

1 Hartree-Fock Theory

The goal of electronic structure theory is to solve the “clamped-nuclei” Schrödinger equation

$$\hat{H}\Psi_k = E_k\Psi_k \quad \hat{H} = V_{\text{nuc}} + \hat{H}_e = \sum_{a<b}^{\text{nuc.}} \frac{Z_a Z_b}{|\mathbf{R}_a - \mathbf{R}_b|} - \frac{1}{2} \sum_i^{\text{elec.}} \nabla_i^2 - \sum_a^{\text{nuc.}} \sum_i^{\text{elec.}} \frac{Z_a}{|\mathbf{R}_a - \mathbf{r}_i|} + \sum_{i<j}^{\text{elec.}} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \quad (1.1)$$

with an optimal balance of accuracy and efficiency for the problem of interest. The most accurate solution possible for a given atomic orbital (AO) basis set¹ results from expanding the wavefunction

$$\Psi_k = \sum_{\mu} \Phi_{\mu} c_{\mu k} \quad (1.2)$$

in terms of all possible Slater determinants (n -electron basis functions $\{\Phi_{\mu}\}$) that can be formed from an orthonormal one-electron basis of spin-orbitals, $\{\psi_p\}$. The expansion coefficients $(\mathbf{c})_k = c_{\mu k}$ are eigenvectors of the matrix $(\mathbf{H})_{\mu\nu} = \langle \Phi_{\mu} | \hat{H} | \Phi_{\nu} \rangle$, which is the matrix representation of the Hamiltonian in the determinant basis. This is called the *full configuration-interaction* (FCI) solution.

Any one-electron basis spans the same “function space” as the AO basis set itself, and the full n -electron basis $\{\Phi_{\mu}\}$ spans the same space of n -electron functions regardless of how one forms spin orbitals from the AO basis set. As a result, one obtains the same FCI solution for any choice of spin-orbitals. In general, however, FCI solutions are completely unfeasible for basis sets of sufficient size to approach the complete basis set limit. One can think of this as a simple counting problem: if there are m functions in the AO basis, then there are $2m$ spin-orbitals in the one-electron basis,² and there are “ $2m$ choose n ”³

$$\binom{2m}{n} \equiv \frac{(2m)!}{n!(2m-n)!}$$

unique Slater determinants in the n -electron basis that can be formed from the spin MOs. The upshot is that we usually have to omit some Slater determinants in order get an answer in a reasonable amount of time.

As soon as we truncate our determinant expansion (1.2), our choice of spin MOs makes a significant difference in the quality of our results. In particular, we need to choose our set of one-electron functions to minimize the number of Slater determinants it takes to “get close to” the exact wavefunction.

1.1 Defining the problem

It can be shown that optimizing $\langle \Psi | \hat{H}_e | \Psi \rangle$ by varying Ψ subject to the normalization constraint $\langle \Psi | \Psi \rangle = 1$ is equivalent to solving the Schrödinger equation. When we further constrain the form of Ψ this is no longer true, but it *does* generally allow us to get the best approximation to Ψ for a given approach (or “Ansatz”).

In order to make the wavefunction expansion converge with a relatively small number of Φ_{μ} s, we wish to find the best single-determinant approximation to Ψ . That is, we wish to optimize

$$\langle \Phi | \hat{H}_e | \Phi \rangle \quad \Phi(1, \dots, n) = \frac{1}{\sqrt{n!}} \begin{vmatrix} \psi_1(1) & \psi_2(1) & \cdots & \psi_n(1) \\ \psi_1(2) & \psi_2(2) & \cdots & \psi_n(2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_1(n) & \psi_2(n) & \cdots & \psi_n(n) \end{vmatrix} \quad (1.3)$$

with respect to variation of the orbitals $\{\psi_p\}$, enforcing the normalization constraint by keeping the spin orbitals orthonormal. Note that here the function argument (i) is shorthand for (\mathbf{r}_i, s_i) where \mathbf{r}_i denotes the position of the i^{th} electron and s_i denotes its spin. This optimization problem is the essence of *Hartree-Fock theory*.

¹cc-pVXZ, 6-31G, ANO1, etc.

² m α -orbitals and m β -orbitals.

³The number of unique sets of n marbles that can be drawn from a bag of $2m$ marbles. See <http://en.wikipedia.org/wiki/Combination>

Once we have solved for the Hartree-Fock optimization problem, the expectation value $\langle \Phi | \hat{H}_e | \Phi \rangle$ is itself a good first approximation to the electronic energy. More importantly, however, when we use this new set of Hartree-Fock spin-orbitals, $\{\psi_p\}$, the FCI expansion tends to converge much more quickly to the true wavefunction. Specifically, when we rewrite equation 1.2 in terms of single $\{\Phi_i^a\}$, double $\{\Phi_{ij}^{ab}\}$, triple $\{\Phi_{ijk}^{abc}\}$, etc. replacements⁴ of the orbitals in the Hartree-Fock determinant Φ with the remaining orbitals in the basis

$$\Psi = \Phi + \sum_i \Phi_i^a c_a^i + \sum_{\substack{a < b \\ i < j}} \Phi_{ij}^{ab} c_{ab}^{ij} + \sum_{\substack{a < b < c \\ i < j < k}} \Phi_{ijk}^{abc} c_{abc}^{ijk} + \dots \quad (1.4)$$

the coefficients tend to be very small, and are often virtually negligible for higher than quadruple replacements.

1.2 Deriving the Hartree-Fock equations

The electronic Hamiltonian \hat{H}_e contains one- and two-electron operators.

$$\hat{H}_e = \sum_i \hat{h}(i) + \sum_{i < j} \hat{g}(i, j) \quad \hat{h}(i) \equiv -\frac{1}{2} \nabla_i^2 + \sum_a \frac{Z_a}{|\mathbf{r}_i - \mathbf{R}_a|} \quad \hat{g}(i, j) \equiv \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \quad (1.5)$$

Its expectation value with respect to a single determinant Φ is given by the first Slater rule

$$\langle \Phi | \hat{H}_e | \Phi \rangle = \sum_i^n \langle \psi_i | \hat{h} | \psi_i \rangle + \frac{1}{2} \sum_{ij}^n \langle \psi_i \psi_j | | \psi_i \psi_j \rangle \quad \langle \psi_p \psi_q | | \psi_r \psi_s \rangle \equiv \langle \psi_p \psi_q | \psi_r \psi_s \rangle - \langle \psi_p \psi_q | \psi_s \psi_r \rangle \quad (1.6)$$

where the one- and two-electron integrals are defined as follows.

$$\langle \psi_p | \hat{h} | \psi_q \rangle \equiv \int d(1) \psi_p^*(1) \hat{h}(1) \psi_q(1) \quad \langle \psi_p \psi_q | | \psi_r \psi_s \rangle \equiv \int d(1) d(2) \psi_p^*(1) \psi_q^*(2) \hat{g}(1, 2) \psi_r(1) \psi_s(2) \quad (1.7)$$

We wish to optimize equation 1.6 with respect to $\{\psi_i\}$, subject to an orthonormality constraint.

$$\langle \psi_i | \psi_j \rangle \stackrel{!}{=} \delta_{ij} \quad (1.8)$$

The corresponding Lagrangian functional is

$$\mathcal{L}[\{\psi_i\}, \{\epsilon_{ij}\}] = \sum_{i=1}^n \langle \psi_i | \hat{h} | \psi_i \rangle + \frac{1}{2} \sum_{i,j=1}^n \langle \psi_i \psi_j | | \psi_i \psi_j \rangle - \sum_{i,j=1}^n \epsilon_{ij} (\langle \psi_i | \psi_j \rangle - \delta_{ij}) \quad (1.9)$$

where $\{\epsilon_{ij}\}$ are our Lagrangian multipliers for the orthonormality constraint (see appendix ?? for a brief explanation the Lagrangian approach to constrained optimization).

⁴It is typical to use dummy indices i, j, k, l to count over the orbitals in the reference determinant Φ – the “occupied orbitals” – and to use a, b, c, d , to count over the orbitals not contained in Φ – the “unoccupied” or “virtual orbitals.” Dummy indices p, q, r, s are generally used to count over the full set of spin-orbitals, whether occupied or not.

A Constrained Optimization

The standard method of optimizing a function or functional subject to a constraint is called Lagrangian optimization. Taking a function of two variables $f(x, y)$ as an example, suppose we want to optimize it subject to a constraint of the form $g(x, y) = c$. In this approach, we define the “Lagrangian function” \mathcal{L} as

$$\mathcal{L}(x, y, \lambda) \equiv f(x, y) - \lambda(g(x, y) - c)$$

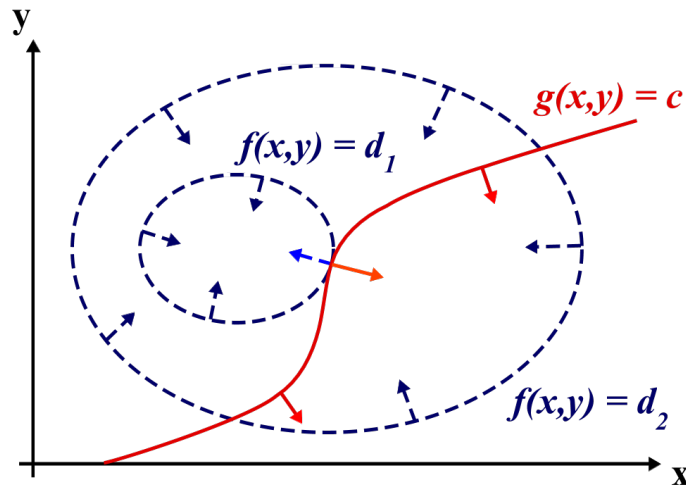
where the parameter λ is called the Lagrange multiplier. The constrained optimization of $f(x, y)$ can be achieved by determining a stationary point of \mathcal{L} in all of its arguments.⁵

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial x} &= \frac{\partial f}{\partial x} - \lambda \frac{\partial g}{\partial x} \stackrel{!}{=} 0 \\ \frac{\partial \mathcal{L}}{\partial y} &= \frac{\partial f}{\partial y} - \lambda \frac{\partial g}{\partial y} \stackrel{!}{=} 0 \\ \frac{\partial \mathcal{L}}{\partial \lambda} &= c - g(x, y) \stackrel{!}{=} 0\end{aligned}$$

The last equation is simply the requirement that the constraint $g(x, y) = c$ be satisfied – i.e. that the point (x, y) lies along the contour of $g(x, y)$ specified by $g(x, y) = c$. The first two equations correspond to the requirement that the gradients of the surface $f(x, y)$ and the constraint surface $g(x, y)$ be parallel

$$\nabla f = \lambda \nabla g \tag{A.1}$$

which is always true at the point (x, y) of closest approach along the line $g(x, y) = c$ to a minimum or maximum of the surface $f(x, y)$. This is best understood visually.



If the gradients were not parallel, we could move along $g(x, y) = c$ to a higher contour of $f(x, y)$ by following the component of ∇f parallel to $g(x, y) = c$.

⁵The $\stackrel{!}{=}$ sign means “must equal” – these are the conditions to be satisfied.

B Functional Derivatives

A functional is just a function of a function – i.e. some rule F that maps a function f to a number $F[f]$. Definite integrals are probably the most familiar example. In order to optimize a functional F with respect to its argument f , one needs to take a *functional derivative*.⁶ To motivate the definition of a functional derivative, first consider the definition of an ordinary derivative

$$\frac{df(x)}{dx} \equiv \lim_{\varepsilon \rightarrow 0} \frac{f(x + \varepsilon) - f(x)}{\varepsilon} \quad (\text{B.1})$$

and note the following identity, which you can verify using $f(x + \varepsilon) = f(x) + \frac{df(x)}{dx}\varepsilon + \mathcal{O}(\varepsilon^2)$.

$$\frac{df(x)}{dx} = \left. \frac{df(x + \varepsilon)}{d\varepsilon} \right|_{\varepsilon=0} \quad (\text{B.2})$$

For multivariate functions, we have the concept of a *directional derivative*

$$\mathbf{y} \cdot \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \lim_{\varepsilon \rightarrow 0} \frac{f(\mathbf{x} + \varepsilon \mathbf{y}) - f(\mathbf{x})}{\varepsilon} \quad (\text{B.3})$$

which measures the change in $f(\mathbf{x})$ in the direction \mathbf{y} . By analogy to equation B.2, the directional derivative can be evaluated as an ordinary scalar derivative with respect to ε .

$$\mathbf{y} \cdot \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \left. \frac{df(\mathbf{x} + \varepsilon \mathbf{y})}{d\varepsilon} \right|_{\varepsilon=0} \quad (\text{B.4})$$

The functional derivative $\frac{\delta F}{\delta f}$ is defined by requiring that it satisfy an equation analogous to B.3

$$\int_{-\infty}^{\infty} dx' g(x') \frac{\delta F[f]}{\delta f(x')} \equiv \lim_{\varepsilon \rightarrow 0} \frac{F[f + \varepsilon g] - F[f]}{\varepsilon} \quad (\text{B.5})$$

which we might refer to as a *functional directional derivative*, giving the change in F upon displacing its argument along g . Here, the integral takes the role of the dot product in B.3. Using the same trick as in equations B.2 and B.4, the functional derivative can be expressed as an ordinary scalar derivative.

$$\int_{-\infty}^{\infty} dx' g(x') \frac{\delta F[f]}{\delta f(x')} = \left. \frac{dF[f + \varepsilon g]}{d\varepsilon} \right|_{\varepsilon=0} \quad (\text{B.6})$$

The standard procedure for evaluating the functional derivative is to first evaluate the right-hand side of equation B.6 for an arbitrary g and then infer what $\frac{\delta F[f]}{\delta f(x)}$ must be by comparing to the left-hand side. Equivalently, $g(x')$ can be replaced with a Dirac delta $\delta(x - x')$ in order arrive at $\frac{\delta F[f]}{\delta f(x)}$ directly.

Using equation B.6 and the Fundamental Lemma of Calculus of Variations (Appendix C) the stationarity condition for a functional

$$\frac{\delta F[f]}{\delta f} \stackrel{!}{=} 0 \quad (\text{B.7})$$

is equivalent to the following condition.

$$\left. \frac{dF[f + \varepsilon g]}{d\varepsilon} \right|_{\varepsilon=0} \stackrel{!}{=} 0 \quad \text{for all } g(x) \quad (\text{B.8})$$

⁶http://en.wikipedia.org/wiki/Functional_derivative

C Fundamental Lemma of Calculus of Variations