



# Numerical methods for computing ROHF ground states. Part I: SCF algorithms

Robert Benda, Laurent Vidal, Emmanuel Giner, Eric Cancès

## ► To cite this version:

Robert Benda, Laurent Vidal, Emmanuel Giner, Eric Cancès. Numerical methods for computing ROHF ground states. Part I: SCF algorithms. 2021. hal-03228618v2

**HAL Id: hal-03228618**

**<https://hal.science/hal-03228618v2>**

Preprint submitted on 5 Oct 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Numerical methods for computing ROHF ground states.

## Part I: SCF algorithms

Robert Benda<sup>1,2</sup>, Laurent Vidal<sup>1</sup>, Emmanuel Giner<sup>3</sup>, and Eric Cancès<sup>1,\*</sup>

<sup>1</sup>CERMICS, Ecole des Ponts and Inria Paris, 6 & 8 avenue Blaise Pascal,  
77455 Marne-la-Vallée, France

<sup>2</sup>LPICM, CNRS, Ecole Polytechnique, Institut Polytechnique de Paris,  
Route de Saclay, 91128 Palaiseau, France

<sup>3</sup>Laboratoire de Chimie Théorique (UMR 7616), Sorbonne Université, CNRS,  
Paris, France

\*Corresponding author : eric.cances@enpc.fr

October 5, 2021

### Abstract

In this article, we propose a simple geometrical derivation of the restricted open-shell Hartree-Fock (ROHF) equations in the density matrix formalism. We then introduce a new, parameter-free, basic fixed-point method to solve these equations, that, in contrast with existing self-consistent field (SCF) schemes, is not based on the introduction of a non-physical, parameter-dependent, composite Hamiltonian. We also introduce a variant of Pulay’s DIIS algorithm specific to ROHF, and extend the Optimal Damping Algorithm to the ROHF framework. We finally present numerical results on challenging systems (complexes with transition metals) demonstrating the performance of the new algorithms we propose.

## 1 Introduction

The ultimate goal of computational chemistry is to propose reliable theoretical tools to describe the chemical properties of any molecular system. The initial step of such a task is always the accurate description of the ground state electronic structure of the system, for which there exist essentially two flavours of approaches: the wave function theory (WFT) and density functional theory (DFT). Although DFT remains certainly the most used theoretical tool for closed-shell systems because of its advantageous ratio between the computational cost and the accuracy of the results, the usual semi-local approximations used in DFT are known to suffer from several issues when open-shell systems need to be considered. For instance, the self-interaction error in open-shell systems is responsible for the over delocalization of electrons in transition metal complexes and has impacts on several chemical properties such as the electronic paramagnetic spectrum, ligand-field excitations or spin-gaps [1, 2, 3, 4, 5]. One major issue in DFT is that there is no systematic way to improve the results, which leads to an inflation of different flavours of approximated functionals tailored for a specific class of systems and/or properties [6]. The situation of WFT is somehow opposite as there exists many ways of systematically refine the results starting from a mean-field description although it comes to the price of a rapidly growing

computational cost. Nevertheless, as remarkable progresses have been obtained in the reduction of the computational cost of correlated WFT methods for open-shell systems (see for instance Ref [7] and references therein), the latter appear more and more as actual computational tools for the treatment of open-shell systems. Even though WFT-based correlated methods are in active development, they all start with a mean-field Hartree Fock (HF) calculation for which there are many convergence problems in the context of open-shell systems. Therefore, improving the reliability of the HF algorithms becomes an important point in order to popularise the correlated WFT methods.

There exists several avatars of the Hartree-Fock method. The most commonly used are the restricted and unrestricted Hartree-Fock methods (RHF and UHF, respectively), which differ by the constraint imposed in the RHF method to have an unique set of spatial orbitals for both up and down spins. For open-shell systems, the constraint of having the same spatial orbitals for the two spins has an important consequence : while the ROHF Slater determinant is an eigenfunction of the  $\hat{S}^2$  operator, the UHF Slater determinant suffers from spin contamination [8]. The latter has a big impact in the post-HF calculations as the correlated wave function built upon a spin-contaminated Slater determinant needs to restore the correct spin symmetry using high-order particle-hole excitations [9, 8, 10]. Moreover, the correlated methods using unrestricted orbitals necessary deal with several types of two-electron integrals types corresponding to the interaction between electrons of different spins, which also induces several complications in the code structure and memory.

From the mathematical point of view, Hartree-Fock methods give rise to constrained optimization problems, whose first-order optimality conditions are the Hartree-Fock equations. As usual in optimization theory, numerical solutions can be obtained either by solving the Hartree-Fock equations by a fixed-point (self-consistent field - SCF) algorithm, or by a direct minimization of the Hartree-Fock energy functional [10, 11, 12].

Many algorithms have been developed for the RHF and UHF frameworks in the past 70 years. Roothaan’s [13], level-shifting [14], and DIIS algorithms [15, 16, 17, 18, 19] belong to the class of SCF algorithms. The direct minimization approach is adopted in *e.g.* Bacsikay’s quadratic convergent algorithm [20], trust-region methods [21], and geometric direct minimization (GDM) methods [10, 11]. Let us also mention the second-order SCF (SOSCF) algorithm [22, 23], and the DIIS-GDM [10, 11], which combine features from both SCF and direct minimization methods. The optimal damping algorithm (ODA) [24] and the EDIIS algorithm [25] solve a relaxed version of the Hartree-Fock optimization problem, whose solutions always coincide with those of the original Hartree-Fock problem for UHF, as well as for the less popular General Hartree-Fock method (GHF) in which each spin-orbital is allowed to have both a spin-up and a spin-down component. For RHF, ODA and EDIIS most often converge to solutions to the RHF problem, but may occasionally converge to one-body density matrices with fractional occupation numbers, which do not correspond to Hartree-Fock states. A robust and efficient method to solve the RHF and UHF problems (which always works for UHF and most of the time for RHF) is to use EDIIS in the first iterations and switch to DIIS to accelerate convergence when the iterates are close enough to the solution [25]. All the above algorithms are relatively well-understood from a mathematical point of view [26]. Roughly speaking, computing RHF and UHF ground states for small and medium-size chemical systems is no longer an issue.

The situation is radically different for ROHF, where existing SCF algorithms fail to converge in many cases, notably for radicals and molecular systems containing transition metals. In [10], the authors extend their GDM method to ROHF, and, based on their implementation in Q-Chem [27], show that it is more robust than the SCF algorithms available at that time. They remark that combining DIIS (for the early iterations) with

GDM (for the last iterations) is an efficient strategy [11, 10]. Unfortunately, the GDM for ROHF is not presented in full detail in [10]. Perhaps for this reason, it is not implemented in most of the popular quantum chemistry packages, so that the latter are often unable to provide ROHF ground states for radicals or open-shell systems with transition metals. The Q-Chem implementation of the GDM for ROHF – the only one we are aware of – performs well in many cases, but does not converge in some cases (see future work). These observations lead us to revisit the problem of computing ROHF ground states.

In this article, we investigate SCF algorithms; a future study will be devoted to direct minimization algorithms. In these contributions, we focus on maximum spin states in order to simplify the presentation, but our approach is valid for any spin state (see Remark 1). In Section 2, we recall the mathematical structure of the ROHF ground state problem in both the density matrix and molecular orbital formalism, and we provide a geometrical derivation of the first-order optimality conditions, the ROHF equations. In contrast with the RHF and UHF settings, the ROHF equations cannot be *naturally* formulated as a nonlinear eigenvalue problem. As a consequence, the simple SCF Roothaan scheme for RHF, “assemble the Fock matrix for the current iterate, diagonalize it, build the next iterate using the *Aufbau* principle, that is by selecting the lowest energy orbitals”, cannot be straightforwardly extended to the ROHF setting. All the existing SCF algorithms we are aware of twist the ROHF equations using coupling operators to transform them into a nonlinear eigenvalue problem. They are based on the construction of a composite, non-physical, effective Hamiltonian obtained by linear combinations of sub-blocks of the Fock matrices  $F_d$  and  $F_s$  respectively associated to the doubly and singly ROHF orbitals. These combinations involve six real coefficients  $A_{tt}$ , and  $B_{tt}$  with  $t$  equal to  $d$  (doubly occupied),  $s$  (singly occupied), or  $v$  (virtual), the choice of which characterizes the SCF scheme. For instance, these six coefficients are all equal to 1/2 in the Guest and Saunders algorithm [28], but are different and depend on the spin state in the Canonical-I and Canonical-II algorithms introduced by Plakhutin and Davidson [29]. From the physical point of view, the choice of  $A_{tt}$  and  $B_{tt}$  coefficients essentially tries to maintain the *Aufbau* principle in order to avoid numerical instabilities of the SCF algorithm induced by swapping of the singly occupied orbital with doubly occupied or virtual orbitals. It is important to stress that, because of the mathematical restriction imposed by the ROHF Slater determinant, the *Aufbau* principle, inspired by the Koopman theorem, is not guaranteed, and therefore a choice of  $A_{tt}$  and  $B_{tt}$  which might work for a given system might break down for another, as illustrated for instance in the numerical results reported here (see Sec. 5.2). In Section 3, we present a new SCF scheme, which better respects the essence of the ROHF equations and is parameter-free. We then study DIIS acceleration. We also discuss the adaptation of the recently introduced adaptive-depth DIIS scheme [19] to the ROHF setting. In Section 4, we extend the ODA to the ROHF setting. In Section 5, we compare the performance of the new algorithms introduced in this article to the state-of-the-art SCF algorithms for some challenging chemical systems, such as organic ligands chelating – or simply interacting with – transition metals.

## 2 The ROHF optimization problem

We present in this section the ROHF model using the density-matrix formalism, before introducing the manifold of ROHF states and deriving, from geometrical arguments, the first-order optimality conditions for the ROHF energy, namely the ROHF equations.

## 2.1 The ROHF model

In ROHF theory, trial wavefunctions  $\Psi$  are not, in general, single Slater determinants, but configuration state functions (CSFs) [29, 30]. The latter are eigenfunctions of the spin operators  $\hat{S}^2$  and  $\hat{S}_z$  and of the number operators  $\hat{n}_i = a_{i\uparrow}^\dagger a_{i\uparrow} + a_{i\downarrow}^\dagger a_{i\downarrow}$ , for a given orthonormal basis of orbitals  $(\varphi_1, \varphi_2, \dots)$  of  $L^2(\mathbb{R}^3; \mathbb{C})$ :

$$\hat{S}^2 \Psi = s(s+1)\Psi, \quad \hat{S}_z \Psi = m_s \Psi, \quad \hat{n}_i \Psi = n_i \Psi,$$

for given  $s \in \frac{1}{2}\mathbb{N}$ ,  $m_s \in \{-s, -s+1, \dots, s-1, s\}$ , and  $n_i \in \{0, 1, 2\}$ . Up to reordering the orbitals, we can assume that  $n_i = 2$  for  $i = 1, \dots, N_d$ ,  $n_i = 1$  for  $i = N_d + 1, \dots, N_d + N_s$ , and  $n_i = 0$  for  $i > N_d + N_s$ . Then,  $\Psi$  is a finite sum of Slater determinants, each of them made of the  $N_d$  doubly occupied orbitals  $\varphi_1, \dots, \varphi_{N_d}$  and  $N_s$  spin-orbitals of the form  $\varphi_{N_d+1} \otimes \eta_1, \dots, \varphi_{N_d+N_s} \otimes \eta_{N_s}$ , the function  $\eta_j$  being equal to either  $\alpha$  (spin-up) or  $\beta$  (spin-down). The numbers  $N_d$ ,  $N_s$ ,  $N$  (number of electrons in the system),  $s$ , and  $m_s$  are such that

$$2N_d + N_s = N, \quad |m_s| \leq s \leq \frac{1}{2}N_s.$$

We also denote by  $N_o := N_d + N_s$  the number of (singly or doubly) occupied orbitals.

For maximum spin states ( $s = \frac{1}{2}N_s$ ) and maximum  $m_s$  value ( $m_s = s$ ), ROHF trial wavefunctions are single Slater determinants built with  $N_d$  doubly occupied orbitals  $\varphi_1, \dots, \varphi_{N_d}$  and  $N_s$  spin-up-orbitals  $\varphi_{N_d+1} \otimes \alpha, \dots, \varphi_{N_o} \otimes \alpha$ , where the  $\varphi_i$ 's satisfy  $\langle \varphi_i | \varphi_j \rangle = \delta_{ij}$  for all  $1 \leq i, j \leq N_o$ . The electronic Hamiltonian

$$H_N = -\frac{1}{2} \sum_{i=1}^N \Delta_{\mathbf{r}_i} + \sum_{i=1}^N V_{\text{nuc}}(\mathbf{r}_i) + \sum_{1 \leq i < j \leq N} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}$$

being real-valued in the absence of external magnetic field and spin-orbit coupling, we can assume without loss of generality that the orbitals  $\varphi_i$  are real-valued. In order to obtain a computationally tractable model, the  $\varphi_i$ 's are expanded in a finite basis set  $\mathcal{X} := (\chi_1, \dots, \chi_{N_b})$  of real-valued functions of the space variable:

$$\varphi_i(\mathbf{r}) = \sum_{\mu=1}^{N_b} C_{\mu i} \chi_\mu(\mathbf{r}).$$

In practice, the  $\chi_\mu$ 's are non-orthogonal atomic orbitals (AO). In order to simplify the presentation, we will however assume here that the basis  $\mathcal{X}$  is orthonormal, or equivalently that the overlap matrix is the identity matrix:

$$S_{\mu\nu} := \int_{\mathbb{R}^3} \chi_\mu(\mathbf{r}) \chi_\nu(\mathbf{r}) d\mathbf{r} = \delta_{\mu\nu}.$$

Let us emphasize that we make this simplification for pedagogical purposes only; extending our arguments to non-orthogonal basis sets is a simple exercise. We denote by

$$\Phi_d = [C_{\mu i}]_{1 \leq \mu \leq N_b, 1 \leq i \leq N_d} \in \mathbb{R}^{N_b \times N_d} \quad \text{and} \quad \Phi_s = [C_{\mu i}]_{1 \leq \mu \leq N_b, N_d+1 \leq i \leq N_o} \in \mathbb{R}^{N_b \times N_s}$$

the rectangular matrices collecting the coefficients of the doubly and singly occupied molecular orbitals in the discretization basis  $\mathcal{X}$ , and define the density matrices  $P_d$  and  $P_s$  as

$$P_d := \Phi_d \Phi_d^T \quad \text{and} \quad P_s := \Phi_s \Phi_s^T. \quad (1)$$

The matrices  $P_d$  and  $P_s$  are the basis representations of the orthogonal projectors on the spaces spanned by the doubly and singly occupied orbitals respectively. Recall that a square matrix  $P$  is an orthogonal projector if  $P^2 = P = P^T$ , and that its rank is the integer  $\text{Tr}(P)$ . These matrices represent the one-body density matrices (projectors) :

$$\gamma_d = \sum_{i=1}^{N_d} |\varphi_i\rangle\langle\varphi_i| \quad \text{and} \quad \gamma_s = \sum_{i=N_d+1}^{N_o} |\varphi_i\rangle\langle\varphi_i| \quad (2)$$

in the basis set  $\mathcal{X}$ :

$$\gamma_d = \sum_{\mu,\nu=1}^{N_b} [P_d]_{\mu\nu} |\chi_\mu\rangle\langle\chi_\nu| \quad \text{and} \quad \gamma_s = \sum_{\mu,\nu=1}^{N_b} [P_s]_{\mu\nu} |\chi_\mu\rangle\langle\chi_\nu|.$$

Denoting by  $I_n$  the identity matrix of rank  $n$ , we have the following equivalences:

$$\langle\varphi_i|\varphi_j\rangle = \delta_{ij} \text{ for all } 1 \leq i, j \leq N_o \quad \Leftrightarrow \quad \Phi_d^T \Phi_d = I_{N_d}, \quad \Phi_s^T \Phi_s = I_{N_s}, \quad \Phi_d^T \Phi_s = 0 \quad (3)$$

$$\Leftrightarrow \quad \begin{cases} P_d^2 = P_d = P_d^T, \quad \text{Tr}(P_d) = N_d, \\ P_s^2 = P_s = P_s^T, \quad \text{Tr}(P_s) = N_s, \\ P_d P_s = 0 \end{cases} \quad (4)$$

The maximum spin ROHF wavefunction  $\Psi$  generated by orthonormal doubly orbitals  $(\varphi_1, \dots, \varphi_{N_b})$  and singly occupied orbitals  $(\varphi_{N_d+1}, \dots, \varphi_{N_o})$  is completely determined (up to an irrelevant global phase) by the one-body density matrices  $\gamma_d$  and  $\gamma_s$  defined by (2). Conversely any pair  $(\gamma_d, \gamma_s)$  of orthogonal projectors satisfying  $\text{Tr}(\gamma_d) = N_d$ ,  $\text{Tr}(\gamma_s) = N_s$ , and  $\gamma_d \gamma_s = 0$  gives rise to a unique ROHF wavefunction  $\Psi_{\gamma_d, \gamma_s}^{\text{ROHF}}$  of maximal spin (up to a global phase), whose energy is a function of  $(\gamma_d, \gamma_s)$ :

$$\mathcal{E}^{\text{ROHF}}(\gamma_d, \gamma_s) := \langle \Psi_{\gamma_d, \gamma_s}^{\text{ROHF}} | H_N | \Psi_{\gamma_d, \gamma_s}^{\text{ROHF}} \rangle.$$

After discretization in the finite basis set  $\mathcal{X}$ , the ROHF energy functional becomes a function of the matrices  $P_d$  and  $P_s$  representing  $\gamma_d$  and  $\gamma_s$  in this basis:

$$E(P_d, P_s) := \mathcal{E}^{\text{ROHF}} \left( \sum_{\mu,\nu=1}^{N_b} [P_d]_{\mu\nu} |\chi_\mu\rangle\langle\chi_\nu|, \sum_{\mu,\nu=1}^{N_b} [P_s]_{\mu\nu} |\chi_\mu\rangle\langle\chi_\nu| \right).$$

Standard algebraic manipulations lead to

$$\begin{aligned} E(P_d, P_s) &= \text{Tr}(h(2P_d + P_s)) + \text{Tr}((2J(P_d) - K(P_d))(P_d + P_s)) \\ &\quad + \frac{1}{2} \text{Tr}((J(P_s) - K(P_s))P_s), \end{aligned} \quad (5)$$

where

$$[h]_{\mu\nu} = \frac{1}{2} \int_{\mathbb{R}^3} \nabla \chi_\mu(\mathbf{r}) \cdot \nabla \chi_\nu(\mathbf{r}) d\mathbf{r} + \int_{\mathbb{R}^3} V_{\text{nuc}}(\mathbf{r}) \chi_\mu(\mathbf{r}) \chi_\nu(\mathbf{r}) d\mathbf{r},$$

$$[J(P)]_{\mu\nu} = \sum_{\kappa,\lambda=1}^{N_b} (\mu\nu|\kappa\lambda) P_{\kappa\lambda}, \quad [K(P)]_{\mu\nu} = \sum_{\kappa,\lambda=1}^{N_b} (\mu\kappa|\nu\lambda) P_{\kappa\lambda},$$

and

$$(\mu\nu|\kappa\lambda) := \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \frac{\chi_\mu(\mathbf{r}) \chi_\nu(\mathbf{r}) \chi_\kappa(\mathbf{r}') \chi_\lambda(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}'.$$

Note that the trace of  $P_d$  is equal to  $N_d$ , the number of doubly-occupied orbitals. The fact that each of these orbitals hosts two electrons is taken into account by the factors 2 in the

first two terms of the right-hand side of Eq. (5). In view of (4), the density matrix (DM) formulation of the ROHF ground state problem in the basis  $\mathcal{X}$  thus reads

$$\mathcal{E}_*^{\text{ROHF}} := \min\{E(P_d, P_s), (P_d, P_s) \in \mathcal{M}_{\text{DM}}\}, \quad (6)$$

where

$$\begin{aligned} \mathcal{M}_{\text{DM}} := \{ & (P_d, P_s) \in \mathbb{R}_{\text{sym}}^{N_b \times N_b} \times \mathbb{R}_{\text{sym}}^{N_b \times N_b} \mid P_d^2 = P_d, P_s^2 = P_s, P_d P_s = 0, \\ & \text{Tr}(P_d) = N_d, \text{Tr}(P_s) = N_s\}. \end{aligned} \quad (7)$$

The only properties of the matrix  $h$  and the functions  $J$  and  $K$  that will be used in the sequel are the following:  $h \in \mathbb{R}_{\text{sym}}^{N_b \times N_b}$ , and  $J, K : \mathbb{R}_{\text{sym}}^{N_b \times N_b} \rightarrow \mathbb{R}_{\text{sym}}^{N_b \times N_b}$  are linear and such that

$$\text{Tr}(J(P)P') = \text{Tr}(J(P')P), \text{Tr}(K(P)P') = \text{Tr}(K(P')P) \text{ for all } P, P' \in \mathbb{R}_{\text{sym}}^{N_b \times N_b}. \quad (8)$$

The set  $\mathcal{M}_{\text{DM}}$  is the set of admissible pairs of doubly and singly occupied density matrices, that are the pairs of matrices actually representing a maximum spin ROHF state in the basis  $\mathcal{X}$ . From a mathematical point of view,  $\mathcal{M}_{\text{DM}}$  is a smooth (*i.e.* infinitely differentiable,  $C^\infty$ ), compact submanifold of the matrix space  $\mathbb{R}_{\text{sym}}^{N_b \times N_b} \times \mathbb{R}_{\text{sym}}^{N_b \times N_b}$ , and problem (6) is a smooth optimization problem on  $\mathcal{M}_{\text{DM}}$ . In order to simplify the notation, we will often denote by  $x$  the points in  $\mathcal{M}_{\text{DM}}$  and rewrite (6) as

$$\mathcal{E}_*^{\text{ROHF}} := \min\{E(x), x \in \mathcal{M}_{\text{DM}}\}. \quad (9)$$

A molecular orbital (MO) formulation of the ROHF ground state problem can then be deduced from (1), (3), and (6):

$$\mathcal{E}_*^{\text{ROHF}} = \min\{\mathcal{E}(\Phi_d, \Phi_s), (\Phi_d, \Phi_s) \in \mathcal{M}_{\text{MO}}\}, \quad (10)$$

where,

$$\mathcal{E}(\Phi_d, \Phi_s) := E(\Phi_d \Phi_d^T, \Phi_s \Phi_s^T),$$

and

$$\mathcal{M}_{\text{MO}} := \{(\Phi_d, \Phi_s) \in \mathbb{R}^{N_b \times N_d} \times \mathbb{R}^{N_b \times N_s} \mid \Phi_d^T \Phi_d = I_{N_d}, \Phi_s^T \Phi_s = I_{N_s}, \Phi_d^T \Phi_s = 0\}.$$

**Remark 1.** The optimization problem (6) with  $E$  and  $\mathcal{M}_{\text{DM}}$  given by (5)-(7) corresponds to the ROHF model for maximum spin states ( $|m_s| = s = \frac{1}{2}N_s$ ). For other spin states ( $|m_s| \leq s < \frac{1}{2}N_s$ ), the ROHF problem still is of the form (6) with  $\mathcal{M}_{\text{DM}}$  given by (7). The energy functional  $E$  has a different expression (due to the Fock exchange term coupling only spin-orbitals having the same spin), but remains a sum of linear and bilinear forms in  $(P_d, P_s)$ . See e.g. Ref. [30] for the derivation of the non-maximal spin energy expressions using the genealogical coupling scheme. Note that the algorithms presented in this article, although formulated for maximum spin state case, can therefore be straightforwardly extended to any spin state.

## 2.2 The manifold of ROHF states

In the DM formalism, the manifold of ROHF states is seen as the submanifold  $\mathcal{M}_{\text{DM}}$  (defined in Eq. (7)) of  $\mathcal{V}_{\text{sym}}$ , the vector space of pairs of real symmetric matrices of size  $N_b$

$$\mathcal{V}_{\text{sym}} := \mathbb{R}_{\text{sym}}^{N_b \times N_b} \times \mathbb{R}_{\text{sym}}^{N_b \times N_b}.$$



As already mentioned in the introduction, there is a one-to-one correspondence between ROHF states and points of  $\mathcal{M}_{\text{DM}}$ , and the ROHF energy functional has a simple form in the DM representation. This makes the DM formalism well-suited for methodological developments. The purpose of this section is to give some insights on the manifold  $\mathcal{M}_{\text{DM}}$  (which, from a mathematical viewpoint, is diffeomorphic to the manifold of ROHF states in the discretization basis  $\mathcal{X}$ ). We denote by  $O(n) = \{U \in \mathbb{R}^{n \times n} \mid U^T U = I_n\}$  the orthogonal group in dimension  $n$ , and by

$$N_v := N_b - N_o$$

the number of virtual orbitals. Let  $x = (P_d, P_s) \in \mathcal{M}_{\text{DM}}$ . Since  $P_d$  is a rank- $N_d$  orthogonal projector (*i.e.* a symmetric matrix fulfilling  $P_d^2 = P_d$  and  $\text{Tr}(P_d) = N_d$ ), it can be diagonalized in an orthonormal basis on  $\mathbb{R}^{N_b}$  and its only eigenvalues are 1 (multiplicity  $N_d$ ) and 0 (multiplicity  $N_s + N_v$ ). Likewise,  $P_s$  is a rank- $N_s$  orthogonal projector. In addition, as  $P_d P_s = 0$ , we also have  $P_s P_d = (P_d P_s)^T = 0$ , which implies that  $P_d$  and  $P_s$  commute and can therefore be co-diagonalized in the same orthonormal basis. Introducing the projector

$$P_v := I_{N_b} - P_d - P_s$$

on the virtual space (the space spanned by the virtual orbitals), which satisfies  $P_v^2 = P_v = P_v^T$ ,  $\text{Tr}(P_v) = N_v$ , and  $P_d P_v = P_s P_v = 0$ , we obtain that there exists a unitary matrix  $U \in O(N_b)$  such that

$$P_d = U \mathcal{I}_d U^T, \quad P_s = U \mathcal{I}_s U^T, \quad P_v = U \mathcal{I}_v U^T, \quad U U^T = I_{N_b}, \quad (11)$$

with

$$\mathcal{I}_d = \begin{pmatrix} I_{N_d} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathcal{I}_s = \begin{pmatrix} 0 & 0 & 0 \\ 0 & I_{N_s} & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathcal{I}_v = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & I_{N_v} \end{pmatrix}. \quad (12)$$

Eqs (11) and (12) are equivalent to finding an orthonormal basis of eigenvectors (which form the unitary matrix  $U$ ) of the projectors and selecting the ones corresponding to the eigenvalue 1. Decomposing  $U$  as three rectangular matrices  $U = (\Phi_d | \Phi_s | \Phi_v)$  with  $\Phi_d \in \mathbb{R}^{N_b \times N_d}$ ,  $\Phi_s \in \mathbb{R}^{N_b \times N_s}$ , and  $\Phi_v \in \mathbb{R}^{N_b \times N_v}$ , we can write

$$\Phi_d = U \mathcal{J}_d, \quad \Phi_s = U \mathcal{J}_s, \quad \Phi_v = U \mathcal{J}_v, \quad U U^T = I_{N_b},$$

with

$$\mathcal{J}_d = \begin{pmatrix} I_{N_d} \\ 0 \\ 0 \end{pmatrix}, \quad \mathcal{J}_s = \begin{pmatrix} 0 \\ I_{N_s} \\ 0 \end{pmatrix}, \quad \mathcal{J}_v = \begin{pmatrix} 0 \\ 0 \\ I_{N_v} \end{pmatrix}.$$

and we have  $P_d = \Phi_d \Phi_d^T$ ,  $P_s = \Phi_s \Phi_s^T$ ,  $P_v = \Phi_v \Phi_v^T$ . In other words, the set  $\Phi_d$  (respectively  $\Phi_s$ ) is the set of  $N_d$  (respectively  $N_s$ ) natural orbitals associated to the density matrix  $P_d$  (respectively  $P_s$ ). The orbitals in  $\Phi_v$  are then the orthogonal complement to  $\Phi_d$  and  $\Phi_s$ .

In order to derive the first-order optimality conditions associated to the minimization problem (9) (a.k.a. the ROHF equations) from a simple geometrical argument, we have to identify the tangent space  $T_x \mathcal{M}_{\text{DM}}$  to a point  $x = (P_d, P_s)$  of the manifold  $\mathcal{M}_{\text{DM}}$ , that is the vector space of velocities  $q = (Q_d, Q_s) = \dot{p}(0)$  at  $t = 0$  for all paths  $p : [-1, 1] \ni t \rightarrow p(t) \in \mathcal{M}_{\text{DM}}$  drawn on  $\mathcal{M}_{\text{DM}}$  and such that  $p(0) = x$  (see Fig. 1).

Let  $p$  be such a path. We have for all  $t \in [-1, 1]$ ,

$$p(t) \in \mathcal{M}_{\text{DM}} \quad \text{and} \quad p(t) = x + tq + O(t^2) = (P_d + tQ_d + o(t), P_s + tQ_s + o(t)). \quad (13)$$



where the  $O(\cdot)$  and  $o(\cdot)$  notations are relative to the usual Euclidean topology. In other words, the conditions (13) are equivalent to defining the tangent space  $T_x \mathcal{M}_{\text{DM}}$  to  $x = (P_d, P_s)$  as the vector space of pairs of symmetric real matrices  $q = (Q_d, Q_s)$  which allow to locally approximate the manifold of density matrices  $\mathcal{M}_{\text{DM}}$  by an affine space, as pictorially represented in Fig. 1. The constraints defining the manifold  $\mathcal{M}_{\text{DM}}$  (see Eq. (7)) are equivalent to the following at first order:

$$p_d(t)^2 = p_d(t), \quad \text{Tr}(p_d(t)) = N_d \quad \Leftrightarrow \quad P_d Q_d + Q_d P_d = Q_d, \quad \text{Tr}(Q_d) = 0, \quad (14)$$

$$p_s(t)^2 = p_s(t), \quad \text{Tr}(p_s(t)) = N_s \quad \Leftrightarrow \quad P_s Q_s + Q_s P_s = Q_s, \quad \text{Tr}(Q_s) = 0, \quad (15)$$

$$p_d(t)p_s(t) = 0 \quad \Leftrightarrow \quad P_d Q_s + Q_d P_s = 0. \quad (16)$$

In the representation (11)-(12), the constraints (14)-(16) are equivalent to

$$Q_d = U \begin{pmatrix} 0 & X & Y \\ X^T & 0 & 0 \\ Y^T & 0 & 0 \end{pmatrix} U^T \quad \text{and} \quad Q_s = U \begin{pmatrix} 0 & -X & 0 \\ -X^T & 0 & Z \\ 0 & Z^T & 0 \end{pmatrix} U^T, \quad (17)$$

where  $X \in \mathbb{R}^{N_d \times N_s}$ ,  $Y \in \mathbb{R}^{N_d \times N_v}$ ,  $Z \in \mathbb{R}^{N_s \times N_v}$  are generic matrices. It follows that for all  $x = (P_d, P_s) \in \mathcal{M}_{\text{DM}}$ ,

$$\begin{aligned} T_x \mathcal{M} &= \{(Q_d, Q_s) \in \mathcal{V}_{\text{sym}} \text{ of the form (17)}\} \\ &= \{(Q_d, Q_s) \in \mathcal{V}_{\text{sym}} \mid P_d Q_d P_d = P_s Q_d P_s = P_v Q_d P_v = P_s Q_s P_s = 0, \\ &\quad P_d Q_s P_d = P_s Q_s P_s = P_v Q_s P_v = P_d Q_s P_v = 0, P_d(Q_d + Q_s)P_s = 0\}. \end{aligned}$$

## 2.3 First-order optimality conditions

### 2.3.1 General considerations on optimization in the DM framework

Finding a point  $x_* = (P_d^*, P_s^*)$  in  $\mathcal{M}_{\text{DM}}$  which minimizes the energy functional defined in (5) requires the definition of the derivative of  $E(P_d, P_s)$  with respect to the pair of density matrices  $x = (P_d, P_s)$ . The ROHF energy functional  $E(P_d, P_s)$  is not only defined for density matrices, but for any pair of real-valued symmetric matrix  $y = (W_d, W_s)$ , which might not be admissible density matrices. Therefore, although the energy gradient  $\nabla E(y)$  with respect to  $y = (W_d, W_s)$  can be easily computed once a topology, allowing to define the later, has been chosen, imposing  $\nabla E(y) = 0$  is not enough to find the optimal ROHF density matrix because of the constraints imposed by the properties of density matrices (see Eq. (7)). The reason for this is that the gradient  $\nabla E(x)$  has a component outside the manifold  $\mathcal{M}_{\text{DM}}$  of density matrices, and following that component of the gradient will necessary lead outside the manifold  $\mathcal{M}_{\text{DM}}$  of density matrices. Therefore, the correct ROHF condition is to find the point  $x_* \in \mathcal{M}_{\text{DM}}$  such that the projection of  $\nabla E(x_*)$  onto the tangent space  $T_{x_*} \mathcal{M}$  is zero. One can also reformulate that the optimal condition is that  $\nabla E(x_*)$  belongs to  $T_{x_*} \mathcal{M}_{\text{DM}}^\perp$ , the space orthogonal to the tangent space  $T_{x_*} \mathcal{M}$ . As we need to define a gradient and an orthogonal projection, we need to define an inner product. As we shall see in a future work, there is no unique choice of inner product and this leads to different minimization algorithms.

### 2.3.2 ROHF-Brillouin condition in the MO and DM framework

Endowing  $\mathcal{V}_{\text{sym}}$  with the Frobenius-like inner product

$$\langle (M_1, N_1), (M_2, N_2) \rangle_{\text{DM}} := \frac{1}{2} (\text{Tr}(M_1 M_2) + \text{Tr}(N_1 N_2)), \quad (18)$$

the manifold  $\mathcal{M}_{\text{DM}}$  can be seen as a smooth submanifold of the Euclidean space  $\mathcal{V}_{\text{sym}}$ . Thanks to this inner product, the critical points of  $E$  on  $\mathcal{M}_{\text{DM}}$  can be characterized in a simple geometric way (see Fig. 1):

$$x_* \text{ critical point of } E \text{ on } \mathcal{M}_{\text{DM}} \quad \Leftrightarrow \quad \nabla E(x_*) \in T_{x_*} \mathcal{M}_{\text{DM}}^\perp, \quad (19)$$

where  $\nabla E(x_*)$  is the gradient of  $E$  for the inner product  $\langle \cdot, \cdot \rangle_{\text{DM}}$ , and  $T_{x_*} \mathcal{M}_{\text{DM}}^\perp$  the orthogonal subspace to  $T_{x_*} \mathcal{M}_{\text{DM}}$ , still for the inner product  $\langle \cdot, \cdot \rangle_{\text{DM}}$ . The condition of Eq. (19) is equivalent to state that, taken at the optimal density matrix  $x_*$ , the component of  $\nabla E(x_*)$  on the tangent plane  $T_{x_*} \mathcal{M}$  is zero. Recall that for any  $x \in \mathcal{V}_{\text{sym}}$ ,  $\nabla E(x)$  is the vector of  $\mathcal{V}_{\text{sym}}$  characterized by

$$E(x + \delta x) = E(x) + \langle \nabla E(x), \delta x \rangle_{\text{DM}} + o(\delta x),$$

whih implies that the gradient depends on the choice of inner product. Also, for any  $x \in \mathcal{M}_{\text{DM}}$ , the vector space  $T_x \mathcal{M}_{\text{DM}}^\perp$  is defined by

$$T_x \mathcal{M}_{\text{DM}}^\perp = \{q' \in \mathcal{V}_{\text{sym}} \mid \forall q \in T_x \mathcal{M}, \langle q, q' \rangle_{\text{DM}} = 0\}.$$

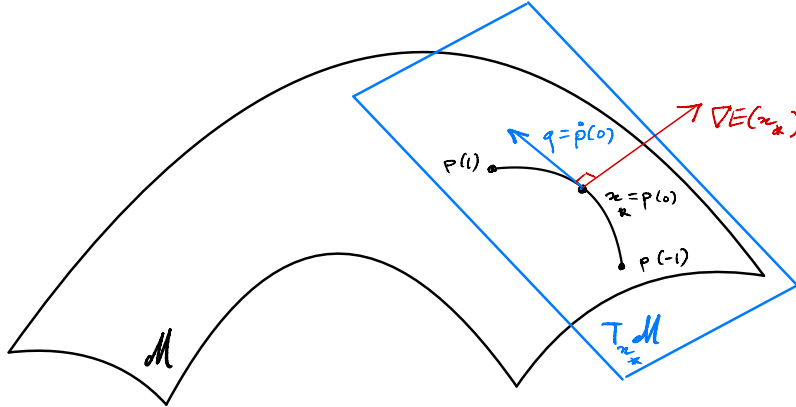


Figure 1: Let  $x_*$  be a critical point of  $E$  on the submanifold  $\mathcal{M}$  of the Euclidean space  $\mathcal{V}$  endowed with the inner product  $\langle \cdot, \cdot \rangle$ . Then for all path  $p : [-1, 1] \rightarrow \mathcal{V}$  drawn on  $\mathcal{M}$  such that  $p(0) = x_*$  and  $\dot{p}(0) = q$ , we have  $E(p(t)) = E(x_*) + o(t)$  (zero first-order variation at critical points). On the other hand,  $E(p(t)) = E(x_* + tq + o(t)) = E(x_*) + t\langle \nabla E(x_*), q \rangle + o(t)$ . Therefore,  $\langle \nabla E(x_*), q \rangle = 0$  for all path  $q \in T_{x_*} \mathcal{M}$ , which means that  $\nabla E(x_*) \in T_{x_*} \mathcal{M}^\perp$ .

Let us now detail the computation of  $\nabla E(x)$  for any ROHF state  $x = (P_d, P_s) \in \mathcal{V}_{\text{sym}}$ . Introducing the Fock operators

$$F_d(P_d, P_s) := h + 2J(P_d) + J(P_s) - K(P_d) - \frac{1}{2}K(P_s), \quad (20)$$

$$F_s(P_d, P_s) := \frac{1}{2} (h + 2J(P_d) + J(P_s) - K(P_d) - K(P_s)), \quad (21)$$

we have for all  $M_d, M_s \in \mathbb{R}_{\text{sym}}^{N_b \times N_b}$ ,

$$\begin{aligned}
E(P_d + M_d, P_s + M_s) &= \text{Tr}(h(2P_d + 2M_d + P_s + M_s)) \\
&\quad + \text{Tr}((2J(P_d + M_d) - K(P_d + M_d))(P_d + M_d + P_s + M_s)) \\
&\quad + \frac{1}{2} \text{Tr}((J(P_s + M_s) - K(P_s + M_s))(P_s + M_s)) \\
&= E(P_d, P_s) + \text{Tr}(2F_d(P_d, P_s)M_d) + \text{Tr}(2F_s(P_d, P_s)M_s) \\
&\quad + \text{Tr}((2J(M_d) - K(M_d))(M_d + M_s)) + \frac{1}{2} \text{Tr}((J(M_s) - K(M_s))M_s) \\
&= E(P_d, P_s) + \langle (4F_d(P_d, P_s), 4F_s(P_d, P_s)), (M_d, M_s) \rangle_{\text{DM}} \\
&\quad + \text{Tr}((2J(M_d) - K(M_d))(M_d + M_s)) + \frac{1}{2} \text{Tr}((J(M_s) - K(M_s))M_s).
\end{aligned}$$

The gradient of  $E$  at  $x = (P_d, P_s)$  for the inner product  $\langle \cdot, \cdot \rangle_{\text{DM}}$  is therefore

$$\nabla E(x) = (4F_d(P_d, P_s), 4F_s(P_d, P_s)) \text{ with } F_d(P_d, P_s) \text{ and } F_s(P_d, P_s) \text{ given by (20)-(21).} \quad (22)$$

*Characterization of  $T_x \mathcal{M}_{\text{DM}}^\perp$ .* Let  $q' = (M_d, M_s) \in \mathcal{V}_{\text{sym}}$ . Using the decomposition

$$M_d = U \begin{pmatrix} M_d^{dd} & M_d^{ds} & M_d^{dv} \\ M_d^{sd} & M_d^{ss} & M_d^{sv} \\ M_d^{vd} & M_d^{vs} & M_d^{vv} \end{pmatrix} U^T \quad \text{and} \quad M_s = U \begin{pmatrix} M_s^{dd} & M_s^{ds} & M_s^{dv} \\ M_s^{sd} & M_s^{ss} & M_s^{sv} \\ M_s^{vd} & M_s^{vs} & M_s^{vv} \end{pmatrix} U^T, \quad (23)$$

we obtain, using the fact that  $M_d$  and  $M_s$  are symmetric matrices, that for all  $q = (Q_d, Q_s) \in T_x \mathcal{M}_{\text{DM}}$  of the form (17),

$$\begin{aligned}
\langle q, q' \rangle_{\text{DM}} &= \frac{1}{2} \text{Tr} \left( U \begin{pmatrix} 0 & X & Y \\ X^T & 0 & 0 \\ Y^T & 0 & 0 \end{pmatrix} U^T U \begin{pmatrix} M_d^{dd} & M_d^{ds} & M_d^{dv} \\ M_d^{sd} & M_d^{ss} & M_d^{sv} \\ M_d^{vd} & M_d^{vs} & M_d^{vv} \end{pmatrix} U^T \right) \\
&\quad + \frac{1}{2} \text{Tr} \left( U \begin{pmatrix} 0 & -X & 0 \\ -X^T & 0 & Z \\ 0 & Z^T & 0 \end{pmatrix} U^T U \begin{pmatrix} M_s^{dd} & M_s^{ds} & M_s^{dv} \\ M_s^{sd} & M_s^{ss} & M_s^{sv} \\ M_s^{vd} & M_s^{vs} & M_s^{vv} \end{pmatrix} U^T \right) \\
\langle q, q' \rangle_{\text{DM}} &= \text{Tr} \left( X^T (M_d^{ds} - M_s^{ds}) \right) + \text{Tr} \left( Y^T M_d^{dv} \right) + \text{Tr} \left( Z^T M_s^{sv} \right). \quad (24)
\end{aligned}$$

Therefore,  $q'$  belongs to the orthogonal subspace  $T_x \mathcal{M}_{\text{DM}}^\perp$  if  $\langle q, q' \rangle_{\text{DM}} = 0$  for all  $q \in T_x \mathcal{M}_{\text{DM}}$ , hence for all  $X, Y, Z$ , or equivalently, according to Eq. (24)

$$q' \in T_x \mathcal{M}_{\text{DM}}^\perp \Leftrightarrow \left( M_d^{ds} - M_s^{ds} = 0, M_d^{dv} = 0, M_s^{sv} = 0 \right). \quad (25)$$

The critical points  $x_* = (P_{d*}, P_{s*})$  of  $E$  on  $\mathcal{M}_{\text{DM}}$  are then characterized by the first-order optimality condition of Eq. (19), which according to Eqs. (22) and (25), leads to

$$(F_{d*} - F_{s*})^{ds} = 0, F_{d*}^{dv} = 0, F_{s*}^{sv} = 0, \quad \text{with } F_{d*} := F_d(P_{d*}, P_{s*}) \text{ and } F_{s*} := F_s(P_{d*}, P_{s*}).$$

We recover the well-known ROHF optimality conditions (see e.g. [29]), which can also be written as

$$\begin{cases} P_{d*}(F_{d*} - F_{s*})P_{s*} = 0, & P_{d*}F_{d*}P_{v*} = 0, & P_{s*}F_{s*}P_{v*} = 0, \\ \text{with } F_{d*} := F_d(P_{d*}, P_{s*}) \text{ and } F_{s*} := F_s(P_{d*}, P_{s*}). \end{cases} \quad (26)$$

We can similarly derive the optimality conditions in the MO representation, by endowing

$$\mathcal{V}_{\text{rect}} := \mathbb{R}^{N_b \times N_d} \times \mathbb{R}^{N_b \times N_s},$$

with the Frobenius inner product

$$\langle (\Xi_d, \Xi_s), (\Xi'_d, \Xi'_s) \rangle_{\text{MO}} := \text{Tr}(\Xi_d^T \Xi'_d) + \text{Tr}(\Xi_s^T \Xi'_s). \quad (27)$$

This inner product is natural since it reproduces the  $L^2$ -inner product. This calculation is detailed in a future work. We just give here the final results:

$$\forall y = (\Phi_d, \Phi_s) \in \mathcal{V}_{\text{rect}}, \quad \nabla \mathcal{E}(y) = (4F_d(\Phi_d \Phi_d^T, \Phi_s \Phi_s^T) \Phi_d, 4F_s(\Phi_d \Phi_d^T, \Phi_s \Phi_s^T) \Phi_s), \quad (28)$$

and  $y_* = (\Phi_{d*}, \Phi_{s*}) \in \mathcal{M}_{\text{MO}}$  is a critical point of  $\mathcal{E}$  on  $\mathcal{M}_{\text{MO}}$  if and only if

$$\begin{cases} F_{d*} \Phi_{d*} = \Phi_{d*} (\Phi_{d*}^T F_{d*} \Phi_{d*}) + \frac{1}{2} \Phi_{s*} (\Phi_{s*}^T (F_{d*} + F_{s*}) \Phi_{d*}), \\ F_{s*} \Phi_{s*} = \Phi_{s*} (\Phi_{s*}^T F_{s*} \Phi_{s*}) + \frac{1}{2} \Phi_{d*} (\Phi_{d*}^T (F_{d*} + F_{s*}) \Phi_{s*}), \\ \text{with } F_{d*} := F_d(\Phi_{d*} \Phi_{d*}^T, \Phi_{s*} \Phi_{s*}^T) \text{ and } F_{s*} := F_s(\Phi_{d*} \Phi_{d*}^T, \Phi_{s*} \Phi_{s*}^T). \end{cases} \quad (29)$$

It can be checked that  $(\Phi_{d*}, \Phi_{s*}) \in \mathcal{M}_{\text{MO}}$  is solution to (29) if and only if  $(P_{d*}, P_{s*}) \in \mathcal{M}_{\text{DM}}$  is solution to (26), where  $P_{d*} := \Phi_{d*} \Phi_{d*}^T$ ,  $P_{s*} := \Phi_{s*} \Phi_{s*}^T$ . An important implication of Eqs. (29) is that, unlike in the RHF and UHF frameworks, the optimal ROHF orbitals in  $\Phi_{d*}$  and  $\Phi_{s*}$  are not eigenfunctions of the Fock operators  $F_{d*}$  and  $F_{s*}$ , because of the second terms of the right hand side of the first two equations in (29). As a consequence, SCF algorithms based on Fock-like operators involve *ad-hoc* effective Hamiltonians for which the *Aufbau* principal is not always satisfied (see for instance Ref. [29]).

Moreover, the representation of the virtual orbitals  $\Phi_v$  does not appear in equations 29.  $\Phi_v$  is thus not needed to formulate the new algorithms, as opposed to the old Aufbau-based methods, that explicitly construct a full-dimensional  $(N_b \times N_b)$  effective Hamiltonian in order to diagonalize it (see section 3). This gain in complexity (from algorithms base on  $N_b \times N_b$ , to  $N_b \times N_v$  only, matrix representations) enabled by this new formalism will be thoroughly detailed in a future work.

### 3 Self-consistent field (SCF) algorithms

In this section, we first present the various basic SCF iterations proposed in the literature, and introduce a new one, which better respects the mathematical structure of the ROHF equations. We then discuss the stabilization and acceleration of basic SCF iterations using Anderson-Pulay (DIIS-type) algorithms.

#### 3.1 Basic SCF iterations

The basic SCF algorithm for RHF was introduced by Roothaan [31]. It consists in assembling the Fock matrix for the current iterate (molecular orbitals or density matrix), diagonalize it (we still assume orthonormality of the basis set for simplicity), and select the lowest energy eigenvectors to form the next iterate (*Aufbau* principle). This idea can be straightforwardly extended to the UHF model, but not to the ROHF model since the ROHF equations (29) cannot be formulated as a nonlinear eigenvalue problem.

Let  $(P_d^{(k)}, P_s^{(k)}) \in \mathcal{M}_{\text{DM}}$  be the current iterate:

$$P_d^{(k)} = U^{(k)} \mathcal{I}_d U^{(k)T}, \quad P_s^{(k)} = U^{(k)} \mathcal{I}_s U^{(k)T}, \quad U^{(k)} U^{(k)T} = I_{N_b},$$

and  $F_d^{(k)} := F_d(P_d^{(k)}, P_s^{(k)})$  and  $F_s^{(k)} := F_s(P_d^{(k)}, P_s^{(k)})$  the associated Fock matrices:

$$F_d^{(k)} = U^{(k)} \begin{pmatrix} F_d^{(k)dd} & F_d^{(k)ds} & F_d^{(k)dv} \\ F_d^{(k)sd} & F_d^{(k)ss} & F_d^{(k)sv} \\ F_d^{(k)vd} & F_d^{(k)vs} & F_d^{(k)vv} \end{pmatrix} U^{(k)T},$$

$$F_s^{(k)} = U^{(k)} \begin{pmatrix} F_s^{(k)dd} & F_s^{(k)ds} & F_s^{(k)dv} \\ F_s^{(k)sd} & F_s^{(k)ss} & F_s^{(k)sv} \\ F_s^{(k)vd} & F_s^{(k)vs} & F_s^{(k)vv} \end{pmatrix} U^{(k)T}.$$

### 3.1.1 Standard approaches

The most popular simple SCF algorithms for ROHF consist in assembling and diagonalizing a composite effective Hamiltonian of the form

$$H_{A,B}^{(k)} := U^{(k)} \begin{pmatrix} R_{dd}^{(k)} & (F_d^{(k)} - F_s^{(k)})_{ds} & F_d^{(k)dv} \\ (F_d^{(k)} - F_s^{(k)})_{sd} & R_{ss}^{(k)} & F_s^{(k)sv} \\ F_d^{(k)vd} & F_s^{(k)vs} & R_{vv}^{(k)} \end{pmatrix} U^{(k)T}, \quad (30)$$

where  $R_{dd}^{(k)}$ ,  $R_{ss}^{(k)}$ , and  $R_{vv}^{(k)}$  are symmetric matrices. The matrices  $R_{tt}^{(k)}$  are of the form

$$R_{tt}^{(k)} = 2A_{tt} \left( F_s^{(k)} \right)^{tt} + 2B_{tt} \left( F_d^{(k)} - F_s^{(k)} \right)^{tt}, \quad t \in \{d, s, v\},$$

where  $A = (A_{dd}, A_{ss}, A_{vv}) \in \mathbb{R}^3$  and  $B = (B_{dd}, B_{ss}, B_{vv}) \in \mathbb{R}^3$  are coefficients characterizing the SCF algorithm (see Table I in [29]). For instance, they are all equal to 1/2 in Guest and Saunders algorithm [28], but are different and depend on the spin state in the Canonical-I and Canonical-II algorithms introduced by Plakhutin and Davidson [29]. The next iterate  $(P_d^{(k+1)}, P_s^{(k+1)})$  is obtained by filling up first the doubly occupied orbitals, then the singly occupied orbitals, using the *Aufbau* principle. The meta-algorithm for the basic SCF iterations can therefore be written as follows:

$$\left\{ \begin{array}{l} \text{compute the Fock matrices } F_d^{(k)} := F_d(P_d^{(k)}, P_s^{(k)}), F_s^{(k)} := F_s(P_d^{(k)}, P_s^{(k)}), \\ \text{assemble the effective Hamiltonian } H_{A,B}^{(k)}: \\ \quad P_v^{(k)} = I_{N_b} - P_d^{(k)} - P_s^{(k)}, \\ \quad R_{tt}^{(k)} := P_t^{(k)} (A_{tt} F_s^{(k)} + B_{tt} (F_d^{(k)} - F_s^{(k)})) P_t^{(k)}, \quad t \in \{d, s, v\}, \\ \quad H_{\text{diag}}^{(k)} := R_{dd}^{(k)} + R_{ss}^{(k)} + R_{vv}^{(k)}, \\ \quad H_u^{(k)} := P_d^{(k)} (F_d^{(k)} - F_s^{(k)}) P_s^{(k)} + P_d^{(k)} F_d^{(k)} P_v^{(k)} + P_s^{(k)} F_s^{(k)} P_v^{(k)}, \\ \quad H_{A,B}^{(k)} := H_{\text{diag}}^{(k)} + H_u^{(k)} + H_u^{(k)T} \\ \text{diagonalize } H_{A,B}^{(k)} \text{ in an orthonormal basis:} \\ \quad H_{A,B}^{(k)} \phi_i^{(k+1)} = \varepsilon_i^{(k+1)} \phi_i^{(k+1)}, \quad \phi_i^{(k+1)T} \phi_j^{(k+1)} = \delta_{ij}, \quad \varepsilon_1^{k+1} \leq \dots \leq \varepsilon_{N_b}^{k+1}, \\ \text{construct the new iterate using the } \textit{Aufbau} \text{ principle:} \\ \quad g_{A,B}(P_d^{(k)}, P_s^{(k)}) := \left( \sum_{i=1}^{N_d} \phi_i^{(k+1)} \phi_i^{(k+1)T}, \sum_{i=N_d+1}^{N_o} \phi_i^{(k+1)} \phi_i^{(k+1)T} \right) \\ \quad (P_d^{(k+1)}, P_s^{(k+1)}) := g_{A,B}(P_d^{(k)}, P_s^{(k)}) \end{array} \right. \quad (31)$$

where  $R_{tt}^{(k)}$  has to be understood as a diagonal block matrix of size  $N_b \times N_b$  with the matrix  $R_{tt}^{(k)}$  as only non zero block (in  $tt$  position). The existing SCF algorithms for ROHF all

use the  $g_{A,B}$  function defined above, for various choices of  $A_{tt}$  and  $B_{tt}$  coefficients. The basic fixed-point iterations  $x^{k+1} = g_{A,B}(x^k)$  being extremely unstable (see section 5), they are stabilized by DIIS schemes.

The iterates are uniquely defined provided

$$\varepsilon_{N_d}^{(k+1)} < \varepsilon_{N_d+1}^{(k+1)} \quad \text{and} \quad \varepsilon_{N_o}^{(k+1)} < \varepsilon_{N_o+1}^{(k+1)} \quad (32)$$

(energy gaps between doubly and single-occupied orbitals on the one-hand, occupied and virtual orbitals on the other hand). If the conditions (32) are not satisfied, iterates are defined by choosing randomly the orbitals among those satisfying the *Aufbau* principle, or by selecting the ones minimizing the ROHF energy functional.

A necessary and sufficient condition for  $x_* = (P_{d*}, P_{s*})$  being a fixed point of these basic SCF iterations, is

$$\left\{ \begin{array}{l} F_{d*} = F_d(P_{d*}, P_{s*}), \quad F_{s*} = F_s(P_{d*}, P_{s*}), \\ P_{v*} = I_{N_b} - P_{d*} - P_{s*}, \\ R_{tt} = P_{t*}(A_{tt}F_{s*} + B_{tt}(F_{d*} - F_{s*}))P_{t*}, \quad t \in \{d, s, v\}, \\ H_{\text{diag}} = R_{dd} + R_{ss} + R_{vv}, \\ H_u = 2P_{d*}(F_{d*} - F_{s*})P_{s*} + P_{d*}F_{d*}P_{v*} + 2P_{s*}F_{s*}P_{v*}, \\ H_{\text{eff}*} = H_{\text{diag}} + H_u + H_u^T \\ H_{\text{eff}*}\phi_{i*} = \varepsilon_{i*}\phi_{i*}, \quad \phi_{i*}^T\phi_{j*} = \delta_{ij}, \quad \varepsilon_{1*} \leq \dots \leq \varepsilon_{N_b*}, \\ P_{d*} = \sum_{i=1}^{N_d} \phi_{i*}\phi_{i*}^T, \quad P_{s*} = \sum_{i=N_d+1}^{N_o} \phi_{i*}\phi_{i*}^T. \end{array} \right. \quad (33)$$

Let  $x_* = (P_{d*}, P_{s*}) \in \mathcal{M}_{\text{DM}}$  be such a fixed point. Since  $P_{d*}P_{s*} = P_{d*}P_{v*} = P_{s*}P_{v*} = 0$ , we have on the one hand

$$P_{d*}H_{\text{eff}*}P_{s*} = \sum_{i=N_d+1}^{N_o} P_{d*}H_{\text{eff}*}\phi_{i*}\phi_{i*}^T = \sum_{i=N_d+1}^{N_o} \varepsilon_{i*} \underbrace{P_{d*}\phi_{i*}}_{=0} \phi_{i*}^T = 0,$$

and on the other hand

$$P_{d*}H_{\text{eff}*}P_{s*} = \underbrace{P_{d*}H_{\text{diag}}P_{s*}}_{=0} + P_{d*}H_uP_{s*} + \underbrace{P_{d*}H_u^TP_{s*}}_{=0} = P_{d*}(F_{d*} - F_{s*})P_{s*}.$$

Therefore,  $P_{d*}(F_{d*} - F_{s*})P_{s*} = 0$ . A similar argument leads to  $P_{d*}F_{d*}P_{v*} = 0$  and  $P_{s*}F_{s*}P_{v*} = 0$ . Hence,  $x_*$  satisfies (26). Conversely, if  $x_*$  satisfies (26), then  $H_u = 0$ , and  $H_{\text{eff}*} = H_{\text{diag}}$  has a block-diagonal structure in the orthogonal decomposition  $\text{Ran}(P_{d*}) \oplus \text{Ran}(P_{s*}) \oplus \text{Ran}(P_{v*})$  of  $\mathbb{R}^{N_b}$ . Therefore, we have

$$\left\{ \begin{array}{l} H_{\text{eff}*}\phi_{i*} = \varepsilon_{i*}\phi_{i*}, \quad \phi_{i*}^T\phi_{j*} = \delta_{ij}, \\ P_{d*} = \sum_{i=1}^{N_d} \phi_{i*}\phi_{i*}^T, \quad P_{s*} = \sum_{i=N_d+1}^{N_o} \phi_{i*}\phi_{i*}^T, \end{array} \right.$$

for some orthonormal basis  $(\phi_{i*})_{1 \leq i \leq N_b}$  of  $\mathbb{R}^{N_b}$  diagonalizing  $H_{\text{eff}}$ . It follows that a point  $(P_{d*}, P_{s*}) \in \mathcal{M}_{\text{DM}}$  is a critical point of  $E$  on  $\mathcal{M}_{\text{DM}}$  if and only if  $(P_{d*}, P_{s*})$  satisfies all the conditions in (33) *except possibly the fact that the doubly-occupied orbitals do not necessarily correspond to the lowest  $N_d$  eigenvalues of  $H_{\text{eff}*}$ , or the singly-occupied orbitals to the next  $N_s$  ones*, which is equivalent to saying that the *Aufbau* principle does not need to be

satisfied *a priori*. As discussed in [29], there are indeed local minima of the ROHF problem for which the *Aufbau* principle is not satisfied for any of the usual choices of  $A$  and  $B$ . We are therefore facing a dilemma. Either the *Aufbau* principle can be kept in the definition of the SCF procedure, leading to a simple iterative scheme, which is however unable to find the ROHF ground state in some cases. Or the *Aufbau* principle can be discarded and replaced by a more complicated construction procedure, to be specified.

### 3.1.2 A new strategy not based on the *Aufbau* principle

A way out of this dilemma is to attack the problem from a different perspective, using another interpretation of the Roothaan scheme: in the RHF setting,  $P^{(k+1)}$  is the point  $P$  of the RHF manifold

$$\mathcal{M}_{\text{RHF}} := \{P \in \mathbb{R}_{\text{sym}}^{N_b \times N_b} \mid P^2 = P, \text{Tr}(P) = N_d\}$$

in the direction along which the slope of the function  $t \mapsto E^{\text{RHF}}(P^{(k)} + t(P - P^{(k)}))$  is minimum [26], *i.e.*

$$P^{(k+1)} \in \underset{P \in \mathcal{M}_{\text{RHF}}}{\text{argmin}} (\nabla E^{\text{RHF}}(P^{(k)}), P)_{\text{F}} = \underset{P \in \mathcal{M}_{\text{RHF}}}{\text{argmin}} \text{Tr}(F^{\text{RHF}}(P^{(k)})P) \quad (34)$$

where  $F^{\text{RHF}}(P) = \frac{1}{2} \nabla E^{\text{RHF}}(P)$  is the Fock matrix associated with the density matrix  $P$ . Here in 34,  $\text{argmin}$  refers to the set of minimizers of the linear form  $P \rightarrow (\nabla E^{\text{RHF}}(P^{(k)}), P)_{\text{F}}$  on  $\mathcal{M}_{\text{RHF}}$ , to which  $P^{(k+1)}$  belongs. This set is always non empty, but may contain several elements. Transposing this characterization to the ROHF setting, we can define a new basic SCF scheme as:  $x^{(k+1)}$  is the point  $x$  of  $\mathcal{M}_{\text{DM}}$  in the direction along which the slope of the function  $t \mapsto E(x^{(k)} + t(x - x^{(k)}))$  is minimum. It is therefore obtained from  $x^{(k)} = (P_d^{(k)}, P_s^{(k)})$  as

$$x^{(k+1)} \in \underset{x \in \mathcal{M}_{\text{DM}}}{\text{argmin}} \langle \nabla E(x^{(k)}), x \rangle_{\text{DM}} = \underset{x = (P_d, P_s) \in \mathcal{M}_{\text{DM}}}{\text{argmin}} \text{Tr}(F_d^{(k)} P_d) + \text{Tr}(F_s^{(k)} P_s), \quad (35)$$

where  $F_d^{(k)} := F_d(P_d^{(k)}, P_s^{(k)})$  and  $F_s^{(k)} := F_s(P_d^{(k)}, P_s^{(k)})$ . This motivates the introduction of the new basic SCF scheme

$$\left\{ \begin{array}{l} \text{compute the Fock matrices } F_d^{(k)} := F_d(P_d^{(k)}, P_s^{(k)}), F_s^{(k)} := F_s(P_d^{(k)}, P_s^{(k)}), \\ \text{solve} \\ g_{\text{new}}(P_d^{(k)}, P_s^{(k)}) = \underset{(P_d, P_s) \in \mathcal{M}_{\text{DM}}}{\text{argmin}} \left\{ \text{Tr}(F_d^{(k)} P_d) + \text{Tr}(F_s^{(k)} P_s), (P_d, P_s) \in \mathcal{M}_{\text{DM}} \right\}, \\ \text{pick} \\ (P_d^{(k+1)}, P_s^{(k+1)}) \in g_{\text{new}}(P_d^{(k)}, P_s^{(k)}). \end{array} \right. \quad (36)$$

The fixed points  $(P_{d*}, P_{s*})$  of the function  $g_{\text{new}}$  satisfy

$$\left\{ \begin{array}{l} F_{d*} := F_d(P_{d*}, P_{s*}), F_{s*} := F_s(P_{d*}, P_{s*}) \\ (P_{d*}, P_{s*}) \in \underset{(P_d, P_s) \in \mathcal{M}_{\text{DM}}}{\text{argmin}} \{E_*(P_d, P_s), (P_d, P_s) \in \mathcal{M}_{\text{DM}}\}, \\ \text{with } E_*(P_d, P_s) = \text{Tr}(F_{d*} P_d) + \text{Tr}(F_{s*} P_s). \end{array} \right. \quad (37)$$

As  $E_*$  is a linear form, its gradient is constant and equal for the inner product  $\langle \cdot, \cdot \rangle_{\text{DM}}$  to  $(4F_{d*}, 4F_{s*})$ . Replacing  $E$  with  $E_*$  in the arguments in Section 2.3.2, we obtain that (37) implies (26), hence that any fixed point  $(P_{d*}, P_{s*})$  of  $g_{\text{new}}$  is a critical point of  $E$  on  $\mathcal{M}_{\text{DM}}$



The inner optimization problem

$$\operatorname{argmin} \left\{ \operatorname{Tr}(F_d^{(k)} P_d) + \operatorname{Tr}(F_s^{(k)} P_s), (P_d, P_s) \in \mathcal{M}_{\text{DM}} \right\} \quad (38)$$

on  $\mathcal{M}_{\text{DM}}$  solved at each step is easier and much cheaper to solve numerically than the original problem (6) since the function  $(P_d, P_s) \mapsto \operatorname{Tr}(F_d^{(k)} P_d) + \operatorname{Tr}(F_s^{(k)} P_s)$  is *linear* while the ROHF energy function  $E(P_d, P_s)$  is nonlinear (see Eq. (5)). In particular, the Coulomb and Fock terms are not recomputed at each iteration. To solve it, we can use a minimization algorithm with initial guess in

$$\operatorname{argmin} \left\{ \operatorname{Tr}(H^{(k)} P_d) + \frac{1}{2} \operatorname{Tr}(H^{(k)} P_s), (P_d, P_s) \in \mathcal{M}_{\text{DM}} \right\}, \quad (39)$$

where  $H^{(k)} = F_d^{(k)}$ , or  $H^{(k)} = H_{A,B}^{(k)}$ , with  $H_{A,B}^{(k)}$  given by (30). The solutions to (39) are easily obtained by diagonalizing  $H^{(k)}$  and applying the *Aufbau* principle. For  $H^{(k)} = H_{A,B}^{(k)}$ , the iterate of the new basic SCF scheme (36) is obtained using  $g_{A,B}(P_d^{(k)}, P_s^{(k)})$  as initial guess for the minimization problem (38). It is also possible, and more efficient in some cases, to use as an initial guess for the minimization problem (38), the previous iterate  $(P_d^{(k-1)}, P_s^{(k-1)})$ . Let us mention however that this approach only provides *local* (non-necessarily global) minima of (38). In practice, we choose for  $(P_d^{(k+1)}, P_s^{(k+1)})$  the approximation of the local minimum of  $(P_d, P_s) \mapsto \operatorname{Tr}(F_d^{(k)} P_d) + \operatorname{Tr}(F_s^{(k)} P_s)$  on  $\mathcal{M}_{\text{DM}}$  obtained by a few iterations of a preconditioned steepest-descent algorithm (see future work) starting from the Roothaan initial guess (i.e.  $g_{A,B}(P_d^{(k)}, P_s^{(k)})$  with  $A = (-\frac{1}{2}, \frac{1}{2}, \frac{3}{2})$  and  $B = (\frac{3}{2}, \frac{1}{2}, -\frac{1}{2})$ ).

In many cases, plain SCF fixed-point iterations using  $g_{A,B}$  or  $g_{\text{new}}$  do not converge (see Section 5) and must be stabilized. This is the matter of the next section.

### 3.2 Anderson-Pulay (DIIS-type) acceleration

Anderson-Pulay acceleration (APA) is a terminology recently proposed in [19] to gather various acceleration schemes, including Anderson's scheme and Pulay's direct inversion in the iteration space (DIIS) schemes, into a general framework. APA can be applied to fixed-point problems of the form

$$\text{find } x_* \in \mathcal{W} \text{ such that } g(x_*) = x_*, \quad (40)$$

where  $g : \mathcal{W} \rightarrow \mathcal{M}$  is a  $C^2$  function from an open subset  $\mathcal{W}$  of  $\mathbb{R}^n$  into a smooth submanifold  $\mathcal{M}$  of  $\mathbb{R}^n$ . In addition to a fixed-point iteration function  $g$ , APA schemes require a residual function  $f : \mathcal{W} \rightarrow \mathbb{R}^p$  of class  $C^2$  with  $p \leq n$  such that for any  $x_* \in \mathcal{W}$ ,  $g(x_*) = x_*$  if and only if  $f(x_*) = 0$  (the residual vanishes at solutions to the fixed point problem and only at those points). A possible choice is  $f_0(x) = x - g(x)$  (in which case  $p = n$ ), but the performance of the algorithm can usually be dramatically improved by resorting to well-suited residual functions. APA schemes are based on linear combinations of the current iterate with the previous ones. We will compare here the classical fixed-depth DIIS scheme (FD-DIIS), with the adaptive-depth DIIS scheme (AD-DIIS) recently introduced in [19].

The AD-DIIS iteration scheme is as follows:

1. **Initialization and first iteration:**  $x^{(0)} \in \mathcal{W}$ ,  $m_{\text{max}} \in \mathbb{N}^*$ ,  $\delta > 0$  and  $\eta > 0$  being given

- (a) set  $r^{(0)} = f(x^{(0)})$  (initial residual)
- (b) set  $x^{(1)} = g(x^{(0)})$  (first iteration: simple fixed-point step)
- (c)  $r^{(1)} = f(x^{(1)})$
- (d) set  $s^{(1)} = r^{(1)} - r^{(0)}$
- (e) set  $k = 1$  (iteration number)
- (f) set  $m_k = 1$  (depth of the algorithm for  $k = 1$ )

2. **Subsequent iterations:** while  $\|r^{(k)}\|_2 > \eta$  do

- (a) solve

$$(\alpha_1^{(k)}, \dots, \alpha_{m_k}^{(k)}) = \underset{(\alpha_1, \dots, \alpha_{m_k}) \in \mathbb{R}^{m_k}}{\operatorname{argmin}} \left\| r^{(k)} - \sum_{i=1}^{m_k} \alpha_i s^{(k-m_k+i)} \right\|_2 \quad (41)$$

- (b) set

$$\tilde{x}^{(k+1)} = x^{(k)} - \sum_{i=1}^{m_k} \alpha_i^{(k)} (x^{(k-m_k+i)} - x^{(k-m_k+i-1)}) \quad (42)$$

- (c) set  $x^{(k+1)} = g(\tilde{x}^{(k+1)})$
- (d) set  $r^{(k+1)} = f(x^{(k+1)})$
- (e) set  $s^{(k+1)} = r^{(k+1)} - r^{(k)}$
- (f) set  $m_{k+1}$  the largest integer  $m \leq \min(m_k + 1, m_{\max})$  such that  
 $\forall i : k + 1 - m \leq i \leq k \implies \delta \|r^{(i)}\|_2 < \|r^{(k+1)}\|_2$
- (g) set  $k = k + 1$ .

In the AD-DIIS scheme, the last  $m_k + 1$  iterates  $x^{(k-m_k)}, \dots, x^{(k-1)}, x^{(k)}$  are linearly combined to form  $\tilde{x}^{(k+1)}$ , and the new iterate  $x^{(k+1)}$  is obtained from  $\tilde{x}^{(k+1)}$  by applying the iteration function  $g$ . The coefficients of the linear combination (42) are obtained by solving the least square problem (41) involving the residuals computed with the function  $f$ . The depth parameter  $m_k$  is adjusted to take into account only the iterates  $x^{(i)}$  at which the Euclidean norm of the residual is lower than  $\delta^{-1}$  times the Euclidean norm of the residual at the current iterate (step 2(f)). The parameter  $m_{\max}$  is the maximum depth allowed. The FD-DIIS is recovered by choosing  $\delta = 0$ , in which case  $m_k = \min(k, m_{\max})$  at each iteration  $k$ .

In summary, an APA scheme is characterized by the choices of

1. an iteration function  $g$ ;
2. a residual function  $f$ ;
3. a scheme for selecting the previous iterates taken into account to compute the new iterate: FD-DIIS vs AD-DIIS, choice of  $m_{\max}$  (as well as  $\delta > 0$  in the case of AD-DIIS);
4. a convergence threshold  $\eta$ .

The first two ingredients are critical: bad choices compromise convergence. The parameter  $m_{\max}$  must be chosen large enough (typically  $m_{\max} = 10$  or  $20$  in quantum chemistry packages) to ensure fast convergence, using sufficient information from previous iterations. One of the limitations of FD-DIIS is that iterates with large residuals (far away from the minimizer) are considered as well, whereas they should be discarded. This is remedied by

the AD-DIIS scheme, but the latter requires choosing an additional empirical parameter  $\delta$ . Selecting the optimal convergence threshold  $\eta$  allowing one to obtain the desired precision on the quantity of interest at a minimum computational cost, is a very challenging problem of numerical analysis. In practice,  $\eta$  is chosen in a conservative manner to make sure that the energy is converged up to, say, a micro-hartree.

**Choice of  $g$ .** We need an iteration function defined in an open neighborhood  $\mathcal{W}$  of  $\mathcal{M}_{\text{DM}}$  since the points  $\tilde{x}^{(k)}$ , which are linear combinations of points of  $\mathcal{M}_{\text{DM}}$ , do not belong to  $\mathcal{M}_{\text{DM}}$  in general. We can directly use one of the basic SCF iteration functions  $g_{A,B}$  (Eq. (31)), or  $g_{\text{new}}$  (Eq. (36)), since they are defined for any point of  $\mathcal{V}_{\text{sym}}$ . Numerical tests seem to show that using  $g_{\text{new}}$  as such is effective.

On the other hand, it can be sometimes beneficial to add a purification step in the construction of the other iteration functions, and replace  $g_{A,B}$  by the function  $\tilde{g}_{A,B}$  defined in a relatively large neighborhood  $\mathcal{W}$  of  $\mathcal{M}_{\text{DM}}$  as follows: for  $\tilde{x} = (\tilde{P}_d, \tilde{P}_s)$  in  $\mathcal{W}$ ,

$$\left\{ \begin{array}{l} \text{compute the Fock matrices } \tilde{F}_d = F_d(\tilde{P}_d, \tilde{P}_s), \quad \tilde{F}_s = F_s(\tilde{P}_d, \tilde{P}_s), \\ \text{purify the density matrices } \tilde{P}_d \text{ and } \tilde{P}_s: \\ \quad P_o := \Theta \left( \tilde{P}_d + \tilde{P}_s - \frac{1}{2} I_{N_b} \right), \quad P_d := \Theta \left( P_o \tilde{P}_d P_o - \frac{1}{2} I_{N_b} \right), \quad (\Theta : \text{Heaviside function}), \\ \quad P_s = P_o - P_d, \quad P_v = I_{N_b} - P_o, \\ \text{assemble the effective Hamiltonian } H_{A,B}: \\ \quad P_v = I_{N_b} - P_d - P_s, \\ \quad R_{tt} := P_t (A_{tt} \tilde{F}_s + B_{tt} (\tilde{F}_d - \tilde{F}_s)) P_t, \quad t \in \{d, s, v\}, \\ \quad H_{\text{diag}} := R_{dd} + R_{ss} + R_{vv}, \\ \quad H_u := P_d (\tilde{F}_d - \tilde{F}_s) P_s + P_d \tilde{F}_d P_v + P_s \tilde{F}_s P_v, \\ \quad H_{A,B} := H_{\text{diag}} + H_u + H_u^T \\ \text{diagonalize } H_{A,B} \text{ in an orthonormal basis:} \\ \quad H_{A,B} \phi_i = \varepsilon_i \phi_i, \quad \phi_i^T \phi_j = \delta_{ij}, \quad \varepsilon_1 \leq \dots \leq \varepsilon_{N_b}, \\ \text{construct the new iterate using the } \textit{Aufbau} \text{ principle:} \\ \quad \tilde{g}_{A,B}(P_d, P_s) := \left( \sum_{i=1}^{N_d} \phi_i \phi_i^T, \sum_{i=N_d+1}^{N_o} \phi_i \phi_i^T \right) \end{array} \right.$$

However, overall, no clear improvement of using  $\tilde{g}_{A,B}$  purification function instead of  $g_{A,B}$  has been unveiled numerically, for the tested systems. We thus only report in the following results using the non-purified  $g_{A,B}$  iteration function (four our implementation of  $g_{A,B}$ ,  $f$  based DIIS – see below).

**Choice of  $f$ .** We take any  $C^2$  function

$$f : \mathcal{W} \rightarrow \mathbb{R}^{N_d \times N_s} \times \mathbb{R}^{N_d \times N_v} \times \mathbb{R}^{N_s \times N_v}$$

whose restriction to  $\mathcal{M}_{\text{DM}}$  is given by

$$f(P_d, P_s) := ((F_d(P_d, P_s) - F_s(P_d, P_s))^{ds}, (F_d(P_d, P_s))^{dv}, (F_s(P_d, P_s))^{sv}). \quad (43)$$

This residual function  $f$  does not depend on  $A_{tt}$ ,  $B_{tt}$  coefficients as it involves only *off-diagonal* blocks of  $H_{A,B}$  (namely, the terms composing the off-diagonal Hamiltonian part  $H_u$ ), but does not involve blocks composing  $H_{\text{diag}}$  ( $R_{tt}$  blocks, which depend on  $A_{tt}$ ,  $B_{tt}$  coefficients).

In DIIS algorithms, the residual function  $f$  is only evaluated at points of the manifold  $\mathcal{M}_{\text{DM}}$ , but must have a  $C^2$  extension to  $\mathcal{W}$  for local convergence to be mathematically

guaranteed [19]. This is obviously the case for the function  $f$  defined by (43) on  $\mathcal{M}_{\text{DM}}$ .

It is important to note that the DIIS algorithms implemented in the quantum chemistry codes that we have tested also use very similar residual functions (namely the commutator-based residual functions *à la* Pulay [16], involving the effective Hamiltonian  $H_{A,B}$ ).

## 4 Relaxed constrained algorithm for ROHF

Relaxed constrained algorithms for the Unrestricted and General Hartree-Fock setting were introduced in [24]. They consist in optimizing the energy functional in the DM formulation on the convex hull of the admissible set. For the UHF and GHF problems, it can be shown that the relaxed constrained problem has the same global minimizers as the original one [32, 33]. The advantage of the relaxed constrained problems is that convex combinations of admissible solutions are admissible solutions as well.

The simplest relaxed constrained algorithm is the optimal damping algorithm (ODA). It generates two sequences of iterates:

- a sequence  $(x^{(k)})$  of points on the admissible manifold  $\mathcal{M}_{\text{DM}}$ ;
- a sequence  $(\tilde{x}^{(k)})$  of points in the convex hull of  $\mathcal{M}_{\text{DM}}$ .

The point  $\tilde{x}^{(k+1)}$  is obtained by doing an optimal convex combination of  $\tilde{x}^{(k)}$  and  $x^{(k+1)}$ :

$$t_k = \operatorname{argmin}_{t \in [0,1]} E(tx^{(k+1)} + (1-t)\tilde{x}^{(k)}), \quad \tilde{x}^{(k+1)} = t_k x^{(k+1)} + (1-t_k)\tilde{x}^{(k)}.$$

The function  $p_k(t) := E(tx^{(k+1)} + (1-t)\tilde{x}^{(k)})$  is a second degree polynomial and we have

$$p_k(0) = E(\tilde{x}^{(k)}) \quad \text{and} \quad p'_k(0) = \langle \nabla E(\tilde{x}^{(k)}), x^{(k+1)} - \tilde{x}^{(k)} \rangle_{\text{DM}}.$$

Computing  $p_k(1) = E(x^{(k+1)})$ , we obtain the value of  $t_k$  explicitly. The point  $x^{(k+1)}$  is chosen so as to minimize the slope  $p'_k(0)$ ; it is therefore obtained from  $\tilde{x}^{(k)}$  as

$$x^{(k+1)} \in \operatorname{argmin}_{x \in \mathcal{M}_{\text{DM}}} \langle \nabla E(\tilde{x}^{(k)}), x \rangle_{\text{DM}} = g_{\text{new}}(\tilde{x}^{(k)}),$$

where  $g_{\text{new}}$  is defined in (36).

The ODA is initialized by choosing an initial guess  $x^{(0)} = (P_d^{(0)}, P_s^{(0)}) \in \mathcal{M}_{\text{DM}}$ , by setting  $\tilde{x}^{(0)} = x^{(0)}$ , and by computing  $(\tilde{F}_d^{(0)}, \tilde{F}_s^{(0)}) = (F_d^{(0)}, F_s^{(0)})$  using (20)-(21) with  $(P_d, P_s) = (P_d^{(0)}, P_s^{(0)})$ . The iterations are as follows:

$$\left\{ \begin{array}{l} \text{solve} \\ \quad g_{\text{new}}(\tilde{P}_d^{(k)}, \tilde{P}_s^{(k)}) = \operatorname{argmin} \left\{ \operatorname{Tr}(\tilde{F}_d^{(k)} P_d) + \operatorname{Tr}(\tilde{F}_s^{(k)} P_s), (P_d, P_s) \in \mathcal{M}_{\text{DM}} \right\}, \\ \text{pick} \\ \quad (P_d^{(k+1)}, P_s^{(k+1)}) \in g_{\text{new}}(\tilde{P}_d^{(k)}, \tilde{P}_s^{(k)}), \\ \text{compute the Fock matrices} \\ \quad F_d^{(k+1)} := F_d(P_d^{(k+1)}, P_s^{(k+1)}), \quad F_s^{(k+1)} := F_s(P_d^{(k+1)}, P_s^{(k+1)}), \\ \text{set} \\ \quad (\tilde{P}_d^{(k+1)}, \tilde{P}_s^{(k+1)}) = (1-t_k)(\tilde{P}_d^{(k)}, \tilde{P}_s^{(k)}) + t_k(P_d^{(k+1)}, P_s^{(k+1)}), \\ \quad (\tilde{F}_d^{(k+1)}, \tilde{F}_s^{(k+1)}) = (1-t_k)(\tilde{F}_d^{(k)}, \tilde{F}_s^{(k)}) + t_k(F_d^{(k+1)}, F_s^{(k+1)}), \\ \quad \text{where } t_k \text{ is the minimizer of the quadratic function} \\ \quad [0, 1] \ni t \mapsto E((1-t)(\tilde{P}_d^{(k)}, \tilde{P}_s^{(k)}) + t(P_d^{(k+1)}, P_s^{(k+1)})). \end{array} \right. \quad (44)$$

Note that the Fock matrices  $(\tilde{F}_d^{(k+1)}, \tilde{F}_s^{(k+1)})$  defined above are actually the Fock matrices associated with  $(\tilde{P}_d^{(k+1)}, \tilde{P}_s^{(k+1)})$  since the Fock matrices are affine functions of the density matrices.

## 5 Numerical results

### 5.1 Methodology

We now analyze the performances of the ODA and the new SCF algorithms introduced in this article using

- the new iteration functions  $\tilde{g}_{A,B}$  (31) or  $g_{\text{new}}$  (36),
- the new DIIS residual function  $f$  given in (43).

We also compare with the SCF algorithms for ROHF available in GAMESS [34]. We have chosen this popular software because all the classical functions  $g_{A,B}$  are implemented, as well as the SOSCF algorithm. We have also run tests with Psi4 [35] (for which only Guest and Saunders'  $g_{A,B}$  is available) and Q-Chem [27], as well as other quantum chemistry packages. The DIIS residual functions implemented in these codes can be slightly different but all also use commutator-based residual functions *à la* Pulay [16], involving the effective Hamiltonian  $H_{A,B}$ .




Convergence behaviors are investigated in two distinct regimes:

- the global convergence regime. The goal here is to reach the vicinity of a minimizer, starting from a bad initial guess obtained in practice by diagonalizing the core Hamiltonian;
- the local convergence regime, when the initial guess is close to a minimizer. We choose in this study the extended Hückel initial guess derived from the Wolfsberg-Helmholtz approximation [36, 37, 38], as implemented in GAMESS.

It appears that in some cases, the local minima found by starting from the core initial guess, are lower in energy than other minima reached from extended Hückel initial guesses. We will elaborate further on this point in a future study.

The new algorithms we introduce, along with the classical SCF schemes, have been implemented in the Julia language [39] as a proof of concept. The best performing algorithms will be added as a plugin within the Quantum Package [40, 41] and made freely available to the community.

Throughout the next sections, qualitative convergence results are tagged by color with the following convention:

-  non-convergence: the energies of the iterates oscillate above the ground state energy by at least  $10^{-2}$  Ha and the residual does not go to zero. In many cases, the oscillations occur between 1 and 100 Ha above the ground state energy;
-  stagnation or small-amplitude oscillations : the algorithm stalls or the iterates display small-amplitude oscillations while the residual is small but not small enough in the sense that the limit values of the energy are  $10^{-4}$  to  $10^{-2}$  Ha higher than the ground state energy (or another local minimum)
-  convergence to a local minimizer.

Throughout the next sections, qualitative convergence results are tagged with the following convention:  $\checkmark$  converged,  $\times$  not-converged.

## 5.2 Basic SCF iterations

We first illustrate the limitations of the classical iteration functions  $g_{A,B}$  computed as in (31), and the relevance of the new iteration function  $g_{\text{new}}$  defined in (36), by analyzing the behavior of the corresponding basic SCF algorithms  $x^{(k+1)} = g(x^{(k)})$  (without any stabilization/acceleration technique) on simple monatomic systems: an oxygen atom in the triplet state, and  $\text{Fe}^{2+}$  and  $\text{Fe}^{3+}$  ions in high-spin configurations (respectively quintet and sextet states).

Recall that the function  $g_{A,B}$  is computed by diagonalizing an effective Hamiltonian depending on the input ROHF state and *ad hoc* coefficients  $A_{tt}$  and  $B_{tt}$ , and constructing the output ROHF state using the *Aufbau* principle (see Section 3). The performance of the basic SCF algorithm  $x^{(k+1)} = g_{A,B}(x^{(k)})$  is found to be very sensitive to the choice of the  $A_{tt}$  and  $B_{tt}$  coefficients; besides, no choice of coefficients provides consistent convergence for the three simple systems. In contrast, the basic fixed-point algorithm built upon the parameter-free iteration function  $g_{\text{new}}$  (36) has been able to converge for the three systems. The results reported in Table 1 have been obtained with the cc-pVDZ basis set and the Hückel initial guess from Quantum Package. Qualitatively similar results have been obtained with the core initial guess and/or other basis sets (6-31G).

	O (triplet)	Fe <sup>2+</sup> (quintet)	Fe <sup>3+</sup> (sextet)
Roothan $A_{tt} = (-\frac{1}{2}, \frac{1}{2}, \frac{3}{2}), B_{tt} = (\frac{3}{2}, \frac{1}{2}, -\frac{1}{2})$	✓(17)	×	✓(18)
McWeeny and Diercksen $A_{tt} = (\frac{1}{3}, \frac{1}{3}, \frac{2}{3}), B_{tt} = (\frac{2}{3}, \frac{1}{3}, \frac{1}{3})$	✓(17)	×	×
Davidson $A_{tt} = (\frac{1}{2}, 1, 1), B_{tt} = (\frac{1}{2}, 0, 0)$	×	×	✓(26)
Guest and Saunders $A_{tt} = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2}), B_{tt} = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$	✓(20)	×	✓(20)
Binkley, Pople and Dobosh $A_{tt} = (\frac{1}{2}, 1, 0), B_{tt} = (\frac{1}{2}, 0, 1)$	×	×	✓(29)
Fægri and Manne $A_{tt} = (\frac{1}{2}, 1, \frac{1}{2}), B_{tt} = (\frac{1}{2}, 0, \frac{1}{2})$	×	×	✓(28)
Euler equations $A_{tt} = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2}), B_{tt} = (\frac{1}{2}, 0, \frac{1}{2})$	✓(21)	×	×
Canonical-ROHF I $A_{tt} = (\frac{2S+1}{2S}, 1, 1), B_{tt} = (-\frac{1}{2S}, 0, 0)$	×	×	✓(26)
Canonical-ROHF II $A_{tt} = (0, 0, -\frac{1}{2S}), B_{tt} = (1, 1, \frac{2S+1}{2S})$	✓(16)	×	×
$g_{\text{new}}$ (parameter free)	✓(11)	✓(33)	✓(16)

Table 1: Convergence of the basic fixed-point algorithm  $x^{(k+1)} = g(x^{(k)})$  for the atomic systems O, Fe<sup>2+</sup>, and Fe<sup>3+</sup> (cc-pVDZ basis set, Hückel initial guess), for (i) the classical  $g_{A,B}$  iteration functions (see Table I in Ref. [29]), and (ii) the  $g_{\text{new}}$  iteration function (this work). The number of iterations needed to reach convergence is specified when the algorithm happens to converge (chosen convergence criterion: the energy of the current iterate is at most  $10^{-6}$  Ha above the ROHF ground state).

### 5.3 Stabilized and accelerated iteration schemes

Table 2 summarizes the benchmark systems considered in this section. They consist of organic molecules bearing aromatic moieties (such as pyridine or porphyrin), interacting with open-shell metallic ions (see Figure 2). These systems are representative of the complexity of open-shell calculations in quantum chemistry as they contain transition metal ions with high spin in interaction with non trivial aromatic organic ligands[42]. The combination of strong repulsion in the  $3d$  shell of the metals together with the very delocalized character of the  $\pi$  system in these organic ligands can lead to SCF instabilities precisely because, according to the choice of the flavour of effective Hamiltonian used in the  $g_{AB}$  function, the *Aufbau* principle is not fulfilled in these systems. We have picked up both systems having space symmetries, such as pyridine-Cu<sup>2+</sup> ( $C_s$  symmetry) and the Porphyrin model-Fe<sup>2+</sup> ( $D_{4h}$  symmetry), and systems with slightly broken symmetry, such as Pyridine-Fe<sup>n+</sup>. We infer the spin multiplicities  $M = 2S + 1$  of these systems (where  $S$  is the total spin) from the corresponding spin multiplicities of the metallic ions, following Hund’s rule. In some cases, it is actually challenging to determine the spin multiplicity of the ground state (*e.g.* triplet or quintet), such as for the iron-porphyrin model system [42]. We have performed some test calculations on a full Porphyrin-Fe<sup>2+</sup> system (37 atoms, 269 basis functions for 6-31G), that yielded qualitatively similar results as for the Porphyrin model-Fe<sup>2+</sup> system. For the sake of brevity, we do not report them here.



System	Number of atoms	$N_d / N_s$	Multiplicity (2s+1)	Basis	Number of basis functions
Pyridine – Cu <sup>2+</sup>	12	34 / 1	2	6-31G	93
Pyridine – Cu <sup>2+</sup>	12	34 / 1	2	cc-pVDZ	164
Pyridine – Fe <sup>2+</sup>	12	31 / 4	5	6-31G	93
Pyridine – Fe <sup>2+</sup>	12	31 / 4	5	cc-pVDZ	164
Pyridine – Fe <sup>3+</sup>	12	30 / 5	6	6-31G	93
Pyridine – Fe <sup>3+</sup>	12	30 / 5	6	cc-pVDZ	164
Porphyrin model – Fe <sup>2+</sup>	29	66 / 4	5	6-31G	197
Porphyrin – Fe <sup>2+</sup>	37	90 / 4	5	6-31G	269

Table 2: Benchmark systems used in Section 5.3.

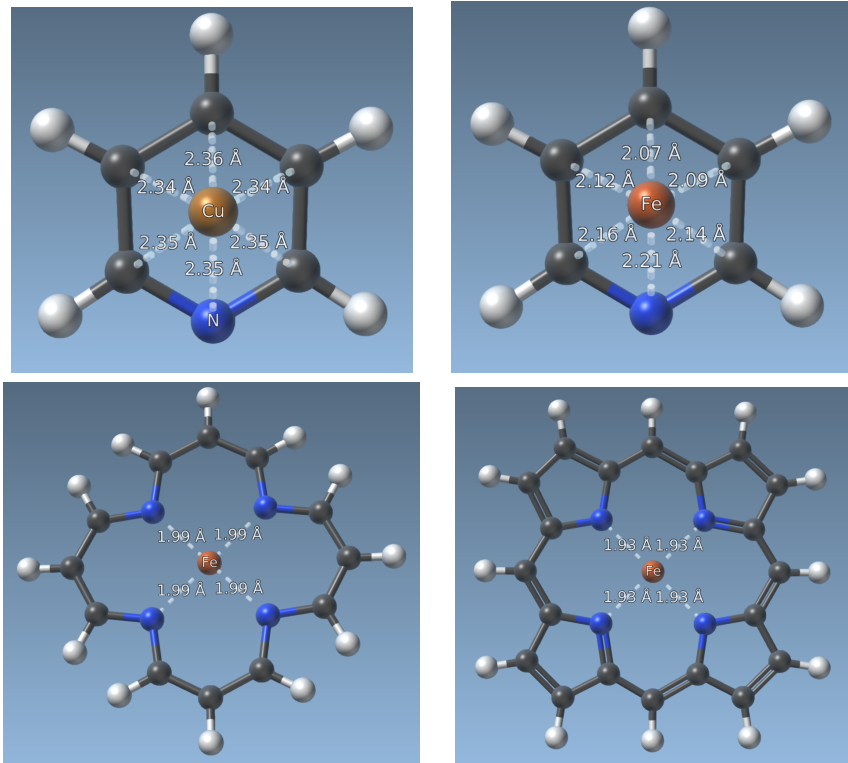


Figure 2: Upper Left : Pyridine - Cu<sup>2+</sup>. Upper Right : Pyridine - Fe<sup>2+</sup>. Lower left : Porphyrin model – Fe<sup>2+</sup> taken from [42]. Lower right : Porphyrin – Fe<sup>2+</sup>. Figures have been generated with SAMSON software [43].

We have tested several families of basis sets representative of quantum chemistry calculations, *i.e.* the 6-31G Pople’s type basis set [44, 45, 46], to the double-zeta correlation-consistent Dunning’s type basis set (cc-pVDZ) [47].

### 5.3.1 Global convergence regime

In this section, we analyze the ability of the various algorithms described in Sections 3 and 4 to reach the vicinity of a local minimizer from the core initial guess. We consider that this

is achieved if the energies of the iterates approach 0.1 Ha from the ROHF ground state energy. We compare the new algorithms proposed in this work with existing algorithms as implemented in GAMESS [34], namely the SOSCF algorithm and the FD-DIIS schemes built from the iteration functions  $g_{A,B}$  and a commutator-based residual function, which we denote by  $f$ . The results for the molecular systems in Table 2 in the 6-31G basis set are gathered in Table 3.

**Algorithms based on  $g_{A,B}$  iteration functions.** We observe in the second and third columns of Table 3 that the results of the GAMESS implementation of DIIS are very close to our  $f$ -DIIS implementation. However, for several choices of coupling coefficients  $A_{tt}$ ,  $B_{tt}$ , both methods fail to converge, and leads to oscillations. For the Pyridine- $\text{Fe}^{2+}$  and  $\text{Fe}^{3+}$  systems, and (respectively) for the Porphyrin model -  $\text{Fe}^{2+}$  system, only three (resp. two) specific choices of  $A_{tt}$ ,  $B_{tt}$  coefficients lead to convergence (notably Guest and Saunders and Roothan). The results for the Pyridine- $\text{Cu}^{2+}$  system (not reported) are qualitatively the same (only Guest and Saunders, Euler, Roothan and Canonical II choices of  $A_{tt}$ ,  $B_{tt}$  coefficients lead to convergence of GAMESS DIIS or of our  $f$ -DIIS implementation). The SOSCF second-order method converges whatever the choice of  $A_{tt}$ ,  $B_{tt}$  coefficients (except one, namely Canonical II, for the Pyridine- $\text{Fe}^{2+}$  system) from the core guess, although always in more than 200 iterations. Forcing DIIS (resp. SOSCF) from the first iterations is needed in GAMESS, as the DIIS residual (resp. gradient norm) is initially much higher than the default threshold for DIIS (resp. SOSCF) activation.

**Algorithms based on the  $g_{\text{new}}$  iteration function.** As shown in the last two columns of Table 3, the FD-DIIS algorithm based on the iteration function  $g_{\text{new}}$  and the residual function  $f$ , as well as the ODA algorithm (44), provide robust schemes for the four systems.

Let us underline that acceleration methods such as DIIS, are designed to accelerate local convergence (*i.e.* convergence when starting close enough to a local minimum). They are now well-understood mathematically in this setting [19]. In contrast, the fact that DIIS can stabilize SCF iterations starting from core initial guess in some cases (this is not always true) remains unexplained to our knowledge.

The iteration numbers reported in Table 3 correspond to a fixed-depth DIIS algorithm with a maximum depth parameter  $m_{\text{max}} = 10$ . Numerical tests show that these numbers of iterations needed to converge are, in some cases, very sensitive to the chosen depth parameter. Taking  $m_{\text{max}} = 20$ ,  $f$ -DIIS (with  $g_{A,B}$  fixed-point function) converges to microHartree accuracy in 135 iterations for Pyridine- $\text{Fe}^{2+}$ , instead of 2741 taking  $m_{\text{max}} = 10$  (see Table 4). Conversely taking  $m_{\text{max}} = 5$ , for the Pyridine- $\text{Cu}^{2+}$  system and  $g_{\text{new}}$  /  $f$ -DIIS method, convergence to microHartree accuracy is achieved in 27 iterations while the algorithm oscillate at about  $10^{-2}$  Ha of the solution for  $m_{\text{max}} = 10, 15$  and 20 (see Table 4).

### 5.3.2 Local convergence

We now compare the different algorithms starting from the extended Hückel initial guess provided by GAMESS, whose energy is about 1 to 2 Ha above the ground state energy for our test cases.

**Algorithms based on  $g_{A,B}$  iteration functions.** Comparing the results in Tables 3 and 5, we observe that  $f$ -DIIS algorithms as implemented in GAMESS barely benefit from a better initial guess. Yet, a detailed analysis of the local minima reached from the core

	GAMESS [34]		This work		
	$g_{A,B}$ (31) - based methods			$g_{\text{new}}$ (36) - based methods	
$A_{tt}, B_{tt}$ (see Table 1)	SOSCF	$f$ -DIIS	$f$ -DIIS	$f$ -DIIS	ODA
Pyridine-Fe <sup>2+</sup>					
Guest and Saunders	✓(244 ; 313)	✓(12 ; 37)	✓(15 ; 53)		
Euler	✓(218 ; 265)	✓(28 ; 95)	✓(17 ; 56)		
Roothan	✓(212 ; 263)	✓(28 ; 109)	✓(34 ; 96)		
Mc Weeny	✓(204 ; 254)	×	×		
Binkley	✓(262 ; 352)	×	×		
Faegri	✓(235 ; 278)	×	×		
Davidson	✓(230 ; 273)	×	×	✓(11; 1840)	✓(7; 112)
Canonical I	✓(262 ; 329)	×	×		
Other coefficients	×	×	×		
Pyridine-Fe <sup>3+</sup>					
Guest and Saunders	✓(236 ; 290)	✓(16 ; 132)	✓(9 ; 81)		
Roothan	✓(221 ; 263)	✓(19 ; 72)	✓(17 ; 193)		
Euler	✓(227 ; 277)	✓(41 ; 181)	✓(12 ; ~ 1000)		
Mc Weeny	✓(217 ; 273)	×	×		
Binkley	✓(216 ; 272)	×	×		
Faegri	✓(323 ; 374)	×	×		
Davidson	✓(259 ; 328)	×	×	✓(9 ; 2944)	✓(6 ; 5046)
Canonical I	✓(246 ; 317)	×	×		
Canonical II	✓(236 ; 305)	×	×		
Porphyrin model-Fe <sup>2+</sup>					
Guest and Saunders	✓(202 ; 215)	✓(15 ; 22)	✓(14; 67)		
Roothan	✓(203 ; 219)	✓(21 ; 34)	×		
Euler	✓(202 ; 218)	×	✓(49 ; 118)		
Mc Weeny	✓(202 ; 219)	×	×		
Binkley	✓(203 ; 213)	×	×		
Faegri	✓(203 ; 216)	×	×		
Davidson	✓(203 ; 221)	×	×	×(15; ∞)	✓(10; 23)
Canonical I	✓(203 ; 212)	×	×		
Canonical II	✓(294 ; 346)	×	×		

Table 3: Convergence results starting from core initial guess (6-31G basis set). The color code is detailed in the introduction to Section 5. The notation  $f$ -DIIS refers to a FD-DIIS method using  $f$  as residual function. The DIIS depth parameter  $m_{\text{max}}$  is fixed to 10 (default value in GAMESS). The notation  $(m; n)$  means that  $m$  iterations are needed to reach 0.1 Ha accuracy, while  $n$  iterations are necessary to reach microHartree accuracy.

FD-DIIS depth		5	10	15	20
System	Method				
Pyridine-Cu <sup>2+</sup>	$g_{A,B}$ , $f$ -DIIS (Guest and Saunders)	✓(22)	✓(20)	✓(21)	✓(23)
Pyridine-Fe <sup>2+</sup>	$g_{A,B}$ , $f$ -DIIS (Guest and Saunders)	✓(481)	✓(2741)	✓(1654)	✓(135)
Pyridine-Cu <sup>2+</sup>	$g_{\text{new}}$ , $f$ -DIIS (Guest and Saunders)	✓(27)	∞	∞	∞
Pyridine-Fe <sup>2+</sup>	$g_{\text{new}}$ , $f$ -DIIS (Guest and Saunders)	✓(1893)	✓(1840)	✓(1865)	✓(69)

Table 4: Number of iterations needed to reach microhartree accuracy, from the core initial guess, depending on the DIIS depth parameter  $m_{\text{max}}$ . Our implementation of DIIS was used.

guess and from the Hückel guess would be needed to assess the quality of the encountered local minima. Note that for the sake of brevity, the results of our implementation of  $f$ -DIIS (qualitatively very close to those of GAMESS implementation) are not reported Table 5. Four different choices of  $A_{tt}$ ,  $B_{tt}$  coefficients lead to convergence for Pyridine- $\text{Fe}^{2+}$  and Porphyrin model –  $\text{Fe}^{2+}$  systems (two for Pyridine- $\text{Fe}^{3+}$ ) for  $f$ -DIIS. Again, the SOSCF second-order method converges whatever the choice of  $A_{tt}$ ,  $B_{tt}$  coefficients, in less than 100 iterations (thanks to the improved starting guess) except for two specific choices of coefficients (106 iterations needed with Binkley and Faegri coefficients, for Pyridine- $\text{Fe}^{2+}$  system).

**Algorithms based on the  $g_{\text{new}}$  iteration function.** Both the DIIS algorithm constructed from  $f$  and the ODA converge for all the four systems. Note that in some cases, the number of iterations needed to converge is very high, *e.g.* for the ODA algorithm applied to the Pyridine- $\text{Fe}^{2+}$  system. In fact, the energy remains on a plateau for about one thousand iterations. We refer to the companion study for a detailed analysis of the origin of this problem, and methods to circumvent it.

	GAMESS		This work	
	$g_{A,B}$ (31) - based methods		$g_{\text{new}}$ (36) - based methods	
$A_{tt}, B_{tt}$ (see Table 1)	SOSCF	$f$ -DIIS	$f$ -DIIS	ODA
Pyridine–Fe <sup>2+</sup>				
Guest and Saunders	✓(78)	✓(82)	✓(52)	✓(∼ 3000)
Euler	✓(40)	✓(59)		
Roothaan	✓(83)	✓(255)		
McWeeny	✓(42)	✓(105)		
Binkley	✓(106)	×		
Faegri	✓(106)	×		
Davidson	✓(87)	×		
Canonical I	✓(88)	×		
Canonical II	✓(42)	×		
Pyridine–Fe <sup>3+</sup>				
Guest and Saunders	✓(78)	✓(178)	✓(∼4500)	✓(188)
Euler	✓(50)	×		
Roothan	✓(88)	✓(185)		
McWeeny	✓(88)	×		
Binkley	✓(93)	×		
Faegri	✓(92)	×		
Davidson	✓(94)	×		
Canonical I	✓(95)	×		
Canonical II	✓(54)	×		
Porphyrin model–Fe <sup>2+</sup>				
Guest and Saunders	✓(22)	✓(17)	✓(29)	✓(∼500)
Euler	✓(29)	✓(25)		
Roothan	✓(23)	✓(37)		
Mc Weeny	✓(36)	✓(32)		
Binkley	✓(23)	×		
Faegri	✓(22)	×		
Davidson	✓(21)	×		
Canonical I	✓(24)	×		
Canonical II	✓(29)	×		

Table 5: Convergence results starting from GAMESS extended Hückel initial guess (6-31G basis set). The color code is detailed in the introduction to Section 5. The notation  $f$ -DIIS refers to a FD-DIIS method using  $f$  as residual function. The DIIS depth parameter  $m_{\text{max}}$  is fixed to 10 (default value in GAMESS). The number of iterations in parentheses is the one needed to reach microHartree accuracy.

## 6 Conclusion and perspectives

In this article, we have provided a geometrical derivation of the ROHF equations based on the density matrix formalism and the appropriate differential geometry framework. A fundamental aspect of that derivation is the characterization of the tangent space of the manifold of ROHF states at a critical point of the ROHF energy functional, as well as its orthogonal complement (for the Frobenius inner product). This analysis lead us to introduce a new, parameter-free, iteration function  $g_{\text{new}}$  (see Eq. (36)), as alternatives to Roothaan-like iteration functions  $g_{A,B}$  based on the construction of a (non-physical)

effective Hamiltonian  $H_{A,B}$ , where  $A = (A_{dd}, A_{ss}, A_{dd})$  and  $B = (B_{dd}, B_{ss}, B_{dd})$  collect six real empirical parameters. An important conceptual difference of the proposed new SCF algorithm with respect to previous works is that it is not based on the usual technique of diagonalization of Fock-like Hamiltonians which can lead to numerical instabilities when the *Aufbau* principle is not fulfilled. Thanks to its geometrical formulation, the present algorithm avoids the ambiguity of the orbital energies for which the Koopman theorem does not apply in the case of the ROHF framework.

The numerical results we report seem to indicate that the (fixed-depth) FD-DIIS algorithm based on the usual  $g_{AB}$  framework with the Guest and Saunders ( $A_{tt} = B_{tt} = \frac{1}{2}$ ) and Roothan ( $A_{tt} = (-\frac{1}{2}, \frac{1}{2}, \frac{3}{2})$ ,  $B_{tt} = (\frac{3}{2}, \frac{1}{2}, -\frac{1}{2})$ ) iteration functions are quite robust and converge in a reasonable number of iterations, even when starting from the core initial guess. However, these observations, made on a small number of test cases (the ones reported in this paper plus a dozen of other challenging cases), do not guarantee that this algorithm will perform well for all systems and basis sets.

The numerical results reported here based on our new parameter-free iteration function  $g_{\text{new}}$  are encouraging as the latter always converges for the set of systems tested here, which involve different open-shell transition metal ions interacting with aromatic ligands. The algorithms based on the parameter-free iteration function  $g_{\text{new}}$  may then provide a useful alternative to the  $g_{A,B}$  iteration functions for very challenging systems. In particular, the ODA (involving  $g_{\text{new}}$ ) seems to be extremely robust and efficient in the early iterations, to reach the attraction basin of a local minimizer.

We also observed that the performance of the FD-DIIS algorithm is very sensitive to the depth parameter  $m_{\text{max}}$ , which should be chosen neither too small, nor too large. For instance, the orange cells in Table 3 (small oscillations close to a minimizer for  $m_{\text{max}} = 10$ ) turn green (convergence to a minimizer) for suitable choices of  $m_{\text{max}}$ . Unfortunately, the optimal value of  $m_{\text{max}}$  is system dependent, and no value seems to work for all systems. The (adaptive-depth) AD-DIIS algorithm was introduced in [19] to remedy this situation, and was shown to be more effective than FD-DIIS for Restricted Hartree-Fock and Kohn-Sham calculations. Unfortunately, we were not able to find a pair of system-independent parameters  $(m_{\text{max}}, \delta)$  (see Section 3.2) for which AD-DIIS has better local convergence properties than FD-DIIS for ROHF.

In a companion study, we will investigate direct minimization algorithms. These algorithms have several advantages: in particular, the convergence of some of them is guaranteed by general mathematical theorems. We will propose new direct minimization algorithms, and compare them with both SCF algorithms, and the existing geometrical direct method (GDM) [10] implemented in Q-Chem. We will also discuss the existence of several (sometimes many) local minima of the ROHF optimization problem, and compare the outputs of correlated post-Hartree-Fock methods when starting from different local minima.

## Acknowledgements

The authors would like to thank Michael F. Herbst, Antoine Levitt, Filippo Lipparini and Tommaso Nottoli for fruitful discussions, as well as the anonymous reviewers for very useful suggestions and comments. This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No 810367).

# Supplementary Material

The Supplementary Material contains the atomic coordinates of the benchmark systems studied in the article.

## Appendix

We provide here the energies at convergence for each system, algorithm, and initial guess. Table 6 corresponds to the energies associated to the results of Table 3 while Table 7 corresponds to the energies associated to the results of Table 5.

	GAMESS		This work		
	$g_{A,B}$ - based methods			$g_{\text{new}}$ - based methods	
$A_{tt}, B_{tt}$	SOSCF	$f$ -DIIS	$f$ -DIIS	$f$ -DIIS	ODA
Pyridine-Fe <sup>2+</sup>					
Guest and Saunders	✓(-1508.134652)	✓(-1508.134652)	✓(-1507.965876)		
Euler	✓(-1508.016536)	✓(-1508.016536)	✓(-1508.016538)		
Roothan	✓(-1508.016536)	✓(-1508.134040)	✓(-1508.130830)		
Mc Weeny	✓(-1508.016536)	×	×		
Binkley	✓(-1508.134652)	×	×		
Fægri	✓(-1508.016536)	×	×		
Davidson	✓(-1508.016536)	×	×		
Canonical I	✓(-1508.134652)	×	×		
Other coefficients	×	×	×	✓(-1508.134654)	✓(-1508.134042)
Pyridine-Fe <sup>3+</sup>					
Guest and Saunders	✓(-1507.414473)	✓(-1507.414091)	✓(-1507.349495)		
Roothan	✓(-1507.414473)	✓(-1507.343997)	✓(-1507.353456)		
Euler	✓(-1507.414473)	✓(-1507.414097)	✓(-1507.414472)		
Mc Weeny	✓(-1507.414473)	×	×		
Binkley	✓(-1507.414473)	×	×		
Fægri	✓(-1507.414473)	×	×		
Davidson	✓(-1507.414473)	×	×		
Canonical I	✓(-1507.414473)	×	×		
Canonical II	✓(-1507.414473)	×	×	✓(-1507.414474)	✓(-1507.414474)
Porphyrin model-Fe <sup>2+</sup>					
Guest and Saunders	✓(-1940.163309)	✓(-1940.513025)	✓(-1940.357814)		
Roothan	✓(-1940.163309)	✓(-1940.335945)	×		
Euler	✓(-1940.163309)	×	✓(-1940.513030)		
Mc Weeny	✓(-1940.163309)	×	×		
Binkley	✓(-1939.977138)	×	×		
Fægri	✓(-1939.977138)	×	×		
Davidson	✓(-1939.977138)	×	×		
Canonical I	✓(-1940.075387)	×	×		
Canonical II	✓(-1940.267466)	×	×	~(-1940.46)	✓(-1940.513031)

Table 6: Energies at convergence starting from core initial guess (6-31G basis set). The color code is detailed in the introduction to Section 5. The notation  $f$ -DIIS refers to a FD-DIIS method using  $f$  as residual function. The DIIS depth parameter  $m_{\text{max}}$  is fixed to 10 (default value in GAMESS). All energies are expressed in Hartrees.

## References

- [1] Eliseo Ruiz, Joan Cano, Santiago Alvarez, and Pere Alemany. Magnetic Coupling in End-On Azido-Bridged Transition Metal Complexes: A Density Functional Study. *J. Am. Chem. Soc.*, 120(43):11122–11129, Nov 1998.



	GAMESS		This work	
	$g_{A,B}$ - based methods		$g_{\text{new}}$ - based methods	
$A_{tt}, B_{tt}$	SOSCF	$f$ -DIIS	$f$ -DIIS	ODA
Pyridine-Fe <sup>2+</sup>				
Guest and Saunders	✓(-1508.134652)	✓(-1508.013967)	✓(-1508.134654)	✓(-1508.134654)
Euler	✓(-1508.134652)	✓(-1508.134652)		
Roothan	✓(-1508.016536)	✓(-1507.967145)		
McWeeny	✓(-1508.134652)	✓(-1508.134652)		
Binkley	✓(-1508.134652)	×		
Faegri	✓(-1508.134652)	×		
Davidson	✓(-1508.134652)	×		
Canonical I	✓(-1508.134652)	×		
Canonical II	✓(-1508.134652)	×		
Pyridine-Fe <sup>3+</sup>				
Guest and Saunders	✓(-1507.414473)	✓(-1507.409935)	✓(-1507.414474)	✓(-1507.357500)
Euler	✓(-1507.357499)	×		
Roothan	✓(-1507.414473)	✓(-1507.409935)		
Mc Weeny	✓(-1507.357499)	×		
Binkley	✓(-1507.414473)	×		
Faegri	✓(-1507.414473)	×		
Davidson	✓(-1507.414473)	×		
Canonical I	✓(-1507.414473)	×		
Canonical II	✓(-1507.357499)	×		
Porphyrin model-Fe <sup>2+</sup>				
Guest and Saunders	✓(-1940.406548)	✓(-1940.513025)	✓(-1940.512761)	✓(-1940.534138)
Euler	✓(-1940.385615)	✓(-1940.513025)		
Roothan	✓(-1940.406548)	✓(-1940.335945)		
Mc Weeny	✓(-1940.650207)	✓(-1940.513025)		
Binkley	✓(-1940.513025)	×		
Faegri	✓(-1940.513025)	×		
Davidson	✓(-1940.513025)	×		
Canonical I	✓(-1940.513025)	×		
Canonical II	✓(-1940.650207)	×		

Table 7: Energies at convergence starting from GAMESS extended Hückel initial guess (6-31G basis set). The color code is detailed in the introduction to Section 5. The notation  $f$ -DIIS refers to a FD-DIIS method using  $f$  as residual function. The DIIS depth parameter  $m_{\text{max}}$  is fixed to 10 (default value in GAMESS). All energies are expressed in Hartrees.

- [2] Robert K. Szilagyi, Markus Metz, and Edward I. Solomon. Spectroscopic Calibration of Modern Density Functional Methods Using [CuCl<sub>4</sub>]<sup>2-</sup>. *J. Phys. Chem. A*, 106(12):2994–3007, Mar 2002.
- [3] A. A. Ramírez-Solís, R. Poteau, A. Vela, and J. P. Daudey. Comparative studies of the spectroscopy of CuCl<sub>2</sub>: DFT versus standard ab initio approaches. *J. Chem. Phys.*, 122(16):164306, 2005.
- [4] Simone Kossmann, Barbara Kirchner, and Frank Neese. Performance of modern density functional theory for the prediction of hyperfine structure: meta-GGA and double hybrid functionals. *Mol. Phys.*, 105(15-16):2049–2071, 2007.
- [5] Mihail Atanasov, Peter Comba, Bodo Martin, Vera Müller, Gopalan Rajaraman, Heidi Rohwer, and Steffen Wunderlich. DFT models for copper(II) bispidine complexes: Structures, stabilities, isomerism, spin distribution, and spectroscopy. *J. Comput. Chem.*, 27(12):1263–1277, Sep 2006.

- [6] Pragya Verma and Donald G. Truhlar. Status and Challenges of Density Functional Theory. *Trends Chem.*, 2(4):302–318, Apr 2020.
- [7] Qianli Ma and Hans-Joachim Werner. Scalable Electron Correlation Methods. 7. Local Open-Shell Coupled-Cluster Methods Using Pair Natural Orbitals: PNO-RCCSD and PNO-UCCSD. *J. Chem. Theory Comput.*, 16(5):3135–3151, May 2020.
- [8] T. Tsuchimochi and G.E. Scuseria. Communication: ROHF theory made simple. *The Journal of Chemical Physics*, 133(141102), 2010.
- [9] George H. Booth, Andreas Grüneis, Georg Kresse, and Ali Alavi. Towards an exact description of electronic wavefunctions in real solids. *Nature*, 493(7432):365–370, Jan 2013.
- [10] B.D. Dunietz, T. Van Voorhis, and M. Head-Gordon. Geometric direct minimization of Hartree-Fock calculations involving open shell wavefunctions with spin restricted orbitals. *Journal of Theoretical and Computational Chemistry*, 01(02):255–261, oct 2002.
- [11] T. Van Voorhis and M. Head-Gordon. A geometric approach to direct minimization. *Molecular Physics*, 100(11):1713–1721, jun 2002.
- [12] E. Cancès, G. Kemlin, and A. Levitt. Convergence analysis of direct minimization and self-consistent iterations. *SIAM Journal of Matrix Analysis*, in press, 2020.
- [13] C.C.J. Roothaan. New developments in molecular orbital theory. *Reviews of Modern Physics*, 23(2):69–89, apr 1951.
- [14] V.R. Saunders and I.H. Hillier. A "level-shifting" method for converging closed shell Hartree-Fock wave functions. *International Journal of Quantum Chemistry*, 7(4):699–705, jul 1973.
- [15] P. Pulay. Convergence acceleration of iterative sequences. the case of SCF iteration. *Chemical Physics Letters*, 73(2):393–398, jul 1980.
- [16] P. Pulay. Improved SCF convergence acceleration. *Journal of Computational Chemistry*, 3(4):556–560, 1982.
- [17] T.P. Hamilton and P. Pulay. Direct inversion in the iterative subspace (DIIS) optimization of open-shell, excited-state, and small multiconfiguration SCF wave functions. *The Journal of Chemical Physics*, 84(10):5728–5734, may 1986.
- [18] T. Rohwedder and R. Schneider. An analysis for the DIIS acceleration method used in quantum chemistry calculations. *J. Math. Chem.*, 49:1889–1914, 2011.
- [19] M. Chupin, M.-S. Dupuy, G. Legendre, and E. Séré. Convergence analysis of adaptive DIIS algorithms with application to electronic ground state calculations. *arXiv preprint arXiv:2002.12850*, 2020.
- [20] G.B. Bacskay. A quadratically convergent Hartree–Fock (QC-SCF) method. application to closed shell systems. *Chemical Physics*, 61(3):385–404, oct 1981.
- [21] L. Thøgersen, J. Olsen, D. Yeager, P. Jørgensen, P. Salek, and T. Helgaker. The trust-region self-consistent field method: Towards a black-box optimization in Hartree–Fock and Kohn–Sham theories. *The Journal of Chemical Physics*, 121(1):16, 2004.

- [22] G. Chaban, M.W. Schmidt, and M.S. Gordon. Approximate second order method for orbital optimization of SCF and MCSCF wavefunctions. *Theoretical Chemistry Accounts*, 97(1):88–95, 1997.
- [23] F. Neese. Approximate second-order SCF convergence for spin unrestricted wavefunctions. *Chemical Physics Letters*, 325(1-3):93–98, 2000.
- [24] E. Cancès and C. Le Bris. Can we outperform the DIIS approach for electronic structure calculations? *International Journal of Quantum Chemistry*, 79(2):82–90, 2000.
- [25] K.N. Kudin, G.E. Scuseria, and E. Cancès. A black-box self-consistent field convergence algorithm: One step closer. *The Journal of Chemical Physics*, 116(19):8255, 2002.
- [26] E. Cancès, M. Defranceschi, W. Kutzelnigg, C. Le Bris, and Y. Maday. *Computational quantum chemistry: a primer*, volume X of *Handbook of Numerical Analysis*, pages 3–270. North-Holland, Amsterdam, 2003.
- [27] Y. Shao, Z. Gan, E. Epifanovsky, A.T.B. Gilbert, M. Wormit, J. Kussmann, A.W. Lange, A. Behn, J. Deng, X. Feng, et al. Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Molecular Physics*, 113(2):184–215, 2015.
- [28] MF Guest and VR Saunders. On methods for converging open-shell Hartree-Fock wave-functions. *Molecular Physics*, 28(3):819–828, 1974.
- [29] B.N. Plakhutin and E.R. Davidson. Canonical form of the Hartree-Fock orbitals in open-shell systems. *The Journal of Chemical Physics*, 140(1):014102, jan 2014.
- [30] T. Helgaker, P. Jørgensen, and J. Olsen. *Molecular Electronic-Structure Theory*. John Wiley & Sons, Ltd, aug 2000.
- [31] C.C.J. Roothaan. Self-consistent field theory for open shells of electronic systems. *Reviews of Modern Physics*, 32(2):179, 1960.
- [32] Eric Cancès. Scf algorithms for hf electronic calculations. In *Mathematical models and methods for ab initio quantum chemistry*, pages 17–43. Springer, 2000.
- [33] Eric Cancès, Gaspard Kemlin, and Antoine Levitt. Convergence analysis of direct minimization and self-consistent iterations. *SIAM Journal on Matrix Analysis and Applications*, 42(1):243–274, 2021.
- [34] M.W. Schmidt, K.K. Baldridge, J.A. Boatz, S.T. Elbert, M.S. Gordon, J.H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T.L. Windus, M. Dupuis, and J.A. Montgomery. General atomic and molecular electronic structure system. *Journal of Computational Chemistry*, 14(11):1347–1363, nov 1993.
- [35] Justin M Turney, Andrew C Simmonett, Robert M Parrish, Edward G Hohenstein, Francesco A Evangelista, Justin T Fermann, Benjamin J Mintz, Lori A Burns, Jeremiah J Wilke, Micah L Abrams, et al. Psi4: an open-source ab initio electronic structure program. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2(4):556–565, 2012.

- [36] M. Wolfsberg and L. Helmholz. The spectra and electronic structure of the tetrahedral ions  $\text{MnO}_4^-$ ,  $\text{CrO}_4^-$ , and  $\text{ClO}_4^-$ . *The Journal of Chemical Physics*, 20(5):837–843, 1952.
- [37] R. Hoffmann. An extended Hückel theory. I. Hydrocarbons. *The Journal of Chemical Physics*, 39(6):1397–1412, 1963.
- [38] J.H. Ammeter, H.B. Bürgi, J.C. Thibeault, and R. Hoffmann. Counterintuitive orbital mixing in semiempirical and ab initio molecular orbital calculations. *Journal of the American Chemical Society*, 100(12):3686–3692, 1978.
- [39] J. Bezanson, A. Edelman, S. Karpinski, and V.B. Shah. Julia: A fresh approach to numerical computing. *SIAM Review*, 59(1):65–98, 2017.
- [40] A. Scemama, T. Applencourt, Y. Garniron, E. Giner, G. David, and M. Caffarel. Quantum package v1. 0. *Zenodo*. <http://dx.doi.org/10.5281/zenodo.200970>, 2016.
- [41] Y. Garniron, T. Applencourt, K. Gasperich, A. Benali, A. Ferté, J. Paquier, B. Pradines, R. Assaraf, P. Reinhardt, J. Toulouse, et al. Quantum package 2.0: An open-source determinant-driven suite of programs. *Journal of chemical theory and computation*, 15(6):3591–3609, 2019.
- [42] G. Li Manni and A. Alavi. Understanding the mechanism stabilizing intermediate spin states in Fe (ii)-porphyrin. *The Journal of Physical Chemistry A*, 122(22):4935–4947, 2018.
- [43] <https://www.samson-connect.net/>.
- [44] P.C. Hariharan and J.A. Pople. Accuracy of  $\text{AH}_n$  equilibrium geometries by single determinant molecular orbital theory. *Molecular Physics*, 27(1):209–214, jan 1974.
- [45] M.M. Francl, W.J. Pietro, W.J. Hehre, J.S. Binkley, M.S. Gordon, D.J. DeFrees, and J.A. Pople. Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. *The Journal of Chemical Physics*, 77(7):3654–3665, oct 1982.
- [46] J.-P. Blaudeau, M.P. McGrath, L.A. Curtiss, and L. Radom. Extension of Gaussian-2 (G2) theory to molecules containing third-row atoms K and Ca. *The Journal of Chemical Physics*, 107(13):5016–5021, oct 1997.
- [47] T.H. Dunning. Gaussian basis sets for use in correlated molecular calculations. i. the atoms boron through neon and hydrogen. *The Journal of Chemical Physics*, 90(2):1007–1023, jan 1989.