Optimization of molecular orbitals

Yann Damour,
Supervisor : Scemama Anthony

21 January 2021

Configuration interaction

The full configuration interaction (CI) gives, for a given atomic basis set, the exact solution of the Schrödinger equation. But for larger molecules, the number of configurations tends toward infinity as the computational cost. The goal is to use selected CI and speed up the convergence by the optimization of the molecular orbitals.

- Natural orbitals, 1955, Löwdin
- Localization, 1963, Edminston and Ruedenberg
- Foster-Boys, Pipek-Mezey

Orbitals optimization

The localization of the orbitals is obtained by minimizing a criterion. This criterion is used to create a rotation matrix in order to apply a linear combination of the orbitals and minimize the criterion.

Umrigar & co (2020), use the idea of minimizing the CI energy as a criterium :

$$E(\mathbf{X}) = \langle \Psi | e^{\hat{\mathbf{X}}} \hat{H} e^{-\hat{\mathbf{X}}} | \Psi \rangle \tag{1}$$

$$min_{(\mathbf{X} \in \mathbb{R}^n)} E(\mathbf{X}) = \langle \Psi | e^{\hat{\mathbf{X}}} \hat{H} e^{-\hat{\mathbf{X}}} | \Psi \rangle \tag{2}$$

With $\hat{X}$ a real antihermitian operator such as :

$$\mathbf{R} = e^{-\hat{\mathbf{X}}} \tag{3}$$

is a rotation matrix

Rotation matrix

We can compute $\mathbf{R} = e^{-\hat{\mathbf{x}}}$ such as :

$X = WxW^{\dagger}$

Let us note $\tau = \sqrt{-x}$

Then we have :
$\mathbf{R} = W\cos(\tau)W^{\dagger} + W\tau^{-1}\sin(\tau)W^{\dagger}X$

Addtionally, $\mathbf{R}$ has the following properties :

$$\mathbf{R}\mathbf{R}^{t} = \mathbf{R}^{t}\mathbf{R} = \mathbf{1} \tag{4}$$

$\rightarrow$ $\mathbf{R}$ produces orthogonal tranformations

Problem

The problem is the following :
We have to optimise the CI and the orbital parameters, with
$n^p_{CI} >>> n^p_{OM} = n_{orb}(n_{orb} - 1)/2$

To solve this problem Umrigar & co, suggest doing the opimization of these parameters alternatly.

Since the cost of the CI is much bigger than the optimization of the orbitals, we need a robust method. This method must ensure the orbitals opimization steps will lower the energy in the right way.

Newton's method

We have to minimize a function, the energy (CI). Taylor development of f around a dummy variable x in one dimension :

$$f(x+dx) = f(x) + \nabla f(x)dx + \frac{1}{2}\nabla^2 f(x)dx^2 + ... \tag{5}$$

Noting the x as the step x_k and $x+dx$ as x_{k+1}, we obtain the Newton method :

$$f(x_{k+1}) = f(x_k) + \nabla f(x_k)(x_{k+1} - x_k) + \frac{1}{2}\nabla^2 f(x_k)(x_{k+1} - x_k)^2 \tag{6}$$

Which leads to :

$$x_{k+1} = x_k - [\nabla^2 f(x_k)]^{-1}\nabla f(x_k) \tag{7}$$

$$x_{k+1} = x_k - H_k^{-1} g_k \tag{8}$$

With $\mathbf{H}_k$ the Hessian and $\mathbf{g}_k$ the gradient

Newton's method

In n-dimensions, the Taylor expansion of f around a dummy position $\vec{x_k}$ is :

$$f(x_k + p) = f(x_k) + \nabla f(x_k)^T p + \frac{1}{2} p^T \nabla^2 f(x_k) p + ... \tag{9}$$

Let us note :
$m_k = f(x_k + p)$, $f(x_k) = f_k$, the gradient $g_k = \nabla f(x_k)$ and the hessian $H_k = \nabla^2 f(x_k)$

$$m_k(p) = f_k + g_k^T p + \frac{1}{2} p^T H_k p \tag{10}$$

$m(p)$ represents the quadratic approximation model of the function around x_k, so the next positions will be obtained using the solution :

$$p = -H_k^{-1} g_k$$

Then

$$x_{k+1} = x_k + p$$

Newton's method

Problem :
$\rightarrow$ the Newton's method using the hessian gives a bad approximation at long range but good approximation at short range.
$\rightarrow$ the gradient method gives a bad approximation at short range but good approximation at long range.

Umrigar & co

Umrigar & co start from natural orbitals and use several algorithms :

- Newton method
- Diagonal Newton method (using only the diagonal terms of the hessian $\Rightarrow$ easy to compute and invert)
- AMSGrad (use only the gradient)
- Accelerated Newton's method

Umrigar & co

Umrigar & co introduce the accerelated Newton's method, which consists to build the diagonal terms of the hessian and apply a "factor f_t determined by the cosine of the angle between the previous step $x_k - x_{k-1}$ and the current step $x_{k+1} - x_k$", such as :

$$f_t = min\left(\frac{1}{2 - \cos(x_k - x_{k-1}, x_{k+1} - x_k)}, \frac{1}{\epsilon}\right) \tag{11}$$

where ϵ is initialized to 0.01 and $\epsilon \leftarrow \epsilon^{0.8}$ each time $\cos(x_k - x_{k-1}, x_{k+1} - x_k) < 0$
And the cosine :

$$cos(v, w) = \frac{\langle v, w \rangle}{\sqrt{\langle v, v \rangle \langle w, w \rangle}} \tag{12}$$

With :

$$\langle v, w \rangle = v^T h w \tag{13}$$

With h, the hessian approximate by his diagonal.

Umrigar & co

They start from HF orbitals at step 0 and build natural orbitals at step 1.

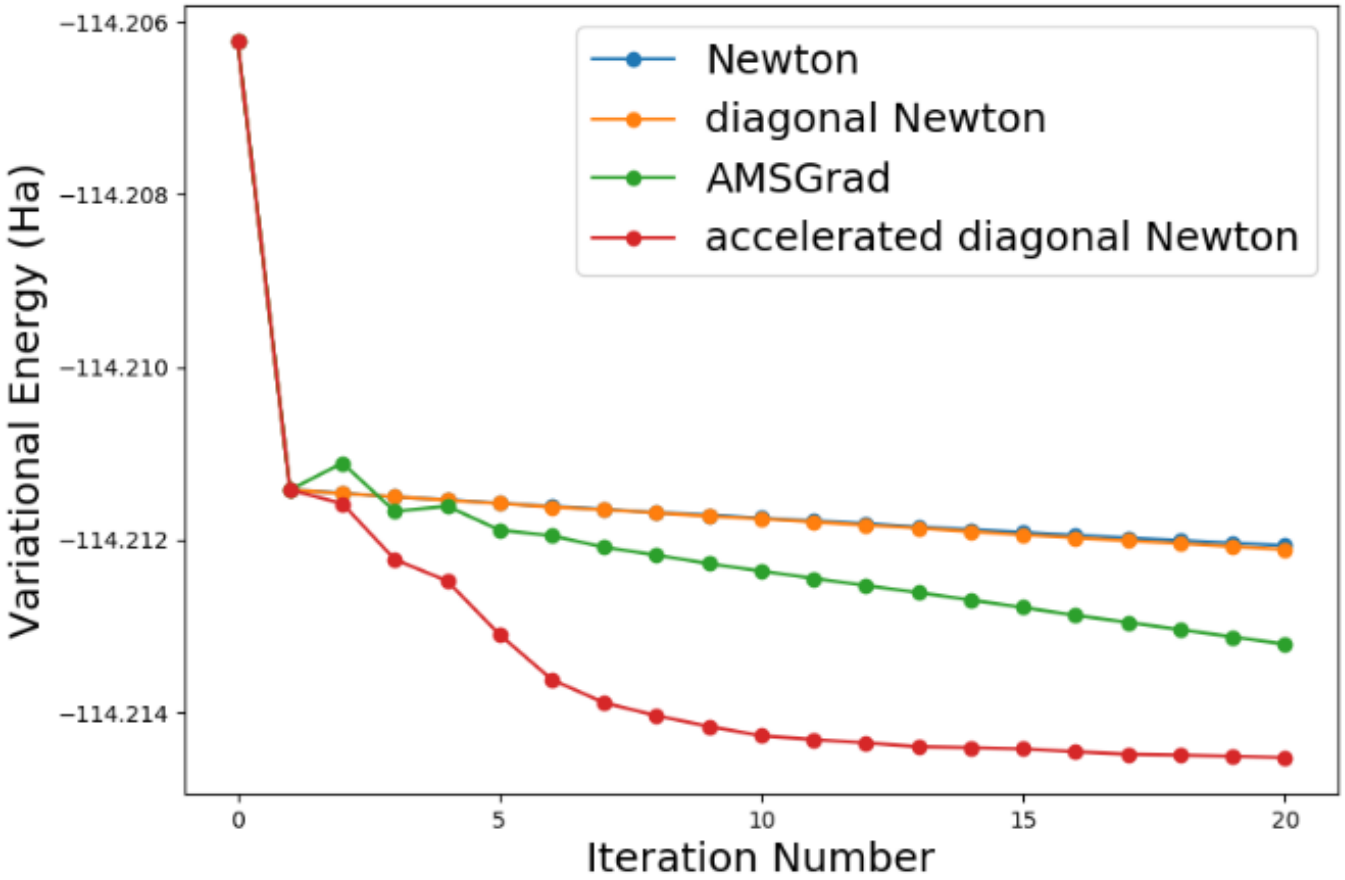

Figure: Convergence of the four orbital optimization schemes, Umrigar & co

Trust region

The idea of the trust region to optimize orbitals was used by Ida-Marie Høyvik (2013) in her thesis.

Basically it is a Newton's method with a constraint on the step range, set by a Lagragian multiplier.

This method is explained in "Trust-Region Methods", Nocedal et Wright (1999). The following slides were made using this book (Chap 4).
DOI : 10.1007/0-387-22742-3_4

We came back at the equation :

$$m_k(p) = f_k + g_k^T p + \frac{1}{2} p^T H_k p$$

We want the solution of the following problem :

$$min_{(p \in \mathbb{R}^n)} m_k(p) = f_k + g_k^T p + \frac{1}{2} p^T H_k p, \quad ||p|| \leq \Delta_k \tag{14}$$

With Δ_k the trust region radius $\Delta_k > 0$

We define the agreement between the model and the real function :

$$\rho_k = \frac{f(x_k) - f(x_k + p_k)}{m_k(0) - m_k(p_k)} \tag{15}$$

Let us note [1]:
$0 < \eta_1 < \eta_2 < 1$ and $0 < \gamma_1 < \gamma_2 < 1 < \gamma_3$

- $\rho_k < \eta_1$, step rejected, $\Delta_k \in [\gamma_1 \Delta_k, \gamma_2 \Delta_k]$
- $\eta_1 \leq \rho_k \leq \eta_2$, step accepted, $\Delta_k \in [\gamma_2 \Delta_k, \Delta_k]$
- $\rho_k \geq \eta_2$, step accepted, $\Delta_k \in [\Delta_k, \gamma_3 \Delta_k]$

The size of the step evolve in function of the agreement between the model m_k and the real function f_k.

[1] Michel BERGMANN, Thesis

Solution of $min_{(p \in \mathbb{R}^n)} m_k(p)$:
if $||p|| = ||H_k^{-1} g_k|| \leq \Delta_k \Rightarrow$ the unconstraint solution p.

else we build the Lagragian :

$$\mathcal{L}(p, \lambda) = f_k + g_k^T p + \frac{1}{2} p^T H_k p + \frac{1}{2} \lambda (p^T p - \Delta^2) \tag{16}$$

With λ the Lagrange multiplier.
By differentiating the Lagragian and set $\nabla_p \mathcal{L}(p, \lambda) = 0$:

$$(H_k + \lambda \mathbf{1}) p = -g_k \tag{17}$$

We use the Lagragian in order to search the minimum value of the model on the boundary of the trust region.

Copy from Nocedal et Wright (1999)
Theroem :
The vector p^* is a global solution of the trust region problem

$$min_{(p\in\mathbb{R}^n)} m_k(p) = f_k + g^T p + \frac{1}{2} p^T H p, \quad ||p|| \leq \Delta \tag{18}$$

if and only if p^* is feasible and there is a scalar λ such that the following conditions are satisfied :

$$(H + \lambda \mathbf{I}) p^* = -g$$
$$\lambda(\Delta - ||p^*||) = 0$$

$(B + \lambda I)$ is positive semidefinite

The previous theorem suggest an algorithm for finding the solution p of (18). Either $\lambda = 0$ satifies the two first conditions of the theorem with $||p|| \leq \Delta$, or else we can define :

$$p(\lambda) = -(H + \lambda I)^{-1} g \tag{19}$$

For λ sufficiently large that $H + \lambda I$ si positive definite, we seek a value $\lambda > 0$ such that :

$$||p(\lambda)|| = \Delta \tag{20}$$

$\Rightarrow$ One dimensional root-finding problem in the variable λ

Since H is symmetric and real, there is an orthonormal matrix $\mathcal{Q}$ and a diagonal matrix Λ such that $H = \mathcal{Q}\Lambda\mathcal{Q}^T$,, where :

$$\Lambda = diag(\lambda_1, \lambda_2, ..., \lambda_n)$$

with $\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_n$ the eigenvalues of H.
So, $H + \lambda I = \mathcal{Q}(\Lambda + \lambda I)\mathcal{Q}^T$, and for $\lambda \neq \lambda_j$ we have :

$$p(\lambda) = -\mathcal{Q}(\Lambda + \lambda I)^{-1}\mathcal{Q}^T = -\sum_{j=1}^{n} \frac{q_j^T g}{\lambda_j + \lambda} \tag{21}$$

Where q_j is the jth column of the matrix $\mathcal{Q}$ contening the eigenvectors.

By orthonormality of the eigenvectors, we have :

$$||p(\lambda)||^2 = \sum_{j=1}^{n} \frac{(q_j^T g)^2}{(\lambda_j + \lambda)^2} \tag{22}$$

If $\lambda > -\lambda_1$, we have $\lambda_j + \lambda > 0$ for all j. So, $||p(\lambda)||$ is a continuous nonincreasing function of λ on $[-\lambda_1; \infty[$. We have :

$$\lim_{\lambda \to \infty} ||p(\lambda)|| = 0 \tag{23}$$

Moreover, we have when $q^T g \neq 0$ that :

$$\lim_{\lambda \to \lambda_j} ||p(\lambda)|| = \infty \tag{24}$$

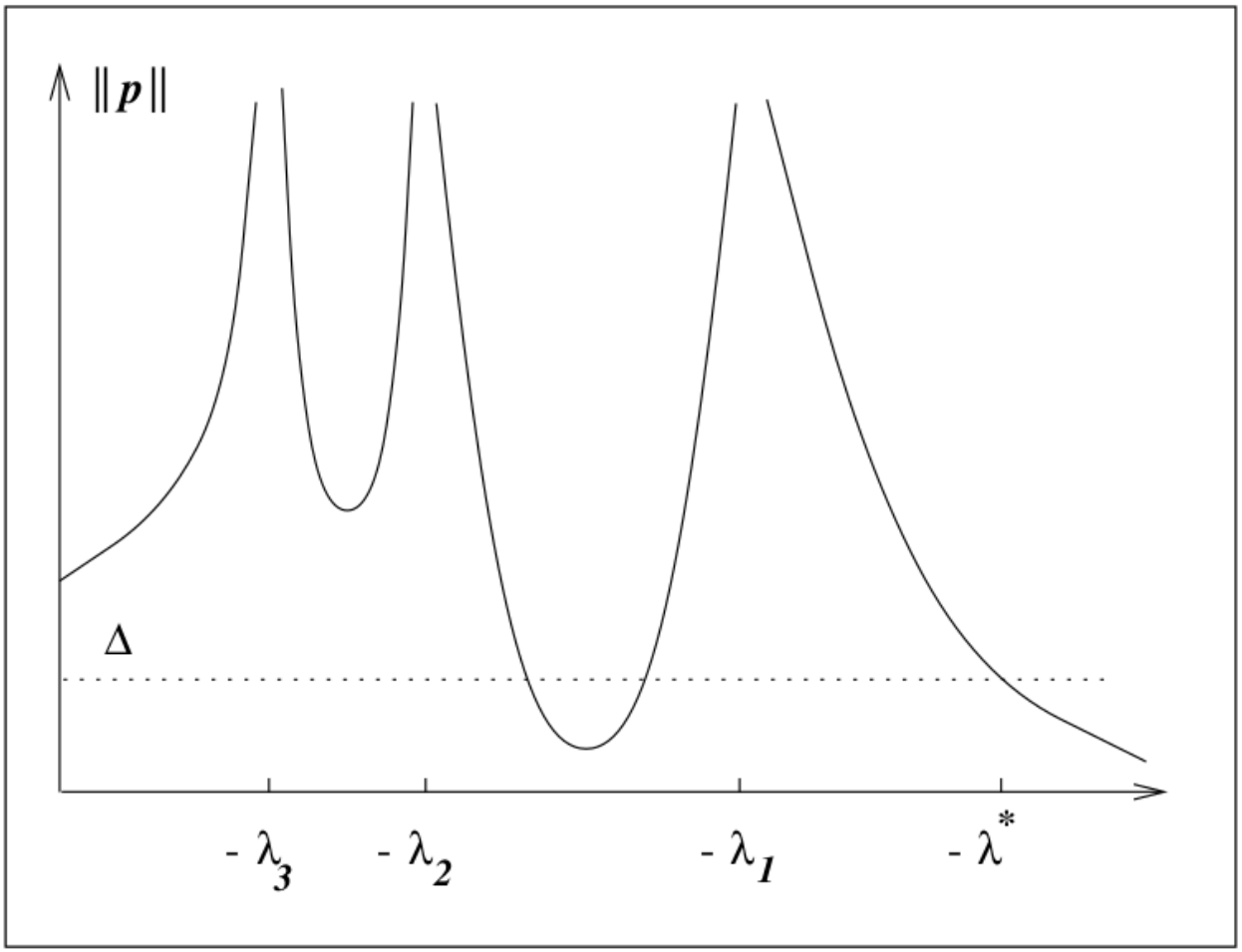

Figure 4.5 $\|p(\lambda)\|$ as a function of λ.

Figure: Nocedal & Wright (1999)

Finally, with (23) and (24), there is a point $\lambda^* \in [-\lambda_1, \infty[$ where $||p(\lambda)|| = \Delta$.

We said before, in the case where H is positive definite and we have $||H^{-1}g|| \leq \Delta$, the value $\lambda = 0$ satisfies the previous theorem and we don't have to search an other value of λ.

And when H is positive definite but $||H^{-1}g|| > \Delta$, there is a strictly positive value of λ for which $||p(\lambda)|| = \Delta$, so we search for the solution of $||p(\lambda)|| = \Delta$ with $\lambda \in]0, \infty[$.

For the case of H indefinite it is more complicated ...

Resolution of the system of equations

So, we have to solve a system of the kind :

$$Hx = -g \tag{25}$$

Yarkony (1981) introduce the idea that it could be computationally better to use the augmented matrix, also called augmented Hessian. In this case we determine x as an eigenvector x^a of the augmented matrix problem :

$$A - EI = 0 \tag{26}$$

With

$$\mathrm{A}= \begin{pmatrix} 0 & g^T \\ g & H \end{pmatrix}$$

Spectral decomposition

We performed a eigendecomposition of the two previous equation such as :

$$\bar{H}\bar{x} = -\bar{g} \tag{27}$$

Where :

$$\bar{H} = U^{-1}HU = \begin{pmatrix} \epsilon_1 & 0 & \dots & 0 \\ 0 & \epsilon_2 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \epsilon_n \end{pmatrix} \tag{28}$$

$$\bar{g} = U^{-1}g \tag{29}$$

Spectral decomposition

And :

$$\bar{A} - EI = 0 \tag{30}$$

Where :

$$\bar{A} = U_a^{-1} A U_a = \begin{pmatrix} 0 & \bar{g}^T \\ \bar{g} & \bar{H} \end{pmatrix} \tag{31}$$

With

$$U_a = \begin{pmatrix} 1 & 0^T \\ 0 & U \end{pmatrix} \tag{32}$$

Properties

Yarkony explain some properties :
"Since H is diagonal :

- the eigenvalues (E_i) of **A** satisfy the betweenness condition $\epsilon_{l-1} \leq E_l \leq \epsilon_l, \forall l$ (MacDonald's theorem)
- The eigenvectors (x^a) of $\bar{A}$ corresponding to the eigenvalue E_k has components :

$$x_0^a = 1, \quad x_l^a = -\bar{g}/(\epsilon_l - E_k) \tag{33}$$

- The solution of $\bar{H}\bar{x} = -\bar{g}$ is

$$x_l = -\bar{g}/\epsilon_l \tag{34}$$

- ...
- As $||g|| \to 0$, $x_l^a \to x_l$, so that the dampening is removed automatically when appropriate."

In her work, she used a similar method but I'm not sure what she did, because in my opinion that things are not well defined. She define the augmented Hessian :

$$A(\alpha) = \begin{pmatrix} 0 & \alpha \mathbf{g}^T \\ \alpha \mathbf{g} & \mathbf{H} \end{pmatrix} \tag{35}$$

In order to solve :

$$A(\alpha) \begin{pmatrix} 1 \\ \mathbf{x}(\alpha) \end{pmatrix} = \mu \begin{pmatrix} 1 \\ \mathbf{x}(\alpha) \end{pmatrix} \tag{36}$$

There are two components :

$$\alpha \mathbf{g}^T \mathbf{x}(\alpha) = \mu \tag{37}$$

$$(\mathbf{H} - \mu \mathbf{I})\mathbf{x}(\alpha) = -\alpha \mathbf{g} \tag{38}$$

In her Lagragian she used $-\mu$ instead of μ, for that reason there is a minus sign before μ.
"The parameter α may be ajusting for the lowest eigenvectors to ensure that

$$\alpha^{-1} ||\mathbf{x}(\alpha)|| = ||\mathbf{R}|| = \Delta" \tag{39}$$

Simplification

We have :

$$(\mathbf{H} + \lambda \mathbf{1}) = -\mathbf{g}$$

with λ the Lagrangian multiplier for the constraint $||\mathbf{R}|| = \Delta$, (Δ represents the radius of the trust region).
Then :

$$\mathbf{H} = \mathcal{Q}\Lambda\mathcal{Q}^T$$
$$\mathbf{H} + \lambda \mathbf{1} = \mathcal{Q}(\Lambda + \lambda \mathbf{1})\mathcal{Q}^T$$

$\mathcal{Q}$: the eigenvectors
Λ : diagonal matrix with the eigenvalues $\lambda_1 \leq \lambda_2 \leq .. \leq \lambda_n$

Simplification

We express p as a function of the constraint λ, with $\lambda \neq \lambda_j$

$$p(\lambda) = -\mathcal{Q}(\Lambda + \lambda \mathbf{1})^{-1}\mathcal{Q}^T g = -\sum_{j=1}^{n} \frac{q_j^T g}{\lambda_j + \lambda} q_j$$

with q_j the jth column of the matrix $\mathcal{Q}$. And by the orthonormality of the vectors $q_1, q_2, ..., q_n$:

$$||p(\lambda)||^2 = \sum_{j=1}^{n} \frac{(q_j^T g)^2}{(\lambda_j + \lambda)^2}$$

This leads to :

$$\lim_{\lambda \to \infty} ||p(\lambda)|| = 0$$

And when $q^T g \neq 0$:

$$\lim_{\lambda \to \lambda_j} ||p(\lambda)|| = \infty$$

Simplification

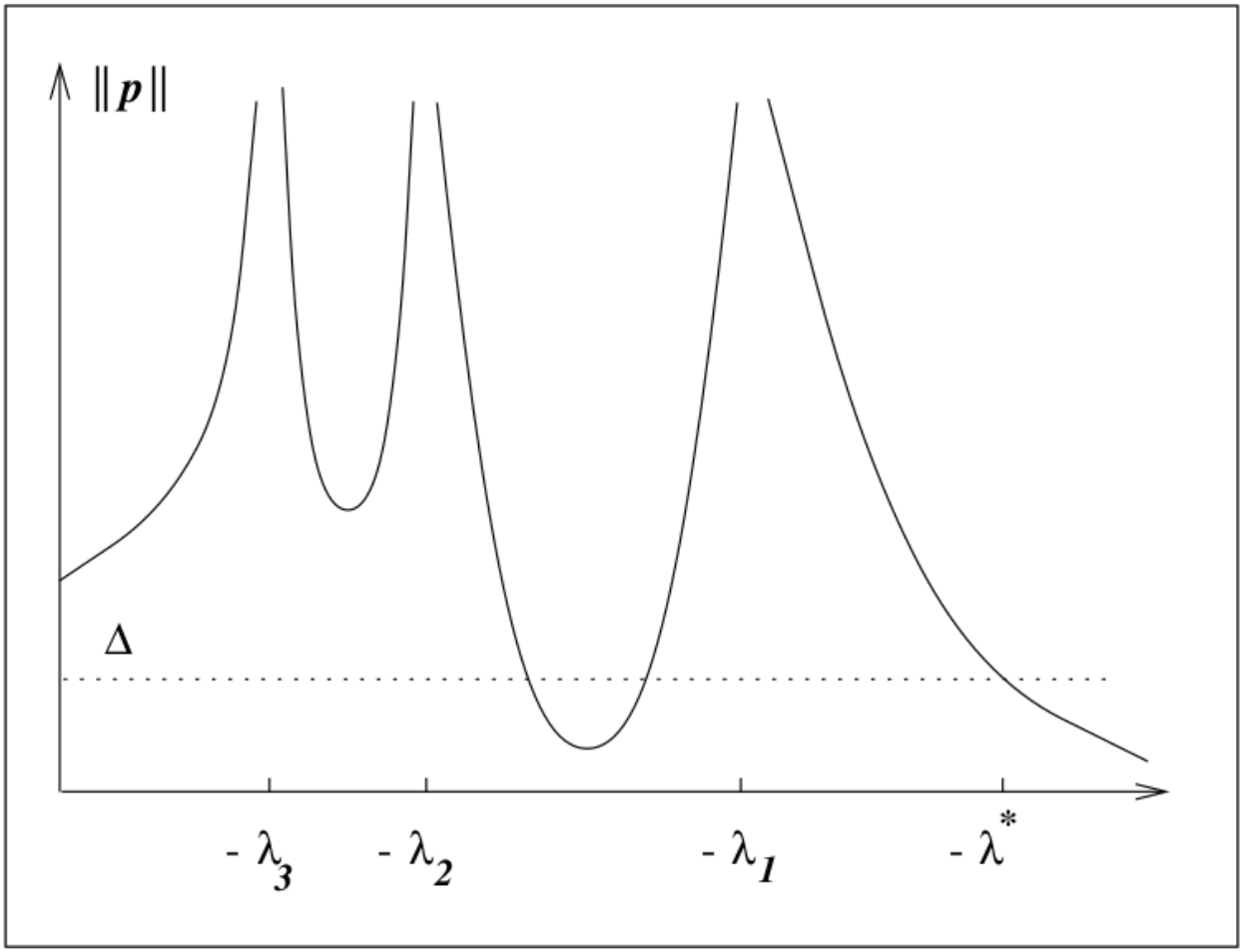

Figure 4.5 $\|p(\lambda)\|$ as a function of λ.

Figure: Nocedal & Wright (1999)

Simplification

If H is positive definite and

$$||H^{-1}g|| \leq \Delta$$

$\Rightarrow \lambda = 0$ (unconstraint solution)

If H is positive definite and

$$||H^{-1}g|| > \Delta$$

$\Rightarrow$ we search the $\lambda \in]0, \infty[$ solution of $||p(\lambda)|| = \Delta$

NB :
A matrix M (n by n) is positive definite if :

$$x^T Mx > 0 \qquad \forall x \in \mathbb{R}^n$$

Simplification

More efficient way to find R $\Rightarrow$ Augmented Hessian (m+1)(m+1)

$$\mathbf{A} = \begin{pmatrix} 0 & \mathbf{g}^T \\ \mathbf{g} & \mathbf{H} \end{pmatrix} = \begin{pmatrix} 0 & g_1 & g_2 & \dots & g_n \\ g_1 & H_{11} & H_{12} & \dots & H_{1n} \\ g_2 & H_{21} & H_{22} & \dots & H_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ g_n & H_{n1} & H_{n2} & \dots & H_{nn} \end{pmatrix}$$

And solve :

$$\mathbf{A} \begin{pmatrix} 1 \\ \mathbf{p} \end{pmatrix} = \mu \begin{pmatrix} 1 \\ \mathbf{p} \end{pmatrix}$$

$\mathbf{p} \rightarrow$ vector of lenght $m = n_{orb}(n_{orb} + 1)/2$
$\Leftrightarrow$ Lower/upper diagonal of the rotation matrix R
$\Rightarrow$ R (antisymmetric matrix)

The hessian matrix

We can compute the hessian matrix from the one and two body density matrix (Henderson, 2014)[2]...

Since we can express energy as :

$$\mathcal{E} = \sum_{pq} \gamma_{pq} \langle p|h|q\rangle + \sum_{pqrs} \Gamma_{pqrs} \langle pq||rs\rangle$$

And the orbital rotation is made by the exponentied one body antihermitian operator $\mathbf{R} = e^{\hat{X}}$, with ($\sigma = \uparrow, \downarrow$):

$$\hat{X} = \sum_{pq} \sum_{\sigma} X_{pq} (c^{\dagger}_{p\sigma} c_{q\sigma} + c^{\dagger}_{q\sigma} c_{p\sigma})$$

The energy is a function of X : $\mathcal{E}(X)$
Then we have :

$$\mathbf{g} = \frac{\partial \mathcal{E}(X)}{\partial X}, \quad \mathbf{H} = \frac{\partial^2 \mathcal{E}(X)}{\partial X^2}$$

[2] Henderson, Bulik, Stein, Scuseria,J. Chem. Phys. 141, 244104 (2014)

The hessian matrix

Henderson (2014) :
$\rightarrow$ use diagonal hessian far from the convergence to avoid the problem of the high energy local minima
$\rightarrow$ use the full hessian near the convergence

Main idea :

- Build the gradient and the hessian matrix
- Try the Newton method and the method of Umrigar & co on H_2
- Build the trust region method
- Use local orbitals as starting point ?
- Davidson diagonalisation
- Bigger systems ! C_5H_5 C_6H_6 ?
- Merge the orbitals optimization library in Quantum Package
- Add different criteria ?

Bibliography

Clyde Edmiston and Klaus Ruedenberg, Rev. Mod. Phys. 35, 457 (1963)
Janos Pipek, Paul G. Mezey, J. Chem. Phys. 90, 4916 (1989)
J. M. Foster, S. F. Boys, Rev. Mod. Phys. 32, 300 (1960)
Y. Yao, E. Giner, J. Li, J. Toulouse, C. J. Umrigar, J. Chem. Phys. 153, 124117 (2020)
Henderson, Bulik, Stein, Scuseria,J. Chem. Phys. 141, 244104 (2014)
Ida-Marie Høyvik, Thesis Local Hartree–Fock orbitals (2013)
Nocedal & Wright, Numerical Optimisation (1999)
Yarkony, Chem. Phys. Lett. 77, 634 (1981)
Michel Bergmann, Thesis (2004)

Thank you for your attention !