mathrm{d}\mathbf{r}_N.$$
(322)

Moreover, when  $|\Psi\rangle$  takes the form of a single Slater determinant, this further simplifies to

$$n(\mathbf{r}) = \sum_{i=1}^{N} |\phi_i(\mathbf{r})|^2.$$
 (323)

In order to compute electronic properties, the perturbation applied is an external electrostatic potential  $v(\mathbf{r})$ . The corresponding interaction energy then reduces to

$$E_{\rm int} = \int n(\mathbf{r})v(\mathbf{r}) \, \mathrm{d}\mathbf{r} \tag{324}$$

We then expand the potential around  $\mathbf{r}_0$ 

$$v(\mathbf{r}) = v(\mathbf{r}_0) + \sum_i \frac{\partial v}{\partial r_i} \Big|_{\mathbf{r}=\mathbf{r}_0} r_i + \frac{1}{2} \sum_{i,j} \frac{\partial^2 v}{\partial r_i \partial r_j} \Big|_{\mathbf{r}=\mathbf{r}_0} r_i r_j + \dots$$
  
=:  $v(\mathbf{r}_0) - \sum_i E_i r_i - \frac{1}{2} \sum_{i,j} F_{i,j} r_i r_j + \dots$  (325)

where  $v(\mathbf{r}_0)$  is a constant potential, **E** is the corresponding electric field, and **F** is the electric field gradient defined as

$$E_{i} = -\frac{\partial v}{\partial r_{i}}\Big|_{\mathbf{r}=\mathbf{r}_{0}} \quad \text{and} \quad F_{i,j} = -\frac{\partial^{2} v}{\partial r_{i} \partial r_{j}}\Big|_{\mathbf{r}=\mathbf{r}_{0}}$$
(326)

The interaction energy can thus be expanded as

$$E_{\text{int}} = v(\mathbf{r}_0) \int n(\mathbf{r}) \, \mathrm{d}\mathbf{r} + E_i \sum_i \int n(\mathbf{r}) r_i \, \mathrm{d}\mathbf{r} + \frac{1}{2} F_{i,j} \sum_{i,j} \int n(\mathbf{r}) r_i r_j \, \mathrm{d}\mathbf{r} + \dots$$

$$= v(\mathbf{r}_0) q - \sum_i E_i \mu_i - \frac{1}{2} \sum_{i,j} F_{i,j} Q_{i,j} + \dots$$
(327)

where q is the electric charge of the system,  $\mu$  is the electric dipole moment, and  $\mathbf{Q}$  is the electric quadrupole moment

$$r = \int n(\mathbf{r}) \, \mathrm{d}\mathbf{r}$$
  

$$\mu_i = \int n(\mathbf{r}) r_i \, \mathrm{d}\mathbf{r}$$
  

$$Q_{i,j} = \int n(\mathbf{r}) r_i r_j \, \mathrm{d}\mathbf{r}$$
(328)

Comparing these quantities with Eq. (315), we note that these properties can be defined as energy derivatives:

$$q = \frac{\partial E_{\text{int}}}{\partial v(\mathbf{r}_0)}$$

$$\mu_i = -\frac{\partial E_{\text{int}}}{\partial E_i}$$

$$Q_{i,j} = -2\frac{\partial E_{\text{int}}}{\partial F_{i,j}}$$
(329)

We can further decompose the dipole moment into permanent and field-induced contributions

$$\mu_i = \mu_{0,1} + \sum_j \alpha_{i,j} E_j + \dots$$
(330)

which leads to the definition of the *permanent electric dipole moment* 

$$\mu_{0,1} = -\frac{\partial E_{\text{int}}}{\partial E_i} \bigg|_{\mathbf{E}=\mathbf{0}}$$
(331)

and the *electric dipole polarizability* 

$$\alpha_{i,j} = -\frac{\partial E_{\text{int}}}{\partial E_i \partial E_i} \bigg|_{\mathbf{E}=\mathbf{0}}$$
(332)

#### 10.2.1 Magnetic properties

Skip this for now...

#### **10.3** Computation of properties

There are two ways of calculating derivatives: (1) by numerical differentiation and (2) by analytical differentiation. In numerical differentiation, the derivatives are calculated by finite differences or by polynomial fitting. In a nutshell:

- It is simple to implement.
- The numerical precision is limited.
- The computational efficiency is low.
- It is not general. Usually, only real-valued and static perturbations can be done.

In analytical differentiation, the derivatives are calculated explicitly from the analytical expressions. In a nutshell:

- It is (potentially) difficult to implement.
- The precision is higher.
- The computational efficiency is higher.
- It is more general. In particular, frequency-dependent perturbations are possible.

#### 10.3.1 Sum-over-state expression for exact wave functions

For "exact" wave functions (full configuration interaction in a basis), we can use straightforward perturbation theory to find the expression of the energy derivatives. Expanding the perturbation operator V(x) in powers of x, we have

$$H(x) = H_0 + V^{(1)}x + \frac{1}{2}V^{(2)}x^2 + \dots$$
(333)

The first-order energy derivative is

Ε

$$\mathcal{E}^{(1)} = \langle \Psi_0 | V^{(1)} | \Psi_0 \rangle \tag{334}$$

where  $|\Psi_0\rangle$  is the exact wave function of the unperturbed Hamiltonian  $H_0$  for the state considered, normalized such that  $\langle \Psi_0 | \Psi_0 \rangle = 1$ . Similarly, the second-order energy derivative is found from the first-order wave function correction, i.e.,

$$|\Psi^{(1)}\rangle = \sum_{n \neq 0} \frac{\langle \Psi_n | V^{(1)} | \Psi_0 \rangle}{\mathcal{E}_0 - \mathcal{E}_n} | \Psi_n \rangle$$
(335)

which yields

where  $|\Psi_n\rangle$  and  $\mathcal{E}_n$  form a set of complete exact eigenstates and associated eigenvalues of  $H_0$ . This last expression is very impractical since we need to know all the exact states of  $H_0$ , which is a complicated many-body Hamiltonian. Moreover, for approximate methods such as Hartree-Fock, it is not clear what we should use for  $|\Psi_n\rangle$  and  $\mathcal{E}_n$ .

#### 10.3.2 General expressions for approximate methods

In an approximate electronic structure method, the total energy  $\mathcal{E}(x)$  is obtained by optimizing parameters  $\mathbf{p} = (p_1, p_2...)$  in an energy function  $E(x, \mathbf{p})$  for each fixed value of x. The final total energy  $\mathcal{E}(x)$  is obtained for the optimal value of the parameters  $\mathbf{p}_*(x)$ , which are functions of x,

$$\mathcal{E}(x) = E(x, p_*(x)). \tag{337}$$

Note that the optimization is not necessarily variational, i.e., the optimization criterion is not necessarily to minimize E(x, p) with respect to the parameters p.

#### Some examples:

- Hatree-Fock (KS-DFT): The parameters **p** are the orbital parameters. (variational)
- Configuration interaction (CI): The parameters **p** are the configuration coefficients. (variational)
- Coupled cluster (CC): The parameters **p** are the cluster amplitudes. (not variational)

#### 10.3.3 First-order energy derivative

The first-order derivative of  $\mathcal{E}(x)$  with respect to x is made of two terms

$$\frac{d\mathcal{E}(x)}{dx} = \frac{\partial E(x, \mathbf{p}_{*}(x))}{\partial x} + \sum_{i} \frac{\partial E(x, \mathbf{p})}{\partial p_{i}} \bigg|_{\mathbf{p}=\mathbf{p}_{*}(x)} \frac{\partial [\mathbf{p}_{*}(x)]_{i}}{\partial x}$$
(338)

where the first term constitutes the explicit dependence on x and the second term constitutes the implicit dependence on x. The derivative

$$\frac{\partial [\mathbf{p}_*(x)]_i}{\partial x} =: [\mathcal{L}(x)]_i \tag{339}$$

is called the wave-function linear-response vector and contains information about how the electronic structure changes when the system is perturbed. It is not straightforward to calculate it since we do not know the explicit dependence of  $p_*$  on x.

Note that when all parameters are variational, then there is an important simplification: The stationary condition (zero electronic gradient)

$$\left. \frac{\partial E(x, \mathbf{p})}{\partial p_i} \right|_{\mathbf{p}=\mathbf{p}_*(x)} = 0 \tag{340}$$

implies that the second term in Eq. (338) vanishes, so the first-order energy derivative reduces to

$$\frac{d\mathcal{E}(x)}{dx} = \frac{\partial E(x, \mathbf{p}_*(x))}{\partial x}$$
(341)

and we do not need the wave-function linear-response vector.

For example, for the calculation of the HF molecular gradient, we need to consider only the explicit dependence on the nuclear positions in the Hamiltonian and in the wave function, but not the implicit dependence of the orbital coefficients on the nuclear positions since these parameters are variational.

#### 10.3.4 Second-order energy derivative

We will only consider the case where all the parameters are variational. In this case, the second-order derivative of  $\mathcal{E}(x)$  is obtained by taking the derivative of Eq. (341), i.e.,

$$\frac{d^2 \mathcal{E}(x)}{dx^2} = \frac{\partial^2 E(x, \mathbf{p}_*(x))}{\partial x^2} + \sum_i \frac{\partial^2 E(x, \mathbf{p})}{\partial x \partial p_i} \bigg|_{\mathbf{p} = \mathbf{p}_*(x)} \frac{\partial [\mathbf{p}_*(x)]_i}{\partial x}$$
(342)

where the perturbed electronic gradient

$$\frac{\partial^2 E(x, \mathbf{p}_*(x))}{\partial x \partial p_i} \Big|_{\mathbf{p}=\mathbf{p}_*(x)}$$
(343)

is not zero. We thus now need the wave-function linear-response vector. Note that we need only the first-order derivative of the wave function to calculate the second-order derivative of the energy. More generally, we have the so-called 2n + 1 rule:

For variational parameters, the derivatives of the wave function to order n determine the energy derivatives to order 2n + 1.

#### 10.3.5 Linear response equations

In order to obtain the wave-function linear-response vector  $\mathcal{L}(x)$  in the case where all the parameters are variational, we start by noting that the stationary condition is true for all x, i.e.,

$$0 = \frac{\partial E(x, \mathbf{p})}{\partial p_i} \bigg|_{\mathbf{p} = \mathbf{p}_*(x)} \qquad \forall x,$$
(344)

which means that we can take the first-order derivative of Eq. (344) with respect to x:

$$\frac{\partial^2 E(x,\mathbf{p})}{\partial x \partial p_i} \bigg|_{\mathbf{p}=\mathbf{p}_*(x)} + \sum_j \frac{\partial^2 E(x,\mathbf{p})}{\partial p_i \partial p_j} \bigg|_{\mathbf{p}=\mathbf{p}_*(x)} \frac{\partial [\mathbf{p}_*(x)]_j}{\partial x} = 0.$$
(345)

We thus arrive at the *linear response equations* 

$$\sum_{j} \frac{\partial^{2} E(x, \mathbf{p})}{\partial p_{i} \partial p_{j}} \Big|_{\mathbf{p}=\mathbf{p}_{*}(x)} [\mathcal{L}(x)]_{j} = \sum_{j} \frac{\partial^{2} E(x, \mathbf{p})}{\partial p_{i} \partial p_{j}} \Big|_{\mathbf{p}=\mathbf{p}_{*}(x)} \frac{\partial [\mathbf{p}_{*}(x)]_{j}}{\partial x} = -\frac{\partial^{2} E(x, \mathbf{p})}{\partial x \partial p_{i}} \Big|_{\mathbf{p}=\mathbf{p}_{*}(x)},$$
(346)

which is a linear system of equations whose solution gives the wave-function linear-response vector  $\mathcal{L}(x)$ .

Remember that, in practice, we are interested in energy derivatives evaluated at x = 0. Therefore, these equations must in fact be solved using the unperturbed electronic Hessian

$$\frac{\partial^2 E(x=0,\mathbf{p})}{\partial p_i \partial p_j} \tag{347}$$

(independent of the perturbation). Since its dimensions are usually large, the response equations are usually solved iteratively without explicitly constructing and storing the Hessian matrix.

#### 10.3.6 Lagrangian formalism for non-variational parameters

For methods with non-variational parameters (CI, CC, MP2, ...), we can simplify the calculation by using the technique of Lagrange's multipliers.

... Include CC here?

#### 10.4 Example of linear-response equations for Hartree-Fock

#### 10.4.1 Exponential parametrization

For optimizing the orbitals in the HF determinant wave function, it is convenient to use an exponential parametrization

$$|\Phi(\boldsymbol{\kappa})\rangle = e^{\boldsymbol{\kappa}}|\Phi_0\rangle \tag{348}$$

where  $e^{\kappa}$  is a unitary operator performing rotations between occupied and virtual spin orbitals in a reference determinant wave function  $|\Phi_0\rangle$ . This rotation operator is constructed from an anti Hermitian single-excitation operator,  $\kappa = -\kappa^{\dagger}$ , which can be written in the secondquantization formalism as

$$\boldsymbol{\kappa} = \sum_{i} \sum_{a} \left( \boldsymbol{\kappa}_{ai} a_{a}^{\dagger} a_{i} - \boldsymbol{\kappa}_{ai}^{*} a_{i}^{\dagger} a_{a} \right)$$
(349)

where  $a_k^{\dagger}$  and  $a_k$  are the creation and annihilation operators of the spin orbital k, respectively, and the indices a and r refer to occupied and virtual spin orbitals in the reference determinant, respectively. The parameters  $\kappa = \kappa_{ar}$  are called the orbital rotation parameters. In comparison to the orbital coefficients on the atomic basis functions, the orbital rotation parameters have the advantage of providing a non-redundant parametrization of the wave function so that one can vary them independently without having to impose any constraints. At each step of the optimization, the orbitals in the reference determinant  $|\Phi_0\rangle$  are updated so that one always considers variations of the orbital rotation parameters around  $\kappa = 0$ . The expansion of the HF wave function with respect to  $\kappa$  thus writes

$$|\Phi(\boldsymbol{\kappa})\rangle = \left(I + \boldsymbol{\kappa} + \frac{1}{2}\boldsymbol{\kappa}^{2} + ...\right)|\Phi_{0}\rangle$$
  
$$= |\Phi_{0}\rangle + \sum_{a}\sum_{i}\kappa_{ai}|\Phi_{i}^{a}\rangle + \frac{1}{2}\sum_{a,b}\sum_{i,j}\kappa_{ai}\kappa_{bj}|\Phi_{ij}^{ab}\rangle + \frac{1}{2}\sum_{a}\sum_{i}\kappa_{ai}^{*}\kappa_{ai}|\Phi_{ij}^{ab}\rangle + ...$$
(350)

where  $|\Phi_i^a\rangle = a_a^{\dagger}a_i|\Phi_0\rangle$  is a singly excited determinant and  $|\Phi_{ij}^{ab}\rangle = a_b^{\dagger}a_a^{\dagger}a_ia_j|\Phi_0\rangle$  is a doubly excited Slater determinant.

#### 10.4.2 Electronic gradient and Hessian

We now want to calculate the HF electronic gradient and Hessian, i.e., the first- and secondorder derivatives of the HF total energy

$$E_{\rm HF}(\boldsymbol{\kappa}) = \frac{\langle \Phi(\boldsymbol{\kappa}) | H_0 | \Phi(\boldsymbol{\kappa}) \rangle}{\langle \Phi(\boldsymbol{\kappa}) | \Phi(\boldsymbol{\kappa}) \rangle}$$
(351)

with respect to  $\kappa$ . Here, we will only consider the case of real-valued orbitals and real-valued orbital parameters  $\kappa_{ar}^* = \kappa_{ar}$ . In this case, we find

$$\frac{\partial |\Phi(\boldsymbol{\kappa})\rangle}{\partial \boldsymbol{\kappa}_{ai}}\Big|_{\boldsymbol{\kappa}=\boldsymbol{0}} = |\Phi_i^a\rangle \tag{352}$$

and

$$\frac{\partial^2 |\Phi(\kappa)\rangle}{\partial \boldsymbol{\kappa}_{ai} \boldsymbol{\kappa}_{bj}}\Big|_{\boldsymbol{\kappa}=\mathbf{0}} = |\tilde{\Phi}_{ij}^{ab}\rangle = \begin{cases} |\Phi_{ij}^{ab}\rangle, & \text{if } a \neq b \text{ or } i \neq j \\ -|\Phi_0\rangle, & \text{if } a = b \text{ and } i = j \end{cases}$$
(353)

Assuming that the wavefunction  $|\Phi_0\rangle$  is normalized, this yields

$$\frac{\partial E_{\rm HF}(\boldsymbol{\kappa})}{\partial \boldsymbol{\kappa}_{ai}}\Big|_{\boldsymbol{\kappa}=0} = 2\langle \Phi_i^a | H_0 | \Phi_0 \rangle.$$
(354)

Thus, at convergence,

$$\langle \Phi_i^a | H_0 | \Phi_0 \rangle = 0 \tag{355}$$

which is known as Brillouin's theorem. For the Hessian, we find

$$\frac{\partial^2 E_{\rm HF}(\boldsymbol{\kappa})}{\partial \boldsymbol{\kappa}_{ai} \partial \boldsymbol{\kappa}_{bj}} \bigg|_{\boldsymbol{\kappa}=0} = 2(A_{ai,bj} + B_{ai,bj})$$
(356)

where

$$A_{ai,bj} = \langle \Phi_i^a | H_0 - E_{\rm HF} | \Phi_j^b \rangle$$
  

$$B_{ai,bj} = \langle \tilde{\Phi}_{ij}^{ab} | H_0 - E_{\rm HF} | \Phi_j^b \rangle = \langle \Phi_{ij}^{ab} | H_0 - E_{\rm HF} | \Phi_j^b \rangle$$
(357)

with  $E_{\rm HF} = E_{\rm HF}(\mathbf{0})$ .

# 11 Special topics

- 11.1 Tailored coupled cluster
- 11.2 Green's functions