Towards Hartree-Fock: Lagrange multipliers

This is a mathematical intermezzo on the "undetermined multiplier method of Lagrange".

The aim: minimizing a function (e.g. the expectation value of H) while applying a constraint (e.g. the orbitals in the wave function should be orthogonal).

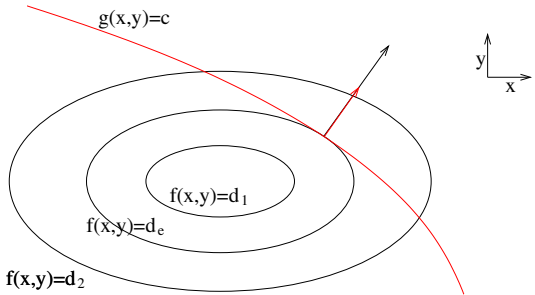
First a graphical explanation for a simple case.

Wanted: extremum of $f(x, y)$ subject to the constraint: $g(x, y) = c$

$g(x, y) = c$ defines a curve, i.e. a single contour of $g(x, y)$ with $g(x, y) = c$.

We look at contours of $f(x, y)$, i.e. curves for which $f(x, y) = d$, with d a constant. Near the extremum there are 2 or no intersections of $g(x, y) = c$ and $f(x, y) = d$, depending on the value of d.

Towards Hartree-Fock: Lagrange multipliers



We change d, from d_1 to d_2 (see plot), and, when we hit the extremal point (x_e, y_e), the two intersections merge into a single point. In this point, the "surfaces" defined by $g(x, y) = c$ and $f(x, y) = d_e$ are parallel, hence their normal vectors, i.e. the gradients, are parallel:

$$\boldsymbol{\nabla} f = \lambda \boldsymbol{\nabla} g \quad \text{i.e.} \quad \begin{pmatrix} \partial f/\partial x \\ \partial f/\partial y \end{pmatrix} = \lambda \begin{pmatrix} \partial g/\partial x \\ \partial g/\partial y \end{pmatrix}$$

Towards Hartree-Fock: Lagrange multipliers

Intermezzo in Intermezzo: The Gradient.

Let's choose a parametrized trajectory $\mathbf{s}(t)$ on the constraint:

$$g(s_x(t), s_y(t)) = c$$

Now differentiate:

$$\frac{d}{dt} g(s_x(t), s_y(t)) = \frac{\partial g}{\partial x}\frac{ds_x}{dt} + \frac{\partial g}{\partial y}\frac{ds_y}{dt} = \begin{pmatrix} \frac{\partial g}{\partial x} \\ \frac{\partial g}{\partial y} \end{pmatrix} \cdot \begin{pmatrix} \frac{ds_x}{dt} \\ \frac{ds_y}{dt} \end{pmatrix} = 0$$

The first vector is the gradient.
The second vector is tangent to the surface $g(x, y) = c$ (Dutch: de vector raakt aan het oppervlak).

Hence the gradient of g is perpendicular to the isosurface of g. This is a general statement.

End of Intermezzo in Intermezzo.

Towards Hartree-Fock: Lagrange multipliers

Solving the system:

$$\begin{cases} \boldsymbol{\nabla} f = \lambda \boldsymbol{\nabla} g \\ g(x, y) = c \end{cases}$$

yields x_e, y_e and λ. These equations can be obtained by demanding that the gradient of the function $\mathcal{L}(x, y, \lambda)$,

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

vanishes in the three dimensional x, y, λ space:

$$\boldsymbol{\nabla} \mathcal{L} = \boldsymbol{\nabla} f - \lambda \boldsymbol{\nabla} g = 0$$
$$\partial \mathcal{L} / \partial \lambda = g(x, y) - c = 0$$

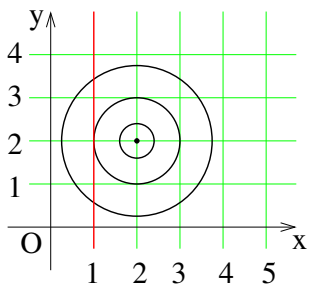
Note that we are only guaranteed to get a stationary/critical point, but not necessarily a (local) minimum or maximum.

Let's look at an example ...

Towards Hartree-Fock: Lagrange multipliers

$$\left.\begin{array}{l} f = (x-2)^2 + (y-2)^2 \\ g = x = 1 \end{array}\right\} \text{ so } \mathcal{L} = (x-2)^2 + (y-2)^2 - \lambda(x-1)$$

We immediately see that $(x, y) = (1, 2)$ yields the minimum $f = 1$ on g.



Lagrange:

$$\frac{\partial \mathcal{L}}{\partial x} = 2(x-2) - \lambda = 0$$
$$\frac{\partial \mathcal{L}}{\partial y} = 2(y-2) = 0 \Rightarrow y = 2$$
$$\frac{\partial \mathcal{L}}{\partial \lambda} = -(x-1) = 0 \Rightarrow x = 1$$

$(x, y, \lambda) = (1, 2, -2)$ is a stationary point in 3D space. Let's look at the "magic" in more detail ...

Towards Hartree-Fock: Lagrange multipliers

Let's try to systematically "complete the square":

$$\mathcal{L} = (x-2)^2 + (y-2)^2 - \lambda(x-1) = x^2 - 4x + 4 - \lambda x + \lambda + (y-2)^2 =$$
$$\left(x - 2 - \frac{\lambda}{2}\right)^2 - \frac{\lambda^2}{4} - 2\lambda + \lambda + (y-2)^2 =$$
$$\left(x - 1 - \frac{1}{2}(\lambda+2)\right)^2 - \frac{1}{4}(\lambda+2)^2 + 1 + (y-2)^2$$

For $(x, y, \lambda) = (1, 2, -2)$ all terms in brackets vanish and this is 1, as it should be. All small excursions of x (or of y, or of λ) away from $(1, 2, -2)$ yield quadratic changes: we have an stationary point. As we know, because the 3D gradient vanishes.

This stationary point is a global minimum on the xy-plane (with $\lambda = -2$ fixed). It is a saddle point in the 3D $xy\lambda$-space. Note: the method provides no guarantees about the nature of the stationary points.

Towards Hartree-Fock: Lagrange multipliers

Now in more, i.e., n dimensions (see Groenenboom notes). We have a function

$$f(\mathbf{x}) = f(x_1, x_2, \dots x_n)$$

We have m constraints:

$$g_i(\mathbf{x}) = g_i(x_1, x_2, \dots x_n) = 0, \quad i = 1, 2, \dots, m$$

This is a $(n - m)$-dimensional surface M.

We are looking for $\mathbf{x}_e$ such that f is stationary in $\mathbf{x}_e$ and on all $g_i(\mathbf{x}_e) = 0$.

Any allowed (by the constraints) excursion $\mathbf{s}$ away from $\mathbf{x}_e$ is perpendicular to all gradients $\boldsymbol{\nabla} g_i$ (calculated in $\mathbf{x}_e$):

$$\mathbf{s} \perp \boldsymbol{\nabla} g_1 \text{ and } \mathbf{s} \perp \boldsymbol{\nabla} g_2 \text{ and } \mathbf{s} \perp \boldsymbol{\nabla} g_3, \dots, \mathbf{s} \perp \boldsymbol{\nabla} g_m. \tag{1}$$

If excursions $\mathbf{s}$ are perpendicular to $\boldsymbol{\nabla} f$ then there is *no* first-order change of f.

We can achieve this by demanding that the gradient $\boldsymbol{\nabla} f$ be a linear combination of the gradients $\boldsymbol{\nabla} g_i$, i.e.

$$\boldsymbol{\nabla} f = \sum_i \lambda_i \boldsymbol{\nabla} g_i.$$

Then, by construction, any vector $\mathbf{s}$ (simultaneously on all constraints) is perpendicular to $\boldsymbol{\nabla} f$ and gives a vanishing first-order change in f: $\mathbf{x}_e$ is a stationary point.

Towards Hartree-Fock: Lagrange multipliers

We define (called $\tilde{f}(\vec{r}, \vec{\lambda})$ in Groenenboom):

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) \equiv f(\mathbf{x}) - \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x})$$

and we demand that the gradient in $(n+m)$-dimensional space vanishes:

$$\frac{\partial \mathcal{L}}{\partial x_j} = \frac{\partial f}{\partial x_j} - \sum_{i=1}^{m} \lambda_i \frac{\partial g_i}{\partial x_j} = 0 \ , \quad j = 1, 2, \ldots, n$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_i} = g_i = 0 \ , \quad i = 1, 2, \ldots, m$$

Note:

- We are allowed to replace λ_i with $-\lambda_i$.
- We did not worry about complex $\mathbf{x}$, ...
- Demanding $g_i(\mathbf{x}) = 0$ is general. Suppose $g_i(\mathbf{x}) = c \neq 0$, then $\tilde{g}_i(\mathbf{x}) = g_i(\mathbf{x}) - c = 0$.

Towards Hartree-Fock: minimization I

Why do we need Lagrange multipliers?

Because we have to minimize our energy expression:

$$\epsilon = 2\sum_{k}^{N/2}\langle\phi_k|\hat{h}|\phi_k\rangle + \sum_{k}^{N/2}\sum_{l}^{N/2} 2J_{kl} - K_{kl} = 2\sum_{k}^{N/2} h_{kk} + \sum_{k}^{N/2}\sum_{l}^{N/2} 2J_{kl} - K_{kl}$$

with the constraints that the orbitals are orthonormal, i.e.

$$\langle\phi_k|\phi_l\rangle = \delta_{lk}$$

These conditions make that the wave function, i.e., the Slater determinant, is normalized and satisfies Pauli.

In order to fulfill our constraints, we introduce Lagrange multipliers Λ_{ij}, one for each constraint, and add terms for our constraints to ϵ to make a new function ϵ' that we will minimize:

$$\epsilon' = 2\sum_{k}^{N/2} h_{kk} + \sum_{k}^{N/2}\sum_{l}^{N/2}(2J_{kl} - K_{kl}) - \sum_{k}^{N/2}\sum_{l}^{N/2}\Lambda_{lk}\left(\langle\phi_k|\phi_l\rangle - \delta_{lk}\right)$$

Towards Hartree-Fock: the basis

Help slide:

So we have to minimize ϵ' by changing the orbitals $|\phi_k\rangle$ while keeping the constraint $\langle\phi_k|\phi_l\rangle = \delta_{lk}$.

This is quite abstract, and not suitable for a computer.

In the next slide we will expand the obitals in basis functions. The basis functions are **fixed**. The expansion coefficients $C_{\lambda k}$ are the **variables**. We need to find those values of $C_{\lambda k}$ that give the minimum of ϵ'.

In principle the number of basis functions M can be infinite. The number of orbitals is uniquely defined. In our closed shell system it is $N/2$, i.e. the number of electrons/2.

Towards Hartree-Fock: the basis

Expand the spatial orbitals in a basis:

$$\phi_k(\mathbf{r}) = \sum_{\lambda}^{M} C_{\lambda k}\, \chi_\lambda(\mathbf{r}) = \sum_{\lambda}^{M} \chi_\lambda(\mathbf{r}) C_{\lambda k} \quad \text{i.e.} \quad |\phi_k\rangle = \sum_{\lambda}^{M} |\chi_\lambda\rangle C_{\lambda k}$$

Typically the ϕ_k are the orbitals of the whole system, i.e. the molecular orbitals (MOs). In chemistry the χ_λ are often chosen to be centered on the individual atoms: atomic orbitals (AOs). The acronym for this approximation: MO-LCAO (MO-Linear Combination of AOs).

In practice the basis is never complete, but should "approximate completeness sufficiently".

The atomic orbitals are, in general, not orthonormal, because

$$S_{\lambda\mu} = \langle \chi_\lambda | \chi_\mu \rangle \neq 0$$

if χ_λ and χ_μ are on different atoms. $\mathbf{S}$ is the overlap matrix.

In rare instances linear dependencies arise (a basis function $|\chi_\lambda\rangle$ is a linear combination of several others). The program will crash. That should be avoided.

Towards Hartree-Fock: the basis

We should minimize:

$$\epsilon' = 2\sum_{k}^{N/2} h_{kk} + \sum_{k}^{N/2}\sum_{l}^{N/2}(2J_{kl} - K_{kl}) - \sum_{k}^{N/2}\sum_{l}^{N/2}\Lambda_{lk}\left(\langle\phi_k|\phi_l\rangle - \delta_{lk}\right)$$

We insert the expansion in the basis (using $|\phi_k\rangle = \sum_{\lambda}^{M}|\chi_\lambda\rangle C_{\lambda k}$):

$$\epsilon' = 2\sum_{k}^{N/2}\sum_{\lambda}^{M}\sum_{\mu}^{M} C^*_{\lambda k} C_{\mu k}\langle\chi_\lambda|\hat{h}|\chi_\mu\rangle$$

$$+\sum_{k}^{N/2}\sum_{l}^{N/2}\sum_{\lambda}^{M}\sum_{\mu}^{M}\sum_{\nu}^{M}\sum_{\tau}^{M} C^*_{\lambda k} C^*_{\mu l} C_{\nu k} C_{\tau l}\langle\chi_\lambda\chi_\mu|(2 - \hat{P}_{12})\hat{g}(1,2)|\chi_\nu\chi_\tau\rangle$$

$$-\sum_{k}^{N/2}\sum_{l}^{N/2}\Lambda_{lk}\left(\sum_{\lambda}^{M}\sum_{\mu}^{M} C^*_{\lambda k} C_{\mu l}\langle\chi_\lambda|\chi_\mu\rangle - \delta_{lk}\right)$$

Help: $\sum_{k}^{N/2}\sum_{l}^{N/2} 2\langle\phi_k(1)\phi_l(2)|\hat{g}(1,2)|\phi_k(1)\phi_l(2)\rangle - \langle\phi_k(1)\phi_l(2)|\hat{g}(1,2)|\phi_l(1)\phi_k(2)\rangle$

$= \sum_{k}^{N/2}\sum_{l}^{N/2} 2\langle\phi_k(1)\phi_l(2)|\hat{g}(1,2)|\phi_k(1)\phi_l(2)\rangle - \langle\phi_k(1)\phi_l(2)|\hat{P}_{12}\hat{g}(1,2)|\phi_k(1)\phi_l(2)\rangle$

Towards Hartree-Fock: "simplification"

Let's try to safe ink:

$$\epsilon' = 2\sum_{k}^{N/2}\sum_{\lambda}^{M}\sum_{\mu}^{M} C^*_{\lambda k} C_{\mu k} \langle \chi_\lambda | \hat{h} | \chi_\mu \rangle$$

$$+\sum_{k}^{N/2}\sum_{l}^{N/2}\sum_{\lambda}^{M}\sum_{\mu}^{M}\sum_{\nu}^{M}\sum_{\tau}^{M} C^*_{\lambda k} C^*_{\mu l} C_{\nu k} C_{\tau l} \langle \chi_\lambda \chi_\mu | (2 - \hat{P}_{12}) \hat{g}(1,2) | \chi_\nu \chi_\tau \rangle$$

$$-\sum_{k}^{N/2}\sum_{l}^{N/2} \Lambda_{lk} \left(\sum_{\lambda}^{M}\sum_{\mu}^{M} C^*_{\lambda k} C_{\mu l} \langle \chi_\lambda | \chi_\mu \rangle - \delta_{ij} \right)$$

$$= \sum_{\lambda\mu} 2 P_{\mu\lambda} h_{\lambda\mu} + \sum_{\lambda\mu\nu\tau} P_{\nu\lambda} P_{\tau\mu} \langle \lambda\mu | (2 - \hat{P}_{12}) \hat{g} | \nu\tau \rangle - \sum_{kl} \Lambda_{lk} \Big(\sum_{\lambda\mu} C^*_{\lambda k} C_{\mu l} S_{\lambda\mu} - \delta_{lk} \Big)$$

The density matrix **P** and (remember) overlap matrix **S** are:

$$P_{\mu\lambda} \equiv \sum_{k}^{N/2} C_{\mu k} C^*_{\lambda k} \quad , \quad S_{\lambda\mu} \equiv \langle \chi_\lambda | \chi_\mu \rangle$$

Note the indices for **P**, the second has the c.c. These are not operators!

Towards Hartree-Fock: minimization II

At the minimum all variations of ϵ' with respect to the real and complex parts of $C_{\rho m}$ must vanish:

$$\frac{\partial \epsilon'}{\partial \mathrm{Re}[C_{\rho m}]} = 0 \quad \text{and} \quad \frac{\partial \epsilon'}{\partial \mathrm{Im}[C_{\rho m}]} = 0 \quad \text{where} \quad C_{\rho m} = \mathrm{Re}[C_{\rho m}] + i\mathrm{Im}[C_{\rho m}]$$

This yields the same as requireing:

$$\frac{\partial \epsilon'}{\partial C_{\rho m}} = 0 \quad \text{and} \quad \frac{\partial \epsilon'}{\partial C^*_{\rho m}} = 0 \quad \text{where } C_{\rho m} \text{ and } C^*_{\rho m} \text{ are considered } \textit{independent}$$

Note: we single out a specific element $C_{\rho m}$ or $C^*_{\rho m}$ in the sum:

$$\frac{\partial P_{\mu\lambda}}{\partial C_{\rho m}} = \frac{\partial}{\partial C_{\rho m}} \left(\sum_k^{N/2} C_{\mu k} C^*_{\lambda k} \right) = \sum_k^{N/2} \delta_{\mu\rho}\delta_{mk} C^*_{\lambda k} = \delta_{\mu\rho} C^*_{\lambda m}$$

$$\frac{\partial P_{\mu\lambda}}{\partial C^*_{\rho m}} = \frac{\partial}{\partial C^*_{\rho m}} \left(\sum_k^{N/2} C_{\mu k} C^*_{\lambda k} \right) = \sum_k^{N/2} \delta_{\lambda\rho}\delta_{mk} C_{\mu k} = \delta_{\lambda\rho} C_{\mu m}$$

Towards Hartree-Fock: minimization II

$$\frac{\partial P_{\mu\lambda}}{\partial C_{\rho m}} = \delta_{\mu\rho} C^*_{\lambda m} \quad \text{and} \quad \frac{\partial P_{\mu\lambda}}{\partial C^*_{\rho m}} = \delta_{\lambda\rho} C_{\mu m}$$

$$\frac{\partial}{\partial C^*_{\rho m}} \left(\sum_{\lambda\mu} 2 P_{\mu\lambda} h_{\lambda\mu} \right) = \sum_{\lambda\mu} 2 \delta_{\lambda\rho} C_{\mu m} h_{\lambda\mu} = \sum_{\mu} 2 C_{\mu m} h_{\rho\mu} = \sum_{\nu} 2 h_{\rho\nu} C_{\nu m}$$

$$\frac{\partial}{\partial C^*_{\rho m}} \left(\sum_{\lambda\mu\nu\tau} P_{\nu\lambda} P_{\tau\mu} \langle \lambda\mu | (2 - \hat{P}_{12}) \hat{g} | \nu\tau \rangle \right)$$

$$= \sum_{\lambda\mu\nu\tau} \delta_{\lambda\rho} C_{\nu m} P_{\tau\mu} \langle \lambda\mu | (2 - \hat{P}_{12}) \hat{g} | \nu\tau \rangle + \sum_{\lambda\mu\nu\tau} P_{\nu\lambda} \delta_{\mu\rho} C_{\tau m} \langle \lambda\mu | (2 - \hat{P}_{12}) \hat{g} | \nu\tau \rangle$$

$$= \sum_{\mu\nu\tau} P_{\tau\mu} \langle \rho\mu | (2 - \hat{P}_{12}) \hat{g} | \nu\tau \rangle C_{\nu m} + \sum_{\lambda\nu\tau} P_{\nu\lambda} \langle \lambda\rho | (2 - \hat{P}_{12}) \hat{g} | \nu\tau \rangle C_{\tau m}$$

$$= \sum_{\mu\nu\tau} P_{\tau\mu} \langle \rho\mu | (2 - \hat{P}_{12}) \hat{g} | \nu\tau \rangle C_{\nu m} + \sum_{\mu\tau\nu} P_{\tau\mu} \langle \mu\rho | (2 - \hat{P}_{12}) \hat{g} | \tau\nu \rangle C_{\nu m}$$

$$= 2 \sum_{\nu} \sum_{\mu\tau} P_{\tau\mu} \langle \rho\mu | (2 - \hat{P}_{12}) \hat{g} | \nu\tau \rangle C_{\nu m}$$

Towards Hartree-Fock: minimization II

$$\frac{\partial}{\partial C^*_{\rho m}}\left(-\sum_{kl}\Lambda_{lk}(\sum_{\lambda\mu}C^*_{\lambda k}C_{\mu l}S_{\lambda\mu}-\delta_{lk})\right)$$
$$=-\sum_{l}\sum_{\mu}\Lambda_{lm}C_{\mu l}S_{\rho\mu}=-\sum_{l}\sum_{\mu}S_{\rho\mu}C_{\mu l}\Lambda_{lm}$$

$$\frac{\partial\epsilon'}{\partial C^*_{\rho m}}=\sum_{\nu}2h_{\rho\nu}C_{\nu m}+2\sum_{\nu}\sum_{\mu\tau}P_{\tau\mu}\langle\rho\mu|(2-\hat{P}_{12})\hat{g}|\nu\tau\rangle C_{\nu m}-\sum_{l}\sum_{\mu}S_{\rho\mu}C_{\mu l}\Lambda_{lm}$$
$$\equiv 2\sum_{\nu}^{M}F_{\rho\nu}C_{\nu m}-\sum_{\mu}^{M}S_{\rho\mu}\sum_{l}^{N/2}C_{\mu l}\Lambda_{lm}=2[\mathbf{FC}]_{\rho m}-[\mathbf{SC\Lambda}]_{\rho m}=0\Rightarrow 2\mathbf{FC}=\mathbf{SC\Lambda}$$

$$\frac{\partial\epsilon'}{\partial C_{\rho m}}=0 \quad \text{yields:} \quad 2\mathbf{C}^\dagger\mathbf{F}=\mathbf{\Lambda}\mathbf{C}^\dagger\mathbf{S}$$

Here we have defined the Fock matrix $\mathbf{F}$. Further we have the overlap matrix $\mathbf{S}$, the coefficient matrix $\mathbf{C}$ and the Lagrange multiplier matrix $\mathbf{\Lambda}$. $\mathbf{F}$, $\mathbf{S}$ and $\mathbf{\Lambda}$ are Hermitian:

$$\mathbf{F}^\dagger=\mathbf{F} \quad \text{i.e.} \quad F^*_{\mu\rho}=F_{\rho\mu} \quad , \quad \mathbf{S}^\dagger=\mathbf{S} \quad \text{and} \quad \mathbf{\Lambda}^\dagger=\mathbf{\Lambda}$$

Towards Hartree-Fock: minimization II

Help slide:

The elements of the Fock matrix:

$$F_{\rho\nu} = h_{\rho\nu} + \sum_{\mu\tau} P_{\tau\mu} \langle \rho\mu | (2 - \hat{P}_{12}) \hat{g} | \nu\tau \rangle$$

So its size is $M \times M$.

Note it depends on the coefficient matrix $\mathbf{C}$ *via* the elements of the density matrix $P_{\tau\mu}$.