

Optimisation of distance in the Grassmannian to an external wave function

Yuri Alexandre Aoto

History:

24 mar 2019 - Start

31 oct 2019 - Adding Absil method

1 Notation

\mathcal{H}	The electronic Hamiltonian
n_α	The number of alpha electrons
n_β	The number of beta electrons
n	The total number of electrons $n = n_\alpha + n_\beta$
\mathcal{V}	The space of spatial one-electron wave functions (The full space of one-electron wave functions is assumed to be a direct sum of two such spaces, the first associated to α -spin and the second to β -spin electrons. A subscript might be used, but the space is the same: $\mathcal{V}_\alpha = \mathcal{V}_\beta = \mathcal{V}$)
N	The dimension of \mathcal{V} : $N = \dim \mathcal{V}$
i, j, k, l	orbital indices for occupied orbitals (the context should make clear the reference determinant)
a, b, c, d	orbital indices for virtual orbitals (the context should make clear the reference determinant)
$\sigma(p)$	spin (α or β) of orbital with index p
\mathcal{H}	The complete Hilbert space of quantum states (within finite basis set approximation) $\mathcal{H} = \Lambda^n(\mathcal{V}_\alpha \oplus \mathcal{V}_\beta)$
Ψ	General Elements of \mathcal{H}
Φ	Elements of $Gr(n, 2K) \in \mathcal{H}$, that is, they can be expressed as a Slater determinant.
ϕ, ψ	Elements of \mathcal{V}

2 Introduction

Let

$$|\Psi_{\text{ext}}\rangle = \sum_I c_I |\Phi_I\rangle \quad (1)$$

be a n -electron normalised CI like wave function represented in the orbital basis

$$|\phi_I\rangle = \phi_{I_1} \wedge \phi_{I_2} \wedge \cdots \wedge \phi_{I_n}, \quad (2)$$

where

$$\{\phi_p\}_{p=1}^K \quad (3)$$

is an orthonormal basis for the space of one-electron wave functions and I is an ordered multi-index. We want to find $|\Phi\rangle \in Gr$ such that $|\langle \Psi_{\text{FCI}} | \Phi \rangle|$ is maximum, where Gr is the image of the Grassmannian in the space of the n -electron wave functions.

Recall that

$$D(\Psi_1, \Psi_2) = \sqrt{2} \sqrt{1 - |\langle \Psi_1 | \Psi_2 \rangle|} \quad (4)$$

is a metric in $\mathbb{P}\mathcal{H}$.[] Also, Gr is the set of all elements of \mathcal{H} that can be written as a single Slater determinant (a decomposable element) for some orthonormal basis of \mathcal{V} . For this reason, such $|\Phi\rangle$ minimises the distance D to the wave function $|\Psi_{\text{ext}}\rangle$ and it will be denoted $|\Phi_{\text{minD}}\rangle$. The subscript “ext” refers to *external*, because we assume that $|\Psi_{\text{ext}}\rangle$ is external to the Grassmannian.

We have developed and implemented two algorithms to calculate $|\Phi_{\text{minD}}\rangle$. The first uses the whole structure of a Full-CI wave function (that is, the full Hilbert space), because the external wave function has to be transformed to the MO basis of $|\Phi_{\text{minD}}\rangle$. Hence, even if we start with a CISD wave function for example, when we update $|\Phi_{\text{minD}}\rangle$ the orbitals change, and the external wave function is no longer just a CISD over that reference. The second uses the Newton method adapted to the Grassmannian, as proposed by Absil *et al.*[]

In any case, this document is intended to give the full formulas that are implemented in the present code. The detailed derivation of these expressions are not given here.

3 Algorithm 1

Using orbital rotations and the structure of a FCI wave function

To find $|\Phi_{\min D}\rangle$, it is equivalent to find the orbitals (namely, a basis of \mathcal{V}) such that

$$|\Psi_{\text{FCI}}\rangle = C_0 |\Phi_{\min D}\rangle + \sum_{\substack{I \\ I \neq \{1,2,\dots,n\}}} c_I |\Phi_I\rangle, \quad (5)$$

and $|C_0|$ is maximum (over all possible coefficients in all possible basis), since $|C_0| = |\langle \Psi_{\text{FCI}} | \Phi \rangle|$.

3.1 Parametrisation by orbital rotations

We parametrise G_r by the orbital rotations as[]

$$|\Phi\rangle = e^{-\hat{K}} |\Phi_0\rangle, \quad (6)$$

where

$$\hat{K} = \sum_{i,a} K_i^a (a_a^\dagger a_i - a_i^\dagger a_a) \quad (7)$$

$$= \sum_{i,a} K_i^a (a_i^a - a_a^i) \quad (8)$$

This parametrisation comes from the most general

$$\hat{K} = \sum_{p,q} K_q^p a_p^\dagger a_q, \quad (9)$$

but using that K_q^p is anti-symmetric (so that $e^{-\hat{K}}$ is orthogonal), and excluding rotations within the occupied or virtual spaces of $|\Phi_0\rangle$, that are redundant (do not alter the Slater determinant with $|\Phi_0\rangle$). For $\hat{K} = 0$ it is clear that $|\Phi\rangle = |\Phi_0\rangle$.

Let

$$f(K_i^a) = \left| \langle \Psi_{\text{FCI}} | e^{-\hat{K}} | \Phi_0 \rangle \right|, \quad (10)$$

where the argument K_i^a represent all the $n_\alpha(N - n_\alpha) + n_\beta(N - n_\beta)$ elements. Note that rotations that mix α and β orbitals. We will also assume that $\langle \Psi_{\text{FCI}} | \Phi_0 \rangle > 0$ and this remains true for all steps of our optimisation. If $\langle \Psi_{\text{FCI}} | \Phi_0 \rangle < 0$ we of course can change the phase of the wave function and if $\langle \Psi_{\text{FCI}} | \Phi_0 \rangle = 0$ for the first or any step of the optimisation, we likely started with a very poor initial guess.

3.2 Jacobian and Hessian

We want to maximise f and we need its Jacobian and Hessian. The expressions at $\hat{K} = 0$ are given below. For the derivation of the expressions, see that hand notes.

$$\frac{\partial f(\hat{K}=0)}{\partial K_i^a} = (-1)^{n_{\sigma(i)}-i+1} C_i^a \quad (11)$$

$$\frac{\partial^2 f(\hat{K}=0)}{\partial K_i^a \partial K_j^b} = \begin{cases} -C_0 & (i=j, a=b) \\ 0 & (i \neq j, a=b) \\ 0 & (i=j, a \neq b) \\ (-1)^{n_{\sigma(i)}+n_{\sigma(j)}-i-j} C_{ij}^{ab} & (\sigma(i) \neq \sigma(j)) \\ (-1)^{i+j+1} C_{ij}^{ab} & (\sigma(i) = \sigma(j), i < j, a < b) \\ (-1)^{i+j} C_{ij}^{ab} & (\sigma(i) = \sigma(j), i < j, a > b) \end{cases} \quad (12)$$

In these equations, C_i^a and C_{ij}^{ab} are the CI coefficients of the single and double excited determinants in the (normalised) wave function $|\Psi_{\text{FCI}}\rangle$. The canonical order of the orbitals is assumed to be “first all α , then all β ”.

3.3 Transformation of the wave function

In the optimisation process, $|\Phi\rangle$ varies and we would need the Jacobian and the Hessian at $\hat{K} \neq 0$. The expressions are much more complicated and we avoid this by making a full transformation of $|\Psi_{\text{FCI}}\rangle$ to the new orbital basis. Let $U = e^{-\hat{K}}$ be the matrix that transform the orbital basis:

$$\phi_p = \sum_q \phi'_q U_{qp}. \quad (13)$$

Given the coefficients C_I of the expansion in the first basis, we want to know the coefficients C'_I such that

$$|\Psi_{\text{FCI}}\rangle = \sum_I c_I |\Phi_I\rangle = \sum_I c'_I |\Phi'_I\rangle. \quad (14)$$

These are given by:

$$C'_I = \sum_J C_J \det(U_{IJ}), \quad (15)$$

where U_{IJ} is the minor of the matrix U with the entries in the rows and columns given by the multi-indices I and J .

Such transformation is the most time consuming step.

3.4 Newton-Raphson step

Starting from a orbital basis $\{\phi_p\}$ such that the first determinant (that is, with the first n_α α orbitals and first n_β β occupied) is $|\Phi_0\rangle$, we calculate the Jacobian \mathbf{J} and the Hessian \mathbf{H} as shown above. The Newton step (in the space of the K_i^a parameters) is

$$\mathbf{z} = -\mathbf{H}^{-1}\mathbf{J}. \quad (16)$$

From this vector, the operator \hat{K} is constructed and the orbital transformation matrix is given by

$$U = e^{-\hat{K}}. \quad (17)$$

This is done for the α and the β orbitals and the wave function $|\Psi_{\text{FCI}}\rangle$ is transformed to the new orbital basis, by equation 15. This proceeds until convergence.

4 Algorithm 2

The Newton-Grassmann method of Absil

This method is based on the work of Absil *et al.*, where they propose a Newton method adapted to the geometry of the Grassmannian. The algorithm is actually on the Stiefel manifold, that is, the manifold of full-rank $K \times n$ matrices. These matrices contain the coefficients of the orbitals in a fixed basis, and thus form a redundant parametrisation for the Grassmannian, in the sense that there are several matrices to represent the same point in Gr , that can be transformed among themselves with multiplication by invertible $n \times n$ matrices (elements of the general linear group of degree n , GL_n).

4.1 General formulation from Absil

Thus, given $|\Phi_0\rangle \in Gr$, let Y be the matrix whose entries are the coefficients of $|\Phi_0\rangle$ in some basis of \mathcal{V} . This is an element of the Stiefel manifold $ST(n, K)$. Let $f : Gr \rightarrow \mathbb{R}$ be any function we might want to optimise. This is defined on the Grassmannian. We will recall some definitions from Absil.[] The version of this function on the Stiefel manifold is defined as:

$$f_\diamond : ST \rightarrow \mathbb{R} \quad (18)$$

$$Y \mapsto f(\text{span}(Y)), \quad (19)$$

where $\text{span}(Y)$ is the vector space spanned by the columns of Y (and thus in Gr). The gradient of f_\diamond at Y is the matrix $K \times n$ whose entries are given by:

$$(\text{grad} f_\diamond(Y))_{ij} = \frac{\partial f_\diamond(Y)}{\partial Y_{ij}}(Y), \quad (20)$$

and the directional derivative of a function any smooth F , at x in the direction of y :

$$DF(x)[y] = \left. \frac{d}{dt} F(x + ty) \right|_{t=0}. \quad (21)$$

Finally, let

$$\Pi_{W^\perp} = I - W(W^T W)^{-1} W^T, \quad (22)$$

be the projection into the orthogonal component of W .

The calculation is done now in the Stiefel manifold. According to Absil:

- First solve the following equation for the unknown $\eta_{\diamond Y} \in H_Y = \{Y_\perp K : K \in \mathbb{R}^{(K-n) \times n}\}$:

$$\Pi_{Y^\perp} D(\Pi_{\perp} \text{grad} f_\diamond(\cdot))(Y) [\eta_{\diamond Y}] = -\Pi_{Y^\perp} \text{grad} f_\diamond(Y); \quad (23)$$

- Update $|\Phi_0\rangle \rightarrow |\Phi\rangle$ by moving along the geodesic on the Grassmannian in the direction of $\text{span}(\eta_{\diamond Y})$, what is done computing a singular value decomposition of $\eta_{\diamond Y} = U \Sigma V^T$ and getting:

$$|\Phi\rangle = \text{span}(YV \cos \Sigma + U \sin \Sigma). \quad (24)$$

4.2 Application to obtain $|\Phi_{\text{minD}}\rangle$

We will now consider our specific function. Similarly to what we did in the first algorithm, we will just work without getting the absolute value. Thus, we will look for stationary points of:

$$f : Gr \rightarrow \mathbb{R} \quad (25)$$

$$|\Phi\rangle \mapsto \langle \Psi_{\text{ext}} | \Phi \rangle. \quad (26)$$

We will assume that the external wave function, $|\Psi_{\text{ext}}\rangle$, is eigenfunction of S_z , with eigenvalue M_s , and belongs to an irreducible representation (irrep, Γ) of the point group of the molecule. This means that, for all determinants of the expansion 1, the following equations hold:

$$n_\alpha - n_\beta = 2M_s \quad (27)$$

$$\bigotimes_i \Gamma_i = \Gamma \quad (28)$$

That is, all determinants have the same M_s (that is half of the difference between alpha and beta orbitals) and belong to the same irrep Γ (that is the direct product of the irrep of all occupied orbitals). We are thus assuming that we are using a basis for \mathcal{V} for which all elements belong to an irrep. Let us further assume that the orbitals are ordered with all alpha coming first, followed by all beta orbitals, and they are ordered by irrep in each of these parts. Under these conditions, the f_\diamond is calculated as follows:

$$f_\diamond(Y) = \frac{1}{\sqrt{\det(Y^T Y)}} \sum_I c_I \det(Y|_I) \quad (29)$$

$$= \frac{1}{\sqrt{\det((Y_\alpha)^T Y_\alpha) \det((Y_\beta)^T Y_\beta)}} \sum_I c_I \det(Y_\alpha|_{I_\alpha}) \det(Y_\beta|_{I_\beta}) \quad (30)$$

$$= \frac{1}{\sqrt{\prod_{i=0}^{g-1} \det((Y_\alpha^{\Gamma_i})^T Y_\alpha^{\Gamma_i}) \det((Y_\beta^{\Gamma_i})^T Y_\beta^{\Gamma_i})}} \sum_{I \text{ occ}(I)=\text{occ}(Y)} c_I \prod_{i=0}^{g-1} \det(Y_\alpha^{\Gamma_i}|_{I_\alpha^{\Gamma_i}}) \det(Y_\beta^{\Gamma_i}|_{I_\beta^{\Gamma_i}}) \quad (31)$$

The above expressions seem very complicated, so let us look at them carefully.

First of all, note that the first expression is general, in the second we consider the separation between α and β orbitals, and in the third we consider the separation between orbitals of different irreps. In all versions, the term with the square root is to guarantee the normalization of the Slater determinant associated to Y . If the coefficients in Y are for an orthonormal set of orbitals, the term in the square root is one. But we note that $f_\diamond(Y) = f(\text{span}(Y))$ must hold for all Y in the Stiefel manifold, and the normalisation factor is important to use equation 23, even if Y is used normalised during computation.

The notation $Y|_I$ means the submatrix of Y formed with the rows that are in the multiindex I . For example, if

$$Y = \begin{pmatrix} Y_{00} & Y_{01} & Y_{02} \\ Y_{10} & Y_{11} & Y_{12} \\ Y_{20} & Y_{21} & Y_{22} \\ Y_{30} & Y_{31} & Y_{32} \\ Y_{40} & Y_{41} & Y_{42} \end{pmatrix}, \quad (32)$$

and $I = \{0, 1, 2\}$, then

$$Y|_I = \begin{pmatrix} Y_{00} & Y_{01} & Y_{02} \\ Y_{10} & Y_{11} & Y_{12} \\ Y_{20} & Y_{21} & Y_{22} \end{pmatrix}, \quad (33)$$

but if $I = \{0, 3, 4\}$, then

$$Y|_I = \begin{pmatrix} Y_{00} & Y_{01} & Y_{02} \\ Y_{30} & Y_{31} & Y_{32} \\ Y_{40} & Y_{41} & Y_{42} \end{pmatrix}, \quad (34)$$

and so on. Thus, the first formula says that, after taking the normalisation into account, f_\diamond is calculated by summing the contribution of all determinants that contribute to $|\Psi_{\text{ext}}\rangle$, and this contribution is the CI coefficient times the determinant of $Y|_I$.

Now, if the orbitals of a Slater determinant $|\Phi\rangle$ can be divided into α and β orbitals, that do not mix, the matrix Y is divided in two blocks, one for each spin:

$$Y = \begin{pmatrix} Y_\alpha & 0 \\ 0 & Y_\beta \end{pmatrix} = Y_\alpha \oplus Y_\beta. \quad (35)$$

The second equation follows easily from the first when we use this division into blocks, because the determinant of block diagonal matrices is the product of the determinant of each block. Furthermore, when we consider symmetry the matrix is further divided:

$$Y = \begin{pmatrix} Y_\alpha^{\Gamma_0} & 0 & \dots & 0 \\ 0 & Y_\alpha^{\Gamma_1} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & Y_\beta^{\Gamma_{g-1}} \end{pmatrix} = Y_\alpha^{\Gamma_0} \oplus Y_\alpha^{\Gamma_1} \oplus \dots \oplus Y_\beta^{\Gamma_{g-1}}, \quad (36)$$

leading to the last formula. In this case, however, the matrices $Y_\sigma^{\Gamma_i}$ might not be square, because by fixing the irrep of the wave function (and thus of its determinants) does not fix the occupation in each irrep block: it is possible to fulfil equation 28 with more than one pattern in the occupation of the irreps Γ_i . In the summation over I , if this multi-index is associated to a determinant with an occupation per irrep different than the occupation of Y , some of the blocks $Y_\sigma^{\Gamma_i}$ are not square, although the direct sum of all of them is again square (because the total number of electrons do not change). However, one can show that the determinant of such block matrices are zero, and thus they do not contribute to f_\diamond . This is indicated in the summation by “ $\text{occ}(I) = \text{occ}(Y)$ ”, that is, we can consider only the terms with the same occupation per irrep of Y .

The notation $I_\sigma^{\Gamma_i}$ indicates the part of the multi-index I associated to σ and Γ_i . Because it is obvious that we can take submatrices of $Y_\alpha^{\Gamma_i}$ only restricting with subindices with same spin and irrep, the notation $Y_\alpha^{\Gamma_i}|_{I_\alpha^{\Gamma_i}}$ is rather redundant. Hence we will use $Y_\alpha^{\Gamma_i}|_I$ or $Y|_{I_\alpha^{\Gamma_i}}$ only.

Because of the separation in spin and irrep blocks, we can show that the unknown matrix $\eta_{\diamond Y}$ has the same division as the matrix Y :

$$\eta_{\diamond Y} = \begin{pmatrix} \eta_\alpha^{\Gamma_0} & 0 & \dots & 0 \\ 0 & \eta_\alpha^{\Gamma_1} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \eta_\beta^{\Gamma_{g-1}} \end{pmatrix} = \eta_\alpha^{\Gamma_0} \oplus \eta_\alpha^{\Gamma_1} \oplus \dots \oplus \eta_\beta^{\Gamma_{g-1}}, \quad (37)$$

where we dropped the subscript $\diamond Y$ in the blocks for a cleaner notation.

After a big exercise of linear algebra, we obtain the following system of linear equations as the main equation of this method (i.e. Equation 23 using the present function f):

$$(X_{\sigma\Gamma}^{\sigma'\Gamma'})_{jk}^{il} (\eta_{\sigma'\Gamma'})_l^k = (C_{\sigma\Gamma})_j^i, \quad (38)$$

with

$$(C_{\sigma\Gamma})_j^i = \sum_{\substack{I \\ \text{occ}(I)=\text{occ}(Y)}} c_I \left(\prod_{\{\sigma',\Gamma'\} \neq \{\sigma,\Gamma\}} F_{I_{\sigma'}^{\Gamma'}} \right) (G_{I_{\sigma}^{\Gamma}} - F_{I_{\sigma}^{\Gamma}} Y_{\sigma}^{\Gamma})_j^i \quad (39)$$

$$(X_{\sigma\Gamma}^{\sigma\Gamma})_{jk}^{il} = \sum_{\substack{I \\ \text{occ}(I)=\text{occ}(Y)}} c_I \left(\prod_{\{\sigma',\Gamma'\} \neq \{\sigma,\Gamma\}} F_{I_{\sigma'}^{\Gamma'}} \right) \left\{ (G_{I_{\sigma}^{\Gamma}} - F_{I_{\sigma}^{\Gamma}} Y_{\sigma}^{\Gamma})_j^i ((Y_{\sigma}^{\Gamma})^T)_k^l - (\Pi_{Y_{\sigma}^{\Gamma} \perp})_{\mu}^i (\tilde{H}_{I_{\sigma}^{\Gamma}})_{jk}^{\mu l} \right\} \quad (40)$$

$$= \sum_{\substack{I \\ \text{occ}(I)=\text{occ}(Y)}} c_I \left(\prod_{\{\sigma',\Gamma'\} \neq \{\sigma,\Gamma\}} F_{I_{\sigma'}^{\Gamma'}} \right) (\Pi_{Y_{\sigma}^{\Gamma} \perp})_{\mu}^i \left((G_{I_{\sigma}^{\Gamma}})_j^{\mu} ((Y_{\sigma}^{\Gamma})^T)_k^l - (\tilde{H}_{I_{\sigma}^{\Gamma}})_{jk}^{\mu l} \right) \quad (41)$$

$$(X_{\sigma\Gamma}^{\sigma'\Gamma'})_{jk}^{il} = - (1 - \delta_{\{\sigma,\Gamma\} \{\sigma',\Gamma'\}}) \sum_{\substack{I \\ \text{occ}(I)=\text{occ}(Y)}} c_I \left(\prod_{\substack{\{\sigma'',\Gamma''\} \neq \{\sigma',\Gamma'\} \\ \{\sigma'',\Gamma''\} \neq \{\sigma,\Gamma\}}} F_{I_{\sigma''}^{\Gamma''}} \right) (G_{I_{\sigma}^{\Gamma}} - F_{I_{\sigma}^{\Gamma}} Y_{\sigma}^{\Gamma})_j^i \left((G_{I_{\sigma'}^{\Gamma'}} - F_{I_{\sigma'}^{\Gamma'}} Y_{\sigma'}^{\Gamma'})^T \right)_k^l \quad (42)$$

where:

$$F_{I_{\sigma}^{\Gamma}} = \det(Y|_{I_{\sigma}^{\Gamma}}) \quad (43)$$

$$(G_{I_{\sigma}^{\Gamma}})_j^i = \det(Y_{\alpha}^{\Gamma} \stackrel{j}{\leftarrow} e_i) |_I \quad (44)$$

$$(H_{I_{\sigma}^{\Gamma}})_{jk}^{il} = \det(Y_{\alpha}^{\Gamma} \stackrel{j}{\leftarrow} e_i \stackrel{l}{\leftarrow} e_k) |_I \quad (45)$$

$$(\tilde{H}_{I_{\sigma}^{\Gamma}})_{jk}^{il} = \begin{cases} (H_{I_{\sigma}^{\Gamma}})_{jk}^{il} & \text{if } l \neq j \\ -F_{I_{\sigma}^{\Gamma}} \delta_{ik} & \text{otherwise} \end{cases} \quad (46)$$

The notation $A \stackrel{j}{\leftarrow} b$ means the matrix A with the j -th column replaced by the vector b . The vector e_i is the i -th element of the canonical basis (0 in every entry, except for the i -th entry where it is 1).