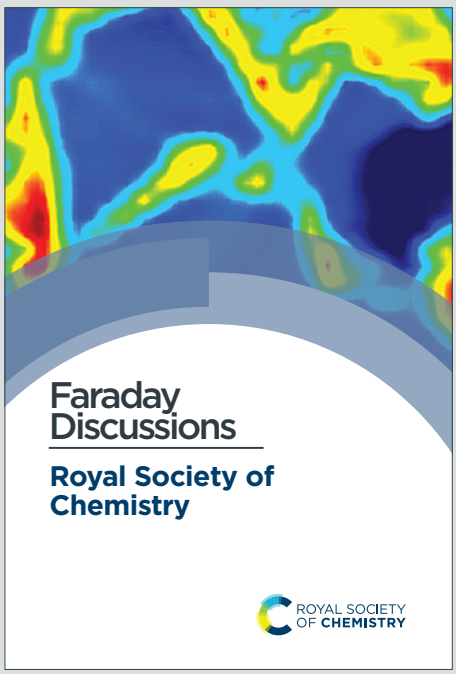
Faraday Discussions

Accepted Manuscript



This article can be cited before page numbers have been issued, to do this please use: G. Levi, A. V. Ivanov and H. Jonsson, *Faraday Discuss.*, 2020, DOI: 10.1039/D0FD00064G.

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/faraday_d

View Article Online
DOI: 10.1039/D0FD00064G

Journal Name

ARTICLE TYPE

Cite this: DOI: 00.0000/xxxxxxxxxx

Variational calculations of excited states via direct optimization of the orbitals in DFT†

Gianluca Levi,*[a] Aleksei V. Ivanov,[a,b] and Hannes Jónsson[a]

Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

A direct optimization method for obtaining excited electronic states using density functionals is presented. It involves selective convergence on saddle points on the energy surface representing the variation of the energy as a function of the electronic degrees of freedom, thereby avoiding convergence to a minimum and corresponding variational collapse to the ground electronic state. The method is based on an exponential transformation of the molecular orbitals, making it possible to use efficient quasi-Newton optimization approaches. Direct convergence on a target nth-order saddle point is guided by an appropriate preconditioner for the optimization as well as the maximum overlap method. Results of benchmark calculations of 52 excited states of molecules indicate that the method is more robust than a standard self-consistent field (SCF) approach especially when degenerate or quasi-degenerate orbitals are involved. The method can overcome challenges arising from rearrangement of closely spaced orbitals in a charge-transfer excitation of the nitrobenzene molecule, a case where the SCF fails to converge. The formulation of the method is general and can be applied to non-unitary invariant functionals, such as self-interaction corrected functionals.

1 Introduction

Accurate modelling of excited electronic states in molecules is essential for an understanding of fundamental photophysical and photochemical processes, such as energy transfer and photocatalytic reactions. Density functional calculations of excited-state properties of molecules are typically carried out using the linear response formulation of time-dependent density functional theory (TDDFT) [1–3], which describes the change of the ground-state electron density upon a weak external perturbation. Standard TDDFT calculations employ ground-state Kohn-Sham (KS) [4,5] exchange-correlation (xc) functionals, thus neglecting the time dependency of the xc kernel, i.e. the functional derivative of the xc potential with respect to the density, the so-called adiabatic approximation. Such TDDFT calculations with KS functionals can provide fairly accurate estimates of excitations to low-lying (valence) excited states [1,6] and do not require large computational effort.

However, TDDFT suffers from well-known deficiencies due to the use of linear response theory and the adiabatic approximation. For example, doubly excited states are not accessible because they require a computationally demanding time-dependent xc kernel [7–9]. Due to the lack of double excitations, the description of conical intersections between ground and excited states is qualitatively incorrect [7]. Furthermore, unless long-range corrected functionals are used, TDDFT shows systematic errors for excitations that involve displacement of electrons between orbitals of different spatial extent (such as excitations to Rydberg states) [6,10,11] or large spatial separation (such as charge-transfer states) [12–14]. There, orbital relaxation effects are important but are missing in practical implementations of TDDFT [15–17].

Alternatively, excited states can be obtained as optimized single Slater determinants corresponding to higher-energy stationary points of the energy surface that represents the variation of the energy of the system as a function of the electronic degrees of freedom. Such methods do not suffer from the same shortcomings as TDDFT and have, therefore, seen a resurgence of interest in recent years [10,11,16,18–33]. Some of these approaches enforce orthogonality of the excited-state solutions to the ground state [23,24,31,34,35], but we consider here the more straightforward version of excited-state DFT§ where a relaxed higher-energy solution is found by ensuring only orthogonality between the orbitals [11,32]. The energy of an excited state obtained in this way does not necessarily represent an upper bound to the energy of

[a] Science Institute and Faculty of Physical Sciences, University of Iceland, 107 Reykjavík, Iceland. [b] Saint Petersburg State University, 199034 Saint Petersburg, Russia. E-mail: giale@hi.is

† Electronic Supplementary Information (ESI) available: Character of the electronic transitions, excitation energies and number of iterations for the benchmark calculations on excited states of small molecules. See DOI: 00.0000/00000000.

§ This methodology is often referred to as ΔSCF in the literature, a terminology that refers to the fact that an excitation energy is computed as the difference between two separate self-consistent field (SCF) calculations. Since this method gives an estimate of other excited-state properties, not just excitation energy, and that stationary points of energy functionals can be obtained without using the SCF procedure, we rather use the more general expression "excited-state DFT".

the exact excited state[11,36,37]. Despite this, such higher-energy DFT solutions obtained with KS functionals have been shown to give useful approximations of excited states[21,32]. The approximation is expected to degrade for states that are intrinsically multideterminantal, but in many cases, such as for open-shell singlets, the excitation energy can be estimated with good accuracy using the Ziegler sum rules[38].

The lack of orthogonality implies that the optimization of a given higher-energy state is liable to variational collapse to a lower-energy solution of the same symmetry. Various techniques have been proposed to avoid variational collapse. The most commonly used approach involves combining the self-consistent field (SCF) procedure with a maximum overlap method (MOM)[11,21,32] that helps retain a given non-aufbau occupation of the orbitals. However, SCF-MOM, or any diagonalization-based optimization methodology, can still be prone to erratic convergence behavior in some cases, in particular when degenerate or near-degenerate orbitals are involved.

Methods based on direct optimization (DO) of the orbitals ensure more robust convergence than SCF[39], but they have not been explored much for excited states. This is because excited states correspond to saddle points on the energy surface[11]. Finding saddle points with an unconstrained optimization strategy, such as a quasi-Newton algorithm, is more difficult than minimization. In a recent implementation of DO for excited-state DFT the squared norm of the gradient is minimized[19]. This method can successfully deal with cases for which SCF-MOM fails, but there are also complications. First of all, the second derivatives of the energy are required in the minimization and this increases the computational effort. Secondly, quasi-Newton squared gradient minimization is not as well conditioned, usually requiring more iterations to converge[19,40]. Finally, such a procedure is not guaranteed to converge on a saddle point as the squared norm of the gradient can have local minima at points that are not stationary points on the energy surface[19]. Therefore, this approach requires a good enough initial guess as a starting point.

Thanks to the above mentioned developments, excited-state DFT using KS functionals has been applied to treat core[18] and double excitations[19,21], Rydberg[11,19] and charge-transfer[21,32,41,42] states, absorption spectra[43] and structural dynamics[22,44] in solution, achieving better results than TDDFT for these problems. However, excitation energies can still be affected by systematic errors. This is because, due to the typically more diffuse character of the excited-state orbitals, many excited states, especially those with Rydberg, charge-transfer and double excitation character, are more affected by the self-interaction error than ground states when using common semi-local KS functionals[15,45]. The self-interaction correction (SIC)[46] applied to KS functionals can alleviate some of these problems by correcting the long-range form of the effective potential[45,47]. However, performing fully variational calculations with SIC functionals is challenging because, contrary to KS functionals, these functionals are not invariant to a unitary transformation of the occupied orbitals[48,49]. As a consequence, implementations of fully variational calculations with SIC functionals for excited states are lacking.

In order to overcome some of the technical and methodological limitations of the excited-state DFT methods mentioned above, we require that a new, improved approach complies with the following:

i The method should avoid variational collapse and ensure robust convergence for a broad range of excited states and configurations;

ii The computational cost should be comparable with ground-state calculations;

iii The nonlinear variational procedure should allow for the use of both unitary (KS) and non-unitary invariant functionals (e.g. SIC functionals).

We pursue these objectives by combining a DO approach with the MOM method. The chosen DO approach makes it possible to use quasi-Newton optimization algorithms suitable for saddle points, while the MOM method ensures that at every iteration the occupied orbitals are consistent with the initial choice of occupation numbers. The Powell Hessian update or variations on it[50,51] is generally considered to give good performance in searches of saddle points for atomic rearrangements. Based on this, we develop a quasi-Newton limited-memory inverse Hessian update scheme using the Powell formula, inspired by the work of Anglada *et al.*[52], and compare its performance with respect to the Limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm, which is a robust quasi-Newton method most commonly employed for minimization.

Using a benchmark set of 52 excited states of molecules, we find that the DO-MOM method can converge on saddle points of any order if supplied with a preconditioner of the appropriate form. DO-MOM is found to be more robust than the SCF-MOM method in that there are fewer convergence failures and fewer iterations are needed, on average, corresponding to less computational effort. We also demonstrate successful application of the DO-MOM method in a particularly challenging case: an excited state of the nitrobenzene molecule, for which failure of SCF-MOM is well known[53]. The DO-MOM method can be used not only with unitary invariant KS functionals but also with non-unitary invariant functionals, such as SIC functionals.

2 Theory

In KS DFT[4,5], the ground-state energy functional for a spin-polarized N-electron system is:

$$E_{\mathrm{KS}}\left[n_\uparrow,n_\downarrow\right] = T_{\mathrm{s}}\left[n_\uparrow,n_\downarrow\right] + \int \mathfrak{v}_{\mathrm{ext}}(\mathbf{r})n(\mathbf{r})d\mathbf{r} + \frac{1}{2}\int\int \frac{n(\mathbf{r})n(\mathbf{r}')}{\mid \mathbf{r}-\mathbf{r}' \mid} d\mathbf{r}d\mathbf{r}' + E_{\mathrm{xc}}\left[n_\uparrow,n_\downarrow\right] \quad (1)$$

where $n(\mathbf{r}) = n_\uparrow(\mathbf{r}) + n_\downarrow(\mathbf{r})$ is the total electron density, and the spin densities $n_\sigma(\mathbf{r})$ and kinetic energy $T_{\mathrm{s}}\left[n_\uparrow,n_\downarrow\right]$ are given in terms of KS orbitals:

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

$$T_s\left[n_\uparrow, n_\downarrow\right] = -\frac{1}{2}\sum_{n\sigma} f_{n\sigma}\int \psi^*_{n\sigma}(\mathbf{r})\nabla^2\psi_{n\sigma}(\mathbf{r})d\mathbf{r} \tag{3}$$

with $0 \leq f_{n\sigma} \leq 1$ being the occupation numbers. The other terms on the right-hand side of eq. 1 represent, respectively, the external potential energy, the Coulomb repulsion between the electrons and the exchange-correlation (xc) functional.

The ground state corresponds to a minimum on the energy surface defined by the variation of the energy of the system with respect to the electronic degrees of freedom according to eq. 1; while saddle points represent excited electronic states[11]. To find such stationary points of the energy functional, eq. 1 is extremized subject to $\int \psi^*_{n\sigma}(\mathbf{r})\psi_{m\sigma'}(\mathbf{r})d\mathbf{r} = \delta_{nm}\delta_{\sigma\sigma'}$, which ensures orthonormality of the orbitals. Using Lagrange multipliers λ^σ_{nm} for the constraints leads to a set of nonlinear coupled equations:

$$\mathbf{H}^\sigma_{\mathrm{KS}}\psi_{n\sigma} = \sum_m \lambda^\sigma_{nm}\psi_{m\sigma} \tag{4}$$

where $\mathbf{H}^\sigma_{\mathrm{KS}}$ is the one-particle KS Hamiltonian:

$$\mathbf{H}^\sigma_{\mathrm{KS}} = -\frac{1}{2}\nabla^2 + \boldsymbol{\upsilon}_{\mathrm{ext}}(\mathbf{r}) + \int \frac{n(\mathbf{r}')}{|\,\mathbf{r}-\mathbf{r}'\,|}d\mathbf{r}' + \boldsymbol{\upsilon}^\sigma_{\mathrm{xc}}(\mathbf{r}) \tag{5}$$

The energy defined by a KS functional of the form of eq. 1 is invariant to a unitary transformation among the equally occupied orbitals. Exploiting the unitary invariance of the functional, the orbitals that make the energy stationary can be chosen as the canonical orbitals obtained from solving the generalized KS eigenvalue equations:

$$\mathbf{H}^\sigma_{\mathrm{KS}}\psi_{n\sigma}(\mathbf{r}) = \varepsilon_{n\sigma}\psi_{n\sigma}(\mathbf{r}) \tag{6}$$

Iteratively solving the KS equations defines the SCF procedure. The choice of occupation numbers in the SCF determines whether the final solution is a minimum (ground state) or a saddle point (excited state). The ground state is obtained if this choice is made according to the aufbau principle, while excited states are obtained for non-aufbau occupations. For excited states, the problem is how to enforce non-aufbau occupations. The MOM method is commonly used for this purpose. There, one selects the occupied orbitals to be those with the biggest overlap with the occupied orbitals of the previous SCF step[32] or with the initial occupied orbitals[11,21] (also known as the initial maximum overlap method (IMOM)[21]).

An alternative formulation of the variational problem in DFT makes use of the exponential transformation of a set of orthonormal reference orbitals to find the optimal orbitals that extremize the energy functional[39,54–56]:

$$\phi_{p\sigma}(\mathbf{r}) = \sum_q \left[e^{\boldsymbol{\theta}}\right]^\sigma_{pq}\psi_{q\sigma}(\mathbf{r}) \tag{7}$$

where $\boldsymbol{\theta}$ is an anti-Hermitian matrix ($\boldsymbol{\theta} = -\boldsymbol{\theta}^\dagger$). The DO with exponential transformation is more general than SCF, since it can be applied to both unitary and non-unitary invariant functionals, such as SIC functionals[46]. Furthermore, the anti-Hermitian matrices form a linear space making it possible to employ quasi-Newton unconstrained optimization strategies[57], with the advantage that gradient-based optimization guarantees more rigorous convergence than SCF[39,58]. Quasi-Newton methods can locate nth-order saddle points of a multidimensional surface if the initial guess is sufficiently good and the formula chosen to compute the search direction is able to build an approximation to the Hessian with the appropriate number of negative eigenvalues. The development of quasi-Newton methods for saddle points has reached a fairly advanced level in the context of searches of transition states of atomic rearrangements[51,59–63]. However, these methodologies have not been explored much for DO of saddle points on electronic energy surfaces, due to the danger of variational collapse to lower solutions. Here, we propose a DO approach to saddle points of electron energy surfaces which relies on a two-fold mechanism to ensure that the search direction follows the target solution throughout the optimization: (1) a preconditioner of the appropriate form is chosen and updated regularly to generate an approximation to the Hessian where the number of negative eigenvalues is consistent with the target nth-order saddle point, and (2) MOM constraints are applied to preserve the character of the occupied orbitals.

3 Implementation

3.1 Exponential Transformation

The exponential transformation is implemented using the localized atomic orbital basis set approach as described by Ivanov *et al.*[64] for energy minimization. We review the implementation here for completeness.

The exponential transformation is applied to each set of spin orbitals separately and, therefore, the spin index will be omitted for simplicity. An initial set of reference orbitals consists of M guess orbitals expressed as linear combinations of M (non-orthogonal) atomic basis functions:

$$\phi_p(\mathbf{r}) = \sum_{\mu=1}^{M} C_{\mu p}\chi_\mu(\mathbf{r}) \tag{8}$$

and satisfying the orthonormality condition:

$$\sum_{\mu\nu} C^*_{\mu p}S_{\mu\nu}C_{\nu q} = \delta_{pq} \tag{9}$$

where:

$$S_{\mu\nu} = \int \chi^*_\mu(\mathbf{r})\chi_\nu(\mathbf{r})d\mathbf{r} \tag{10}$$

The coefficients of the orbitals that extremize the energy functional (optimal orbitals) are found through a unitary exponential transformation of the $C_{\mu p}$ coefficients:

$$O_{\mu k} = \sum_{p=1}^{M} C_{\mu p}\left[e^{\boldsymbol{\theta}}\right]_{pk} \tag{11}$$

where the $M \times M$ anti-Hermitian matrix $\boldsymbol{\theta}$ is parametrized as follows:

$$\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\theta}_{\mathrm{oo}} & \boldsymbol{\theta}_{\mathrm{ov}} \\ -\boldsymbol{\theta}^\dagger_{\mathrm{ov}} & \mathbf{0} \end{pmatrix} \tag{12}$$

In eq. 12, the occupied-occupied (oo) block is an $N \times N$ anti-Hermitian matrix describing rotations among the occupied orbitals, while the occupied-virtual (ov) blocks have dimension $N \times (M-N)$ and describe rotations that mix occupied and unoccupied orbitals. Since the functional never depends on unoccupied orbitals, rotations among them do not change the energy and, therefore, the virtual-virtual (vv) block can be set to zero without loss of generality[56]. Due to the anti-Hermiticity of $\boldsymbol{\theta}$, the energy depends only on the upper-triagonal and diagonal elements. $\boldsymbol{\theta}_{\mathrm{oo}}$ contains N^2 variables, $N(N-1)/2$ real and $N(N+1)/2$ imaginary parts (the real part of the diagonal elements is zero), while $\boldsymbol{\theta}_{\mathrm{ov}}$ contains $2N(M-N)$ variables, $N(M-N)$ real and imaginary parts. Therefore, the total number of independent variational parameters is $N(2M-N)$. To evaluate the matrix exponential $e^{\boldsymbol{\theta}}$ we use the scaling and squaring algorithm of Al-Mohy and Higham[65] as implemented in the SciPy library[66]. For unitary invariant functionals the $\boldsymbol{\theta}_{\mathrm{oo}}$ block can be set to zero[56]. In this case the number of degrees of freedom is $2N(M-N)$ and the formulas for calculating the matrix exponential given by Hutter *et al.*[56] can be used, with a computational cost that scales as $\mathscr{O}(N^2M)$.

Stationary solutions are obtained when the gradient of the energy with respect to the real and imaginary parts of the $\{\theta_{ij}\}$ $i=0,\ldots,N$, $j=i,\ldots,M$ rotational variables is zero. By noting that the derivative of a function of complex variables can be written as:

$$\frac{\partial E}{\partial \theta_{ij}} = \frac{1}{2}\left(\frac{\partial E}{\partial \mathrm{Re}(\theta_{ij})} - i\frac{\partial E}{\partial \mathrm{Im}(\theta_{ij})}\right) \tag{13}$$

the stationarity conditions are:

$$\frac{\partial E}{\partial \theta_{ij}} = \frac{\partial E}{\partial \theta_{ij}^*} = 0 \tag{14}$$

The gradient of the energy with respect to the $\{\theta_{ij}\}$ can be computed using the chain rule and the definition of the derivative of a matrix exponential:

$$\frac{\partial E}{\partial \theta_{ij}} = \frac{2-\delta_{ij}}{2}\left[\int_0^1 e^{t\boldsymbol{\theta}}\mathbf{L}e^{-t\boldsymbol{\theta}}dt\right]_{ji} \tag{15}$$

where $\mathbf{L}$ is the following commutator:

$$L_{lk} = [\mathbf{F},\mathbf{H}]_{lk} \tag{16}$$

i.e. the commutator of the diagonal matrix $\mathbf{F}$ of the occupation numbers f_l and the Hamiltonian matrix expressed in the optimal orbitals basis:

$$H_{lk} = \sum_{\mu\nu} O^*_{\mu l} H_{\mu\nu} O_{\nu k}, \qquad H_{\mu\nu} = \int \chi^*_\mu(\mathbf{r})\mathbf{H}_{\mathrm{KS}}\chi_\nu(\mathbf{r})d\mathbf{r} \tag{17}$$

When the generator of a rotation is small ($\|\boldsymbol{\theta}\| \ll 1$), the integral in eq. 15 can be expended in a series:

$$\int_0^1 e^{t\boldsymbol{\theta}}\mathbf{L}e^{-t\boldsymbol{\theta}}dt = \mathbf{L} + \frac{1}{2!}[\boldsymbol{\theta},\mathbf{L}] + \frac{1}{3!}[\boldsymbol{\theta},[\boldsymbol{\theta},\mathbf{L}]] + \ldots \tag{18}$$

and only the first term on the right-hand side of eq. 18 used to efficiently estimate the gradient. During the iterative DO, in order to avoid the norm of $\boldsymbol{\theta}$ becoming too large, approaching 1, the reference orbital coefficients are updated with the coefficients of the optimal or canonical orbitals and $\boldsymbol{\theta}$ reset to zero every Nth iterations. Alternatively, one can calculate the gradient from eq. 15 exactly using methods based on the eigendecompostion of the matrix $\boldsymbol{\theta}$ or scaling and squaring algorithms for computing exponential integrals[67].

3.2 Quasi-Newton Step

We have implemented a quasi-Newton step based on updating an approximation to the inverse of the Hessian matrix:

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \mathbf{B}^{(k)}\mathbf{g}^{(k)} \tag{19}$$

where $\mathbf{B}^{(k)}$ is the approximation to the inverse Hessian at step k, and $\mathbf{x}^{(k)}$ and $\mathbf{g}^{(k)}$ are the vector of elements of the anti-Hermitian matrix $\boldsymbol{\theta}$ and analytical gradient, respectively. To avoid too large steps at the beginning of the optimization, we prevent the step from exceeding a maximum allowed step length, $p_{\max}$. We have found that, in most cases, a $p_{\max}$ of 0.20 ensures stable convergence without significantly affecting the rate of convergence.

In order to allow calculations of large systems, we adopt a limited-memory approach to compute the search direction, which uses only vectors and scalars from the m most recent steps to update $\mathbf{B}^{(k)}$. Currently, the most widely used limited-memory quasi-Newton method is L-BFGS. But, it is mainly used for minimization of a function. In order to asses its performance in the search of saddle points on DFT energy surfaces, we have implemented L-BFGS following reference[57]. In optimizations of atomic structures, other update formulas are preferred over BFGS for saddle point searches, because BFGS has a tendency to build approximate Hessians that are positive definite. One of these alternative formulas that has proven most successful for saddle-point optimization is the Powell Hessian update[50,59,61]. To our knowledge, Powell inverse Hessian update has not been formulated in a limited-memory version yet. We present here a limited-memory algorithm that uses the Powell formula to update an approximate inverse Hessian. The formulation of the inverse Hessian update draws upon the approach used by Anglada *et al.*[52] for a limited memory update of the Hessian. The Powell inverse Hessian update formula can be written as[68]:

$$\mathbf{B}_{\mathrm{P}}^{(k+1)} = \mathbf{j}^{(k)}\mathbf{u}^{T(k)} + \mathbf{u}^{(k)}\left[\mathbf{j}^{T(k)} - \left(\mathbf{y}^{T(k)}\mathbf{j}^{(k)}\right)\mathbf{u}^{T(k)}\right] \tag{20}$$

with:

$$\mathbf{j}^{(k)} = \mathbf{s}^{(k)} - \mathbf{B}^{(k)}\mathbf{y}^{(k)}, \quad \mathbf{u}^{(k)} = \frac{\mathbf{y}^{(k)}}{\mathbf{y}^{T(k)}\mathbf{y}^{(k)}} \tag{21}$$

and:

$$\mathbf{s}^{(k)} = \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}, \quad \mathbf{y}^{(k)} = \mathbf{g}^{(k+1)} - \mathbf{g}^{(k)} \tag{22}$$

To formulate the limited-memory Powell update of the inverse Hessian, we make use of the following recursive formula for com-

puting the product $\mathbf{B}_{\mathrm{P}}^{(k)}\mathbf{v}^{(k)}$ between $\mathbf{B}_{\mathrm{P}}^{(k)}$ and a vector $\mathbf{v}^{(k)}$:

$$\mathbf{B}_{\mathrm{P}}^{(k)}\mathbf{v}^{(k)} = \mathbf{B}_0^{(k)}\mathbf{v}^{(k)} + \sum_{i=k-m}^{k-1} \mathbf{j}^{(i)}\mathbf{u}^{T(i)}\mathbf{v}^{(k)} + \sum_{i=k-m}^{k-1} \left\{ \mathbf{u}^{(i)} \left[\mathbf{j}^{T(i)}\mathbf{v}^{(k)} - \left(\mathbf{y}^{T(i)}\mathbf{j}^{(i)} \right) \mathbf{u}^{T(i)}\mathbf{v}^{(k)} \right] \right\} \quad (23)$$

Using eq. 23, the quasi-Newton algorithm with limited-memory Powell inverse Hessian update is formulated as 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. 23;
  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 pair $\{\mathbf{j}^{(k-m)}, \mathbf{u}^{(k-m)}\}$ and scalar $r^{(k-m)}$;
  end
  $\mathbf{s}^{(k)} \leftarrow \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}$ and $\mathbf{y}^{(k)} \leftarrow \mathbf{g}^{(k+1)} - \mathbf{g}^{(k)}$;
  $\mathbf{u}^{(k)} \leftarrow \frac{\mathbf{y}^{(k)}}{\mathbf{y}^{T(k)}\mathbf{y}^{(k)}}$;
  Compute $\mathbf{j}^{(k)} \leftarrow \mathbf{B}^{(k)}\mathbf{y}^{(k)}$ using eq. 23;
  $\mathbf{j}^{(k)} \leftarrow \mathbf{s}^{(k)} - \mathbf{j}^{(k)}$;
  $r^{(k)} \leftarrow \mathbf{y}^{T(k)}\mathbf{j}^{(k)}$;
  Store vector pair $\{\mathbf{j}^{(k)}, \mathbf{u}^{(k)}\}$ and scalar $r^{(k)}$;
  $k \leftarrow k+1$;
end

Algorithm 1: Limited-memory quasi-Newton algorithm based on the Powell inverse Hessian update. If $\mathbf{B}_0^{(k)}$ is chosen as a diagonal matrix, the computational cost scales linearly with the dimensionality n of the optimization problem.

3.3 Choice of Preconditioner

The inverse of an approximate Hessian ($\mathbf{B}_0^{(k)}$, according to the terminology in Algorithm 1) represents a preconditioner of the quasi-Newton step, which can be allowed to vary during the optimization. In the basis of canonical orbitals the approximate Hessian is diagonal and proportional to the difference in eigenvalues (ε_i)[55]:

$$\frac{\partial^2 E}{\partial^2 \theta_{ij}} \approx -2(\varepsilon_i - \varepsilon_j)(f_i - f_j) \quad (24)$$

Therefore, the preconditioner for the optimization problem can be written as:

$$\mathbf{B}_0^{(k)} = \frac{1}{-2(\varepsilon_i^{(k)} - \varepsilon_j^{(k)})(f_i^{(k)} - f_j^{(k)})} \quad (25)$$

Previous experience has shown that this form of the preconditioner can accelerate the convergence of a ground-[39] and excited-state[19] optimization (the latter through minimization of the squared norm of the gradient) in Hartree-Fock (HF) and DFT calculations with BFGS update.

In the present calculations the preconditioner of eq. 25 is used for saddle-point searches using DO. Since eq. 25 is not defined for oo terms and for degenerate ov pairs, the preconditioner is not used in these cases, i.e. the corresponding elements of $\mathbf{B}_0^{(k)}$ are set to 1. To ensure convergence on the target nth-order saddle-point, it is important that the approximate Hessian has n negative eigenvalues. When the initial guess orbitals are generated by promoting an electron from an occupied to a virtual orbital of the ground state, the corresponding approximate diagonal Hessian has one negative element for every ov pair for which the occupied orbital lies above in energy with respect to the virtual orbital. In some cases, due to relaxation effects, which cause changes in the energy ordering of the orbitls, the final state can correspond to a saddle point of different order than the one predicted by the initial approximate Hessian. In these cases, updating $\mathbf{B}_0^{(k)}$ during the optimization can be important to converge on the target saddle point, as will be demonstrated later for an excited state of nitrobenzene. We have found that, in most cases, it is sufficient to update $\mathbf{B}_0^{(k)}$ together with the DO reference orbitals every Nth iterations, where N is a predefined number, which avoids the costly diagonalization of the Hamiltonian matrix at each step. For difficult cases, updating the preconditioner after the MOM method has detected a change of the character of the orbitals with respect to the initial guess might improve the convergence, but we have found that it is in general not required.

3.4 Maximum Overlap Method

The MOM method is applied at each step of the DO and when choosing the canonical orbitals to update the preconditioner according to eq. 25. At each step k, given the overlap matrix:

$$\Omega_{pl}^{(k)} = \sum_{\nu\mu} C_{p\mu}^{*} S_{\mu\nu} O_{\nu l}^{(k)} \quad (26)$$

with $S_{\mu\nu}$ defined in eq. 10, the optimal orbitals that have the largest projections $\omega_l^{(k)}$ into the occupied subspace of the initial reference orbitals get occupied:

$$\omega_l^{(k)} = \left[\sum_{p=1}^{N} \left(\Omega_{pl}^{(k)} \right)^2 \right]^{\frac{1}{2}} \quad (27)$$

Note that the reference orbitals for the MOM method are chosen as the guess orbitals, which coincide with the initial reference orbitals of the DO; but while the DO reference orbitals are updated during the iterative procedure, the MOM reference is fixed. Similar expressions are used to find the occupation numbers of the canonical orbitals when updating the preconditioner.

4 Computational Methods

The DO-MOM method has been implemented in a developmental version of the Grid-based Projector Augmented Wave (GPAW)

program[69–71] that is used for all the calculations presented here. GPAW is based on the PAW approach[72] for describing the regions close to the nuclei and a frozen-core approximation. Uncontracted functions are removed from all basis sets of Gaussian-type orbitals used in the present work. A grid spacing of 0.15 Å is chosen and the simulations box made large enough to avoid any effect due to truncation of the basis functions. The spin-unrestricted formalism is used and each (spin-mixed) excited state is described by a single-electron transition within the same spin channel. All the calculations use the PBE generalized gradient approximation to the energy functional.[73] A calculation is considered to be converged when the integrated value of the square of the residuals of the KS equations is smaller than 10^{-10} eV^2 per valence electron. Each calculation has a maximum of 300 iterations.

Convergence tests are carried out for 52 excited states of 18 molecules with atomic coordinates taken from the benchmark set of Loos *et al.*[74]. Frozen-geometry calculations are performed with DO-MOM using either L-BFGS or the limited-memory Powell inverse Hessian formulation developed in the present work, as well as calculations based on standard direct diagonalization of the Hamiltonian matrix[70] together with the MOM method with fixed reference orbitals[21] (SCF-MOM). The initial guess for each excited state is obtained by promoting an electron from an occupied to a virtual orbital obtained from a ground-state SCF calculation. The SCF-MOM calculations use GPAW default parameters for density mixing, while in the DO-MOM calculations the update of the reference orbitals used in eq. 7 occurs every 20 steps. The memory m for the limited-memory computation of the quasi-Newton search direction is chosen as 20, while the maximum step length, $p_{\max}$, is 0.20. The Dunning *aug*-cc-pVDZ basis set is used in these calculations[75–77].

Calculations are also carried out for nitrobenzene in the C_{2v} geometry taken from reference[53]. The initial guess used for converging to the $A_1(\pi' \rightarrow \pi^*)$ excited state is generated using the converged orbitals from an SCF calculation of the ground state by promoting an electron from the second highest occupied π orbital (π' according to the notation used in reference[53]) to the lowest unoccupied π^* orbital. The calculations make use of the def2-TZVP[78] basis set, as in reference[53].

5 Results

5.0.1 Benchmarks on Small Molecules

A summary of the resuls of the 52 convergence tests is given in Figures 1 and 2. The SCF-MOM does not converge within the maximum allowed number of iterations of 300 for 16 out of the total of 52 states. DO-MOM with L-BFGS is able to converge in all cases except for one, the $1\Delta(\pi \rightarrow \pi^*)$ state of carbon monoxide, and it requires on average ∼9 fewer iterations than SCF-MOM. DO-MOM with limited-memory Powell inverse Hessian update is not as robust, failing to reach convergence for 10 excited states, all cases where SCF-MOM also does not converge.

The performance of DO-MOM with L-BFGS in the cases for which SCF-MOM fails is summarized in Figure 3. A comparison with Figure 2 shows that a larger number of iterations is on average needed in these cases. For most of them, convergence difficulties arise because excitation occurs from or to degenerate or quasi-degenerate orbitals. The extreme situation is represented by the Δ states of carbon monoxide and dinitrogen since both the donor (π) and acceptor (π^*) orbitals belong to a degenerate pair. In these cases, the challenges faced by SCF-based techniques are well known: since the single-determinant solutions obtained by occupying either of the two orbitals of a degenerate pair have the same energy, the character of these orbitals can undergo large changes from one iteration to another, leading to convergence failure. Tuning modifications, including level shifting, damping and smearing can enforce convergence of the SCF when degenerate orbitals are unequally occupied[79]. However, they often require a careful choice of parameters to ensure converge to the target state without introducing artefacts[80]. On the other hand, the DO-MOM approach can converge these difficult cases more easily, without the need for tuning modifications. Previously, Van Voorhis and Head-Gordon[39] have shown the advantage of a gradient-based unconstrained optimization in ground-state calculations where the HOMO-LUMO gap is small, using a geometric approach to direct minimization. The present results demonstrate the advantage of DO in finding excited states as saddle points on the energy surface.

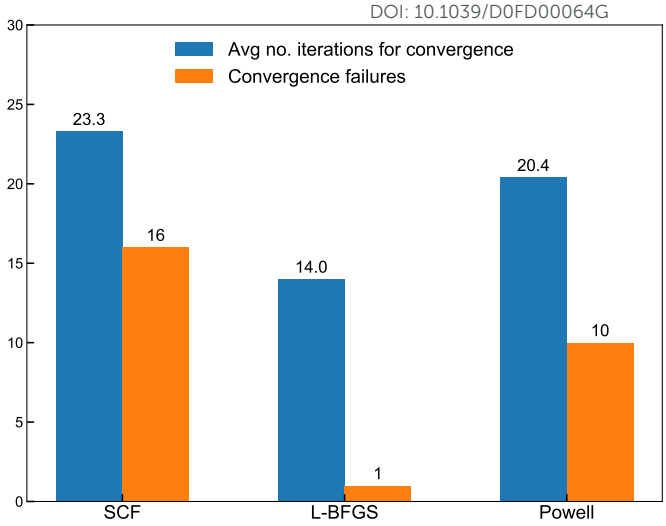


Fig. 1 Average number of iterations needed for convergence in cases where all three methods converge, and the number of failed calculations. One iteration corresponds to an energy and gradient evaluation for the DO-MOM methods, and to a Hamiltonian diagonalization and energy evaluation for the SCF-MOM method.

L-BFGS is known to be an efficient quasi-Newton method for minimization, but it is not as well suited for calcultions of saddle points[81]. Hence, it might appear surprising that DO-MOM performs better with L-BFGS than the limited-memory Powell inverse Hessian update. This is in part because the systems considered here include several highly symmetric molecules where the gradient along directions of negative curvature is often zero by construction of the initial guess as the respective ov pairs are made of orbitals belonging to different irreducible representations of point group symmetry. For the excited states represented by a first-order saddle point, except for the $1A_1$ state of ammonia, the

View Article Online
DOI: 10.1039/D0FD00064G

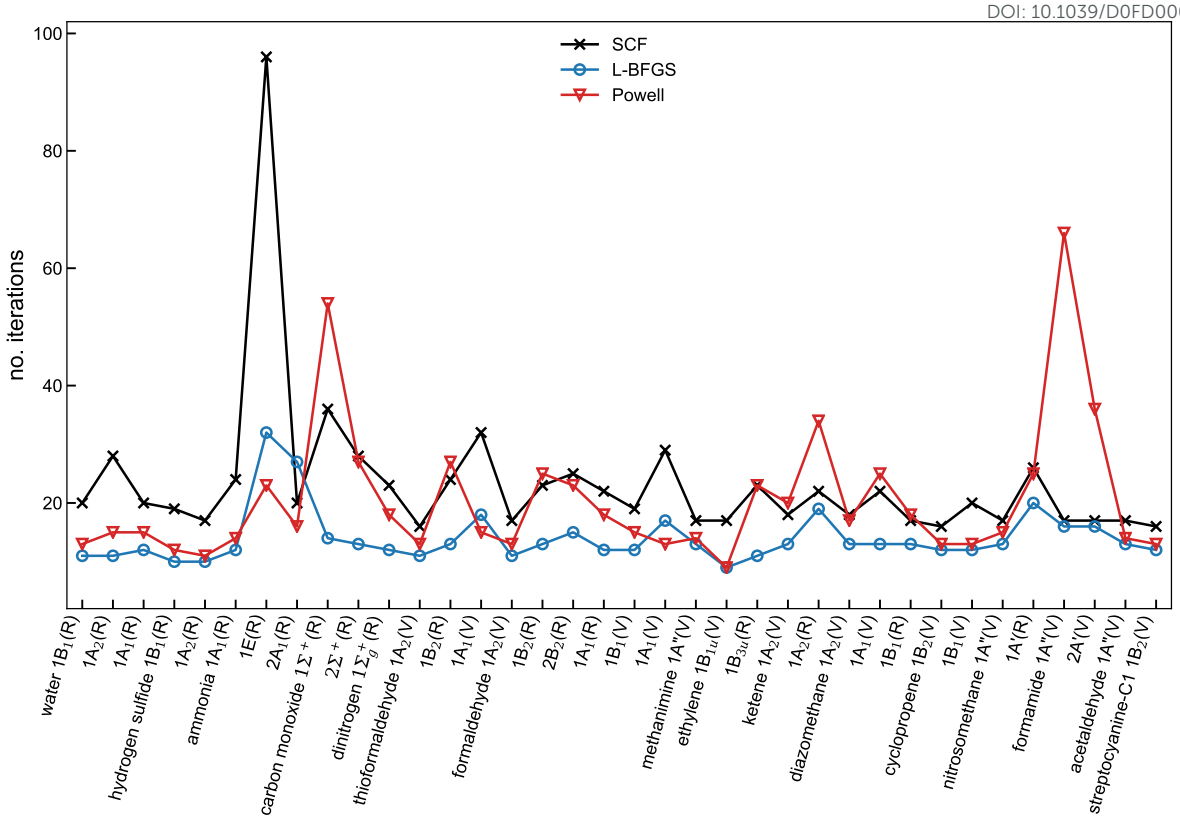


Fig. 2 Number of iterations required in the calculations of the excited states in the convergence test set where all three methods lead to convergence. The averages are given in Figure 1. Labeling: V=Valence, R=Rydberg, CT=Charge-transfer (see the ESI† for details on the character of each state).

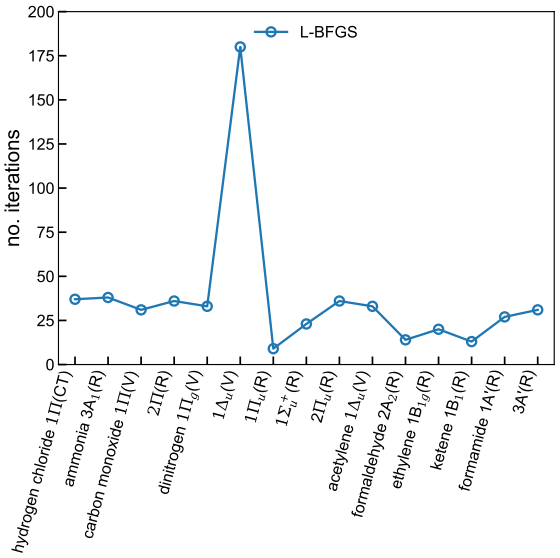


Fig. 3 Number of iterations required in the calculations where DO-MOM with L-BFGS converged on the excited states in the benchmark set but SCF-MOM did not converge.

unique gradient along a negative curvature direction is zero and, therefore, the excited-state optimization corresponds to a minimization. On the other hand, for more complex cases, where the orbitals associated with directions of negative curvature are allowed to mix and orbital energy levels change during the relaxation, the ability of L-BFGS to converge on the target saddle point stems from the use of the MOM constraints together with the updates of the precondtioner. This is illustrated in the following section.

For the molecules considered here, DO-MOM and SCF-MOM have similar computational effort per iteration because neither the arithmetic operations involved in the quasi-Newton step nor the diagonalization of the Hamiltonian are a bottleneck. The scaling and squaring algorithm employed here to compute the matrix exponential in DO-MOM calculations can outperform matrix diagonalization for sparse matrices; while the cost of the limited-memory quasi-Newton update scales linearly as $\mathcal{O}(mn)$, where m is the memory and n is the dimensionality of the problem. Therefore, if the memory can be kept low without affecting the convergence properties of the algorithm, the computational effort in each iteration of DO-MOM can be less than that for SCF-MOM where a matrix diagonalization needs to be carried out.

5.0.2 Nitrobenzene

The $A_1(\pi' \rightarrow \pi^*)$ excited state of nitrobenzene has an important role in the photophysics of this molecule since it is the lowest excited state with large oscillator strength[53,82]. It has charge-transfer character because it involves the transfer of an electron from the benzene ring to the nitro group. Mewes *et al.*[53] have reported failure of the SCF-MOM method to converge to this state when the occupation numbers at each SCF iteration are updated

View Article Online
DOI: 10.1039/D0FD00064G

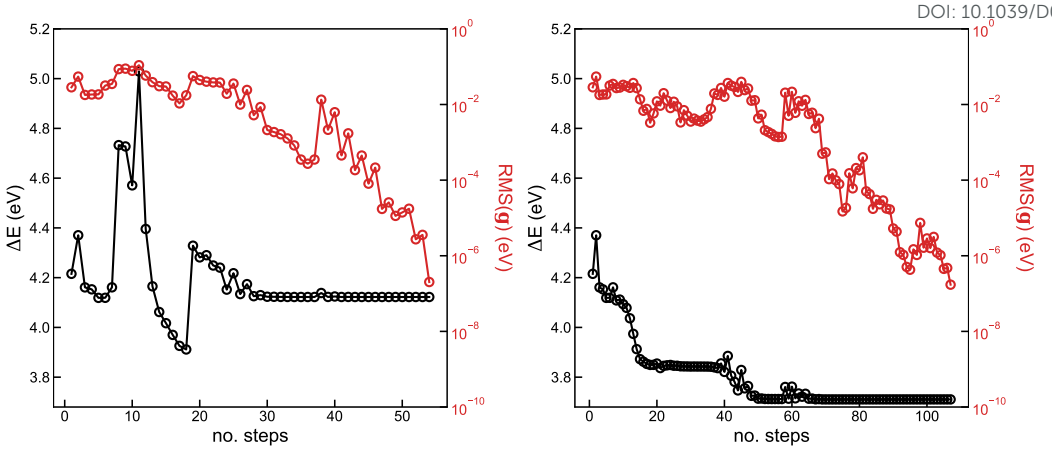


Fig. 4 (Left) Energy difference with respect to the ground state and root mean square of the gradient in a DO-MOM calculation converging to the $A_1(\pi' \rightarrow \pi^*)$ excited state of nitrobenzene. (Right) Energy difference and root mean square of the gradient in a DO calculation where MOM is not used and the preconditioner is fixed according to the initial approximation to the inverse Hessian.

according to the overlaps with the orbitals of the previous iteration. Such procedure leads to gradual drifting from the initial guess and collapse to the ground state because the hole and excited orbital are allowed by symmetry to mix. The SCF-MOM procedure employed here avoids gradual drifting to the ground state since the overlaps are determined with respect to a fixed set of orbitals; however, there is oscillatory behavior and ultimately failure to converge on the $A_1(\pi' \rightarrow \pi^*)$ state. Recently, Hait *et al.*[19] have observed a similar erratic convergence behavior of the SCF-MOM with fixed reference orbitals when applied to another excited state of nitrobenzene with A_1 symmetry. On the other hand, DO-MOM with L-BFGS is able to converge the energy of the $A_1(\pi' \rightarrow \pi^*)$ state to 10^{-8} eV within ∼50 iterations, as indicated in Figure 4.

To shed light on the failure of SCF-MOM and the success of DO-MOM, Figure 5 shows the orbitals of the initial guess constructed from the converged ground state and of the converged $A_1(\pi' \rightarrow \pi^*)$ state. First of all, we note that the five highest-energy occupied orbitals of the ground state (from the n_π to the n orbital) are closely spaced, spanning an energy range of ∼1 eV. Then, we observe that, as a result of the charge transfer, the three highest n orbitals localized at the nitro group are destabilized in the relaxed $A_1(\pi' \rightarrow \pi^*)$ state compared to the initial guess, while the π orbitals are stabilized. The presence of near degeneracies between the hole and the occupied orbitals, and the significant rearrangement in the energy levels of the orbitals contribute to the difficulties encountered by SCF-MOM in converging to the $A_1(\pi' \rightarrow \pi^*)$ state.

A challenge for DO-MOM arises from the fact that while the initial approximation to the Hessian has three negative eigenvalues (one for each of the three ov pairs where v is the hole and o is the π, n or π^* orbital), the relaxed $A_1(\pi' \rightarrow \pi^*)$ state corresponds to a fourth-order saddle point (see Figure 5). The approximation to the Hessian obtained from the initial guess has fewer negative eigenvalues as a consequence of the rearrangement in the ordering of the orbitals from the initial (based on ground state) to the relaxed excited state. The performance of a quasi-Newton optimization can be poor if the preconditioner is not good enough. Figure 4 shows the convergence of a DO calculation where the preconditioner is not updated and the MOM is not applied during the optimization. Then the character of the hole changes from the original π' orbital to a mix of π and n orbitals, and the optimization converges to a third-order saddle point, which is consistent with the less than optimal preconditioner. From Figure 4 we also see how DO-MOM can overcome the risk of converging to a lower saddle point of the wrong order. In the first twenty iterations, there are two cases where the MOM detects a change in the character of the hole and promotes a flip of the orbital occupations that forces the hole to have a character consistent with the initial guess. These events are indicated by jumps in the energy in Figure 4. Following the application of the MOM constraints, the optimization approaches a region of the electron configuration space where an approximation to the Hessian has the structure of the target solution. By updating the preconditioner using the canonical orbitals with occupation numbers determined through the MOM, L-BFGS is guided to ultimately converge on a fourth-order saddle point consistent with the $A_1(\pi' \rightarrow \pi^*)$ state. Finally, we note that for this system updating the preconditioner (together with the reference orbitals) after DO-MOM has detected a change of the character of the occupied orbitals is not necessary but speeds up the convergence.

6 Concluding Remarks

Despite the fact that DO has long been known to be an alternative to SCF for energy minimization[39,83,84], DO calculations of saddle points have so far been considered to be too difficult, due to the risk of variational collapse. Here, a DO method for converging to single-determinant DFT excited states is presented. The approach uses an updated preconditioner that guides convergence on the desired saddle point as well as the MOM method to avoid variational collapse. Since only one gradient calculation is performed at each iteration, there is no added computational cost with respect to DO ground-state calculations. This is an advantage over squared gradient minimization approaches, which re-

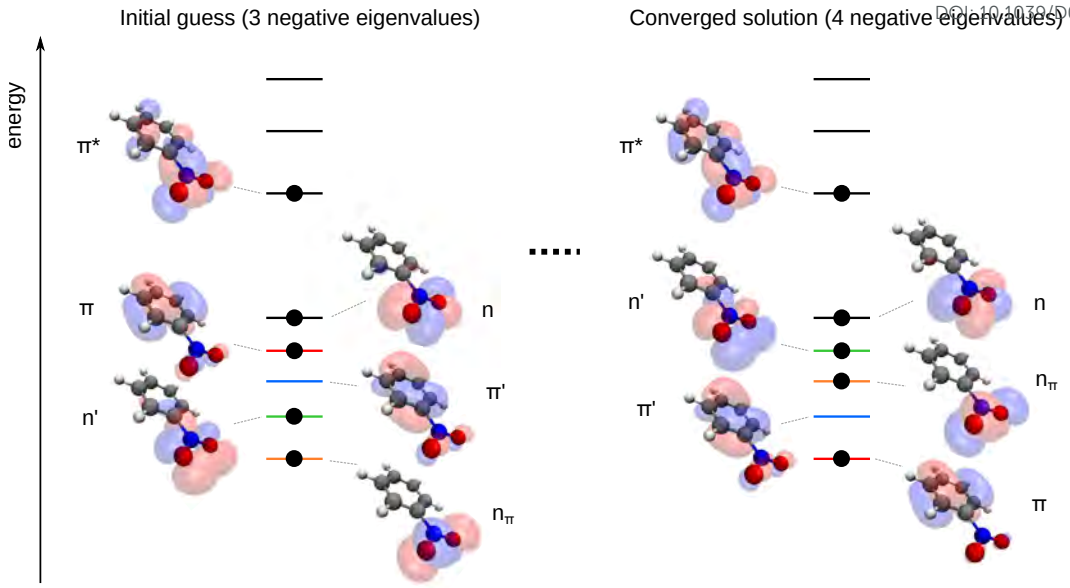


Fig. 5 Illustration of the frontier molecular orbitals of nitrobenzene ordered according to the energy. Left: initial guess for the DO-MOM calculation based on the ground-state orbitals. Right: relaxed $A_1(\pi' \rightarrow \pi^*)$ excited state. Isosurfaces are drawn at an isovalue of 0.1 $Å^{-3/2}$. The labelling of the orbitals follows the notation used in reference [53]. In each case the energy ordering of the orbitals is represented schematically.

quire explicit computation of second derivatives of the energy at each iteration [19]. We further note that the DO-MOM method can be applied more generally than just KS DFT, provided that the appropriate expression for **L** is used in eq. 15 for the gradient. This, for example, opens up the possibility of performing fully variational calculations of excited states with SIC functionals.

An implementation of DO-MOM using a localized basis set representation is found here to outperform a standard implementation of the widely used SCF-MOM approach for two tested quasi-Newton schemes both in terms of robustness and rate of convergence for a 52 excited-state benchmark set. The best performance is obtained with the L-BFGS optimization algorithm with a memory of 20. DO-MOM with L-BFGS is found to be more robust than SCF-MOM for the challenging case of a charge-transfer state of nitrobenzene, which is characterized by excitation from closely spaced orbitals and large electronic rearrangement. This example further illustrates the importance of updating the preconditioner in the DO-MOM as the initial excited state constructed from orbitals of a ground-state calculation has an approximate Hessian with three negative eigenvalues, while the converged excited state is a fourth-order saddle point.

These are preliminary results. The Powell update, which here is found to perform worse than L-BFGS, is not the only quasi-Newton method used for saddle-point searches. For atomic structural rearrangements, the symmetric rank-one (SR1) or a combination of SR1 and Powell updates proposed by Bofill [50,51] are other commonly employed quasi-Newton algorithms. In order to assess the performance of these methods within DO-MOM, an extension of the limited-memory inverse Hessian update algorithm presented in this work is needed. Moreover, the tests should be extended to larger molecules and to excited states of different character, such as long-range charge-transfer states. It would be particularly valuable to assess the performance of DO-MOM with respect to electronic transitions in transition metal complexes, which are usually challenging for SCF methodologies [26].

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The present work was funded by the Icelandic Research Fund (grant number 196070-052) and the University of Iceland Research Fund. AVI is supported by a doctoral fellowship from the University of Iceland.

Notes and references

1 A. Dreuw and M. Head-Gordon, *Chem. Rev.*, 2005, **105**, 4009–4037.
2 M. E. Casida, in *Time-Dependent Density Functional Response Theory for Molecules*, ed. D. P. Chong, World Scientific, 1995, pp. 155–192.
3 E. Runge and E. K. U. Gross, *Physical Review Letters*, 1984, **52**, 997–1000.
4 W. Kohn and L. J. Sham, *Physical Review*, 1965, **140**, 1133–1138.
5 P. Hohenberg and W. Kohn, *Phys. Rev.*, 1964, **136**, B864.
6 R. Van Meer, O. V. Gritsenko and E. J. Baerends, *Journal of Chemical Theory and Computation*, 2014, **10**, 4432–4441.
7 B. G. Levine, C. Ko, J. Quenneville and T. J. Martínez, *Molecular Physics*, 2006, **104**, 1039–1051.
8 N. T. Maitra, F. Zhang, R. J. Cave and K. Burke, *The Journal of Chemical Physics*, 2004, **120**, 5932–5937.
9 D. J. Tozer and N. C. Handy, *Physical Chemistry Chemical Physics*, 2000, **2**, 2117–2121.
10 I. Seidu, M. Krykunov and T. Ziegler, *Journal of Physical Chem-*

istry A, 2015, **119**, 5107–5116.

11 C. L. Cheng, Q. Wu and T. Van Voorhis, *Journal of Chemical Physics*, 2008, **129**, 124112.

12 E. J. Baerends, O. V. Gritsenko and R. Van Meer, *Physical Chemistry Chemical Physics*, 2013, **15**, 16408–16425.

13 A. Dreuw and M. Head-Gordon, *Journal of the American Chemical Society*, 2004, **126**, 4007–4016.

14 A. Dreuw, J. L. Weisman and M. Head-Gordon, *Journal of Chemical Physics*, 2003, **119**, 2943–2946.

15 L. Zhao and E. Neuscamman, *Journal of Chemical Theory and Computation*, 2019, **16**, 164–178.

16 H. R. Zhekova, M. Seth and T. Ziegler, *International Journal of Quantum Chemistry*, 2014, **114**, 1019–1029.

17 T. Ziegler, M. Seth, M. Krykunov and J. Autschbach, *Journal of Chemical Physics*, 2008, **129**, 1–10.

18 D. Hait and M. Head-Gordon, *Journal of Physical Chemistry Letters*, 2020, **11**, 775–786.

19 D. Hait and M. Head-Gordon, *Journal of Chemical Theory and Computation*, 2020, **16**, 1699–1710.

20 E. Pradhan, K. Sato and A. V. Akimov, *Journal of Physics: Condensed Matter*, 2018, **30**, 484002.

21 G. M. Barca, A. T. Gilbert and P. M. Gill, *Journal of Chemical Theory and Computation*, 2018, **14**, 1501–1509.

22 G. Levi, M. Papai, N. E. Henriksen, A. O. Dohn and K. B. Møller, *Journal of Physical Chemistry C*, 2018, **122**, 7100–7119.

23 P. Ramos and M. Pavanello, *The Journal of Chemical Physics*, 2018, **148**, 144103.

24 Y. C. Park, F. Senn, M. Krykunov and T. Ziegler, *Journal of Chemical Theory and Computation*, 2016, **12**, 5438–5452.

25 M. Hemanadhan, M. Shamim and M. K. Harbola, *Journal of Physics B: Atomic, Molecular and Optical Physics*, 2014, **47**, 115005.

26 B. Peng, B. E. Van Kuiken, F. Ding and X. Li, *Journal of Chemical Theory and Computation*, 2013, **9**, 3933–3938.

27 M. K. Harbola, M. Hemanadhan, M. Shamim and P. Samal, *Journal of Physics: Conference Series*, 2012, **388**, 0–9.

28 B. Himmetoglu, A. Marchenko, I. Dabo and M. Cococcioni, *Journal of Chemical Physics*, 2012, **137**, 154309.

29 T. Kowalczyk, S. R. Yost and T. V. Voorhis, *Journal of Chemical Physics*, 2011, **134**, 054128.

30 R. J. Maurer and K. Reuter, *Journal of Chemical Physics*, 2011, **135**, 224303.

31 T. Baruah and M. R. Pederson, *Journal of Chemical Theory and Computation*, 2009, **5**, 834–843.

32 A. T. B. Gilbert, N. A. Besley and P. M. W. Gill, *The journal of physical chemistry. A*, 2008, **112**, 13164–71.

33 J. Gavnholt, T. Olsen, M. Engelund and J. Schiøtz, *Physical Review B*, 2008, **78**, 075441.

34 Y. C. Park, M. Krykunov and T. Ziegler, *Molecular Physics*, 2015, **113**, 1636–1647.

35 R. Pederson and B. M. Klein, *Physical Review B*, 1988, **37**, 319–331.

36 T. Helgaker, P. Jørgensen and J. Olsen, *Molecular Electronic-Structure Theory*, John Wiley & Sons, Ltd, 2014, pp. 107–141.

37 J. P. Perdew and M. Levy, *Physical Review B*, 1985, **31**, 6264–6272.

38 T. Ziegler, A. Rauk and E. J. Baerends, *Theoretica Chimica Acta*, 1977, **43**, 261–271.

39 T. V. Voorhis and M. Head-Gordon, *Molecular Physics*, 2002, **100**, 1713–1721.

40 J. A. R. Shea, E. Gwin and E. Neuscamman, *Journal of Chemical Theory and Computation*, 2020, **16**, 1526–1540.

41 J. Liu, Y. Zhang, P. Bao and Y. Yi, *Journal of Chemical Theory and Computation*, 2017, **13**, 843–851.

42 E. A. Briggs and N. A. Besley, *The Journal of Physical Chemistry A*, 2015, 2902–2907.

43 E. A. Briggs, N. A. Besley and D. Robinson, *Journal of Physical Chemistry A*, 2013, **117**, 2644–2650.

44 G. Levi, E. Biasin, A. O. Dohn and H. Jónsson, *Physical Chemistry Chemical Physics*, 2019, **22**, 748–757.

45 H. Gudmundsdóttir, Y. Zhang, P. M. Weber and H. Jónsson, *The Journal of Chemical Physics*, 2013, **139**, 194102.

46 J. P. Perdew and A. Zunger, *Physical Review B*, 1981, **23**, 5048–5079.

47 Y. Zhang, P. M. Weber and H. Jónsson, *Journal of Physical Chemistry Letters*, 2016, **7**, 2068–2073.

48 S. Lehtola and H. Jónsson, *J. Chem. Theory Comput.*, 2014, **10**, 5324–5337.

49 S. Lehtola, M. Head-Gordon and H. Jónsson, *Journal of Chemical Theory and Computation*, 2016, **12**, 3195–3207.

50 J. M. Bofill and M. Comajuan, *Journal of Computational Chemistry*, 1995, **16**, 1326–1338.

51 J. M. Bofill, *Journal of Computational Chemistry*, 1994, **15**, 1–11.

52 J. M. Anglada, E. Besalú, J. M. Bofill and J. Rubio, *Journal of Mathematical Chemistry*, 1999, **25**, 85–92.

53 J. M. Mewes, V. Jovanović, C. M. Marian and A. Dreuw, *Physical Chemistry Chemical Physics*, 2014, **16**, 12393–12406.

54 J. Douady, Y. Ellinger, R. Subra and B. Levy, *The Journal of Chemical Physics*, 1980, **72**, 1452–1462.

55 M. Head-Gordon and J. A. Pople, *Journal of Physical Chemistry*, 1988, **92**, 3063–3069.

56 J. Hutter, M. Parrinello and S. Vogel, *The Journal of Chemical Physics*, 1994, **101**, 3862–3865.

57 J. Nocedal and S. Wright, *Numerical Optimization*, Springer, New York, 2006.

58 S. Lehtola, F. Blockhuys and C. Van Alsenoy, *Molecules*, 2020, **25**, 1–23.

59 R. A. Olsen, G. J. Kroes, G. Henkelman, A. Arnaldsson and H. Jónsson, *Journal of Chemical Physics*, 2004, **121**, 9776–9792.

60 P. Culot, G. Dive, V. H. Nguyen and J. M. Ghuysen, *Theoretica Chimica Acta*, 1992, **82**, 189–205.

61 J. Baker, *Journal of Computational Chemistry*, 1986, **7**, 385–395.

62 J. Simons, P. Jørgensen, H. Taylor and J. Ozment, *Journal of*

Physical Chemistry, 1983, **87**, 2745–2753.

63 C. J. Cerjan and W. H. Miller, *The Journal of Chemical Physics*, 1981, **75**, 2800–2806.

64 A. Ivanov, E. Ö. Jónsson, T. Vegge and H. Jónsson, *Implementation of a Direct Minimisation Method using Exponential Transformation in the Localised Basis Set Approach*, 2020, In preparation.

65 A. Al-Mohy and N. Higham, *SIAM Journal on Matrix Analysis and Applications*, 2009, **31**, 970–989.

66 P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. Jarrod Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt and SciPy 1. 0 Contributors, *Nature Methods*, 2020, **17**, 261–272.

67 A. Al-Mohy and N. Higham, *SIAM Journal on Matrix Analysis and Applications*, 2009, **30**, 1639–1657.

68 W. Sun and Y.-X. Yuan, *Optimization Theory and Methods*, Springer, Boston, 2006.

69 J. Enkovaara, C. Rostgaard, J. J. Mortensen, J. Chen, M. Dulak, 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 and K. W. Jacobsen, *Journal of Physics: Condensed matter*, 2010, **22**, 253202.

70 A. H. Larsen, M. Vanin, J. J. Mortensen, K. S. Thygesen and K. W. Jacobsen, *Phys. Rev. B*, 2009, **80**, 195112.

71 J. Mortensen, L. Hansen and K. W. Jacobsen, *Physical Review B*, 2005, **71**, 035109.

72 P. E. Blöchl, *Phys. Rev. B*, 1994, **50**, 17953.

73 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865.

74 P. F. Loos, A. Scemama, A. Blondel, Y. Garniron, M. Caffarel and D. Jacquemin, *Journal of Chemical Theory and Computation*, 2018, **14**, 4360–4379.

75 D. E. Woon and T. H. Dunning, *J. Chem. Phys.*, 1993, **98**, 1358–1371.

76 R. A. Kendall, T. H. Dunning and R. J. Harrison, *J. Chem. Phys.*, 1992, **96**, 6796–6806.

77 T. H. Dunning, *J. Chem. Phys.*, 1989, **90**, 1007–1023.

78 F. Weigend and R. Ahlrichs, *Physical Chemistry Chemical Physics*, 2005, **7**, 3297–3305.

79 A. D. Rabuck and G. E. Scuseria, *The Journal of Chemical Physics*, 1999, **110**, 695–700.

80 V. A. Biasuk, *2011*, 2011, **111**, 4197–4205.

81 J. M. Anglada and J. M. Bofill, *Journal of Computational Chemistry*, 1998, **19**, 349.

82 S. Nagakura, M. Kojima and Y. Maruyama, *Journal of Molecular Spectroscopy*, 1964, **13**, 174 – 192.

83 K. Baarman and J. Vandevondele, *Journal of Chemical Physics*, 2011, **134**, 244104.

84 J. VandeVondele and J. Hutter, *Journal of Chemical Physics*, 2003, **118**, 4365–4369.