Calculations of Ground and Excited Electronic States Using Self-Interaction Corrected Density Functionals

Aleksei V. Ivanov

Faculty of Physical Sciences
University of Iceland
2021

Calculations of Ground and Excited Electronic States Using Self-Interaction Corrected Density Functionals

Aleksei V. Ivanov

Dissertation submitted in partial fulfillment of a
Philosophiae Doctor degree in Chemistry

Advisor
Hannes Jónsson

PhD Committee
Hannes Jónsson
Egill Skúlason
Elvar Örn Jónsson

Opponents
Thomas Olsen
Andrei Manolescu

Faculty of Physical Sciences
School of Engineering and Natural Sciences
University of Iceland
Reykjavik, April 2021

Calculations of Ground and Excited Electronic States Using Self-Interaction Corrected Density Functionals
Ground and Excited Electronic States from SIC-DFT
Dissertation submitted in partial fulfillment of a *Philosophiae Doctor* degree in Chemistry

Faculty of Physical Sciences
School of Engineering and Natural Sciences
University of Iceland
Dunhaga 5
107, Reykjavik
Iceland

Telephone: 525-4700

Bibliographic information:
Aleksei V. Ivanov, 2021, *Calculations of Ground and Excited Electronic States Using Self-Interaction Corrected Density Functionals*, PhD dissertation, Faculty of Physical Sciences, University of Iceland, 196 pp.

ISBN 978-9935-9579-4-8

Printing: Haskolaprent
Reykjavik, Iceland, April 2021

Abstract

Direct optimization methods for the calculation of ground and excited electronic states are presented for both total density and orbital-density-dependent functionals. The methods have been developed for various types of basis sets including localized atomic orbitals, plane waves and real space grid. The algorithms have been implemented in combination with the projector-augmented-wave method to represent inner electrons of the atoms. The direct optimization method is shown to be more robust and faster than the conventional self-consistent field approach in calculations of both ground and excited states. An assessment of the Perdew-Zunger self-interaction correction (PZ-SIC) to the energy functional has also been made and its performance compared to the commonly used generalized gradient approximation (GGA). PZ-SIC is found to systematically improve the description of the atomization and ionization energy as well as the band gaps of insulators, but needs to be scaled by a half. PZ-SIC can be especially important for the accurate description of systems containing transition metals as is illustrated by the excellent results obtained for the Mn dimer, a system where results of GGA calculations are qualitatively incorrect. However, PZ-SIC does not substantially improve the excitation energy of small organic molecules as the correction there tends to cancel out when the energy of ground and excited states is compared. The efficient and practical implementation of PZ-SIC presented here paves the way for the development of more accurate orbital-density-dependent functionals.

Útdráttur

Aðferðir til að reikna út grunn og örvuð rafeindaástönd með beinni bestun hafa verið þróaðar, bæði fyrir felli háð heildarrafeindaþéttni sem og almennari felli háð þéttni svigrúmanna. Aðferðirnar hafa verið þróaðar fyrir ýmsar gerðir grunna svo sem staðbundin atómsvigrúm, planbylgjur og grind í raunrúminu. Þær hafa verið innleiddar með 'projector-augmented-wave' aðferðinni til að lýsa áhrifum innri rafeinda atómanna. Beina bestunin reynist vera áreiðanlegri og hraðvirkari en þær aðferðir sem áður hafa verið notaðar bæði hvað varðar reikninga á grunnástöndum sem og örvuðum ástöndum. Áhrif sjálfsvíxverkunarleiðréttingar Perdew og Zunger (PZ-SIC) á útkomu reikninganna hefur einnig verið könnuð og niðurstöðurnar bornar saman við almennu stigulnálgunina (GGA). PZ-SIC leiðréttingin reynist bæta kerfisbundið útreiknaða sundrunarorku sameinda sem og jónunarorku og einnig orkugeil einangrara, en þarf að skalast niður í helming. Reikningar á eiginleikum kerfa með hliðarmálmatóm geta batnað sérlega mikið með PZ-SIC leiðréttingunni og þar er Mn tvennan sérlega afgerandi tilfelli þar sem reikningar með PZ-SIC gefa mjög góða niðurstöðu á sama tíma og GGA gefur verulega rangar niðurstöður. Hins vegar hefur PZ-SIC ekki mikil áhrif á útreiknaða örvunarorku lítilla lífrænna sameinda þar eð leiðréttingin á grunn og örvaða ástandinu styttist þar að miklu leiti út. Sú skilvirka innsetning á PZ-SIC sem er framsett hér opnar leiðina fyrir þróun á nákvæmari fellum háðum svigrúmaþéttni.

Table of Contents

List of Figures

List of Tables

List of Original Articles

Articles included in this thesis:

Article I: Ivanov A. V., Jónsson E. Ö., Vegge T., Jónsson H., 2021, Direct Energy Minimization Based on Exponential Transformation in Density Functional Calculations of Finite and Extended Systems, *Submitted to Comput. Phys. Commun., arXiv:2101.12597 [physics.comp-ph].*

Article II: Levi G., Ivanov A.V., Jónsson H., 2020, Variational Density Functional Calculations of Excited States via Direct Optimization, *J. Chem. Theory Comput.* **16***, 6968-6982.*

Article III: Ivanov A.V., Levi G., Jónsson E. Ö, Jónsson H., 2021, Direct Optimization Method for Variational Excited-State Density Functional Calculations Using Real Space Grid or Plane Waves, *Submitted to J. Chem. Theory Comput., arXiv:2102.06542 [physics.comp-ph]*

Article IV: Ivanov A. V., Ghosh T. K., Jónsson E. Ö., Jónsson H., 2021, Mn Dimer can be Described Accurately with Density Functional Calculations when Self-interaction Correction is Applied, *Accepted to J. Phys. Chem. Lett., arxiv:2102.00890 [physics.chem-ph].*

Other articles published during PhD study but not included in this thesis:

Article V: Ivanov A.V., Uzdin V.M., Jónsson H., 2021, Fast and robust algorithm for energy minimization of spin systems applied in an analysis of high temperature spin configurations in terms of skyrmion density, *Comput. Phys. Commun.* **260** *107749*

Article VI: Levi G., Ivanov A.V., Jónsson H., 2020, Variational calculations of excited states via direct optimization of the orbitals in DFT, *Faraday Discuss.* **224** *448*

Article VII: Ivanov A.V., Dagbartsson D., Tranchida J., Uzdin V. M., Jónsson H., 2020, Efficient optimization method for finding minimum energy paths of magnetic transitions, *J. Phys.: Condens. Matter* **32** *345901*

Article VIII: Ivanov A.V., Bessarab P.F., Jónsson H., Uzdin V.M, 2020, Fully self-consistent calculations of magnetic structure within non-collinear Alexander-Anderson model, *Nanosystems: Physics, Chemistry, Mathematics* **11(1)** *65–77*

Article IX: Ivanov A.V., Bessarab P.F., Uzdin V.M., Jónsson H., 2017, Magnetic exchange force microscopy: Theoretical analysis of induced magnetization reversals, *Nanoscale* **9** *13320*

Acknowledgments

I am deeply indebted to my academic supervisor Prof. Hannes Jónsson for his unstinting support, supervision, valuable research advice and the opportunities he provided to visit fellow research groups for collaborative projects. He is a great mind and I feel incredibly fortunate to have had the opportunity to work with him. I would also like to thank Dr. Pavel Bessarab and Prof. Valery Uzdin from St. Petersburg State University for introducing me to Prof. Hannes Jónsson and their frequent advice during my study. I am also grateful to Dr. Elvar Örn Jónsson for the fruitful collaboration and generous assistance with GPAW software as well as Dr. Gianluca Levi for his invaluable insights into the study of excited states. Finally, I would like to extend a special thanks to my parents and sister for their unwavering support as well as my wife, Sally, for her encouragement, infinite patience and the incredible years we have spent together in Iceland.

1 Introduction

Quantum chemical calculations have become commonplace in science and industry thanks to the development of electronic structure algorithms, density functionals, and software established throughout both the previous and current centuries. Moreover, the increase in computational performance due to semiconductor miniaturization, exemplified by Moore's law, has made it possible to perform electronic structure calculations even on a regular laptop. However, with the foreseeable end of Moore's law (Waldrop, 2016; Leiserson et al., 2020), algorithm and software engineering will play an even more crucial role in achieving further growth in computational performance (Leiserson et al., 2020), not only in quantum chemical simulations, but also in other areas of computational physics and chemistry. Therefore, the development of new and efficient approaches in electronic structure calculations are of the upmost importance, particularly due to the increasing usage of computation for the interpretation of sophisticated experimental data and the subsequent formulation of theory.

Modern advanced computational methods based on wave function formalism and many-body theory allow for an accurate description of quantum systems. However, the single-determinant mean-field approach to the electronic structure problem remains the tool of choice for many calculations, especially those of large systems. This is due to the fact that such an approach has lower computational complexity as compared to other electronic structure methods and, at the same time, can either provide an accurate approximation to the true many-body problem or be used as an initial guess for advanced quantum chemical calculations.

The two main mean-field *ab initio* methods are Hartree-Fock theory (HFT) and density functional theory (DFT) (Hohenberg and Kohn, 1964; Kohn and Sham, 1965). While DFT has enjoyed considerable success in predicting the total energy and geometry of many quantum systems, its practical implementations based on electron density and its gradient do not deliver a sufficiently accurate description of ionization energies, band gaps, localized electronic states and some magnetic systems (Peng and Perdew, 2017). Meanwhile, HFT tends to produce errors of opposite sign as schematically shown in figure 1.1. This has led to the development of an evergrowing number of hybrid energy functionals where calculations are based on a combination of the two approaches, the mixture of which is often treated as an empirical parameter.

The solution to the problem outlined above may be achieved through the introduction of explicit orbital density dependence (ODD) in the energy functional and by carrying out variational minimization of the total energy with respect to the orbitals. An example of such a model is the Perdew-Zunger self-interaction correction (Perdew and Zunger, 1981) (PZ-SIC), where the spurious self-interaction error inherent to density functional approximations is exactly cancelled out in the one-electron limit. Although PZ-SIC

	KS-DFT (Commonly used approximations)	Hartree-Fock Theory
Atomization Energies	✓	✗
Delocalized states	✓	✗
Localized states	✗	✓
Orbital Energies (HOMO, LUMO)	✗	✗
Magnetic Systems	✗	✗

Figure 1.1. A schematic representation of the successes and failures of both density functional theory (DFT) and Hartree-Fock theory (HFT). Semilocal functionals (SLF) can provide a reliable estimate of atomization energy, while HFT typically underbinds molecules. SLF favors delocalized states, while HFT often localizes electron wave functions. Koopman's theorem states that absolute value of the HOMO energy of HFT is the first ionization energy of a molecule. However, this does not take into account the relaxation of ionic orbitals and therefore, this tends to overestimate the ionization energy. Conversely, SLF underestimates the ionization energy unless Koopman's theorem is enforced through the introduction of additional constraints. Finally, some magnetic systems can be described with SLF, but due to the presence of both localized (d, f electrons) and delocalized (s, p electrons) states both SLF and HFT struggle to accurately describe such systems.

was proposed approximately 40 years ago, the accurate variational assessment of this functional has only recently started to appear in literature (Bylaska et al., 2004, 2006; Klüpfel et al., 2011; Jonsson, 2011; Klüpfel et al., 2012b; Gudmundsdóttir et al., 2015; Lehtola et al., 2016). This is due to the fact that the PZ-SIC is an orbital-density-dependent functional lacking the unitary invariant symmetry and therefore, conventional SCF algorithms developed for DFT functionals cannot be applied to PZ-SIC without significant modifications.

While methods for solving KS self-consistent field equations for electronic ground state are in principle well-established in the literature (Payne et al., 1992; Furthmüller and Kresse, 1996; Van Voorhis and Head-Gordon, 2002; VandeVondele and Hutter, 2003; Garza and Scuseria, 2012), the development of efficient and robust algorithms for the calculation of excited states within a time-independent DFT framework is currently of great interest (Ye et al., 2017; Hait and Head-Gordon, 2020; Levi et al., 2020b,a; Carter-Fenk and Herbert, 2020; Ivanov et al., 2021c). However, this also poses a significant challenge as a solution which describes an excited state does not typically correspond to a local minimum on an electronic energy surface but rather to a saddle point. As a result, finding such solutions requires a distinctly different approach. Due to the fact that even conventional KS excited state calculations are hindered by the limitations of commonly used self-consistent field procedures, a variational assessment

of the PZ-SIC functional has not yet been carried out.

Thus, an evaluation of PZ-SIC on large molecular systems and solids in its ground and excited states requires the development of new optimization methods and its efficient implementation in modern software. The development of such algorithms and the assessment of these myriad approaches is explored throughout the published papers (Ivanov et al., 2021a,b,c; Levi et al., 2020b) and complementary information is provided in this thesis.

This thesis is organized as follows. The theoretical background for ground state DFT as well as the formal justification of orbital-density-dependent non-unitary invariant functionals are presented in Chapter 2. The direct minimization algorithms for ground state PZ-SIC as well as for regular DFT calculations are given in Chapter 3. An extension of the direct minimization method for excited state calculations within both DFT and PZ-SIC is given in Chapter 4. Finally, conclusions are presented in Chapter 5.

2 Density and Orbital Density Functional Theories

The Hamiltonian of electrons interacting with nuclei within the Born-Oppenheimer approximation is:

$$\hat{H} = \sum_{i=1..N} \hat{v}(\mathbf{r}_i) + \hat{T} + \hat{V}_{ee} \tag{1}$$

where $\hat{T}$ is the kinetic energy operator:

$$\hat{T} = \sum_{i=1}^{N} -\frac{1}{2}\nabla_i^2 \tag{2}$$

$\hat{V}_{ee}$ is the Coulomb electron-electron interaction:

$$\hat{V}_{ee} = \frac{1}{2} \sum_{i,j=1;i\neq j}^{N} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \tag{3}$$

and $\hat{v}(\mathbf{r}_i)$ is the one-body potential which describes the interaction of electrons with the nuclei and possibly also with an external field. When the external field is zero, this potential can be written as follows:

$$\hat{v}(\mathbf{r}_i) = \sum_a \frac{Z_a}{|\mathbf{r}_i - \mathbf{R}_a|} \tag{4}$$

where $\mathbf{R}_a$ and Z_a are a position and a charge of a nucleus.

The ground state electronic energy can be found as a minimal expectation value of the Hamiltonian (1) (hereafter referred to as energy) among all antisymmetric wave functions $\Psi(\mathbf{r}_1, \ldots, \mathbf{r}_N)$ normalized to unity:

$$\hat{E}_{GS} = \min_{\Psi} \langle \Psi | \hat{H} | \Psi \rangle = \langle \Psi_{GS} | \hat{H} | \Psi_{GS} \rangle , \tag{5}$$

At the minimum of the energy, the gradient must equal zero. Taking into account the normality constraints $\langle \Psi | \Psi \rangle = 1$ through the Lagrange multiplier, one can obtain a time-independent Schrödinger equation:

$$\hat{H} | \Psi \rangle = E | \Psi \rangle \tag{6}$$

The lowest eigenvalue of the operator $\hat{H}$ corresponds to the ground state energy. Finding the exact solution of the Schrödinger equation or carrying out the direct minimization of the energy presents an extremely challenging problem for small systems and such problems is practically impossible to solve for a large number of electrons even using modern supercomputers. However, theories which allow one to solve such a challenging problem in an efficient way are described in sections below.

2.1 Hohenberg-Kohn Theorem and Constrained Search Generalization

In 1964, Hohenberg and Kohn proved that in the case of nondegenerate ground state there exists a unique map between the external potential $v(\mathbf{r})$ and one-electron density:

$$\rho(\mathbf{r}_1) = N \int \cdots \int d^3\mathbf{r}_2 \ldots d^3\mathbf{r}_N \, \Psi^*(\mathbf{r}_1,..,\mathbf{r}_N)\Psi(\mathbf{r}_1,..,\mathbf{r}_N) \tag{7}$$

Therefore, the ground state energy is in fact a functional of the one-electron density (Hohenberg and Kohn, 1964). Rather than carrying out the minimization with respect to many-body wave functions, one can search for the energy minimum in the space of one-electron densities:

$$E_{GS} = \min_{\rho} E_v[\rho] = \min_{\rho} \left\{ \int v(\mathbf{r})\rho(\mathbf{r})\, d^3\mathbf{r} + F_{HK}[\rho] \right\} \tag{8}$$

where $F_{HK}[\rho]$ is the universal Hohenberg-Kohn functional which does not depend on the external potential:

$$F_{HK}[\rho] = \langle \Psi_\rho^{HK} | \, \hat{T} + \hat{V}_{ee} \, | \Psi_\rho^{HK} \rangle \, . \tag{9}$$

Generally speaking, there is an infinite number of wave functions which correspond to the same density ρ. However, there can only be one ground state wave function Ψ_ρ^{HK} of a Hamiltonian H' which corresponds to the density ρ. In order to define this wave function, one needs to find the potential v' of the Hamiltonian H' for which such a solution exists. Therefore, in Hohenberg-Kohn theory, it is assumed that the density ρ is v-representable.

Levy (1979) expanded Hohenberg-Kohn theory and conducted a constrained-search generalization of the universal density functional for a large class of densities (N-representable densities) as well as for degenerate ground states. In this approach, the minimization (5) is carried out in two steps: firstly, searching for the optimal wave function for a given density and secondly, performing the minimization of the obtained functional with respect to the density:

$$E_{GS} = \min_{\rho} \min_{\Psi \to \rho} \langle \Psi | \sum_{i=1..N} \hat{v}(\mathbf{r}_i) + \hat{T} + \hat{V}_{ee} \, | \Psi \rangle \tag{10}$$

$$E_{GS} = \min_{\rho} \left\{ \int v(\mathbf{r})\rho(\mathbf{r})\, d^3\mathbf{r} + F[\rho] \right\} \tag{11}$$

where $F[\rho]$ is the universal functional:

$$F[\rho] = \min_{\Psi \to \rho} \langle \Psi | \, \hat{T} + \hat{V}_{ee} \, | \Psi \rangle = \langle \Psi_\rho | \, \hat{T} + \hat{V}_{ee} \, | \Psi_\rho \rangle \, . \tag{12}$$

$F[\rho] = F_{HK}[\rho]$ when the density is v-representable. It can be shown that for the ground state density, the wave function $\Psi_{\rho_{GS}}$ from Eq. (12) is the ground state wave function Ψ_{GS} (Levy, 1979). The number of electrons is a simple functional of the density, $N = \int \rho(\mathbf{r})\, d^3\mathbf{r}$, and therefore, the operators $\hat{T}$, $\hat{V}_{ee}$ are also functionals of the density. Since $\hat{T}$, $\hat{V}_{ee}$ and Ψ_{GS} are defined, the external potential can be found from

the Schrödinger equation, and given that this potential decays to zero at infinity, the ground state energy can be determined as well. Thus, the ground state density defines *all* properties of the many-body electron system.

The constrained-search generalization provides not only the formal justification of the universal density functional on N-representable densities, but it also presents an exact and explicit equation for this functional through Eq. (12). This explicit equation can be used for deriving constraints which the exact functional must satisfy and therefore, Eq. (12) is of great value when developing approximate density functionals.

2.2 Kohn-Sham Density Functional Theory

In Kohn-Sham density functional theory (Kohn and Sham, 1965), the energy is separated into the following terms:

$$E[\rho] = T_s[\rho] + \int d^3\mathbf{r}\rho(\mathbf{r})v(\mathbf{r}) + U[\rho] + E_{xc}[\rho] \tag{13}$$

where T_s is the kinetic energy of independent electrons which have the same density ρ as the interacting electrons. U is the Hartree energy:

$$U[\rho] = \frac{1}{2}\iint d^3\mathbf{r}\, d^3\mathbf{r}' \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \tag{14}$$

and the exchange-correlation energy, E_{xc}, is defined as follows:

$$E_{xc} = F - U - T_s \tag{15}$$

The Euler-Lagrange equation for such partitioning of the functional (13) is:

$$\frac{\delta T_s[\rho]}{\delta\rho} + v_H + v_{xc} + v = 0, \tag{16}$$

Since the Hohenberg-Kohn theorem is also valid for non-interacting electrons moving in an external potential v_s, the energy of the non-interacting electrons is a functional of density and the Euler-Lagrange equation for such a system can be written as follows:

$$\frac{\delta T_s[\rho]}{\delta\rho} + v_s = 0, \tag{17}$$

In order to ensure that the ground state density of the non-interacting electrons is the same as that of the true electron system, one can choose $v_s = v_H + v_{xc} + v$ so that the solutions of Eq. (16) and Eq. (17) coincide. In order to obtain single-particle equations, which are easier to solve than the true many-body problem, it is assumed that the ground state density can be further obtained from orbitals which satisfy the single-particle Schrödinger equations with local multiplicative potential. Such equations also describe the non-interacting electrons and therefore, the self-consistent equations are:

$$\left\{-\frac{1}{2}\Delta + v_s\right\}\phi_i = \varepsilon_i\phi_i, \quad \text{with } v_s = v_H + v_{xc} + v. \tag{18}$$

The kinetic energy of the free electrons and the total ground state density can be calculated using orbitals (18):

$$T_s = \sum_{i=1}^{\infty} f_i \int d^3\mathbf{r} \frac{|\nabla \phi_i|^2}{2} \tag{19}$$

and:

$$\rho(\mathbf{r}) = \sum_{i=1}^{\infty} f_i |\phi_i|^2, \tag{20}$$

where f_i are occupation numbers $0 \leq f_i \leq 2$ and $\sum_i f_i = N$. We should note here that the solution of Eq. (18) does not correspond to the stationary point of the ground state functional for *any* distribution of occupation numbers (Perdew and Levy, 1985). Some stationary points of Eq. (18) can represent excited states when occupation numbers are not distributed according to aufbau principle as will be discussed in Chapter 4. In this case, an excited state density may not be a stationary density of the ground state functional (Perdew and Levy, 1985).

2.3 Generalized Kohn-Sham Density Functional Theory

The central part of Kohn-Sham theory is the mapping of the true many-body electron system described by the wave function Ψ into an auxiliary system of non-interacting electrons moving in a local potential and described by a single Slater determinant $\Phi = |\phi_1, \phi_2, \ldots, \phi_N|$ which gives the same density as Ψ. However, one can also map the true electron system into an auxiliary system where the electron interaction is partially and explicitly taken into account but the system is still described by a single Slater determinant Φ. Such a generalization of the Kohn-Sham scheme is called generalized Kohn-Sham (GKS) density functional theory as was proposed by Seidl et al. (1996).

Let us define the functional of the auxiliary electron system moving in local potential $v_s(\mathbf{r})$:

$$E^s = F^s[\rho] + \int d\mathbf{r} v_s(\mathbf{r}) \rho(\mathbf{r}) \tag{21}$$

and:

$$F^s[\rho] = \min_{\Phi \to \rho} S[\Phi] \tag{22}$$

where the functional $S[\Phi]$ depends on a single Slater determinant and may include part of the electron interaction. For example:

$$S[\Phi] = \langle \Phi | \hat{T} + \hat{V}_{ee} | \Phi \rangle = T_s[\Phi] + U[\Phi] + E_x[\Phi] \tag{23}$$

Now let us define $R[\rho]$ as the functional which describes the 'missing' interaction in $F^s[\rho]$ that is:

$$R[\rho] = F[\rho] - F^s[\rho] \tag{24}$$

Given the definition of the ground state energy (11), one can obtain:

$$\begin{aligned} E_{GS} = & \min_{\rho} \left\{ \int v(\mathbf{r})\rho(\mathbf{r})\, d\mathbf{r} + F[\rho] \right\} = \\ & \min_{\rho} \left\{ \int v(\mathbf{r})\rho(\mathbf{r})\, d\mathbf{r} + F^{s}[\rho] + R[\rho] \right\} = \\ & \min_{\rho} \left\{ \int v(\mathbf{r})\rho(\mathbf{r})\, d\mathbf{r} + \min_{\Phi\to\rho} S[\Phi] + R[\rho] \right\} = \\ & \min_{\rho} \min_{\Phi\to\rho} \left\{ \int v(\mathbf{r})\rho(\mathbf{r})\, d\mathbf{r} + S[\Phi] + R[\rho] \right\} = \\ & \min_{\Phi} \left\{ \int v(\mathbf{r})\rho[\Phi](\mathbf{r})\, d\mathbf{r} + S[\Phi] + R[\rho[\Phi]] \right\} \end{aligned} \tag{25}$$

Thus, the ground state energy can be obtained through minimization with respect to the single Slater determinant Φ:

$$E_{GS} = \min_{\Phi} \left\{ \int v(\mathbf{r})\rho[\Phi](\mathbf{r})\, d\mathbf{r} + S[\Phi] + R[\rho[\Phi]] \right\} \tag{26}$$

Setting the first variational derivative to zero and taking the orthonormality constraint of orbitals through Lagrange multipliers into account, one can obtain self-consistent equations in canonical representation given that the functional is unitary invariant:

$$\left\{ \hat{O}[\{\phi_i\}] + v_R(\mathbf{r}) + v(\mathbf{r}) \right\} \phi_i(\mathbf{r}) = \varepsilon_i \phi(\mathbf{r}) \tag{27}$$

where $\hat{O}[\{\phi_i\}]$ is a non-local operator which depends on orbitals $\{\phi_i\}$ and is defined as:

$$\frac{\delta S}{\delta \phi_i^*(\mathbf{r})} = \hat{O}[\{\phi_i\}]\phi_i(\mathbf{r}) \tag{28}$$

$v_R(\mathbf{r}) = \delta R/\delta\rho$ – local operator. Thus, the external potential from Eq. (21) equals:

$$v_s = v_R(\mathbf{r}) + v(\mathbf{r}) \tag{29}$$

The generalized Kohn-Sham scheme provides the theoretical background for hybrid functionals where part of exchange energy is treated through exact exchange in functional S and the rest of the exchange energy is approximated with the semi-local functional of total density ρ in R.

2.4 Orbital-Density-Dependent Functional Theory

One of the reasons for the failures of approximate density functionals is often attributed to the self-interaction error. Perdew and Zunger proposed orbital-by-orbital corrections to the Hartree and the exchange correlation energies which make any approximate functional self-interaction free in the one electron limit (Perdew and Zunger, 1981). As a result of this procedure, the functional becomes orbital-density-dependent lacking

unitary invariance and such a formulation does not fall into either Kohn-Sham or generalized Kohn-Sham theories. Here we propose a scheme similar to generalized Kohn-Sham theory in order to provide a more rigorous theoretical basis for such an approach.

In Hartree theory, the electrons interact with each other only through classical electrostatic potential and in such a theory they do not interact with themselves. The Hartree functional can be written as:

$$E_H = \int d^3\mathbf{r}\, v(\mathbf{r})\rho(\mathbf{r}) - \frac{1}{2}\sum_i \int d^3\mathbf{r}\, \xi_i^*(\mathbf{r})\Delta\xi_i(\mathbf{r}) + \frac{1}{2}\sum_{i\neq j}\iint d^3\mathbf{r}\, d^3\mathbf{r}'\, \frac{\rho_i(\mathbf{r})\rho_j(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \quad (30)$$

or:

$$E_H = \langle\Xi| \sum_{i=1..N} \hat{v}(\mathbf{r}_i) + \hat{T} + \hat{V}_{ee}|\Xi\rangle \quad (31)$$

where $|\Xi\rangle$ is a Hartree wave function:

$$\Xi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = \xi_1(\mathbf{r}_1)\xi_2(\mathbf{r}_2)\ldots\xi_N(\mathbf{r}_N) \quad (32)$$

and $\rho_i(\mathbf{r}) = |\xi_i(\mathbf{r})|^2$. Note that the electron interaction in Eqs. (30) can be written as:

$$\sum_{i\neq j}\iint d^3\mathbf{r}\, d^3\mathbf{r}'\, \frac{\rho_i(\mathbf{r})\rho_j(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} = \iint d^3\mathbf{r}\, d^3\mathbf{r}'\, \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} - \sum_{i=1..N}\iint d^3\mathbf{r}\, d^3\mathbf{r}'\, \frac{\rho_i(\mathbf{r})\rho_i(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \quad (33)$$

and therefore, in Eqs. (30) and (31) the electrostatic interaction U is self-interaction corrected by subtraction of orbital self-interaction.

We can define a functional $L[\Xi]$ which depends not on a single Slater determinant of orbitals but rather on Hartree wave function and one can further map the true many-body problem into an auxiliary partially interacting electron system described by the Hartree wave function with the same density as the true ground state density. Here, we *assume* similar to the KS or GKS schemes that such mapping exists. Let us define the functional:

$$F_H[\rho] = \min_{\Xi\to\rho} L[\Xi] \quad (34)$$

and:

$$J[\rho] = F[\rho] - F_H[\rho] \quad (35)$$

Similar to GKS we can express the ground state energy as:

$$E_{GS} = \min_{\Xi}\left\{\int v(\mathbf{r})\rho[\Xi](\mathbf{r})\, d^3\mathbf{r} + L[\Xi] + J[\rho[\Xi]]\right\} \quad (36)$$

Now if $L = E_H$ from Eq. (31), we obtain:

$$\begin{aligned} E[\{\rho_i\}] = &\int v(\mathbf{r})\rho(\mathbf{r})\, d^3\mathbf{r} - \frac{1}{2}\sum_i \int d^3\mathbf{r}\, \xi_i^*(\mathbf{r})\Delta\xi_i(\mathbf{r}) + \\ &\frac{1}{2}\iint d^3\mathbf{r}\, d^3\mathbf{r}'\, \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} - \frac{1}{2}\sum_{i=1..N}\iint d^3\mathbf{r}\, d^3\mathbf{r}'\, \frac{\rho_i(\mathbf{r})\rho_i(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + J[\rho] \end{aligned} \quad (37)$$

Note that the Hartree term in energy (37) is already self-interaction free. The functional (37) is not unitary invariant due to its explicit dependence on orbital densities. This is an artefact due to the Hartree wave function which is not always anti-symmetric.

In order to obtain a Perdew-Zunger self-interaction corrected functional, let us choose:

$$\hat{L} = \hat{V}_{ee} - \sum_{i=1..N} \hat{\mathscr{E}}_{xc}(\mathbf{r}_i) \tag{38}$$

where $\hat{\mathscr{E}}_{xc}(\mathbf{r}_i)$ is a one-body non-linear exchange-correlation operator. For example, if the exchange-correlation energy is approximated using a semi-local functional:

$$E_{xc}[\rho] = \int d^3\mathbf{r}\rho(\mathbf{r})\varepsilon(\rho(\mathbf{r})) \tag{39}$$

where ε is the energy density, then the action of $\hat{\mathscr{E}}_{xc}$ can be defined as

$$\langle \mathbf{r}|\hat{\mathscr{E}}_{xc}|\xi_i\rangle = \varepsilon(\rho_i(\mathbf{r}))\xi_i(\mathbf{r}) \tag{40}$$

so that

$$\langle \xi_i|\hat{\mathscr{E}}_{xc}|\xi_i\rangle = E_{xc}[\rho_i] \tag{41}$$

For the exact $\hat{\mathscr{E}}_{xc}$ and the Hartree wave function, the one-body contributions from $\sum_{i=1..N}\hat{\mathscr{E}}_{xc}(\mathbf{r}_i)$ should exactly be cancelled by self-Hartree energy and therefore, $J[\rho] = E_{xc}[\rho]$. Finally, substituting L, J and Ξ into Eq. (36), we arrive at the PZ-SIC functional:

$$E^{PZ-SIC}[\rho_1,\ldots,\rho_N] = \int v(\mathbf{r})\rho(\mathbf{r})\,d^3\mathbf{r} - \frac{1}{2}\sum_i \int d^3\mathbf{r}\,\xi_i^*(\mathbf{r})\Delta\xi_i(\mathbf{r}) + \frac{1}{2}\iint d^3\mathbf{r}\,d^3\mathbf{r}'\frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} - \sum_{i=1..N}\left(\frac{1}{2}\iint d^3\mathbf{r}\,d^3\mathbf{r}'\frac{\rho_i(\mathbf{r})\rho_i(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + E_{xc}[\rho_i]\right) + E_{xc}[\rho] \tag{42}$$

3 Ground-State SIC-DFT Calculations: Direct Minimization Approach

In order to find the orbitals that provide the ground state density of the exact functional, one has to find the solution of the KS self-consistent equations (18). To solve these equations, one can use iterative algorithms which are usually separated into two steps. In the first step, for a fixed density and for a fixed KS potential, one finds the canonical orbitals and the eigenvalues using an iterative algorithm such as the Davidson algorithm (Davidson, 1975) or using the direct diagonalization of the Hamiltonian matrix when the size of the basis set representing the orbitals is not too large as in the case of linear combination of atomic orbitals (LCAO) basis sets. In the second step, one uses extrapolation techniques such as Pulay mixing (Pulay, 1980; Garza and Scuseria, 2012) - which is also known as direct inversion in iterative subspace (DIIS) algorithm - or Broyden mixing (Furthmüller and Kresse, 1996) in order to increase robustness and accelerate convergence towards the solution. These methods give the optimal density (or Hamiltonian matrix) which is spanned in a subspace of densities (or Hamiltonian matrices) that can be obtained from the orbitals obtained in the previous steps. While such an approach is the one most commonly used today, it often suffers from convergence issues, for example, for systems with degenerate HOMO-LUMO orbitals or systems containing transition metal atoms. Furthermore, the user of software developed for KS-DFT calculations is often presented with a large choice of density-mixing parameters which cannot be used universally in all systems. These algorithms cannot be applied to non-unitary invariant functionals as the canonical orbitals that diagonalize the Hamiltonian matrix are not necessarily optimal orbitals, i.e. providing the minimum of the energy. Then, a different approach is required.

An alternative method for solving self-consistent equations (18) is to find the minimum of the KS or PZ-SIC energy functional taking into account the orthonormality constraints of the orbitals using a direct minimization technique with efficient quasi-Newton algorithms (Nocedal and Wright, 2006). This approach will be described below for both the KS and PZ-SIC functionals.

3.1 General considerations for the minimization of energy functionals

Let E be the functional of interest which depends on a set of orthonormal orbitals $\{|\psi_i\rangle\}$. Optimal orbitals corresponding to the minimum of the energy functional E must satisfy

the Euler-Lagrange equations that can be written as follows:

$$\hat{h}_i |\psi_i\rangle = \sum_j |\psi_j\rangle \lambda_{ji}, \tag{43}$$

where $\hat{h}_i$ is a single-particle Hamiltonian defined from the equation:

$$\hat{h}_i |\psi_i\rangle = \frac{\partial E}{\partial \langle\psi_i|} \tag{44}$$

For example, for KS-DFT, the single-particle Hamiltonian is:

$$\hat{h}_i = f_i \left[-\frac{1}{2}\nabla^2 + \hat{v}_{ext}(\rho) + \hat{v}_H(\rho) + \hat{v}_{xc}(\rho) \right] = f_i \hat{h} \tag{45}$$

and for PZ-SIC it is:

$$\hat{h}_i = f_i \left[-\frac{1}{2}\nabla^2 + \hat{v}_{ext}(\rho) + \hat{v}_H(\rho) + \hat{v}_{xc}(\rho) - \hat{v}_H(\rho_i) - \hat{v}_{xc}(\rho_i) \right] = f_i(\hat{h} + \hat{v}_i) \tag{46}$$

When the optimal orbitals are found, the Lagrange matrix $\{\lambda_{ij}\}$ may not be diagonal in equation (43) which is unlike the canonical representation of Eq. (18). For density functionals, the canonical and optimal representations are equivalent – that is, they both correspond to the same density and energy. Conversely, the canonical orbitals may not correspond to an extremum at all in the case of the PZ-SIC functional. In order to demonstrate this, let us consider the infinitesimal rotation of orbitals in the vicinity of the energy minimum:

$$|\phi_i\rangle = \sum_j \left(e^{-A}\right)_{ij} |\psi_j\rangle, \tag{47}$$

where A is a skew-hermitian matrix, $A^\dagger = -A$. The derivative of energy with respect to the matrix elements $a_{ij} = (A)_{ij}$ at $A = 0$ is (see Appendix A for details)

$$\left.\frac{\partial E}{\partial a_{ij}}\right|_{A=0} = \langle\psi_i|\hat{h}_j|\psi_j\rangle - \langle\psi_i|\hat{h}_i|\psi_j\rangle = \lambda_{ij} - \lambda_{ji}^* = 0 \tag{48}$$

From equation (48) follows that the Lagrange matrix must be Hermitian. By substituting Eq. (45) into Eq. (48) one obtains:

$$(f_j - f_i)\,\langle\psi_i|\hat{h}|\psi_j\rangle = 0 \tag{49}$$

Thus, for KS-DFT calculations, a matrix element of Hamiltonian must be zero if it corresponds to overlap between orbitals with different occupation numbers $f_j \neq f_i$,

$$\langle\psi_i|\hat{h}|\psi_j\rangle = 0, \quad \text{when } f_j \neq f_i \tag{50}$$

For a subspace of equally occupied orbitals, one can consider any linear combination of orbitals spanned in this subspace and, as a result, they can be chosen as those which diagonalize this part of the Hamiltonian. Therefore, the solution for Eq. (43) is equivalent to the solution for Eq. (18) for KS-DFT.

This is not the case, however, for PZ-SIC due to the orbital-density dependence in the one-electron Hamiltonian. Indeed, by substituting Eq. (46) into Eq. (48), one can obtain:

$$(f_j - f_i)\,\langle\psi_i|\hat{h}|\psi_j\rangle + \langle\psi_i|f_j\hat{v}_j - f_i\hat{v}_i|\psi_j\rangle = 0 \tag{51}$$

Thus, even orbitals with the same occupancies have a uniquely defined amplitudes ($|\psi_i|$) and must satisfy the localization (Pederson) equation (Pederson et al., 1985):

$$\langle\psi_i|\hat{v}_j - \hat{v}_i|\psi_j\rangle = 0 \tag{52}$$

and, as a result, optimal orbitals cannot be chosen as canonical orbitals.

Further details of the implementation of direct minimization depends on the representation of the wave-function and are briefly outlined below.

3.2 Minimization Algorithm for a Linear Combination of Atomic Orbitals (LCAO) Basis Set

The wave functions are expanded in a series of M non-orthogonal atomic basis functions:

$$|\psi_i\rangle = \sum_{\mu=1..M} O_{\mu i}\,|\chi_\mu\rangle\,. \tag{53}$$

The orthonormal condition of wave functions leads to the generalized orthonormality constraints imposed on the expansion coefficients:

$$O^\dagger S O = I \tag{54}$$

where:

$$S_{\mu\nu} = \langle\chi_\mu|\chi_\nu\rangle \tag{55}$$

In order to avoid a constrained energy minimization with respect to the coefficients O, they can be parametrized in the following way:

$$O = Ce^A \tag{56}$$

where C is a coefficient matrix of auxiliary orbitals which represent an initial guess for the wave function and A is a skew-hermitian matrix, $A = -A^\dagger$. When the reference orbitals are fixed, the energy can be considered as a function defined on a linear space of skew-hermitian matrices. This is a major simplification of the problem as the minimization of a function defined in a linear space is easier, and efficient quasi-Newton algorithms with an inexact line search have already been developed for such problems (Nocedal and Wright, 2006). In order to perform the minimization efficiently, one needs to calculate the gradient of the energy with respect to A (see Appendix A):

$$\frac{\partial E}{\partial a_{ji}} = \left(\int_0^1 e^{tA} L e^{-tA}\,dt\right)_{ij} = L_{ji} + o(A) \tag{57}$$

where:

$$L_{ij} = -(f_j - f_i)H_{ij} - (f_j V_{ij} - f_i V_{ji}^*), \tag{58}$$

$$H_{ij} = \sum_{\mu\nu} O^*_{\mu i} H_{\mu\nu} O_{\nu j}, \tag{59}$$

$$V_{ij} = \sum_{\mu\nu} O^*_{\mu i} V^j_{\mu\nu} O_{\nu j}, \tag{60}$$

$$H_{\mu\nu} = \langle \chi_\mu | -\frac{1}{2}\nabla^2 + v_s | \chi_\nu \rangle \tag{61}$$

$$V^j_{\mu\nu} = \langle \chi_\mu | v_H(\rho_j) + v_{xc}(\rho_j) | \chi_\nu \rangle \tag{62}$$

This gradient expression is used in combination with the limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm with the following preconditioning (Ivanov et al., 2021b):

$$P^{(k)}_{ij} = \frac{1}{-2(1-\gamma)(\varepsilon_i - \varepsilon_j)(f_i - f_j) + \gamma\beta^{(k)}_{ij}}, \tag{63}$$

where k is the iteration number, the parameter $\beta^{(k)}_{ij}$ is a scaling factor recommended when using an approximate initial Hessian in the L-BFGS algorithm (Nocedal and Wright, 2006):

$$\beta^{(k)}_{ij} = \frac{\|\nabla_{\vec{a}} E(\vec{a}^{\,(k)}) - \nabla_{\vec{a}} E(\vec{a}^{\,(k-1)})\|^2}{(\vec{a}^{\,(k)} - \vec{a}^{\,(k-1)}) \cdot (\nabla_{\vec{a}} E(\vec{a}^{\,(k)}) - \nabla_{\vec{a}} E(\vec{a}^{\,(k-1)}))} \tag{64}$$

and $\gamma = 0.25$. For $\gamma = 0$, this preconditioner can be obtained by taking the derivative of a linear expansion of the energy gradient with respect to the skew-hermitian matrix, and by neglecting the derivative of the KS potential with respect to density (Appendix B). The details of the implementation and its application to KS-DFT are given in Ref. (Ivanov et al., 2021b). It was shown that the proposed direct minimization method has the ability to outperform the conventional SCF procedure based on the DIIS algorithm and, at the same time, provides more robust convergence. Application of the developed methodology to PZ-SIC calculations of Mn dimer are presented in Ref. (Ivanov et al., 2021a) and shortly described in section 3.5.

3.3 Minimization Algorithm for Real Space and Plane Wave Representations

The transformation of orbitals according to equation (56) requires calculating the $M \times M$ exponential matrix, where M is the number of basis functions. For finite-difference (FD) real space grid (RSG) or plane wave (PW) representations, the number of basis functions can easily reach 10^6 and the algorithm presented for LCAO calculations in the previous section is, therefore, not applicable here. A different approach is required which is described below.

Let $\mathbf{\Psi} = (|\psi_1\rangle, .., |\psi_M\rangle)^T$, and M be the number of orbitals, and

$$\mathscr{M} = \{\mathbf{\Psi} : \langle \psi_i | \psi_j \rangle = \delta_{ij}, i \neq j = 1 \ldots M\} \tag{65}$$

be a manifold spanned by a set of orthonormal wave functions $\mathbf{\Psi}$. Let $\hat{R}$ be an operator which transforms any wave functions to a set of orthonormal wave functions. Then the tangent space to $\mathscr{M}$ at $\mathbf{\Psi}$ is defined as:

$$\boldsymbol{V}_{\mathbf{\Psi}}(\boldsymbol{G}) = \lim_{\varepsilon \to 0} \frac{\hat{R}[\mathbf{\Psi} + \varepsilon \boldsymbol{G}] - \mathbf{\Psi}}{\varepsilon} = \hat{R}'_{\varepsilon}[\mathbf{\Psi} + \varepsilon \boldsymbol{G}]\big|_{\varepsilon=0} \tag{66}$$

The operator $\hat{R}$ can be chosen so that the orthonormalization procedure corresponds to the symmetric Löwdin transformation. Let $S_{\boldsymbol{X},\boldsymbol{Y}}$ be the overlap matrix between the two vectors $\boldsymbol{X},\boldsymbol{Y}$ in the Hilbert space. Then:

$$\begin{aligned} &S_{\mathbf{\Psi}+\varepsilon\boldsymbol{G},\mathbf{\Psi}+\varepsilon\boldsymbol{G}} = \int d\mathbf{r}\,(\mathbf{\Psi}+\varepsilon\boldsymbol{G})(\mathbf{\Psi}+\varepsilon\boldsymbol{G})^{\dagger} = \\ &\int d\mathbf{r}\left[\mathbf{\Psi}\mathbf{\Psi}^{\dagger} + \varepsilon\left(\mathbf{\Psi}\boldsymbol{G}^{\dagger} + \boldsymbol{G}\mathbf{\Psi}^{\dagger}\right) + \varepsilon^2 \boldsymbol{G}\boldsymbol{G}^{\dagger}\right] = \\ &I + \varepsilon\left(S_{\mathbf{\Psi},\boldsymbol{G}} + S_{\boldsymbol{G},\mathbf{\Psi}}\right) + \varepsilon^2 S_{\boldsymbol{G},\boldsymbol{G}} = I + \varepsilon W_{\mathbf{\Psi},\boldsymbol{G}} + \varepsilon^2 S_{\boldsymbol{G},\boldsymbol{G}} \end{aligned} \tag{67}$$

where $W_{\mathbf{\Psi},\boldsymbol{G}} = S_{\mathbf{\Psi},\boldsymbol{G}} + S_{\boldsymbol{G},\mathbf{\Psi}}$. Therefore, $S^{-1/2}$ can be approximated as:

$$S^{-1/2}_{\mathbf{\Psi}+\varepsilon\boldsymbol{G},\mathbf{\Psi}+\varepsilon\boldsymbol{G}} = I - \frac{\varepsilon}{2} W_{\mathbf{\Psi},\boldsymbol{G}} + o(\varepsilon) \tag{68}$$

If $\hat{R}[\mathbf{\Psi}+\varepsilon\boldsymbol{G}] = S^{-1/2}_{\mathbf{\Psi}+\varepsilon\boldsymbol{G},\mathbf{\Psi}+\varepsilon\boldsymbol{G}}(\mathbf{\Psi}+\varepsilon\boldsymbol{G})$ and then the tangent space is defined as follows:

$$\boldsymbol{V}_{\mathbf{\Psi}}(\boldsymbol{G}) = \boldsymbol{G} - \frac{1}{2} W_{\mathbf{\Psi},\boldsymbol{G}}\mathbf{\Psi} \tag{69}$$

Note that for wave functions which belong to the tangent space, the hermitian part of the overlap matrix is zero, $W_{\mathbf{\Psi},\boldsymbol{V}} = 0$, and therefore, the equation $W_{\mathbf{\Psi},\boldsymbol{V}} = 0$ is a constraint on the tangent vectors. When the displacement of Ψ belongs to the tangent space, the $S^{-1/2}$ can further be approximated with higher-order terms as:

$$S^{-1/2}_{\mathbf{\Phi}+\varepsilon\boldsymbol{V},\mathbf{\Phi}+\varepsilon\boldsymbol{V}} = I - \frac{\varepsilon^2}{2} S_{\boldsymbol{V},\boldsymbol{V}} + o(\varepsilon^2) \tag{70}$$

and therefore:

$$\mathbf{\Phi} := \hat{R}[\mathbf{\Psi}+\varepsilon\boldsymbol{V}] = \mathbf{\Psi} + \varepsilon\boldsymbol{V} - \frac{\varepsilon^2}{2} S_{\boldsymbol{V},\boldsymbol{V}}\mathbf{\Psi} + o(\varepsilon^2) \tag{71}$$

Thus, when the gradient of energy, $\boldsymbol{G}$, is calculated, it should be projected on the tangent space to $\mathbf{\Psi}$ using the equation (69). Then, the wave functions, $\mathbf{\Psi}$, are modified according to (71) with $\varepsilon < 0$ in the opposite direction to the energy gradient. This procedure constitutes the steepest descent algorithm (SDA). However, SDA exhibits slow convergence especially near the energy minimum when the gradient is small and, as a result, it requires many evaluations of the energy. A more efficient approach can be emplyoed in which the energy is approximated with a quadratic function in the tangent space at each minimization step. This approach is known as the Newton-Raphson algorithm. In order to simplify the analysis and demonstrate how a quadratic

approximation to the energy in curved space can be constructed, let us consider free electrons as a model system:

$$E_s = \int d\mathbf{r}\, \mathbf{\Phi}^\dagger \left(\hat{\boldsymbol{T}} + \hat{\boldsymbol{v}}_s(\mathbf{r})\right) \mathbf{\Phi}, \tag{72}$$

where:

$$(\hat{\boldsymbol{T}} + \hat{\boldsymbol{v}}_s)_{ij} = \left(\hat{T} + v_s(\mathbf{r})\right) \delta_{ij} \tag{73}$$

Using Eq. (71) with $\varepsilon = 1$, one can obtain:

$$\begin{aligned} E_s = &\int d\mathbf{r}\, \mathbf{\Phi}^\dagger \left(\hat{\boldsymbol{T}} + \hat{\boldsymbol{v}}_s\right) \mathbf{\Phi} \approx \\ &\int d\mathbf{r}\, \mathbf{\Psi}^\dagger \left(\hat{\boldsymbol{T}} + \hat{\boldsymbol{v}}_s\right) \mathbf{\Psi} + \int d\mathbf{r}\, \boldsymbol{V}^\dagger \left(\hat{\boldsymbol{T}} + \hat{\boldsymbol{v}}_s\right) \mathbf{\Psi} + \int d\mathbf{r}\, \mathbf{\Psi}^\dagger \left(\hat{\boldsymbol{T}} + \hat{\boldsymbol{v}}_s\right) \boldsymbol{V} + \\ &+ \int d\mathbf{r}\, \boldsymbol{V}^\dagger \left(\hat{\boldsymbol{T}} + \hat{\boldsymbol{v}}_s\right) \boldsymbol{V} - \\ &\frac{1}{2} \int d\mathbf{r}\, \mathbf{\Psi}^\dagger \left(\hat{\boldsymbol{T}} + \hat{\boldsymbol{v}}_s\right) S_{\boldsymbol{V},\boldsymbol{V}} \mathbf{\Psi} + \mathbf{\Psi}^\dagger S^\dagger_{\boldsymbol{V},\boldsymbol{V}} \left(\hat{\boldsymbol{T}} + \hat{\boldsymbol{v}}_s\right) \mathbf{\Psi} + o(\varepsilon^2) \end{aligned} \tag{74}$$

The last term can be rewritten as:

$$\frac{1}{2} \int d\mathbf{r}\, \mathbf{\Psi}^\dagger \left(\hat{\boldsymbol{T}} + \hat{\boldsymbol{v}}_s\right) S_{\boldsymbol{V},\boldsymbol{V}} \mathbf{\Psi} + \mathbf{\Psi}^\dagger S^\dagger_{\boldsymbol{V},\boldsymbol{V}} \left(\hat{\boldsymbol{T}} + \hat{\boldsymbol{v}}_s\right) \mathbf{\Psi} = -\int d\mathbf{r}\, \boldsymbol{V}^\dagger B_s \boldsymbol{V} \tag{75}$$

with B_s being:

$$B_s = -\frac{1}{2} W_{\left(\hat{\boldsymbol{T}} + \hat{\boldsymbol{v}}_s\right)\mathbf{\Psi},\mathbf{\Psi}} \tag{76}$$

Thus we obtain a quadratic approximation to the energy at $\mathbf{\Psi}^\dagger$:

$$E_s[\boldsymbol{V}] \approx E_s[\mathbf{\Psi}] + \int d\mathbf{r}\, \boldsymbol{V}^\dagger \frac{\delta E_s[\mathbf{\Psi}]}{\delta \mathbf{\Psi}^\dagger} + \int d\mathbf{r}\, \frac{\delta E_s[\mathbf{\Psi}]}{\delta \mathbf{\Psi}} \boldsymbol{V} + \int d\mathbf{r}\, \boldsymbol{V}^\dagger \hat{\boldsymbol{H}} \boldsymbol{V} \tag{77}$$

with:

$$\hat{\boldsymbol{H}}_s = \hat{\boldsymbol{T}} + \hat{\boldsymbol{v}}_s + B_s \tag{78}$$

$\hat{\boldsymbol{H}}_s$ is the Hessian of energy and the matrix B_s takes into account the curvature of the manifold. At first glance, the approximation of the energy in the tangent space did not simplify the problem. However, the advantage is that the constraints imposed on the tangent vectors are *linear* constraints:

$$W_{\mathbf{\Psi},\boldsymbol{V}} = 0 \tag{79}$$

and because of this, the Lagrange multipliers can be found exactly through:

$$\Lambda = -\frac{1}{2} \int d\mathbf{r}\, \frac{\delta E_s}{\delta \boldsymbol{V}^\dagger} \mathbf{\Psi}^\dagger - \frac{1}{2} \int d\mathbf{r}\, \mathbf{\Psi} \frac{\delta E_s}{\delta \boldsymbol{V}} = -\frac{1}{2} W_{\mathbf{\Psi},\, \frac{\delta E_s}{\delta \mathbf{\Psi}} + \hat{\boldsymbol{H}} \boldsymbol{V}} \tag{80}$$

Now, taking the derivative of

$$L_s = E_s + \mathrm{Tr}[\Lambda W_{\mathbf{\Psi},\boldsymbol{V}}] \tag{81}$$

with respect to $\boldsymbol{V}^{\dagger}$ and setting it to zero, one obtains:

$$\frac{\delta E_s[\boldsymbol{\Psi}]}{\delta \boldsymbol{\Psi}^{\dagger}} + \hat{\boldsymbol{H}}_s \boldsymbol{V} + \Lambda \boldsymbol{\Psi} = 0 \tag{82}$$

From Eq. (82), it follows that the search direction for new wave functions is:

$$\boldsymbol{V} = -\hat{\boldsymbol{H}}_s^{-1} \left(\frac{\delta E_s[\boldsymbol{\Psi}]}{\delta \boldsymbol{\Psi}^{\dagger}} + \Lambda \boldsymbol{\Psi} \right) \tag{83}$$

For interacting electrons, the above consideration is also valid but $\hat{\boldsymbol{H}}_s$ should be replaced with the projected Hessian of the interacting electrons and E_s should be replaced with the KS energy functional. A calculation of the Hessian matrix usually requires large computational effort and an approximation to the Hessian matrix is instead constructed iteratively at each step (quasi-Newton algorithms). There, the energy gradient and wave functions from previous steps are used and $\hat{\boldsymbol{H}}_s$ is employed as an initial approximation to the true Hessian. Alternatively, one can employ $\hat{\boldsymbol{H}}_s$ as a preconditioner to the conjugate gradients and steepest descent algorithms. In practice, $\hat{\boldsymbol{H}}_s$ is usually approximated as the kinetic energy operator (Briggs et al., 1995; Kresse and Furthmüller, 1996). Thus, the iterative process in wave function minimization using quasi-Newton methods is:

$$\boldsymbol{V}^{(k)} = -\hat{\boldsymbol{H}}^{-1\,(k)} \left(\frac{\delta E[\boldsymbol{\Psi}^{(k)}]}{\delta \boldsymbol{\Psi}^{(k)\dagger}} + \Lambda^{(k)} \boldsymbol{\Psi}^{(k)} \right) \tag{84}$$

$$\boldsymbol{\Psi}^{(k+1)} = \boldsymbol{\Psi}^{(k)} + \alpha \boldsymbol{V}^{(k)} \tag{85}$$

Further practical simplification is made for Lagrange multipliers: they are calculated using Eq. (80) at $\boldsymbol{V} = 0$. In this case, the energy gradient is simply projected on to the tangent space at $\boldsymbol{\Psi}$ using Eq. (69):

$$\frac{\delta E[\boldsymbol{\Psi}^{(k)}]}{\delta \boldsymbol{\Psi}^{(k)\dagger}} + \Lambda^{(k)} \boldsymbol{\Psi}^{(k)} = \boldsymbol{V}_{\boldsymbol{\Psi}^{(k)}} \left(\frac{\delta E[\boldsymbol{\Psi}^{(k)}]}{\delta \boldsymbol{\Psi}^{(k)\dagger}} \right) \tag{86}$$

This simplification leads to the fact that the wave functions will not be orthonormal after each step of the minimization algorithm and therefore, one needs to employ an explicit orthonormalization procedure. One can also add higher order terms to the iteration process using the following equation:

$$\boldsymbol{\Psi}^{(k+1)} = \boldsymbol{\Psi}^{(k)} + \alpha \boldsymbol{V}^{(k)} - \frac{\alpha^2}{2} S[\boldsymbol{V}^{(k)}, \boldsymbol{V}^{(k)}] \boldsymbol{\Psi}^{(k)} \tag{87}$$

which violates the orthonormality constraints less than Eq. (85) for sufficiently small $\alpha^{(k)}$. However, we found that even usage of the first order term in α with a following application of the orthonormalization procedure provides a robust convergence.

The energy gradient for Kohn-Sham energy functionals and PZ-SIC are defined from equations (44), (45) and (46).

3.4 Preconditioning for the Perdew-Zunger Self-Interaction Correction

The PZ-SIC functional includes terms that depend on the orbital densities and the functional is, therefore, not invariant with respect to unitary transformations among equally occupied orbitals. As a result, it depends on a larger number of degrees of freedom as compared to Kohn-Sham functionals and this, in turn, leads to a larger number of iterations needed to achieve convergence. In order to improve the convergence, an additional *inner* loop among occupied orbitals is included in the minimization of the PZ-SIC functional (Klüpfel et al., 2012a; Lehtola and Jónsson, 2014; Borghi et al., 2015). This is achieved by an explicit unitary transformation that applies to occupied orbitals only:

$$|\phi_i\rangle = \sum_{ij} U_{ij} |\psi_j\rangle \tag{88}$$

In this case, the PZ-SIC functional can be considered to be a functional of both the unitary matrix U and the orbitals $|\psi_j\rangle$:

$$E^{SIC} = E^{SIC}[U, \mathbf{\Psi}] \tag{89}$$

Minimization with respect to the unitary matrix leads to a new unitary invariant functional (Stengel and Spaldin, 2008; Lehtola and Jónsson, 2014; Borghi et al., 2015):

$$F^{SIC}[\mathbf{\Psi}] = \min_{U:U^\dagger U=I} E^{SIC}[U, \mathbf{\Psi}] \tag{90}$$

Minimization among unitary matrices can be achieved by employing the exponential transformation described in Sec. 3.2, $U = \exp(-A)$, which leads to the energy gradient as shown in equation (48). Thus, the minimization of the PZ-SIC functional consists of a double loop minimization: the inner loop representing a minimization with respect to U, and and the outer loops representing the minimization with respect to the wave function $\mathbf{\Psi}$ as described in section 3.3.

3.5 Performance of the Scaled Perdew-Zunger Self-interaction Correction

The PZ-SIC is shown to improve the energy of atoms and molecules when a complex domain of wave functions is used (Klüpfel et al., 2011, 2012b; Lehtola et al., 2016). However, PZ-SIC tends to overcorrect the energetics of the base functional, and scaling down the correction is necessary for a more accurate estimate of orbital energies in both molecules and solids (Bylaska et al., 2004, 2006; Vydrov et al., 2006; Dabo et al., 2010; Jonsson, 2011; Klüpfel et al., 2012b; Gudmundsdóttir et al., 2015; Colonna et al., 2018; Nguyen et al., 2018). Here, we will use the scaling factor of 1/2 which has been shown to work well for different types of systems (Jonsson, 2011; Klüpfel et al., 2012b; Gudmundsdóttir et al., 2015):

$$E^{PZ-SIC/2} = E - \frac{1}{2} \sum_{i=1..N} \left(\frac{1}{2} \iint d^3\mathbf{r}\, d^3\mathbf{r}' \frac{\rho_i(\mathbf{r})\rho_i(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + E_{xc}[\rho_i] \right) \tag{91}$$

This will be referred to as PZ-SIC/2 calculations. The direct minimization algorithm has been implemented in combination with SIC in the GPAW software (Mortensen et al., 2005; Enkovaara et al., 2010; Larsen et al., 2017). The implementation works for the three available representations implemented in GPAW: an LCAO basis set (Larsen et al., 2009), a PW basis set and FD RSG representation (Mortensen et al., 2005). The calculations are carried out using the frozen core approximation and the projector-augmented-wave method (Blöchl, 1994).

The test sets presented here include both molecular and solid state systems. All the molecules were placed in a rectangular box with boundaries at least 7 Å vacuum away from any nucleus in the molecule. The calculations of solids were performed using periodic boundary conditions with k-point sampling at Γ-point only. The grid mesh spacing was at most 0.15 Å.

The GPAW software uses two real space (or k-space) grids, one of which has twice as small grid mesh spacing as the other. The finer grid is used for representing the pseudo-density and calculating the exchange-correlation and Hartree energies and potentials. For PZ-SIC, one needs to estimate the one-electron density, Hartree and exchange correlation energies N times, where N is the number of occupied bands. However, the overall computational effort still scales as N^3, as for KS functionals. In order to reduce the cost of PZ-SIC calculations, one can use the coarse grid instead of the fine grid for calculating one-electron density, energies and potentials. Such an approach reduces the computational time by a factor of 8 and, at the same time, does not significantly downgrade the accuracy of the calculations. In order to validate this statement, we have calculated the atomization energy of 55 small molecules from the G2 set (Curtiss et al., 1997) using both coarse and fine grid meshes for the SIC term, and the results are presented in Fig. 3.2. The mean absolute error (MAE) for both calculations is 0.238 eV with a maximum energy deviation between the two calculations of 0.006 eV.

Atomization Energy and Ionization Potentials

Fig. 3.3 shows the performance of the PZ-SIC/2 employed with PBE functional (PBE-SIC/2). For reference atomization energies, we have used the experimental energies collected in the work of Paier et al. (2005), while for ionization potentials ($-\varepsilon_{HOMO}$), the self-consistent GW estimation from the work of Rostgaard et al. (2010) has been used as reference. The PBE-SIC/2 functional improves the atomization energy by a factor of 1.6. Since the SIC improves the energy of atoms (Klüpfel et al., 2011) as well as the atomization energy, the total energy of molecules is also improved. This result is in agreement with previous calculations (Klüpfel et al., 2012b; Lehtola et al., 2016) and, thus, validates our implementation of SIC. For the ionization potentials, there is an even larger improvement (by a factor of 2.5) than for the atomization energy, but the root mean square error (RMSE) is still large: 1.78 eV for the PBE-SIC/2.

Band gaps

The effect of SIC on the band gaps of solids is presented in Fig. 3.4. For small band gap semiconductors, the PBE-SIC/2 provides an error of opposite sign as compared

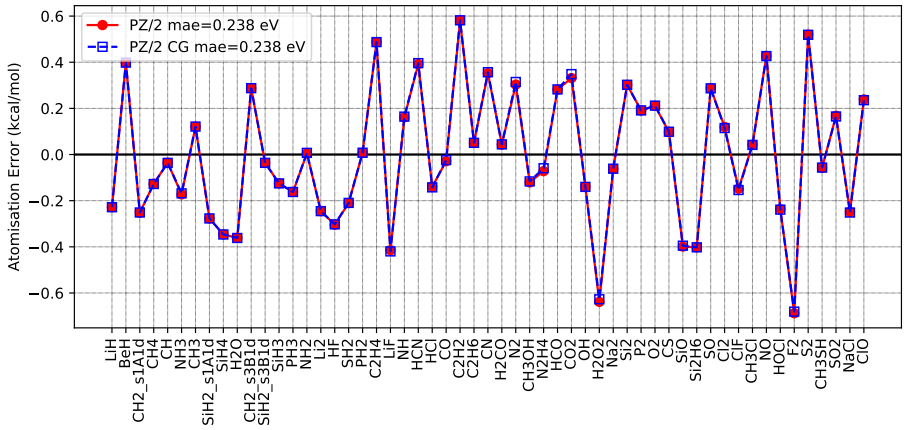

Figure 3.2. The atomization energy error obtained from the PZ-SIC functional for molecules in the G2-1 (Curtiss et al., 1997) set. The red line represents the PZ-SIC calculations when the SIC term is estimated on a coarse grid of 0.15 Å. The blue line represents the PZ-SIC calculations when the SIC term is estimated on a fine grid (0.075 Å). The PBE functional has been used as the base functional. The error is calculated with respect to the experimental atomization energies collected in the work of Paier et al. (2005).

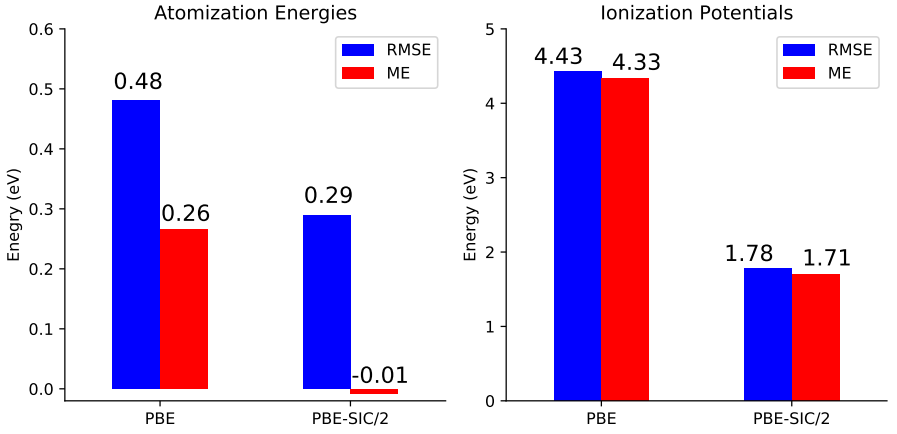

Figure 3.3. The performance of the PBE, PBE-SIC/2. (Left) The atomization energy errors are estimated on 55 molecules from a G2-1 molecular set (Curtiss et al., 1997). (Right) Ionization potential errors are estimated as $-\varepsilon_{HOMO}$ for 34 molecules from G2-1. The reference energy is taken from GW calculations from the work of Rostgaard et al. (2010).

to the PBE. This means that the SIC, even with a scaling of 1/2, 'overcorrects' the HOMO-LUMO gap. For example, for GaAs, the PBE predicts too low a band gap of 0.6 eV while the band gap estimated from the PBE-SIC/2 is 2.19 eV, and the experimental band gap is 1.52 eV(Kittel, 2005, p.190). At the same time, for large band gap materials such as Ar, half SIC is not enough to sufficiently open the band gap and a large scaling factor for SIC would be required. Thus, the PBE functional always underestimates the HOMO-LUMO gap while SIC/2 overestimates the gap for small band-gap materials and underestimates the gap for large band-gap materials. The PBE-SIC/2 functional still greatly improves performance as compared to the PBE functional in calculating the HOMO-LUMO gaps of solids, especially of those with a band gap larger than 4 eV as can be seen in Fig. 3.4.

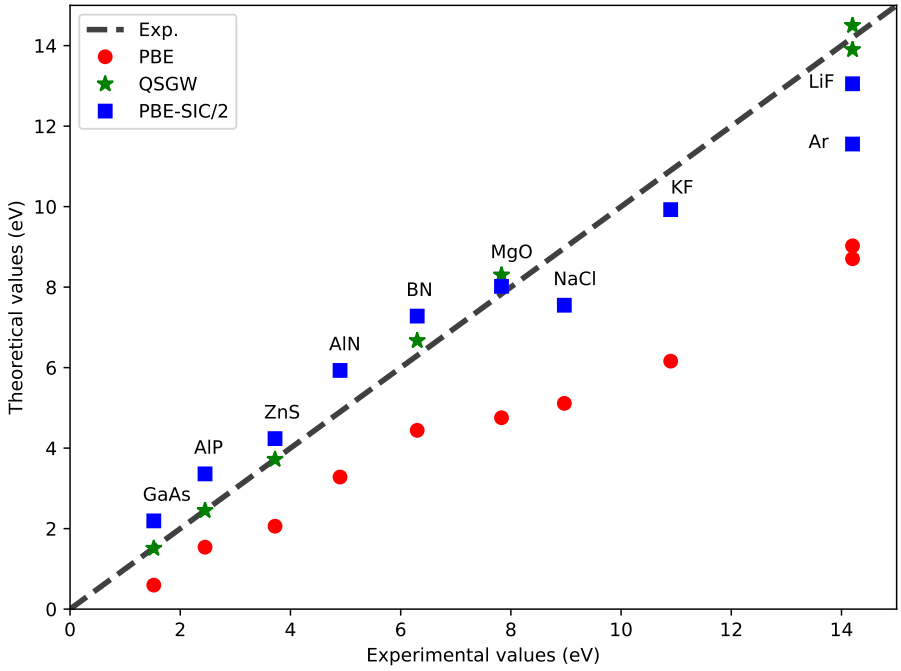

Figure 3.4. Band gaps calculated as HOMO-LUMO gaps for various levels of theory. [a] *The QSGW calculations are taken from the works of Shishkin et al. (2007), and Chen and Pasquarello (2015). For the experimental values, see references therein.*

Mn Dimer

The manganese dimer exhibits intricate chemical bonding in the ground state. Experimental measurements reveal a weak antiferromagnetic bonding interaction ($^1\Sigma_u^+$), as is reflected in the relatively long bond length of 3.4 Å (Baumann et al., 1982; Cheeseman et al., 1990) and a small binding energy of 0.13±0.1 eV (Kant et al., 1968). Similar

results are predicted by CASPT2 (Wang and Chen, 2004) and MCQDPT2 (Yamamoto et al., 2006) theories. Even though these theories predict a similar binding energy of 0.12eV and 0.14eV, respectively, their estimate of a bond distance differs by 0.35 Å (3.64 Å and 3.29 Å, respectively). DFT calculations with semilocal functionals provide a qualitatively incorrect ground state. The most recent all-electron calculations at the generalized gradient approximation (GGA) level of theory predict a ferromagnetic ground state ($^{11}\Pi_u$) with a bond length of 2.57-2.61Å and a binding energy of around 0.9 eV (Barborini, 2016). The description of this system, however, tends to show improvement when a large percentage of exact exchange is included in calculations using hybrid functionals (Yamanaka et al., 2007; Barborini, 2016).

The PBE and PBE-SIC/2 calculations of the potential energy curves for $^{1}\Sigma_g$, $^{11}\Sigma_u^+$ and $^{11}\Pi_u$ states are presented in Fig. 3.5. The calculations have been performed using an LCAO basis set which includes primitive Gaussians from the def2-TZVPD basis set (Weigend and Ahlrichs, 2005; Rappoport and Furche, 2010; Feller, 1996; Schuchardt et al., 2007; Pritchard et al., 2019) augmented with a single-zeta basis set of the GPAW software. As can be seen, the PBE-SIC/2 functional gives the correct ground state of the manganese dimer with a bond distance 3.32 Å and a binding energy 0.184 eV close to both experimental results and high-level quantum chemistry calculations. The $^{11}\Pi_u$ state is dominated by bonding between *d*-*d* electrons at a short distance. The PBE over-stabilises this orbital due to over-compensation of the classical self-Coulomb interaction by the exchange-correlation term (Ivanov et al., 2021a) and, as a result, predicts that $^{11}\Pi_u$ is the ground state. Removing this error by using the PBE-SIC/2 functional leads to a significant increase in the energy of this electronic state by around 1.0 eV. A much smaller self-interaction error is obtained for $^{1}\Sigma_g$, $^{11}\Sigma_u^+$ states which, instead of a bonding HOMO π molecular orbital formed by $3d$ atomic orbitals, have an anti-bonding HOMO σ^* state formed by $4s$ electrons. This result demonstrates that SIC accurately describes the magnetic configurations, and predicts a bond distance and a binding energy of the manganese dimer in agreement with the experimental measurements and high-level quantum chemistry calculations. A more detailed analysis is presented in article (Ivanov et al., 2021a) in this thesis.

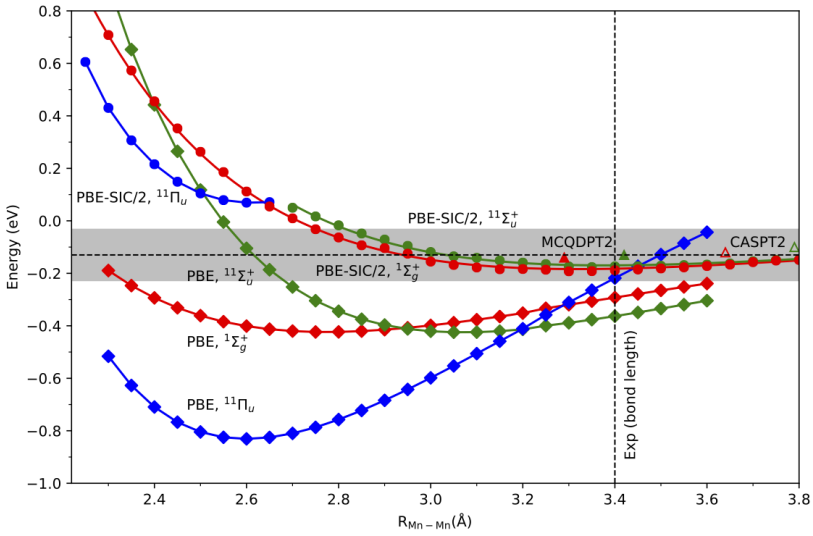

Figure 3.5. Binding curves of the Mn dimer. The filled diamonds and circles are the PBE and PBE-SIC/2 calculations, respectively. The red curve is the anti-ferromagnetic state, $^{1}\Sigma_g$, while the green and blue curves are the ferromagnetic states, $^{11}\Sigma_u^+$ and $^{11}\Pi_u$, respectively. The filled and open triangles are the MCQDPT2 (Yamamoto et al., 2006) and CASPT2 (Wang and Chen, 2004) results, respectively. The shaded gray area depicts the experimental binding energy, while the vertical dashed line represents the experimental bond length (Kant et al., 1968; Baumann et al., 1982; Cheeseman et al., 1990).

4 Direct Optimization Approach for Calculating Excited States

A commonly used method for calculating the excited-state properties of a quantum system is time-dependent DFT (TDDFT) (Runge and Gross, 1984; Casida, 1995; Dreuw and Head-Gordon, 2005). TDDFT is an exact method for describing electronic excitations in theory, but in practice, it requires several approximations. TDDFT is usually employed with linear-response theory and the adiabatic approximation which neglects the time dependence of the exchange-correlation kernel. Within these approximations, TDDFT provides quite an accurate description of low-lying valence excitations (Dreuw and Head-Gordon, 2005), but fails to describe double excitations (Tozer and Handy, 2000; Maitra et al., 2004; Levine et al., 2006), conical intersections (Levine et al., 2006) and Rydberg states (Cheng et al., 2008; Van Meer et al., 2014; Seidu et al., 2015). TDDFT equations depend on the exchange-correlation kernel and, in practice, this kernel is approximated using the ground-state energy functional. When semilocal functionals are used, the TDDFT equations lack coupling between spatially separated orbitals which leads to a catastrophic underestimation of the excitation energy of the long-range charge transfer reaction (Dreuw et al., 2003). This flaw can be remedied by using hybrid functionals which have a non-local exchange-correlation kernel. However, the amount of the exact exchange in hybrid functionals is an empirical parameter and therefore, it is not clear how accurately this approach would perform for excitations outside the set of molecules for which they were designed. Charge transfer excitations are of the upmost importance for example in the context of solar-energy conversion (Yeh et al., 2000; Bakulin et al., 2012; Harlang et al., 2015) and therefore, other efficient methods for simulating charge transfer processes are required.

An alternative method for calculating excited states is based on a *time-independent* density functional framework. In this approach, the electronic configuration of the excited state is calculated as a solution of the Kohn-Sham equations (18) for non-aufbau occupation numbers using a ground-state exchange-correlation functional. Perdew and Levy (1985) demonstrated that the stationary points of the exact ground-state density functional represent excited states, and the exact functional applied to all other excited-state densities, which are not extrema of the functional, provides the lower bound of the excited-state energy. At the same time, the KS energy functional can have more stationary points than a functional which depends on density only, and, at least in practice, a large number of excited states can be found as its extrema. It has been shown that such an approach provides a useful approximation for excited states and delivers accurate potential energy surfaces (Barca et al., 2018; Levi et al., 2018) including regions around conical intersections (Barca et al., 2018; Pradhan et al., 2018; Mališ and Luber, 2020). Furthermore, this method can accurately describe the

charge-transfer excitations (Zhekova et al., 2014) as well as double excitations (Hait and Head-Gordon, 2020) already at a semilocal level of approximation to the exchange-correlation functionals.

While DFT is a ground-state theory, there are also several generalizations of DFT for excited states (Görling, 1999; Levy and Nagy, 1999; Ayers et al., 2012, 2015, 2018). Due to the fact that several external potentials can provide the same excited-state density (Gaudoin and Burke, 2004), the energy functional cannot be a functional of the density only for an arbitrary external potential. In this case, the energy functional can be formulated as a bifunctional (Görling, 1999; Levy and Nagy, 1999) which, for example, can depend on both the excited-state density and the ground-state density (Levy and Nagy, 1999). However, in practice, the most commonly used external potential is the Coulomb potential which describes the electrostatic interaction of nuclei with electrons. The Coulomb systems have several important properties: the cusps of the density define the position of nuclei, and the derivative of spherically averaged density at the position of nuclei provides the nuclei's charges (Ayers et al., 2018), which is known as Kato's theorem (Kato, 1957). Therefore, the external Coulomb potential can be determined from the density only, which is Bright Wilson's argument for Coulomb ground-state DFT (Bicout and Field, 1996, p.2)). Furthermore, for finite systems, two different excited-state wave functions cannot have the same density because the asymptotic region of the density is defined by the ionization potential which differs for different excited states (Ayers et al., 2012). Based on these properties, it was shown that the energy functional of the density exists (Ayers et al., 2012, 2018) which is exact for both excited and ground states, and, as with the ground-state DFT, the Kohn-Sham equations can also be introduced (Ayers et al., 2015). The approximation made in practice is then the choice of the exchange-correlation functional, which is often set to the ground-state energy functional approximation.

The variational calculations typically used in DFT are problematic when calculating excited states due to the fact that the solution of the KS equations for excited states in many cases corresponds to a saddle point on the electronic energy surface. Conventional SCF algorithms based on interpolation techniques (Pulay, 1980; Kresse and Furthmüller, 1996; Garza and Scuseria, 2012) often collapse to the ground state (Gilbert et al., 2008; Barca et al., 2018) or exhibit oscillatory behaviour around the targeted solution (Mewes et al., 2014; Hait and Head-Gordon, 2020; Carter-Fenk and Herbert, 2020; Levi et al., 2020a,b). An alternative method to SCF algorithms is the direct minimization (DM) which often shows a more robust convergence in the case of ground-state calculations (Head-Gordon and Pople, 1988; Gillan, 1989; Payne et al., 1992; Hutter et al., 1994; Ismail-Beigi and Arias, 2000; Van Voorhis and Head-Gordon, 2002; VandeVondele and Hutter, 2003; Weber et al., 2008; Freysoldt et al., 2009; Ivanov et al., 2021b). This approach can be extended to calculations of excited states as is demonstrated below both for KS and PZ-SIC functionals.

4.1 LCAO Excited-State Algorithm

In order to use DM for calculating excited states, several changes need to be made. First, the minimization line search cannot be used because the system may need to climb up in energy to the saddle point. Instead, one needs to use a trust region optimization which, in its simple implementation, just restricts the step size along the search direction by some threshold value. Secondly, the curvature of an electronic energy surface at the saddle point is convex along some degrees of freedom and concave along others. Thus, the optimization algorithm has to be able to identify those degrees of freedom along which the energy must be maximized. This can be achieved using quasi-Newton methods that can build a non-positive definite matrix approximation to the true Hessian. Examples of such quasi-Newton methods are the symmetric rank-one (SR1) method (Nocedal and Wright, 2006, p.144) and the Powell algorithm (Anglada et al., 1999). The L-BFGS algorithm used for direct minimization in Sec. 3 usually provides a positive definite approximations to the Hessian (Nocedal and Wright, 2006, p.135) and cannot, therefore, be applied to saddle-point searches unless a special preconditioner is used.

In LCAO calculations, when the rotation matrix mixes orbitals between occupied and virtual states, the preconditioner which can estimate unstable modes of the Hessian is given in Eq. (63) with $\gamma = 0$:

$$P_{ij} = \frac{1}{-2(\varepsilon_i - \varepsilon_j)(f_i - f_j)}, \tag{92}$$

Indeed, when an electron is promoted from the i^{th} orbital to the j^{th} orbital and $\varepsilon_i < \varepsilon_j$, then $-(\varepsilon_i - \varepsilon_j)(f_i - f_j) < 0$ since $f_i = 0$ and $f_j = 1$. Therefore, the matrix element of the preconditioner for a rotation of the ij orbital pair is negative indicating that the energy for this rotation needs to be maximized.

The changes that need to be made in order to use the direct minimization algorithm proposed in Sec. 3.2 for excited-state calculations are:

1. Instead of an inexact line search, use constrained step length update algorithm. Namely, calculate a search direction D and if

   $$\|D\| > \alpha_{max} \tag{93}$$

   then:

   $$D \leftarrow \frac{D}{\|D\|}\alpha_{max} \tag{94}$$

   where α_{max} is an empirical parameter defined by the user. Using this method, it was found that $\alpha = 0.2 - 0.25$ provides good performance.

2. Instead of the L-BFGS algorithm, use the limited-memory SR1 (L-SR1) algorithm which has the ability to develop negative eigenvalues in the approximated Hessian matrix.

3. Use the preconditioner (92).

4. In order to prevent a variational collapse to the ground state, use the maximum overlap method (MOM) (Gilbert et al., 2008; Barca et al., 2018) which occupies the orbitals of the excited states.

This approach has been described in detail in the work of Levi et al. (2020b) and will hereafter be referred to as the direct optimization (DO-MOM) method. The performance of DO-MOM within KS-DFT was compared to that of the direct diagonalization SCF algorithm employed with MOM (SCF-MOM) on molecular set including 88 excitations. The proposed L-SR1 algorithm was able to reach the convergence for all excitations while the conventional SCF algorithm based on the DIIS method failed in 27 cases. Furthermore, the average number of energy and gradient evaluations in the L-SR1 algorithm was lower by a factor of 2 than that in the SCF. This demonstrates that not only is the proposed methodology more robust, but it is also faster than the method commonly used today.

While the developed DO is useful for LCAO calculations, the large size of virtual space and the large amount of memory used in the L-SR1 algorithm (20 previous optimization steps) prevent its usage for a RS or PW basis set. For the latter, a different approach can be developed which combines both the efficient saddle-point search discussed here and the direct minimization algorithm from Sec 3.3.

4.2 Real Space Grid and Plane Wave Excited-State Algorithm

Let $E[\mathbf{\Psi}]$ be the KS or preconditioned PZ-SIC functional as defined in Sec. 3.4. Both functionals are unitary invariant with respect to a rotation of occupied orbitals. Let $\mathbf{\Phi}$ be some auxiliary orthonormal orbitals. Then the optimal orbitals, $\mathbf{\Psi}$, can be obtained by applying a unitary transformation to $\mathbf{\Phi}$:

$$\mathbf{\Psi} = U\mathbf{\Phi} \tag{95}$$

In this case, the energy functional $E[\mathbf{\Psi}]$ can be considered as the functional of both the matrix U and orbitals $\mathbf{\Phi}$. Therefore, the stationary point of E can be found through iterations involving extremization with respect to the matrix U followed by minimization with respect to the auxiliary orbitals $\mathbf{\Phi}$:

$$\operatorname*{stat}_{\mathbf{\Psi}} E[\mathbf{\Psi}] = \min_{\mathbf{\Phi}} \operatorname*{stat}_{U} E[U\mathbf{\Phi}] \tag{96}$$

The auxiliary orbitals can be chosen to be the ground-state KS orbitals or to any other initial guess for an excited state. They can also be thought of as the basis functions and the matrix U as the expansion coefficients of the optimal orbitals. If $\mathbf{\Phi}$ represents the complete basis set, then U is the identity matrix because the stationary point is found with respect to all degrees of freedom. In practice, $\mathbf{\Phi}$ consists of occupied and only a few virtual orbitals and a second step is, therefore, necessary which changes the basis set so as to minimize the total energy. This is contrary to LCAO calculations, where the basis set is always fixed and the full basis-set limit can only be reached by systematically increasing the size of $\mathbf{\Phi}$.

Thus, the DO algorithm for PW and RSG representations consists of two nested loops. The inner loop is based on the LCAO method from Sec. 4.1 and is used to find the unitary matrix U for a fixed $\mathbf{\Phi}$. The outer loop is a minimization of $L[\mathbf{\Phi}] = \operatorname*{stat}_{U} E[U\mathbf{\Phi}]$

with respect to $\mathbf{\Phi}$ using an algorithm from Sec. 3.3. A combination of both loops is described in detail in the article (Ivanov et al., 2021c) in this thesis.

4.3 Performance of the Scaled PZ-SIC for Excited States

The performance of the scaled PZ-SIC for excited states has been assessed using 9 valence and 9 Rydberg excitations in 13 molecules for both singlet and triplet states. These molecules are acetaldehyde, acetylene, ammonia, carbon-monoxide, diazomethane, ethylene, formaldehyde, formamide, hydrogensulfide, ketene, methanimine, thioformaldehyde, and water. We have used reference energies and geometries from the work of Loos et al. (2018) in order to assess the PBE, PBE-SIC/2. Since it is known that complex orbitals are needed to calculate the ground state properties we have also performed calculations using real orbitals (the PBE-RSIC/2 functional) in order to see what effect they have on the excitation energies. Due to the fact that the singlet determinant has been used in calculations, the open shell singlet states are spin contaminated (mixed spin states). Therefore, a spin purification formula (Ziegler et al., 1977) has been applied in order to estimate the excitation energy of a singlet state:

$$E = 2E(\uparrow\downarrow) - E(\uparrow\uparrow), \tag{97}$$

where $E(\uparrow\downarrow)$ is the energy of the mixed spin state and $E(\uparrow\uparrow)$ is the energy of the triplet state. The energy of the mixed spin state is compared here with that of the singlet reference state.

Table 4.1. A summary of the performance of the PBE, scaled self-interaction correction restricted to real orbitals (PBE-RSIC/2) and scaled self-interaction correction (PBE-SIC/2).

Multiplicity	Functional	ME	MAE	RMSE	Max Abs Error
Triplet State					
	PBE	-0.27	0.27	0.31	0.55
	PBE-RSIC/2	-0.38	0.39	0.44	0.8
	PBE-SIC/2	-0.25	0.26	0.29	0.51
Mixed Spin State					
	PBE	-0.73	0.73	0.95	2.33
	PBE-RSIC/2	-0.76	0.76	0.89	2.02
	PBE-SIC/2	-0.66	0.66	0.83	2.02
Purified Singlet State					
	PBE	-0.46	0.46	0.54	1.20
	PBE-RSIC/2	-0.40	0.46	0.50	0.76
	PBE-SIC/2	-0.33	-0.40	0.43	0.75

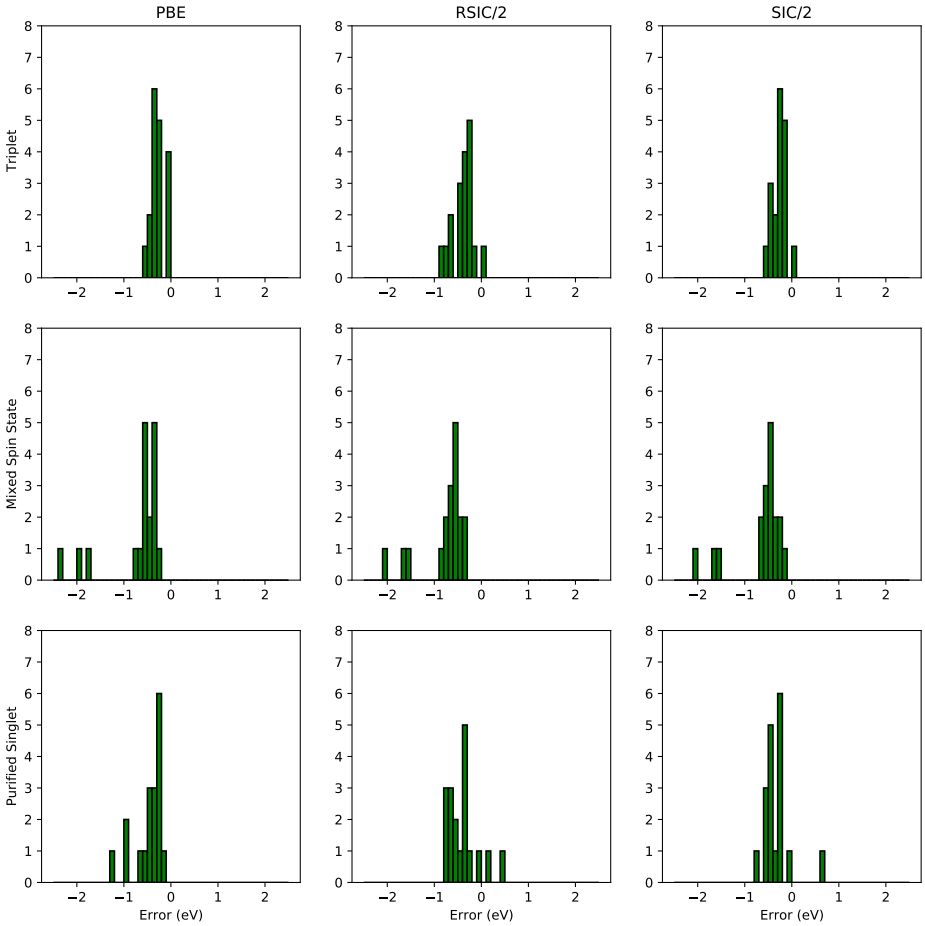

Figure 4.6. Histograms of errors obtained with different functionals for triplet, mixed spin and purified singlet excitations. Left column: PBE, Middle column: PBE-RSIC/2. Right column: PBE-SIC/2.

The results of the calculations are summarised in Table 4.1 and Fig. 4.6. As can be seen from Fig. 4.6 the PBE functional always underestimates the excitation energy. It performs better for triplet excitations than for mixed spin excitations with the root mean square error (RMSE) of 0.31 eV and 0.95 eV, respectively. The maximum error is for mixed spin excitations, as large as 2.33 eV. As anticipated, the spin purification formula (97) provides a better estimation of a singlet energy than the mixed spin determinant reducing the RMSE to 0.54 eV and the maximum error to 1.20 eV.

When SIC/2 is applied to the PBE functional and optimization is restricted to real valued orbitals, the error in the triplet excitation energy increases RMSE by 0.13 eV. At the same time, the PBE-RSIC/2 performs better for the mixed spin state and purified singlet states by around 0.05 eV according to the RMSE. The maximum error is also

smaller than in the PBE functional and it equals 2.02 eV and 0.76 eV for mixed spin states and purified singlets, respectively.

The PBE-SIC/2 functional performs better than the PBE and PBE-RSIC/2 functionals for both triplet and singlet states. For triplet excitations the improvement is marginal and the RMSE equals 0.29 eV which is close to the PBE results. The RMSE for mixed-spin and purified singlet excitation is smaller by around 0.1 eV as compared to PBE.

In conclusion, fully variational algorithms for the calculation of excited states within the KS-DFT and PZ-SIC formalisms have been developed and, for the first time, a variational assessment of the PZ-SIC functional has been performed on excited states. PZ-SIC using complex-valued orbitals performs better than in the case of real-valued orbitals. This result is on par with calculations of the total energy of atoms (Klüpfel et al., 2011), and geometric and energetic properties of molecules (Klüpfel et al., 2012b). The effect of PZ-SIC on the excitation energy of both valence and Rydberg states is small on the order of 0.01 - 0.1 eV. This can be attributed to the fact that PZ-SIC is a one-electron self-interaction correction (Ivanov et al., 2021c). However, it is expected that PZ-SIC will have a larger effect on excitations with charge localization as in the example of the diamine molecule (N,N' -Dimethylpiperazine) where semilocal and commonly used hybrid functionals fail to stabilise the localized state (Cheng et al., 2016).

5 Conclusion

Optimization methods for calculating both ground and excited states have been developed for the PZ-SIC functional and implemented. The implementation can make use of three choices of basis sets: LCAO, PW and the FD RSG. PZ-SIC provides an efficient alternative to hybrid functionals (Ivanov et al., 2021a) in PW and finite-difference RSG representations. The methods presented here are fully variational and, as a result, atomic forces are available through the Hellmann-Feynman theorem, which, in turn, makes it possible to explore both ground- and excited-state atomic energy surfaces in an efficient way.

The assessment of the scaled PZ-SIC showed that it performs systematically better than the PBE functional in calculating the atomization and ionization energies as well as band gaps. Furthermore, it can accurately describe systems with transition metals, such as the Mn dimer for which semilocal functionals fail dismally even for the ground state. For excited states, the scaled PZ-SIC does not significantly reduce the excitation-energy error of the PBE functional in the Franck-Condon region. However, scaled PZ-SIC reproduces the shape of the potential energy surface for excited states more accurately than the PBE functional (Ivanov et al., 2021c). These findings indicate that a further improvement in excited-state energy calculations require the development of a new SIC approach which goes beyond one-electron corrections.

As often occurs in research, an investigation into a new problem can have a positive impact on previous works and other related areas. Not only are the direct optimization algorithms applicable for variational calculations using the PZ-SIC functional, they can also be applied to regular DFT calculations. The direct minimization approach for KS-DFT based on the exponential transformation of molecular orbitals can outperform the conventional SCF algorithm based on the DIIS method by a factor of up to 2 in some cases and, moreover, it gives a more robust convergence (Ivanov et al., 2021b). A modification of this method for calculating excited states also produces a superior performance (Levi et al., 2020b; Ivanov et al., 2021c) and thus, provides an appealing approach for studying complex processes in ultrafast experiments (Ge et al., 1998; Yeh et al., 2000; Bakulin et al., 2012).

6 Appendix A: Derivative of the Matrix Exponential

Let A be an $M \times M$ skew-hermitian matrix, that is $A^\dagger = -A$, and a_{ij} are matrix elements of A. Let us define the derivative of the matrix exponential with respect to the real part of the complex non-diagonal elements of A as:

$$\frac{\partial e^A}{\partial \mathrm{Re}(a_{ij})} = \lim_{h\to 0} \frac{e^{A+h(E_{ij}-E_{ji})} - e^A}{h} \tag{98}$$

and with respect to the imaginary part as:

$$\frac{\partial e^A}{\partial \mathrm{Im}(a_{ij})} = i \lim_{h\to 0} \frac{e^{A+ih(E_{ij}+E_{ji})} - e^A}{ih} \tag{99}$$

where the matrix E_{ij} is such that $(E_{ij})_{mn} = \delta_{im}\delta_{jn}$

The last two formulae equal [see Eq.(2) from Ref.(Najfeld and Havel, 1995)]:

$$\frac{\partial e^A}{\partial \mathrm{Re}(a_{ij})} = \int_0^1 e^{tA} \left(E_{ij} - E_{ji}\right) e^{(1-t)A}\, dt \tag{100}$$

$$\frac{\partial e^A}{\partial \mathrm{Im}(a_{ij})} = i \int_0^1 e^{tA} \left(E_{ij} + E_{ji}\right) e^{(1-t)A}\, dt, \tag{101}$$

Using:

$$\frac{\partial e^A}{\partial a_{ij}} = \frac{1}{2}\left(\frac{\partial e^A}{\partial \mathrm{Re}(a_{ij})} - i\frac{\partial e^A}{\partial \mathrm{Im}(a_{ij})}\right) \tag{102}$$

one can obtain:

$$\frac{\partial e^A}{\partial a_{ij}} = \int_0^1 e^{tA} E_{ij} e^{(1-t)A}\, dt \tag{103}$$

For the diagonal elements of the skew-symmetric matrix, the gradient with respect to the real part equals zero and therefore, let us define the derivative with respect to the imaginary part only:

$$\frac{\partial e^A}{\partial \mathrm{Im}(a_{ii})} = i \lim_{h\to 0} \frac{e^{A+ihE_{ii}} - e^A}{ih} \tag{104}$$

$$\frac{\partial e^A}{\partial a_{ii}} = \frac{\partial e^A}{\partial \mathrm{Im}(a_{ii})} = \int_0^1 e^{tA} E_{ii} e^{(1-t)A}\, dt \tag{105}$$

Thus, the equation for the derivative with respect to the matrix element is:

$$\frac{\partial e^A}{\partial a_{ij}} = \int_0^1 e^{tA} E_{ij} e^{(1-t)A} dt \tag{106}$$

and:

$$\frac{\partial e^A}{\partial a_{ij}} = E_{ij} + o(A) \tag{107}$$

Similarly, one can obtain:

$$\frac{\partial e^{-A}}{\partial a_{ij}} = -\int_0^1 e^{-tA} E_{ij} e^{(-1-t)A} dt \tag{108}$$

and:

$$\frac{\partial e^{-A}}{\partial a_{ij}} = -E_{ij} + o(A) \tag{109}$$

7 Appendix B: Preconditioning for the Exponential Transformation

Let us denote:

$$\left(\frac{\partial E}{\partial a}\right)_{ij} = \frac{\partial E}{\partial a_{ij}} = \frac{2-\delta_{ij}}{2}\left(\int_0^1 e^{tA} L e^{-t}\, dt\right)_{ji} \tag{110}$$

and consider only a real-valued case. Expanding the integral into a series, one obtains:

$$\left(\frac{\partial E}{\partial a}\right)^T = 2L + [A,L] + O(A^2) \tag{111}$$

Simplifying:

$$L = FO^T HO - O^T HOF \tag{112}$$

$$O = C\exp(A), \quad \text{let } H^0 = C^T HC \rightarrow \tag{113}$$

$$L = [F, e^{-A} H^0 e^A], \tag{114}$$

$$e^{-A} H^0 e^A = H^0 - [A, H^0] + O(A^2) \tag{115}$$

$$L = [F, H^0] - [F, [A, H^0]], \tag{116}$$

one can obtain:

$$\left(\frac{\partial E}{\partial a}\right)^T = 2\left[F, H^0\right] - 2\left[F, [A, H^0]\right] + \left[A, [F, H^0]\right] + O(A^2) \tag{117}$$

$$\left[A, [F, H^0]\right] = -\left[F, [H^0, A]\right] - \left[H^0, [A, F]\right] \tag{118}$$

$$\left(\frac{\partial E}{\partial a}\right)^T = 2\left[F, H^0\right] - \left[F, [A, H^0]\right] + \left[H^0, [F, A]\right] + O(A^2) \tag{119}$$

$$\frac{\partial E}{\partial a} = -2\left[F, H^0\right] + \left[F, [A, H^0]\right] - \left[H^0, [F, A]\right] + O(A^2) \tag{120}$$

Let us consider the first derivative of the second term, $\left[F, [A, H^0]\right]$ from Eq. (120):

$$\frac{\partial^2 E^{(2)}}{\partial a_{lm} \partial a_{ij}} = \left[F, \left[E_{lm} - E_{ml}, H^0\right]\right]_{ij} = (f^i - f^j)\left[E_{lm} - E_{ml}, H^0\right]_{ij} = \tag{121}$$

$$= (f^i - f^j) \sum_k \left(\delta_{il}\delta_{mk} - \delta_{im}\delta_{lk}\right) H^0_{kj} - \left(\delta_{kl}\delta_{mj} - \delta_{km}\delta_{lj}\right) H^0_{ik} \tag{122}$$

$$= (f^i - f^j)\left(\delta_{li} H^0_{mj} - \delta_{mi} H^0_{lj} - \delta_{mj} H^0_{il} + \delta_{lj} H^0_{im}\right) \tag{123}$$

Now, let us have a look at the first derivative of the third term in Eq. (120):

$$[F, E_{lm} - E_{ml}]_{xy} = (f^x - f^y)\left(E^{xy}_{lm} - E^{xy}_{ml}\right) = (f^x - f^y)\left(\delta_{lx}\delta_{my} - \delta_{mx}\delta_{ly}\right) \tag{124}$$

$$\sum_x H^0_{ix}[F, E_{lm} - E_{ml}]_{xj} = \sum_x H^0_{ix}(f^x - f^j)\left(\delta_{lx}\delta_{mj} - \delta_{mx}\delta_{lj}\right) = \tag{125}$$

$$= H^0_{im}(f^j - f^m)\delta_{lj} - H^0_{il}(f^j - f^l)\delta_{mj} \tag{126}$$

$$\sum_y [F, E_{lm} - E_{lm}]_{iy} H^0_{yj} = \sum_y (f^i - f^y)\left(\delta_{li}\delta_{my} - \delta_{mi}\delta_{ly}\right) H^0_{yj} = \tag{127}$$

$$= H^0_{mj}(f^i - f^m)\delta_{li} - H^0_{lj}(f^i - f^l)\delta_{mi} \tag{128}$$

$$\frac{\partial^2 E^{(3)}}{\partial a_{lm}\partial a_{ij}} = -\left[H^0, [F, E_{lm} - E_{ml}]\right]_{ij} = \tag{129}$$

$$H^0_{il}(f^j - f^l)\delta_{mj} - H^0_{im}(f^j - f^m)\delta_{lj} + H^0_{mj}(f^i - f^m)\delta_{li} - H^0_{lj}(f^i - f^l)\delta_{mi} \tag{130}$$

Thus, the total contribution:

$$\frac{\partial^2 E}{\partial a_{lm}\partial a_{ij}} = \delta_{li}H^0_{mj}(f^i - f^j) - \delta_{mi}H^0_{lj}(f^i - f^j) - \delta_{mj}H^0_{il}(f^i - f^j) + \delta_{lj}H^0_{im}(f^i - f^j) + \tag{131}$$

$$+ \delta_{mj}H^0_{il}(f^j - f^l) - \delta_{lj}H^0_{im}(f^j - f^m) + \delta_{li}H^0_{mj}(f^i - f^m) - \delta_{mi}H^0_{lj}(f^i - f^l) \tag{132}$$

or:

$$\frac{\partial^2 E}{\partial a_{ij}\partial a_{lm}} = \delta_{il}H^0_{jm}(f^l + f^i - f^j - f^m) \tag{133}$$

$$+\delta_{jl}H^0_{im}(f^m + f^i - f^l - f^j) \tag{134}$$

$$+\delta_{jm}H^0_{li}(f^m - f^i - f^l + f^j) \tag{135}$$

$$+\delta_{im}H^0_{lj}(f^l - f^m - f^i + f^j) \tag{136}$$

If one chooses canonical orbitals, then only the elements with $i = l, j = m$ are not zeros:

$$\frac{\partial^2 E}{\partial^2 a_{ij}} = -2(H^0_{ii} - H^0_{jj})(f_i - f_j) \tag{137}$$

References

Anglada, J. M., Besalú, E., Bofill, J. M., and Rubio, J. (1999). Another way to implement the Powell formula for updating Hessian matrices related to transition structures. *J. Math. Chem.*, 25(1):85–92.

Ayers, P. W., Levy, M., and Nagy, A. (2012). Time-independent density-functional theory for excited states of Coulomb systems. *Phys. Rev. A*, 85(4):042518.

Ayers, P. W., Levy, M., and Nagy, Á. (2015). Communication: Kohn-Sham theory for excited states of Coulomb systems. *J. Chem. Phys.*, 143(19):191101.

Ayers, P. W., Levy, M., and Nagy, Á. (2018). Time-independent density functional theory for degenerate excited states of Coulomb systems. *Theor. Chem. Acc.*, 137(11):152.

Bakulin, A. A., Rao, A., Pavelyev, V. G., Van Loosdrecht, P. H., Pshenichnikov, M. S., Niedzialek, D., Cornil, J., Beljonne, D., and Friend, R. H. (2012). The role of driving energy and delocalized states for charge separation in organic semiconductors. *Science*, 335(6074):1340–1344.

Barborini, M. (2016). Neutral, Anionic, and Cationic Manganese Dimers through Density Functional Theory. *J. Phys. Chem. A*, 120(10):1716–1726.

Barca, G. M., Gilbert, A. T., and Gill, P. M. (2018). Simple Models for Difficult Electronic Excitations. *J. Chem. Theory Comput.*, 14(3):1501–1509.

Baumann, C. A., Van Zee, R. J., Bhat, S. V., and Weltner, W. (1982). ESR of Mn2 and Mn5 molecules in rare-gas matrices. *J. Chem. Phys.*, 78(1):190–199.

Bicout, D. and Field, M. E. (1996). *Quantum Mechanical Simulation Methods for Studying Biological Systems*. Springer Berlin Heidelberg, Berlin, Heidelberg.

Blöchl, P. E. (1994). Projector augmented-wave method. *Phys. Rev. B*, 50(24):17953–17979.

Borghi, G., Park, C. H., Nguyen, N. L., Ferretti, A., and Marzari, N. (2015). Variational minimization of orbital-density-dependent functionals. *Phys. Rev. B*, 91(15).

Briggs, E. L., Sullivan, D. J., and Bernholc, J. (1995). Large-scale electronic-structure calculations with multigrid acceleration. *Phys. Rev. B*, 52(8):R5471—-R5474.

Bylaska, E., Tsemekhman, K., and Jonsson, H. (2004). Self-consistent self-interaction corrected DFT: The method and applications to extended and confined systems. *APS*, 2004:L38—-004.

Bylaska, E. J., Tsemekhman, K., and Gao, F. (2006). New development of self-interaction corrected DFT for extended systems applied to the calculation of native defects in 3C-SiC. *Phys. Scr. T*, T124:86–90.

Carter-Fenk, K. and Herbert, J. M. (2020). State-Targeted Energy Projection: A Simple and Robust Approach to Orbital Relaxation of Non-Aufbau Self-Consistent Field Solutions. *J. Chem. Theory Comput.*, 16(8):5067–5082.

Casida, M. E. (1995). *Time-Dependent Density Functional Response Theory for*

Molecules, pages 155–192. World Scientific.

Cheeseman, M., Van Zee, R. J., Flanagan, H. L., and Weltner, W. (1990). Transition-metal diatomics: Mn 2 , Mn + 2 , CrMn. *J. Chem. Phys.*, 92(3):1553–1559.

Chen, W. and Pasquarello, A. (2015). Accurate band gaps of extended systems via efficient vertex corrections in GW. *Phys. Rev. B*, 92(4):1–5.

Cheng, C.-L., Wu, Q., and Van Voorhis, T. (2008). Rydberg energies using excited state density functional theory. *J. Chem. Phys.*, 129(12):124112.

Cheng, X., Zhang, Y., Jónsson, E., Jónsson, H., and Weber, P. M. (2016). Charge localization in a diamine cation provides a test of energy functionals and self-interaction correction. *Nat. Commun.*, 7(1):11013.

Colonna, N., Nguyen, N. L., Ferretti, A., and Marzari, N. (2018). Screening in Orbital-Density-Dependent Functionals. *J. Chem. Theory Comput.*, 14(5):2549–2557.

Curtiss, L. A., Raghavachari, K., Redfern, P. C., and Pople, J. A. (1997). Assessment of Gaussian-2 and density functional theories for the computation of enthalpies of formation. *J. Chem. Phys.*, 106(3):1063–1079.

Dabo, I., Ferretti, A., Poilvert, N., Li, Y., Marzari, N., and Cococcioni, M. (2010). Koopmans' condition for density-functional theory. *Phys. Rev. B*, 82(11):1–16.

Davidson, E. R. (1975). The iterative calculation of a few of the lowest eigenvalues and corresponding eigenvectors of large real-symmetric matrices. *J. Comput. Phys.*, 17(1):87–94.

Dreuw, A. and Head-Gordon, M. (2005). Single-Reference ab Initio Methods for the Calculation of Excited States of Large Molecules. *Chem. Rev.*, 105(11):4009–4037.

Dreuw, A., Weisman, J. L., and Head-Gordon, M. (2003). Long-range charge-transfer excited states in time-dependent density functional theory require non-local exchange. *J. Chem. Phys.*, 119(6):2943–2946.

Enkovaara, J., Rostgaard, C., Mortensen, J. J., Chen, J., Dułak, M., Ferrighi, L., Gavnholt, J., Glinsvad, C., Haikola, V., Hansen, H. A., Kristoffersen, H. H., Kuisma, M., Larsen, A. H., Lehtovaara, L., Ljungberg, M., Lopez-Acevedo, O., Moses, P. G., Ojanen, J., Olsen, T., Petzold, V., Romero, N. A., Stausholm-Møller, J., Strange, M., Tritsaris, G. A., Vanin, M., Walter, M., Hammer, B., Häkkinen, H., Madsen, G. K. H., Nieminen, R. M., Nørskov, J. K., Puska, M., Rantala, T. T., Schiøtz, J., Thygesen, K. S., and Jacobsen, K. W. (2010). Electronic structure calculations with GPAW: a real-space implementation of the projector augmented-wave method. *J. Phys. Condens. Matter*, 22(25):253202.

Feller, D. (1996). The role of databases in support of computational chemistry calculations. *J. Comput. Chem.*, 17(13):1571–1586.

Freysoldt, C., Boeck, S., and Neugebauer, J. (2009). Direct minimization technique for metals in density functional theory. *Phys. Rev. B*, 79(24):1–4.

Furthmüller, J. and Kresse, G. (1996). Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.*, 6(1):15–50.

Garza, A. J. and Scuseria, G. E. (2012). Comparison of self-consistent field convergence acceleration techniques. *J. Chem. Phys.*, 137(5).

Gaudoin, R. and Burke, K. (2004). Lack of Hohenberg-Kohn Theorem for Excited States. *Phys. Rev. Lett.*, 93(17):173001.

Ge, N.-H., Wong, C. M., Lingle, R. L., McNeill, J. D., Gaffney, K. J., and Harris, C. B.

(1998). Femtosecond Dynamics of Electron Localization at Interfaces. *Science*, 279(5348):202–205.

Gilbert, A. T., Besley, N. A., and Gill, P. M. (2008). Self-consistent field calculations of excited states using the maximum overlap method (MOM). *J. Phys. Chem. A*, 112(50):13164–13171.

Gillan, M. J. (1989). Calculation of the vacancy formation energy in aluminium. *J. Phys. Condens. Matter*, 1(4):689–711.

Görling, A. (1999). Density-functional theory beyond the Hohenberg-Kohn theorem. *Phys. Rev. A*, 59(5):3359–3374.

Gudmundsdóttir, H., Jónsson, E. Ö., and Jónsson, H. (2015). Calculations of Al dopant in α -quartz using a variational implementation of the Perdew–Zunger self-interaction correction. *New J. Phys.*, 17(8):083006.

Hait, D. and Head-Gordon, M. (2020). Excited state orbital optimization via minimizing the square of the gradient: General approach and application to singly and doubly excited states via density functional theory. *J. Chem. Theory Comput.*, 16:1699–1710.

Harlang, T. C. B., Liu, Y., Gordivska, O., Fredin, L. A., Ponseca, C. S., Huang, P., Chábera, P., Kjaer, K. S., Mateos, H., Uhlig, J., Lomoth, R., Wallenberg, R., Styring, S., Persson, P., Sundström, V., and Wärnmark, K. (2015). Iron sensitizer converts light to electrons with 92% yield. *Nat. Chem.*, 7(11):883–889.

Head-Gordon, M. and Pople, J. A. (1988). Optimization of wave function and geometry in the finite basis Hartree-Fock method. *J. Phys. Chem.*, 92(11):3063–3069.

Hohenberg, P. and Kohn, W. (1964). Inhomogeneous electron gas. *Phys. Rev.*, 136(3B):B864.

Hutter, J., Parrinello, M., and Vogel, S. (1994). Exponential transformation of molecular orbitals. *J. Chem. Phys.*, 101(5):3862–3865.

Ismail-Beigi, S. and Arias, T. (2000). New algebraic formulation of density functional calculation. *Comput. Phys. Commun.*, 128(1-2):1–45.

Ivanov, A., Gosh, T., Jónsson E., and Jónsson, H. (2021a). Mn Dimer can be Described Accurately with Density Functional Calculations when Self-interaction Correction is Applied. *arxiv:2102.00890 [physics.chem-ph]*.

Ivanov, A. V., Jónsson E., Vegge, T., and Jónsson, H. (2021b). Direct Energy Minimization Based on Exponential Transformation in Density Functional Calculations of Finite and Extended Systems. *arXiv:2101.12597 [physics.comp-ph]*.

Ivanov, A. V., Levi, G., Jónsson E., and Jónsson, H. (2021c). Direct Optimization Method for Variational Excited-State Density Functional Calculations Using Real Space Grid or Plane Waves. *Submitt. to J. Chem. Theory Comput.*

Jonsson, H. (2011). Simulation of surface processes. *Proc. Natl. Acad. Sci.*, 108(3):944–949.

Kant, A., Lin, S., and Strauss, B. (1968). Dissociation Energy of Mn_2. *J. Chem. Phys.*, 49(4):1983–1985.

Kato, T. (1957). On the eigenfunctions of many-particle systems in quantum mechanics. *Commun. Pure Appl. Math.*, 10(2):151–177.

Kittel, C. (2005). *Introduction to Solid State Physics*. Wiley, 8th edition.

Klüpfel, P., Klüpfel, S., Tsemekhman, K., and Jónsson, H. (2012a). Optimization of Functionals of Orthonormal Functions in the Absence of Unitary Invariance. In *Lect. Notes Comput. Sci.*, volume 7134 LNCS, pages 23–33.

Klüpfel, S., Klüpfel, P., and Jónsson, H. (2011). Importance of complex orbitals in calculating the self-interaction-corrected ground state of atoms. *Phys. Rev. A*, 84(5):050501.

Klüpfel, S., Klüpfel, P., and Jónsson, H. (2012b). The effect of the Perdew-Zunger self-interaction correction to density functionals on the energetics of small molecules. *J. Chem. Phys.*, 137(12):124102.

Kohn, W. and Sham, L. J. (1965). Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.*, 140(4A):1133–1138.

Kresse, G. and Furthmüller, J. (1996). Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B*, 54(16):11169–11186.

Larsen, A. H., Mortensen, J. J., Blomqvist, J., Castelli, I. E., Christensen, R., Dułak, M., Friis, J., Groves, M. N., Hammer, B., Hargus, C., Hermes, E. D., Jennings, P. C., Jensen, P. B., Kermode, J., Kitchin, J. R., Kolsbjerg, E. L., Kubal, J., Kaasbjerg, K., Lysgaard, S., Maronsson, J. B., Maxson, T., Olsen, T., Pastewka, L., Peterson, A., Rostgaard, C., Schiøtz, J., Schütt, O., Strange, M., Thygesen, K. S., Vegge, T., Vilhelmsen, L., Walter, M., Zeng, Z., and Jacobsen, K. W. (2017). The atomic simulation environment—a Python library for working with atoms. *J. Phys. Condens. Matter*, 29(27):273002.

Larsen, A. H., Vanin, M., Mortensen, J. J., Thygesen, K. S., and Jacobsen, K. W. (2009). Localized atomic basis set in the projector augmented wave method. *Phys. Rev. B*, 80(19):195112.

Lehtola, S., Jónsson, E. Ö., and Jónsson, H. (2016). Effect of Complex-Valued Optimal Orbitals on Atomization Energies with the Perdew–Zunger Self-Interaction Correction to Density Functional Theory. *J. Chem. Theory Comput.*, 12(9):4296–4302.

Lehtola, S. and Jónsson, H. (2014). Variational, Self-Consistent Implementation of the Perdew–Zunger Self-Interaction Correction with Complex Optimal Orbitals. *J. Chem. Theory Comput.*, 10(12):5324–5337.

Leiserson, C. E., Thompson, N. C., Emer, J. S., Kuszmaul, B. C., Lampson, B. W., Sanchez, D., and Schardl, T. B. (2020). There's plenty of room at the Top: What will drive computer performance after Moore's law? *Science*, 368(6495):eaam97444.

Levi, G., Ivanov, A. V., and Jónsson, H. (2020a). Variational calculations of excited states via direct optimization of the orbitals in DFT. *Faraday Discuss.*, 224:448–466.

Levi, G., Ivanov, A. V., and Jónsson, H. (2020b). Variational Density Functional Calculations of Excited States via Direct Optimization. *J. Chem. Theory Comput.*, 16(11):6968–6982.

Levi, G., Pápai, M., Henriksen, N. E., Dohn, A. O., and Møller, K. B. (2018). Solution Structure and Ultrafast Vibrational Relaxation of the PtPOP Complex Revealed by ΔSCF-QM/MM Direct Dynamics Simulations. *J. Phys. Chem. C*, 122(13):7100–7119.

Levine, B. G., Ko, C., Quenneville, J., and Martinez, T. J. (2006). Conical intersections and double excitations in time-dependent density functional theory. *Mol. Phys.*, 104(5-7):1039–1051.

Levy, M. (1979). Universal variational functionals of electron densities, first-order density matrices, and natural spin-orbitals and solution of the v-representability problem. *Proc. Natl. Acad. Sci. USA*, 76(12):6062–6065.

Levy, M. and Nagy, Á. (1999). Variational Density-Functional Theory for an Individual

Excited State. *Phys. Rev. Lett.*, 83(21):4361–4364.

Loos, P.-F., Scemama, A., Blondel, A., Garniron, Y., Caffarel, M., and Jacquemin, D. (2018). A Mountaineering Strategy to Excited States: Highly Accurate Reference Energies and Benchmarks. *J. Chem. Theory Comput.*, 14(8):4360–4379.

Maitra, N. T., Zhang, F., Cave, R. J., and Burke, K. (2004). Double excitations within time-dependent density functional theory linear response. *J. Chem. Phys.*, 120(13):5932–5937.

Mališ, M. and Luber, S. (2020). Trajectory Surface Hopping Nonadiabatic Molecular Dynamics with Kohn-Sham ΔSCF for Condensed-Phase Systems. *J. Chem. Theory Comput.*, 16(7):4071–4086.

Mewes, J. M., Jovanović, V., Marian, C. M., and Dreuw, A. (2014). On the molecular mechanism of non-radiative decay of nitrobenzene and the unforeseen challenges this simple molecule holds for electronic structure theory. *Phys. Chem. Chem. Phys.*, 16(24):12393–12406.

Mortensen, J. J., Hansen, L. B., and Jacobsen, K. W. (2005). Real-space grid implementation of the projector augmented wave method. *Phys. Rev. B*, 71(3):035109.

Najfeld, I. and Havel, T. F. (1995). Derivatives of the matrix exponential and their computation. *Adv. Appl. Math.*, 16(3):321–375.

Nguyen, N. L., Colonna, N., Ferretti, A., and Marzari, N. (2018). Koopmans-Compliant Spectral Functionals for Extended Systems. *Phys. Rev. X*, 8(2):021051.

Nocedal, J. and Wright, S. J. (2006). *Numerical Optimization*. Springer, New York, NY, USA, second edition.

Paier, J., Hirschl, R., Marsman, M., and Kresse, G. (2005). The Perdew-Burke-Ernzerhof exchange-correlation functional applied to the G2-1 test set using a plane-wave basis set. *J. Chem. Phys.*, 122(23):234102.

Payne, M. C., Teter, M. P., Allan, D. C., Arias, T. A., and Joannopoulos, J. D. (1992). Iterative minimization techniques for ab initio total-energy calculations: molecular dynamics and conjugate gradients. *Rev. Mod. Phys.*, 64(4):1045–1097.

Pederson, M. R., Heaton, R. A., and Lin, C. C. (1985). Density-functional theory with self-interaction correction: Application to the lithium molecule. *J. Chem. Phys.*, 82(6):2688–2699.

Peng, H. and Perdew, J. P. (2017). Synergy of van der Waals and self-interaction corrections in transition metal monoxides. *Phys. Rev. B*, 96(10):100101.

Perdew, J. P. and Levy, M. (1985). Extrema of the density functional for the energy: Excited states from the ground-state theory. *Phys. Rev. B*, 31(10):6264–6272.

Perdew, J. P. and Zunger, A. (1981). Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B*, 23(10):5048–5079.

Pradhan, E., Sato, K., and Akimov, A. V. (2018). Non-adiabatic molecular dynamics with ΔsCF excited states. *J. Phys. Condens. Matter*, 30(48).

Pritchard, B. P., Altarawy, D., Didier, B., Gibson, T. D., and Windus, T. L. (2019). New Basis Set Exchange: An Open, Up-to-Date Resource for the Molecular Sciences Community. *J. Chem. Inf. Model.*, 59(11):4814–4820.

Pulay, P. (1980). Convergence acceleration of iterative sequences. the case of scf iteration. *Chem. Phys. Lett.*, 73(2):393–398.

Rappoport, D. and Furche, F. (2010). Property-optimized Gaussian basis sets for molecular response calculations. *J. Chem. Phys.*, 133(13):134105.

Rostgaard, C., Jacobsen, K. W., and Thygesen, K. S. (2010). Fully self-consistent GW calculations for molecules. *Phys. Rev. B*, 81(8):1–10.

Runge, E. and Gross, E. K. U. (1984). Density-functional theory for time-dependent systems. *Phys. Rev. Lett.*, 52(12):997–1000.

Schuchardt, K. L., Didier, B. T., Elsethagen, T., Sun, L., Gurumoorthi, V., Chase, J., Li, J., and Windus, T. L. (2007). Basis Set Exchange: A Community Database for Computational Sciences. *J. Chem. Inf. Model.*, 47(3):1045–1052.

Seidl, A., Görling, A., Vogl, P., Majewski, J. A., and Levy, M. (1996). Generalized Kohn-Sham schemes and the band-gap problem. *Phys. Rev. B*, 53(7):3764–3774.

Seidu, I., Krykunov, M., and Ziegler, T. (2015). Applications of Time-Dependent and Time-Independent Density Functional Theory to Rydberg Transitions. *J. Phys. Chem. A*, 119(21):5107–5116.

Shishkin, M., Marsman, M., and Kresse, G. (2007). Accurate quasiparticle spectra from self-consistent GW calculations with vertex corrections. *Phys. Rev. Lett.*, 99(24):14–17.

Stengel, M. and Spaldin, N. A. (2008). Self-interaction correction with Wannier functions. *Phys. Rev. B*, 77(15):155106.

Tozer, D. J. and Handy, N. C. (2000). On the determination of excitation energies using density functional theory. *Phys. Chem. Chem. Phys.*, 2(10):2117–2121.

Van Meer, R., Gritsenko, O. V., and Baerends, E. J. (2014). Physical meaning of virtual Kohn-Sham orbitals and orbital energies: An ideal basis for the description of molecular excitations. *J. Chem. Theory Comput.*, 10(10):4432–4441.

Van Voorhis, T. and Head-Gordon, M. (2002). A geometric approach to direct minimization. *Mol. Phys.*, 100(11):1713–1721.

VandeVondele, J. and Hutter, J. (2003). An efficient orbital transformation method for electronic structure calculations. *J. Chem. Phys.*, 118(10):4365–4369.

Vydrov, O. A., Scuseria, G. E., Perdew, J. P., Ruzsinszky, A., and Csonka, G. I. (2006). Scaling down the Perdew-Zunger self-interaction correction in many-electron regions. *J. Chem. Phys.*, 124(9).

Waldrop, M. M. (2016). The chips are down for Moore's law. *Nature*, 530(7589):144–147.

Wang, B. and Chen, Z. (2004). Magnetic coupling interaction under different spin multiplets in neutral manganese dimer: CASPT2 theoretical investigation. *Chem. Phys. Lett.*, 387(4-6):395–399.

Weber, V., Vandevondele, J., Hutter, J., and Niklasson, A. M. (2008). Direct energy functional minimization under orthogonality constraints. *J. Chem. Phys.*, 128(8).

Weigend, F. and Ahlrichs, R. (2005). Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.*, 7(18):3297.

Yamamoto, S., Tatewaki, H., Moriyama, H., and Nakano, H. (2006). A study of the ground state of manganese dimer using quasidegenerate perturbation theory. *J. Chem. Phys.*, 124(12):1–8.

Yamanaka, S., Ukai, T., Nakata, K., Takeda, R., Shoji, M., Kawakami, T., Takada, T., and Yamaguchi, K. (2007). Density functional study of manganese dimer. *Int. J. Quantum Chem.*, 107(15):3178–3190.

Ye, H.-Z., Welborn, M., Ricke, N. D., and Van Voorhis, T. (2017). σ -SCF:

A direct energy-targeting method to mean-field excited states. *J. Chem. Phys.*, 147(21):214104.

Yeh, A. T., Shank, C. V., and McCusker, J. K. (2000). Ultrafast Electron Localization Dynamics Following Photo-Induced Charge Transfer. *Science*, 289(5481):935–938.

Zhekova, H. R., Seth, M., and Ziegler, T. (2014). A perspective on the relative merits of time-dependent and time-independent density functional theory in studies of the electron spectra due to transition metal complexes. An illustration through applications to copper tetrachloride and plastocyanin. *Int. J. Quantum Chem.*, 114(15):1019–1029.

Ziegler, T., Rauk, A., and Baerends, E. J. (1977). On the calculation of multiplet energies by the Hartree-Fock-Slater method. *Theor. Chim. Acta*, 43(3):261–271.

Article I

Direct Energy Minimization Based on Exponential Transformation in Density Functional Calculations of Finite and Extended Systems

Ivanov A. V., Jónsson E. Ö., Vegge T., Jónsson H.

Submitted to Computer Physics Communication,
arXiv:2101.12597 [physics.comp-ph].

Direct Energy Minimization Based on Exponential Transformation in Density Functional Calculations of Finite and Extended Systems

Aleksei V. Ivanov[a,b], Elvar Ö. Jónsson[a], Tejs Vegge[c], Hannes Jónsson[a]

[a] *Science Institute and Faculty of Physical Sciences, University of Iceland VR-III,107 Reykjavík, Iceland*
[b] *St. Petersburg State University, 199034, St. Petersburg, Russia*
[c] *Department of Energy Conversion and Storage, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark*

Abstract

The energy minimization involved in density functional calculations of electronic systems can be carried out using an exponential transformation that preserves the orthonormality of the orbitals. The energy of the system is then represented as a function of the elements of a skew-Hermitian matrix that can be optimized directly using unconstrained minimization methods. An implementation based on the limited memory Broyden-Fletcher-Goldfarb-Shanno approach with inexact line search and a preconditioner is presented and the performance compared with that of the commonly used self-consistent field approach. Results are presented for the G2 set of 148 molecules, liquid water configurations with up to 576 molecules and some insulating crystals. A general preconditioner is presented that is applicable to systems with fractional orbital occupation as is, for example, needed in the k-point sampling for periodic systems. This exponential transformation direct minimization approach is found to outperform the standard implementation of the self-consistent field approach in that all the calculations converge with the same set of parameter values and it requires less computational effort on average. The formulation of the exponential transformation and the gradients of the energy presented here are quite general and can be applied to energy functionals that are not unitary invariant such as self-interaction corrected functionals.

1. Introduction

There are several different approaches for finding optimal orbitals corresponding to the minimum of an energy functional in the context of Kohn–Sham density functional theory (KS-DFT) [1, 2]. The most commonly used method is based on a self-consistent field (SCF) algorithm consisting of two steps. In

Email addresses: alxvov@gmail.com (Aleksei V. Ivanov), hj@hi.is (Hannes Jónsson)

Preprint submitted to Computer Physics Communications *January 28, 2021*

the first step and for a given density, one finds eigenvalues and eigenfunctions using an iterative algorithm such as the Davidson algorithm [3] or even direct diagonalization of the full Hamiltonian matrix when the size of the basis set is not too large. In the second step, the electron density or Hamiltonian matrix is updated using, for example, direct inversion in the iterative subspace (DIIS) method [4, 5]. The SCF approach is widely used and has proven to be efficient for both finite (molecules/clusters) and extended systems, but can, nevertheless, suffer from convergence problems. Various density and Hamiltonian mixing schemes have been introduced to address such cases [6, 7]. As a result, the user of typical software developed for KS-DFT calculations is often presented with the task of choosing values of various parameters and select between various types of eigensolvers. Systems with similar chemical and physical properties may even call for different choices. A further problem of the SCF method in calculations of ground electronic states is that it may converge on a saddle point of the energy surface rather than a minimum [8].

Another approach to this optimization problem is based on direct minimization of the energy with respect to the electronic degrees of freedom [9, 10, 11, 12, 13, 14, 15, 16, 17, 18]. The challenge then is to incorporate the constraint of orthonormality of the orbitals (the single electron wave functions). One way to approach this is to follow the energy gradient projected on the subspace tangent to the orbitals [10, 11]. After such an adjustment of the orbitals within this tangent space, the orthonormality constraints will be violated and, therefore, an explicit orthonormalization of the orbitals needs to be applied after each iteration. This approach is often used in calculations with a plane wave basis set. Alternatively, when the basis set is compact, as in calculations using linear combination of atomic orbitals, a unitary transformation can be applied to a set of orthonormal reference orbitals that includes all occupied and virtual orbitals, and the energy is then minimized by optimizing the elements of the transformation matrix. The orthonormality constraints will then be satisfied, but, due to the constraints imposed by the unitary matrix, the energy is defined on a curved space. As a result, minimization algorithms need to be modified to take the curvature into account. This can be achieved by performing a line search along geodesics [19]. Alternatively, the unitary matrix can be parameterized using an exponential transformation [9, 12, 14] in which case the energy becomes a function of the elements of a skew-Hermitian matrix in linear space. Well-established, unconstrained minimization strategies can then be applied including inexact line searches that can give robust convergence. We will refer to this approach as exponential transformation direct minimization (ETDM). Furthermore, it has been used in calculations of molecules using KS-DFT [15] and previously in the context of Hartree-Fock theory [9, 20, 21, 22]. There, the occupation numbers for the orbitals have been restricted to integers so that unitary invariance with respect to rotation within the space of occupied orbitals is ensured. Preconditioners to accelerate convergence have been presented for such systems and found to be important in order to achieve good performance [9, 15].

In this article, a generalization and efficient implementation of the ETDM approach is presented as well as applications to both finite and extended sys-

tems. The method can be applied to systems with fractional occupation, for example, where k-point sampling of the Brillouine zone (BZ) is carried out. The formulation presented here is also applicable to energy functionals that are not unitary invariant, such as self-interaction corrected functionals [23]. Tests of the performance of this ETDM implementation and comparison with the SCF method including density mixing are carried out for the G2 set (a total of 148 molecules), liquid configurations consisting of up to 576 water molecules and several insulating crystals.

The article is organised as follows. In section 2, the ETDM method is formulated in a general way and equations provided for the derivative of the energy with respect to the matrix elements in the exponential transformation. In section 3, an efficient preconditioner is presented, applicable to systems with non-integer occupation numbers, as well as methods for evaluating the gradient of the energy and ways to choose the search direction as well as step-length in an inexact line-search procedure. In section 4, performance tests are presented with comparison to conventional SCF calculations. Finally, discussion and conclusions are presented in section 5.

2. General formulation

In KS-DFT, the energy functional is

$$E = \sum_{i,\mathbf{k}} f_i(\mathbf{k}) \int d^3\mathbf{r} \frac{|\nabla \phi_{i\mathbf{k}}(\mathbf{r})|^2}{2} + \int d^3\mathbf{r} \rho(\mathbf{r}) v_{ext}(\mathbf{r}) + \tag{1}$$

$$+ \frac{1}{2} \iint d^3\mathbf{r}\, d^3\mathbf{r}' \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + E_{xc}[\rho(\mathbf{r})]. \tag{2}$$

where the ϕ are orbitals of the non-interacting electron system that has total electron density

$$\rho(\mathbf{r}) = \sum_{i,\mathbf{k}} f_i(\mathbf{k}) |\phi_{i\mathbf{k}}(\mathbf{r})|^2, \tag{3}$$

equal to that of the interacting electron system, the $f_i(\mathbf{k})$ are orbital occupation numbers for the k-th point of the BZ with $0 \leq f_i(\mathbf{k}) \leq 1$, $v_{ext}(\mathbf{r})$ is the external potential corresponding to electron-nuclei interaction, and E_{xc} is the exchange-correlation energy. The orbitals are expanded in terms of a possibly non-orthogonal basis set consisting of M basis functions

$$\phi_{i\mathbf{k}}(\mathbf{r}) = \sum_{\mu=1}^{M} O_{\mu i}(\mathbf{k}) \chi_{\mu\mathbf{k}}(\mathbf{r}), \tag{4}$$

and the task is to find optimal values of the coefficients $O_{\mu i}(\mathbf{k})$ that minimize the energy $E[\{O(\mathbf{k})\}_\mathbf{k}]$ subject to the orthonormality constraints:

$$O^\dagger(\mathbf{k}) S(\mathbf{k}) O(\mathbf{k}) = I \quad \mathbf{k} \in BZ, \tag{5}$$

with $S_{\mu\nu}(\mathbf{k}) = \int \chi^*_{\mu\mathbf{k}}(\mathbf{r})\chi_{\nu\mathbf{k}}(\mathbf{r})\, d\mathbf{r}$ being the overlap matrix.

The basis functions for periodic systems are Bloch states and in a localised basis set approach they can be written as

$$\chi_{\mu\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{N}} \sum_{\mathbf{R}} \exp(i\mathbf{k}\cdot\mathbf{R})\eta_\mu(\mathbf{r}-\mathbf{R}-\mathbf{d}_\mu) \tag{6}$$

where $\eta_\mu(\mathbf{r}-\mathbf{R}-\mathbf{d}_\mu)$ is an atomic orbital centered on an atom in the simulated cell. The subscript μ enumerates the atomic orbitals and $\mathbf{R}$ belongs to the Bravais lattice. An initial guess for the orbitals is expressed as a linear combination of the basis functions

$$\psi_{m\mathbf{k}}(\mathbf{r}) = \sum_{\mu=1..M} C_{\mu m}(\mathbf{k})\chi_{\mu\mathbf{k}}(\mathbf{r}). \tag{7}$$

Given an initial guess for the orbitals, $C_{\mu m}(\mathbf{k})$, which we will refer to as the reference orbitals, the optimal orbital coefficients $O_{\mu m}(\mathbf{k})$ that provide minimal energy can be found through a unitary transformation as

$$O(\mathbf{k}) = C(\mathbf{k})e^{A(\mathbf{k})} \tag{8}$$

where $A(\mathbf{k})$ is a skew-Hermitian matrix, $A(\mathbf{k})^\dagger = -A(\mathbf{k})$. For a set of N_k vectors used to represent the BZ, a set of matrices $\{A(\mathbf{k})\}_\mathbf{k}$ is needed. For a given set of reference orbitals, a set of unitary matrices, $U(\mathbf{k}) = \exp(A(\mathbf{k}))$, exists so that the reference orbitals are transformed to the optimal orbitals. Thus, the ground-state energy of the system is a function of the upper triangular elements of a set of matrices $A(\mathbf{k})$,

$$E[n] = E[\{a_{11}, \ldots, a_{1M}, a_{22}, \ldots, a_{2M}, \ldots, a_{MM}\}_\mathbf{k}] \tag{9}$$

where $a_{ij} = (A)_{ij}$ and $\mathbf{k}$ denotes the set of N_k vectors. The real part of the diagonal elements of the matrices are zeros and therefore the energy is a function of $N_k M^2$ variables. There are $M(M-1)/2$ real elements and $M(M+1)/2$ imaginary elements for every k-point. The energy needs to be minimized with respect to the real and imaginary parts of the matrix elements $\{a_{ij}(\mathbf{k})\}_{i\leq j}$. Introducing the derivative

$$\frac{\partial}{\partial a_{ij}(\mathbf{k})} = \frac{1}{2}\left(\frac{\partial}{\partial \mathrm{Re}(a_{ij}(\mathbf{k}))} - i\frac{\partial}{\partial \mathrm{Im}(a_{ij}(\mathbf{k}))}\right) \tag{10}$$

the gradient of the energy can be evaluated as

$$\frac{\partial E}{\partial a_{ij}(\mathbf{k})} = \sum_{\mu\nu} H_{\mu\nu}(\mathbf{k})\frac{\partial \rho_{\mu\nu}(\mathbf{k})}{\partial a_{ij}(\mathbf{k})} \tag{11}$$

where the Hamiltonian matrix is

$$H_{\mu\nu}(\mathbf{k}) = \int d\mathbf{r}\, \chi^*_{\mu\mathbf{k}}(\mathbf{r})\left(-\frac{1}{2}\nabla + v(\mathbf{r})\right)\chi_{\nu\mathbf{k}}(\mathbf{r}). \tag{12}$$

Here, $v(\mathbf{r})$ is the single electron Kohn-Sham potential, and the density matrix is given in terms of the optimal coefficient matrix as

$$\rho_{\mu\nu}(\mathbf{k}) = \sum_m f_m(\mathbf{k}) O_{\mu m}(\mathbf{k}) \overline{O}_{\nu m}(\mathbf{k}). \tag{13}$$

By defining the commutator

$$L_{mk}(\mathbf{k}) = \left[F(\mathbf{k}), H(\mathbf{k})\right]_{mk}, \tag{14}$$

where $H(\mathbf{k})$ is the Hamiltonian matrix represented in terms of the optimal orbitals

$$H(\mathbf{k})_{mk} = \sum_{\mu\nu} \overline{O}_{\mu m}(\mathbf{k}) H_{\mu\nu}(\mathbf{k}) O_{\nu k}(\mathbf{k}),$$

and $F(\mathbf{k})$ is a diagonal matrix with occupation numbers $f_m(\mathbf{k})$ as diagonal elements, the derivatives in Eq. (11) can be written as

$$\frac{\partial E}{\partial a_{ij}(\mathbf{k})} = \frac{2-\delta_{ij}}{2} \left(\int_0^1 e^{tA(\mathbf{k})} L(\mathbf{k}) e^{-tA(\mathbf{k})}\, dt \right)_{ji}. \tag{15}$$

For the optimal orbitals, the gradient $\partial E / \partial a_{ij}(\mathbf{k})$ must be zero so

$$\int_0^1 e^{tA(\mathbf{k})} L(\mathbf{k}) e^{-tA(\mathbf{k})}\, dt = 0, \quad \mathbf{k} \in BZ. \tag{16}$$

These non-linear equations can be used to find the skew-Hermitian matrix that provides the energy minimum. For the remainder of this article, the k-point index $\mathbf{k}$ is omitted for simplicity.

Eq. (15) is general and can be applied to an objective function that depends explicitly on the orbitals as well as the total density, but then the definition of L needs to be be changed accordingly. For example, for the Perdew-Zunger self-interaction correction (PZ-SIC) [23], the matrix L for a single k-point calculation is

$$L_{mk} = \left[F, H\right]_{mk} + f_k \overline{V}_{km} - f_m V_{mk}, \tag{17}$$

where V_{km} is a matrix element of the SIC potential§

$$V_{mk} = \sum_{\mu\nu} \overline{O}_{\mu m} V^k_{\mu\nu} O_{\nu k}, \tag{18}$$

$$V^k_{\mu\nu} = \int \chi^*_\mu(\mathbf{r}) \left[\int d^3\mathbf{r}' \frac{\rho_k(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + v_{xc}(\rho_k(\mathbf{r})) \right] \chi_\nu(\mathbf{r}) d\mathbf{r}. \tag{19}$$

Equation (16) can be expanded in a series as

$$\int_0^1 e^{tA} L e^{-tA}\, dt = L + \frac{1}{2!}\left[A, L\right] + \frac{1}{3!}\left[A, \left[A, L\right]\right] + \ldots \tag{20}$$

If $\|L\| \gg \frac{1}{2}\|\left[A, L\right]\|$, which corresponds to $\|A\| \ll 1$ since $\frac{1}{2}\|\left[A, L\right]\| \geq \|A\|\|L\|$, then the first term on the right hand side can be used to estimate the gradient.

This limit of 'small rotations' corresponds to the geometric approach used by Van Voorhis and Head-Gordon [15] and has also been used in the context of orbital-density dependent functionals [24, 25, 26]. The higher order terms can also be included to increase the accuracy of the gradient estimate, but each iteration then requires more computational effort.

The minimization procedure is performed with respect to the real and imaginary parts of matrix elements using the energy gradient given by Eq. (15)

$$\frac{\partial E}{\partial \mathrm{Re}(a_{ij})} = 2\mathrm{Re}\left(\frac{\partial E}{\partial a_{ij}}\right) \tag{21}$$

and

$$\frac{\partial E}{\partial \mathrm{Im}(a_{ij})} = -2\mathrm{Im}\left(\frac{\partial E}{\partial a_{ij}}\right). \tag{22}$$

Computational algorithms for the evaluation of the the matrix exponential and gradient of the energy are presented in Sec. 3.4

3. Algorithms and Computational Parameters

In order to find the optimal orbitals, O, corresponding to minimal energy, the appropriate exponential transformation of the reference orbitals, C,

$$O = Ce^{A} \tag{23}$$

needs to be determined. The reference orbitals can be chosen to be any set of orthonormal orbitals spanned by the basis set and they are held fixed during the minimization of the energy for a given number of steps while only the matrix A is varied. The closer the reference orbitals are to the optimal orbitals, the faster the iterative procedure will converge.

A line search method has been implemented where the $(k+1)$th iteration step is

$$\vec{a}^{\,(k+1)} = \vec{a}^{\,(k)} + \alpha^{\,(k)}\,\vec{p}^{\,(k)}. \tag{24}$$

Here, $\vec{a}^{\,(k+1)}$ is a vector consisting of the real and imaginary part of the upper triangular elements of matrix A at the kth step of the minimization algorithm,

$$\begin{aligned}\vec{a} = (&\mathrm{Re}(a_{12}), \ldots, \mathrm{Re}(a_{1M}),\\ &\mathrm{Re}(a_{23}), \ldots, \mathrm{Re}(a_{2M}), \ldots, \mathrm{Re}(a_{M-1M}),\\ &\mathrm{Im}(a_{11}), \mathrm{Im}(a_{12}), \ldots, \mathrm{Im}(a_{1M}),\\ &\mathrm{Im}(a_{22}), \mathrm{Im}(a_{23}), \ldots, \mathrm{Im}(a_{2M}), \ldots, \mathrm{Im}(a_{MM}))^{T},\end{aligned} \tag{25}$$

and $\vec{p}_k$ is the search direction while α_k is the step length.

3.1. Choice of search direction

The search direction can be chosen according to the steepest descent method, various Quasi-Newton methods, or nonlinear conjugate gradient (CG) methods. The calculation of the search direction involves algebraic operations associated with the particular method plus the evaluation of the energy and gradient for the given energy functional. The dimensionality of the minimization problem scales as NM, where N is the number of occupied orbitals and M is the number of basis functions. While Quasi-Newton methods such as the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm require fewer iterations than limited-memory BFGS (L-BFGS) or CG, the algebraic operations become a bottleneck even for systems of moderate size (the BFGS algorithm scales as $\mathcal{O}(N^2M^2)$ [27]). However, every iteration of the L-BFGS algorithm, in which the approximate inverse Hessian matrix is updated, can be computed with the cost of $\mathcal{O}(mNM)$ operations, where m is the number of previous steps stored in memory. In the present implementation, the L-BFGS algorithm as described in Ref. [27] is used and $m = 3$ in the benchmark calculations.

3.2. Choice of step length

The step length α_k is chosen in such a way that it satisfies the strong Wolfe conditions [28, 29, 27]

$$E(\vec{a}^{\,(k)} + \alpha^{\,(k)}\vec{p}^{\,(k)}) \leq E(\vec{a}^{\,(k)}) + c_1\alpha^{\,(k)}\nabla_{\vec{a}}E(\vec{a}^{\,(k)}) \cdot \vec{p}^{\,(k)} \tag{26}$$

and

$$|\nabla E(\vec{a}^{\,(k)} + \alpha^{\,(k)}\vec{p}^{\,(k)}) \cdot \vec{p}^{\,(k)}| \leq c_2|\nabla_{\vec{a}}E(\vec{a}^{\,(k)}) \cdot \vec{p}^{\,(k)}| \tag{27}$$

with $0 < c_1 < c_2 < 1$. A trial step of $\alpha^{\,(k)} = 1$ is always used first to test the conditions. After several iterations, a step length of 1 guarantees that the strong Wolfe conditions are satisfied in the L-BFGS algorithm [27]. This is appealing since it reduces the number of energy and gradient calculations which are computationally most intensive in KS-DFT calculations. If $\alpha^{\,(k)} = 1$ is not satisfied by the strong Wolfe conditions, then the inexact line search based on the interpolation of the energy along the search direction is used [27]. When the energy of the system is evaluated, the KS-DFT potential needs to be obtained and, as a result, there is little additional effort involved in evaluating the gradient. Therefore, the energy along the search direction is always interpolated by a cubic function using information about the energy values and gradient at the boundaries of the search interval $[a, b]$. Alongside the strong Wolfe conditions, approximate Wolfe conditions are also checked [30] at the minimum of the interpolated cubic function

$$(2\delta - 1)\nabla_{\vec{a}}E(\vec{a}^{\,(k)})^T\vec{p}^{\,(k)} \geq \nabla_{\vec{a}}E(\vec{a}^{\,(k)} + \alpha^{\,(k)}\vec{p}^{\,(k)}) \cdot \vec{p}^{\,(k)} \geq \sigma\nabla_{\vec{a}}E(\vec{a}^{\,(k)}) \cdot \vec{p}^{\,(k)}, \tag{28}$$

and the condition

$$E(\vec{a}^{\,(k)} + \alpha^{\,(k)}\vec{p}^{\,(k)}) \leq E(\vec{a}^{\,(k)}) + \epsilon|E(\vec{a}^{\,(k)})| \tag{29}$$

where $\delta < \min\{0.5, \sigma\}$, $0 < \sigma < 1$ and ϵ is a small fixed number. Thus, the line search algorithm is terminated when either the strong Wolfe conditions of Eqs. (26)-(27) or the approximate Wolfe conditions of Eq. (28) along with the condition in Eq (29) holds. The parameter values are set to [27, 30]

$$c_1 = 10^{-4}, c_2 = 0.9, \delta = 0.1, \sigma = 0.9, \epsilon = 10^{-6}. \tag{30}$$

3.3. Preconditioning

A preconditioner speeds up convergence of this iterative algorithm. It is constructed as the inverse of an approximate Hessian matrix that can be obtained by taking the derivative of a linear expansion of the gradient (Eq. (15)) with respect to the skew-Hermitian matrix, and neglecting first order derivatives of the effective potential. For the real valued case, the Hessian can be approximated as

$$\begin{aligned}
\frac{\partial^2 E}{\partial a_{ij} \partial a_{lm}} \approx & \delta_{il} H_{jm} (f_l + f_i - f_j - f_m) && (31)\\
& + \delta_{jl} H_{im} (f_m + f_i - f_l - f_j) && (32)\\
& + \delta_{jm} H_{li} (f_m - f_i - f_l + f_j) && (33)\\
& + \delta_{im} H_{lj} (f_l - f_m - f_i + f_j) && (34)\\
& + \beta_{ij} \delta_{il} \delta_{jm} && (35)
\end{aligned}$$

where the matrix β_{ij} must be chosen according to the following two principles: (1) the approximate Hessian must be positive definite, and (2) it must provide a good estimate of the true Hessian along the search direction such that a step size of 1 satisfies the strong Wolfe conditions.

If the orbitals are chosen as eigenvectors of the Hamiltonian then the approximate Hessian is diagonal

$$\frac{\partial^2 E}{\partial^2 a_{ij}} = -2(\epsilon_{ii} - \epsilon_{jj})(f_i - f_j) + \beta_{ij}. \tag{36}$$

The first term on the right hand side coincides with the preconditioner that has previously been used for molecular systems with integer occupation numbers [9]. There, an extra term was added in cases of degeneracy, $\epsilon_{ii} = \epsilon_{jj}$, but here the initial approximation of the Hessian in the L-BFGS algorithm [27] β_{ij} is used. Since the approximate Hessian is diagonal, the preconditioner is simply

$$P_{ij} = \frac{1}{-2(\epsilon_{ii} - \epsilon_{jj})(f_i - f_j) + \beta_{ij}}. \tag{37}$$

In the present implementation, the preconditioner is updated iteratively and for iteration k it is

$$P_{ij}^{(k)} = \frac{1}{-2(1-\gamma)(\epsilon_{ii} - \epsilon_{jj})(f_i - f_j) + \gamma \beta_{ij}^{(k)}}, \tag{38}$$

where

$$\beta_{ij}^{(k)} = \frac{\|\nabla_{\vec{a}} E(\vec{a}^{\,(k)}) - \nabla_{\vec{a}} E(\vec{a}^{\,(k-1)})\|^2}{(\vec{a}^{\,(k)} - \vec{a}^{\,(k-1)}) \cdot (\nabla_{\vec{a}} E(\vec{a}^{\,(k)}) - \nabla_{\vec{a}} E(\vec{a}^{\,(k-1)}))}. \tag{39}$$

The parameter γ in Eq.(38) is a number that determines the mixing of the two approximate Hessians: the one obtained from a linear expansion of the gradient, Eq.(20), and the one based on the LBFGS estimate, Eq.(39). In the calculations presented here, $\gamma = 0.25$ was found empirically to give a good compromise between the rate of convergence and robustness. When k-point sampling is included for the periodic systems, $\beta_{ij}^{(k)}$, needs to be multiplied by the numerical weight of the corresponding k-point.

With this preconditioner, a step length of 1 is almost always accepted and it works well for both finite and extended systems. It is used for both the real and imaginary parts of the skew-Hermitian matrix. We note that the eigenvalues in Eq. (38) are not updated at every iteration of the minimization algorithm but only at the beginning, thereby avoiding the costly diagonalization of the Hamiltonian matrix at each step.

3.4. Evaluation of the matrix exponential and energy gradient

The evaluation of the exponential of the skew-Hermitian matrix, $\exp(A)$, is carried out using eigendecomposition of iA. Let Ω be a diagonal, real-valued matrix with elements corresponding to the eigenvalues of the matrix iA and let U be a column matrix of the eigenvectors of iA. Then the matrix exponential of A is

$$\exp(A) = U \exp(-i\Omega) U^{\dagger}. \tag{40}$$

This computation requires diagonalization of a $M \times M$ matrix and becomes a computational bottleneck for large systems. However, for unitary invariant energy functionals (such as Kohn-Sham functionals), Hutter *et.al.* [12] have shown that A can be parametrised without loss of generality as

$$A = \begin{pmatrix} 0 & A_{ov} \\ -A_{ov}^{\dagger} & 0 \end{pmatrix}, \tag{41}$$

where A_{ov} is a $N \times (M - N)$ matrix (N - number of occupied states) and the matrix exponential can be calculated as

$$\exp(A) = \begin{pmatrix} \cos(P) & P^{-1/2}\sin(P^{1/2})A_{ov} \\ -A_{ov}^{\dagger}P^{-1/2}\sin(P^{1/2}) & I_{M-N} + A_{ov}^{\dagger}\cos(P^{1/2} - I_N)P^{-1}A_{ov}) \end{pmatrix}, \tag{42}$$

where $P = A_{ov}A_{ov}^{\dagger}$. In this case the computational effort scales as $\mathcal{O}(N^2 M)$.

An alternative and more general approach is provided by the scaling and squaring algorithm based on the equation

$$\exp(A) = \exp(A/2^m)^{2^m} \tag{43}$$

and on $[q, q]$ Páde approximant to the matrix $\exp(A/2^m)$, where m and q are positive integer constants [31]. The algorithm of Al-Mohy and Higham is used

here [32, 33]. The two approaches are compared in the benchmark calculations presented below.

If the matrix exponential is evaluated using the eigendecomposition of iA, then one can calculate the gradient of the energy using the matrices U and Ω as

$$G^T = U\left(\left(U^\dagger L U\right) \otimes D\right) U^\dagger, \tag{44}$$

where the matrix D is

$$D_{ij} = \frac{e^{-i(\Omega_{ii}-\Omega_{jj})} - 1}{i(\Omega_{ii} - \Omega_{jj})} \tag{45}$$

and the matrix G is

$$G_{ij} = \frac{\partial E}{\partial a_{ij}} \frac{2}{(2 - \delta_{ij})}. \tag{46}$$

However, due to the sparsity of the matrix A and if the norm is $\|A\| \ll 1$, the gradients can be evaluated more efficiently using only the first term on the right hand side of Eq. (20)

$$G \approx L^T. \tag{47}$$

If the norm of the matrix A is larger than 1, then the reference orbitals can be updated $C \leftarrow C\exp(A)$ in which case $A \leftarrow 0$ and then Eq. (47) can be used. Namely, during the iterative process,

$$O^{(k)} = C\exp(A^{(k)})$$

check if $\|A^{(k)}\| > \epsilon$ then set $C' = C\exp(A^{(k)})$, $A^{(k)} = 0$ and continue with

$$O^{(k+1)} = C'\exp(A^{(k+1)}).$$

For small systems, the performance is similar for the various methods for evaluating the matrix exponential and energy gradient since the calculation of the effective Kohn-Sham potential and the total energy then dominates the computational effort. For larger systems, a difference in performance becomes evident, as illustrated below for configurations of liquid water with up to 576 molecules.

3.5. Implementation and parameter values

We have implemented the ETDM algorithm using a numerical localized atomic basis set and the projector augmented-wave formalism (PAW) [34] to take into account the frozen, inner electrons of the atoms within the open-source GPAW software [35]. An SCF algorithm based on the eigendecomposition of the Hamiltonian in a localised atomic basis set representation is already available there and is frequently used in KS-DFT calculations [36]. To compare the efficiency of the two approaches, single-point ground-state energy calculations are performed for the G2 [37] data set of small molecules, five ionic solids, as well as liquid water configurations including 32, 64, 128, 256, 384 and 576 molecules subject to periodic boundary conditions. The double-zeta polarized basis set

(which is the default basis set in GPAW) and the generalized gradient approximation (GGA) parametrized by Perdew-Burke-Ernzerhof [38] is used. An initial guess for the orbitals is taken to be the eigenvectors of the Hamiltonian obtained from a superposition of atomic densities.

Convergence is considered achieved for both the SCF and the ETDM methods when the inequality

$$\frac{1}{N_e}\sum_{i=1}^{N_b}\int d\,\mathbf{r} f_i|\hat{H}_{KS}\psi_i(\mathbf{r}) - \sum_{j=1}^{N_b}\lambda_{ij}\psi_j(\mathbf{r})|^2 \;<\; 10^{-10}\ \mathrm{eV}^2 \tag{48}$$

is satisfied. In the equation above, the λ_{ij} are Lagrange multipliers and for an SCF algorithm this is a diagonal matrix. N_b is the number of occupied orbitals. Default values in GPAW are used, for example the Pulay density mixing parameters. We note that in cases where the SCF method fails to converge, it could in principle be made to converge by using, for example, other, non-default values of the density mixing parameter. Failure to reach convergence here means that convergence is not obtained in the default maximum number of iteration steps, which is 333.

4. Results

4.1. Molecules

The average number of energy and gradient evaluations for the ETDM method and the average number of energy and diagonalization calculations for the SCF method are presented in Table 1 and Fig. 1. The ETDM method converges for all the 148 molecules in the G2 set using the parameter values specified in Sec. 3. The SCF method, however, fails to converge for five of the molecules: CH, SH, ClO, NO, and OH. These five molecules are also challenging for the ETDM method as it requires more iterations to reach convergence there than the average for the whole G2 set (see Fig. 1). For the molecules where SCF converges, it requires a similar number of iterations as ETDM. On average 18 and 17 iterations are required by the SCF and ETDM methods, respectively.

The reason for the lack of convergence for SCF and slow convergence of ETDM in the five problematic cases could be the presence of nearby saddle points or near-degenerate higher energy states. In the SCF calculations, the orbitals obtained from the diagonalization of the Hamiltonian matrix at subsequent iterations can 'jump' between different energy surfaces or oscillate around a saddle point. Analogous convergence issues for the DIIS method have been reported for the G2 molecular set and transition metal complexes [15].

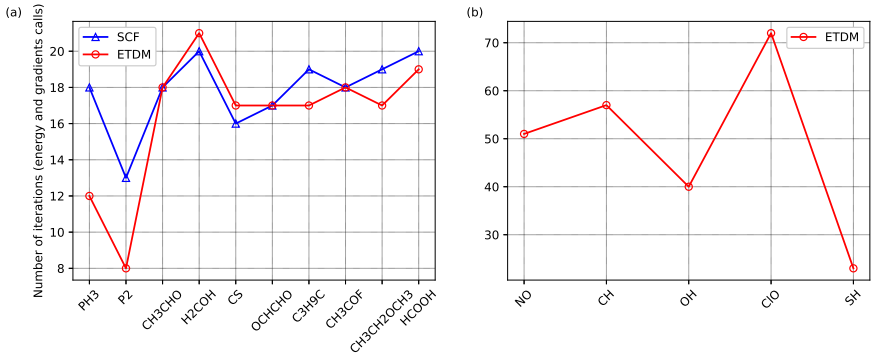

Figure 1: (a) Number of SCF iterations and energy/gradient evaluations in the exponential transform direct minimization needed to reach convergence according to criterion Eq. (48) for a representative set of 10 molecules from the G2 set. (b) Energy/gradient evaluations in the exponential transform direct minimization for the molecules for which the SCF method failed to converge.

Table 1: Comparison of the performance of the exponential transform direct minimization, ETDM, and self-consistent field, SCF, methods for the G2 set of molecules (a total of 148 molecules). The average number of energy and gradient evaluations is reported for the former method, but the average number of energy and diagonalization calculations for the latter (in both cases denoted e/g(d)). In the column labeled ETDM*, the five molecules for which the SCF calculations did not converge are excluded.

	SCF	ETDM	ETDM*
average e/g(d)	18	17	16
min e/g(d)	12	6	6
max e/g(d)	26	72	25
did not converge	5	-	-

For these small molecules, the evaluation of the matrix exponential and energy gradient, i.e. the diagonalization of the Hamiltonian matrix, is not the dominant computational effort. The various algorithms presented in Sec 3.4 therefore involve similar computational effort.

4.2. Periodic Systems

As examples of extended systems subject to periodic boundary conditions, calculations have been carried out for five crystalline solids: NaCl, NaF, LiCl, LiF and MgO. A cubic unit cell is chosen consisting of 8 atoms and Γ-centered $3 \times 3 \times 3$ Monkhorst-Pack meshes are used for the BZ sampling. The lattice constants are set to the optimal values obtained from PBE calculations [39].

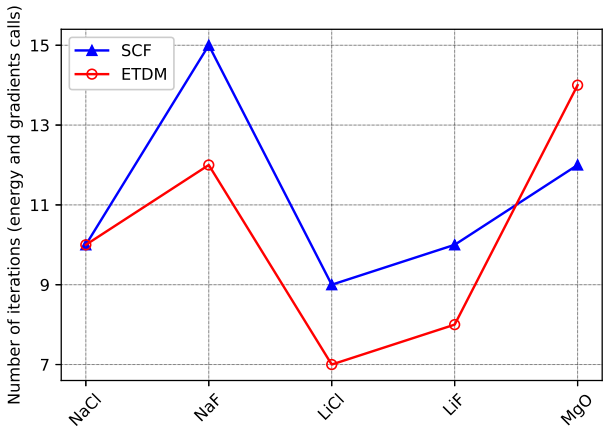

Figure 2: Number of SCF iterations and energy/gradient evaluations in the exponential transform direct minimization needed to reach convergence according to criterion Eq. (48) for NaCl, NaF, LiCl, LiF and MgO crystals.

The number of iterations required to reach convergence is presented in Fig. 2. The results show that the ETDM and the SCF algorithms have similar rate of convergence for these systems. This is an important test of the preconditioner given in equation (38) and shows that it is suitable for solids as well as molecules.

Tests were also carried out for another set of extended systems representing snapshots of liquid water. The systems contain 32, 64, 128, 256, 384 and 576 water molecules subject to periodic boundary conditions. The efficiency of the two approaches for evaluating the matrix exponential in the ETDM method discussed in Sec 3.4 is compared, also in relation to SFC, and reported in Fig.3. One of the approaches is based on Eq. (42) and makes use of the fact that the energy is invariant with respect to unitary rotations of the occupied orbitals. In this case, the computation of the matrix exponential requires diagonalization of an $N \times N$ matrix and involves less computational time as compared to the SCF algorithm where the first N eigenvectors of a $M \times M$ Hamiltonian matrix need to be calculated. The other approach, the scaling and squaring algorithm Eq.(43), is more general and does not rely on the parameterization of the skew-Hermitian matrix based on Eq. (43). For dense matrices, this approach is generally slower than the one based on eigendecomposion of the skew-Hermitian matrix Eq. (40), but for sparse matrices this algorithm can outperform the eigendecomposion approach. The energy gradient is calculated according to Eq. (47).

The ratio of the CPU time required by calculations using the SCF method and the ETDM method is shown as a function of the number of water molecules in Fig. 3. When the matrix exponential is evaluated using Eq. (42) the ETDM method outperforms SCF by a factor of two if more than 200 water molecules are included in the system. Also, the more general implementation of ETDM using

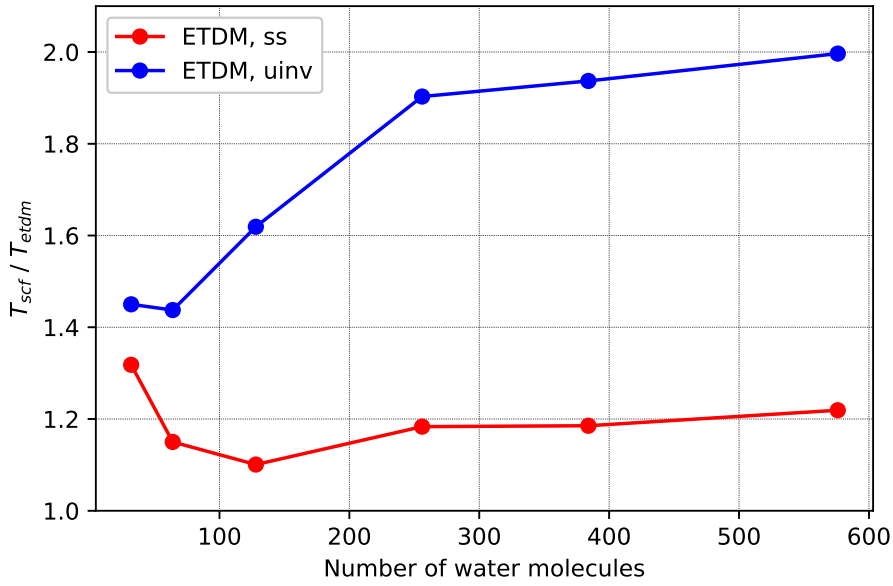

Figure 3: Ratio of the CPU time used by the SCF method and the exponential transform direct minimization, ETDM, method based on either the scaling and squaring algorithm, Eq. (43) (ss, red curve) or the evaluation of the matrix exponential by diagonalization, Eq. (40), (uinv, blue curve), as a function of the number of water molecules in liquid configurations subject to periodic boundary conditions. For the largest system, the direct minimization based on matrix diagonalization outperforms the SFC method by a factor of two while the implementation based on the scaling and squaring algorithm is 20% faster than SCF.

scaling and squaring, Eq. (43), is faster than SCF by 20% for these relatively large systems. It has the advantage of being applicable to energy functionals lacking unitary invariance, unlike the SCF algorithm.

5. Discussion and Conclusion

The main advantage of the ETDM implementation presented here, based on a general preconditioner, L-BFGS algorithm and inexact line search is robustness. For small molecules the computational effort is similar to the standard SCF approach when the latter converges, but the ETDM is found to converge for all the molecules in the G2 set with the same set of parameter values, a set that also works for extended liquid configurations and insulating solids. This demonstrates the transferability of the ETDM algorithm as implemented here. For the large systems considered here, liquid water configurations with 200 and up to 576 molecules, the ETDM outperforms the direct SCF method up to by a factor of two when special parametrization of skew-Hermitian matrix is used and by around 20% when the more general scaling and squaring method is used. The latter is more general and can be applied to any type of orbital dependent energy functional such as self-interaction corrected functionals [23].

The ETDM method involves minimization of the energy with respect to the elements of a skew-Hermitian matrix and, therefore, the number of degrees of freedom scales as M^2, where M is the number of basis functions. However, for energy functionals that are unitary invariant with respect to the occupied orbitals, the skew-Hermitian matrices can be parametrized using $N \times (M - N)$ degrees of freedom,[12] where N is the number of occupied orbitals. Therefore, taking into account the sparsity of the matrices, the algorithm can be implemented in such a way that the computational effort scales as $\mathcal{O}(N^2M)$. The scaling and squaring algorithm for evaluating the matrix exponential is not as efficient but is more generally applicable and can still outperform the SCF method as was found for the large liquid water configurations.

Future work will involve generalization of the ETDM method to finite temperature KS-DFT, i.e. thermal smearing, where an additional inner loop for variational optimization of the occupation numbers is included, analogous to the direct minimization method used in ensemble DFT [13]. This is needed for calculations of metallic systems. A more efficient preconditioner could also likely be developed, especially for orbital density dependent functionals. Finally, we point out that the ETDM method is also useful in other types of electronic structure calculations, such as studies of excited states [40, 41].

6. Acknowledgement

The authors thank Gianluca Levi for fruitful discussion and valuable comments on the manuscript. This work was supported by the University of Iceland Research Fund and the Icelandic Research Fund (grant no. 174082-053). AVI is supported by a doctoral fellowship from the University of Iceland. HJ, AVI and

EÖJ thank the Department of Energy Conversion and Storage at the Technical University of Denmark for hospitality during an extended visit and access to computational resources.

[1] P. Hohenberg, W. Kohn, Inhomogeneous electron gas, Phys. Rev. 136 (1964) B864.

[2] W. Kohn, L. J. Sham, Self-consistent equations including exchange and correlation effects, Phys. Rev. 140 (1965) 1133.

[3] E. Davidson, The iterative calculation of a few of the lowest eigenvalues and corresponding eigenvectors of large real-symmetric matrices, J. Comput. Phys. 17 (1975) 87.

[4] P. Pulay, Convergence acceleration of iterative sequences. the case of scf iteration, Chem. Phys. Lett. 73 (1980) 393.

[5] P. Pulay, Improved scf convergence acceleration, J. Comput. Chem. 3 (1982) 556.

[6] G. Kresse, J. Furthmüller, Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set, Phys. Rev. B 54 (1996) 11169.

[7] A. J. Garza, G. E. Scuseria, Comparison of self-consistent field convergence acceleration techniques, J. Chem. Phys. 137 (2012) 054110.

[8] A. C. Vaucher, M. Reiher, Steering orbital optimization out of local minima and saddle points toward lower energy, J. Chem. Theory Comput. 13 (2017) 1219.

[9] M. Head-Gordon, J. Pople, Optimization of wave function and geometry in the finite basis hartree-fock method, J. Phys. Chem. 92 (1988) 3063.

[10] M. J. Gillan, Calculation of the vacancy formation energy in aluminium, J. Phys. Condens. Matter 1 (1989) 689.

[11] M. Payne, M. Teter, D. Allan, T. Arias, J. Joannopoulos, Iterative minimization techniques for ab initio total-energy calculations: Molecular dynamics and conjugate gradients, Rev. Mod. Phys. 64 (1992) 1045.

[12] J. Hutter, M. Parrinello, S. Vogel, Exponential transformation of molecular orbitals, J. Chem. Phys 101 (1994) 3862.

[13] N. Marzari, D. Vanderbilt, M. C. Payne, Ensemble density-functional theory for ab initio molecular dynamics of metals and finite-temperature insulators, Phys. Rev. Lett. 79 (1997) 1337.

[14] S. Ismail-Beigi, T. Arias, New algebraic formulation of density functional calculation, Comput. Phys. Commun 128 (2000) 1.

[15] T. Van Voorhis, M. Head-Gordon, A geometric approach to direct minimization, Molecular Physics 100 (2002) 1713.

[16] J. VandeVondele, J. Hutter, An efficient orbital transformation method for electronic structure calculations, J. Chem. Phys. 118 (2003) 4365.

[17] V. Weber, J. Vandevondele, J. Hutter, A. Niklasson, Direct energy functional minimization under orthogonality constraints, J. Chem. Phys. 128 (2008) 084113.

[18] C. Freysoldt, S. Boeck, J. Neugebauer, Direct minimization technique for metals in density functional theory, Phys. Rev. B 79 (2009) 241103.

[19] A. Edelman, T. Arias, S. Smith, The geometry of algorithms with orthogonality constraints, SIAM J. Matrix Anal. Appl. 20 (1998) 303.

[20] J. F. Rico, J. M. G. De La Vega, J. I. F. Alonso, P. Fantucci, Restricted Hartree–Fock approximation. I. Techniques for the energy minimization, J. Comput. Chem. 4 (1983) 33.

[21] J. Á. Fern Rico, M. Paniagua, J. I. Fern Alonso, P. Fantucci, Restricted Hartree–Fock approximation. II. Computational aspects of the direct minimization procedure, J. Comput. Chem. 4 (1983) 41.

[22] J. Douady, Y. Ellinger, R. Subra, B. Levy, Exponential transformation of molecular orbitals: A quadratically convergent SCF procedure. I. General formulation and application to closed-shell ground states, J. Chem. Phys. 72 (1980) 1452.

[23] J. P. Perdew, A. Zunger, Self-interaction correction to density-functional approximations for many-electron systems, Phys. Rev. B 23 (1981) 5048.

[24] S. Lehtola, H. Jónsson, Variational, Self-Consistent Implementation of the Perdew–Zunger Self-Interaction Correction with Complex Optimal Orbitals, J. Chem. Theory Comput. 10 (12) (2014) 5324.

[25] G. Borghi, C.-H. Park, N. Nguyen, A. Ferretti, N. Marzari, Variational minimization of orbital-density-dependent functionals, Phys. Rev. B 91 (2015) 155112.

[26] S. Lehtola, M. Head-Gordon, H. Jónsson, Complex orbitals, multiple local minima, and symmetry breaking in perdew-zunger self-interaction corrected density functional theory calculations, J. Chem. Theory Comput. 12 (2016) 3195.

[27] J. Nocedal, S. J. Wright, Numerical Optimization, 2nd Edition, Springer, New York, NY, USA, 2006.

[28] P. Wolfe, Convergence conditions for ascent methods, SIAM Review 11 (1969) 226.

[29] P. Wolfe, Convergence conditions for ascent methods. ii: Some corrections, SIAM Review 13 (1971) 185.

[30] W. Hager, H. Zhang, A new conjugate gradient method with guaranteed descent and an efficient line search, SIAM J. Optim 16 (2006) 170.

[31] C. Moler, C. Van Loan, Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later, SIAM Review 45 (2003) 3.

[32] A. Al-Mohy, N. Higham, A new scaling and squaring algorithm for the matrix exponential, SIAM J. Matrix Anal. Appl. 31 (2009) 970.

[33] E. Jones, T. Oliphant, P. Peterson, et al., SciPy: Open source scientific tools for Python (2001–).

[34] P. E. Blöchl, Projector augmented-wave method, Phys. Rev. B 50 (1994) 17953.

[35] J. Enkovaara, C. Rostgaard, J. J. Mortensen, J. Chen, M. Dułak, L. Ferrighi, J. Gavnholt, C. Glinsvad, V. Haikola, H. A. Hansen, H. H. Kristoffersen, M. Kuisma, A. H. Larsen, L. Lehtovaara, M. Ljungberg, O. Lopez-Acevedo, P. G. Moses, J. Ojanen, T. Olsen, V. Petzold, N. A. Romero, J. Stausholm-Møller, M. Strange, G. A. Tritsaris, M. Vanin, M. Walter, B. Hammer, H. Häkkinen, G. K. H. Madsen, R. M. Nieminen, J. K. Nørskov, M. Puska, T. T. Rantala, J. Schiøtz, K. S. Thygesen, K. W. Jacobsen, Electronic structure calculations with GPAW: a real-space implementation of the projector augmented-wave method, J. Phys. Condens. Matter 22 (2010) 253202.

[36] A. H. Larsen, M. Vanin, J. J. Mortensen, K. S. Thygesen, K. W. Jacobsen, Localized atomic basis set in the projector augmented wave method, Phys. Rev. B 80 (2009) 195112.

[37] L. Curtiss, K. Raghavachari, P. Redfern, J. Pople, Assessment of gaussian-2 and density functional theories for the computation of enthalpies of formation, J. Chem. Phys. 106 (1997) 1063.

[38] J. P. Perdew, K. Burke, M. Ernzerhof, Generalized gradient approximation made simple, Phys. Rev. Lett. 77 (1996) 3865.

[39] V. N. Staroverov, G. E. Scuseria, J. Tao, J. P. Perdew, Tests of a ladder of density functionals for bulk solids and surfaces, Phys. Rev. B 69 (2004) 075102.

[40] G. Levi, A. V. Ivanov, H. Jónsson, Variational calculations of excited states via direct optimization of the orbitals in DFT, Faraday Discuss. 224 (2020) 448.

[41] G. Levi, A. V. Ivanov, H. Jónsson, Variational Density Functional Calculations of Excited States via Direct Optimization, J. Chem. Theory Comput. 16 (2020) 6968.

Article II

Variational Density Functional Calculations of Excited States via Direct Optimization

Levi G., Ivanov A.V., Jónsson H.

J. Chem. Theory Comput. **16**, 6968-6982 2020

Variational density functional calculations of excited states via direct optimization

Gianluca Levi,*,† Aleksei V. Ivanov,†,‡ and Hannes Jónsson†

†*Science Institute and Faculty of Physical Sciences, University of Iceland, 107 Reykjavík, Iceland*

‡*Saint Petersburg State University, 199034 Saint Petersburg, Russia.*

E-mail: giale@hi.is

Abstract

The development of variational density functional theory approaches to excited electronic states is impeded by limitations of the commonly used self-consistent field (SCF) procedure. A method based on a direct optimization approach as well as the maximum overlap method is presented and the performance compared with previously proposed SCF strategies. Excited-state solutions correspond to saddle points of the energy as a function of the electronic degrees of freedom. The approach presented here makes use of a preconditioner determined with the help of the maximum overlap method to guide the convergence on a target nth-order saddle point. The method is found to be more robust and to converge faster than previously proposed SCF approaches for a set of 89 excited states of molecules. A limited-memory formulation of the symmetric rank-one method for updating the inverse Hessian is found to give the best performance. A conical intersection for the carbon monoxide molecule is calculated without resorting to fractional occupation numbers. Calculations on excited states of the hydrogen atom and a doubly excited state of the dihydrogen molecule using a self-interaction corrected functional are presented. For these systems, the self-interaction correction is found to

improve the accuracy of density functional calculations of excited states.

1 Introduction

In the light of recent and rapid advancements in fields such as photocatalysis and ultrafast spectroscopies, the availability of efficient and accurate computational methods to model electronic excited-state properties of molecules has become increasingly important. A widely used methodology to obtain excited-state properties of molecules is time-dependent density functional theory (TDDFT).[1–3] Practical applications of TDDFT rely on (i) linear response to describe the perturbation of the electron density due to an external field, and (ii) the adiabatic approximation, which neglects the time dependency of the functional derivative of the exchange-correlation (xc) potential with respect to the density, the so-called xc kernel. With those approximations, the computations can be carried out with local and semi-local ground-state Kohn-Sham (KS)[4,5] functionals without excessive computational requirements and this has been found to give an adequate description of valence excitations in many cases.[1,6] On the other hand, the neglect of the time dependency of the xc kernel limits the applicability of this approach and makes it, for example, inadequate for the description of double excitations[7–9] and conical intersections between ground and excited states.[7,10] Moreover, due to the incorrect form of the potential at long range and to the lack of orbital relaxation effects,[11–13] TDDFT with KS functionals suffers from systematic errors when applied to excited states that are diffuse, such as Rydberg states,[6,14,15] or involve transfer of charge between spatially separated regions.[16–18]

Some of these issues can be solved employing alternative DFT formulations where excited states are obtained as single Slater determinant wave functions optimized for non-aufbau occupations using ground-state functionals. Here, one seeks a saddle point on the energy surface instead of a minimum. Thanks to the inclusion of state-specific orbital relaxation

effects, these methods can describe a broader range of excited states than linear-response TDDFT in the adiabatic approximation, and have, therefore, seen a revival of interest recently.[12,14,15,19–35]

The excited-state DFT methodology that we consider here does not enforce orthogonality constraints between the different excited-state solutions and the ground state, and, therefore, represents a straightforward extension of ground-state DFT[4]. Higher-energy stationary points of ground-state density functionals obtained in this way do not necessarily represent rigorous approximations to the exact stationary states.[15,36] On the other hand, practice has shown that excited-state DFT calculations are usually able to deliver useful approximations to excited-state properties of molecules, such as excitation energies and potential energy surfaces.[23,34] Some studies have also highlighted how the method can satisfactorily treat cases, such as conical intersections, with strong static correlation, despite the single-determinant approximation.[19,22,22,33]

From a more practical point of view, the lack of orthogonality and the single-determinant approximation give rise to difficulties in the convergence of higher-energy solutions. First of all, when lower-energy states of the same symmetry are present, variational collapse can occur due to mixing of occupied and virtual orbitals with the same symmetry. The commonly used self-consistent field (SCF) approach can be combined with a maximum overlap method (MOM),[15,23,34] which prevents variational collapse. However, SCF convergence can still be problematic when dealing with single determinants that include unequally occupied degenerate or near-degenerate orbitals. This situation is analogous to what happens for ground states with vanishing HOMO-LUMO gap[37] and can arise, for example, close to conical intersections.[38] One strategy that is often adopted is electronic smearing to obtain convergence on an average occupied configuration.[39] This, however, comes with the risk of introducing artifacts in the calculated excited-state properties, as will be demonstrated below.

[4]Sometimes, this method is referred to as Δ self-consistent field (ΔSCF),[21,24,28,32,33,35] but here we prefer the more general term "excited-state DFT", following Cheng *et al.*,[15] avoiding the risk of relating the method to a specific nonlinear variational procedure (such as SCF) and to the computation of a specific excited-state property (the excitation energy through the energy difference, Δ, between excited and ground state).

There exist alternative nonlinear variational procedures for finding stationary points of energy functionals based on direct optimization (DO) of the orbitals through unitary transformations.[40–43] Implemented with gradient-based unconstrained optimization algorithms, this approach has been shown to be a more robust strategy for converging ground states with DFT than SCF-based methodologies.[40,44,45] However, the risk of variational collapse impedes straightforward application of DO methods for locating saddle points of the energy surface. One way of circumventing this problem is to convert the saddle-point optimization to a minimization of the squared norm of the gradient of the energy with respect to the electronic degrees of freedom.[21] Variational collapse is avoided by squared gradient minimization but there is a series of drawbacks that have to be considered. First, the computational cost is increased with respect to ground-state calculations, because the gradient of the squared norm of the gradient is needed. Furthermore, this strategy requires more iterations than SCF-MOM (when convergence can be reached),[21] because squared gradient minimization is less well conditioned than energy minimization.[21,46] Lastly, this approach can converge on points where the squared norm of the gradient has a minimum but is not zero. The initial guess, therefore, needs to be sufficiently good.[21]

When the above-mentioned practical issues have not represented a problem, excited-state calculations using KS functionals have given more accurate results than linear-response TDDFT for a number of challenging excited states. These include doubly excited states,[21,23] core excitations,[20] Rydberg[15,21] and charge-transfer[23,26,34,47] transitions, absorption spectra[48] and structural dynamics[24,49] in solution, including nonadiabatic dynamics.[19,22,33] However, it has been pointed out[11,50] that many excited states, such as Rydberg, charge-transfer and doubly excited states, are affected more by self-interaction error (SIE) than ground states at the level of the commonly employed semi-local KS functionals. An unbalanced treatment of self interaction can, for example, lead to systematic errors in calculations of excitation energy.[11] Self-interaction correction (SIC)[51] applied to KS functionals corrects the long-range form of the effective potential, as has been demonstrated, for example, for Rydberg states[50]

and dipole bound anions;[52] thus, it can improve the description of the excited states.[53] However, it is challenging to perform fully variational calculations with SIC functionals since they are explicitly orbital-density dependent and the energy is not invariant to unitary tranformations among the equally occupied orbitals. While fully variational implementations of SIC functionals has been developed for ground states,[54–57] the excited-state calculations have so far not been fully variational.[50]

Here, we present a DO approach with the aim of improving on already existing excited-state DFT methodologies in two ways: (1) ensuring convergence for different types of excited states, including cases with unequally occupied degenerate orbitals, while avoiding variational collapse and without increasing the computational cost with respect to ground-state DFT calculations; (2) allowing the use of non-unitary invariant functionals, such as SIC functionals, in variational excited-state calculations. The proposed method uses a quasi-Newton algorithm to directly converge on saddle points of any order with the help of a preconditioner built from the eigenvalues of the Hamiltonian matrix at given intervals during the optimization, and MOM constraints to prevent variational collapse. A preliminary evaluation of the convergence properties of the DO-MOM method when using the Limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm and a new limited-memory formulation of Powell inverse Hessian update (L-Powell) is presented in a conference proceeding.[58] L-BFGS is a quasi-Newton method commonly employed for minimization, and it was shown that the application in the present context crucially depends on updates of the preconditioner and on the MOM constraints in order to converge on a saddle point. L-Powell was found to be less robust than L-BFGS,[58] despite its ability to generate indefinite Hessian approximations. It would be advantageous to attain convergence on a target nth-order saddle point without relying on updates of the preconditioner, since it requires costly diagonalization of the Hamiltonian matrix. In the present work, we extend the limited-memory inverse Hessian update algorithm presented in reference[58] to the symmetric rank-one (SR1) formula. SR1 can develop negative eigenvalues[59] and therefore has the desired characteristics to minimize

the dependency on the preconditioner.

The convergence properties of the DO-MOM method[58] are tested on 55 singlet and 34 triplet excited states of small and medium size molecules, including tests of the new limited-memory SR1 (L-SR1) inverse Hessian update algorithm. Furthermore, we test the convergence with respect to two challenging charge-transfer states of the nitrobenzene molecule, for which SCF-MOM has been reported to fail,[21,60] demonstrating that improved robustness and reduced dependency on the preconditioner can be achieved with the new L-SR1 method. Finally, we show how the DO-MOM method can converge for systems with unequally occupied (near-)degenerate orbitals without tuning modifications, taking the conical intersection of two excited states of carbon monoxide as a representative example. In each case, the performance of DO-MOM is compared to that of a standard SCF-MOM method.

The DO-MOM method can be used for non-unitary invariant functionals such as SIC functionals, as well as the unitary invariant KS functionals. We perform fully variational excited-state calculations with SIC on the hydrogen atom and dihydrogen molecule and show that the application of SIC in both ground- and excited-state calculations leads to significant improvement in the excitation energy.

2 Theory

2.1 Excited-State DFT

2.1.1 Kohn-Sham Formulation

Within KS DFT,[4,5] excited states of a spin-polarized system of $N = N_\uparrow + N_\downarrow$ electrons with density $n(\mathbf{r}) = n_\uparrow(\mathbf{r}) + n_\downarrow(\mathbf{r})$ can be found as saddle points of the energy surface defined by the dependence of the ground-state energy on the electronic degrees of freedom:[15]

$$E_{\mathrm{KS}}\left[n_\uparrow, n_\downarrow\right] = T_{\mathrm{s}}\left[n_\uparrow, n_\downarrow\right] + V_{\mathrm{ext}}\left[n\right] + J\left[n\right] + E_{\mathrm{xc}}\left[n_\uparrow, n_\downarrow\right] \tag{1}$$

where $T_{\mathrm{s}}[n_\uparrow, n_\downarrow]$ is the kinetic energy of the non-interacting $N-$electron system, $V_{\mathrm{ext}}[n]$ and $J[n]$ are the energy due to the external potential and the Hartree electrostatic energy, respectively:

$$V_{\mathrm{ext}}[n] = \int \boldsymbol{\upsilon}_{\mathrm{ext}}(\mathbf{r}) n(\mathbf{r}) d\mathbf{r} \tag{2}$$

$$J[n] = \frac{1}{2} \int \int \frac{n(\mathbf{r}) n(\mathbf{r}')}{\mid \mathbf{r} - \mathbf{r}' \mid} d\mathbf{r} d\mathbf{r}' \tag{3}$$

while $E_{\mathrm{xc}}[n_\uparrow, n_\downarrow]$ is the exchange-correlation (xc) functional. The KS kinetic energy and the spin densities $n_\sigma(\mathbf{r})$ are given in terms of orthonormal KS orbitals $\psi_{n\sigma}(\mathbf{r})$:

$$T_{\mathrm{s}}[n_\uparrow, n_\downarrow] = -\frac{1}{2} \sum_{n\sigma} f_{n\sigma} \int \psi^*_{n\sigma}(\mathbf{r}) \nabla^2 \psi_{n\sigma}(\mathbf{r}) d\mathbf{r} \tag{4}$$

$$n_\sigma(\mathbf{r}) = \sum_n f_{n\sigma} \mid \psi_{n\sigma}(\mathbf{r}) \mid^2 \tag{5}$$

in which $0 \leq f_{n\sigma} \leq 1$ is the occupation number for orbital n with σ spin quantum number ($\uparrow$ or $\downarrow$).

Stationary states of the non-interacting $N-$electron system can be obtained by finding extrema of the energy, eq. 1, subject to orbital orthonormality constraints:

$$\int \psi^*_{n\sigma}(\mathbf{r}) \psi_{m\sigma'}(\mathbf{r}) d\mathbf{r} = \delta_{nm} \delta_{\sigma\sigma'} \tag{6}$$

For a fixed set of $f_{\sigma n}$, the stationarity condition leads to a set of nonlinear coupled equations:

$$f_{n\sigma} \mathbf{H}^\sigma_{\mathrm{KS}} \psi_{n\sigma} = \sum_m \lambda^\sigma_{nm} \psi_{m\sigma} \tag{7}$$

where the λ^σ_{nm} are Lagrange multipliers for the constraints, and $\mathbf{H}^\sigma_{\mathrm{KS}}$ is the one-particle KS

Hamiltonian:

$$\mathbf{H}^{\sigma}_{\mathrm{KS}} = -\frac{1}{2}\nabla^2 + \boldsymbol{\upsilon}_{\mathrm{ext}}(\mathbf{r}) + \int \frac{n(\mathbf{r}')}{\mid \mathbf{r} - \mathbf{r}' \mid} d\mathbf{r}' + \boldsymbol{\upsilon}^{\sigma}_{\mathrm{xc}}(\mathbf{r}) \tag{8}$$

For a functional with a form given by eq. 1, any unitary transformation that mixes equally occupied orbitals among themselves leaves the total energy unchanged. Therefore, the set of orbitals that makes the energy stationary for given set of occupation numbers is not unique.

2.1.2 Self-Interaction Correction

In KS functionals, the Coulomb interaction between the electrons is estimated from the total electron density, and hence it includes non-local self interaction. While the xc functional also includes self interaction of opposite sign, a local or semi-local functional form cannot cancel out the self Coulomb interaction and a SIE remains, as can be seen most clearly for one-electron systems. Perdew and Zunger[51] proposed the following procedure for removing self interaction from a KS functional:

$$E_{\mathrm{SIC}}[n_{\uparrow}, n_{\downarrow}] = E_{\mathrm{KS}}[n_{\uparrow}, n_{\downarrow}] - \sum_{n\sigma} \left(J[n_{n\sigma}] + E_{xc}[n_{n\sigma}, 0] \right) \tag{9}$$

where $n_{n\sigma} = |\psi_{n\sigma}|^2$ is an orbital density. This represents an orbital-by-orbital estimate of the SIE that is exact for one-electron systems.

For a SIC functional, the stationarity condition leads to a set of nonlinear coupled equations:

$$f_{n\sigma} \left(\mathbf{H}^{\sigma}_{\mathrm{KS}} - \mathbf{V}_{n\sigma} \right) \psi_{n\sigma} = \sum_{m} \lambda^{\sigma}_{nm} \psi_{m\sigma} \tag{10}$$

where the Hamiltonian contains an orbital-density dependent part:

$$\mathbf{V}_{n\sigma} = \int \frac{n_{n\sigma}(\mathbf{r}')}{\mid \mathbf{r} - \mathbf{r}' \mid} d\mathbf{r}' + \boldsymbol{\upsilon}_{\mathrm{xc}}(n_{n\sigma}) \tag{11}$$

In contrast to KS functionals, SIC functionals are not invariant under unitary transformations of the equally occupied orbitals, and the optimal orbitals are uniquely defined as those that extremize the energy of the given SIC functional.[54–57,61–63] This corresponds to maximizing the self-interaction correction, and involves unitary optimization within the manifold of occupied orbitals.

2.2 Self-Consistent Field

For unitary invariant functionals, eq. 7 can be simplified by choosing a unitary transformation that diagonalizes $\boldsymbol{\lambda}^\sigma$ while leaving the energy unchanged, leading to the generalized KS eigenvalue equations in the canonical form:

$$\mathbf{H}^\sigma_{\mathrm{KS}}\psi_{n\sigma}(\mathbf{r}) = \epsilon_{n\sigma}\psi_{n\sigma}(\mathbf{r}) \tag{12}$$

For the non-unitary invariant SIC functionals presented in the previous section, the Lagrange matrix is not diagonal for the optimal orbitals that extremize the total SIC-DFT energy due to the orbital-density dependence.[56,61,64]

Solutions to the KS equations are found iteratively, defining the SCF procedure. The ground state corresponds to a minimum of the energy given by the functional and is obtained if at each SCF iteration the orbitals are occupied according to the aufbau principle. Saddle points on the energy surface are obtained for non-aufbau occupations and are interpreted as excited states.[15,21,65] Non-aufbau occupations can be enforced during the SCF cycle through the MOM method: at each iteration, the occupied orbitals are selected as those that overlap most with the occupied orbitals of the previous iteration[34] or with a set of fixed reference orbitals[15,23] (the latter strategy is also known as initial maximum overlap method (IMOM)[23]).

2.3 Direct Optimization

Alternatively, the variational problem can be formulated as an optimization of the orbitals through application of a unitary transformation to a set of orthonormal reference orbitals:[40–43]

$$\phi_{p\sigma}(\mathbf{r}) = \sum_q U^{\sigma}_{pq} \psi_{q\sigma}(\mathbf{r}) \tag{13}$$

The unitary matrix $\mathbf{U}$ can be parametrized as the matrix exponential:[41,42]

$$\mathbf{U} = e^{\boldsymbol{\theta}} \tag{14}$$

where $\boldsymbol{\theta}$ is required to be anti-Hermitian ($\boldsymbol{\theta} = -\boldsymbol{\theta}^{\dagger}$) in order to preserve the orbital orthonormality. In this way, the energy functional can be directly extremized in the linear space formed by anti-Hermitian matrices, which makes it possible to use well-established local unconstrained optimization strategies.[59] The exponential transformation of molecular orbitals can be applied to both KS and SIC functionals, since it does not require the functional to be unitary invariant (unitary optimization for SIC functionals means that the elements of $\boldsymbol{\theta}$ that mix occupied orbitals are non-zero in contrast to KS functionals, as explained in the next section). Moreover, gradient-based direct optimization (DO) ensures more rigorous convergence compared to SCF.[40,65]

For excited states, the unconstrained search can be done with quasi-Newton methods that are able to locate saddle points. Compared to minimization, the search for a saddle point is arguably a more challenging task, requiring an initial guess that is sufficiently close to the wanted solution and a good approximation to the Hessian. Nevertheless, quasi-Newton methods for saddle points have long been employed with some success in various contexts, most notably transition-state searches on potential energy surfaces for atomic rearrangements.[66–71] Here, we explore a strategy for DO of saddle points of KS and SIC density

functionals using quasi-Newton search directions starting from a guess obtained by promoting one or more electrons from occupied to unoccupied orbitals of a converged ground-state calculation.

3 Implementation

We have implemented DO-MOM with KS and SIC functionals in a development branch of the Grid-based Projector Augmented Wave (GPAW)[72–74] software using localized atomic basis sets to represent the molecular orbitals. The implementation of the exponential transformation for KS functionals is presented in Reference.[58] A review of this implementation and its extension to SIC functionals are given in Appendix A. The MOM is based on a standard implementation using fixed reference orbitals[23] as shown in Appendix C. In the following, we describe the new L-SR1 algorithm, including the choice of preconditioner.

3.1 Quasi-Newton Step

The computational effort of a quasi-Newton step based on updating the Hessian matrix scales as $\mathcal{O}(n^3)$,[59] where n is the dimensionality of the optimization problem (the present DO implementation based on exponential transformation scales as NM, where N is the number of occupied orbitals and M the number of basis set functions). A less computationally demanding approach is to update the inverse Hessian instead of the Hessian, since this does not involve any matrix-matrix operation or solution of a linear system of equations. The quasi-Newton step with inverse Hessian update is:

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \mathbf{B}^{(k)}\mathbf{g}^{(k)} \tag{15}$$

where $\mathbf{B}^{(k)}$ is the approximate inverse Hessian at iteration k, and $\mathbf{x}^{(k)}$ and $\mathbf{g}^{(k)}$ are the vectors of the $\{\theta_{ij}\}$ independent variables and the analytical gradient, respectively.

When the inverse Hessian is updated, the arithmetic operations scale as $\mathcal{O}(n^2)$,[59] which

can become a bottleneck for systems with a moderate number of electrons and/or large basis sets. To circumvent the costly operations embedded in the explicit update and storage of the Hessian matrix, quasi-Newton algorithms can be formulated in a limited-memory version by storing only vectors and scalars carrying the information necessary to propagate $\mathbf{B}$ implicitly. In this case, the operations involved in one iteration scale linearly as $\mathcal{O}(mn)$, where m is the number of previous steps used to update the current step.

L-BFGS is a commonly used limited-memory version of BFGS, which is generally considered to be the most effective inverse Hessian update for minimization. The L-BFGS method has been implemented here as described in reference.[59] In calculations of atomic structures, the Powell or SR1 Hessian update formulas, or a combination of the two,[67,75] are preferred for saddle-point searches, because they are able to develop negative eigenvalues contrary to the BFGS formula. Therefore, we have formulated and implemented limited-memory variants of the Powell and SR1 inverse Hessian updates (L-Powell and L-SR1) by extending the approach based on Powell Hessian updates presented by Anglada *et al.*[76] The L-Powell method is described in reference[58] and is reviewed in Appendix B.

The inverse Hessian SR1 update formula, written in a compact form, is:[77]

$$\mathbf{B}_{\mathrm{SR1}}^{(k+1)} = \mathbf{B}^{(k)} + \frac{\mathbf{j}^{(k)}\mathbf{j}^{T(k)}}{\mathbf{j}^{T(k)}\mathbf{y}^{(k)}} \tag{16}$$

where:

$$\mathbf{j}^{(k)} = \mathbf{s}^{(k)} - \mathbf{B}^{(k)}\mathbf{y}^{(k)}, \tag{17}$$

and:

$$\mathbf{s}^{(k)} = \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}, \quad \mathbf{y}^{(k)} = \mathbf{g}^{(k+1)} - \mathbf{g}^{(k)} \tag{18}$$

For any vector $\mathbf{v}^{(k)}$ and approximation $\mathbf{B}_0^{(k)}$ to the inverse Hessian ($\mathbf{B}_0^{(k)}$ can in principle be

allowed to vary at each iteration), $\mathbf{B}_{\mathrm{SR1}}^{(k)}\mathbf{v}^{(k)}$ can be computed using the following recursive formula:

$$\mathbf{B}_{\mathrm{SR1}}^{(k)}\mathbf{v}^{(k)} = \mathbf{B}_{0}^{(k)}\mathbf{v}^{(k)} + \sum_{i=k-m}^{k-1} \frac{\mathbf{j}^{(i)}\mathbf{j}^{T(i)}\mathbf{v}^{(k)}}{\mathbf{j}^{T(i)}\mathbf{y}^{(i)}} \tag{19}$$

which takes into account the implicit information contained in the m most recent steps. Using this result, the L-SR1 algorithm can be formulated as shown in Algorithm 1.

Choose $\mathbf{x}^{(0)}$, m and $p_{\max}$;
$k \leftarrow 0$;
while *not converged* **do**
  Choose $\mathbf{B}_{0}^{(k)}$;
  Compute $\mathbf{p}^{(k)} \leftarrow \mathbf{B}^{(k)}\mathbf{g}^{(k)}$ using eq. 19;
  if $\|\mathbf{p}^{(k)}\| \geq p_{\max}$ **then**
    $\mathbf{p}^{(k)} \leftarrow \frac{p_{\max}}{\|\mathbf{p}^{(k)}\|}\mathbf{p}^{(k)}$
  end
  $\mathbf{x}^{(k+1)} \leftarrow \mathbf{x}^{(k)} - \mathbf{p}^{(k)}$;
  if $k > m$ **then**
    discard vector $\mathbf{j}^{(k-m)}$ and scalar $r^{(k-m)}$;
  end
  $\mathbf{s}^{(k)} \leftarrow \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}$ and $\mathbf{y}^{(k)} \leftarrow \mathbf{g}^{(k+1)} - \mathbf{g}^{(k)}$;
  Compute $\mathbf{j}^{(k)} \leftarrow \mathbf{B}^{(k)}\mathbf{y}^{(k)}$ using eq. 19;
  $\mathbf{j}^{(k)} \leftarrow \mathbf{s}^{(k)} - \mathbf{j}^{(k)}$;
  $r^{(k)} \leftarrow \mathbf{j}^{T(k)}\mathbf{y}^{(k)}$;
  Store vector $\mathbf{j}^{(k)}$ and scalar $r^{(k)}$;
  $k \leftarrow k + 1$;
end

Algorithm 1: Quasi-Newton algorithm with limited-memory SR1 inverse Hessian update. The computational cost of the operations involved scales linearly with n if $\mathbf{B}_{0}^{(k)}$ is selected to be diagonal.

Among the quasi-Newton schemes that are commonly used in optimizations of saddle points for atomic rearrangements, some update the Hessian through a combination of the SR1 and Powell updates. Algorithm 1 is readily generalized to use an analogous update

formula that combines the SR1 and Powell inverse Hessian updates:

$$\mathbf{B}_{\phi}^{(k+1)} = (1-\phi^k)\mathbf{B}_{\mathrm{SR1}}^{(k+1)} + \phi^k \mathbf{B}_{\mathrm{P}}^{(k+1)} \tag{20}$$

where $\mathbf{B}_{\mathrm{SR1}}^{(k+1)}$ and $\mathbf{B}_{\mathrm{P}}^{(k+1)}$ are given by eqs. 16 and B.1, respectively. Following Bofill,[67,75] the factor ϕ^k can be taken as:

$$\phi^k = 1 - \frac{(\mathbf{y}^{T(k)}\mathbf{j}^{(k)})^2}{(\mathbf{y}^{T(k)}\mathbf{y}^{(k)})(\mathbf{j}^{T(k)}\mathbf{j}^{(k)})} \tag{21}$$

In Algorithm 1 we have also introduced a maximum allowed step length, $p_{\max}$. This is because, due to the approximate nature of the initial approximation to the Hessian (see next section), initial steps may be too large, causing departure from the basin of attraction of the desired saddle point. We have found that $p_{\max} = 0.20$ provides an adequate balance between stability and speed of convergence in most cases. The SR1 update can become unstable if the denominator in eq. 16 is small. To avoid such instabilities, the following procedure is adopted: if $|\mathbf{j}^{T(i)}\mathbf{y}^{(i)}| < \varepsilon$, where ε is a small number, then $\mathbf{j}^{T(i)}\mathbf{y}^{(i)}$ is set to ε. When using $\varepsilon = 10^{-12}$, we have found that this procedure prevents L-SR1 from becoming unstable, without affecting the rate of convergence.

3.2 Preconditioner

The preconditioner for the quasi-Newton step, represented by the matrix $\mathbf{B}_0^{(k)}$ introduced in the previous section, is chosen as the inverse of the following diagonal approximation to the Hessian matrix:[42]

$$\frac{\partial^2 E}{\partial^2 \theta_{ij}} \approx -2(\epsilon_i - \epsilon_j)(f_i - f_j) \tag{22}$$

where the ϵ_i are the eigenvalues of the Hamiltonian matrix. Eq. 22 is obtained by taking the derivative of a linear expansion of the gradient (eq. A.6) and neglecting second-order

derivatives of the potential. Previously, it has been shown that this type of preconditioner can improve the convergence of Hartree-Fock (HF) and DFT calculations based on direct minimization of the energy,[21,40] when using the BFGS method.

At the beginning of the optimization, the preconditioner is generated using the eigenvalues and occupation numbers of the guess obtained by promoting electrons from occupied to virtual orbitals of the ground state. As will be shown for the excited states of nitrobenzene, it can happen that the number of negative eigenvalues of this initial approximate Hessian is not consistent with the order of the saddle point corresponding to the target excited state. To ensure that the preconditioner has the appropriate structure to guide the convergence towards the target nth-order saddle point, the approximate Hessian of eq. 22 is recomputed at regular intervals during the optimization and $\mathbf{B}_0^{(k)}$ updated together with the reference orbitals. In order to find the occupation numbers of the canonical orbitals, which are needed to compute the preconditioner based on eq. 22, the MOM method is employed (see Appendix C). Close to the target solution, the update of the preconditioner is not needed and can be avoided using a threshold on the magnitude of the energy gradient, which reduces the computational cost by avoiding unnecessary diagonalization of the Hamiltonian matrix. Finally, we note that the preconditioner derived from eq. 22 is not defined for oo terms, since in this case $f_i^{(k)} = f_j^{(k)}$, and for degenerate ov pairs. For these cases, the preconditioner is not used, corresponding to setting the elements of $\mathbf{B}_0^{(k)}$ to 1.

4 Computational Methods

All the calculations presented in this work are performed with a development version of GPAW where the DO-MOM method for KS and SIC xc functionals has been implemented. The PAW method[78] is used to treat the regions near the nuclei, core electrons of each atom are frozen to the result of a reference scalar relativistic calculation of the isolated atom, and valence electrons are represented in a basis of linear combination of atomic orbitals. For

all the basis sets considered in this work, the uncontracted functions are removed, as the nodal structure of the orbitals around the nuclei is accounted for by the PAW correction. The simulation cell has a uniform grid with grid spacing of 0.15 Å, while the dimensions of the box are chosen in such a way as to avoid effects due to truncation of the numerical representation of the basis functions. For the DO-MOM calculations, a maximum allowed step length, $p_{\max}$, of 0.20 is utilized, while the memory m of L-BFGS, L-Powell and L-SR1 is chosen as equal to 20. At every 20th step the preconditioner based on eq. 22 is updated unless the root mean square of the gradient is less than 10^{-3} eV. The SCF-MOM method is based on direct diagonalization of the Hamiltonian matrix in the basis of atomic orbitals and updating the electron density using a direct inversion in the iterative subspace (DIIS) procedure (Pulay mixing of the density[72,79]). We use GPAW default parameters for the Pulay mixing of the density: the number of old densities used in the mixing is 3, the coefficient used in the linear mixing of the density with the density residual vector is 0.15, and no damping of short wavelength density changes is used.[72] Unless otherwise stated, convergence (both in SCF or DO calculations) is considered achieved if the integrated value of the square of the residuals of eq. 7 (for KS functionals) or 10 (for SIC functionals) is less than $4.0 \cdot 10^{-8}$ eV^2 per electron. All calculations are performed within the spin-unrestricted formalism. Since each state is described by a single determinant, open-shell singlets are not pure-spin states. Both the KS and SIC calculations use the generalized gradient approximation (GGA) functional PBE.[80]

4.1 Convergence Tests

The robustness and rate of convergence of the DO-MOM method is assessed by performing single-point calculations of the excited states of small and medium size molecules. The tests include 52 singlet and 34 triplet excited states of 18 small compounds from the benchmark set of reference,[81] and the lowest singlet excited states of 3 medium organic compounds (acetone, benzene and naphtalene) from reference,[82] for a total of 89 states generated by

single electron transition from the ground state. Lowest triplet states have been excluded from this benchmark set since they correspond to minima on the energy surface and not saddle points. These states are chosen because highly accurate reference data is available making reliable assignment of the states possible, and due to the diverse character of the electronic transitions. The test set includes 35 valence (V) excitations ($n \rightarrow \pi^*$, $\sigma \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ transtions), 53 Rydberg (R), and 1 charge-transfer (CT) states (the lowest singlet excited state of hydrogen chloride). The geometries are taken from reference[81] and reference.[82] For the DO-MOM calculations three different inverse Hessian update schemes are compared: L-BFGS, L-Powell and L-SR1 (the latter two according to the limited-memory algorithm presented in section 3.1). We further compare DO-MOM to a standard SCF-MOM method based on direct diagonalization of the Hamiltonian matrix, as implemented in GPAW.[73] For each molecule, the ground state is first converged using SCF. Then, the initial guess for an excited state is generated by a one-electron excitation involving the occupied and unoccupied orbitals that define the character and symmetry of the excited state. Convergence is obtained when the square of the residuals is less than 10^{-10}eV^2. The maximum number of iterations for a calculation is 300. The *aug*-cc-pVDZ basis set[83–85] is used.

The calculations of nitrobenzene test both SCF-MOM and DO-MOM with L-BFGS and L-SR1 with respect to convergence to the singlet $^1\text{A}_1(\text{n}_\pi \rightarrow \pi'^*)$ and $^1\text{A}_1(\pi' \rightarrow \pi^*)$ excited states. Using the ground-state orbitals, the initial guess for the $^1\text{A}_1(\text{n}_\pi \rightarrow \pi'^*)$ state is generated by promoting an electron from the highest energy π lone pair (n_π) to the second lowest π^* orbital (π'^*), while for the $^1\text{A}_1(\pi' \rightarrow \pi^*)$ state excitation is from the second highest occupied π orbital (π') to the lowest unoccupied π^* orbital. The calculations are perfomed at the C_{2v} geometry used in references.[21,60] The basis set is def2-TZVP,[86] as in the calculations presented in references.[21,60]

To further assess the robustness of the DO-MOM method in cases of orbital degeneracy, the potential energy curves (PECs) of the lowest $^1\Pi(\sigma \rightarrow \pi^*)$ and $^1\Delta(\pi \rightarrow \pi^*)$ excited states

of carbon monoxide are calculated around the conical intersection. The DO-MOM PECs and analytical atomic forces are compared with PECs and forces obtained using an SCF-MOM method where convergence is attained through Gaussian smearing of both the hole and the excited electron.[24] Let N denote the number of valence electrons described explicitly and M the total number of orbitals included in the calculation. At each SCF step, the hole i and the excited orbital a are determined through the maximum overlap criterion and the occupation numbers of the n lowest N orbitals and those of the m orbitals from $N+1$ to M are modified according to:

$$f_n(\epsilon_n) = 1 - s_i(\epsilon_n) \tag{23}$$

$$f_m(\epsilon_m) = s_a(\epsilon_m) \tag{24}$$

where $s_i(\epsilon_n)$ and $s_a(\epsilon_m)$ are Gaussian functions of the KS eigenvalues:

$$s_i(\epsilon_n) = \frac{1}{A_i} \exp\left[-\frac{(\epsilon_n - \epsilon_i)^2}{2\sigma^2}\right], \qquad s_a(\epsilon_m) = \frac{1}{A_a} \exp\left[-\frac{(\epsilon_m - \epsilon_a)^2}{2\sigma^2}\right] \tag{25}$$

with the normalization factors being such that the total number of electrons is conserved. The width σ is chosen as 0.01 eV at the beginning of the SCF and then it is increased by 0.02 eV every 40th iterations, until convergence is reached. A similar electronic smearing technique has been used before to stabilize the SCF convergence in DFT calculations of PECs[33] and Born-Oppenheimer molecular dynamics simulations with DFT atomic forces.[24,49,87] For all the calculated points of both DO-MOM and SCF-MOM PECs, the guess orbitals are from a ground-state calculation at the reference geometry,[81] where the interatomic distance is 1.134 Å. The DO-MOM and SCF-MOM calculations of the PECs of carbon monoxide use a dzp basis[73] (default in GPAW).

4.2 Self-Interaction Corrected Calculations

DO-MOM calculations were carried out of the ground and first three lowest excited states of the hydrogen atom and of the ground and $^1\Sigma_g^+(1\sigma_g^2 \rightarrow 1\sigma_u^2)$ doubly excited state of the dihydrogen molecule using both PBE and SIC-PBE. The basis sets are *daug*-cc-pV6Z excluding *g*- and *h*-type functions for hydrogen, and *aug*-mcc-pVQZ excluding *f*-type functions for dihydrogen, which leads to an excitation energy converged to within ∼0.01 eV (see Figure S1 in the Supporting Information) The interatomic distance in dihydrogen is set to 1.4 Å as in reference.[23]

5 Results

5.1 Convergence Tests

5.1.1 Benchmarks on Small and Medium Size Molecules

The results of the convergence tests on 55 singlet and 34 triplet excited states of small and medium size molecules are reported in Tables 1 and 2. The average, maximum and

Table 1: Convergence properties of the SCF-MOM and DO-MOM methods for 52 singlet excited states of molecules from the benchmark set in reference[81] plus the lowest excited states of acetone, benzene and naphtalene. For the DO-MOM methods, one iteration corresponds to one energy and gradient evaluation, while for SCF-MOM it represents one energy evaluation and finding the solution for the eigendecomposition of the Hamiltonian matrix.

	SCF-MOM	DO-MOM		
		L-BFGS	L-Powell	L-SR1
Convergence failures	17	2	10	0
Avg no. iterations	22.9	13.9	20.5	12.3
Max no. iterations	96	32	69	17
Min no. iterations	15	9	9	9
Local saddle points	0	1	1	1

minimum number of iterations are reported after excluding the cases that do not converge

Table 2: Convergence properties of the SCF-MOM and DO-MOM methods for 34 triplet states of molecules from the benchmark set in reference.[81] The calculations corresponding to one iteration are the same as in Table 1.

	SCF-MOM	DO-MOM		
		L-BFGS	L-Powell	L-SR1
Convergence failures	10	0	7	0
Avg no. iterations	26.6	15.1	23.0	12.4
Max no. iterations	121	35	44	16
Min no. iterations	15	9	11	10
Local saddle points	0	2	1	4

for any of the methods. Figure 1 shows the number of iterations needed to converge the singlet states.

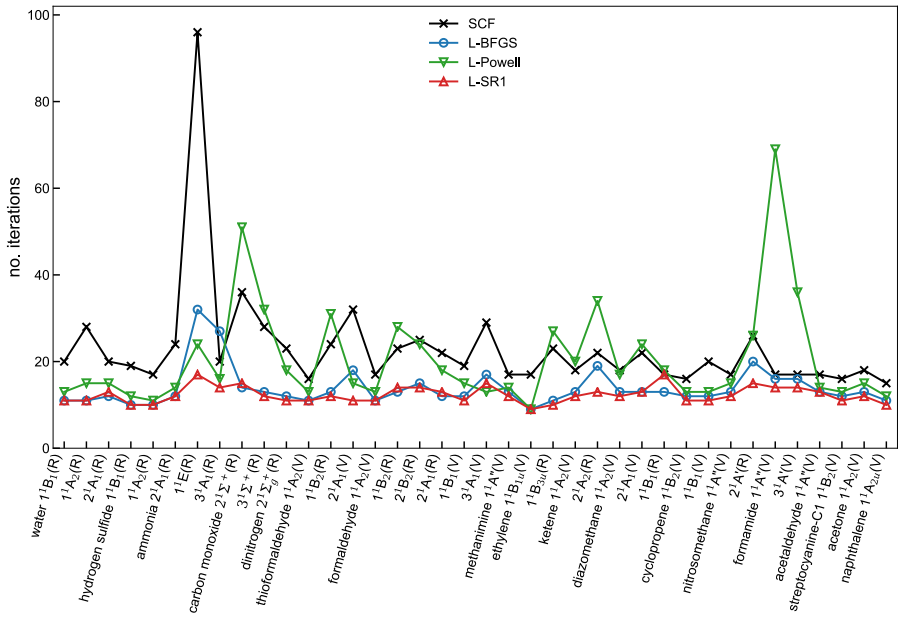

Figure 1: Number of iterations needed to reach convergence of the singlet excited states.

SCF-MOM fails to converge within the maximum number of iterations threshold in about

30% of the cases. All the quasi-Newton algorithms employed within the DO-MOM framework are more robust and show a faster rate of convergence than SCF-MOM. The best performance is obtained with L-SR1, for which all calculations converge, and convergence takes on average about 11 and 14 iterations less than SCF-MOM for singlet and triplet states, respectively. L-BFGS also performs well, being able to converge in all cases except two (the $^1\Delta(\pi \rightarrow \pi^*)$ states of carbon monoxide and dinitrogen). The limited-memory Powell inverse Hessian update in DO-MOM is considerably less efficient than L-SR1 and L-BFGS. L-Powell can have a slow rate of convergence close to a stationary point, which in many cases precludes convergence within the maximum number of allowed iterations. We have tested different combinations of limited-memory inverse Hessian updates by considering some of the cases that are most difficult to converge (see Figure 2). The combination of L-SR1 and L-Powell according to the Bofill approach (see eqs. 20 and 21) is found to have similar performance as L-Powell. This is consistent with the fact that in the Bofill approach ϕ^k, representing the weight of the Powell update, tends to 1 when the optimization approaches a stationary point. We have also tested a combination of L-SR1 and L-BFGS updates using the Bofill factor.[88] This approach does not lead to better convergence compared to L-SR1.

Figure 2 shows the number of iterations needed by DO-MOM with SR1 update to converge the singlet excited states for which the other methods fail. On average these difficult cases require more iterations than the cases presented in Figure 1. Among the states that are difficult to converge are those where excitation occurs from or to a degenerate pair of π orbitals, such as the Π states of hydrogen chloride, carbon monoxide and dinitrogen, while others are high-lying Rydberg states, most of which involve near-degenerate p-type Rydberg orbitals. A particularly challenging situation arises when both the donor and the acceptor orbitals involved in the excitation belong to degenerate pairs, as for the $\Delta(\pi \rightarrow \pi^*)$ states of carbon monoxide and dinitrogen. In this case, all methods except L-SR1 fail to converge. The other DO-MOM methods exhibit oscillations between different critical points, failing to converge to the desired solution. The failure of SCF when degenerate orbitals are un-

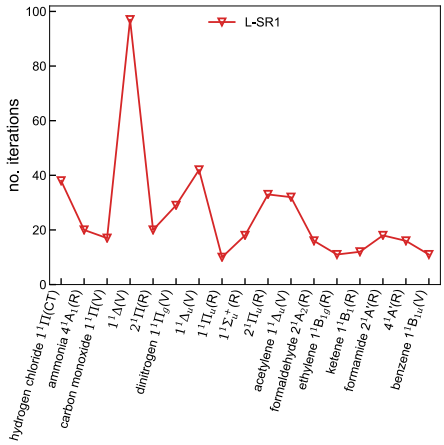

Figure 2: Total number of iterations when DO-MOM with L-SR1 is used to converge the singlet excited states for which SCF-MOM fails.

equally occupied is discussed in detail below. In about 30% of the calculations that do not converge with SCF-MOM, occupied orbitals can mix with lower-lying empty orbitals of the same symmetry. On the other hand, the L-SR1 method is able to converge all these cases. The properties of DO-MOM when orbitals involved in the excitation are allowed to mix are analyzed in more detail in the following section, where calculations of two totally symmetric excited states of the nitrobenzene molecule are presented.

The SCF convergence problems for states where degenerate orbitals are unequally occupied arise because the electron density represented by a single determinant of KS orbitals obtained at each step is not well defined and this can lead to large changes in the orbitals involved in the excitation from one step to another, and oscillations between different critical points.[40] Several options can be used to help convergence of the SCF: different mixing of the density, alternative DIIS extrapolation techniques,[89] or tuning modifications such as damping, level shifting of the iterations, or electronic smearing.[37] However, DO is able to follow the same solution more consistently without such modifications. It can obtain convergence in these difficult cases if the chosen quasi-Newton method guarantees sufficiently accurate Hes-

sian updates for the given form of the preconditioner, as shown here. The robustness of the DO approach in calculations involving orbital degeneracies has been previously recognized for ground states of systems with vanishing HOMO-LUMO gap.[40]

For some excited states, a DO method can converge on a solution with higher energy than the solution obtained by SCF or another DO method for the same excited state. For example, the solution obtained for the 1^1E(n $\rightarrow$ 3p) state of ammonia with L-BFGS or L-SR1 DO-MOM lies $\sim$0.03 eV higher in energy with respect to the SCF-MOM solution. The occurrence of higher-energy solutions, which we refer to as "local saddle points", is indicated in Tables 1 and 2. We stress that the multiple solutions that are obtained for a particular case are all saddle points of the same order and correspond to the same excited state. Multiple solutions corresponding to the same excited state are found to differ in the orientation of the highest occupied molecular orbitals (see Figures S2 to S6 in the Supporting Information). Similar to what is observed here for saddle points, the geometric direct minimization method of reference[40] exhibits a tendency to converge on local minima of energy functionals compared to SCF minimizers. Defining the "optimal" approximation to an excited state among multiple variational solutions might not be trivial. Indeed, variational solutions of a nonlinear optimization are in general not orthogonal to one another, and hence higher solutions are not necessarily upper bounds to the exact excited states, but only upper bounds to the ground state.[90] Besides, for many practical applications, such as calculations of PECs or molecular dynamics, one is usually only interested in consistently converging on the same stationary point. For these cases, DO-MOM can be used without modifications. For cases in which the lowest energy saddle point of a given excited state is desired, a possible strategy could be to combine the DO approach with techniques for guiding the convergence towards a global solution, such as the one presented in reference.[91]

Finally, we note that due to the small size of the molecules considered here, the computational effort of SCF-MOM and DO-MOM is comparable, as indicated by similar values of the elapsed time per iteration. For larger systems, care must be taken that the memory of

the quasi-Newton algorithm used within DO-MOM, which here is chosen as $m = 20$, does not degrade the computational performance of the method. From test calculations, where we compare the convergence of L-BFGS and L-SR1 with different levels of memory, we find that L-SR1 tends to become less robust with lower memory faster than L-BFGS. Therefore, for large systems, L-BFGS might represent the best compromise between speed of convergence and computational effort among the various limited-memory inverse Hessian update schemes.

5.1.2 Nitrobenzene

Figure 3 illustrates the frontier molecular orbitals involved in the electronic transitions that lead to the $^1A_1(n_\pi \rightarrow \pi'^*)$ and $^1A_1(\pi' \rightarrow \pi^*)$ excited states of nitrobenzene. Both states

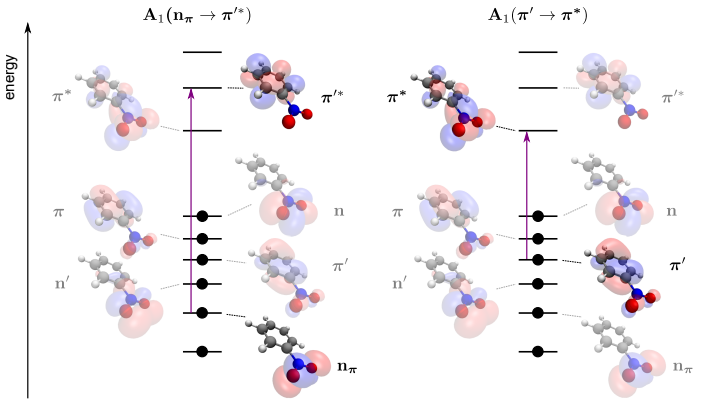

Figure 3: Ground-state frontier molecular orbitals of nitrobenzene and depiction of the electronic transitions involved in the $^1A_1(n_\pi \rightarrow \pi'^*)$ (Left) and $^1A_1(\pi' \rightarrow \pi^*)$ (Right) excited states. The labels of the orbitals are according to the notation from reference.[60] The orbital surfaces are drawn at an isovalue of 0.1 $Å^{-3/2}$.

have charge-transfer character: in the case of the $^1A_1(n_\pi \rightarrow \pi'^*)$ state, one electron moves from the nitro group to the benzene ring, while in the case of the $^1A_1(\pi' \rightarrow \pi^*)$ state, the direction of the charge transfer is reversed. Figure 3 also schematically illustrates that the highest occupied orbitals, including the orbital from which excitation occurs, are all closely

spaced in energy, covering a range of around 1 eV, despite being localized on different regions of the molecule. Charge transfer from such a subset of closely spaced orbitals is expected to be accompanied by a change of the energy ordering of the occupied orbitals.

Hait *et al.*[21] and Mewes *et al.*[60] have shown that SCF-MOM-based techniques fail to converge to the $^1A_1(n_\pi \rightarrow \pi'^*)$ and $^1A_1(\pi' \rightarrow \pi^*)$ states, respectively. When the overlaps used to find the occupation numbers with the MOM at one step are computed with respect to the orbitals from the previous step, collapse to the ground state occurs; while if the overlaps are computed with respect to the initial set of orbitals, the iterative procedure does not converge. In accord with this, our SCF-MOM calculations exhibit large and rapid oscillations without convergence in 300 iterations. This failure is likely caused by the presence of orbitals energetically close to the n_π and π' orbitals from which excitation occurs, and to rearrangements in the order of the orbital energy levels. DO-MOM, however, is able to converge both of these challenging cases.

Figure 4 shows the convergence of energy and gradient in a DO-MOM calculation of the $^1A_1(n_\pi \rightarrow \pi'^*)$ state using the L-BFGS method, where the preconditioner is updated after the MOM determines a change in the occupation numbers. After 13 steps of the

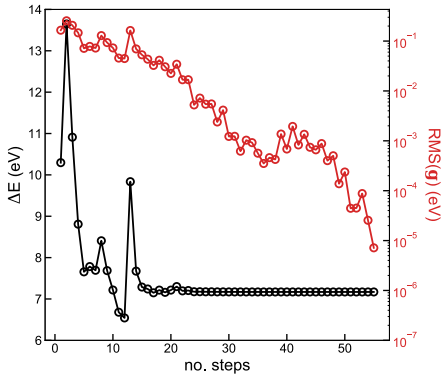

Figure 4: Convergence of excitation energy and root mean square of the gradient in a DO-MOM calculation of the $^1A_1(n_\pi \rightarrow \pi'^*)$ excited state of nitrobenzene using L-BFGS.

optimization, a change of the character of the occupied orbitals is detected and, as a result, the MOM induces a change in the occupation numbers, which restores the character of the initial guess. Application of the MOM constraints is accompanied by a jump in the energy as can be observed from Figure 4. After that, the energy is converged to 10^{-6} eV in ∼50 steps. While the approximate Hessian at the initial guess has six negative eigenvalues, the converged solution is a ninth-order saddle point. This is a consequence of a significant rearrangement in the ordering of the orbitals induced by the charge transfer, which stabilizes the orbitals localized on the nitro group, including the hole, and destabilizes the orbitals localized on the benzene ring. When L-BFGS is used, it is essential to apply the MOM constraints and update the preconditioner in order to achieve convergence to the target excited state. This is illustrated in Figure 5, which shows a DO calculation with L-BFGS starting from the same initial guess as in Figure 4 but where the MOM is not applied and the preconditioner is not updated. In this case, the hole which has an initial n_π character, acquires during the

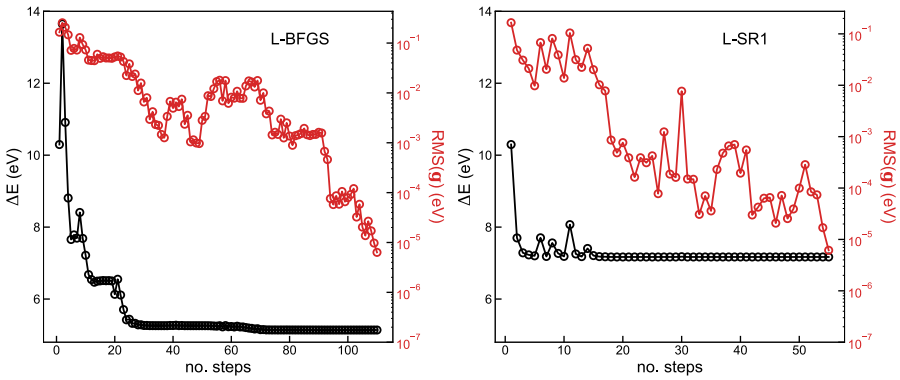

Figure 5: Convergence of excitation energy and root mean square of the gradient in DO calculations using L-BFGS (Left) and L-SR1 (Right) without the MOM and with a preconditioner fixed at the guess for the $^1A_1(n_\pi \rightarrow \pi'^*)$ state of nitrobenzene.

DO the character of the π orbital depicted in Figure 3 (the n_π and π orbitals are allowed to mix because they both belong to the A_2 irreducible representation in the C_{2v} point group symmetry), and the calculation eventually collapses to a third-order saddle point. Figure 5

also shows a DO calculation without the MOM and with a fixed preconditioner when the approximate inverse Hessian is updated using L-SR1. Despite an initial approximate Hessian with a lower number of negative eigenvalues compared to the Hessian of the target solution, the DO with L-SR1 is able to converge to the ninth-order saddle point corresponding to the $^1A_1(n_\pi \rightarrow \pi'^*)$ state. This can be explained with the ability of L-SR1 to develop negative eigenvalues, while L-BFGS cannot. The squared gradient minimization method of reference[21] is also able to converge to the $^1A_1(n_\pi \rightarrow \pi'^*)$ state of nitrobenzene. However, due to the need to compute the derivative of the squared norm of the gradient at each step, the minimization involves larger computational effort per iteration than the present DO-MOM calculations.

In the case of the $^1A_1(\pi' \rightarrow \pi^*)$ excited state, the converged solution is found to be a fourth-order saddle point, while the approximate Hessian at the initial guess generated from the ground-state orbitals (see Figure 3) has three negative eigenvalues. As for the $^1A_1(n_\pi \rightarrow \pi'^*)$ state, a DO calculation with L-BFGS can converge to the target solution only if the MOM is used and the preconditioner updated during the optimization.[58] Figure 6 shows the convergence of DO-MOM calculations using L-SR1 with and without preconditioner. It

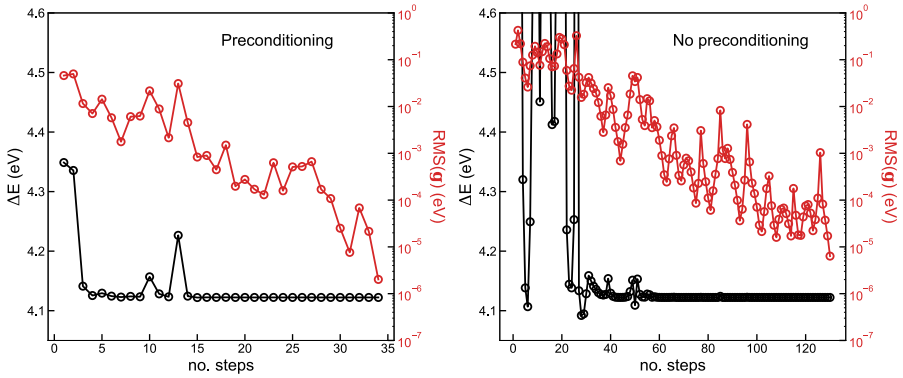

Figure 6: Convergence of excitation energy and root mean square of the gradient in DO-MOM calculations of the $^1A_1(\pi' \rightarrow \pi^*)$ excited state of nitrobenzene using L-SR1 with (Left) and without (Right) preconditioner.

is found that DO-MOM with L-SR1 is able to converge to the $^1A_1(\pi' \rightarrow \pi^*)$ state even

without the use of a preconditioner, although large oscillations of the energy are observed at the beginning of the optimization and almost four times as many steps are required to achieve convergence compared to a calculation that uses the preconditioner. These results show that the L-SR1 method developed in the present work is less sensitive to the quality of the preconditioner and is able to build a better approximation to the inverse electronic Hessian when used in optimizations of excited states within DFT compared to a standard implementation of the most used L-BFGS quasi-Newton algorithm.

All calculations presented above use a maximum step length, $p_{\max}$, of 0.20, which is the value found optimal in most of the cases. However, a $p_{\max}$ of 0.25 leads to smaller oscillations at the beginning of the optimization and faster convergence in the case of the DO-MOM calculations with L-SR1 (see Figures S7 and S8 in the Supporting Information). The use of a fixed allowed step length is a limitation of the current implementation. To ensure smooth and monotonic convergence for a broad range of systems, a trust region scheme could be introduced.

5.1.3 Potential Energy Curves of Carbon Monoxide

The electron configuration of the ground state of carbon monoxide is $(1\sigma)^2(1\sigma^*)^2(2\sigma)^2(2\sigma^*)^2(1\pi)^4(3\sigma)^2(1\pi^*)^0(3\sigma^*)^0$. The lowest singlet excited states arise from $\sigma \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ single-electron excitations. Among the states with these configurations, the $1^1\Pi(\text{n}\rightarrow \pi^*)$ and $1^1\Delta(\pi \rightarrow \pi^*)$ can be approximated using a single determinant.

KS DFT has several difficulties describing the $1^1\Pi(\text{n}\rightarrow \pi^*)$ and $1^1\Delta(\pi \rightarrow \pi^*)$ states and their conical intersection. Firstly, the determinant obtained from a single-electron transition between orbitals of the same spin has a broken spin symmetry, since the pure singlet open-shell state is a symmetry-adapted linear combination of two determinants with the same configuration. Secondly, KS DFT neglects the multireference character of the wave functions arising from mixing of configurations involving the degenerate pairs of 1π and $1\pi^*$

orbitals. Finally, at the conical intersection the 1π orbitals become degenerate with the 3σ orbital, further increasing the multireference character of the states. The strong static correlation prevents the SCF-MOM method with integer occupation numbers from converging. Convergence can be achieved by smearing the hole and excited electron over the degenerate orbitals. We emphasize that the aim here is not to assess the accuracy of DFT with KS functionals in the description of the excited states, for which highly accurate multireference calculations are available when the molecules are small, but rather to demonstrate the ability of the DO-MOM method to handle a challenging case without *ad hoc* modifications to achieve convergence.

The PECs of the $1^1\Pi(\mathrm{n}\rightarrow \pi^*)$ and $1^1\Delta(\pi \rightarrow \pi^*)$ states of carbon monoxide around the conical intersection computed using SCF-MOM with Gaussian smearing and DO-MOM are shown in Figure 7 together with the analytical atomic forces for selected points on the $1^1\Delta(\pi \rightarrow \pi^*)$ curves. For the Gaussian smearing SCF-MOM calculations, far from the conical intersection, the occupation numbers of the 1π orbitals are 1 ($1^1\Pi(\mathrm{n}\rightarrow \pi^*)$ state) or 0.5 ($1^1\Delta(\pi \rightarrow \pi^*)$), while the occupation of 3σ is either 0 ($1^1\Pi(\mathrm{n}\rightarrow \pi^*)$) or 1 ($1^1\Delta(\pi \rightarrow \pi^*)$). Close to the conical intersection, the hole can be smeared over both the 1π and 3σ orbitals (see Tables S3 and S4 in the Supporting Information). When this happens, the SCF-MOM PECs display some artefacts. The PEC of $1^1\Pi(\mathrm{n}\rightarrow \pi^*)$ shows discontinuities around the three points for which the smearing is largest. The only point of the PEC of $1^1\Delta(\pi \rightarrow \pi^*)$ for which the hole is smeared over three orbitals coincides approximately with the point of crossing of the two curves (R$\sim$1.56 Å). The analytical forces computed at this point are not consistent with the slope of the $1^1\Delta(\pi \rightarrow \pi^*)$ curve.

On the other hand, the curves obtained with DO-MOM PECs and integer occupation numbers do not exhibit discontinuities and the computed atomic forces are consistent with the slopes of the curves. We further note that Gaussian smearing SCF-MOM converges on higher-energy solutions, with the $1^1\Pi(\mathrm{n}\rightarrow \pi^*)$ and $1^1\Delta(\pi \rightarrow \pi^*)$ PECs computed with DO-MOM lying respectively $\sim$0.1 and $\sim$0.3 eV lower in energy with respect to the SCF-MOM

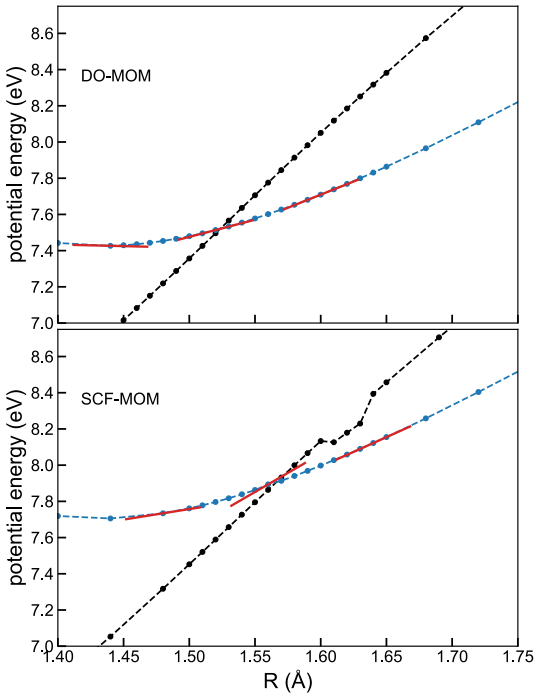

Figure 7: Potential energy curves of the $1^1\Pi(\mathrm{n}\rightarrow\pi^*)$ (black) and $1^1\Delta(\pi\rightarrow\pi^*)$ (blue) excited states of carbon monoxide computed with the DO-MOM method (top) and a variant of SCF-MOM using Gaussian smearing of the hole and excited electron (bottom). The red lines represent the analytical atomic forces at selected points.

PECs. This also affects the relative position of the two conical intersections.

Electronic smearing is often employed together with the SCF method to converge excited-state DFT calculations, especially in molecular dynamics simulations.[22,24,87,92,93] The results presented here show that one needs to carefully check whether artefacts are introduced due to the smearing. The DO-MOM method can converge energy and forces even in cases of

degeneracies without the need of smearing.

5.2 Calculations with Self-Interaction Correction

5.2.1 Hydrogen Atom

Table 3 reports the energy and eigenvalue of the occupied orbital for the ground state and each of the three lowest excited states of the hydrogen atom computed with PBE and SIC-PBE using DO-MOM, as compared to the experimental values of the ionization energy. The PBE functional displays a well-known systematic underestimation of the energy of the

Table 3: Total energies (E) and orbital eigenvalues (ϵ) of the ground and the three lowest excited states of the hydrogen atom computed using DO-MOM with PBE and SIC-PBE and experimental values of the ionization (I) energy.[94] The values in parenthesis represent the differences with respect to the experimental energies. All values are in eV.

	PBE		SIC-PBE		Exp.[94]
Electronic state	E	ϵ	E	ϵ	-I
1s	-13.60(0.00)	-7.59(6.01)	-13.60(0.00)	-13.60(0.00)	-13.60
2s	-3.70(-0.30)	-2.23(1.17)	-3.40(0.00)	-3.40(0.00)	-3.40
2p	-3.81(-0.41)	-1.91(1.49)	-3.40(0.00)	-3.40(0.00)	-3.40
3s	-1.73(-0.22)	-1.13(0.38)	-1.50(0.01)	-1.50(0.01)	-1.51

excited states (linear-response TDDFT with PBE predicts no bound Rydberg states for the hydrogen atom).[15] The inability of excited-state DFT with KS semi-local functionals to describe Rydberg series of atoms has been traced back to the fact that the long-range form of the effective potential is incorrect (see, for example, reference[15]).

The SIE of a one-electron system cancels exactly for the SIC-PBE functional. As a result, the SIC-PBE energy values are accurate for the basis set used. Furthermore, for a given state the eigenvalue of the occupied orbital coincides with the total energy and is independent of the occupation number, i.e. for a one-electron system the SIC functional restores the derivative discontinuity that is missing in the approximate functional.[51]

5.2.2 Dihydrogen Molecule

Gill *et al.*[23] have recently reported SCF-MOM calculations of the $^1\Sigma_g^+(1\sigma_g^2 \rightarrow 1\sigma_u^2)$ doubly excited state of dihydrogen using xc fuctionals for several choices of the fraction of exact exchange. Their results show that GGA and hydrid functionals with small fraction of HF exchange severely underestimate the excitation energy because the SIE in the excited state is significantly larger than in the ground state. The DO-MOM PBE calculation of the $^1\Sigma_g^+(1\sigma_g^2 \rightarrow 1\sigma_u^2)$ state is in line with this observation. The PBE excitation energy is 27.25 eV, with a deviation of 1.50 eV from the full configuration interaction (CI) result of reference[23] (28.75 eV). The one-electron SIE calculated according to eq. 9 using the density and the orbitals converged with PBE is ∼-1.69 eV, compared to an SIE of ∼-0.10 eV for the ground state. Therefore, most of the error in the excitation energy comes from an imbalance in the SIEs. If the self-interaction correction is applied non-variationally, the resulting excitation energy is equal to 28.83 eV, which is closer to the full CI result. Further improvement is obtained with the fully variational SIC-PBE calculations giving an excitation energy of 28.79 eV, only 0.04 eV larger than the full CI energy. The remaining error is due to the approximate treatment of correlation and to the use of different basis sets in the DO-MOM SIC-PBE and full CI calculations.

These results illustrate how self-interaction correction in variational DFT calculations of excited states can be an effective route to correct the unbalanced SIE between ground and excited states in calculations based on semi-local functionals.

6 Concluding Remarks

DO has long been known to be a robust and computationally competitive alternative to SCF in ground-state calculations.[40,44,45] Calculations using single-determinant excited-state DFT and DO have been limited to minimization of the squared norm of the gradient, while DO of saddle points has been considered to be too difficult, due to the need of a better approxi-

mation to the Hessian and the risk of variational collapse. Here, a DO method is presented that overcomes these challenges by: (1) employing a newly developed limited-memory formulation of quasi-Newton SR1 inverse Hessian update (L-SR1) that is able to build a better approximation to the Hessian for saddle-point searches than the L-BFGS update commonly employed in minimization, and (2) avoiding variational collapse by using MOM constraints. Since only one gradient evaluation is required at each step, the computational cost is the same as for ground-state calculations. We further note that even if DO-MOM has been presented here in the context of excited-state DFT, it can be applied with any other method where the objective is to optimize a set of orthogonal orbitals, provided that the appropriate form of the **L** matrix is used in the expression of the gradient, eq. A.6.

We find that DO-MOM in combination with a localized basis set representation of the orbitals outperforms the conventional SCF-MOM approach both in terms of robustness and speed of convergence for a benchmark set of 89 excited states. The best performance is obtained with the L-SR1 algorithm when using a memory of 20 iterations. Furthermore, tests on challenging charge-transfer excited states of nitrobenzene show that L-SR1 is more robust than L-BFGS for saddle-point optimization, being less dependent on the preconditioner. Therefore, DO-MOM with L-SR1 is a promising method for excited-state calculations of large systems, where diagonalization of the Hamiltonian matrix needed to compute the preconditioner is prohibitive. These tests were limited to valence and Rydberg excited states of small and medium size molecules. In future work these tests will be extended to include larger molecules and long-range charge-transfer states.

DO-MOM is able to converge single-determinant excited states close to conical intersections, which often require fractional occupations in SCF approaches, as demonstrated here for the first two singlet excited states of carbon monoxide. This makes it possible to assess more rigorously the applicability of single-determinant density functional methods for modelling conical intersections as compared to methods that explicitly take into account static correlation effects. Crucially, such benchmarks are currently lacking despite

the fact that excited-state DFT has been proposed in the context of nonadiabatic dynamics simulations.[19,22,33] Formally, the single-determinant approximation is a clear limitation of excited-state DFT. Multiconfigurational effects can be taken into account within, for example, ensemble DFT.[95] Extending the applicability of DO-MOM requires handling the simultaneous optimization of the orbitals and the occupation numbers.[96]

Finally, our implementation of DO-MOM can be used with non-unitary invariant functionals, such as SIC functionals. As pointed out earlier,[11,50] the accuracy of excitation energies obtained with semi-local functionals can be affected by different amounts of SIE in the ground and excited state. The accurate results obtained from the calculations on the lowest doubly excited state of dihydrogen represent a preliminary indication that SIC functionals can help alleviate this issue. However, tests on more complex systems are needed to draw a general conclusion on the performance of SIC functionals. Benchmarks on excited states of molecules, including Rydberg states, are currently ongoing.

Appendix

A Exponential Transformation

The spin index is omitted here for simplicity as the exponential transformation does not mix orbitals with different spin quantum number. An initial guess for the optimal orbitals (reference orbitals) is expanded into a linear combination of M localised basis functions:

$$\phi_p(\mathbf{r}) = \sum_{\mu=1}^{M} C_{\mu p}\chi_\mu(\mathbf{r}) \tag{A.1}$$

The coefficients of this expansion must satisfy the orthonormality constraints:

$$\sum_{\mu\nu} C^*_{\mu p} S_{\mu\nu} C_{\nu q} = \delta_{pq} \tag{A.2}$$

with:

$$S_{\mu\nu} = \int \chi_\mu^*(\mathbf{r})\chi_\nu(\mathbf{r})d\mathbf{r} \tag{A.3}$$

The optimal orbital coefficients corresponding to an extremum of the energy functional can be found through a unitary transformation of the $C_{\mu p}$:

$$O_{\mu k} = \sum_{p=1}^{M} C_{\mu p} \left[e^{\boldsymbol{\theta}}\right]_{pk} \tag{A.4}$$

The $M \times M$ anti-Hermitian matrix $\boldsymbol{\theta}$ contains the parameters that describe rotations of the orbitals and is parametrized as:

$$\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\theta}_{\mathrm{oo}} & \boldsymbol{\theta}_{\mathrm{ov}} \\ -\boldsymbol{\theta}_{\mathrm{ov}}^\dagger & \mathbf{0} \end{pmatrix} \tag{A.5}$$

where the $N \times N$ block $\boldsymbol{\theta}_{\mathrm{oo}}$ contains the parameters that describe rotations mixing occupied-occupied (oo) orbitals, while the $N \times (M-N)$ blocks $\boldsymbol{\theta}_{\mathrm{ov}}$ mix occupied-virtual (ov) orbitals. The total energy does not depend on rotations among the virtual orbitals and, as a result, the virtual-virtual (vv) block is set to zero. Since $\boldsymbol{\theta}$ is anti-Hermitian, the total number of free parameters is $N(2M-N)$. For KS functionals, the energy is invariant with respect to unitary transformation of equally occupied orbitals and, therefore, the $\boldsymbol{\theta}_{\mathrm{oo}}$ block can be set to zero without loss of generality.[41] In this case, the number of degrees of freedom is reduced to $2N(M-N)$ and the matrix exponential can be calculated using the equation given by Hutter *et al.*[41] For SIC functionals, $\boldsymbol{\theta}_{\mathrm{oo}}$ cannot be set to zero.[54] In this case, the scaling and squaring algorithm of Al-Mohy and Higham[97] as implemented in the SciPy library[98] is used to evaluate the matrix exponential.

In order to carry out the optimization efficiently, using a quasi-Newton method, or any other gradient-based algorithm, the gradient of the energy with respect to the $\{\theta_{ij}\}$ rotation

parameters is needed:

$$\frac{\partial E}{\partial \theta^*_{ij}} = \frac{2-\delta_{ij}}{2}\left[\int_0^1 e^{t\boldsymbol{\theta}}\mathbf{L}e^{-t\boldsymbol{\theta}}dt\right]_{ij} \tag{A.6}$$

where the matrix $\mathbf{L}$ has elements:

$$L_{lk} = (f_l - f_k)H_{lk} + f_k V_{lk} - f_l V^*_{kl} \tag{A.7}$$

In eq. A.7, the H_{lk} are the elements of the Hamiltonian matrix in the basis of optimal orbitals:

$$H_{lk} = \sum_{\mu\nu} O^*_{\mu l} H_{\mu\nu} O_{\nu k}, \qquad H_{\mu\nu} = \int \chi^*_\mu(\mathbf{r})\mathbf{H}_{\mathrm{KS}}\chi_\nu(\mathbf{r})d\mathbf{r} \tag{A.8}$$

while the V_{lk} are the elements of orbital-density dependent potentials due to SIC:

$$V_{lk} = \sum_{\mu\nu} O^*_{\mu l} V^k_{\mu\nu} O_{\nu k}, \qquad V^k_{\mu\nu} = \int \chi^*_\mu(\mathbf{r})\mathbf{V}_k\chi_\nu(\mathbf{r})d\mathbf{r} \tag{A.9}$$

For KS functionals, the V_{lk} become zero.

The integral in eq. A.6 can be expanded in a series:

$$\int_0^1 e^{t\boldsymbol{\theta}}\mathbf{L}e^{-t\boldsymbol{\theta}}dt = \mathbf{L} + \frac{1}{2!}[\boldsymbol{\theta},\mathbf{L}] + \frac{1}{3!}[\boldsymbol{\theta},[\boldsymbol{\theta},\mathbf{L}]] + \ldots \tag{A.10}$$

When the norm of the matrix $\boldsymbol{\theta}$ is small ($\|\boldsymbol{\theta}\| \ll 1$), the energy gradient can be estimated accurately using only the first term of this series. During the optimization, the coefficients of the reference orbitals are updated with those of the optimal or canonical orbitals at regular step intervals and, in addition, each time the MOM (see next section) changes the orbital occupations. At every update, the $\boldsymbol{\theta}$ matrix is reset to zero; therefore, these updates avoid $\|\boldsymbol{\theta}\|$ becoming too large, thus allowing to use only the first term of the series in eq. A.10 to estimate the gradient.

B Limited-Memory Powell Update

The Powell inverse Hessian update formula in compact form is:[77]

$$\mathbf{B}_{\mathrm{P}}^{(k+1)} = \mathbf{B}^{(k)} + \mathbf{j}^{(k)}\mathbf{u}^{T(k)} + \mathbf{u}^{(k)}\left[\mathbf{j}^{T(k)} - \left(\mathbf{y}^{T(k)}\mathbf{j}^{(k)}\right)\mathbf{u}^{T(k)}\right] \tag{B.1}$$

where:

$$\mathbf{u}^{(k)} = \frac{\mathbf{y}^{(k)}}{\mathbf{y}^{T(k)}\mathbf{y}^{(k)}} \tag{B.2}$$

and $\mathbf{j}^{(k)}$ and $\mathbf{y}^{(k)}$ are defined as in eqs. 17 and 18, respectively. The product $\mathbf{B}_{\mathrm{P}}^{(k)}\mathbf{v}^{(k)}$, where $\mathbf{v}^{(k)}$ is a vector, can be computed using the following recursive formula:

$$\begin{aligned}\mathbf{B}_{\mathrm{P}}^{(k)}\mathbf{v}^{(k)} =& \mathbf{B}_{0}^{(k)}\mathbf{v}^{(k)} + \sum_{i=k-m}^{k-1}\mathbf{j}^{(i)}\mathbf{u}^{T(i)}\mathbf{v}^{(k)} \\ &+ \sum_{i=k-m}^{k-1}\left\{\mathbf{u}^{(i)}\left[\mathbf{j}^{T(i)}\mathbf{v}^{(k)} - \left(\mathbf{y}^{T(i)}\mathbf{j}^{(i)}\right)\mathbf{u}^{T(i)}\mathbf{v}^{(k)}\right]\right\}\end{aligned} \tag{B.3}$$

The L-Powell algorithm is obtained by replacing the use of eq. 19 with eq. B.3 in Algorithm 1, which requires storing the vector $\mathbf{u}^{(k)}$ in addition to $\mathbf{j}^{(k)}$ at each step (see also Algorithm 1 in reference[58]).

C Maximum Overlap Method

The MOM method is used to ensure that the character of the occupied optimal orbitals is consistent with the initial guess and to choose the occupation numbers of the canonical orbitals whenever they are needed, e.g. when updating the preconditioner according to eq. 22. The coefficients of the reference orbitals for the MOM, which are used to compute the overlaps with the orbitals at a given step, are chosen as the coefficients $C_{\mu p}$ of the orbitals

of the initial guess, and are fixed. Accordingly, the overlap matrix at step k has elements:

$$\Omega_{pl}^{(k)} = \sum_{\nu\mu} C_{p\mu}^{*} S_{\mu\nu} O_{\nu l}^{(k)} \tag{C.1}$$

where $S_{\mu\nu}$ is defined according to eq. A.3. The occupied orbitals are chosen as those with the largest projections onto the occupied subspace of the initial guess orbitals:

$$\omega_{l}^{(k)} = \left[\sum_{p=1}^{N} \left(\Omega_{pl}^{(k)} \right)^{2} \right]^{\frac{1}{2}} \tag{C.2}$$

If the MOM detects a change of the character of the occupied optimal orbitals, the reference orbitals for the DO are updated. Analogous expressions are used to obtain the occupation numbers of the canonical orbitals when a Hamiltonian diagonalization is performed.

Acknowledgement

This work was funded by the Icelandic Research Fund (grant number 196070) and the University of Iceland Research Fund. AVI is supported by a doctoral fellowship from the University of Iceland. The authors thank Elvar Ö. Jónsson, Asmus O. Dohn and Myneni Hemanadhan for useful discussions.

Supporting Information Available

Basis set convergence tests, data of the convergence tests, additional DO-MOM calculations on nitrobenzene, additional information on the SCF-MOM calculations of the excited-state potential energy curves of carbon monoxide.

References

(1) Dreuw, A.; Head-Gordon, M. Single-Reference ab Initio Methods for the Calculation of Excited States of Large Molecules. *Chem. Rev.* **2005**, *105*, 4009–4037.

(2) Casida, M. E. In *Recent Advances in Density Functional Methods*; Chong, D. P., Ed.; World Scientific, 1995; pp 155–192.

(3) Runge, E.; Gross, E. K. U. Density-functional theory for time-dependent systems. *Phys. Rev. Lett.* **1984**, *52*, 997–1000.

(4) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, 1133–1138.

(5) Hohenberg, P.; Kohn, W. Inhomogeneous electron gas. *Phys. Rev.* **1964**, *136*, B864.

(6) Van Meer, R.; Gritsenko, O. V.; Baerends, E. J. Physical meaning of virtual kohn-sham orbitals and orbital energies: An ideal basis for the description of molecular excitations. *J. Chem. Theory Comput.* **2014**, *10*, 4432–4441.

(7) Levine, B. G.; Ko, C.; Quenneville, J.; Martínez, T. J. Conical intersections and double excitations in time-dependent density functional theory. *Mol. Phys.* **2006**, *104*, 1039–1051.

(8) Maitra, N. T.; Zhang, F.; Cave, R. J.; Burke, K. Double excitations within time-dependent density functional theory linear response. *J. Chem. Phys.* **2004**, *120*, 5932–5937.

(9) Tozer, D. J.; Handy, N. C. On the determination of excitation energies using density functional theory. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2117–2121.

(10) Huix-Rotllant, M.; Nikiforov, A.; Thiel, W.; Filatov, M. In *Density-Functional Methods for Excited States*; Ferré, N., Filatov, M., Huix-Rotllant, M., Eds.; Springer International Publishing: Cham, 2016; pp 445–476.

(11) Zhao, L.; Neuscamman, E. Density Functional Extension to Excited-State Mean-Field Theory. *J. Chem. Theory Comput.* **2019**, *16*, 164–178.

(12) Zhekova, H. R.; Seth, M.; Ziegler, T. A perspective on the relative merits of time-dependent and time-independent density functional theory in studies of the electron spectra due to transition metal complexes. An illustration through applications to copper tetrachloride and plastocyanin. *Int. J. Quantum Chem.* **2014**, *114*, 1019–1029.

(13) Ziegler, T.; Seth, M.; Krykunov, M.; Autschbach, J. A revised electronic Hessian for approximate time-dependent density functional theory. *J. Chem. Phys.* **2008**, *129*, 1–10.

(14) Seidu, I.; Krykunov, M.; Ziegler, T. Applications of time-dependent and time-independent density functional theory to Rydberg transitions. *J. Phys. Chem. A* **2015**, *119*, 5107–5116.

(15) Cheng, C. L.; Wu, Q.; Van Voorhis, T. Rydberg energies using excited state density functional theory. *J. Chem. Phys.* **2008**, *129*, 124112.

(16) Baerends, E. J.; Gritsenko, O. V.; Van Meer, R. The Kohn-Sham gap, the fundamental gap and the optical gap: The physical meaning of occupied and virtual Kohn-Sham orbital energies. *Phys. Chem. Chem. Phys.* **2013**, *15*, 16408–16425.

(17) Dreuw, A.; Head-Gordon, M. Failure of Time-Dependent Density Functional Theory for Long-Range Charge-Transfer Excited States: The Zincbacteriochlorin-Bacteriochlorin and Bacteriochlorophyll-Spheroidene Complexes. *J. Am. Chem. Soc.* **2004**, *126*, 4007–4016.

(18) Dreuw, A.; Weisman, J. L.; Head-Gordon, M. Long-range charge-transfer excited states in time-dependent density functional theory require non-local exchange. *J. Chem. Phys.* **2003**, *119*, 2943–2946.

(19) Mališ, M.; Luber, S. Trajectory Surface Hopping Nonadiabatic Molecular Dynamics with Kohn-Sham ΔSCF for Condensed-Phase Systems. *J. Chem. Theory Comput* **2020**, *16*, 4071–4086.

(20) Hait, D.; Head-Gordon, M. Highly Accurate Prediction of Core Spectra of Molecules at Density Functional Theory Cost: Attaining Sub-electronvolt Error from a Restricted Open-Shell Kohn-Sham Approach. *J. Phys. Chem. Lett.* **2020**, *11*, 775–786.

(21) Hait, D.; Head-Gordon, M. Excited state orbital optimization via minimizing the square of the gradient: General approach and application to singly and doubly excited states via density functional theory. *J. Chem. Theory Comput.* **2020**, *16*, 1699–1710.

(22) Pradhan, E.; Sato, K.; Akimov, A. V. Non-adiabatic molecular dynamics with ΔSCF excited states. *J. Phys.: Condens. Matter* **2018**, *30*, 484002.

(23) Barca, G. M.; Gilbert, A. T.; Gill, P. M. Simple Models for Difficult Electronic Excitations. *J. Chem. Theory Comput.* **2018**, *14*, 1501–1509.

(24) Levi, G.; Papai, M.; Henriksen, N. E.; Dohn, A. O.; Møller, K. B. Solution structure and ultrafast vibrational relaxation of the PtPOP complex revealed by ΔSCF-QM/MM Direct Dynamics simulations. *J. Phys. Chem. C* **2018**, *122*, 7100–7119.

(25) Ramos, P.; Pavanello, M. Low-lying excited states by constrained DFT. *J. Chem. Phys.* **2018**, *148*, 144103.

(26) Liu, J.; Zhang, Y.; Bao, P.; Yi, Y. Evaluating Electronic Couplings for Excited State Charge Transfer Based on Maximum Occupation Method ΔSCF Quasi-Adiabatic States. *J. Chem. Theory Comput.* **2017**, *13*, 843–851.

(27) Park, Y. C.; Senn, F.; Krykunov, M.; Ziegler, T. Self-Consistent Constricted Variational Theory RSCF-CV(∞)-DFT and Its Restrictions to Obtain a Numerically Stable

ΔsCF-DFT-like Method: Theory and Calculations for Triplet States. *J. Chem. Theory Comput.* **2016**, *12*, 5438–5452.

(28) Park, Y. C.; Krykunov, M.; Ziegler, T. On the relation between adiabatic time dependent density functional theory (TDDFT) and the ΔSCF-DFT method. Introducing a numerically stable ΔSCF-DFT scheme for local functionals based on constricted variational DFT. *Mol. Phys.* **2015**, *113*, 1636–1647.

(29) Peng, B.; Van Kuiken, B. E.; Ding, F.; Li, X. A guided self-consistent-field method for excited-state wave function optimization: Applications to ligand-field transitions in transition-metal complexes. *J. Chem. Theory Comput.* **2013**, *9*, 3933–3938.

(30) Maurer, R. J.; Reuter, K. Excited-state potential-energy surfaces of metal-adsorbed organic molecules from linear expansion Δ-self-consistent field density-functional theory (Δ-DFT). *J. Chem. Phys.* **2013**, *139*, 014708–014718.

(31) Himmetoglu, B.; Marchenko, A.; Dabo, I.; Cococcioni, M. Role of electronic localization in the phosphorescence of iridium sensitizing dyes. *J. Chem. Phys.* **2012**, *137*, 154309.

(32) Kowalczyk, T.; Yost, S. R.; Voorhis, T. V. Assessment of the ΔSCF density functional theory approach for electronic excitations in organic dyes. *J. Chem. Phys.* **2011**, *134*, 054128.

(33) Maurer, R. J.; Reuter, K. Assessing computationally efficient isomerization dynamics: ΔSCF density-functional theory study of azobenzene molecular switching. *J. Chem. Phys.* **2011**, *135*, 224303.

(34) Gilbert, A. T. B.; Besley, N. A.; Gill, P. M. W. Self-consistent field calculations of excited states using the maximum overlap method (MOM). *J. Phys. Chem. A* **2008**, *112*, 13164–71.

(35) Gavnholt, J.; Olsen, T.; Engelund, M.; Schiøtz, J. Delta Self-Consistent Field as a Method to Obtain Potential Energy Surfaces of Excited Molecules on Surfaces. *Phys. Rev. B* **2008**, *78*, 075441.

(36) Perdew, J. P.; Levy, M. Extrema of the density functional for the energy: Excited states from the ground-state theory. *Phys. Rev. B* **1985**, *31*, 6264–6272.

(37) Rabuck, A. D.; Scuseria, G. E. Improving self-consistent field convergence by varying occupation numbers. *J. Chem. Phys.* **1999**, *110*, 695–700.

(38) Huix-Rotllant, M.; Filatov, M.; Gozem, S.; Schapiro, I.; Olivucci, M.; Ferré, N. Assessment of density functional theory for describing the correlation effects on the ground and excited state potential energy surfaces of a retinal chromophore model. *J. Chem. Theory Comput.* **2013**, *9*, 3917–3932.

(39) Dickson, R. M.; Ziegler, T. A density functional study of the electronic spectrum of permanganate. *Int. J. Quantum Chem.* **1996**, *58*, 681–687.

(40) Voorhis, T. V.; Head-Gordon, M. A geometric approach to direct minimization. *Mol. Phys.* **2002**, *100*, 1713–1721.

(41) Hutter, J.; Parrinello, M.; Vogel, S. Exponential transformation of molecular orbitals. *J. Chem. Phys.* **1994**, *101*, 3862–3865.

(42) Head-Gordon, M.; Pople, J. A. Optimization of wave function and geometry in the finite basis Hartree-Fock method. *J. Phys. Chem.* **1988**, *92*, 3063–3069.

(43) Douady, J.; Ellinger, Y.; Subra, R.; Levy, B. Exponential transformation of molecular orbitals: A quadratically convergent SCF procedure. I. General formulation and application to closed-shell ground states. *J. Chem. Phys.* **1980**, *72*, 1452–1462.

(44) Baarman, K.; Vandevondele, J. A comparison of accelerators for direct energy minimization in electronic structure calculations. *J. Chem. Phys.* **2011**, *134*, 244104.

(45) VandeVondele, J.; Hutter, J. An efficient orbital transformation method for electronic structure calculations. *J. Chem. Phys.* **2003**, *118*, 4365–4369.

(46) Shea, J. A. R.; Gwin, E.; Neuscamman, E. A Generalized Variational Principle with Applications to Excited State Mean Field Theory. *J. Chem. Theory Comput.* **2020**, *16*, 1526–1540.

(47) Briggs, E. A.; Besley, N. A. Density Functional Theory Based Analysis of Photoinduced Electron Transfer in a Triazacryptand Based K+ Sensor. *J. Phys. Chem. A* **2015**, 2902–2907.

(48) Briggs, E. A.; Besley, N. A.; Robinson, D. QM/MM excited state molecular dynamics and fluorescence spectroscopy of BODIPY. *J. Phys. Chem. A* **2013**, *117*, 2644–2650.

(49) Levi, G.; Biasin, E.; Dohn, A. O.; Jónsson, H. On the interplay of solvent and conformational effects in simulated excited-state dynamics of a copper phenanthroline photosensitizer. *Phys. Chem. Chem. Phys.* **2019**, *22*, 748–757.

(50) Gudmundsdóttir, H.; Zhang, Y.; Weber, P. M.; Jónsson, H. Self-interaction corrected density functional calculations of molecular Rydberg states. *J. Chem. Phys.* **2013**, *139*, 194102.

(51) Perdew, J. P.; Zunger, A. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B* **1981**, *23*, 5048–5079.

(52) Zhang, Y.; Weber, P. M.; Jónsson, H. Self-Interaction Corrected Functional Calculations of a Dipole-Bound Molecular Anion. *J. Phys. Chem. Lett.* **2016**, *7*, 2068–2073.

(53) Hemanadhan, M.; Shamim, M.; Harbola, M. K. Testing an excited-state energy density functional and the associated potential with the ionization potential theorem. *J. Phys. B: At., Mol. Opt. Phys.* **2014**, *47*, 115005.

(54) Lehtola, S.; Head-Gordon, M.; Jónsson, H. Complex orbitals, multiple local minima, and symmetry breaking in perdew-zunger self-interaction corrected density functional theory calculations. *J. Chem. Theory Comput.* **2016**, *12*, 3195–3207.

(55) Borghi, G.; Park, C. H.; Nguyen, N. L.; Ferretti, A.; Marzari, N. Variational minimization of orbital-density-dependent functionals. *Phys. Rev. B* **2015**, *91*, 155112.

(56) Lehtola, S.; Jónsson, H. Variational, Self-Consistent Implementation of the Perdew-Zunger Self-Interaction Correction with Complex Optimal Orbitals. *J. Chem. Theory Comput.* **2014**, *10*, 5324–5337.

(57) Goedecker, S.; Umrigar, C. J. Critical assessment of the self-interaction-corrected–local-density-functional method and its algorithmic implementation. *Phys. Rev. A* **1997**, *55*, 1765.

(58) Levi, G.; Ivanov, A. V.; Jónsson, H. Variational Calculations of Excited States Via Direct Optimization of Orbitals in DFT. *Faraday Discuss.* **2020**, `https://doi.org/10.1039/D0FD00064G`.

(59) Nocedal, J.; Wright, S. *Numerical Optimization*; Springer, New York, 2006.

(60) Mewes, J. M.; Jovanović, V.; Marian, C. M.; Dreuw, A. On the molecular mechanism of non-radiative decay of nitrobenzene and the unforeseen challenges this simple molecule holds for electronic structure theory. *Phys. Chem. Chem. Phys.* **2014**, *16*, 12393–12406.

(61) Klüpfel, P.; Klüpfel, S.; Tsemekhman, K.; Jónsson, H. Optimization of Functionals of Orthonormal Functions in the Absence of Unitary Invariance. *Lect. Notes Comput. Sci.* **2012**, *7134*, 23.

(62) Messud, J.; Dinh, P. M.; Reinhard, P.-G.; Suraud, E. On the exact treatment of time-dependent self-interaction correction. *Ann. Phys.* **2009**, *324*, 955.

(63) Svane, A. Electronic structure of cerium in the self-interaction-corrected local-spin-density approximation. *Phys. Rev. B* **1995**, *53*, 4275–4286.

(64) Lehtola, S.; Jónsson, H. Correction to Variational, Self-Consistent Implementation of the Perdew–Zunger Self-Interaction Correction with Complex Optimal Orbitals. *J. Chem. Theory Comput.* **2015**, *11*, 5052–5053.

(65) Lehtola, S.; Blockhuys, F.; Van Alsenoy, C. An overview of self-consistent field calculations within finite basis sets. *Molecules* **2020**, *25*, 1–23.

(66) Olsen, R. A.; Kroes, G. J.; Henkelman, G.; Arnaldsson, A.; Jónsson, H. Comparison of methods for finding saddle points without knowledge of the final states. *J. Chem. Phys.* **2004**, *121*, 9776–9792.

(67) Bofill, J. M. Updated Hessian matrix and the restricted step method for locating transition structures. *J. Comput. Chem.* **1994**, *15*, 1–11.

(68) Culot, P.; Dive, G.; Nguyen, V. H.; Ghuysen, J. M. A quasi-Newton algorithm for first-order saddle-point location. *Theor. Chim. Acta* **1992**, *82*, 189–205.

(69) Baker, J. An algorithm for the location of transition states. *J. Comput. Chem.* **1986**, *7*, 385–395.

(70) Simons, J.; Jørgensen, P.; Taylor, H.; Ozment, J. Walking on potential energy surfaces. *J. Phys. Chem.* **1983**, *87*, 2745–2753.

(71) Cerjan, C. J.; Miller, W. H. On finding transition states. *J. Chem. Phys.* **1981**, *75*, 2800–2806.

(72) Enkovaara, J.; Rostgaard, C.; Mortensen, J. J.; Chen, J.; Dulak, M.; Ferrighi, L.; Gavnholt, J.; Glinsvad, C.; Haikola, V.; Hansen, H. A.; Kristoffersen, H. H.; Kuisma, M.; Larsen, A. H.; Lehtovaara, L.; Ljungberg, M.; Lopez-Acevedo, O.; Moses, P. G.; Ojanen, J.; Olsen, T.; Petzold, V.; Romero, N. A.; Stausholm-Møller, J.; Strange, M.;

Tritsaris, G. A.; Vanin, M.; Walter, M.; Hammer, B.; Häkkinen, H.; Madsen, G. K. H.; Nieminen, R. M.; Nørskov, J. K.; Puska, M.; Rantala, T. T.; Schiøtz, J.; Thygesen, K. S.; Jacobsen, K. W. Electronic structure calculations with GPAW: a real-space implementation of the projector augmented-wave method. *J. Phys.: Condens. Matter* **2010**, *22*, 253202.

(73) Larsen, A. H.; Vanin, M.; Mortensen, J. J.; Thygesen, K. S.; Jacobsen, K. W. Localized atomic basis set in the projector augmented wave method. *Phys. Rev. B* **2009**, *80*, 195112.

(74) Mortensen, J.; Hansen, L.; Jacobsen, K. W. Real-space grid implementation of the projector augmented wave method. *Phys. Rev. B* **2005**, *71*, 035109.

(75) Bofill, J. M.; Comajuan, M. Analysis of the updated Hessian matrices for locating transition structures. *J. Comput. Chem.* **1995**, *16*, 1326–1338.

(76) Anglada, J. M.; Besalú, E.; Bofill, J. M.; Rubio, J. Another way to implement the Powell formula for updating Hessian matrices related to transition structures. *J. Math. Chem.* **1999**, *25*, 85–92.

(77) Sun, W.; Yuan, Y.-X. *Optimization Theory and Methods*; Springer, Boston, 2006.

(78) Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **1994**, *50*, 17953.

(79) Pulay, P. Convergence acceleration of iterative sequences. the case of scf iteration. *Chem. Phys. Lett.* **1980**, *73*, 393–398.

(80) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865.

(81) Loos, P. F.; Scemama, A.; Blondel, A.; Garniron, Y.; Caffarel, M.; Jacquemin, D. A Mountaineering Strategy to Excited States: Highly Accurate Reference Energies and Benchmarks. *J. Chem. Theory Comput.* **2018**, *14*, 4360–4379.

(82) Schreiber, M.; Silva-Junior, M. R.; Sauer, S. P.; Thiel, W. Benchmarks for electronically excited states: CASPT2, CC2, CCSD, and CC3. *J. Chem. Phys.* **2008**, *128*, 134110.

(83) Woon, D. E.; Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. *J. Chem. Phys.* **1993**, *98*, 1358–1371.

(84) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **1992**, *96*, 6796–6806.

(85) Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

(86) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.

(87) Abedi, M.; Levi, G.; Zederkof, D. B.; Henriksen, N. E.; Pápai, M.; Møller, K. B. Excited-state solvation structure of transition metal complexes from molecular dynamics simulations and assessment of partial atomic charge methods. *Phys. Chem. Chem. Phys.* **2019**, *21*, 4082–4095.

(88) Hratchian, H. P.; Schlegel, H. B. In *Theory and Applications of Computational Chemistry*; Dykstra, C. E., Frenking, G., Kim, K. S., Scuseria, G. E., Eds.; Elsevier: Amsterdam, 2005; pp 195 – 249.

(89) Garza, A. J.; Scuseria, G. E. Comparison of self-consistent field convergence acceleration techniques. *J. Chem. Phys* **2012**, *137*, 54110.

(90) Helgaker, T.; Jørgensen, P.; Olsen, J. *Molecular Electronic-Structure Theory*; John Wiley & Sons, Ltd, 2014; Chapter 4, pp 107–141.

(91) Vaucher, A. C.; Reiher, M. Steering Orbital Optimization out of Local Minima and Saddle Points Toward Lower Energy. *J. Chem. Theory Comput.* **2017**, *13*, 1219–1228.

(92) Dohn, A. O.; Kjær, K. S.; Harlang, T. B.; Canton, S. E.; Nielsen, M. M.; Møller, K. B. Electron Transfer and Solvent-Mediated Electronic Localization in Molecular Photocatalysis. *Inorg. Chem.* **2016**, *55*, 10637–10644.

(93) Dohn, A. O.; Jónsson, E. O.; Kjær, K. S.; B. van Driel, T.; Nielsen, M. M.; Jacobsen, K. W.; Henriksen, N. E.; Møller, K. B. Direct Dynamics Studies of a Binuclear Metal Complex in Solution: The Interplay Between Vibrational Relaxation, Coherence, and Solvent Effects. *J. Phys. Chem. Lett.* **2014**, *5*, 2414–2418.

(94) Kramida, A.; Yu. Ralchenko,; Reader, J.; NIST ASD Team, NIST Atomic Spectra Database (ver. 5.7.1), [Online]. Available: `https://physics.nist.gov/asd` [2020, April 23]. National Institute of Standards and Technology, Gaithersburg, MD., 2019.

(95) Theophilou, A. K. The energy density functional formalism for excited states. *J. Phys. C: Solid State Phys.* **1979**, *12*, 5419–5430.

(96) Nygaard, C. R.; Olsen, J. A second-order unconstrained optimization method for canonical-ensemble density-functional methods. *J. Chem. Phys.* **2013**, *138*, 1–13.

(97) Al-Mohy, A.; Higham, N. A new scaling and squaring algorithm for the matrix exponential. *SIAM J. Matrix Anal. Appl.* **2009**, *31*, 970–989.

(98) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Jarrod Millman, K.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C.; Polat, İ.; Feng, Y.; Moore, E. W.; VanderPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; van Mulbregt, P.; SciPy 1. 0 Contrib-

utors, SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17*, 261–272.

Article III

Direct Optimization Method for Variational Excited-State Density Functional Calculations Using Real Space Grid or Plane Waves

Ivanov A.V., Levi G., Jónsson E. Ö, Jónsson H.

Submitted to Journal of Chemical Theory and Computation.

Direct Optimization Method for Variational Excited-State Density Functional Calculations Using Real Space Grid or Plane Waves

Aleksei V. Ivanov, Gianluca Levi, Elvar Ö. Jónsson, and Hannes Jónsson*

Science Institute and Faculty of Physical Sciences, University of Iceland, VR-III, 107 Reykjavík, Iceland

E-mail: hj@hi.is

Abstract

A direct optimization method is presented for density functional calculations of excited electronic states using either a real space grid or a plane wave basis set. The method is variational, provides atomic forces in the excited states, and can be applied to Kohn-Sham (KS) functionals as well as orbital-density dependent functionals (ODD) including explicit self-interaction correction. The implementation for KS functionals involves two nested loops: (1) An inner loop for finding a stationary point in a subspace spanned by the occupied and a few virtual orbitals corresponding to the excited state; (2) an outer loop for minimizing the energy in a tangential direction. For ODD functionals, a third loop is used to find the unitary transformation that minimizes the energy functional among occupied orbitals only. Combined with the maximum overlap method, the algorithm converges in challenging cases where conventional self-consistent field algorithms tend to fail. The benchmark tests presented include two charge-transfer excitations in nitrobenzene and an excitation of CO to degenerate π^* orbitals where the importance of complex orbitals is illustrated. An application of the method to several

metal-to-ligand charge-transfer and metal-centred excited states of an Fe^{II} photosensitizer complex is described and the results compared to reported experimental estimates. The method is also used to study the effect of Perdew-Zunger self-interaction correction on valence and Rydberg excited states of several molecules, both singlet and triplet states. The correction is found to improve the description of molecular bond stretching but calculated values of the excitation energy are improved only slightly, by *ca.* 0.1 eV, due to cancellation of the estimated self-interaction error in the ground and excited states.

1 Introduction

Density functional theory (DFT) is commonly used in computational studies of molecules and materials as it can in many cases give reasonable accuracy without too much computational effort. Calculations of ground electronic states can be performed even by non-experts thanks to well-established algorithms and software implementations. This does not, however, apply to calculations of excited electronic states, although such states are of great importance in many rapidly developing fields such as ultrafast spectroscopy, solar energy conversion and photocatalysis. Even though DFT is formulated theoretically as a ground state theory, it can also be used to provide useful estimates of excited states. The most commonly used excited-state extension of DFT is time-dependent density functional theory (TDDFT).[1–3] In practical implementations, TDDFT calculations are carried out using some ground-state density functional, linear-response theory and an adiabatic approximation that neglects the time dependence of the exchange-correlation (XC) kernel. Within these approximations, TDDFT typically provides a fairly good description of low-lying valence excitations,[3] but often fails to describe higher excitations[4–6] and conical intersections between the ground and excited states.[4]

Alternative approaches with wide applicability and similar computational effort can be based on time-independent DFT such as ensemble DFT,[7–9] excited-state DFT (eDFT)[10–17]

(also sometimes referred to as Δ self-consistent field, ΔSCF), the quasi-particle energy DFT (QE-DFT) approach[18], as well as constrained DFT,[19–21] orthogonality constrained,[22] and constricted DFT approaches.[23,24] There, an excited state is found from a stationary solution of self-consistent equations and the computational robustness strongly depends on the algorithm used. In eDFT, the excited state is a solution of the Kohn-Sham (KS) equations for non-aufbau orbital occupation numbers. Commonly used methods in ground-state calculations are based on some iterative eigensolver such as the Davidson algorithm[25] or the residual minimization method–direct inversion in the iterative subspace (RMM-DIIS)[26–28] enhanced in various ways to improve robustness and rate of convergence.[29,30] However, these algorithms are not specifically designed for calculations of excited states. The maximum overlap method (MOM)[13] can be used to reduce the probability of convergence on the ground state in the iterative calculation, but convergence problems often occur.[16] The basic problem lies in the fact that excited-state calculations do not involve finding the global minimum of the energy as a function of the electronic degrees of freedom, but rather a more general stationary point on the high-dimensional electronic energy surface. A method that can in general converge on a saddle point on the energy surface rather than a minimum is required. To find an N-th order saddle point, one needs to maximize the energy with respect to N degrees of freedom while minimizing with respect to all the others. The degrees of freedom along which the energy needs to be maximized are not known *a priori*. This makes a search for a saddle point significantly more difficult than a search for a minimum. Therefore, while calculations based on SCF-MOM-type algorithms can in principle converge on excited states, they are in practice not reliable.

Alternatively, direct optimization (DO) can be used to converge on solutions of the KS equations. While this approach has mainly been used in energy minimization to find a ground-state solution[31–39], it can be extended to calculations of excited states.[15,16,40] One possible approach is to formulate the problem as a search for a minimum of the norm of the gradient.[16] But, it is then important to also ensure that the norm of the gradient is

zero at the minimum and the evaluation of the gradient of this objective function increases the computational effort significantly. A different approach is based on direct optimization of the energy using a quasi-Newton method that can develop negative eigenvalues of the Hessian consistent with the type of saddle point searched for.[15,41] A preconditioner is then needed to estimate the degrees of freedom for which the energy needs to be maximized and thereby ensure convergence on the desired saddle point. When combined with MOM, this approach can perform better than conventional SCF-MOM algorithms.[15,41] The number of degrees of freedom for which the optimization needs to be carried out is, however, an important consideration and so far the DO-MOM approach has only been formulated and implemented in the context of the linear combination of atomic orbitals (LCAO) basis set where the number of degrees of freedom is relatively small.

Real space grid (RSG) and plane wave (PW) basis sets have the advantage that the complete basis set limit can be reached by varying systematically a single parameter such as the mesh spacing or plane wave cutoff. They are also more universal and can easily be applied to diffuse states such as Rydberg and metallic states where a typical LCAO basis set needs to be supplemented by specially tailored diffuse basis functions. This is demonstrated for a Rydberg state of NH_3 in section 3 of this article. In calculations relevant to, for example, ultrafast experiments, the system may evolve through a series of localized and delocalized states, making it challenging to design an LCAO basis set that is complete enough for all the relevant states. It is clearly more convenient to use an RSG or PW basis set in these cases.

In this article, a DO-MOM algorithm for calculating excited electronic states that can be used with both RSG and PW basis sets is presented. The calculations are variational and, since the Hellmann-Feynman theorem is satisfied not only at the minimum but also at any stationary point on the electronic energy surface, they can be used to evaluate the atomic forces, thereby providing a powerful tool for exploring excited-state potential energy surfaces (PESs) in, for example, simulations of the dynamics or minimum energy path calculations.[42]

The DO-MOM method is tested and its performances compared with that of SCF-MOM with either Davidson or RMM-DIIS methods combined with Pulay density mixing.[26] The tests include charge-transfer excitations in nitrobenzene that are known to be challenging cases for conventional algorithms. While the SCF-MOM algorithms show erratic behavior, likely because of the presence of several orbitals with similar energy, the DO-MOM calculation converges in a robust way. Another test involves calculations of an electronic excitation of the CO molecule to degenerate states. Again, SCF-MOM shows erratic behavior while the DO-MOM calculation converges smoothly. There, the advantage of using complex orbitals instead of real orbitals is, furthermore, demonstrated.

Two applications of the DO-MOM method are presented. The first one involves calculations of metal-to-ligand charge-transfer (MLCT) and metal-centered (MC) excitations of the $[Fe(bmip)_2]^{2+}$ (bmip=2,6-bis(3-methyl-imidazole-1-ylidine)-pyridine) complex, a prototype of a class of Fe-based photosensitizers.[43–45] This complex has been studied experimentally using X-ray emission and scattering with femtosecond resolution[46] and theoretically using TDDFT.[47,48] In order to optimize its performance as a photosensitizer, an understanding of the transitions between the various states and the way they are affected by the ligands is needed. An initial excitation to a singlet MLCT state is believed to be followed in part by a relaxation to a lower energy triplet MLCT state while another part decays to a triplet MC state where it generates a vibrational wavepacket along a metal-ligand bond stretching coordinate.[47,48] This branching occurs on ultrafast time scale, on the order of 100 fs, and influences the performance of the complex as photosensitizer. Dynamics of the molecule in the lowest-lying, dark singlet MLCT state has been simulated using energy surfaces calculated with TDDFT but a higher energy, bright singlet MLCT state is likely populated in the experiments.[47,48] As a result, direct comparison with the ultrafast branching observed in the experiment could not be made. Here, the DO-MOM method is used to calculate six excited states that are close in energy, including the bright singlet MLCT state along the metal-ligand bond stretching coordinate that is believed to be activated during the photoinduced

dynamics. The calculated excitation energy agrees well with the experimentally observed value and, furthermore, an estimate of the vibrational period in the lowest triplet MC state is also found to be in good agreement with experimental observations. This demonstrates that the DO-MOM method could, in future work, be used in dynamics simulations to help interpret the experiments on this and similar photosensitizer complexes.

A second application of the DO-MOM method, which illustrates its applicability to ODD functionals, is a study of the effect of Perdew-Zunger self-interaction correction (PZ-SIC)[49] on the estimated excitation energy in 13 transitions to singlet and triplet excited states in 9 molecules. The results are compared with theoretical best estimates as well as experimental values. Also, O-H bond stretching curves in the water molecule and water dimer are calculated. It is found that PZ-SIC improves the shape of the curves, producing a local minimum analogous to what has been found in high level wave function calculations while the uncorrected functional does not.

The article is organized as follows. In section 2, the DO-MOM algorithm for variational density functional calculations of excited states is presented. Section 3 shows the results of the numerical tests. In sections 4 and 5, the applications of the DO-MOM method to excited-state energy curves of the Fe^{II} complex and the effect of PZ-SIC on excited states of molecules are presented, respectively. Finally, a discussion and conclusions are presented in section 6.

2 Methodology

In generalized KS-DFT, the energy of an electronic system is given by

$$\mathcal{E}[\mathbf{\Psi}] = \mathcal{T}_s + \int d^3\mathbf{r}\rho(\mathbf{r})v_{ext}(\mathbf{r}) + \mathcal{U}[\rho] + \mathcal{E}_{xc}, \tag{1}$$

where $\mathcal{T}_s$ is the kinetic energy of a system of non-interacting electrons that have the same density $\rho(\mathbf{r})$ as the interacting electron system

$$\mathcal{T}_s = -\frac{1}{2}\sum_{i=1}^{M}\sum_{\sigma=0,1} f_{i\sigma}\left\langle \psi_{i\sigma}\right|\nabla^2\left|\psi_{i\sigma}\right\rangle \tag{2}$$

and

$$\sum_{i=1}^{M}\sum_{\sigma=0,1} f_{i\sigma}\left\langle \mathbf{r}|\psi_{i\sigma}\right\rangle\left\langle \psi_{i\sigma}|\mathbf{r}\right\rangle = \rho(\mathbf{r}). \tag{3}$$

with occupation numbers $\{f_{i\sigma}\}$. The occupation numbers can be chosen to be non-aufbau in order to represent an excited state. M is the number of bands (orbitals) in the calculations and σ is the spin index. $v_{ext}(\mathbf{r})$ is the external potential and $\mathcal{U}$ is the classical Coulomb energy

$$\mathcal{U}[\rho] = \frac{1}{2}\iint d^3\mathbf{r}\, d^3\mathbf{r}' \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}. \tag{4}$$

$\mathcal{E}_{xc}$ is the XC energy, approximated in practice as a semilocal functional of the density and its gradient, but can also include an explicit dependence on the orbitals (as in meta-generalized gradient functionals and hybrid functionals).

Excited states are obtained when the total energy is stationary with respect to the orbitals $\mathbf{\Psi} = \left(|\psi_1\rangle, \cdots, |\psi_M\rangle\right)^T$ with non-aufbau occupation numbers and can correspond to saddle points. The electronic energy surface has dimensionality MN_b, where N_b is the number of grid points in an RSG or number of PW coefficients. Even for a small molecule, the number of grid points can easily become large, on the order of 10^6. In order to facilitate the stationary-point search problem, the orbitals $\mathbf{\Psi}$ are expanded in a linear combination of some auxiliary orbitals $\mathbf{\Phi}$

$$\mathbf{\Psi} = U\mathbf{\Phi}, \tag{5}$$

where U is an $M \times M$ unitary matrix. The auxiliary orbitals, $\mathbf{\Phi}$, can be chosen to be the ground-state orbitals or any set of orbitals that represents an initial guess for the excited

state. The energy can then be considered as a functional of both U and $\mathbf{\Phi}$

$$\mathcal{E}[\mathbf{\Psi}] = \mathcal{E}[U\mathbf{\Phi}] = \mathcal{F}[U, \mathbf{\Phi}]. \tag{6}$$

Therefore, stationary points of $\mathcal{E}[\mathbf{\Psi}]$ can be found in two steps: first by *extremizing* $\mathcal{F}$ with respect to the expansion coefficients U and then by *minimizing* the functional

$$\mathcal{L}[\mathbf{\Phi}] = \operatorname*{stat}_{U} \mathcal{F}[U, \mathbf{\Phi}] \tag{7}$$

with respect to $\mathbf{\Phi}$:

$$\min_{\mathbf{\Phi}} \mathcal{L}[\mathbf{\Phi}] = \min_{\mathbf{\Phi}} \operatorname*{stat}_{U} \mathcal{F}[U, \mathbf{\Phi}] = \operatorname*{stat}_{\mathbf{\Psi}} \mathcal{E}[\mathbf{\Psi}]. \tag{8}$$

The introduction of the additional functional $\mathcal{L}[\mathbf{\Phi}]$ reduces the stationary-point search problem into two simpler tasks. First, instead of finding a saddle point in a wave function space, one finds the saddle-point in the space of unitary matrices U of low dimensionality. The reduction of dimensionality can further be achieved by decreasing the number of the virtual orbitals. For example, the first excited state of ammonia can be obtained by including only 4 virtual orbitals and, therefore, the dimensionality of the problem equals 24, within a frozen core approximation. This is a significant simplification as compared to the original problem of finding a saddle point in a MN_b dimensional space. Furthermore, an efficient algorithm based on a recently proposed[15] quasi-Newton method for finding saddle points in the space of unitary matrices can readily be used with minor modifications. The second simpler task is the outer loop minimization where conventional energy function minimization algorithms[50] generalized for a wave function optimization can be employed.

The division of the original optimization problem into separate minimizations of different degrees of freedom is a standard technique employed for ground-state calculations of metals,[31] self-interaction corrected density functional calculations[51–55] and ensemble density functional calculations.[56] The inner and outer loops are further described below. The optimization of

$\mathcal{E}$ is further performed in conjunction with the maximum overlap method (MOM)[13] used to distribute the occupation numbers $\{f_m\}$, similar to what has previously been done in the context of LCAO,[15,41] and it is described in Appendix B.

Additional considerations need to be addressed when a non-unitary invariant functional is used as in the case of PZ-SIC.[49] Semilocal approximations of the XC energy possess a spurious self-interaction error due to the inability of semilocal functionals to cancel the non-local self-interaction error in the classical Coulomb energy. Perdew and Zunger proposed an orbital-by-orbital correction[49]

$$\mathcal{E}^{\mathrm{SIC}}[\mathbf{\Psi}] = \mathcal{E}[\mathbf{\Psi}] - \sum_{i\sigma} \left(\mathcal{U}[\rho_{i\sigma}] + \mathcal{E}_{xc}[\rho_{i\sigma}, 0]\right) \tag{9}$$

making any approximation of the energy functional self-interaction free for one-electron systems. The functional $\mathcal{E}^{\mathrm{SIC}}$ is not invariant with respect to unitary transformations of the occupied orbitals and, therefore, an additional inner loop needs to be included in order to find the optimal orbitals in the occupied subspace minimizing the self-interaction corrected energy.[49] In this case, the functional defined on the occupied subspace becomes unitary invariant.[51,55,57] Thus, finding a solution that corresponds to the SIC excited state is achieved in a three-loop optimization:

$$\operatorname*{stat}_{\mathbf{\Psi}} \mathcal{E}^{\mathrm{SIC}}[\mathbf{\Psi}] = \min_{\mathbf{\Phi}} \operatorname*{stat}_{U} \min_{O}{}' \mathcal{E}^{\mathrm{SIC}}[UO\mathbf{\Phi}] = \min_{\mathbf{\Phi}} \operatorname*{stat}_{U} \mathcal{F}^{SIC}[U, \mathbf{\Phi}] \tag{10}$$

where the unitary minimization $\min'$ is performed among occupied orbitals only. More details on this additional inner loop are given in Appendix A.

2.1 Inner loop: Finding a stationary point of $\mathcal{F}[U, \mathbf{\Phi}]$ with respect to U.

To find a stationary point of $\mathcal{F}[U, \mathbf{\Phi}]$ for a given $\mathbf{\Phi}$, an exponential transformation is made, analogous to what has previously been done for energy minimization in wave-function based calculations[58–61] and density functional calculations,[36,53,62] and, more recently, for saddle-point searches .[15,41] The unitary matrix is parametrized as

$$U = e^A \tag{11}$$

where A is a skew-Hermitian matrix, $A^\dagger = -A$. The gradient of $\mathcal{F}$ with respect to the elements of A is

$$\frac{\partial \mathcal{F}}{\partial A^*_{ij}} = \frac{2 - \delta_{ij}}{2} \left[\int_0^1 e^{tA} L e^{-tA} dt \right]_{ij}, \tag{12}$$

where the elements of L are given by

$$L_{ij} = \overline{\langle \phi_j | \hat{h}_i | \phi_i \rangle} - \langle \phi_i | \hat{h}_j | \phi_j \rangle \tag{13}$$

with $\hat{h}_j$ defined as

$$\hat{h}_j \left| \psi_j \right\rangle = \frac{\delta \mathcal{E}}{\delta \left\langle \psi_j \right|} = f_j \left[-\frac{1}{2}\nabla^2 + \hat{v}_{ext} + \hat{v}_H + \hat{v}_{xc} \right] \left| \psi_j \right\rangle \tag{14}$$

During the optimization, the elements of A are found iteratively using a limited-memory version of the symmetric rank one quasi-Newton algorithm.[15] The initial inverse Hessian is preconditioned with a diagonal matrix with elements[61]

$$K_{ij} = \frac{1}{-2(\epsilon_i - \epsilon_j)(f_i - f_j)} \tag{15}$$

where the ϵ_i are the eigenvalues of the KS Hamiltonian. Since this preconditioner is valid only for the canonical representation of the Hamiltonian, the auxiliary orbitals are updated to the canonical orbitals every Xth iteration of the outer loop if the inner loop reaches a maximum number of iterations

$$\mathbf{\Phi} \leftarrow C\mathbf{\Psi} = CU\mathbf{\Phi}, \tag{16}$$

and set

$$U = I \tag{17}$$

where C is the unitary matrix that transforms the auxiliary orbitals to the canonical orbitals. Further implementation details of the inner loop can be found in Refs.[15,41]

2.2 Outer loop: Minimization of $\mathcal{L}[\mathbf{\Phi}]$.

Let $\mathcal{M}$ be a manifold in the Hilbert space such that

$$\mathcal{M} = \{\mathbf{\Phi} : \int d\mathbf{r}\, \phi_i^*(\mathbf{r})\phi_j(\mathbf{r}) = \delta_{ij}, i, j = 1 \dots M\}. \tag{18}$$

The tangent space to this manifold at $\mathbf{\Phi}$ is defined as

$$\boldsymbol{V}_{\mathbf{\Phi}}(\boldsymbol{G}) = \lim_{\varepsilon \to 0} \frac{\hat{R}[\mathbf{\Phi} + \varepsilon \boldsymbol{G}] - \mathbf{\Phi}}{\varepsilon} = \hat{R}'_{\varepsilon}[\mathbf{\Phi} + \varepsilon \boldsymbol{G}]\big|_{\varepsilon=0} \tag{19}$$

where $\hat{R}$ is the orthonormalization operator such that $\hat{R}[\mathbf{\Phi} + \varepsilon \boldsymbol{G}] \in \mathcal{M}$, and $\boldsymbol{G}$ is a vector in the Hilbert space. For example, $\hat{R}$ can be chosen as the Löwdin transformation. Let $S_{\boldsymbol{X},\boldsymbol{Y}}$ be the overlap matrix between two vectors $\boldsymbol{X}, \boldsymbol{Y}$ from the Hilbert space. Then

$$\begin{aligned} &S_{\mathbf{\Phi}+\varepsilon\boldsymbol{G},\mathbf{\Phi}+\varepsilon\boldsymbol{G}} = \int d\mathbf{r}\, (\mathbf{\Phi} + \varepsilon \boldsymbol{G})(\mathbf{\Phi} + \varepsilon \boldsymbol{G})^\dagger = \\ &I + \varepsilon\, (S_{\mathbf{\Phi},\boldsymbol{G}} + S_{\boldsymbol{G},\mathbf{\Phi}}) + \varepsilon^2 S_{\boldsymbol{G},\boldsymbol{G}} = I + \varepsilon\, W_{\mathbf{\Phi},\boldsymbol{G}} + \varepsilon^2 S_{\boldsymbol{G},\boldsymbol{G}} \end{aligned} \tag{20}$$

with $W_{\boldsymbol{\Phi},\boldsymbol{G}} = S_{\boldsymbol{\Phi},\boldsymbol{G}} + S_{\boldsymbol{G},\boldsymbol{\Phi}}$. For the Löwdin transformation

$$\hat{R}[\boldsymbol{\Psi} + \varepsilon\boldsymbol{G}] = S^{-1/2}_{\boldsymbol{\Psi}+\varepsilon\boldsymbol{G},\boldsymbol{\Psi}+\varepsilon\boldsymbol{G}}\,(\boldsymbol{\Psi} + \varepsilon\boldsymbol{G}) \tag{21}$$

and therefore, the tangent space at $\boldsymbol{\Phi}$ is

$$\boldsymbol{V}_{\boldsymbol{\Phi}}(\boldsymbol{G}) = \boldsymbol{G} - \frac{1}{2} W_{\boldsymbol{\Phi},\boldsymbol{G}}\,\boldsymbol{\Phi}, \tag{22}$$

obtained after substituting eqs (20) and (21) into eq (19) keeping only first order terms with respect to ε.

The gradient of $\mathcal{L}$ can be calculated using the chain rule:

$$\frac{\delta\mathcal{L}}{\delta\boldsymbol{\Phi}^*} = U\frac{\delta\mathcal{E}}{\delta\boldsymbol{\Psi}^*}, \tag{23}$$

where

$$\left(\frac{\delta\mathcal{E}}{\delta\boldsymbol{\Psi}^*}\right)_j = \frac{\delta\mathcal{E}}{\delta\,\langle\psi_j|} \tag{24}$$

After defining the tangent space in equation (19) and the gradient in equations (23) and (24), the minimization of $\mathcal{L}$ can be written as

Minimization algorithm.

- Set $k \leftarrow 0$, choose initial guess $\boldsymbol{\Phi}^{(k)}$; calculate gradient $\boldsymbol{G}^{(k)} = \delta\mathcal{L}/\delta\boldsymbol{\Phi}^{*(k)}$ using eqs (23) and (24).
- Project gradient on the tangent space $\boldsymbol{V}^{(k)} = \boldsymbol{V}(\boldsymbol{G}^{(k)})$ at $\boldsymbol{\Phi}^{(k)}$. ε and calculate residual error $\Delta^{(k)} = \|\boldsymbol{V}^{(k)}\|$.
- While $\Delta^{(k)} > \varepsilon$:
    1. Compute search direction $\boldsymbol{P}^{(k)}$ according to the chosen minimization algorithm and apply preconditioning (for example, for gradient descent, $\boldsymbol{P}^{(k)} = -\boldsymbol{G}^{(k)}$ and

inverse kinetic energy operator as preconditioner[28,63]).

2. Project the search direction on the tangent space $\boldsymbol{V}^{(k)} = \boldsymbol{V}(\boldsymbol{P}^{(k)})$ at $\boldsymbol{\Phi}^{(k)}$.
3. Choose optimal step length $\alpha^{(k)}$ along $\boldsymbol{V}^{(k)}$ and compute

$$\boldsymbol{\Phi}^{(k+1)} \leftarrow \boldsymbol{\Phi}^{(k)} + \alpha^{(k)}\boldsymbol{V}^{(k)} \tag{25}$$

4. Orthonormalize the wave functions, $\boldsymbol{\Phi}^{(k+1)} \leftarrow \hat{R}[\boldsymbol{\Phi}^{(k+1)}]$.
5. Compute new gradient $\boldsymbol{G}^{(k+1)}$ and project it on the tangent space $\boldsymbol{V}^{(k+1)} = \boldsymbol{V}(\boldsymbol{G}^{(k+1)})$ at $\boldsymbol{\Phi}^{(k+1)}$.
6. Calculate residual $\Delta^{(k+1)} = \|\boldsymbol{V}^{(k+1)}\|$.
7. $k \leftarrow k + 1$.

- End.

The search direction $\boldsymbol{P}^{(k)}$ can be chosen using a conjugate gradient[64] or a limited memory quasi-Newton algorithm. Here, the limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm as described in Ref.[65] is used. For minimization, this algorithm is known to give fast and robust convergence.

3 Implementation and Numerical Tests

The DO-MOM algorithm has been implemented in a development branch of the Grid-based projector augmented wave (GPAW) software[66] and can be used with either a finite-difference RSG[67] or PW basis set. The calculations are carried out using the frozen core approximation and the projector-augmented wave method.[68] The iterative SCF algorithms used here are based on either the Davidson algorithm[25] or the RMM-DIIS algorithm[28] as implemented in GPAW. Both versions of the SCF algorithm make use of Pulay density mixing[26] and MOM[13] (see Appendix B). Default values of the convergence parameters are used. The Pulay density

mixing uses densities from three previous steps and the coefficient used in the linear mixing of the density with the density residual vector is 0.15. No damping of short-wavelength density changes is used.[66] In the DO-MOM calculations, Pulay density mixing is not used. Instead, the density is calculated from the orbitals obtained at each iteration. The gradient of the energy projected on the tangent space in the outer loop in DO, or the residual vector in SCF, is preconditioned with the inverse kinetic energy operator.[28,63]

In the outer loop of the DO-MOM algorithm, the search direction is calculated according to the L-BFGS algorithm, using only the previous step to estimate the Hessian matrix. For a quasi-Newton algorithm a step length of 1 is a natural choice. However, the first step in the optimization corresponds to the gradient descent algorithm and the following maximum step length update is used in order to avoid too large changes in the orbitals: if the norm of the search direction

$$\Delta = \boldsymbol{P}^{\dagger}\boldsymbol{P} \tag{26}$$

is larger than α_{max} ($\Delta > \alpha_{max}$) then

$$\boldsymbol{P} \leftarrow \frac{\alpha_{max}}{\Delta}\boldsymbol{P}. \tag{27}$$

The value $\alpha_{max} = 0.25$ is found to give reliable convergence. For the inner loop optimization, the limited-memory symmetric rank one update[15] is used.

The calculations were carried out in the following way if not stated otherwise: The molecule is placed in a rectangular box with at least 7 Å vacuum space in all directions from the nuclei to the boundary of the box. Open boundary conditions are used. A grid mesh spacing of 0.2 Å is employed. All the calculations are spin-unrestricted and use the PBE functional.[69]

The advantage of RSG over LCAO is illustrated in a calculation of the $3p_z$ Rydberg state of NH_3 for excitation from the HOMO, see Fig. 1. An LCAO calculation using the aug-cc-pVDZ basis set is clearly not sufficient to reproduce the RSG calculation while the

expanded d-aug-cc-pVDZ basis set, which includes additional diffuse functions, gives close agreement.

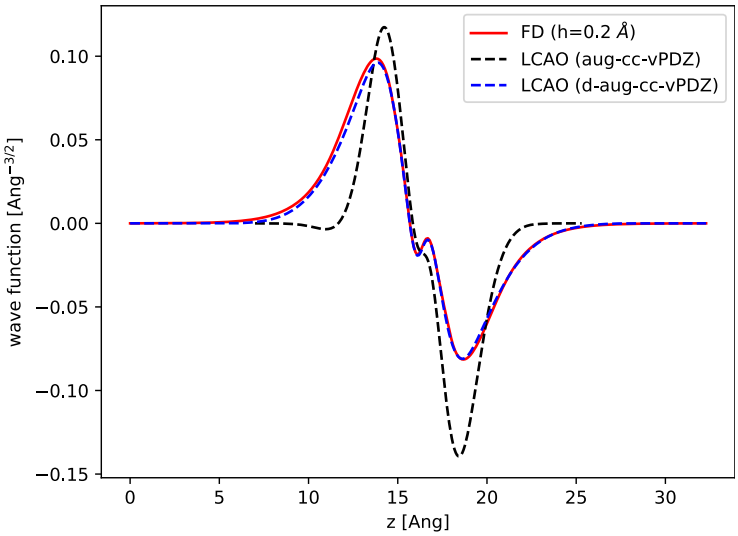

Figure 1: Comparison of three basis sets in a PBE calculation of the $3p_z$ excited state of NH_3 for excitation from the HOMO. The $3p_z$ orbital is shown along the rotational axis of the molecule (z-axis). The finite-difference real-space grid calculation with grid spacing of 0.2 Å (red) is closely reproduced by an LCAO calculation when the expanded d-aug-cc-pVDZ basis set including an extra set of diffuse functions is used (blue dashed), but the aug-cc-pVDZ basis set is clearly not sufficient (black dashed).

The energy of an open-shell singlet state is calculated using the spin purification:[70]

$$E^s = 2E(\uparrow\downarrow) - E(\uparrow\uparrow), \tag{28}$$

where $E(\uparrow\uparrow)$ is the energy of the triplet state and $E(\uparrow\downarrow)$ is the energy of the mixed spin state. Both states are calculated independently and variationally. The singlet excited-state energy will hereafter be referred to as the energy calculated according to eq (28). This has been found to give a better estimate of the singlet excited-state energy in eDFT calculations using semilocal KS functionals compared to the estimate obtained from the mixed spin state.[14]

3.1 Test I: Charge-transfer excitations in nitrobenzene

Charge-transfer excitations in nitrobenzene are known to be challenging cases for conventional algorithms[16,71] and are often used as benchmark tests.[15–17,41] The $^1A_1(\pi' \rightarrow \pi^*)$ excitation transfers an electron from the benzene ring to the nitro group while in the $^1A_1(n_\pi \rightarrow \pi'^*)$ excitation the transfer is in the opposite direction.

A ground-state calculation including 9 virtual orbitals was first performed to obtain the initial orbitals. The $^1A_1(\pi' \rightarrow \pi^*)$ excited state was then calculated by promoting an electron from the HOMO-2 to the LUMO orbital, while the $^1A_1(n_\pi \rightarrow \pi'^*)$ state was calculated by promoting an electron from the HOMO-4 to the LUMO+1 orbital.

An analysis of the performance of the DO-MOM calculation and comparison with the two versions of the SCF-based methods is presented in Fig. 2. Both the Davidson and RMM-DIIS implementations of SCF-MOM quickly approach the excitation energy of the target solution but then show erratic behaviour. This is attributed to the presence of several orbitals with energy close to that of the orbital from which excitation occurs.[15,41] The energy difference between HOMO-4 and HOMO-1 is only 0.56 eV and a change in the ordering of the orbitals occurs during the optimization of the excited state.[15,41] It is known that for orbitals that are energetically close, SCF algorithms have a difficulty converging unless smearing of occupation numbers is used or the parameters in the Pulay mixing are fine tuned. In contrast, the DO-MOM algorithm shows robust convergence. Tight convergence is obtained within 30 to 45 outer loop iterations as shown in Fig. 2(b) and (e). Initially, during each outer loop iteration, several inner loop iterations are performed as shown in Fig. 2(c) and (f). Towards the minimum of the energy functional $\mathcal{L}$, no inner-loop iterations are performed, only outer-loop iterations.

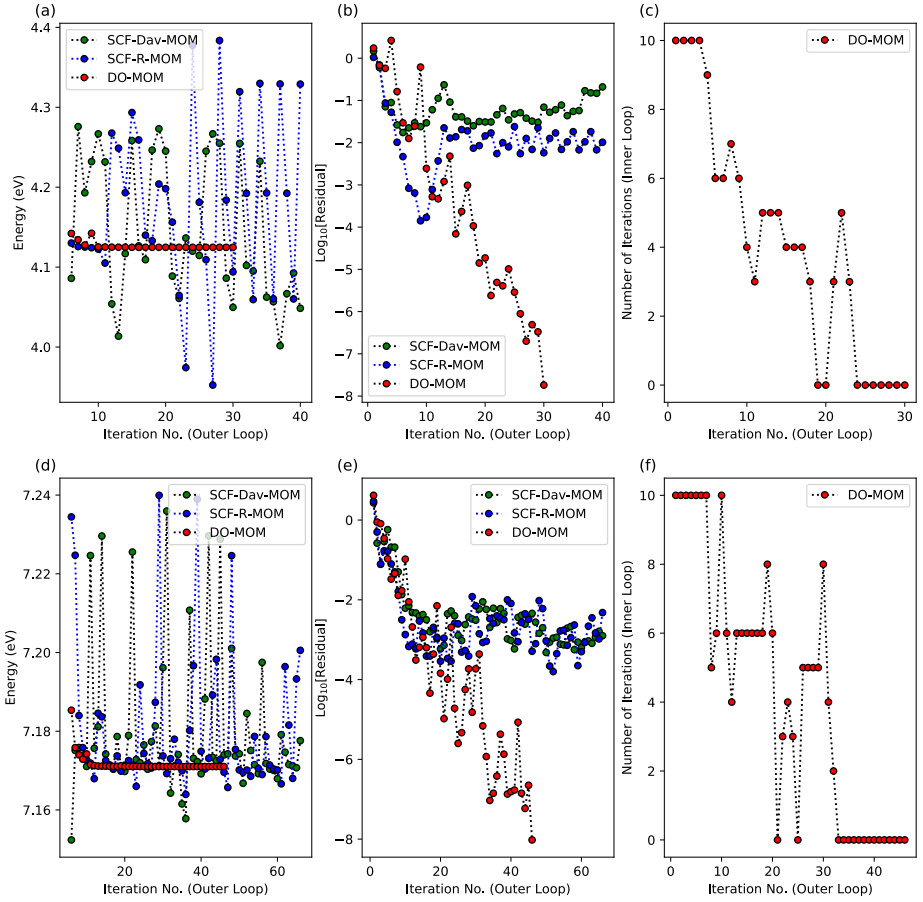

Figure 2: Comparison of the performance of various algorithms in calculations of the $^1A_1(\pi' \rightarrow \pi^*)$ excitation (a,b,c) and $^1A_1(n_\pi \rightarrow \pi'^*)$ excitation (d,e,f) of nitrobenzene. The SCF-MOM method, based either on Davidson (SCF-Dav-MOM) or RMM-DIIS (SFC-R-MOM) algorithms, does not converge well, while DO-MOM gives robust convergence. (a) and (d) show the convergence of the energy; (b) and (e) show the norm of the residual, and (c) and (f) show the number of inner-loop iterations for each outer-loop iteration in the DO-MOM calculation.

3.2 Test II: Excitation to degenerate orbitals

Degenerate electronic states need to be represented by multi-determinant wave functions. When a single determinant is used in a KS-DFT calculation two problems occur. The first is a technical problem, as the SCF algorithm has difficulty converging unless large enough smearing of the occupation numbers is used. The second is a conceptual problem, in that different single determinants which should in principle be degenerate can give different electron densities and, as a result, different total energy. This occurs, for example, in calculations of open-shell atoms.[72,73] In the case of degenerate excited states, an additional problem arises. In order to unambiguously assign an excited state, the wave function needs to have the symmetry of the excited state. With real-valued orbitals, which are most commonly employed in electronic structure calculations, this requirement is not necessarily satisfied due to symmetry breaking, as is demonstrated below.

Consider the lowest valence excited state in carbon monoxide. Using a single determinant, this excited state can be described by the promotion of an electron from a σ orbital (ground-state HOMO) to one of the two lowest degenerate π^* orbitals, (π_x^* or π_y^*, the ground-state LUMO, in the case of real wave functions). For such an excitation, the SCF method with the Davidson algorithm does not converge with integer occupation numbers. The energy oscillates around the excited-state solution, as shown in Fig. 3(a). The DO-MOM algorithm, however, gives smooth convergence to the excited-state solution.

Since the chosen orbitals are real, they are not eigenstates of the z-component of the angular momentum operator, and the angular momentum of the single-determinant wave function around the internuclear axis (z-axis) is not defined. In addition, the resulting electron density lacks uniaxial symmetry. It has instead an elliptic shape in the x-y plane with orientation depending on which orbital is occupied, π_x^* or π_y^*, see Fig. 3(b). This is inconsistent with the symmetry of the molecule. In the DO-MOM calculation, the orbitals can be chosen to be complex valued functions without any modifications of the algorithm. If the LUMO is chosen as a complex π_{+1}^* or π_{-1}^* orbital, where +1 or -1 is the eigenvalue

of the z-component angular momentum operator, the single-determinant excited-state wave function has a well-defined angular momentum and can unambiguously be identified as a Π state. The solution DO-MOM converges to using complex orbitals has 0.15 eV higher energy compared to the real valued solution, but it is more accurate since the total density then has rotational symmetry around the internuclear axis [see Fig. 3(b)]. Thus, the use of complex orbitals not only allows one to properly represent the total angular momentum of the excited state, but it also provides a density with the correct symmetry. The spin symmetry is still broken, however, in the unrestricted calculation.

The importance of using complex orbitals in order to provide correct description of the ground state has been emphasized in calculations of atoms and molecules using self-interaction corrected functionals[74–76] as well as within restricted Hartree-Fock theory[77] and KS formalism.[78] In particular, in the work of Lee *et. al*[78] it was shown that real orbitals break the cylyndrical symmetry of the density in the singlet ground state of O_2 while complex orbitals restore such symmetry within the restricted KS formalism. Here, it is shown that a similar situation occurs in the excited states of open-shell singlets and that the symmetry can be restored in the spin-unrestricted formalism using complex orbitals.

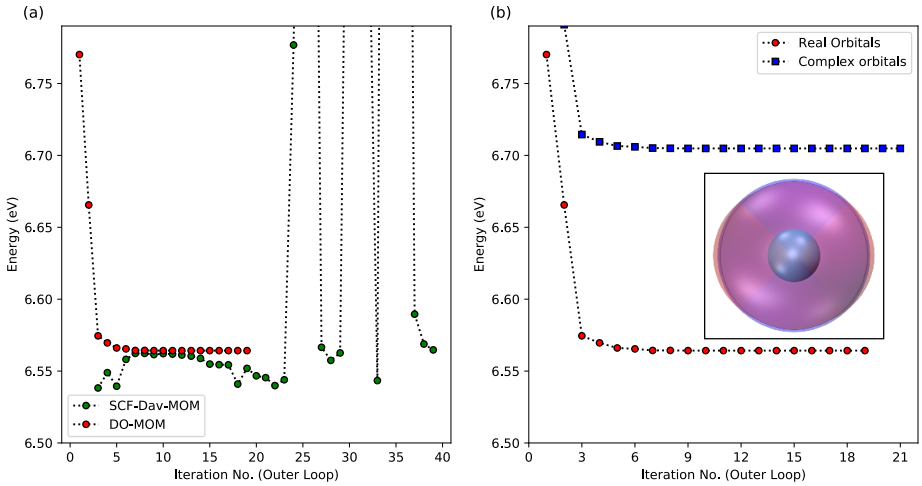

Figure 3: (a) Comparison of the performance of the SCF-MOM method implemented with Davidson and the DO-MOM method in calculations of the lowest n→ π^* excitation of a carbon monoxide molecule. (b) Comparison of the DO-MOM calculations with real and complex orbitals. The inset shows the total electron density obtained with real orbitals (red) and complex orbitals (blue). The internuclear axis is perpendicular to the plane of the image. The density obtained using real orbitals lacks the uniaxial symmetry.

4 Application I: Excited states of an Fe^{II} carbene photosensitizer

The first application of the DO-MOM method involves calculations of four MLCT and two MC excited states of the $[Fe(bmip)_2]^{2+}$ complex that consists of 63 atoms (see Fig. 4). The calculations are carried out with the BLYP functional.[79,80]

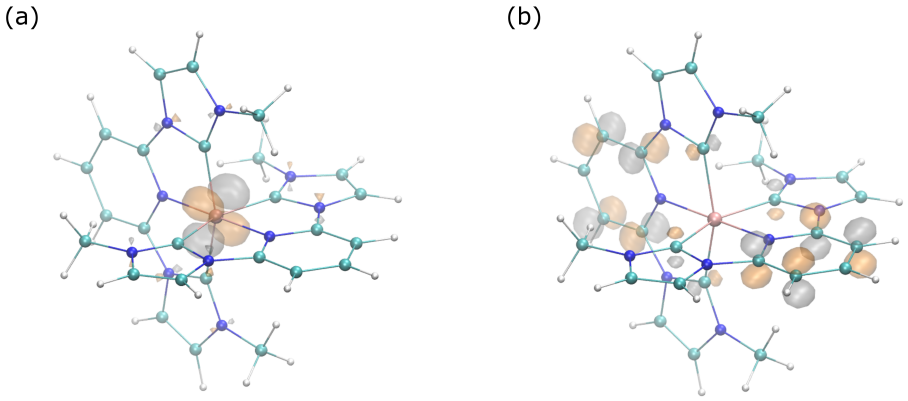

Figure 4: Atomic structure of the $[Fe(bmip)_2]^{2+}$ complex together with the orbitals involved in the transition to the bright MLCT state (C atoms green, N atoms blue, H atoms white, Fe atom orange). (a) HOMO orbital of the ground state localized on the Fe atom. (b) LUMO+2 orbital of the ground state delocalized over the ligands. Promotion of the electron from HOMO to LUMO+2 corresponds to a charge-transfer excitation (denoted as $^1MLCT_2$ state in the text). The isosurfaces are shown with grey and orange colors and correspond to isovalues of ± 0.16 Å$^{-3/2}$.

Fig. 5 shows the energy of the various states calculated along the metal-ligand bond stretching coordinate (the Q_6 breathing normal mode according to Ref.[48]) that is believed to account for the nuclear dynamics following photoexcitation.[46,47] The singlet state labelled $^1MLCT_2$ in Fig. 5, corresponding to a HOMO-to-LUMO transition (see Fig. 4), has an excitation energy of 2.58 eV, only 0.13 eV lower than the position of the maximum of the experimental UV/VIS absorption spectrum of the complex dissolved in acetonitrile.[43] Indeed, this state has the same character as the state with largest oscillator strength in TDDFT calculations;[48] thus, confirming that $^1MLCT_2$ corresponds to the bright MLCT state with a calculated excitation energy in good agreement with experiment. The triplet with same orbital occupancy (labelled $^3MLCT_2$ in Fig. 5), the lower-lying singlet ($^1MLCT_1$) and triplet MLCT ($^3MLCT_1$) states arising from the HOMO-to-LUMO excitation are also shown, as well as the lowest triplet MC state and the corresponding singlet with same character (arising from HOMO-1 to LUMO+4 excitation). The orbitals involved in the transitions to these

states are shown in the Supporting Information.

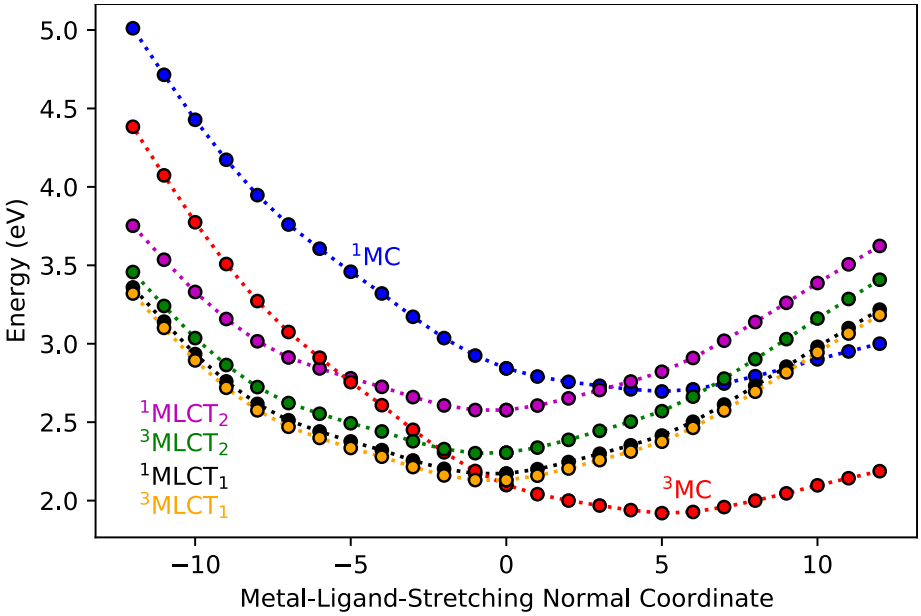

Figure 5: Potential energy curves along a metal-ligand normal coordinate (the breathing mode Q_6 as defined in Ref.[48]) that is expected to be responsible for the nuclear dynamics during the relaxation from the initially photoexcited ^{1}MLCT state to the ^{3}MC state. MLCT: metal-to-ligand charge-transfer excitation, MC: metal-centered excitation.

The combined ultrafast X-ray emission and scattering experiments have detected vibrational wavepacket dynamics along the metal-ligand stretching coordinate in a ^{3}MC state with a period of 278 fs.[46] The curvature of the energy curve for the ^{3}MC state calculated with DO-MOM is used to estimate the vibrational period, obtaining a value of 280 fs (see Supporting Information). The shape of the PES predicted by DO-MOM with BLYP, therefore, appears to agree well with the experiment in this respect, lending support for the use of DO-MOM in future dynamics simulations to study this and other photosensitizer complexes.

5 Application II: Assessment of Perdew-Zunger self-interaction correction

Calculations are carried out for 9 valence and 9 Rydberg excitations in 13 molecules involving both singlet and triplet excited states. The results of the calculations are compared to theoretical best estimates from Ref.,[81] which include corrections for basis set limitations and "all-electron" effects. The "all-electron" relaxation effects are estimated to be small, around $0.01 - 0.02$ eV,[81] in the present cases. On the other hand, the basis set correction can be significant, and is important for a consistent comparison with the finite-difference RSG results obtained in the present work. The molecules are placed in a rectangular box with at least 9 Å vacuum space in in all directions from any of the nuclei to the edge of the simulation box. This is found to be large enough to correctly describe even the diffuse Rydberg orbitals. The atomic coordinates of the molecules are those given in Ref.[81] The number of virtual orbitals is set to 8. The excitation energy is calculated with respect to the energy of the singlet ground state. The excited states are calculated using the lowest energy single Slater determinant for a given transition without enforcing point-group symmetry constraints on the total density. The SIC calculations are carried out using the PBE functional with full PZ-SIC (PBE-SIC) and with the PBE functional where the correction is scaled by a half (PBE-SIC/2) as such scaling has previously been found to provide better estimates of atomization energy of molecules and band gaps of solids.[75,82]

The DO-MOM calculated values of the excitation energy are given in Table 1 for the triplet states and in Table 2 for the singlet states. Since mixed spin states are often used in practice as an approximation to a singlet energy surface, such calculations are also performed and the results are given in Table S1 in the Supporting Information.

The calculations using the PBE functional give a mean error (ME) of -0.27 eV and a root mean square error (RMSE) of 0.31 eV with respect to the theoretical best estimates for the excitations to triplet states while a larger error is obtained for excitations to singlet states,

ME of -0.46 eV and RMSE of 0.54 eV. If the spin purification for singlet states is not applied, the error in the excitation energy is significantly larger, with the RMSE being 0.95 eV.

The calculations using self-interaction correction, i.e. the PBE-SIC functional, give slightly better values, the magnitude of the ME with respect to the theoretical best estimates being 0.2 eV smaller for the singlet state excitations and 0.1 eV smaller for the triplet state excitations, compared to PBE. The MAE is not improved as much. As has been shown previously in ground state calculations, it is important to use complex orbitals in SIC calculations.[74–76] This is also found to be the case here in the calculated values of the excitation energy. If real orbitals are used, the calculated values of the excitation energy become worse than those obtained with the PBE functional (the RMSE being 0.44 eV for triplet excitations, see Table S2 in Supporting Information).

For the triplet excitations, the scaled self-interaction correction, PBE-SIC/2 functional, gives smaller improvement while for the singlet excitations the MAE and RMSE with respect to the theoretical best estimates is a bit smaller than for full correction PBE-SIC. For mixed spin states, PBE-SIC performs better (see Table S1 in Supporting Information).

The small effect PZ-SIC has on the excitation energy is a result of cancellation of the estimated self-interaction energy in the ground and excited states. This orbital-by-orbital estimate of the self-interaction energy, which is most appropriate for a single electron system, turns out to be of similar magnitude for the ground and excited states. This is a surprising result since the classical self-Coulomb energy of a diffuse, Rydberg orbital is known to be smaller than that of a more localized ground-state orbital. Table 3 shows an analysis of this for the water molecule. A near cancellation of the total self-interaction energy still occurs because there is a simultaneous change in the self-XC term that offsets the difference in the classical self-Coulomb energy.

Table 1: Energy of excitations to triplet states calculated with the DO-MOM method and comparison with theoretical best estimates as well as experimental values. The calculations make use of a generalized gradient approximation Kohn-Sham functional (PBE), with scaled self-interaction correction (SIC/2) and with full self-interaction correction (SIC). The mean error (ME), mean absolute error (MAE) and root mean square error (RMSE) are given with respect to theoretical best estimates and with respect to experimental values at the bottom of the table.

molecule	excitation	PBE	SIC/2	SIC	TBE[a]	EXP[b]
acetaldehyde	$1^3A''$ ($n \rightarrow \pi^*$; V)	3.65	3.75	3.79	3.98	3.97
acetylene	$1^3\Delta_u$($\pi \rightarrow \pi^*$; V)	6.33	5.89	6.04	6.40	6.0
ammonia	2^3A_1(n→3s; R)	6.16	6.10	6.06	6.37	6.02
carbon monoxide	$1^3\Pi$(n→ π^*; V)	5.91	5.84	5.66	6.28	6.32
diazomethane	1^3A_2($\pi \rightarrow \pi^*$; V)	2.76	2.38	1.88	2.80	
ethylene	1^3B_{3u}(n→3s; R)	7.01	7.05	7.07	7.28	6.98
	1^3B_{1u}($\pi \rightarrow \pi^*$; V)	4.46	4.63	4.75	4.54	4.36
formaldehyde	1^3B_2(n→3s; R)	6.69	6.99	7.12	7.14	6.83
formamide	$1^3A"$(n→ π^*; V)	5.14	5.23	5.27	5.37	5.2
hydrogen sulfide	1^3A_2(n→4p; R)	5.39	5.44	5.43	5.74	5.8
ketene	1^3B_1(π →3s; R)	5.64	5.77	5.79	5.85	5.8
methanimine	$1^3A"$(n→ π^*; V)	4.20	4.35	4.41	4.64	
thioformaldehyde	1^3A_2(n→ π^*; V)	1.71	1.81	1.88	1.94	
	1^3B_2(n→4s; R)	5.31	5.54	5.67	5.76	
	2^3A_1($\pi \rightarrow \pi^*$; V)	3.36	3.33	3.28	3.44	3.28
water molecule	1^3B_1(n→3s; R)	7.10	7.09	7.08	7.33	7.2
	1^3A_2(n→3p; R)	8.75	8.87	8.97	9.30	8.9
	2^3A_1(n→3s; R)	9.28	9.25	9.23	9.59	9.46
	ME (TBE)	-0.27	-0.25	-0.18		
	ME (EXP)	-0.09	-0.06	-0.02		
	MAE (TBE)	0.27	0.26	0.21		
	MAE (EXP)	0.19	0.16	0.21		
	RMSE (TBE)	0.31	0.29	0.25		
	RMSE (EXP)	0.22	0.21	0.27		

[a]Theoretical best estimates as given in Ref.[81] [b]Experimental values listed in Ref.[81] (see references therein).

Table 2: Energy of excitations to singlet states (spin purified) calculated with the DO-MOM method and comparison with theoretical best estimates as well as experimental values. The calculations make use of a generalized gradient approximation Kohn-Sham functional (PBE), with scaled self-interaction correction (SIC/2) and with full self-interaction correction (SIC). The mean error (ME), mean absolute error (MAE) and root mean square error (RMSE) are given with respect to theoretical best estimates and with respect to experimental values at the bottom of the table.

molecule	excitation	PBE	SIC/2	SIC	TBE[a]	EXP[b]
acetaldehyde	$1^1A''$ ($n \rightarrow \pi^*$; V)	3.94	3.74	3.59	4.31	4.27
acetylene	$1^1\Delta_u$($\pi \rightarrow \pi^*$; V)	6.69	7.72	7.76	7.10	7.2
ammonia	2^1A_1(n→3s; R)	6.44	6.40	6.37	6.66	6.38
carbon monoxide	$1^1\Pi$(n→ π^*; V)	7.48	7.96	9.36	8.48	8.51
diazomethane	1^1A_2($\pi \rightarrow \pi^*$; V)	2.94	2.56	2.14	3.13	3.14
ethylene	1^1B_{3u}(n→3s; R)	7.14	7.18	7.2	7.43	7.11
	1^1B_{1u}($\pi \rightarrow \pi^*$; V)	6.72	7.17	7.64	7.92	7.6
formaldehyde	1^1B_2(n→3s; R)	6.89	7.10	7.30	7.11	
formamide	$1^1A''$(n→ π^*; V)	5.38	5.17	5.01	5.63	5.8
hydrogen sulfide	1^1A_2(n→4p; R)	5.63	5.63	5.55	6.10	
ketene	1^1B_1(π →3s; R)	5.87	5.97	6.11	6.06	5.86
methanimine	$1^1A''$(n→ π^*; V)	4.65	4.77	4.89	5.21	
thioformaldehyde	1^1A_2(n→ π^*; V)	1.91	1.74	1.57	2.20	2.03
	1^1B_2(n→4s; R)	5.64	5.72	5.77	5.99	5.85
	2^1A_1($\pi \rightarrow \pi^*$; V)	5.36	6.02	6.66	6.34	6.2
water molecule	1^1B_1(n→3s; R)	7.46	7.46	7.41	7.70	7.41
	1^1A_2(n→3p; R)	8.91	9.02	9.11	9.47	9.2
	2^1A_1(n→3s; R)	9.73	9.71	9.69	9.97	9.67
	ME (TBE)	-0.46	-0.33	-0.21		
	ME (EXP)	-0.30	-0.17	-0.06		
	MAE (TBE)	0.46	0.40	0.44		
	MAE (EXP)	0.33	0.27	0.36		
	RMSE (TBE)	0.54	0.43	0.51		
	RMSE (EXP)	0.46	0.35	0.49		

[a]Corrected theoretical best estimates as given in Ref.[81] [b]Experimental values listed in Ref.[81] (see references therein).

Table 3: One-electron self-interaction energy in eV calculated for the ground- and excited-state optimal orbitals of the water molecule. The excited-state $3s$ orbital has smaller classical self-Coulomb energy than the valence ground-state (GS) orbitals. However, the estimate of the total self-interaction energy is of similar magnitude for all the orbitals.

orbital	Coulomb	XC	Coulomb + XC
GS_1 [a]	10.06	-10.46	-0.40
GS_2 [a]	10.98	-11.23	-0.26
$3s$ orbital [b]	3.34	-3.70	-0.36

[a] Two orbitals from a ground-state calculation with PBE-SIC/2 (see Figure S5 in the Supplementary Information), giving different SIC estimates. [b] Orbital obtained from a DO-MOM excited-state calculation.

In order to further investigate the effect of PZ-SIC on excited states, the change in energy as an O-H bond is stretched in a water molecule and a dimer of water molecules is calculated for the ground and lowest singlet excited states. The spin-purification correction is not applied in this case. First, the ground-state geometry is optimized until the maximum of the force on the atoms has magnitude below 0.01 eV/Å. Then, the hydrogen-bonded O-H bond is stretched by changing the position of the hydrogen atom in increments of 0.1 Å, while keeping the positions of all other atoms frozen. The results of these calculations with both PBE and PBE-SIC/2 are presented in Fig. 6 and compared with the results of CR-EOMCCSD(T),ID/aug-cc-pVTZ calculations from Ref.[83] The energy curves obtained with PBE and PBE-SIC/2 lie close to each other. However, the PBE-SIC/2 energy curve reproduces better the S-shape of the high level reference curves for both the monomer and the dimer. For the water dimer in the lowest excited state, PBE-SIC/2 reproduces the local minimum at short bond length while PBE predicts a barrierless path towards the second constrained energy minimum near 1.8 Å O-H distance.

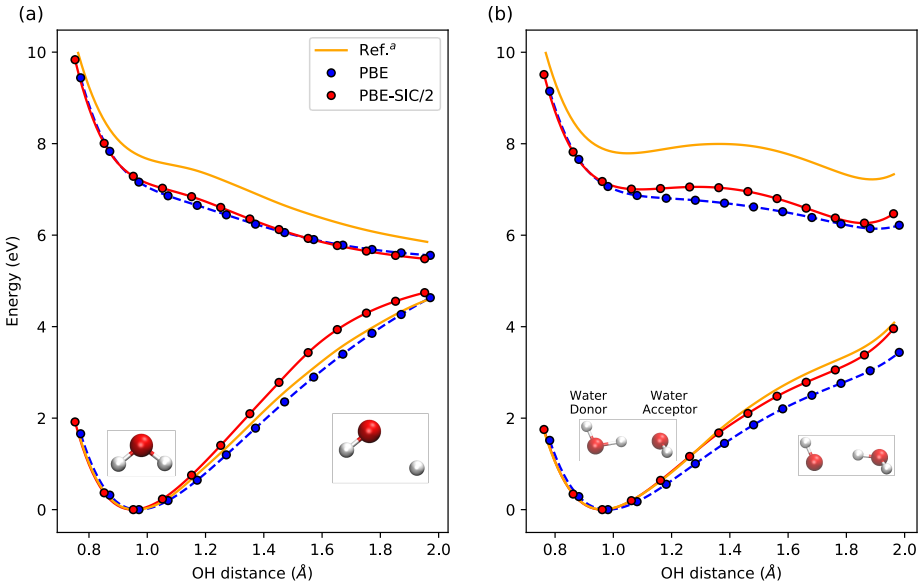

Figure 6: Potential energy curves calculated for ground and lowest excited singlet states using mixed-spin determinant along the hydrogen-bonded O-H bond with all other molecular geometry parameters fixed. (a) Water monomer, (b) Water dimer. The insets show the atomic configurations at the energy minimum and the endpoint of the scan in the O-H bond length. [a]CR-EOMCCD(T),ID/aug-cc-pVTZ calculations from Ref.[83]

6 Discussion and Conclusions

The RSG or PW representations have an advantage over LCAO in that full basis set limit can be reached by systematically changing only a single parameter. However, such calculations involve larger computational cost than LCAO and are limited to the frozen core approximation. A useful strategy for reducing the computational cost is to obtain initial orbitals from an LCAO calculation of the excited state and then switch to RSG or PW mode. All three types of representations can be available in the same software, as is the case with our implementation in the GPAW software, making such a hybrid approach relatively straightforward.

The search of saddle points on multidimensional electronic energy surfaces is a considerably more difficult problem than minimization due to the fact that the energy must be maximized along a few degrees of freedom that are not known *a priori*. Even if the order of a saddle point is known, the success of the extremization of the energy depends on how close the initial guess for the orbitals is to the target solution. The ground-state orbitals can represent a good enough initial guess for the iterative convergence on an excited state, but in some cases another excited-state solution provides a better choice. This can be especially important when the lowest energy excited state does not correspond to a HOMO-LUMO excitation. For example, in the water dimer the lowest excited state corresponds to an electron-hole pair localized on the hydrogen bond donor molecule but the HOMO-LUMO excitation corresponds to the transfer of an electron from the donor to the acceptor molecule with the hole localized on the donor and the excited electron localized on the acceptor. In the calculations of the potential energy curve for the lowest excited state in the water dimer presented in Sec. 5, a two-step procedure was used: First, the HOMO-LUMO excitation was computed and the obtained solution then used as an initial guess in the calculation on the lowest excited state.

In conclusion, we have developed a formulation of the DO-MOM algorithm for density functional calculations of excited states that can be applied to RSG and PW representations, where it is not possible to include all virtual orbitals, unlike with LCAO basis sets. This implementation of the DO-MOM algorithm is found to be robust and able to converge on excited states that are challenging for commonly used SCF-MOM algorithms. The importance of complex-valued orbitals in calculations of excited states with a KS-DFT functional is also demonstrated. In the case of the lowest excited state of CO, it was shown that real-valued orbitals break the uniaxial symmetry of the electron density, while complex-valued orbitals restore the symmetry and correspond to higher energy solutions. The fact that complex orbitals provide a better description of the electronic system is in par with findings from ground-state calculations.[74–78]

The application of DO-MOM in calculations of the excited states of the $[Fe(bmip)_2]^{2+}$ complex shows that: (1) the approach is robust enough to allow the calculation of several excited states that are close in energy, including regions where the potential energy curves cross; and (2) the predicted excited-state properties are in good agreement with experimental values. As for any single-determinant method, the quality of the excited states obtained with DO-MOM is expected to degrade when the states have a multi-determinantal character and this can represent a limitation for the application of the methodology to dynamics simulations. However, when explicit solvent effects are included in the simulations, the symmetry breaking induced by the solvent can lift degeneracy, thereby reducing the presence of multi-determinantal states. So far, excited-state dynamics simulations of systems as large as transition metal complexes using a variational eDFT approach have been limited to a single Born-Oppenheimer surface.[84–86] The present results represent a preliminary indication that the DO-MOM method is viable for nonadiabatic molecular dynamics simulations including multiple excited states. Current efforts in this direction include evaluating nonadiabatic couplings between the excited states obtained variationally, using numerical or analytical approaches such as the one recently presented in Ref.[87] A simulation of the excited-state relaxation of the $[Fe(bmip)_2]^{2+}$ complex starting from an excitation to the bright MLCT state could help explain the experimentally observed ultrafast ^{3}MLCT/^{3}MC branching that affects its performance as a photosensitizer.

An assessment of self-interaction correction, the PZ-SIC, applied to the PBE functional has been performed for excited states. For valence and Rydberg excitations, the correction improves the value obtained for the excitation energy, but only by *ca.* 0.1 eV. This small effect is due to a cancellation of the correction to the energy of the ground and excited states and may be attributed to the one-electron nature of this orbital-by-orbital estimate of the correction. Both the PBE value and the corrected value typically underestimate the excitation energy. Even though the effect of SIC on the excitation energy is small, it can lead to improved shape of the potential energy surface as is seen for the lowest excited state

of the water monomer and dimer. There, the dependence of the excited-state energy on the O-H bond length agrees better with results of high-level, CR-EOMCCD(T),ID/aug-cc-pVTZ calculations[83] better than results obtained with the PBE functional. Further improvement in excited-state energy calculations may require going beyond the self-interaction correction based on orbital densities, for example by including the complementary density in the error estimate.[88]

An important advantage of the RSG and PW implementation presented here is the possibility of doing calculations for extended systems that are subject to periodic boundary conditions. This includes, for example, excited states of defects in crystals and solvent effects on electronic excitations of molecular complexes. This will be the topic of future studies with the DO-MOM method.

Acknowledgement

This work was supported by the University of Iceland Research Fund and the Icelandic Research Fund. AVI is supported by a doctoral fellowship from the University of Iceland. GL thanks the Icelandic Research Fund for financial support under grant number 196070. The authors are grateful to Mátyás Pápai for discussions on the Fe photosensitizer complex. The calculations were carried out at the Icelandic Research High Performance Computing (IRHPC) facility at the University of Iceland.

Appendix

A Inner-Loop Minimizaion for ODD functional

The unitary matrix that minimizes an ODD functional, such as the PZ-SIC among occupied orbitals is found using an exponential transformation:

$$|\psi_i\rangle = \sum_{ij} (e^{-A})_{ij} \, |\phi_j\rangle \tag{A.1}$$

where A is a skew-hermitian matrix, $A^\dagger = -A$. Let

$$\lambda_{ij} = \left\langle \phi_i \left| \, \hat{h}_j + \hat{v}_j \, \right| \phi_j \right\rangle \tag{A.2}$$

with $\hat{h}_j$ defined in Eq. (14) and

$$\hat{v}_j = -\frac{\delta \left(\mathcal{U}[\rho_j] + \mathcal{E}_{xc}[\rho_j, 0] \right)}{\delta \rho_j} \tag{A.3}$$

The gradient of the energy with respect to matrix elements of A is:

$$\frac{\partial \mathcal{E}^{ODD}}{\partial A_{ij}} = \lambda_{ij} - \lambda_{ji}^* + o(\|A\|) \tag{A.4}$$

This expression must be zero at the minimum and for equally occupied orbitals this leads to the Pederson (or localization) conditions:[89]

$$\langle \phi_i \, | \, \hat{v}_j - \hat{v}_i \, | \, \phi_j \rangle = 0 \tag{A.5}$$

The L-BFGS algorithm with inexact line search is used to find the optimal matrix A and, thereby, the optimal unitary matrix $O = e^{-A}$. Details of the implementation of the exponential transformation are given in Ref.[62]

B Maximum Overlap Method within the Projector-Augmented Wave Approach

Let $\{|\psi_n\rangle\}$ be the N orbitals used as initial guess for an excited-state calculation and $\{|\psi_m^{(k)}\rangle\}$ the M orbitals at the k^{th} iteration of the optimization. Within the PAW approach, the elements of the overlap matrix between orbitals $\{|\psi_n\rangle\}$ and $\{|\psi_m^{(k)}\rangle\}$ are:[68]

$$S_{nm}^{(k)} = \langle \psi_n \,|\, \psi_m^{(k)} \rangle = \left\langle \tilde{\psi}_n \,\middle|\, \tilde{\psi}_m^{(k)} \right\rangle + \sum_{a,i_1,i_2} \left\langle \tilde{\phi}_n \,\middle|\, \tilde{p}_{i_1}^a \right\rangle \left(\langle \phi_{i_1}^a \,|\, \phi_{i_2}^a \rangle - \left\langle \tilde{\phi}_{i_1}^a \,\middle|\, \tilde{\phi}_{i_2}^a \right\rangle \right) \left\langle \tilde{p}_{i_2}^a \,\middle|\, \tilde{\psi}_m^{(k)} \right\rangle \tag{B.1}$$

where $|\tilde{\psi}_n\rangle$ are pseudo orbitals, $|\tilde{p}_{i_1}^a\rangle$ ($|\tilde{p}_{i_2}^a\rangle$) are projector functions localized on atom a, while $|\phi_{i_1}^a\rangle$ ($|\phi_{i_2}^a\rangle$) and $|\tilde{\phi}_{i_1}^a\rangle$ ($|\tilde{\phi}_{i_2}^a\rangle$) are partial and pseudo partial waves localized on atom a, respectively. After computing the overlap matrix at iteration k, the occupation numbers $\{f_m\}$ of the orbitals $\{|\psi_m^{(k)}\rangle\}$ are chosen as following. An occupation number of 1 is given to the first N orbitals with the highest numerical weights, evaluated from the projection onto the manifold $\{|\psi_n\rangle\}$:

$$\omega_m^{(k)} = \left(\sum_{n=1}^{N} |S_{nm}^{(k)}|^2 \right)^{1/2} \tag{B.2}$$

The remaining $M - N$ orbitals are left unoccupied (occupation number of 0).

References

(1) Runge, E.; Gross, E. K. U. Density-functional theory for time-dependent systems. *Physical Review Letters* **1984**, *52*, 997–1000.

(2) Casida, M. E. In *Recent Advances in Density Functional Methods*; Chong, D. P., Ed.; World Scientific, 1995; pp 155–192.

(3) Dreuw, A.; Head-Gordon, M. Single-Reference ab Initio Methods for the Calculation of Excited States of Large Molecules. *Chem. Rev.* **2005**, *105*, 4009–4037.

(4) Levine, B. G.; Ko, C.; Quenneville, J.; Martínez, T. J. Conical intersections and double excitations in time-dependent density functional theory. *Mol. Phys.* **2006**, *104*, 1039–1051.

(5) Maitra, N. T.; Zhang, F.; Cave, R. J.; Burke, K. Double excitations within time-dependent density functional theory linear response. *J. Chem. Phys.* **2004**, *120*, 5932–5937.

(6) Tozer, D. J.; Handy, N. C. On the determination of excitation energies using density functional theory. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2117–2121.

(7) Oliveira, L. N.; Gross, E. K. U.; Kohn, W. Ensemble-Density functional theory for excited states. *Int. J. Quantum Chem.* **1990**, *38*, 707–716.

(8) Yang, Z.-h.; Pribram-Jones, A.; Burke, K.; Ullrich, C. A. Direct Extraction of Excitation Energies from Ensemble Density-Functional Theory. *Phys. Rev. Lett.* **2017**, *119*, 33003.

(9) Deur, K.; Mazouin, L.; Fromager, E. Exact ensemble density functional theory for excited states in a model system: Investigating the weight dependence of the correlation energy. *Phys. Rev. B* **2017**, *95*, 35120.

(10) Hellman, A.; Razaznejad, B.; Lundqvist, B. I. Potential-energy surfaces for excited states in extended systems. *J. Chem. Phys.* **2004**, *120*, 4593–4602.

(11) Gavnholt, J.; Olsen, T.; Engelund, M.; Schiøtz, J. Δ self-consistent field method to obtain potential energy surfaces of excited molecules on surfaces. *Phys. Rev. B* **2008**, *78*, 75441.

(12) Cheng, C.-L.; Wu, Q.; Van Voorhis, T. Rydberg energies using excited state density functional theory. *J. Chem. Phys.* **2008**, *129*, 124112.

(13) Gilbert, A. T.; Besley, N. A.; Gill, P. M. Self-consistent field calculations of excited states using the maximum overlap method (MOM). *J. Phys. Chem. A* **2008**, *112*, 13164–13171.

(14) Kowalczyk, T.; Yost, S. R.; Voorhis, T. V. Assessment of the ΔSCF density functional theory approach for electronic excitations in organic dyes. *J. Chem. Phys.* **2011**, *134*.

(15) Levi, G.; Ivanov, A. V.; Jónsson, H. Variational Density Functional Calculations of Excited States via Direct Optimization. *Journal of Chemical Theory and Computation* **2020**, *16*, 6968–6982.

(16) Hait, D.; Head-Gordon, M. Excited state orbital optimization via minimizing the square of the gradient: General approach and application to singly and doubly excited states via density functional theory. *J. Chem. Theory Comput.* **2020**, *16*, 1699–1710.

(17) Carter-Fenk, K.; Herbert, J. M. State-Targeted Energy Projection: A Simple and Robust Approach to Orbital Relaxation of Non-Aufbau Self-Consistent Field Solutions. *J. Chem. Theory Comput.* **2020**, *16*, 5067–5082.

(18) Mei, Y.; Yang, W. Excited-State Potential Energy Surfaces, Conical Intersections, and Analytical Gradients from Ground-State Density Functional Theory. *J. Phys. Chem. Lett.* **2019**, *10*, 2538–2545.

(19) Ramos, P.; Pavanello, M. Low-lying excited states by constrained DFT. *J. Chem. Phys.* **2018**, *148*, 144103.

(20) Roychoudhury, S.; Sanvito, S.; O'Regan, D. D. Neutral excitation density-functional theory: an efficient and variational first-principles method for simulating neutral excitations in molecules. *Sci. Rep.* **2020**, *10*, 8947.

(21) Karpinski, N.; Ramos, P.; Pavanello, M. Capturing multireference excited states by constrained-density-functional theory. *Phys. Rev. A* **2020**, *101*, 32510.

(22) Evangelista, F. A.; Shushkov, P.; Tully, J. C. Orthogonality Constrained Density Functional Theory for Electronic Excited States. *J. Phys. Chem. A* **2013**, *117*, 7378–7392.

(23) Ziegler, T.; Krykunov, M.; Cullen, J. The implementation of a self-consistent constricted variational density functional theory for the description of excited states. *J. Chem. Phys.* **2012**, *136*, 124107.

(24) Park, Y. C.; Senn, F.; Krykunov, M.; Ziegler, T. Self-Consistent Constricted Variational Theory RSCF-CV(inf)-DFT and Its Restrictions to Obtain a Numerically Stable ΔsCF-DFT-like Method: Theory and Calculations for Triplet States. *J. Chem. Theory Comput.* **2016**, *12*, 5438–5452.

(25) Davidson, E. The iterative calculation of a few of the lowest eigenvalues and corresponding eigenvectors of large real-symmetric matrices. *J. Comput. Phys.* **1975**, *17*, 87.

(26) Pulay, P. Convergence acceleration of iterative sequences. the case of scf iteration. *Chem. Phys. Lett.* **1980**, *73*, 393–398.

(27) Pulay, P. Improved SCF convergence acceleration. *J. Comput. Chem.* **1982**, *3*, 556–560.

(28) Kresse, G.; Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **1996**, *54*, 11169–11186.

(29) Furthmüller, J.; Kresse, G. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. **1996**, *6*, 15–50.

(30) Garza, A. J.; Scuseria, G. E. Comparison of self-consistent field convergence acceleration techniques. *J. Chem. Phys.* **2012**, *137*.

(31) Gillan, M. J. Calculation of the vacancy formation energy in aluminium. *J. Phys. Condens. Matter* **1989**, *1*, 689–711.

(32) Payne, M.; Teter, M.; Allan, D.; Arias, T.; Joannopoulos, J. Iterative minimization techniques for ab initio total-energy calculations: Molecular dynamics and conjugate gradients. *Reviews of Modern Physics* **1992**, *64*, 1045–1097.

(33) Hutter, J.; Parrinello, M.; Vogel, S. Exponential transformation of molecular orbitals. *The Journal of Chemical Physics* **1994**, *101*, 3862–3865.

(34) Marzari, N.; Vanderbilt, D.; Payne, M. C. Ensemble Density-Functional Theory for Ab Initio Molecular Dynamics of Metals and Finite-Temperature Insulators. *Phys. Rev. Lett.* **1997**, *79*, 1337–1340.

(35) Ismail-Beigi, S.; Arias, T. A. New algebraic formulation of density functional calculation. *Comput. Phys. Commun.* **2000**, *128*, 1–45.

(36) Van Voorhis, T.; Head-Gordon, M. A geometric approach to direct minimization. *Molecular Physics* **2002**, *100*, 1713–1721.

(37) VandeVondele, J.; Hutter, J. An efficient orbital transformation method for electronic structure calculations. *Journal of Chemical Physics* **2003**, *118*, 4365–4369.

(38) Weber, V.; Vandevondele, J.; Hutter, J.; Niklasson, A. Direct energy functional minimization under orthogonality constraints. *Journal of Chemical Physics* **2008**, *128*.

(39) Freysoldt, C.; Boeck, S.; Neugebauer, J. Direct minimization technique for metals in density functional theory. *Phys. Rev. B* **2009**, *79*, 241103.

(40) Ye, H.-Z.; Welborn, M.; Ricke, N. D.; Van Voorhis, T. σ-SCF: A direct energy-targeting method to mean-field excited states. *J. Chem. Phys.* **2017**, *147*, 214104.

(41) Levi, G.; Ivanov, A. V.; Jónsson, H. Variational calculations of excited states via direct optimization of the orbitals in DFT. *Faraday Discuss.* **2020**, *224*, 448–466.

(42) Ásgeirsson, V.; Jónsson, H. In *Handbook of Materials Modeling: Methods: Theory and Modeling*; Andreoni, W., Yip, S., Eds.; Springer International Publishing: Cham, 2020; pp 689–714.

(43) Liu, Y.; Harlang, T.; Canton, S. E.; Chábera, P.; Suárez-Alcántara, K.; Fleckhaus, A.; Vithanage, D. A.; Göransson, E.; Corani, A.; Lomoth, R.; Sundström, V.; Wärnmark, K. Towards longer-lived metal-to-ligand charge transfer states of iron(ii) complexes: an N-heterocyclic carbene approach. *Chem. Commun.* **2013**, *49*, 6412.

(44) Harlang, T. C.; Liu, Y.; Gordivska, O.; Fredin, L. A.; Ponseca, C. S.; Huang, P.; Chábera, P.; Kjaer, K. S.; Mateos, H.; Uhlig, J.; Lomoth, R.; Wallenberg, R.; Styring, S.; Persson, P.; Sundström, V.; Wärnmark, K. Iron sensitizer converts light to electrons with 92% yield. *Nature Chemistry* **2015**, *7*, 883–889.

(45) Lindh, L.; Chábera, P.; Rosemann, N. W.; Uhlig, J.; Wärnmark, K.; Yartsev, A.; Sundström, V.; Persson, P. Photophysics and Photochemistry of Iron Carbene Complexes for Solar Energy Conversion and Photocatalysis. *Catalysts* **2020**, *10*, 315.

(46) Kunnus, K.; Vacher, M.; Harlang, T. C. B.; Kjær, K. S.; Haldrup, K.; Biasin, E.; van Driel, T. B.; Pápa, M.; Chabera, P.; Liu, Y.; Tatsuno, H.; Timm, C.; Källman, E.; Delcey, M.; Hartsock, R. W.; Reinhard, M. E.; Koroidov, S.; Laursen, M. G.; Hansen, F. B.; Vester, P.; Christensen, M.; Sandberg, L.; Németh, Z.; Szemes, D. S.; Bajnóczi, É.; Alonso-Mori, R.; Glownia, J. M.; Nelson, S.; Sikorski, M.; Sokaras, D.; Lemke, H. T.; Canton, S.; Møller, K. B.; Nielsen, M. M.; Vankó, G.; Wärnmark, K.; Sundström, V.; Persson, P.; Lundberg, M.; Uhlig, J.; Gaffney, K. J. Origin of vibrational wavepacket dynamics in Fe carbene photosensitizer determined with femtosecond X-ray emission and scattering. *Nature Communications* **2020**, *11*, 1–11.

(47) Pápai, M.; Vankó, G.; Rozgonyi, T.; Penfold, T. J. High-Efficiency Iron Photosensitizer

Explained with Quantum Wavepacket Dynamics. *J. Phys. Chem. Lett.* **2016**, *7*, 2009–2014, and private communication with the first author.

(48) Pápai, M.; Rozgonyi, T.; Penfold, T. J.; Nielsen, M. M.; Møller, K. B. Simulation of ultrafast excited-state dynamics and elastic x-ray scattering by quantum wavepacket dynamics. *The Journal of Chemical Physics* **2019**, *151*, 104307.

(49) Perdew, J. P.; Zunger, A. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B* **1981**, *23*, 5048–5079.

(50) Nocedal, J.; Wright, S. *Numerical Optimization*; Springer, New York, 2006.

(51) Stengel, M.; Spaldin, N. A. Self-interaction correction with Wannier functions. *Phys. Rev. B* **2008**, *77*, 155106.

(52) Klüpfel, P.; Klüpfel, S.; Tsemekhman, K.; Jónsson, H. Optimization of functionals of orthonormal functions in the absence of unitary invariance. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **2012**, *7134 LNCS*, 23–33.

(53) Lehtola, S.; Head-Gordon, M.; Jónsson, H. Complex orbitals, multiple local minima, and symmetry breaking in perdew-zunger self-interaction corrected density functional theory calculations. *J. Chem. Theory Comput.* **2016**, *12*, 3195–3207.

(54) Lehtola, S.; Parkhill, J.; Head-Gordon, M. Orbital optimisation in the perfect pairing hierarchy: applications to full-valence calculations on linear polyacenes. *Mol. Phys.* **2018**, *116*, 547–560.

(55) Borghi, G.; Park, C. H.; Nguyen, N. L.; Ferretti, A.; Marzari, N. Variational minimization of orbital-density-dependent functionals. *Phys. Rev. B* **2015**, *91*.

(56) Marzari, N.; Vanderbilt, D.; Payne, M. C. Ensemble density-functional theory for ab

initio molecular dynamics of metals and finite-temperature insulators. *Phys. Rev. Lett.* **1997**, *79*, 1337–1340.

(57) Lehtola, S.; Jónsson, H. Variational, Self-Consistent Implementation of the Perdew–Zunger Self-Interaction Correction with Complex Optimal Orbitals. *J. Chem. Theory Comput.* **2014**, *10*, 5324–5337.

(58) Fern Rico, J. Á.; Paniagua, M.; Fern Alonso, J. I.; Fantucci, P. Restricted Hartree–Fock approximation. II. Computational aspects of the direct minimization procedure. *Journal of Computational Chemistry* **1983**, *4*, 41–47.

(59) Rico, J. F.; De La Vega, J. M.; Alonso, J. I.; Fantucci, P. Restricted Hartree–Fock approximation. I. Techniques for the energy minimization. *Journal of Computational Chemistry* **1983**, *4*, 33–40.

(60) Douady, J.; Ellinger, Y.; Subra, R.; Levy, B. Exponential transformation of molecular orbitals: A quadratically convergent SCF procedure. I. General formulation and application to closed-shell ground states. *The Journal of Chemical Physics* **1980**, *72*, 1452–1462.

(61) Head-Gordon, M.; Pople, J. Optimization of wave function and geometry in the finite basis Hartree-Fock method. *Journal of Physical Chemistry* **1988**, *92*, 3063–3069.

(62) Ivanov, A. V.; Jónsson E.; Vegge, T.; Jónsson, H. Direct Energy Minimization Based on Exponential Transformation in Density Functional Calculations of Finite and Extended Systems. *arXiv:2101.12597 [physics.comp-ph]* **2021**,

(63) Briggs, E. L.; Sullivan, D. J.; Bernholc, J. Large-scale electronic-structure calculations with multigrid acceleration. *Phys. Rev. B* **1995**, *52*, R5471–R5474.

(64) Ref. 50, p. 121.

(65) Ref. 50, p. 177.

(66) Enkovaara, J.; Rostgaard, C.; Mortensen, J. J.; Chen, J.; Dulak, M.; Ferrighi, L.; Gavnholt, J.; Glinsvad, C.; Haikola, V.; Hansen, H. A.; Kristoffersen, H. H.; Kuisma, M.; Larsen, A. H.; Lehtovaara, L.; Ljungberg, M.; Lopez-Acevedo, O.; Moses, P. G.; Ojanen, J.; Olsen, T.; Petzold, V.; Romero, N. A.; Stausholm-Møller, J.; Strange, M.; Tritsaris, G. A.; Vanin, M.; Walter, M.; Hammer, B.; Häkkinen, H.; Madsen, G. K. H.; Nieminen, R. M.; Nørskov, J. K.; Puska, M.; Rantala, T. T.; Schiøtz, J.; Thygesen, K. S.; Jacobsen, K. W. Electronic structure calculations with GPAW: a real-space implementation of the projector augmented-wave method. *J. Phys.: Condens. Matter* **2010**, *22*, 253202.

(67) Mortensen, J.; Hansen, L.; Jacobsen, K. W. Real-space grid implementation of the projector augmented wave method. *Phys. Rev. B* **2005**, *71*, 035109.

(68) Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **1994**, *50*, 17953.

(69) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(70) Ziegler, T.; Rauk, A.; Baerends, E. J. On the calculation of multiplet energies by the hartree-fock-slater method. *Theor. Chim. Acta* **1977**, *43*, 261–271.

(71) Mewes, J. M.; Jovanović, V.; Marian, C. M.; Dreuw, A. On the molecular mechanism of non-radiative decay of nitrobenzene and the unforeseen challenges this simple molecule holds for electronic structure theory. *Phys. Chem. Chem. Phys.* **2014**, *16*, 12393–12406.

(72) Becke, A. D. Current density in exchange-correlation functionals: Application to atomic states. *The Journal of Chemical Physics* **2002**, *117*, 6935–6938.

(73) Johnson, E. R.; Dickson, R. M.; Becke, A. D. Density functionals and transition-metal atoms. *The Journal of Chemical Physics* **2007**, *126*, 184104.

(74) Klüpfel, S.; Klüpfel, P.; Jónsson, H. Importance of complex orbitals in calculating the self-interaction-corrected ground state of atoms. *Phys. Rev. A* **2011**, *84*, 050501.

(75) Klüpfel, S.; Klüpfel, P.; Jónsson, H. The effect of the Perdew-Zunger self-interaction correction to density functionals on the energetics of small molecules. *J. Chem. Phys.* **2012**, *137*, 124102.

(76) Lehtola, S.; Jónsson, E. Ö.; Jónsson, H. Effect of Complex-Valued Optimal Orbitals on Atomization Energies with the Perdew–Zunger Self-Interaction Correction to Density Functional Theory. *J. Chem. Theory Comput.* **2016**, *12*, 4296–4302.

(77) Small, D. W.; Sundstrom, E. J.; Head-Gordon, M. Restricted Hartree Fock using complex-valued orbitals: A long-known but neglected tool in electronic structure theory. *J. Chem. Phys.* **2015**, *142*, 24104.

(78) Lee, J.; Bertels, L. W.; Small, D. W.; Head-Gordon, M. Kohn-Sham Density Functional Theory with Complex, Spin-Restricted Orbitals: Accessing a New Class of Densities without the Symmetry Dilemma. *Phys. Rev. Lett.* **2019**, *123*, 113001.

(79) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098–3100.

(80) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785–789.

(81) Loos, P.-F.; Scemama, A.; Blondel, A.; Garniron, Y.; Caffarel, M.; Jacquemin, D. A Mountaineering Strategy to Excited States: Highly Accurate Reference Energies and Benchmarks. *J. Chem. Theory Comput.* **2018**, *14*, 4360–4379.

(82) Jonsson, H. Simulation of surface processes. *Proc. Natl. Acad. Sci.* **2011**, *108*, 944–949.

(83) Chipman, D. M. Stretching of hydrogen-bonded OH in the lowest singlet excited electronic state of water dimer. *J. Chem. Phys.* **2006**, *124*, 44305.

(84) Levi, G.; Biasin, E.; Dohn, A. O.; Jónsson, H. On the interplay of solvent and conformational effects in simulated excited-state dynamics of a copper phenanthroline photosensitizer. *Phys. Chem. Chem. Phys.* **2019**, *22*, 748–757.

(85) Levi, G.; Papai, M.; Henriksen, N. E.; Dohn, A. O.; Møller, K. B. Solution structure and ultrafast vibrational relaxation of the PtPOP complex revealed by ΔSCF-QM/MM Direct Dynamics simulations. *J. Phys. Chem. C* **2018**, *122*, 7100–7119.

(86) Dohn, A. O.; Jónsson, E. O.; Kjær, K. S.; B. van Driel, T.; Nielsen, M. M.; Jacobsen, K. W.; Henriksen, N. E.; Møller, K. B. Direct Dynamics Studies of a Binuclear Metal Complex in Solution: The Interplay Between Vibrational Relaxation, Coherence, and Solvent Effects. *J. Phys. Chem. Lett.* **2014**, *5*, 2414–2418.

(87) Ramos, P.; Pavanello, M. Nonadiabatic couplings from a variational excited state method based on constrained DFT. *The Journal of Chemical Physics* **2021**, *154*, 014110.

(88) Lundin, U.; Eriksson, O. Novel method of self-interaction corrections in density functional calculations. *Int. J. Quantum Chem.* **2001**, *81*, 247–252.

(89) Pederson, M. R.; Heaton, R. A.; Lin, C. C. Local-density Hartree-Fock theory of electronic states of molecules with self-interaction correction. *J. Chem. Phys.* **1984**, *80*, 1972–1975.

Supporting Information: Direct Optimization Method for Variational Excited-State Density Functional Calculations using Real-Space Grid or Plane Waves

Aleksei V. Ivanov, Gianluca Levi, Elvar Ö. Jónsson, and Hannes Jónsson*

Science Institute and Faculty of Physical Sciences, University of Iceland, VR-III, 107 Reykjavík, Iceland

E-mail: hj@hi.is

Contents

1 Ground-state molecular orbitals of $[Fe(bmip)_2]^{2+}$

Below are the molecular orbitals of the $[Fe(bmip)_2]^{2+}$ complex involved in the electronic transitions investigated in the present work. The isofurfaces are represented with gray and orange colors and correspond to isovalues of ± 0.16 Å$^{-3/2}$.

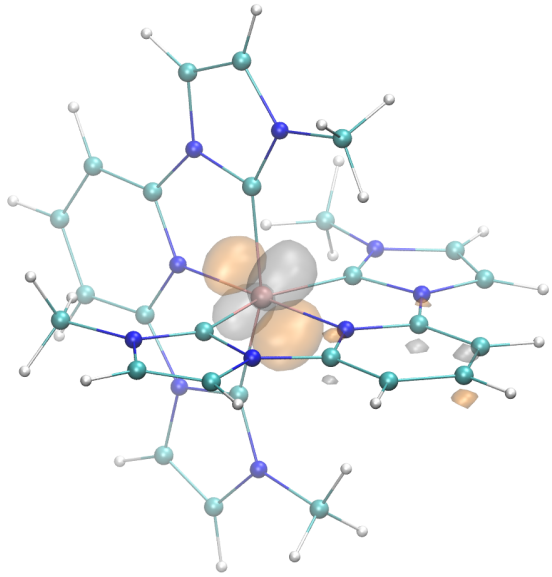

Figure S1: HOMO-1

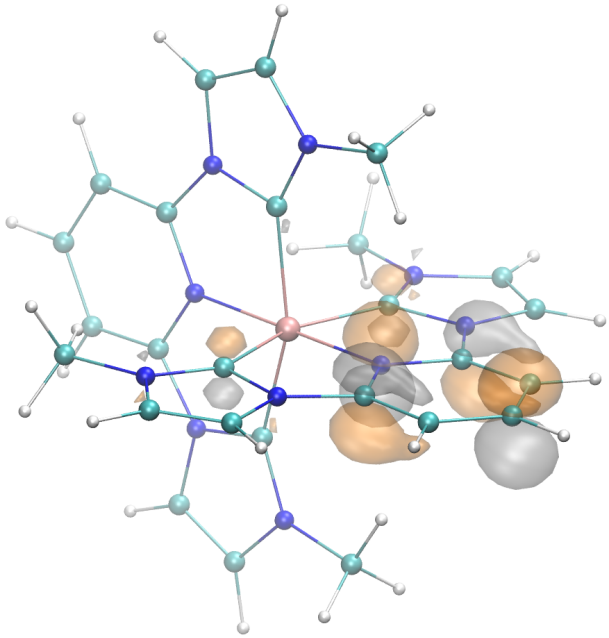

Figure S2: LUMO

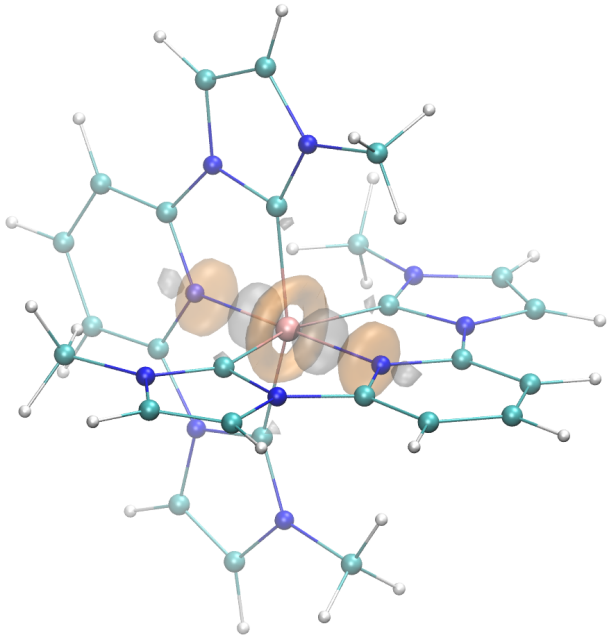

Figure S3: LUMO+4

2 Curvature of the ^{3}MC potential energy curve of $[Fe(bmip)_2]^{2+}$

The potential energy curves of the lowest ^{3}MC state of $[Fe(bmip)_2]^{2+}$ calculated with DO-MOM (BLYP) in the present work and linear response TDDFT (B3LYP*) from Ref.[S1] are fitted using a fourth order polynomial:

$$E(Q) = -d_e + \frac{k}{2}(Q - r_e)^2 - \alpha(Q - r_e)^3 + \beta(Q - r_e)^4, \quad (1)$$

where k is the curvature. Below are the results of the fitting of the curve calculated with DO-MOM (BLYP):

$$d_e = -1.934$$
$$k = 1.332 \times 10^{-2}$$
$$r_e = 5.060$$
$$\alpha = 1.265 \times 10^{-4}$$
$$\beta = -1.652 \times 10^{-6}$$

and of the curve calculated with TDDFT (B3LYP*):

$$d_e = -1.821$$
$$k = 1.263 \times 10^{-2}$$
$$r_e = 5.514$$
$$\alpha = 9.256 \times 10^{-4}$$
$$\beta = 2.461 \times 10^{-7}$$

The fitting curves are shown in Fig. S4 together with the points obtained from the DO-MOM (BLYP) and TDDFT (B3LYP*) calculations. The curvatures of the fitting curves differ by around 5%. The oscillation period estimated on the basis of TDDFT is 285 fs,[S1] and thus,

we obtain an estimation of an oscillation period on the basis of eDFT as

$$T_{eDFT} = T_{TDDFT}\sqrt{\frac{k_{TDDFT}}{k_{eDFT}}} \approx 280\,(\mathrm{fs})$$

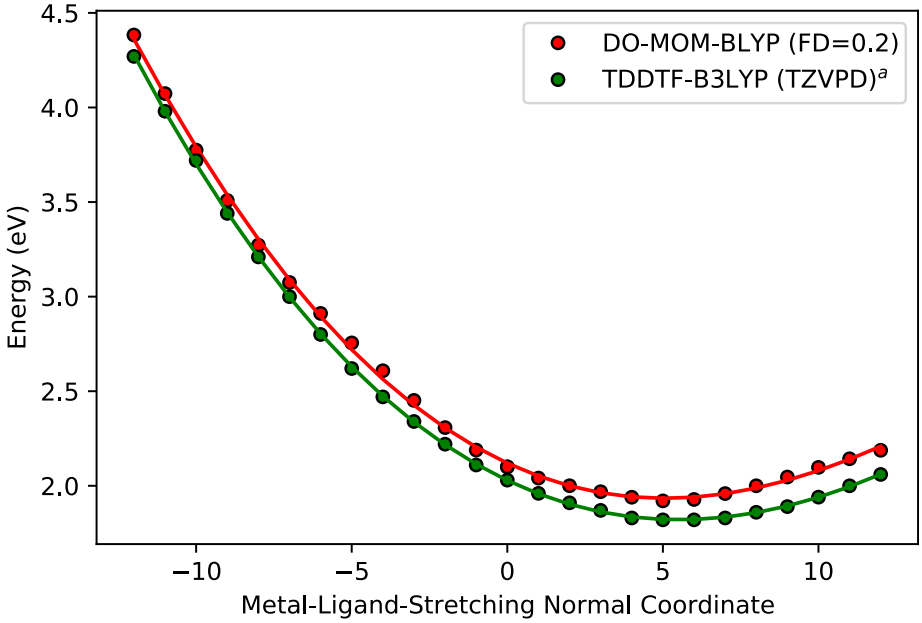

Figure S4: Potential energy curves along a metal-ligand normal coordinate (the breathing mode Q_6 as defined in Ref.[S1]) of the lowest ^{3}MC state. The DO-MOM (BLYP) calculations were performed using a finite-difference real-space basis set as described in the main text of the article. [a]Calculations from Ref.[S1]

3 Excitation Energy obtained from the Mixed-Spin Determinant

Table S1: Mixed-Spin States

species	excitation	PBE[a]	SIC/2[b]	SIC[b]	TBE[c]	EXP[d]
acetaldehyde	$1^1A''$ ($n \rightarrow \pi^*$; V)	3.79	3.74	3.69	4.31	4.27
acetylene	$1^1\Delta_u$($\pi \rightarrow \pi^*$; V)	6.51	6.81	6.90	7.10	7.2
ammonia	2^1A_1(n→3s; R)	6.30	6.25	6.23	6.66	6.38
carbon monoxide	$1^1\Pi$(n→ π^*; V)	6.70	6.90	7.51	8.48	8.51
diazomethane	1^1A_2($\pi \rightarrow \pi^*$; V)	2.85	2.47	2.01	3.13	3.14
ethylene	1^1B_{3u}(n→3s; R)	7.07	7.12	7.14	7.43	7.11
	1^1B_{1u}($\pi \rightarrow \pi^*$; V)	5.59	5.90	6.20	7.92	7.6
formaldehyde	1^1B_2(n→3s; R)	6.79	7.04	7.16	7.30	7.11
formamide	1^1A"(n→ π^*; V)	5.26	5.20	5.14	5.63	5.8
hydrogen sulfide	1^1A_2(n→4p; R)	5.51	5.53	5.49	6.10	
ketene	1^1B_1(π →3s; R)	5.75	5.87	5.95	6.06	5.86
methanimine	1^1A"(n→ π^*; V)	4.43	4.56	4.65	5.21	
thioformaldehyde	1^1A_2(n→ π^*; V)	1.81	1.78	1.72	2.20	2.03
	1^1B_2(n→4s; R)	5.47	5.63	5.72	5.99	5.85
	2^1A_1($\pi \rightarrow \pi^*$; V)	4.36	4.67	4.97	6.34	6.2
water	1^1B_1(n→3s; R)	7.28	7.27	7.24	7.70	7.41
	1^1A_2(n→3p; R)	8.83	8.94	9.04	9.47	9.2
	2^1A_1(n→3s; R)	9.50	9.48	9.46	9.97	9.67
	ME (TBE)	-0.73	-0.66	-0.59		
	ME (EXP)	-0.59	-0.52	-0.49		
	MAE (TBE)	0.73	0.66	0.59		
	MAE (EXP)	0.59	0.52	0.50		
	RMSE (TBE)	0.95	0.83	0.73		
	RMSE (EXP)	0.88	0.76	0.68		

[a] Employing real-valued orbitals; [b] Initial guess for the wave functions is PBE real-valued orbitals followed by complex Rudenberg-Edmiston localization; [c] Corrected theoretical best estimates as given in Ref.;[S2] [d] Listed in Ref.[S2] (see references therein);

4 The Perdew Zunger Self-Interaction Correction Restricted to Real Orbitals (rSIC)

Table S2: Excitation energy from the PBE-rSIC/2 functional

species	excitation	Mixed Spin	Triplet	Singlet
acetaldehyde	$1^1A''$ ($n \rightarrow \pi^*$; V)	3.71	3.73	3.69
acetylene	$1^1\Delta_u$($\pi \rightarrow \pi^*$; V)	6.56	5.60	7.52
ammonia	2^1A_1(n→3s; R)	6.15	6.01	6.29
carbon monoxide	$1^1\Pi$(n→ π^*; V)	6.94	5.79	8.09
diazomethane	1^1A_2($\pi \rightarrow \pi^*$; V)	2.26	2.15	2.36
ethylene	1^1B_{3u}(n→3s; R)	6.87	6.80	6.93
	1^1B_{1u}($\pi \rightarrow \pi^*$; V)	5.90	4.62	7.19
formaldehyde	1^1B_2(n→3s; R)	6.90	6.85	6.96
formamide	1^1A"(n→ π^*; V)	5.01	5.06	4.96
hydrogen sulfide	1^1A_2(n→4p; R)	5.46	5.42	5.49
ketene	1^1B_1(π →3s; R)	5.73	5.24	6.22
methanimine	1^1A"(n→ π^*; V)	4.66	4.39	4.92
thioformaldehyde	1^1A_2(n→ π^*; V)	1.75	1.80	1.69
	1^1B_2(n→4s; R)	5.60	5.51	5.69
	2^1A_1($\pi \rightarrow \pi^*$; V)	4.72	3.15	6.29
water	1^1B_1(n→3s; R)	7.00	6.84	7.16
	1^1A_2(n→3p; R)	8.69	8.63	8.76
	2^1A_1(n→3s; R)	9.43	9.21	9.66
	ME (TBE[a])	-0.73	-0.38	-0.40
	ME (EXP[b])	-0.59	-0.23	-0.24
	MAE (TBE[a])	0.73	0.39	0.46
	MAE (EXP[a])	0.59	0.26	0.34
	RMSE (TBE[a])	0.95	0.44	0.50
	RMSE (EXP[a])	0.88	0.31	0.41

[a] Deviation from corrected theoretical best estimates as given in Ref.;[S2] [b] Deviation from experimental values collected in Ref.[S2] (see references therein);

5 Ground-State SIC Complex Orbitals of the Water Monomer

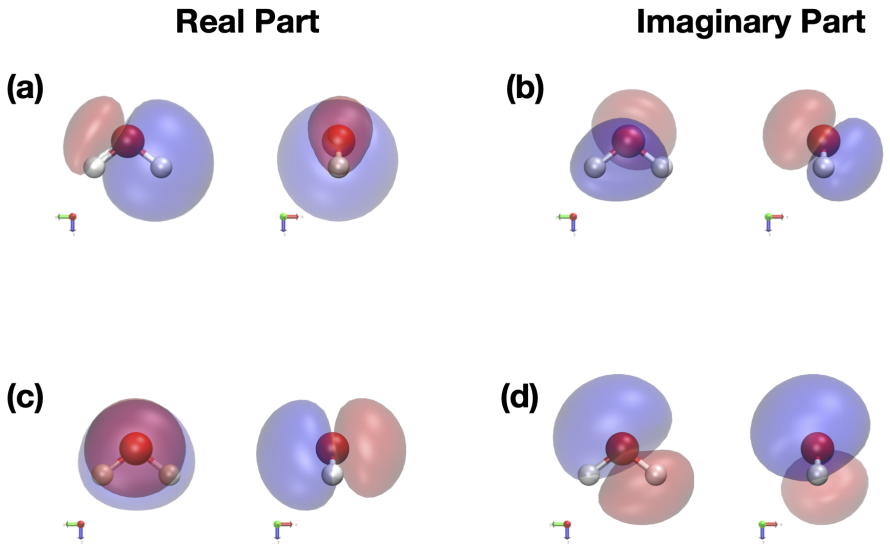

Figure S5: Real and imaginary parts of optimal ground-state SIC complex orbitals of the water monomer. The isofurfaces are represented with blue and red colors and correspond to isovalues of ± 0.05 Å$^{-3/2}$. There are two degenerate pairs and only one orbital of each degenerate pair is shown. One optimal orbital (GS_1 as denoted in Table 3 of the main text) is made of real part (a) and imaginary part (b). The other optimal orbital (GS_2 as denoted in Table 3 of the main text) is made of real part (c) and imaginary part (d). Each figure (a), (b), (c), and (d) contains its own isosurfaces from two different views as shown by the axis in insets.

References

(S1) Pápai, M.; Rozgonyi, T.; Penfold, T. J.; Nielsen, M. M.; Møller, K. B. Simulation of ultrafast excited-state dynamics and elastic x-ray scattering by quantum wavepacket dynamics. *The Journal of Chemical Physics* **2019**, *151*, 104307.

(S2) Loos, P.-F.; Scemama, A.; Blondel, A.; Garniron, Y.; Caffarel, M.; Jacquemin, D. A Mountaineering Strategy to Excited States: Highly Accurate Reference Energies and Benchmarks. *J. Chem. Theory Comput.* **2018**, *14*, 4360–4379.

Article IV

Mn Dimer can be Described Accurately with Density Functional Calculations when Self-interaction Correction is Applied

Ivanov A. V., Ghosh T. K., Jónsson E. Ö., Jónsson H.

Accepted to The Journal of Physical Chemistry Letters,
arxiv:2102.00890 [physics.chem-ph].

Mn Dimer can be Described Accurately with Density Functional Calculations when Self-interaction Correction is Applied

Aleksei V. Ivanov,[†,‡] Tushar K. Ghosh,[†,¶] Elvar Ö. Jónsson,[†,¶] and Hannes Jónsson[*,†,¶]

†*Science Institute and Faculty of Physical Sciences, University of Iceland VR-III, 107 Reykjavík, Iceland.*

‡*St. Petersburg State University, 199034, St. Petersburg, Russia*

¶*Department of Applied Physics, Aalto University, Espoo, FI-00076, Finland.*

E-mail: hj@hi.is

Abstract

Qualitatively incorrect results are obtained for the Mn dimer in density functional theory calculations using the generalized gradient approximation (GGA) and similar results are obtained from local density and meta-GGA functionals. The coupling is predicted to be ferromagnetic rather than antiferromagnetic and the bond between the atoms is predicted to be an order of magnitude too strong and about an Ångstrøm too short. Explicit, self-interaction correction (SIC) applied to a commonly used GGA energy functional, however, provides close agreement with both experimental data and high-level, multi-reference wave function calculations. These results show that the failure is not due to strong correlation but rather the single electron self-interaction that is necessarily introduced in estimates of the classical Coulomb and exchange-correlation energy when only the total electron density is used as input. The corrected functional depends explicitly on the orbital densities and can, therefore, avoid the introduction of self-Coulomb interaction. The error arises because of over-stabilization of bonding *d*-states in the minority spin channel resulting from an overestimate of the *d*-electron self-interaction in the semi-local exchange-correlation functionals. Since the computational effort in the self-interaction corrected calculations scales with system size in the same way as for regular semi-local functional calculations, this approach provides a way to calculate properties of Mn nanoclusters as well as biomolecules and extended solids where Mn dimers and larger cluster are present, while multi-reference wave function calculations can only be applied to small systems.

Manganese atoms have large magnetic moment and are, therefore, of interest for various technical applications. Complexes of Mn atoms are also found in several metalloenzymes,[1,2] for example in the oxygen-evolving photo-system II. The properties of such systems are of great interest and theoretical calculations could in principle provide valuable information to help gain an understanding of the role Mn atoms play. However, theoretical calculations prove to be particularly challenging for these systems. The Mn dimer is the simplest manganese complex and it represents an important test system for theoretical methods that could ultimately be used for the larger and more complex systems. Its properties are quite well known from electron spin resonance and optical absorption measurements of dimers confined in a rare gas matrix. The ground state is found to be antiferromagnetic[3–5] with a small bond energy of 0.13±0.1 eV[6] and a large bond length of 3.4 Å.[4,7] Resonance Raman spectra give vibrational frequency of 76 cm^{-1} in a Kr matrix[8] and 68 cm^{-1} in a Xe matrix.[5]

High-level wave function calculations of an isolated Mn dimer give results in close agreement with the experimental measurements. Both complete-active-space self-consistent field in combination with second-order perturbation theory,[9] as well as multi-reference[9–13] calculations have been carried out. They predict bond energy in the range 0.10 - 0.14 eV and bond length of 3.3 - 3.8 Å in an antiferromagnetic ground state with a coupling constant of J=-5.8 cm^{-1}. It is clear from the close agreement between these calculations and the experimental measurements that the effect of the confining rare gas matrix is small and the measurements indeed probe the properties of the Mn dimer. Such high-level, wave function based calculations become, however, impractical for larger systems due to the strong scaling of the computational effort with system size.

Kohn-Sham density functional theory (KS-DFT)[14] can provide a valuable tool for theoretical studies of large systems with up to several hundred atoms, free – in principle – of adjustable parameters with unknown values. Unfortunately, the results of such calculations for the Mn dimer with commonly used energy functionals such as local density approximation (LDA) and generalized gradient approximation (GGA), are in strong disagreement with the

experimental measurements and the high-level, wave function calculations.[15–17] The ground state is predicted to be ferromagnetic rather than antiferromagnetic with the bond between the Mn atoms being much too strong and too short. All electron calculations with various GGA functionals give a binding energy of around 0.9 eV and bond length of 2.57-2.61 Å.[17] Below, we present results using an elaborate meta-GGA functional, SCAN,[18] where the Mn dimer is also found to be poorly described. We will from now on refer to calculations using LDA, GGA or meta-GGA functionals collectively as semi-local DFT calculations.

Similar failure in DFT calculations of magnetic coupling constants has also been reported for various manganese binuclear complexes.[19] Shortcomings of DFT calculations are often ascribed to 'strong correlation' and systems where large errors are obtained are often characterized as 'highly correlated systems' (for a recent discussion of a possible meaning of these terms, see Ref.[20]). The Mn dimer is an example of such a system. It has, however, been shown that calculated results for the Mn dimer can be improved when exact exchange is added to a semi-local DFT functional to form a hybrid functional, but the weight of the exact exchange in this blend needs to be significantly larger than the range of 0.20 to 0.25 in commonly used functionals of this form.[17,19,21] Semi-local DFT functionals tend to give errors of opposite sign to those of Hartree-Fock calculations, so some mix of the two can often be tuned to give the desired result.

An alternative reason for the failure of semi-local DFT calculations is the self-interaction error that is necessarily introduced in the estimate of the classical Coulomb interaction between the electrons when only the total electron density is used as input. This error is highly non-local. In the exchange-correlation part of the semi-local functionals a correction is estimated, i.e. a self-interaction contribution of opposite sign, but the cancellation can mathematically not be complete and, therefore, self-interaction error remains and can lead to erratic results. Previously it has been speculated that the self-interaction error is not responsible for the poor performance of semi-local DFT functionals for the Mn dimer but rather strong correlation.[15,16] So, the question remains whether the large error in semi-local

DFT calculations of the Mn dimer is due to strong correlation or self-interaction, or possibly some other source of error.

In this letter, the results of variational, self-consistent calculations are presented where the self-interaction error is removed explicitly as proposed by Perdew and Zunger.[22] As described below, the results are found to be in remarkably good agreement with the high-level wave function calculations. While any form of correlation necessarily reflects interaction between two or more electrons, the self-interaction error in semi-local DFT functionals is present even for systems containing a single electron and the correction applied here involves terms that depend only on one-electron densities. The results presented here therefore demonstrate that the problem in the semi-local DFT calculations of the Mn dimer is not related to correlation, but rather self-interaction error. In addition to providing important insight into the reason for the failure of DFT for the Mn dimer, this opens up an avenue for accurate calculations of larger systems containing Mn complexes because the computational effort of the GGA calculations with self-interaction correction scales with system size in the same way as regular GGA calculations.[23]

In Kohn-Sham density functional theory,[14] the energy of an electronic system is estimated from

$$E^{\mathrm{KS}}[\rho] = T_s + \int v_{\mathrm{ext}}(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} + E_{\mathrm{C}}[\rho] + E_{xc}[\rho_\uparrow, \rho_\downarrow] \tag{1}$$

were, T_s is the kinetic energy of an independent electron system described by spin-orbitals ϕ and the electron density

$$\rho_\sigma(\mathbf{r}) \;=\; \sum_{i_\sigma} \rho_{i_\sigma}(\mathbf{r}) \;=\; \sum_{i_\sigma} |\phi_{i_\sigma}(\mathbf{r})|^2 \tag{2}$$

corresponds to the ground state electron density of the interacting electron system for each spin channel, $\sigma = \{\uparrow, \downarrow\}$. The energy due to the electron-nuclei interaction, described by the external potential v_{ext}, can be evaluated correctly from the total electron density, $\rho(\mathbf{r}) =$

$\rho_\uparrow(\mathbf{r}) + \rho_\downarrow(\mathbf{r})$, but the estimate of the classical Coulomb interaction between the electrons

$$E_{\mathrm{C}}[\rho] \;=\; \int\int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r} d\mathbf{r}' \tag{3}$$

includes spurious interaction of each electron with itself. This is most clearly seen for a system with a single electron where non-zero interaction energy is obtained from this estimate. A more accurate estimate can be obtained by using the spin-orbital densities, ρ_i, where interaction of a spin-orbital with itself is avoided

$$E_{\mathrm{C}}^{\mathrm{SIC}}[\rho_1, \ldots, \rho_N] \;=\; E_{\mathrm{C}}[\rho] - \sum_i \int\int \frac{\rho_i(\mathbf{r})\rho_i(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r} d\mathbf{r}' = \sum_i \sum_{j>i} \int\int \frac{\rho_i(\mathbf{r})\rho_j(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r} d\mathbf{r}' \tag{4}$$

Here, the summation indices run over both spin channels, $i = \{i_\uparrow, i_\downarrow\}$. The exchange-correlation energy term in the KS-DFT functional, E_{xc}, attempts to provide such a correction but, because of the semi-local form, cannot accurately cancel out the non-local self-interaction error in $E_{\mathrm{C}}[\rho]$.

Perdew and Zunger proposed a procedure where the net self-interaction error is estimated for each spin-orbital and the sum subtracted from the Kohn-Sham functional[22]

$$E^{\mathrm{SIC}}[\rho_1, \ldots, \rho_N] = E^{\mathrm{KS}}[\rho] - \sum_i \left(E_{\mathrm{C}}[\rho_i] + E_{xc}[\rho_i, 0] \right). \tag{5}$$

This provides the correction to the classical Coulomb energy as in Eqn. (4) and also addresses the extent to which the exchange-correlation functional is able to cancel out the self-interaction by evaluating the net self-interaction for each spin-orbital separately. For a one electron system, the corrected functional is guaranteed to be self-interaction free, but for many electron systems, this correction procedure is approximate. While this approach was originally proposed for the LDA functional, it can also be applied to GGA functionals but there it has been found to give an overcorrection and a scaling by 1/2 has been shown to give good results for a wide range of systems and properties such as atomization energy of

molecules, band gaps of solids and the balance between localized and delocalized electronic states.[24–26] We choose here to use the PBE functional,[27] a GGA functional approximation that is commonly used in calculations of condensed matter and refer to the corrected functional as PBE-SIC/2. As discussed below, the scaling of 1/2 is not essential here, similar results are obtained for the Mn dimer without the scaling.

The corrected functional is not unitary invariant as it depends explicitly on the orbital densities, as indicated in Eqs. (4-5), and it turns out that the optimal spin-orbitals that minimize the energy of the system are hybrid orbitals, i.e. linear combinations of the canonical orbitals that are eigenfunctions of the Hamiltonian. The calculations need to make use of complex-valued functions to represent the optimal spin-orbitals.[28–30] Here, a real-space grid has been used[31] combined with projector augmented wave (PAW) to represent the effect of frozen core-electrons of the atoms.[32] A localized atomic orbital basis set is used including primitive Gaussians from the def2-TZVPD basis set[33–37] augmented with a single-zeta basis.[38] Tests against calculations using full flexibility of the real-space grid with mesh size of 0.15 Å give nearly identical results, showing that the atomic basis set is sufficient. Starting from localized orbitals,[39] the energy is variationally minimized using an exponential transform direct optimization method described elsewhere.[40,41]

Fig. 1 shows the energy of the three lowest energy electronic states of the Mn dimer as a function of the distance between the atoms, calculated with the PBE functional with and without self-interaction correction. Three relevant, low lying electronic states are found in the calculations: two ferromagnetic states, $^{11}\Sigma_u^+$ and $^{11}\Pi_u$, and an antiferromagnetic $^{1}\Sigma_g^+$ state. The self-interaction corrected functional predicts the antiferromagnetic state to be the ground state, and gives a binding energy of 0.18 eV and bond length of 3.32 Å, in close agreement with the experimental results as well as the high-level wave function calculations. The ferromagnetic $^{11}\Sigma_u^+$ state is only 0.04 eV higher in energy at the optimal bond length. A vibrational frequency of 69 cm^{-1} is obtained from the ground state energy curve, in good agreement with the experimental measurements.[5,8] The magnetic coupling constant is

calculated using the relationship[17]

$$J = \frac{E_{AF} - E_F}{\langle S_F^2 \rangle - \langle S_{AF}^2 \rangle}, \tag{6}$$

where E_{AF} and E_F are the energy of the antiferromagnetic and ferromagnetic states, and $< S^2 >$ is evaluated taking into account spin contamination.[42] This gives J= -6.0 cm^{-1}, close to the value obtained in the MCQDPT2 calculations.[9] The scaling of the self-interaction correction is not important in this case. When the full correction is used as in Eqn. (5), similar results are obtained, namely bond energy of 0.12 eV, bond length of 3.5 Å and vibrational frequency of 56 cm^{-1}. Most importantly, the self-interaction correction, scaled or not scaled, gives the right antiferromagnetic ground state.

Different results are obtained with the uncorrected PBE functional, as can be seen from Fig. 1. There, the ferromagnetic state $^{11}\Pi_u$ is lowest in energy and even the other ferromagnetic state, $^{11}\Sigma_u^+$, turns out to be lower than the antiferromagnetic, $^1\Sigma_g^+$ state at the experimental bond length. The calculated binding energy in the ground state is 0.90 eV with a bond length of 2.6 Å in strong disagreement with best estimates. These results are consistent with previously reported calculations using semi-local functionals.[17] At a large distance between the Mn atoms, the $^{11}\Pi_u$ state becomes higher in energy than the others as it dissociates into a d^6s^1 configuration for one of the Mn atoms. Calculations with the meta-GGA SCAN functional[18] were also carried out using the VASP software[43] and the results are qualitatively similar to the PBE results in that they also give the ferromagnetic $^{11}\Pi_u$ state as the ground electronic state, but the binding energy is smaller than for PBE, 0.46 eV. The SCAN functional can produce an antiferromagnetic ground state if a U-term is added, effectively mimicking a self-interaction correction (see supporting information in Ref.[44]).

In order to analyse this failure of the semi-local functional calculations, a molecular orbital diagram is shown in Fig. 2. Interestingly, the occupation of minority-spin molecular

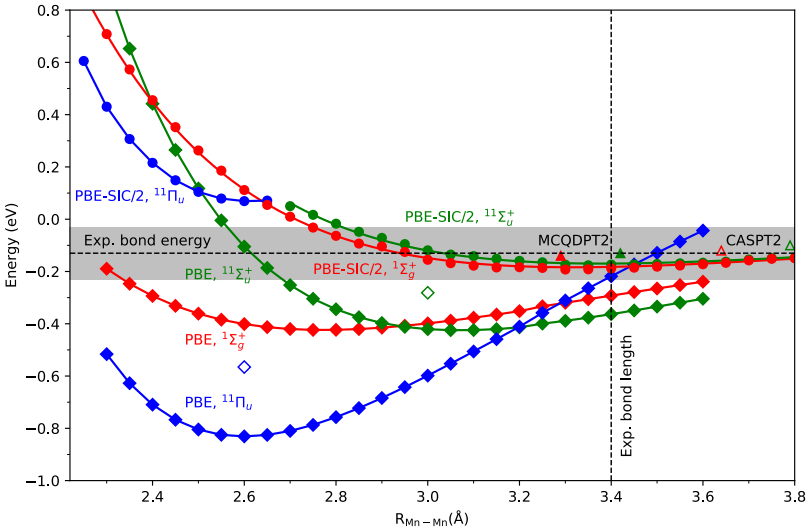

Figure 1: Energy of the Mn dimer in the three lowest lying electronic states as a function of distance between the atoms calculated with the PBE functional (filled diamonds) and self-interaction corrected PBE-SIC/2 (filled circles). The zero of energy is twice the energy of an Mn atom in the ^{6}S electron configuration. The ferromagnetic states $^{11}\Sigma_u^+$ and $^{11}\Pi_u$ are shown in blue and green, respectively, and the antiferromagnetic state $^1\Sigma_g^+$ in red. The ground state in the PBE calculations is ferromagnetic with a binding energy of nearly 0.9 eV and bond length of 2.6 Å while the antiferromagnetic state is the ground state in the self-interaction corrected PBE-SIC/2 calculations, with binding energy of 0.18 eV and bond length of 3.32 Å. Experimental estimates[4,6,7] of the bond length and bond energy (shaded grey area indicating the estimated error bar) are shown with dashed lines. The triangles show results of high-level calculations using MCQDPT2[9] (filled) and CASPT2[10] (open). The self-interaction corrected PBE-SIC/2 calculation is in close agreement with experimental measurements as well as the high-level calculations, while the PBE results are qualitatively incorrect. Open diamonds show results of calculations where the PBE energy is evaluated for the PBE-SIC/2 electron density, showing that the dominant error is in the self-interaction rather than the electron density.

orbitals turns out to play an important role here. In the PBE calculation, the bonding minority spin orbital $\pi(3d)$ is lower in energy than the anti-bonding $\sigma^*(4s)$, and this leads to d-d bond formation in the minority spin states. When the self-interaction correction is applied, the relative energy of $\pi(3d)$ and $\sigma^*(4s)$ is reversed and the anti-bonding spin-orbital

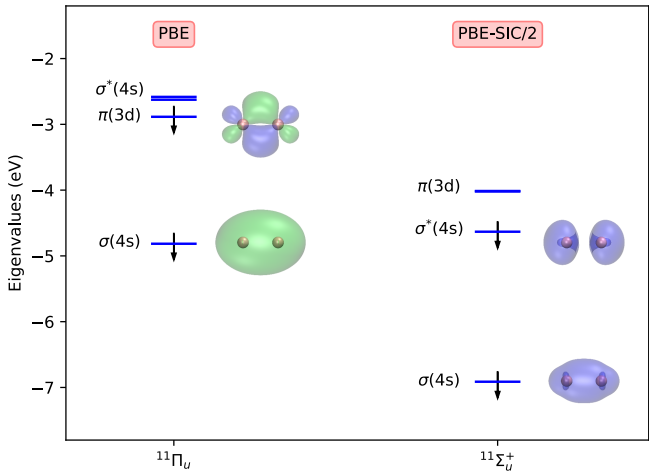

Figure 2: Molecular orbital diagram for the lowest energy ferromagnetic state of the Mn dimer calculated with the PBE functional and with the self-interaction corrected PBE-SIC/2 functional, based on orbital energy of the canonical orbitals. In the PBE calculations, the bonding $\pi(3d)$ minority spin orbital has lower energy than the antibonding $\sigma^*(4s)$ orbital, and becomes populated, resulting in a $^{11}\Pi_u$ ground state for the Mn dimer. When self-interaction correction is applied, in the PBE-SIC/2 functional, the relative energy of these two molecular spin-orbitals is reversed, and $\sigma^*(4s)$ becomes populated resulting in a $^{11}\Sigma_u^+$ state of the dimer. The magnetic coupling then makes the $^1\Sigma_g^+$ state slightly lower in energy than the $^{11}\Sigma_u^+$ state, by 0.04 eV, resulting in an antiferromagnetic ground state in the PBE-SIC/2 calculation. The surfaces illustrating the molecular spin-orbitals in the insets correspond to a value of ± 0.08 Å$^{-3/2}$, where different colors indicate the sign for the PBE orbitals, but for PBE-SIC/2 the amplitude of the orbitals is shown as they are complex.

is occupied instead. The relative energy of the d and s atomic orbitals is an important issue here.

It is well known that semi-local DFT functionals do not describe well the energy balance between localized and delocalized electrons, and this is reflected in the relative energy of d and s atomic orbitals. The repulsive self-interaction error in the classical Coulomb interaction is larger the more localized the electrons are. One might, therefore, expect that d electrons are calculated to be high in energy compared with s electrons. However, the results of the calculations presented here using the PBE functional show the opposite trend, as illustrated

in the molecular orbital diagram of Fig. 2 where the PBE functional ends up placing electrons in a molecular spin-orbital formed from d atomic orbitals rather than the one formed from s atomic orbitals. The answer lies in the estimate of the self-interaction in the exchange-correlation part of the PBE functional, a contribution that should cancel out the repulsive self-interaction in the classical Coulomb term. This estimate is based on an analysis of smoothly varying electron gas. The larger the deviation is from the uniform electron gas, the larger an error can be expected in the exchange-correlation functional. The extent to which the self-interaction is cancelled out by the two contributions for the various atomic orbitals in the Mn atom as well as a few other atoms is shown in Fig. 3. The cancellation of the self-interaction energy for the $3s$ atomic orbital is quite good, the net self-interaction being only 0.03 eV. But, the net self-interaction has a larger magnitude and is negative for the $3d$ orbitals of the Mn atom, -2.8 eV. This might be related to the fact that the d electrons exhibit a multi-center character which the semi-local functionals often fail to describe.[45] Similar trend is observed for orbitals of other atoms, as shown in the figure. The self-interaction correction in the semi-local exchange-correlation functional is, therefore, an overcorrection and makes the d atomic orbitals too low in energy as compared with the s atomic orbitals. This leads to the population of a bonding $\pi(3d)$ molecular spin-orbital in the Mn dimer instead of an anti-bonding $\sigma^*(4s)$ spin-orbital. As a result, the Mn dimer is overbound in an incorrect electronic ground state in the semi-local DFT calculations.

There are two aspects of the error in DFT calculations:[46] (1) an error in the self-consistent electron density, and (2) an error in the energy obtained for a given, possibly correct, electron density. A calculation using the PBE functional with the PBE-SIC/2 electron density as input for the Mn dimer is also shown in Fig. 1. The binding energy for both ferromagnetic states is reduced, but the $^{11}\Pi_u$ state is still the lowest energy state and there is still large overbinding of the dimer. This shows that the self-interaction is the main source of error rather than the electron density.

To analyze this further, the electron interaction terms are evaluated using PBE for each

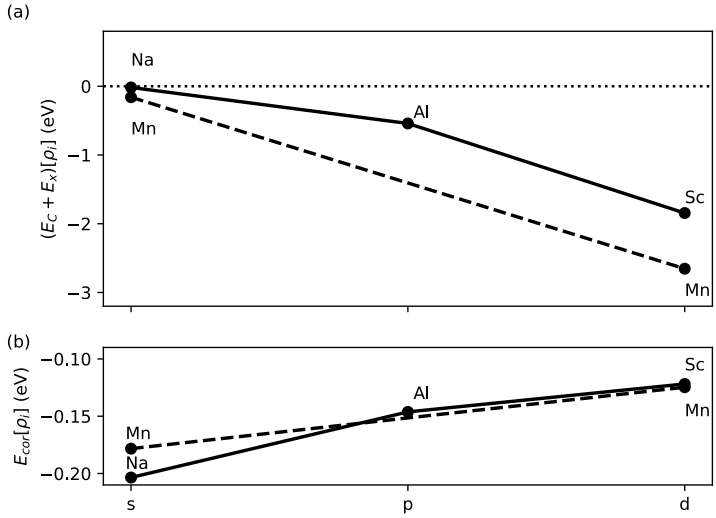

Figure 3: (a) Net self-interaction, $E_{\mathrm{C}}[\rho_i]+E_x[\rho_i,0]$, evaluated for atomic orbitals using orbital densities obtained from PBE calculations of the atoms. Results are shown for the Mn atom, as well as a few other atoms (Na, Al, Sc) to illustrate the trend of increasing magnitude of the negative net self-interaction as the orbital angular momentum increases. For s orbitals, the negative self-interaction energy in the exchange part of the PBE functional compensates well the positive self-interaction energy in the classical Coulomb part, but for d orbitals it overcorrects. The d orbitals in the Mn atom are, therefore, too low in energy compared to the s orbitals and this leads to incorrect ordering of the bonding $\pi(3d)$ molecular spin-orbital and the anti-bonding $\sigma^*(4s)$ spin-orbital in the minority-spin channel, as illustrated in Fig. 2. (b) Self-interaction in the correlation part of the PBE functional evaluated for the atomic orbitals as in (a). Here, the largest contribution is obtained for the s orbitals, but this contribution is an order of magnitude smaller than the one shown in (a).

of the Mn_2 molecular spin-orbitals separately using the optimal PBE-SIC/2 orbital densities (see Table 1). The net self-interaction energy estimated for the majority-spin $^{11}\Pi_u$ and $^{11}\Sigma_u^+$ states differs only by *ca.* 20 meV so the relative energy of these states is not affected significantly by the self-interaction error. The reason is that these two states have similar type of bonding, both involve ten molecular orbitals formed from d electrons and two molecular orbitals formed from s electrons. However, for the minority-spin states, the magnitude of the self-interaction energy is quite different, being 1.2 eV larger for the $^{11}\Pi_u$ state than the $^{11}\Sigma_u^+$

state. The reason is that the former involves an orbital formed from d electrons, $\pi(3d)$, as well the $\sigma(4s)$ orbital formed from s electrons, while the latter involves only orbitals formed from s electrons. As a result, the subtraction of the self-interaction error from the PBE functional has a large effect and reverses the relative energy of the $^{11}\Pi_u$ and $^{11}\Sigma_u^+$ states.

Table 1: The sum of the classical Coulomb and exchange-correlation self-interaction energy evaluated with PBE for each molecular spin-orbital separately, using optimal spin-orbitals from the PBE-SIC/2 calculation. Only the 14 valence electrons are included. For the majority-spin orbitals, the difference is just 0.02 eV, because they both involve ten molecular orbitals formed from d electrons and two molecular orbitals formed from s electrons. But, for the minority-spin, the $^{11}\Sigma_u^+$ state involves two orbitals formed from s electrons, $\sigma(4s)$ and $\sigma^*(4s)$, while the $^{11}\Pi_u$ state involves an orbital formed from d electrons, $\pi(3d)$, as well as the $\sigma(4s)$ orbital. The self-interaction has, therefore, different magnitude for the minority-spin, being 1.2 eV larger for the $^{11}\Pi_u$ state than for the $^{11}\Sigma_u^+$ and, thereby, reverses the relative stability of these states.

	Majority spin			Minority spin		
	$\sum E_C[\rho_i]$	$\sum E_{xc}[\rho_i]$	total_{maj}	$\sum E_C[\rho_i]$	$\sum E_{xc}[\rho_i]$	total_{min}
$^{11}\Pi_u$	114.9	-126.0	-11.10	8.8	-10.6	-1.8
$^{11}\Sigma_u^+$	118.6	-129.7	-11.08	7.2	-7.8	-0.6

In summary, KS-DFT calculations with the LDA, GGA or meta-GGA functional approximations give qualitatively incorrect results for the Mn dimer. The reason for this failure is not strong correlation but rather the one-electron self-interaction error that results from incomplete cancellation of the non-local self-interaction that is necessarily included in an estimate of the classical Coulomb interaction using only the total electron density as input and the semi-local estimate of the compensating self-interaction in the exchange correlation term in these functionals. It turns out that the net self-interaction error is particularly large for the d-electrons and artificially lowers their energy with respect to that of s-electrons because of an overestimate of the self-interaction in the semi-local exchange correlation functionals. As a result, an electronic state that dissociates into an Mn atom with a d^6s^1 electron configuration becomes the ground state and much too strong bonding is obtained because of constructive overlap of the d-orbitals in the minority spin channel. The results presented here show that an explicit self-interaction correction for each spin-orbital applied to the PBE

functional accurately describes the Mn dimer, predicting bond distance, binding energy, vibrational frequency and magnetic coupling constant in close agreement with experimental measurements and high-level quantum chemistry calculations. Since the computational effort in the self-interaction corrected calculations scales with system size in the same way as DFT calculations with semi-local functionals, i.e. as the system size to the third power, the method can be applied to large systems including extended solids.[24,26] There are several other formulations of self-interaction correction and different implementations[47] and it would be interesting in future work to see how well they perform in calculations of the Mn dimer.

Acknowledgement

This work was supported by the Icelandic Research Fund and the Academy of Finland. AVI is supported by a doctoral fellowship from the University of Iceland and thanks Tuomas Rossi and Valery Uzdin for helpful discussions.

References

(1) Hanson, G., Berliner, L., Eds. *Metals in Biology: Applications of High-Resolution EPR to Metalloenzymes*; Springer: New York, NY, 2010; Vol. 29.

(2) Dismukes, G. C. Manganese Enzymes with Binuclear Active Sites. *Chem. Rev.* **1996**, *96*, 2909–2926.

(3) Rivoal, J.-C.; Emampour, J. S.; Zeringue, K. J.; Vala, M. Ground-state exchange energy of the Mn_2 antiferromagnetic molecule. *Chem. Phys. Lett.* **1982**, *92*, 313–316.

(4) Baumann, C. A.; Zee, R. J. V.; Bhat, S. V.; Jr., W. W. ESR of Mn_2 and Mn_5 molecules in rare-gas matrices. *J. Chem. Phys.* **1983**, *78*, 190–199.

(5) Kirkwood, A. D.; Bier, K. D.; Thompson, J. K.; Haslett, T. L.; Huber, A. S.;

Moskovits, M. Ultraviolet-visible and Raman spectroscopy of diatomic manganese isolated in rare-gas matrixes. *J. Phys. Chem.* **1991**, *95*, 2644–2652.

(6) Kant, A.; Lin, S.; Strauss, B. Dissociation Energy of Mn_2. *J. Chem. Phys.* **1968**, *49*, 1983–1985.

(7) Cheeseman, M.; Van Zee, R. J.; Flanagan, H. L.; Weltner, W. Transition metal diatomics: Mn_2, Mn_2^+, CrMn. *J. Chem. Phys.* **1990**, *92*, 1553–1559.

(8) Bier, K. D.; Haslett, T. L.; Kirkwood, A. D.; Moskovits, M. The resonance Raman and visible absorbance spectra of matrix isolated Mn_2 and Mn_3. *J. Chem. Phys.* **1988**, *89*, 6–12.

(9) Yamamoto, S.; Tatewaki, H.; Moriyama, H.; Nakano, H. A study of the ground state of manganese dimer using quasidegenerate perturbation theory. *J. Chem. Phys.* **2006**, *124*, 124302.

(10) Wang, B.; Chen, Z. Magnetic coupling interaction under different spin multiplets in neutral manganese dimer: CASPT2 theoretical investigation. *Chem. Phys. Lett.* **2004**, *387*, 395 – 399.

(11) Camacho, C.; Yamamoto, S.; Witek, H. A. Choosing a proper complete active space in calculations for transition metal dimers: ground state of Mn_2 revisited. *Phys. Chem. Chem. Phys.* **2008**, *10*, 5128–5134.

(12) Angeli, C.; Cavallini, A.; Cimiraglia, R. An ab initio multireference perturbation theory study on the manganese dimer. *J. Chem. Phys.* **2008**, *128*, 244317.

(13) Buchachenko, A. A.; Chałasiński, G.; Szczęśniak, M. M. Electronic structure and spin coupling of the manganese dimer: The state of the art of ab initio approach. *J. Chem. Phys.* **2010**, *132*, 024312.

(14) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133–A1138.

(15) Pederson, M. R.; Reuse, F.; Khanna, S. N. Magnetic transition in Mn_n ($n = 2 - 8$) clusters. *Phys. Rev. B* **1998**, *58*, 5632–5636.

(16) Yanagisawa, S.; Tsuneda, T.; Hirao, K. An investigation of density functionals: The first-row transition metal dimer calculations. *J. Chem. Phys.* **2000**, *112*, 545–553.

(17) Barborini, M. Neutral, Anionic, and Cationic Manganese Dimers through Density Functional Theory. *J. Phys. Chem. A* **2016**, *120*, 1716–1726.

(18) Sun, J.; Ruzsinszky, A.; Perdew, J. P. Strongly Constrained and Appropriately Normed Semilocal Density Functional. *Phys. Rev. Lett.* **2015**, *115*, 036402.

(19) Pantazis, D. Assessment of Double-Hybrid Density Functional Theory for Magnetic Exchange Coupling in Manganese Complexes. *Inorganics* **1999**, *7*, 57.

(20) Perdew, J. P.; Ruzsinszky, A.; Sun, J.; Nepal, N. K.; Kaplan, A. D. Interpretations of ground-state symmetry breaking and strong correlation in wavefunction and density functional theories. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, 1–6.

(21) Yamanaka, S.; Ukai, T.; Nakata, K.; Takeda, R.; Shoji, M.; Kawakami, T.; Takada, T.; Yamaguchi, K. Density functional study of manganese dimer. *Int. J. Quantum Chem.* **2007**, *107*, 3178–3190.

(22) Perdew, J. P.; Zunger, A. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B* **1981**, *23*, 5048–5079.

(23) Klüpfel, P.; Klüpfel, S.; Tsemekhman, K.; Jónsson, H. Optimization of functionals of orthonormal functions in the absence of unitary invariance. *Lecture Notes in Computer Science* **2012**, *7134*, 23.

(24) Jónsson, H. Simulation of surface processes. *Proc. Natl. Acad. Sci.* **2011**, *108*, 944–949.

(25) Klüpfel, S.; Klüpfel, P.; Jónsson, H. The effect of the Perdew-Zunger self-interaction correction to density functionals on the energetics of small molecules. *J. Chem. Phys.* **2012**, *137*, 124102.

(26) Gudmundsdóttir, H.; Jónsson, E. O.; Jónsson, H. Calculations of Al dopant in α-quartz using a variational implementation of the Perdew-Zunger self-interaction correction. *New J. Phys.* **2015**, *17*, 083006.

(27) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(28) Klüpfel, S.; Klüpfel, P.; Jónsson, H. Importance of complex orbitals in calculating the self-interaction-corrected ground state of atoms. *Phys. Rev. A* **2011**, *84*, 050501.

(29) Lehtola, S.; Jónsson, H. Unitary optimization of localized molecular orbitals. *J. Chem. Theo. Comput.* **2013**, *9*, 5365–5372.

(30) Lehtola, S.; Head-Gordon, M.; Jónsson, H. Complex orbitals, multiple local minima, and symmetry breaking in Perdew–Zunger self-interaction corrected density functional theory calculations. *J. Chem. Theo. Comput.* **2016**, *12*, 3195–3207.

(31) Enkovaara, J.; *et. al.*, Electronic structure calculations with GPAW: a real-space implementation of the projector augmented-wave method. *J. Phys. Condens. Matter* **2010**, *22*, 253202.

(32) Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **1994**, *50*, 17953–17979.

(33) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.

(34) Rappoport, D.; Furche, F. Property-optimized Gaussian basis sets for molecular response calculations. *J. Chem. Phys.* **2010**, *133*, 134105.

(35) Feller, D. The role of databases in support of computational chemistry calculations. *J. Comput. Chem.* **1996**, *17*, 1571–1586.

(36) Schuchardt, K. L.; Didier, B. T.; Elsethagen, T.; Sun, L.; Gurumoorthi, V.; Chase, J.; Li, J.; Windus, T. L. Basis Set Exchange: A Community Database for Computational Sciences. *J. Chem. Inf. Model.* **2007**, *47*, 1045–1052.

(37) Pritchard, B. P.; Altarawy, D.; Didier, B.; Gibson, T. D.; Windus, T. L. New Basis Set Exchange: An Open, Up-to-Date Resource for the Molecular Sciences Community. *J. Chem. Inf. Model.* **2019**, *59*, 4814–4820.

(38) Rossi, T.; Lehtola, S.; Sakko, A.; Puska, M.; Nieminen, R. *J. Chem. Phys.* **2015**, *142*, 094114.

(39) Jónsson, E. O.; Lehtola, S.; Puska, M.; Jónsson, H. Theory and Applications of Generalized Pipek-Mezey Wannier Functions. *J. Chem. Theo. Comput.* **2017**, *13*, 460–474.

(40) Ivanov, A. V.; Jónsson E.; Vegge, T.; Jónsson, H. Implementation of a Direct Minimisation Method using Exponential Transformation in the Localised Basis Set Approach. *(submitted, manuscript available at arXiv:2101.12597)* **2021**,

(41) Levi, G.; Ivanov, A. V.; Jónsson, H. Variational density functional calculations of excited states via direct optimization. *J. Chem. Theo. Comput.* **2020**, *16*, 6968.

(42) Wang, J.; Becke, A. D.; Smith Jr., V. H. Evaluation of $\langle S^2 \rangle$ in restricted, unrestricted Hartree–Fock, and density functional based theories. *J. Chem. Phys.* **1995**, *102*, 3477–3480.

(43) Kresse, G.; Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **1996**, *54*, 11169–11186.

(44) Pulkkinen, A.; Barbiellini, B.; Nokelainen, J.; Sokolovskiy, V.; Baigutlin, D.; Miroshkina, O.; Zagrebin, M.; Buchelnikov, V.; Lane, C.; Markiewicz, R. S.; Bansil, A.; Sun, J.; Pussi, K.; Lähderanta, E. Coulomb correlation in noncollinear antiferromagnetic α-Mn. *Phys. Rev. B* **2020**, *101*, 075115.

(45) Shahi, C. et al. Stretched or noded orbital densities and self-interaction correction in density functional theory. *J. Chem. Phys.* **2019**, *150*.

(46) Kim, M.-C.; Sim, E.; Burke, K. Understanding and Reducing Errors in Density Functional Calculations. *Phys. Rev. Lett.* **2013**, *111*, 73003.

(47) Perdew, J. P.; Ruzsinszky, A.; Sun, J.; Pederson, M. R. In *Chapter One - Paradox of Self-Interaction Correction: How Can Anything So Right Be So Wrong?*; Arimondo, E., Lin, C. C., Yelin, S. F., Eds.; Advances In Atomic, Molecular, and Optical Physics; Academic Press, 2015; Vol. 64; pp 1–14.