May 16, 2026 STRASBOURG STUDENTS PHYSICAL LETTERS **2**, 0101

Towards a description of molecular excited states: Analytical derivations for Orthogonally Constrained CASSCF*

Loris Delafosse†
Laboratoire de Chimie Quantique de Strasbourg, Institute of Chemistry - UMR 7177,
CNRS/University of Strasbourg, 4 rue Blaise Pascal, 67000 Strasbourg, France

Despite decades of research, obtaining the electronic structure of relatively large molecules remains an open problem, as describing the correlation between the motions of many electrons is extremely difficult. The most successful post-Hartree-Fock methods use some variational ansatz to tackle this problem, often retrieving excellent approximations of the ground state energy and wavefunction. Obtaining the excited states, which are necessary to photochemistry or molecular dynamics computations, is however out of the scope of such methods. In a previous work by YALOUZ and ROBERT (*J. Chem. Theory Comput.* **19**, 5, 2023), an orthogonality constraint was successfully introduced in the variational problem to account for the excited states. The present paper develops this approach by obtaining analytical expressions for the Gradients and Hessians needed to solve the variational problem. A theoretical lower bound on the Lagrange multipliers is then calculated, while considerations on error control allow us to set them to a known and physically sensible value.

INTRODUCTION

It is well-known that finding analytical solutions to the quantum eigenvalue problem of systems larger than the non-relativistic hydrogenoid atom is impossible. For systems that remain small (molecules of a few light atoms), it is possible to write and numerically diagonalize the exact Hamiltonian (though in a chosen finite basis). This brute-force method, known as *Full Configuration Interaction* (FCI), becomes inoperable for molecules consisting of tens of atoms, as the dimension of the Hilbert space grows exponentially with the number of electrons.

Fortunately, for most chemical systems, only the ground state is of interest, allowing us to make use of a variational principle. Indeed, the system's energy is known to be a functional of the wavefunction

$$E[\Psi] = \frac{\langle\Psi|\hat{H}|\Psi\rangle}{\langle\Psi|\Psi\rangle} \tag{1}$$

where $\hat{H}$ is the system's Hamiltonian. Now, Ritz' theorem states that the stationary wavefunctions (*i.e.* the Ψ for which the functional differential vanishes: $\delta E = 0$) are exactly the energy eigenfunctions. Therefore, minimising $E[\Psi]$ yields the ground state energy and wavefunction. This minimisation is often implemented by a *Self-Consistent Field* (SCF) approach, which consists in solving iteratively a set of equations resulting from the minimisation condition [1]. More details regarding the choice of manifold in which Ψ is allowed to vary, and the numerical implementation in general, are given in Section I.

However, the knowledge of the ground state is not sufficient for all applications, notably in photochemistry and molecular dynamics. Assuming a known ground state Ψ^0, YALOUZ and ROBERT proposed to adapt the usual variational method by minimising the energy under the constraint of orthogonality with Ψ^0 [2, 3]. Then, according to Ritz' theorem, the variational method should yield the first excited state. This method is described in Section II A. Its numerical implementation requires the calculation of the Gradients and Hessians of the cost function, whose analytical derivations are given in Section II B.

Section III is devoted to a discussion of some mathematical aspects of the proposed optimisation algorithm, and most of all to the calculation of the Lagrange multipliers associated with the orthogonality constraint. Issues related to preventing error propagation from the ground to the excited states are investigated.

NOMENCLATURE

Throughout this article, we will use a slight variant of Ricci's notation for (anti)symmetrising tensors:

$$A_{(pq)} = A_{pq} + A_{qp}, \tag{2a}$$

$$A_{[pq]} = A_{pq} - A_{qp}. \tag{2b}$$

Latin lower-case indices run over orbitals, or over active configurations if they are underlined, and Latin upper-case over energy eigenstates.

We also define the symmetric transpose operator $\mathcal{T}$, which is a rank-8 tensor acting on rank-4 tensors:

$$\mathcal{T}^{(pq)}_{(rs)}(a_{pqrs}) \overset{\text{def}}{=} \frac{1}{2}(a_{pqrs} + a_{rspq}). \tag{3}$$

---

[*] Project under the mentorship of Dr. Saad Yalouz and Pr. Vincent Robert from the Strasbourg Quantum Chemistry Laboratory.

[†] loris.delafosse@unistra.fr

I. THE COMPLETE ACTIVE SPACE SELF-CONSISTENT FIELD

A. Notion of configuration

The notion of *configuration* is of major importance in quantum chemistry, as it produces an interpretation of quantum states, and provides a guideline for many electronic structure methods.

We call *orbital* a one-electron spatial wavefunction. A single orbital can be occupied by up to two electrons because of spin degeneracy. A *configuration* (a.k.a. a Slater determinant, or a Fock state) is a quantum state of the system in which each electron is assigned to a particular orbital. Of course, a general quantum state should be a superposition of configurations. The inability of electrons to belong to one precise configuration is called *electron correlation* and accounts for the deviation from the mean-field motion of the electrons.

B. Multi-configurational approaches

As mentioned before, Self-Consistent Field approaches consist in minimising the energy functional of eq. (1). In practice, Ψ cannot span the full infinite-dimensional Hilbert space during the optimisation process, so we need to choose a manifold in which it will be allowed to vary. Imposing Ψ to be a single configuration yields the *Hartree-Fock* (HF) method [1]. However, a single configuration means we do not take into account electron correlation. We should allow Ψ to be a superposition of configurations, an extension of HF known as *Multi-Configurational Self-Consistent Field* (MCSCF) [4].

However, as Ψ could contain an arbitrary number of configurations, each weighted with a varying coefficient, this approach makes the variational manifold extremely large. An astute strategy to reduce the manifold's size consists in considering only the configurations in which certain orbitals are necessarily doubly occupied (the *frozen* orbitals), others are necessarily empty (the *virtual* orbitals), and a few only can be freely occupied (or not) by the electrons (the *active* orbitals). Such configurations (consisting of frozen, active and virtual orbitals) are called active configurations in the following.

This approach, known as *Complete Active Space Self-Consistent Field* (CASSCF), amounts to considering only the valence electrons, because the other ones occupy fixed orbitals of lower energy. Details on MCSCF and CASSCF can be found in ref. [4].

Let us write the expression of the CAS variational ansatz

$$|\Psi(\mathbf{c},\boldsymbol{\kappa})\rangle = \mathrm{e}^{-\hat{\kappa}}\,|\psi_{\mathbf{c}}\rangle \,. \tag{4}$$

where $\mathbf{c},\boldsymbol{\kappa}$ are the variational parameters, respectively the configuration coefficients c_I and the orbital coefficients κ_{pq} (upper-case Latin indices I,J run over active configurations, while lower-case Latin indices p,q,r,s run over orbitals). Configuration coefficients are contained within $|\psi_{\mathbf{c}}\rangle$, while orbital coefficients are contained within the anti-Hermitian operator $\hat{\kappa}$, making $\mathrm{e}^{-\hat{\kappa}}$ a unitary operator. Setting all these parameters to zero gives the reference state, which is the initial guess from which we try to determine a better approximation of the exact state. We write it $|\Psi(\mathbf{0},\mathbf{0})\rangle \overset{\text{def}}{=} |\psi\rangle$. Equation (4) implies the following expression of the energy:

$$E(\boldsymbol{\xi}) = \langle\psi_{\mathbf{c}}|\mathrm{e}^{\hat{\kappa}}\hat{H}\mathrm{e}^{-\hat{\kappa}}|\psi_{\mathbf{c}}\rangle \tag{5}$$

where $\hat{H}$ is the Hamiltonian, and $\boldsymbol{\xi} = (\mathbf{c},\boldsymbol{\kappa})$ are the variational parameters. More details on the form of the ansatz are given in Appendix A.

The CASSCF method involves two simultaneous optimisation processes. On the one hand, it finds the configuration coefficients $\mathbf{c}$ that weight each of the active configurations in $|\psi_{\mathbf{c}}\rangle$. On the other hand, it adjusts the orbital coefficients $\boldsymbol{\kappa}$ that rotate the orbitals and thus change the meaning of the configurations. Indeed, it is possible for a quantum state to be composed of a superposition of, *e.g.* ten configurations in an orbital basis $\mathcal{B}_1$, and of only two configurations in another orbital basis $\mathcal{B}_2$. In this case, the CASSCF method may return good results even if there are only two active configurations, because the orbital optimisation will rotate the orbital basis from $\mathcal{B}_1$ to $\mathcal{B}_2$. From a quantum information perspective, simultaneously adjusting the configuration coefficients and the shape of the orbitals in CASSCF methods provides a way to condense quantum information into a physically significant ansatz.

C. Molecular integrals and reduced density matrices

Before we present the numerical approach we will follow to implement the CASSCF method, we need to introduce a few mathematical objects which will allow us to express all quantities of interest.

First of all, we define the one-electron *excitation operator* $\hat{E}_{pq}$, which is used to excite electrons from the q-th orbital to the p-th orbital. Similarly, the two-electron excitation operator $\hat{e}_{pqrs}$ excites electrons from the q-th to the p-th orbital, and from the s-th to the r-th orbital. More details are given in Appendix A.

A quantum state $|\psi\rangle$ can now be characterised by the one- and two-electron *reduced density matrices* (RDMs), defined as

$$\gamma^{\psi,\psi}_{pq} = \langle\psi|\hat{E}_{pq}|\psi\rangle = \gamma^{\psi,\psi}_{qp}, \tag{6a}$$

$$\Gamma^{\psi,\psi}_{pqrs} = \langle\psi|\hat{e}_{pqrs}|\psi\rangle = \Gamma^{\psi,\psi}_{qpsr}. \tag{6b}$$

Any quantity involving two quantum states $|\psi^0\rangle$ and $|\psi^1\rangle$ can be described by the one- and two-electron *tran-*

sition reduced density matrices (TRDMs), defined as

$$\begin{aligned}\gamma_{pq}^{\psi^1,\psi^0} &= \left\langle \psi^1 \middle| \hat{E}_{pq} \middle| \psi^0 \right\rangle \\ &= \left\langle \psi^0 \middle| \hat{E}_{qp} \middle| \psi^1 \right\rangle = \gamma_{qp}^{\psi^0,\psi^1},\end{aligned} \tag{7a}$$

$$\begin{aligned}\Gamma_{pqrs}^{\psi^1,\psi^0} = \Gamma_{rspq}^{\psi^1,\psi^0} &= \left\langle \psi^1 \middle| \hat{e}_{pqrs} \middle| \psi^0 \right\rangle \\ &= \left\langle \psi^0 \middle| \hat{e}_{srqp} \middle| \psi^1 \right\rangle = \Gamma_{srqp}^{\psi^0,\psi^1} = \Gamma_{qpsr}^{\psi^0,\psi^1}.\end{aligned} \tag{7b}$$

(We assume all integrals to be real.)

Finally, a quantum chemical system is completely characterised by the one- and two-electron *molecular integrals* $\left(h_{pq}\right)_{pq}$ and $\left(g_{pqrs}\right)_{pqrs}$, whose precise definition can be found in ref. [4].

D. The Newton step algorithm

Minimising the energy function of eq. (5) is a challenge that can be addressed through various methods. One of the most commonly used is the Newton step algorithm, described in this subsection. It consists first of all in computing the first- and second-order derivatives (Gradient and Hessian) of the energy, with respect to the various variational parameters, at the point $\boldsymbol{\xi} = \mathbf{0}$. Derivatives corresponding to different variational parameters must be computed separately [4], that is:

- $\mathbf{G}^c$ the configurational Gradient $\left(G_I^{\mathrm{E}}\right)_I$

$$G_I^{\mathrm{E}} \overset{\mathrm{def}}{=} \left.\frac{\partial E}{\partial c_I}\right|_{\boldsymbol{\xi}=\mathbf{0}}, \tag{8}$$

- $\mathbf{G}^o$ the orbital Gradient $\left(G_{pq}^{\mathrm{E}}\right)_{p,q}$

$$G_{pq}^{\mathrm{E}} \overset{\mathrm{def}}{=} \left.\frac{\partial E}{\partial \kappa_{pq}}\right|_{\boldsymbol{\xi}=\mathbf{0}}, \tag{9}$$

- $\mathbf{H}^{cc}$ the configurational Hessian $\left(H_{IJ}^{\mathrm{E}}\right)_{I,J}$

$$H_{IJ}^{\mathrm{E}} \overset{\mathrm{def}}{=} \left.\frac{\partial^2 E}{\partial c_I \partial c_J}\right|_{\boldsymbol{\xi}=\mathbf{0}}, \tag{10}$$

- $\mathbf{H}^{oo}$ the orbital Hessian $\left(H_{pqrs}^{\mathrm{E}}\right)_{p,q,r,s}$

$$H_{pqrs}^{\mathrm{E}} \overset{\mathrm{def}}{=} \left.\frac{\partial^2 E}{\partial \kappa_{pq} \partial \kappa_{rs}}\right|_{\boldsymbol{\xi}=\mathbf{0}}, \tag{11}$$

- $\mathbf{H}^{oc} = \left(\mathbf{H}^{co}\right)^T$ the mixed Hessian $\left(H_{pqI}^{\mathrm{E}}\right)_{p,q,I}$

$$H_{pqI}^{\mathrm{E}} \overset{\mathrm{def}}{=} \left.\frac{\partial^2 E}{\partial \kappa_{pq} \partial c_I}\right|_{\boldsymbol{\xi}=\mathbf{0}}. \tag{12}$$

To compute the orbital Gradient and Hessian, we take the Taylor and Campbell expansions of the energy to the second order in $\boldsymbol{\kappa}$ [5]:

$$E(\mathbf{c}, \boldsymbol{\kappa}) \approx E(\mathbf{c}, \mathbf{0}) + \boldsymbol{\kappa}^T \mathbf{G}^o + \frac{1}{2} \boldsymbol{\kappa}^T \mathbf{H}^{oo} \boldsymbol{\kappa}, \tag{13a}$$

$$E(\mathbf{c}, \boldsymbol{\kappa}) \approx \left\langle \psi_{\mathbf{c}} \middle| \hat{H} + \left[\hat{\kappa}, \hat{H}\right] + \left[\hat{\kappa}, \left[\hat{\kappa}, \hat{H}\right]\right] \middle| \psi_{\mathbf{c}} \right\rangle. \tag{13b}$$

Identifying each order leads to the following expressions of the orbital Gradient and Hessian:

$$G_{pq}^{\mathrm{E}} = \left\langle \psi \middle| \left[\hat{E}_{[pq]}, \hat{H}\right] \middle| \psi \right\rangle, \tag{14a}$$

$$H_{pqrs}^{\mathrm{E}} = \mathcal{T}_{(rs)}^{(pq)} \left\langle \psi \middle| \left[\hat{E}_{[pq]}, \left[\hat{E}_{[rs]}, \hat{H}\right]\right] \middle| \psi \right\rangle. \tag{14b}$$

Furthermore, by differentiating eq. (5) with respect to $\mathbf{c}$, we find expressions for the configuration Gradient and Hessians [4]:

$$G_I^{\mathrm{E}} = 2 \left\langle \psi \middle| \hat{H} - E(\boldsymbol{\xi} = \mathbf{0}) \middle| I \right\rangle, \tag{15a}$$

$$H_{IJ}^{\mathrm{E}} = 2 \left\langle I \middle| \hat{H} - E(\boldsymbol{\xi} = \mathbf{0}) \middle| J \right\rangle - S^{\psi,I} G_J^{\mathrm{E}} - S^{\psi,J} G_I^{\mathrm{E}}, \tag{15b}$$

$$H_{pqI}^{\mathrm{E}} = 2 \left\langle \psi \middle| \left[\hat{E}_{[pq]}, \hat{H}\right] \middle| I \right\rangle - 2 S^{\psi,I} G_{pq}^{\mathrm{E}}. \tag{15c}$$

From there, more practical expressions can be obtained as a function of the RDMs and of the molecular integrals. For the sake of brevity, we do not discuss them here, but the interested reader can refer to Appendix B. These expressions can be tidily stored in two matrices as follows:

$$\mathbf{G}^{\mathrm{E}} = \begin{pmatrix} \mathbf{G}^c \\ \mathbf{G}^o \end{pmatrix}, \tag{16a}$$

$$\mathbf{H}^{\mathrm{E}} = \begin{pmatrix} \mathbf{H}^{cc} & \mathbf{H}^{co} \\ \mathbf{H}^{oc} & \mathbf{H}^{oo} \end{pmatrix}. \tag{16b}$$

Finally, to minimize the energy, we differentiate the second-order Taylor expansion with respect to $\boldsymbol{\xi}$ and set the result to zero, which yields the updated value:

$$\boldsymbol{\xi} = -\left(\mathbf{H}^{\mathrm{E}}\right)^{-1} \mathbf{G}^{\mathrm{E}}. \tag{17}$$

The procedure is iterated until it converges towards a value $\boldsymbol{\xi}_{\mathrm{final}}$. Then, $E(\boldsymbol{\xi}_{\mathrm{final}})$ is the energy and $\left|\Psi(\boldsymbol{\xi}_{\mathrm{final}})\right\rangle$ is the CASSCF ground state.

II. THE ORTHOGONALLY CONSTRAINED CASSCF METHOD

A. The orthogonality constraint

Suppose we already applied an MCSCF method to find an optimised ground state $\left|\psi^0\right\rangle$, and we now want to obtain the first excited state $\left|\Psi(\mathbf{c}, \boldsymbol{\kappa})\right\rangle$, parameterised as in eq. (5). Then, we introduce a cost function containing not only the energy, but also a penalty term

$\left| \langle \psi^0 | \mathrm{e}^{-\hat{\kappa}} | \psi_{\mathbf{c}} \rangle \right|^2$, which is the squared modulus of the overlap between the variational ansatz and the ground state [2, 3]. To minimize this term, the variational ansatz must be orthogonal to the ground state, ensuring that the final optimised wavefunction should be the least-energetic state that is not the ground state, *i.e.* the first excited state. The cost function can thus be written

$$\begin{aligned}\mathcal{E}(\mathbf{c}, \boldsymbol{\kappa}) &= \langle \psi_{\mathbf{c}} | \mathrm{e}^{\hat{\kappa}} \hat{H} \mathrm{e}^{-\hat{\kappa}} | \psi_{\mathbf{c}} \rangle + \lambda_0 \left| \langle \psi^0 | \mathrm{e}^{-\hat{\kappa}} | \psi_{\mathbf{c}} \rangle \right|^2 \\ &= E(\mathbf{c}, \boldsymbol{\kappa}) + \lambda_0 \mathcal{E}^0(\mathbf{c}, \boldsymbol{\kappa})\end{aligned} \tag{18}$$

where λ_0 is some (unknown for the time being) Lagrange multiplier. Let us introduce $\hat{\Pi}^0 = |\psi^0\rangle\langle\psi^0|$, the projector on the ground state, and rewrite the cost function as

$$\mathcal{E}(\mathbf{c}, \boldsymbol{\kappa}) = \left\langle \psi_{\mathbf{c}} \middle| \mathrm{e}^{\hat{\kappa}} \left(\hat{H} + \lambda_0 \hat{\Pi}^0 \right) \mathrm{e}^{-\hat{\kappa}} \middle| \psi_{\mathbf{c}} \right\rangle. \tag{19}$$

This reveals the penalty term to be of the same mathematical form as the energy term.

Our goal is to compute an *overlap Gradient* $\mathbf{G}^0$ and an *overlap Hessian* $\mathbf{H}^0$ associated to $\mathcal{E}^0$, such that the optimised coefficients become

$$\boldsymbol{\xi} = -\left(\mathbf{H}^{\mathrm{E}} + \lambda \mathbf{H}^0\right)^{-1} \left(\mathbf{G}^{\mathrm{E}} + \lambda \mathbf{G}^0\right). \tag{20}$$

B. First excited state Gradients and Hessians

Because of the similarity between the two terms in the cost function, we can follow the same steps as in Section I. Identifying the Taylor and Campbell expansions to the second order in $\boldsymbol{\kappa}$ leads to the following formulae for the overlap Gradient and Hessian:

$$G^0_{pq} \overset{\text{def}}{=} \left. \frac{\partial \mathcal{E}^0}{\partial \kappa_{pq}} \right|_{\boldsymbol{\xi}=\mathbf{0}} = \left\langle \psi \middle| \left[\hat{E}_{[pq]}, \hat{\Pi}^0 \right] \middle| \psi \right\rangle, \tag{21a}$$

$$\begin{aligned}H^0_{pqrs} &\overset{\text{def}}{=} \left. \frac{\partial^2 \mathcal{E}^0}{\partial \kappa_{pq} \partial \kappa_{rs}} \right|_{\boldsymbol{\xi}=\mathbf{0}} \\ &= \mathcal{T}^{(pq)}_{(rs)} \left\langle \psi \middle| \left[\hat{E}_{[pq]}, \left[\hat{E}_{[rs]}, \hat{\Pi}^0 \right] \right] \middle| \psi \right\rangle.\end{aligned} \tag{21b}$$

Again, we assume all integrals to be real, and after some calculations (detailed in Appendix C), we can express these quantities as a function of the TRDMs:

$$G^0_{pq} = 2 S^{\psi,\psi^0} \gamma^{\psi,\psi^0}_{[pq]}, \tag{22a}$$

$$\begin{aligned}H^0_{pqrs} = S^{\psi,\psi^0} &\left(2\Gamma^{\psi,\psi^0}_{[pq][rs]} + \delta_{[r|[q} \left(\gamma^{\psi,\psi^0}_{p]|s]} + \gamma^{\psi,\psi^0}_{s]|p]} \right) \right) \\ &+ 2\gamma^{\psi,\psi^0}_{[pq]} \gamma^{\psi,\psi^0}_{[rs]}.\end{aligned} \tag{22b}$$

where we have introduced the overlap integral $S^{\psi^1,\psi^0} = \langle \psi^1 | \psi^0 \rangle = S^{\psi^0,\psi^1}$.

Again, as in Section I, we seek the overlap configuration Gradient and Hessians by starting from eq. (18) and evaluating the following derivatives:

$$G^0_I \overset{\text{def}}{=} \left. \frac{\partial \mathcal{E}^0}{\partial c_I} \right|_{\boldsymbol{\xi}=\mathbf{0}} = 2 \left\langle \psi \middle| \hat{\Pi}^0 - \left| S^{\psi,\psi^0} \right|^2 \middle| I \right\rangle, \tag{23a}$$

$$\begin{aligned}H^0_{IJ} &\overset{\text{def}}{=} \left. \frac{\partial^2 \mathcal{E}^0}{\partial c_I \partial c_J} \right|_{\boldsymbol{\xi}=\mathbf{0}} \\ &= 2 \left\langle I \middle| \hat{\Pi}^0 - \left| S^{\psi,\psi^0} \right|^2 \middle| J \right\rangle - S^{\psi,I} G^0_J - S^{\psi,J} G^0_I,\end{aligned} \tag{23b}$$

$$\begin{aligned}H^0_{pqI} &\overset{\text{def}}{=} \left. \frac{\partial^2 \mathcal{E}^0}{\partial \kappa_{pq} \partial c_I} \right|_{\boldsymbol{\xi}=\mathbf{0}} \\ &= 2 \left\langle \psi \middle| \left[\hat{E}_{[pq]}, \hat{\Pi}^0 \right] \middle| I \right\rangle - 2 S^{\psi,I} G^0_{pq}.\end{aligned} \tag{23c}$$

Let us now compute

$$\left\langle \psi \middle| \left[\hat{E}_{pq}, \hat{\Pi}^0 \right] \middle| I \right\rangle = S^{\psi^0,I} \gamma^{\psi,\psi^0}_{pq} - S^{\psi,\psi^0} \gamma^{\psi^0,I}_{pq} \tag{24}$$

yielding

$$G^0_I = 2 S^{\psi,\psi^0} S^{\psi^0,I} - 2 \left| S^{\psi,\psi^0} \right|^2 S^{\psi,I}, \tag{25a}$$

$$H^0_{IJ} = 2 S^{\psi^0,I} S^{\psi^0,J} - 2 \left| S^{\psi,\psi^0} \right|^2 \delta_{IJ} - S^{\psi,I} G^0_J - S^{\psi,J} G^0_I, \tag{25b}$$

$$H^0_{pqI} = 2 S^{\psi^0,I} \gamma^{\psi,\psi^0}_{[pq]} - 2 S^{\psi,\psi^0} \gamma^{\psi^0,I}_{[pq]} - 2 S^{\psi,I} G^0_{pq}. \tag{25c}$$

C. Complete active space reduced density matrices

The calculation of RDMs and TRDMs can be simplified in the context of a CAS ansatz. If i is an inactive index, v a virtual index, p, q, r, s any indices, and $|\psi\rangle$ a CAS state, then

$$\hat{E}_{iq} |\psi\rangle = 2\delta_{iq} |\psi\rangle, \quad \langle\psi| \hat{E}_{pi} = 2\delta_{pi} \langle\psi|, \tag{26}$$

$$\hat{E}_{pv} |\psi\rangle = 0, \quad \langle\psi| \hat{E}_{vq} = 0. \tag{27}$$

This immediately leads to the well-known expressions:

$$\gamma^{\psi,\psi}_{iq} = \gamma^{\psi,\psi}_{qi} = \langle \psi | \hat{E}_{iq} | \psi \rangle = 2\delta_{iq}, \tag{28}$$

$$\gamma^{\psi,\psi}_{pv} = \gamma^{\psi,\psi}_{vp} = \langle \psi | \hat{E}_{pv} | \psi \rangle = 0. \tag{29}$$

2-electron RDM elements with at least one virtual index are all zero. 2-electron RDM elements without virtual indices but with at least one inactive index can be computed by using:

$$\hat{e}_{pqrs} = \hat{E}_{pq} \hat{E}_{rs} - \delta_{qr} \hat{E}_{ps}, \tag{30}$$

thus:

$$\Gamma^{\psi,\psi}_{pqis} = 2\delta_{is} \gamma^{\psi,\psi}_{pq} - \delta_{iq} \gamma^{\psi,\psi}_{ps}. \tag{31}$$

Equivalent computations can be conducted for TRDM elements, when $\left|\psi^0\right\rangle$ is however not *a priori* a CAS state:

$$\gamma_{pi}^{\psi,\psi^0} = \left\langle\psi\middle|\hat{E}_{pi}\middle|\psi^0\right\rangle = 2\delta_{pi}S^{\psi,\psi^0}, \tag{32}$$

$$\gamma_{vq}^{\psi,\psi^0} = \left\langle\psi\middle|\hat{E}_{vq}\middle|\psi^0\right\rangle = 0, \tag{33}$$

$$\Gamma_{pqri}^{\psi,\psi^0} = \Gamma_{ripq}^{\psi,\psi^0} = 2\delta_{ri}\gamma_{pq}^{\psi,\psi^0} - \delta_{ip}\gamma_{rq}^{\psi,\psi^0}, \tag{34}$$

$$\Gamma_{pqvs}^{\psi,\psi^0} = \Gamma_{vspq}^{\psi,\psi^0} = 0. \tag{35}$$

However, neither $\gamma_{iq}^{\psi,\psi^0}$, $\gamma_{pv}^{\psi,\psi^0}$, $\Gamma_{pqis}^{\psi,\psi^0} = \Gamma_{ispq}^{\psi,\psi^0}$ nor $\Gamma_{pqrv}^{\psi,\psi^0} = \Gamma_{rvpq}^{\psi,\psi^0}$ can be simplified in general.

III. HIGHER EXCITED STATES AND THE LAGRANGE MULTIPLIERS

A. Generalisation of the OC-CASSCF method

Suppose we have optimised K states $\left(\left|\psi^A\right\rangle\right)_{0\leq A\leq K-1}$, and we want to find the K^{th} excited state $\left|\psi^K\right\rangle$. Unlike the indices I, J that run over active configurations, the index A runs over energy eigenstates.

We write the general cost function

$$\mathcal{E}(\boldsymbol{\xi}) = E(\boldsymbol{\xi}) + \sum_{A=0}^{K-1}\lambda_A\left\langle\Psi(\boldsymbol{\xi})|\hat{\Pi}^A|\Psi(\boldsymbol{\xi})\right\rangle, \tag{36}$$

$$\hat{\Pi}^A = \left|\psi^A\middle\rangle\!\middle\langle\psi^A\right|. \tag{37}$$

Then, the total Gradient and Hessian consist in a sum

$$\mathbf{G}^{\text{tot}} = \mathbf{G}^{\text{E}} + \sum_{A=0}^{K-1}\lambda_A\mathbf{G}^A, \tag{38a}$$

$$\mathbf{H}^{\text{tot}} = \mathbf{H}^{\text{E}} + \sum_{A=0}^{K-1}\lambda_A\mathbf{H}^A \tag{38b}$$

and the optimised parameters are

$$\boldsymbol{\xi} = -\left(\mathbf{H}^{\text{tot}}\right)^{-1}\mathbf{G}^{\text{tot}}. \tag{39}$$

Following the same steps, the various components of the Gradients and Hessians can be computed through the formulae:

$$G_{pq}^A = 2S^{\psi,\psi^A}\gamma_{[pq]}^{\psi,\psi^A}, \tag{40a}$$

$$H_{pqrs}^A = S^{\psi,\psi^A}\left(2\Gamma_{[pq][rs]}^{\psi,\psi^A} + \delta_{[r|[q}\left(\gamma_{p]|s]}^{\psi,\psi^A} + \gamma_{s]|p]}^{\psi,\psi^A}\right)\right) + 2\gamma_{[pq]}^{\psi,\psi^A}\gamma_{[rs]}^{\psi,\psi^A}, \tag{40b}$$

$$G_I^A = 2S^{\psi,\psi^A}S^{\psi^A,I} - 2\left|S^{\psi,\psi^A}\right|^2 S^{\psi,I}, \tag{40c}$$

$$H_{IJ}^A = 2S^{\psi^A,I}S^{\psi^A,J} - 2\left|S^{\psi,\psi^A}\right|^2\delta_{IJ} - S^{\psi,I}G_J^A - S^{\psi,J}G_I^A, \tag{40d}$$

$$H_{pqI}^A = 2S^{\psi^A,I}\gamma_{[pq]}^{\psi,\psi^A} - 2S^{\psi,\psi^A}\gamma_{[pq]}^{\psi^A,I} - 2S^{\psi,I}G_{pq}^A. \tag{40e}$$

B. A lower bound on the Lagrange multipliers

It is worth noting that the exact Hamiltonian is of the form

$$\hat{H} = \sum_{A=0}^{\infty}E_A\left|\psi^A\middle\rangle\!\middle\langle\psi^A\right|. \tag{41}$$

Since the states are bound, there is a supremum E_∞ to the set of eigenvalues. The cost function is

$$\mathcal{E}(\boldsymbol{\xi}) = \sum_{A=0}^{\infty}E_A\left\langle\Psi(\boldsymbol{\xi})|\hat{\Pi}^A|\Psi(\boldsymbol{\xi})\right\rangle + \sum_{A=0}^{K-1}\lambda_A\left\langle\Psi(\boldsymbol{\xi})|\hat{\Pi}^A|\Psi(\boldsymbol{\xi})\right\rangle \tag{42}$$

$$= \left\langle\Psi(\boldsymbol{\xi})\middle|\sum_{A=0}^{K-1}(E_A+\lambda_A)\hat{\Pi}^A + \sum_{A=K}^{\infty}E_A\hat{\Pi}^A\middle|\Psi(\boldsymbol{\xi})\right\rangle. \tag{43}$$

Imposing $\lambda_A = E_\infty - E_A$ effectively takes away the energies we already know in the Hamiltonian and brings them to the top of the spectrum, creating a new system in which these known states are almost ionised. Then, the cost function is nothing but the expectation value of the K-th augmented Hamiltonian

$$\hat{H}_{\text{aug}}^K = \sum_{A=0}^{K-1}E_\infty\hat{\Pi}^A + \sum_{A=K}^{\infty}E_A\hat{\Pi}^A \tag{44}$$

whose ground state is $\left|\psi^K\right\rangle$ since $E_K \leq E_A \leq E_\infty$.

Rigorously, writing $|\Psi(\boldsymbol{\xi})\rangle = \sum_A k_A \left|\psi^A\right\rangle$:

$$\mathcal{E}(\boldsymbol{\xi}) = \langle\Psi(\boldsymbol{\xi})|\hat{H}^K_{\text{aug}}|\Psi(\boldsymbol{\xi})\rangle \tag{45a}$$

$$= \sum_{A=0}^{K-1} E_\infty |k_A|^2 + \sum_{A=K}^{\infty} E_A |k_A|^2 \tag{45b}$$

$$\geq E_K \sum_{A=0}^{\infty} |k_A|^2 = E_K \tag{45c}$$

where the equality case happens when $|\Psi(\boldsymbol{\xi})\rangle = \left|\psi^K\right\rangle$.

This adapted version of Ritz' theorem shows that $(\lambda_A)_A = (E_\infty - E_A)_A$ correctly solves the variational problem, and thus is a correct choice of Lagrange multipliers. However, the same reasoning applies with the weaker condition $\lambda_A \geq E_\infty - E_A$, preventing us to conclude on the exact value of the Lagrange multipliers.

C. Convergence and error control

Numerical considerations must however also be taken into account. Along the optimisation process, the penalty term becomes ever so small (since we are trying to cancel it). If the penalty term becomes very small with respect to the energy term, it can prevent the method from converging. Conversely, if the Lagrange multipliers are chosen too large, the method will indeed minimise the overlap, but will (at least at first) ignore the energy. We need to choose the Lagrange multipliers so that the energy and penalty terms are of the same order of magnitude.

We know moreover that the state vector describing the ground state will not be exactly recovered by the CASSCF method, yet we depend on this state vector to create the penalty term. Let us write the numerical ground state $\left|\psi^0_{\text{CASSCF}}\right\rangle$ as a slight deviation from the exact ground state:

$$\left|\psi^0_{\text{CASSCF}}\right\rangle = \left|\psi^0\right\rangle + \varepsilon \left|\psi^{\text{err}}\right\rangle \tag{46}$$

where ε is a small quantity, and $|\psi^{\text{err}}\rangle$ an unknown normalised ket, orthogonal to $\left|\psi^0\right\rangle$. Then, using the same notations as in Section III B:

$$\left|\left\langle\Psi(\boldsymbol{\xi})\middle|\psi^0_{\text{CASSCF}}\right\rangle\right|^2 = |k_0 + \varepsilon k_{\text{err}}|^2 = |k_0|^2 + 2\varepsilon k_0 k_{\text{err}} + |\varepsilon k_{\text{err}}|^2 \tag{47}$$

where $k_{\text{err}} = \langle\Psi(\boldsymbol{\xi})|\psi^{\text{err}}\rangle$. The cost function for the first excited state reads:

$$\mathcal{E}(\boldsymbol{\xi}) = E_0|k_0|^2 + E_1|k_1|^2 + \lambda_0|k_0|^2 + 2\lambda_0\varepsilon k_0 k_{\text{err}} + \lambda_0|\varepsilon k_{\text{err}}|^2. \tag{48}$$

We see that the larger we take λ_0, the larger the error terms become. Optimising with large λ_0 may lead the computation to first minimise k_{err} (which is unphysical). Even if we correctly orthogonalise $|\Psi(\boldsymbol{\xi})\rangle$ with respect to the true ground state, we get:

$$\mathcal{E}(\boldsymbol{\xi}) = E_1|k_1|^2 + \lambda_0|\varepsilon k_{\text{err}}|^2. \tag{49}$$

so λ_0 controls the error on the energy. In fact, our best interest is to take the smallest possible value of λ_0. The same reasoning applies to any excited state multiplier λ_A.

CONCLUSION

We have summarised the main points of the variational methods used in quantum chemistry to solve the electronic structure problem, from the basic Hartree-Fock theory to the more complete CASSCF method that is able to account for electron correlation with numerical efficiency.

We then explained the idea of Yalouz and Robert to adapt this method to the description of excited states by adding an orthogonality constraint to the CASSCF optimisation problem. Building on this idea, we derived analytical expressions for the Gradients and Hessians needed for the numerical implementation of the minimisation algorithm.

Finally, we investigated the value of the Lagrange multipliers that appear in the theory, and proved a rigorous lower bound from an adapted version of Ritz' theorem. The necessity to control error propagation from the ground to the excited states led us to conclude that this lower bound was also the best choice for the multipliers' value, as long as typical values of the orthogonality constraint term remain in the same orders of magnitude as the energies.

Appendix A: CAS variational ansatz

The complete expression of the CAS variational ansatz reads [4]:

$$|\Psi(\mathbf{c}, \boldsymbol{\kappa})\rangle = \mathrm{e}^{-\hat{\kappa}} \left|\psi_{\mathbf{c}}\right\rangle = \mathrm{e}^{-\hat{\kappa}} \frac{|\psi\rangle + \hat{\pi}\left|\mathbf{c}\right\rangle}{\sqrt{1 + \langle\mathbf{c}|\hat{\pi}|\mathbf{c}\rangle}} \tag{A1}$$

with

$$\hat{\pi} \overset{\text{def}}{=} \nVdash - |\psi\rangle\langle\psi|\,, \tag{A2}$$

$$|\mathbf{c}\rangle \overset{\text{def}}{=} \sum_I c_I \left|I\right\rangle, \tag{A3}$$

$$\hat{\kappa} \overset{\text{def}}{=} \sum_{p>q} \kappa_{pq} \hat{E}_{[pq]}, \tag{A4}$$

$$\hat{E}_{pq} \overset{\text{def}}{=} \sum_\sigma \hat{a}^\dagger_{p\sigma}\hat{a}_{q\sigma} = \hat{E}^\dagger_{qp}, \tag{A5}$$

$$\hat{e}_{pqrs} = \hat{e}_{rspq} \overset{\text{def}}{=} \sum_{\sigma,\tau} \hat{a}^\dagger_{p\sigma}\hat{a}^\dagger_{r\tau}\hat{a}_{s\tau}\hat{a}_{q\sigma} = \hat{e}^\dagger_{qpsr} = \hat{e}^\dagger_{srqp}. \tag{A6}$$

$|\psi\rangle$ is the reference state, the initial guess from which we try to elaborate. $\mathrm{e}^{-\hat{\kappa}}$ is simply the unitary operator that rotates the orbitals during the orbital optimisation, and thus changes the meaning of the configurations. $\hat{\pi}\left|\mathbf{c}\right\rangle$ modifies the weight of the various configurations during the configuration optimisation. The denominator is just a normalisation factor.

Appendix B: Electronic orbital Gradient and Hessian

After some calculations, we have the following expressions for the orbital Gradient and Hessian [4]:

$$G^{\mathrm{E}}_{pq} = 2F_{[pq]}, \tag{B1}$$

$$H^{\mathrm{E}}_{pqrs} = 2h_{[q|[s}\gamma^{\psi,\psi}_{p]|r]} - \delta_{[s|[q}F_{p]|r]} + \delta_{[q|[s}F_{r]|p]} + 2Y_{[pq][rs]} \tag{B2}$$

with

$$F_{mn} = \sum_q h_{nq}\gamma^{\psi,\psi}_{mq} + \sum_{qrs} g_{nqrs}\Gamma^{\psi,\psi}_{mqrs}, \tag{B3}$$

$$Y_{pqrs} = \sum_{mn}\left(g_{qmns}\Gamma^{\psi,\psi}_{pm(rn)} + g_{qsmn}\Gamma^{\psi,\psi}_{prmn}\right) \tag{B4}$$

where $(h_{pq})_{pq}$ and $(g_{pqrs})_{pqrs}$ are the molecular one- and two-electron integrals.

We can also compute, for the mixed orbital-configuration Hessian,

$$\left\langle\psi\middle|\left[\hat{E}_{pq},\hat{H}\right]\middle|I\right\rangle = F^{\psi,I}_{pq} - F^{I,\psi}_{qp} \tag{B5}$$

where we have introduced the transition generalised Fock matrix

$$F^{\psi^1,\psi^2}_{mn} = \sum_q h_{nq}\gamma^{\psi^1,\psi^2}_{mq} + \sum_{qrs} g_{nqrs}\Gamma^{\psi^1,\psi^2}_{mqrs}. \tag{B6}$$

Appendix C: Overlap orbital Gradient and Hessian

Let us first compute

$$\left\langle\psi\middle|\left[\hat{E}_{pq},\hat{\Pi}^0\right]\middle|\psi\right\rangle = S^{\psi,\psi^0}\gamma^{\psi,\psi^0}_{[pq]} \tag{C1}$$

where we have introduced the overlap integral $S^{\psi^1,\psi^0} = \left\langle\psi^1\middle|\psi^0\right\rangle = S^{\psi^0,\psi^1}$. Noticing that the overlap Gradient of eq. (21a) is the antisymmetrisation of this (already antisymmetric) expression yields

$$G^0_{pq} = 2S^{\psi,\psi^0}\gamma^{\psi,\psi^0}_{[pq]}. \tag{C2}$$

For the Hessian, we first compute the bracket

$$\begin{aligned}\left\langle\psi\middle|\left[\hat{E}_{pq},\left[\hat{E}_{rs},\hat{\Pi}^0\right]\right]\middle|\psi\right\rangle &= \left\langle\psi\middle|\hat{E}_{pq}\hat{E}_{rs}\hat{\Pi}^0\middle|\psi\right\rangle - \left\langle\psi\middle|\hat{E}_{pq}\hat{\Pi}^0\hat{E}_{rs}\middle|\psi\right\rangle - \left\langle\psi\middle|\hat{E}_{rs}\hat{\Pi}^0\hat{E}_{pq}\middle|\psi\right\rangle + \left\langle\psi\middle|\hat{\Pi}^0\hat{E}_{rs}\hat{E}_{pq}\middle|\psi\right\rangle \\ &= S^{\psi,\psi^0}\left(\daleth^{\psi,\psi^0}_{pqrs} + \daleth^{\psi,\psi^0}_{qpsr}\right) - \gamma^{\psi,\psi^0}_{pq}\gamma^{\psi,\psi^0}_{sr} - \gamma^{\psi,\psi^0}_{rs}\gamma^{\psi,\psi^0}_{qp}\end{aligned} \tag{C3}$$

where

$$\begin{aligned}\daleth^{\psi,\psi^0}_{pqrs} &\overset{\text{def}}{=} \left\langle\psi\middle|\hat{E}_{pq}\hat{E}_{rs}\middle|\psi^0\right\rangle = \Gamma^{\psi,\psi^0}_{pqrs} + \delta_{rq}\gamma^{\psi,\psi^0}_{ps} &\text{(C4)}\\ &= \left\langle\psi^0\middle|\hat{E}_{sr}\hat{E}_{qp}\middle|\psi\right\rangle = \daleth^{\psi^0,\psi}_{srqp} &\text{(C5)}\end{aligned}$$

is the “improper” transition RDM. Again, we antisymmetrise this expression:

$$\left\langle\psi\middle|\left[\hat{E}_{[pq]},\left[\hat{E}_{[rs]},\hat{\Pi}^0\right]\right]\middle|\psi\right\rangle = 2S^{\psi,\psi^0}\daleth^{\psi,\psi^0}_{[pq][rs]} + 2\gamma^{\psi,\psi^0}_{[pq]}\gamma^{\psi,\psi^0}_{[rs]} \tag{C6}$$

yielding for the overlap Hessian:

$$\begin{aligned}H^0_{pqrs} &= \mathcal{T}^{(pq)}_{(rs)}\left\langle\psi_{\mathbf{c}}\middle|\left[\hat{E}_{[pq]},\left[\hat{E}_{[rs]},\hat{\Pi}^0\right]\right]\middle|\psi_{\mathbf{c}}\right\rangle \\ &= S^{\psi,\psi^0}\left(\daleth^{\psi,\psi^0}_{[pq][rs]} + \daleth^{\psi,\psi^0}_{[rs][pq]}\right) + 2\gamma^{\psi,\psi^0}_{[pq]}\gamma^{\psi,\psi^0}_{[rs]}.\end{aligned} \tag{C7}$$

ACKNOWLEDGMENTS

The author would like to thank CNRS researcher Saad Yalouz, not only for correcting this article, but also for his precious guiding and support throughout the last years; and Professor Vincent Robert for his help in entering the very technical domain of quantum chemistry.

This work is supported by the Interdisciplinary Thematic Institute QMat, as part of the ITI 2021-2028 program of the University of Strasbourg, CNRS and Inserm, and was supported by IdEx Unistra (ANR-10-IDEX-0002), SFRI STRAT'US project (ANR-20-SFRI-0012), and EUR QMAT ANR-17-EURE-0024 under the framework of the French Investments for the Future Program.

LICENSING

[1] A. Szabó and N. S. Ostlund, *Modern quantum chemistry : introduction to advanced electronic structure theory* (Dover Publications, 1982).

[2] S. Yalouz and V. Robert, Orthogonally constrained orbital optimization: Assessing changes of optimal orbitals for orthogonal multireference states, J. Chem. Theory Comput. **19**, 1381 (2023), doi:10.1021/acs.jctc.2c01144.

[3] L. Delafosse, V. Robert, and S. Yalouz, Orthogonally constrained casscf framework: Newton-raphson orbital optimization and nuclear gradients (2026), arXiv:2601.19699.

[4] T. Helgaker, P. Jørgensen, and J. Olsen, *Molecular electronic-structure theory* (John Wiley & Sons, Ltd, 2000).

[5] S. Yalouz, E. Koridon, B. Senjean, B. Lasorne, F. Buda, and L. Visscher, Analytical nonadiabatic couplings and gradients within the state-averaged orbital-optimized variational quantum eigensolver, J. Chem. Theory Comput. **18** (2022), doi:10.1021/acs.jctc.1c00995.