LoFEx — A local framework for calculating excitation energies: Illustrations using RI-CC2 linear response theory

Pablo Baudin and Kasper Kristensen

Citation: The Journal of Chemical Physics 144, 224106 (2016); doi: 10.1063/1.4953360
View online: http://dx.doi.org/10.1063/1.4953360
View Table of Contents: http://aip.scitation.org/toc/jcp/144/22
Published by the American Institute of Physics

Articles you may be interested in

Towards a pair natural orbital coupled cluster method for excited states
The Journal of Chemical Physics 145, 034102 (2016); 10.1063/1.4958734

Automatic active space selection for the similarity transformed equations of motion coupled cluster method
The Journal of Chemical Physics 146, 074103 (2017); 10.1063/1.4976130

CC2 oscillator strengths within the local framework for calculating excitation energies (LoFEx)
The Journal of Chemical Physics 146, 144107 (2017); 10.1063/1.4979713

A pair natural orbital implementation of the coupled cluster model CC2 for excitation energies
The Journal of Chemical Physics 139, 084114 (2013); 10.1063/1.4819071

Reduced-cost linear-response CC2 method based on natural orbitals and natural auxiliary functions
The Journal of Chemical Physics 146, 194102 (2017); 10.1063/1.4983277

Correlated natural transition orbitals for core excitation energies in multilevel coupled cluster models
The Journal of Chemical Physics 146, 144109 (2017); 10.1063/1.4979908
LoFEx — A local framework for calculating excitation energies: Illustrations using RI-CC2 linear response theory

Pablo Baudin and Kasper Kristensen
qLEAP Center for Theoretical Chemistry, Department of Chemistry, Aarhus University, Langelandsgade 140, DK-8000 Aarhus C, Denmark

(Received 11 March 2016; accepted 24 May 2016; published online 10 June 2016)

We present a local framework for the calculation of coupled cluster excitation energies of large molecules (LoFEx). The method utilizes time-dependent Hartree-Fock information about the transitions of interest through the concept of natural transition orbitals (NTOs). The NTOs are used in combination with localized occupied and virtual Hartree-Fock orbitals to generate a reduced excitation orbital space (XOS) specific to each transition where a standard coupled cluster calculation is carried out. Each XOS is optimized to ensure that the excitation energies are determined to a predefined precision. We apply LoFEx in combination with the RI-CC2 model to calculate the lowest excitation energies of a set of medium-sized organic molecules. The results demonstrate the black-box nature of the LoFEx approach and show that significant computational savings can be gained without affecting the accuracy of CC2 excitation energies. Published by AIP Publishing. [http://dx.doi.org/10.1063/1.4953360]

I. INTRODUCTION

Coupled cluster (CC) theory \(^1,2\) is considered one of the greatest successes of electronic structure theory for the determination of accurate energies and molecular properties of systems dominated by a single configuration. Its strength relies on the hierarchy of CC models that provides a fast and systematic convergence towards the full configuration interaction (FCI) results. The main problem of CC is its computational cost and in particular its steep scaling with the system size. The dynamic correlation effects described by CC theory are spatially local and the steep scaling of the models can be attributed to the use of canonical molecular orbitals (CMOs) that are highly delocalized. This issue prevents a straightforward application of CC theory to large molecules and approximate CC models have to be developed.

The need for accurate and reliable methods for the calculation of excitation energies of large molecular systems can be seen from the number of attempts to do so in the last decade. All recently developed methods try to take advantage of the locality of correlation effects and electronic transitions. The concept of natural transition orbitals (NTOs) \(^3\) was, for example, used in the incremental scheme devised by Mata and Stoll \(^4\) where they used a mixed occupied orbital space composed of NTOs and localized molecular orbitals (LMOs). The virtual orbitals were left intact, and they relied on a many-body expansion of the excitation energies with a fragmentation of the occupied molecular orbital space to reduce the cost of the equation-of-motion coupled cluster singles and doubles (EOM-CCSSD) model.

Other local CC methods focus on reformulations of the second-order approximate coupled cluster singles and doubles (CC2) model. \(^5\) The CC2 model has proven to be a good compromise between accuracy and computational cost for the calculation of frequency-dependent molecular properties and its reformulation using density fitting techniques (e.g., the resolution of the identity, RI-CC2) has significantly extended the application range of the method. \(^6,7\) To further reduce the cost of the CC2 model, Helmich and Hättig proposed a local version of CC2 for excitation energies where they used orbital-specific virtuals (OSVs) and pair natural orbitals (PNOs) to reduce the dimension of the virtual orbital space. \(^8\) Along the same lines, the local CC implementation of Kats, Korona, and Schütz uses information from the coupled cluster singles (CCS) model to select the relevant occupied LMOs to describe each transition, combined with projected atomic orbitals (PAOs) for the virtual space. This strategy was first applied to the CCSD model and involved local approximations in both the singles and the doubles amplitudes, \(^9\) while the singles amplitudes were left intact in their local CC2 implementation. \(^10\) In a more recent work, \(^11\) they proposed a multistate implementation relying on a Laplace transformation of the orbital energy denominators which resulted in an improved description of the CC2 excitation spectrum.

The multi-level coupled cluster theory of Myhre \emph{et al.} \(^12,13\) is another CC-based method employing a different strategy. In their methods, the orbital space is partitioned based on chemical intuition and different CC models are used to treat different parts of the molecule. In this way, high accuracy (CCSD, CC3) \(^14,15\) can be achieved at a reduced cost by treating the less important part of the system with CCS and CC2 models.

In this paper, we present a local framework for calculating excitation energies (LoFEx) at the CC level of theory with emphasis on the RI-CC2 model. The overall goal is to enable the calculation of accurate excitation energies of large molecular systems by maintaining the CC error control with a computational cost that can compete against time-dependent density functional theory (TD-DFT).
In the same way as for most local CC models, the strategy used by LoFEx is to reduce the dimension of the orbital space in which the CC equations have to be solved. As we show in Section II B, this is done by considering a mixed orbital space composed of NTOs and LMOs for both the occupied and virtual orbital spaces. Information about the main characteristics of the electronic transition is included in the orbital space through the NTOs, while the LMOs enable a compact description of correlation effects at the CC level. In order to reduce the dimension of the mixed NTO/LMO space for the correlated calculation, the LoFEx method considers only the LMOs in the vicinity of the NTOs. The size of the orbital space is optimized in a black-box manner such that the computed excitation energies reproduce the conventional CC results to a predefined accuracy (cf. Section II C). We note that restricting both the occupied and the virtual orbital spaces to a (small) fraction of the total orbital space is a fundamental requirement for the calculation of excitation energies of (very) large molecules. LoFEx differs from the aforementioned local schemes for calculating excitation energies since localized orthogonal virtual orbitals are used; the local excitation space of occupied and virtual orbitals is optimized dynamically, and no many-body expansion of the excitation energy is performed (as is done in the incremental scheme).

We present numerical RI-CC2 results for the lowest transitions in a set of organic molecules and analyse the convergence of excitation energies with the size of the mixed NTO/LMO space in Section III B. The errors and the cost-reduction of the black-box LoFEx procedure when compared to standard RI-CC2 implementations are investigated in Section III C.

II. THEORETICAL FOUNDATIONS OF LoFEx

Over the last few decades, many theoretical methods and tools have been developed to calculate excitation energies and analyse electronic transitions. More recently, the focus has been directed towards developing algorithms with lower computational costs. In this section, we present some of those concepts and describe how we have combined them to develop the LoFEx procedure.

A. The RI-CC2 model for excitation energies

In this paper, we target excitation energies calculated at the CC2 level of theory. The CC2 model was introduced by Christiansen et al.\(^5\) as an intermediate model between the CCS and CCSD models in the CC hierarchy for the calculation of frequency-dependent properties. The CC2 amplitude equations in the absence of external perturbations are given by

\[
\Omega_{\mu_1} = \langle \mu_1 | \hat{H} + [\hat{H},T_2] | \text{HF} \rangle = 0, \tag{1}
\]

\[
\Omega_{\mu_2} = \langle \mu_2 | \hat{H} + [F,T_2] | \text{HF} \rangle = 0, \tag{2}
\]

where \(|\text{HF}\rangle,|\mu_1\rangle,|\mu_2\rangle\) denote the Hartree-Fock (HF) ground state and the set of singles and doubles excitation manifolds. \(F\) is the Fock operator and \(\hat{H}\) is a similarity transformed Hamiltonian,

\[
\hat{H} = \exp(-T_1) \hat{H} \exp(T_1), \tag{3}
\]

where \(T_1 = \sum \tau_{\mu_1} \hat{\tau}_{\mu_1}\) is a cluster operator, \(\tau_{\mu_1}\) is a cluster amplitude, \(\hat{\tau}_{\mu_1}\) is an excitation operator, and \(i\) denotes the excitation level. In the CC2 model, the doubles amplitudes are only correct through first order in the fluctuation potential \((\Phi = H - F)\). This approximation leads to a particularly simple form of the doubles amplitudes, which for closed shell systems reads

\[
t''_{ij}^{ab} = \frac{1}{1 + \delta_{ia} \delta_{jb}} \frac{\tilde{g}_{aibj}}{\epsilon_i - \epsilon_a + \epsilon_j - \epsilon_b}, \tag{4}
\]

where \(i,j \ (a,b)\) denote occupied (virtual) CMOs, \(\epsilon_p\) is the energy associated with orbital \(p\) of unspecified occupancy, and \(\tilde{g}_{aibj}\) is a \(T_1\)-transformed two-electron repulsion integral in Mulliken notation.\(^{16}\) This simple form of the doubles amplitudes can lead to a CCS-like formulation of the CC2 equations where the doubles amplitudes are calculated on-the-fly. In order to take full advantage of this formulation and avoid the storage of any four-index quantity (amplitudes or integrals), Hättig and Weigend used the resolution of the identity approximation for the two-electron integrals.\(^8\) Batches of doubles amplitudes are then contracted on-the-fly with three-center integrals to form intermediates that are used in the CC2 vector function. As shown in Ref. 6, this strategy can be generalized to the calculation of excitation energies at the RI-CC2 level. CC excitation energies are usually obtained as eigenvalues of the CC Jacobian matrix. However, to avoid storing four-index quantities also in the excitation energy part of the calculation, an effective CC2 Jacobian has to be considered,

\[
A^\text{eff}(\omega)b = \omega b, \tag{5}
\]

\[
A^\text{eff}_{\mu_1\nu_1}(\omega) = A_{\mu_1\nu_1} - \sum_{\gamma_2} A_{\mu_1\gamma_2} A_{\gamma_2\nu_1} \epsilon_{\gamma_2} - \omega, \tag{6}
\]

where \(\omega\) is the excitation energy of interest, \(b\) is a singles excitation vector, and \(A_{\mu_1\nu_1} = \partial \Omega_{\mu_1}/\partial \tau_{\nu_1}\) is an element of the full CC2 Jacobian. The non-linearity in the excitation energy \(\omega\) introduced by the effective Jacobian in Eq. (6) implies that a standard subspace diagonalization procedure cannot be employed to obtain the RI-CC2 excitation energies. In practice, we construct the effective Jacobian with an initial guess for the eigenvalue \(\omega\). It is then diagonalized using standard Davidson techniques,\(^{17–19}\) and the optimized eigenvalue is used to build a new effective Jacobian until self-consistency is reached. This process is repeated for each excitation of interest.\(^7\)

B. A mixed NTO/LMO space

In order to provide a compact description of electronic transitions, it is important to work with an optimal orbital space both for occupied and virtual orbitals. LoFEx is therefore relying on a mixed orbital representation composed of the most important pair of NTOs, while the rest of the orbital space is localized. The dominant pair of NTOs provides a compact description of the main characteristics of the transition, while the LMOs enable an efficient description
of electron correlation effects. In this section, we detail the generation of the mixed NTO/LMO space.

NTOs are usually generated by singular value decomposition (SVD) of the transition density matrix \( \rho^{\text{CBS}}_{ai} \) between the HF ground state and an excited state obtained from the CCS model.\(^3\)

\[ \rho^{\text{CBS}}_{ai} = \langle \text{CCS} | \tau_{ai} | \text{HF} \rangle. \tag{7} \]

We note for later that NTOs can also be generated at the correlated level by performing an SVD of the singles excitation vector \( \rho^{\text{CBS}}_{ai} \) only. Our current pilot implementation also generates NTOs from time-dependent Hartree-Fock theory (TDHF). The TDHF transition density matrix in the CMO basis can be expressed as\(^20\)

\[ \bar{b}_{ai} = \frac{X_{ai} + Y_{ai}}{|X_{ai} + Y_{ai}|} \sqrt{X_{ai}^2 + Y_{ai}^2}, \tag{8} \]

where \( X \) and \( Y \) are solutions of the TDHF equation for a given excitation energy \( \tilde{\omega} \).

\[
\begin{pmatrix}
A & B \\
B & A
\end{pmatrix}
\begin{pmatrix}
X \\
Y
\end{pmatrix}
= \tilde{\omega}
\begin{pmatrix}
1 & 0 \\
0 & -1
\end{pmatrix}
\begin{pmatrix}
X \\
Y
\end{pmatrix}. \tag{9}
\]

The transformation matrices from the CMO basis to NTOs are then obtained by solving the following eigenvalue equations:

\[
\bar{b} \bar{b}^k u_k = \lambda_k u_k, \quad k = 1, 2, \ldots, N_o, \tag{10}
\]

\[
\bar{b} \bar{b}^k v_k = \lambda_k v_k, \quad k = 1, 2, \ldots, N_v, \tag{11}
\]

where \( N_o \) is the number of occupied valence orbitals (in order to prevent the core and valence spaces from mixing), while \( N_v \) is the number of virtual orbitals. Finally, unitary matrices transforming to the NTO basis are given by

\[
U = (u_1, u_2, \ldots, u_{N_o}), \tag{12}
\]

\[
V = (v_1, v_2, \ldots, v_{N_v}), \tag{13}
\]

for occupied and virtual orbitals, respectively. Note also that \( \lambda_k \equiv \lambda_k^1 \) for \( k = 1, 2, \ldots, N_o \), and that \( \lambda_k^1 = 0 \) for \( k = N_o + 1, \ldots, N_v \), assuming \( N_o \leq N_v \). The first \( N_o \) eigenvalues give information about the importance of a given pair of occupied-virtual NTOs for the description of the transition represented by \( \bar{b} \). For excited states dominated by single-electron replacements—which we are targeting in this paper—this procedure results in one occupied-virtual pair of NTOs with a singular value very close to one, while the remaining NTOs have much smaller singular values and thus provide little information about the electronic transition. We therefore choose to keep only the dominant pair of NTOs, while the rest of the orbital space is localized to enable a compact description of electron correlation effects in the CC part of the calculation. Different localization techniques have been investigated, including Boys localization\(^21,22\) and minimization of the second power of the second\(^23\) and fourth\(^24\) moments of the orbitals. No significant differences in the results presented in Section III were observed and we have therefore chosen a minimization of the second power of the second moment of the orbitals which is known to provide a good compromise between computational cost and spatial locality.\(^25\) It is also important to emphasize that the use of NTOs makes the orbital space specific to a given transition, the one associated with the transition density matrix used to generate the set of NTOs. A different mixed orbital space should thus be constructed for each transition of interest. The generation of the mixed orbital space is summarized in Fig. 1.

C. Space optimization procedure

Once the full orbital space has been transformed to the mixed orbital space to enable an efficient description of the transition process, we need to build and optimize a reduced excitation orbital space (XOS) where the RI-CC2 calculation is carried out to provide the target excitation energy to a predefined accuracy. The main character of the excitation is described by the dominant pair of NTOs already included in the mixed orbital space. The XOS should thus be constructed by considering the dominant NTOs as well as a reduced set of important LMOs. We therefore setup a priority list of all LMOs ordered based on their estimated importance for the description of correlation effects in the vicinity of the NTOs. The importance of a given LMO \( p \) (occupied or virtual) is evaluated by the effective distance \( \tilde{r}_p \) given by

\[
\tilde{r}_p = \min_A \left( \frac{r_{Ap}}{Q_A^{\text{NTO,o}}}, \frac{r_{Ap}}{Q_A^{\text{NTO,v}}} \right), \tag{14}
\]

where index \( A \) denotes atomic centers, \( r_{Ap} \) corresponds to the distance between the center of charge of a local orbital \( p \) and atomic center \( A \), and \( Q_A^{\text{NTO,o}} \) and \( Q_A^{\text{NTO,v}} \) are the Löwdin atomic charges of the occupied and virtual NTOs on center \( A \), respectively. The Löwdin atomic charges of the NTOs (with values between 0 and 1) are used to modify the \( r_{Ap} \) distance in Eq. (14) as an attempt to quantify the distance
between the LMOs and the NTOs, even in the case where the NTOs are delocalized over several atomic centers. The smaller the modified distance \( \tilde{r}_p \) is, the more important orbital \( \rho \) is expected to be.

A first guess for the XOS (XOS\(_1\)) is constructed by including the NTOs as well as a given number of LMOs from the priority list defined by Eq. (14). The standard RI-CC2 amplitude and Jacobian eigenvalue equations are then solved as described in Section II A, except that the excitation manifold is restricted to XOS\(_1\) instead of the full orbital space. For simplicity, the Fock matrix is diagonalized in XOS\(_1\) to solve the RI-CC2 equations in a pseudo-canonical basis. Because of the restriction on the orbital space, the ground state correlation energy associated with the RI-CC2 amplitudes is meaningless and therefore not calculated. However, the excitation energy obtained by diagonalization of the Jacobian in XOS\(_1\) corresponds to a first estimate of the exact RI-CC2 value (in the full orbital space). The quality of this estimate is of course strongly dependent on the size of XOS\(_1\) and on the nature of the excitation.

In order to provide accurate excitation energies in a black-box manner, an optimization of the XOS is required. To this end, XOS\(_2\) is constructed by including more orbitals based on the priority list defined by Eq. (14). The correlated calculation is then carried out in XOS\(_2\) (solution of the RI-CC2 amplitude and Jacobian eigenvalue equations), and the new excitation energy is compared to the previous one. The orbital space is considered converged when the difference between two subsequent excitation energies is smaller than a given threshold (\( \tau_{\text{XOS}} \)). The whole procedure is described step by step in Algorithm 1 and summarized schematically in Fig. 2. Default values for the main LoFEx threshold (\( \tau_{\text{XOS}} \)) and the number of orbitals added to the XOS in each iteration will be discussed in Section III.

A couple technical complications arise from the simple procedure described so far. The first issue concerns the computation of several excitation energies. The Davidson algorithm usually employed for such eigenvalue problems converges to the lowest excitation energies present in the subspace.\(^{17-19}\) However, if one is interested in the two lowest transitions for a given system, the mixed orbital spaces in LoFEx will be different for each transition, and consequently, the XOSs for the two transitions will also be different and might not even overlap. This means that in the calculation of the second excitation energy, depending on the size of the XOS, the second transition might correspond to the first or the second Jacobian eigenvalue.

In order to circumvent this issue, we employ a projection strategy where, for a given transition \( k \), all previously optimized excitation vectors \( \{b_{\text{opt}}^1, b_{\text{opt}}^2, \ldots, b_{\text{opt}}^{k-1}\} \) are projected out of the Davidson subspace. This prevents the optimized vector for transition \( k \) from collapsing into one of the previous solutions. Because the previously optimized excitation vectors are determined in different XOSs, they need to be transformed to the current mixed orbital space and projected onto the current XOS. As a consequence, the set of previously optimized vectors is usually not orthonormal in the current XOS and a given trial vector \( b' \) in the Davidson algorithm needs to be orthonormalized with respect to the previous excitation vectors using

\[
b' \leftarrow (1 - P)b',
\]

\[
P = \sum_{j=1}^{k-1} \sum_{l=1}^{k-1} b_{\text{opt}}^j S^{-1} T_{jl} (b_{\text{opt}}^l)^T,
\]

\[
S_{jl} = (b_{\text{opt}}^j)^T b_{\text{opt}}^l,
\]

where all quantities are expressed in the current mixed orbital space as well as restricted to the current XOS. Because of the difference in dimensions between the different XOSs, a given excitation vector might have only very small components in the current XOS (e.g., if the current and old transitions occur in different extremities of the molecular system). In such a case the excitation vector is discarded from the set of \( k - 1 \) previous vectors to avoid problems in the inversion of the overlap matrix \( S \).

One possible complication is that the lowest state in the Davidson subspace (after projection) might not have any resemblance with the TDHF state used to calculate NTOs (e.g., in cases where the TDHF and CC2 spectra differ significantly). The current mixed orbital space is then ill-suited to describe the CC2 transition of interest. This issue can be fixed by recalculating NTOs from the optimized CC2 singles excitation vector (\( b_{\text{CC2}}^{\text{opt}} \)) obtained in the current XOS. This is done whenever

\[
|b^T b_{\text{CC2}}^{\text{opt}}| < \tau_{\text{overlap}},
\]

where \( \tilde{\mathbf{b}} \) is the last starting guess (usually from TDHF) and \( \tau_{\text{overlap}} \) is fixed to 0.5 for normalized vectors. A new mixed orbital space is then generated using CC2 NTOs and the optimization procedure is restarted as shown in step 12 of Algorithm 1.

---

**FIG. 2.** Schematic representation of the generation of the excitation orbital space (XOS) in LoFEx. The orbitals in the mixed orbital space (left box) are reordered (middle box) according to the measure in Eq. (14), and the XOS (right box) used for the calculation of CC excitation energies is optimized as detailed in Algorithm 1. Gray: natural transition orbitals (NTOs); stripes: Localized molecular orbitals (LMOs).
ALGORITHM 1. LoFEx pseudo-algorithm, where \( \tau_{\text{XOS}} \) is the main threshold controlling the accuracy of the final excitation energies.

1: Solve TDHF problem for the \( n \) target excitation energies: \( \tilde{\omega}_1, \ldots, \tilde{\omega}_n \).
2: for state \( i = 1, n \) do
3:   Generate mixed orbital space using \( \tilde{b}_i \) (cf. Fig. 1).
4:   Build XOS, based on distance measure [Eq. (14)].
5:   Solve RI-CC2 amplitude and Jacobian equations in XOS and get \( \omega_i^{(1)} \).
6:   \( j \leftarrow 1 \)
7:   repeat
8:     \( j \leftarrow j + 1 \)
9:     Increase orbital space based on distance measure [Eq. (14)] and get XOS.
10:   until \( |\omega_i^{(j)} - \omega_i^{(j-1)}| < \tau_{\text{XOS}} \)
11: if \( |\tilde{b}_i^{(CC2)}| < \tau_{\text{overlap}} \) then \( b_i \leftarrow b_i^{(CC2)} \) and go to step 3.
12: Store optimized excitation energy, \( \omega_i \leftarrow \omega_i^{(j)} \).
13: end for

A second issue concerns the number of atomic orbitals (AOs) needed to describe the molecular orbitals of the XOS. Indeed, both local orbitals and NTOs are constructed as linear combinations of AOs and have in principle small but non-zero coefficients on atoms far from their center of charge. As a consequence, even a small XOS would result in many two-electron integrals to be calculated in the AO basis and would destroy the performance of the method. We therefore reduce the number of AOs \( \{ \chi_{\mu} \} \) involved in the construction of a given MO \( \phi_p \),

\[
\phi_p = \sum_{\mu}^{\chi_{\mu}} \chi_{\mu} c_{\mu p},
\]

by introducing a modified MO \( \tilde{\phi}_p \) according to

\[
\tilde{\phi}_p = \sum_{\mu}^{\chi_{\mu}} \tilde{\chi}_{\mu} \tilde{c}_{\mu p},
\]

where the summation over AO indices, \( \tilde{\chi}_{\mu} \), has been restricted to a subset of AOs \( \{ \chi_{\mu} \} \). This restriction is done based on the Löwdin charges of orbital \( p \) and the coefficients \( \tilde{c}_{\mu p} \) are determined such that \( \tilde{\phi}_p \) resembles the true MO \( \phi_p \) as much as possible in a least square sense. The details about this procedure are described in Ref. 26 in the context of the divide–expand–consolidate local CC framework. Similarly, in connection with the RI approximation only a subset of auxiliary basis functions are considered, as described in Ref. 27.

III. NUMERICAL ILLUSTRATIONS

In this section, we present proof of concept calculations to support the theoretical foundations developed in Section II. We first investigate the convergence of excitation energies with the size of the XOS and then proceed to the application of the LoFEx algorithm as a black-box method.

A. Molecules and computational details

In Section III B we consider the following test molecules which have been optimized at the RI-MP2\(^{29}\)/cc-pVTZ(cc-pVTZ-RI) level of theory:

- caprylic acid,
- lauric acid,
- palmitic acid,
- 15-oxopentadecenoic acid (15-OPDA), and
- Dec-1,3,5,7,9-pentaene (C\(_{10}\)H\(_{12}\)).

For the application of LoFEx in Sections III C and III D, the above set is augmented with the following molecules:

- prostacyclin,
- an \( \alpha \)-helix composed of 8 glycine residues (\( \alpha \)-Gly\(_8\)),
- leupeptin,
- latanoprost,
- metenkephalin, and
- Phenothiazine-isooalloxazine dyad (dyad).

Those larger molecules were optimized at the RI-MP2/cc-pVDZ(cc-pVDZ-RI) level of theory. All geometry optimizations were performed using the ORCA program\(^{29}\) and Cartesian coordinates are available in the supplementary material.\(^{30}\)

For the computation of vertical singlet excitation energies, the cc-pVDZ(cc-pVDZ-RI) basis sets were used on the hydrogen atoms, while diffuse functions were added to the other atoms via the aug-cc-pVDZ(aug-cc-pVDZ-RI) basis sets.\(^{31–33}\) We denote this combination of basis sets with a prime (aug-cc-pVDZ'). For the calculations on the dyad system in Section III D, the cc-pVDZ/cc-pVDZ-RI) basis sets were used without adding diffuse functions. The frozen core approximation was used in all calculations. The LoFEx algorithm as described in Section II has been implemented in a local version of the LSDAlx program.\(^{34,35}\) The RI-CC2 algorithm used in LoFEx relies on a self-consistent Davidson algorithm\(^7\) where the norm of the residual is converged below \( \tau_{\text{residual}} = 10^{-4} \) a.u. and the self-consistent energy threshold is set to \( \tau_{\text{exc}} = 10^{-4} \) a.u. (\( \approx 2.7 \text{ meV} \)). This last parameter should not be confused with the main LoFEx threshold, \( \tau_{\text{XOS}} \) which will be specified in the following investigation. In practical applications of LoFEx, it is of course important to have \( \tau_{\text{XOS}} \geq \tau_{\text{exc}} \).
B. Convergence of excitation energies

1. Saturated fatty acids

For analysis purposes, we consider the lowest transition in a set of saturated fatty acids with a chain of 8 (caprylic acid), 12 (lauric acid), and 16 (palmitic acid) carbon atoms. Fatty acids are simple systems, for which the lowest excitation is known to be localized on the carboxyl group. It is therefore a typical case where a canonical CC calculation of the excitation energy is unnecessarily expensive, since most of the carbon chain does not affect the transition. In order to confirm this statement, we examined the convergence of the lowest excitation energy with the size of the XOS in LoFEx. A fixed number of orbitals were added to the XOS in each LoFEx iteration based on the priority list defined by Eq. (14) and the convergence threshold for the excitation energy ($\tau_{XOS}$) was set to zero, such that the full orbital space is included in the last iteration, which therefore provides the target RI-CC2 excitation energy.

In Fig. 3, we have plotted the dominant pair of TDHF NTOs for the lowest transition of the caprylic, lauric, and palmitic acids. These plots clearly support our initial expectation that — at least at the TDHF level — the transition of interest is localized on the carboxyl group and the carbon chain has no important effects on the main characteristics of the electronic transition. This is confirmed at the RI-CC2 level by the plots presented in Fig. 4, which illustrate the effects of the size of the XOS on the excitation energy when MOs are included based on the priority list defined by Eq. (14). It shows a fast and smooth convergence towards the standard RI-CC2 excitation energy. In particular, the error in the excitation energy is below 0.01 eV for all three systems already when...
less than 230 MOs are included in the orbital space (third point in the curves). This demonstrates that a significant amount of computational resources can be saved for such small but simple systems.

2. 15-oxopentadecanoic acid (15-OPDA)

As a complement to this preliminary investigation, we now consider the three lowest excitation energies of the 15-oxopentadecanoic acid (15-OPDA). By looking at the dominant TDHF NTOs of the three transitions in Fig. 5, it is seen that the lowest and third transitions of 15-OPDA are localized on the aldehyde group, while the second lowest transition occurs on the carboxyl group at the other end of the carbon chain. This system is thus a good candidate to test the projection strategy presented in Section II C. The same analysis as before is performed by including a fixed number of orbitals to the XOS until the complete orbital space is included, and this is done independently for the three transitions of interest.

FIG. 5. Occupied (top) and virtual (bottom) TDHF natural transition orbitals (NTOs) for the three lowest transitions of 15-oxopentadecanoic acid (15-OPDA) using the aug-cc-pVDZ′ basis set. The contour plot value was set to 0.02 a.u. (a) Dominant pair of NTOs for \( S_1 \). (b) Dominant pair of NTOs for \( S_2 \). (c) Dominant pair of NTOs for \( S_3 \).

The challenge of computing several excitation energies can be exemplified as follows: when computing the second lowest transition on the carboxyl group of 15-OPDA, the first XOS (XOS\(_1\)) will not include any component of the lowest excitation vector on the aldehyde group. The standard Davidson procedure would therefore provide the excitation energy of interest (the second lowest for the whole system) by converging to the lowest Jacobian eigenvalue. However, if the XOS corresponds to the complete orbital space, the lowest eigenvalue obtained with the standard Davidson procedure would most likely be the lowest transition on the aldehyde group. The projection strategy described in Section II C is therefore necessary to ensure that the excitation vector is not collapsing to the lowest transition as the XOS is increased.

In Fig. 6, we have plotted the three lowest excitation energies of 15-OPDA against the size of the XOS. It is important to emphasize that the three excitation energies on a specific vertical line in Fig. 6 are determined in different and independent XOSs that contain the same amount of orbitals. The curves in Fig. 6 show a smooth convergence behaviour for all three excitation energies without ever collapsing to another transition which shows that the projection strategy presented in Section II C is enough to ensure convergence to the desired excitation energies. Fig. 6 also shows that the two lowest excitation energies of 15-OPDA converge as fast as the lowest excitation energy of the fatty acids in Fig. 4, while the convergence of the third excitation energy is significantly slower. This is due to the more diffuse character of the third transition (cf. Fig. 5(c)) which is rather poorly described in the small aug-cc-pVDZ′ basis.

3. Dec-1,3,5,7,9-pentaene (C\(_{10}\)H\(_{12}\))

Electronic transitions are not always as local as for the systems treated so far. In some interesting cases, the excitation

FIG. 6. Convergence of the three lowest excitation energies of the 15-oxopentadecanoic acid (15-OPDA) using the aug-cc-pVDZ′ basis as a function of the size of the excitation orbital space (XOS) when MOs are included based on the priority list defined by Eq. (14). The last points of the curves and the dotted horizontal lines correspond to the reference RI-CC2 values (complete orbital space).
can be spread over a large part of the molecule, for example, due to electronic delocalization in conjugated systems or in the case of charge-transfer (CT) transitions. Such examples would necessarily require a larger XOS and in some special cases the complete orbital space needs to be included. To illustrate further the impact of the character of the transition on the size of the XOS, we have applied LoFEx to the C_{10}H_{12} conjugated molecule. The TDHF NTOs plotted in Fig. 7 show that the first transition of C_{10}H_{12} is delocalized over the whole system. Fig. 8 represents the decay of the lowest excitation energy of C_{10}H_{12} with the size of the XOS. As expected, the convergence is stable but much slower than for the systems analyzed previously. For this small molecule, the complete orbital space has to be included in the XOS to reproduce the standard RI-CC2 result, but we can expect that for larger systems, even if the NTOs are delocalized over a part of the molecule, computational savings should still be possible by ignoring all parts of the system not affected by the transition. In short, C_{10}H_{12} represents a worst case scenario for LoFEx and no computational savings can be obtained. Nonetheless, as we shall see in Section III C, the standard RI-CC2 excitation energy is still obtained by the black-box LoFEx algorithm.

4. Convergence behaviour

Finally, we note that, for all examples treated in this section, the excitation energies are systematically lowered when the dimension of the XOS increases. This behaviour is also usually observed in basis set convergence studies of excitation energies \(^3\) and can be partly explained by Cauchy’s interlacing theorem. Part of Cauchy’s theorem states that, for a symmetric matrix, the eigenvalues \(\omega^B_k\) obtained with a basis \(\{B_1\}\) will always be larger than the corresponding eigenvalues \(\omega^B_k\) obtained with a basis \(\{B_2\}\) when \(\{B_1\} \subset \{B_2\}\) (i.e., \(\omega^B_k \geq \omega^B_k\) for each eigenvalue \(k\)). This is for example the case in the configuration interaction (CI) excitation energy problem where a symmetric matrix has to be diagonalized. However, it does not strictly hold in CC theory, partly because the Jacobian matrix is not Hermitian and therefore the eigenvalue problem is not ensured to have (real) solutions, but also because the Jacobian matrix depends on the CC amplitudes which are different from one basis to another meaning that the Jacobian in the basis \(\{B_1\}\) is not a true submatrix of the Jacobian in the \(\{B_2\}\) basis. In spite of these formal differences between CI and CC, we have always observed that the energies are lowered when the XOS is increased, indicating that in practice, Cauchy’s theorem effectively holds also for the CC Jacobian. This feature is important in the context of LoFEx since it lowers the risk of premature convergence when the excitation energies of two subsequent XOSs are compared (step 11 in Algorithm 1).

C. LoFEx as a black-box method

It has now been demonstrated with simple examples that the method described in Section II can lead to significant savings. In this section we apply LoFEx as a black-box method and provide default parameters for the optimization procedure that result in accurate excitation energies using a reduced orbital space. The main challenge here is to find the number of orbitals to add to the XOS between two iterations, such that it is large enough to avoid premature convergence and small enough to prevent the XOS from becoming unnecessarily large. This quantity needs to be conservative as the convergence rate of the excitation energies with the size of the XOS is strongly dependent on the nature of the transitions and the chemical structure of the molecules and premature convergence could in principle occur in some difficult cases. In this investigation, the number of orbitals added to the XOS in each LoFEx iteration was set to ten times the average number of MOs per atom (i.e., \(10 \times N_{MO}/N_{atom}\)) and the main LoFEx threshold was chosen to be \(\tau_{XOS} = 0.02\) eV. The number of orbitals added has been chosen based on a series of tests and seems like a conservative choice but more extensive testing of LoFEx will tell if this value is optimal. The LoFEx threshold \(\tau_{XOS}\) is more of a tunable parameter and depends on the required accuracy. A value of 0.02 eV is conservative and is expected to provide excitation energies of standard CC2 quality (CC2 is known to result in excitation energies where the typical errors are an order of magnitude larger).
In Table I we report the RI-CC2-LoFEx excitation energies as well as measures of the size of the converged XOS and the number of LoFEx iterations (the number of different XOSs considered) for all molecules presented in Section III A. Speed-ups of LoFEx compared to canonical RI-CC2 implementations are also reported. For calculations containing more than one transition, the current RI-CC2 implementation in the LSDATON program does not allow for fair comparisons. Speed-ups are therefore not reported for those cases.

Comparison with canonical RI-CC2 energies — available in the supplementary material\cite{footnote5} — results in errors in the LoFEx excitation energies below 0.01 eV for all molecules. This indicates that the LoFEx parameters chosen above are conservative enough to provide excitation energies of CC2 quality and to avoid false convergence. Concerning the size of the optimized XOS and the reduction in computational cost, one can see from Table I that it is strongly dependent on the nature of the electronic transition and on the size of the molecular system. However, even when the full orbital space has to be included in the XOS (e.g., $S_1$ of prostacyclin), the last RI-CC2 calculation will in general be time-dominating and the overall cost is of the same order as for canonical RI-CC2 calculations (for C$_{10}$H$_{12}$ with a speed-up of 0.43, the system is so small that the full calculation is very fast anyway). On the other hand, significant savings are usually obtained when the LoFEx algorithm converges in only two or three iterations. For example, a LoFEx calculation of the lowest transition of metenkephalin provides a speed-up of almost 60, which is metenkephalin provides a speed-up of almost 60, which is challenging for many computational methods and in particular for TD-DFT.

The five lowest excitation energies of the dyad have been calculated using LoFEx and are reported in Table I. We note that the error in the excitation energies is as low as for the other systems (below 0.01 eV) but that due to the complexity of the transition, the full molecular orbital space has to be included in the XOS for all but the first transition. It is also important to note that for those states (S$_1$, S$_3$, S$_4$, and S$_5$), CC2 NTOs were calculated in the course of the calculation because the overlap with the starting guess was below $\tau_{\text{overlap}} = 0.5$ [cf.

### Table I. Excitation energies from RI-CC2-LoFEx ($\omega_{\text{LoFEX}}$) and size of the XOS given by the fraction of the complete set of MOs and AOs included in the XOS. The number of iterations used in LoFEx (number of different XOSs considered) as well as speed-ups compared to canonical RI-CC2 calculations is also reported. All the calculations have been done using the aug-cc-pVDZ basis set (except for the dyad which uses cc-pVDZ) and the frozen core approximation. The main LoFEx threshold was set to $\tau_{\text{XOS}} = 0.02$ eV.

<table>
<thead>
<tr>
<th>System</th>
<th>State</th>
<th>$\omega_{\text{LoFEX}}$</th>
<th>% MOs</th>
<th>% AOs</th>
<th>No. iterations</th>
<th>Speed-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caprylic acid</td>
<td>S$_1$</td>
<td>6.06</td>
<td>82.3</td>
<td>94.8</td>
<td>2</td>
<td>1.11</td>
</tr>
<tr>
<td>Lauric acid</td>
<td>S$_1$</td>
<td>6.05</td>
<td>83.9</td>
<td>96.2</td>
<td>3</td>
<td>3.49</td>
</tr>
<tr>
<td>Palmitic acid</td>
<td>S$_1$</td>
<td>6.06</td>
<td>63.5</td>
<td>72.6</td>
<td>3</td>
<td>4.55</td>
</tr>
<tr>
<td>15-OPDA</td>
<td>S$_1$</td>
<td>4.44</td>
<td>46.1</td>
<td>57.0</td>
<td>2</td>
<td>...</td>
</tr>
<tr>
<td></td>
<td>S$_2$</td>
<td>6.06</td>
<td>46.5</td>
<td>52.3</td>
<td>2</td>
<td>...</td>
</tr>
<tr>
<td></td>
<td>S$_3$</td>
<td>6.19</td>
<td>100</td>
<td>100</td>
<td>5</td>
<td>...</td>
</tr>
<tr>
<td>C$<em>{10}$H$</em>{12}$</td>
<td>S$_1$</td>
<td>4.09</td>
<td>100</td>
<td>100</td>
<td>3</td>
<td>0.43</td>
</tr>
<tr>
<td>Prostacyclin</td>
<td>S$_1$</td>
<td>4.98</td>
<td>92.3</td>
<td>99.6</td>
<td>5</td>
<td>0.98</td>
</tr>
<tr>
<td>$\alpha$-Gly$_8$</td>
<td>S$_1$</td>
<td>5.42</td>
<td>71.6</td>
<td>93.4</td>
<td>4</td>
<td>1.81</td>
</tr>
<tr>
<td>Leupeptin</td>
<td>S$_1$</td>
<td>4.27</td>
<td>46.7</td>
<td>91.0</td>
<td>3</td>
<td>7.08</td>
</tr>
<tr>
<td>Latanoprost</td>
<td>S$_1$</td>
<td>5.08</td>
<td>44.7</td>
<td>58.5</td>
<td>3</td>
<td>8.12</td>
</tr>
<tr>
<td>Metenkephalin</td>
<td>S$_1$</td>
<td>4.78</td>
<td>42.3</td>
<td>48.9</td>
<td>3</td>
<td>58.9</td>
</tr>
<tr>
<td>Dyad</td>
<td>S$_1$</td>
<td>3.10</td>
<td>82.3</td>
<td>88.7</td>
<td>4</td>
<td>...</td>
</tr>
<tr>
<td></td>
<td>S$_2$</td>
<td>3.47</td>
<td>100</td>
<td>100</td>
<td>8\textsuperscript{b}</td>
<td>...</td>
</tr>
<tr>
<td></td>
<td>S$_3$</td>
<td>3.51</td>
<td>100</td>
<td>100</td>
<td>6\textsuperscript{b}</td>
<td>...</td>
</tr>
<tr>
<td></td>
<td>S$_4$</td>
<td>3.71</td>
<td>100</td>
<td>100</td>
<td>6\textsuperscript{b}</td>
<td>...</td>
</tr>
<tr>
<td></td>
<td>S$_5$</td>
<td>4.06</td>
<td>100</td>
<td>100</td>
<td>6\textsuperscript{b}</td>
<td>...</td>
</tr>
</tbody>
</table>

\textsuperscript{a} All errors in the LoFEx excitation energies with respect to canonical RI-CC2 values are below 0.01 eV.

\textsuperscript{b} Extra iterations due to the calculation of CC2 NTOs.

\textbf{D. A difficult case: The phenothiazine-isoalloxazine dyad}

As reported by Kats and Schütz in Ref. 11, the few lowest transitions of the phenothiazine-isoalloxazine dyad are particularly difficult to describe at the CC2 level and especially with local approximations. The problem originates from the very poor starting guess that non-correlated methods (CCSD/TDHF) provide for the CC2 calculation in that particular case. This observation could indicate that CC2 is not appropriate for describing the electronic spectrum of this molecule and that larger basis sets (double-zeta basis sets were used) as well as higher level models (CCSD or CC3) might provide significantly different results. However, it is interesting to see how the LoFEx method is performing with such a difficult case. In particular, one of the lowest transitions of the dyad is known to exhibit a CT character which is challenging for many computational methods and in particular for TD-DFT.
Eq. (18) and the associated discussion]. This did not occur for any of the other molecules in Table I and it is a confirmation of the difference between the optimized TDHF and CC2 states. In general, we have observed that the use of CC2 NTOs does not reduce the size of the optimized XOS which is why CC2 NTOs are only used when the TDHF NTOs do not represent properly the transition of interest. With the geometry of the dyad used in this paper (which is different from the one used in Ref. 11), the CT transition corresponds to the fourth excitation energy and is properly described with LoFEx. However, with the parameters used in the black-box algorithm, it does require that the XOS contains the full orbital space, and consequently, no savings are obtained.

Even though the full orbital space is included for most of the transitions of the dyad, this application of LoFEx shows that the method provides correct CC2 results even for difficult cases such as CT transitions. We also note that significant savings are expected if the dyad is considered in a larger environment (e.g., including solvent).

IV. CONCLUSION AND PERSPECTIVES

We have introduced the LoFEx approach for the calculation of vertical CC excitation energies where the CC amplitude and Jacobian eigenvalue equations are solved in a transition-specific reduced orbital space (the XOS) composed of NTOs and LMOs. The method can be used as a black-box where the error in the excitation energies is controlled by a single parameter (\(\tau_{\text{XOS}}\)). LoFEx was tested on a set of molecules, including small and simple systems as well as larger and more real-life compounds. It was demonstrated at the RI-CC2 level that significant savings can be obtained without loss of accuracy. The computational cost of the method is dominated by the last RI-CC2 calculation using the largest XOS and is expected to be constant with the system size for a given type of transition. For large enough systems, and in particular when the excitation of interest is local compared to the size of the molecule, the cost of the method will be dominated by the underlying TDHF calculation on the full system, resulting in excitation energies of CC2 quality with a computational cost comparable to TD-DFT.

LoFEx relies on TDHF to provide a qualitatively correct description of the excitation process, i.e., the main characteristics of the transition should be described by the NTOs built from the TDHF transition density matrix. If the TDHF and CC2 spectra are significantly different, CC2 NTOs are constructed in the course of the calculation. This results in a more appropriate mixed orbital space which improves the description of the CC2 spectrum. We have shown that LoFEx can handle the description of complex electronic spectra exemplified with the phenothiazine-isoalloxazine dyad molecule for which the CT transition was described properly. This was however achieved without cost-reduction and a more elaborate algorithm is required to achieve that goal. Future works will thus be focused on the development of a more efficient RI-CC2 algorithm and on improving the orbital space optimization procedure together with more extensive testing of LoFEx. The calculation of transition moments and oscillator strengths will also be investigated.

The LoFEx procedure can in principle be straightforwardly applied to more accurate CC models (CCSD, CC3, etc.). However, the size of the XOS needed for a proper description of most transitions would probably prevent most practical applications without introducing further approximations on the underlying CC models. The dependence on the NTOs could also be an issue if CCSD or CC3 are used to describe doubles dominated transitions, and a generalization of the LoFEx strategy would be necessary.

LoFEx is conceptually and practically simple and can be seen as a general framework to select and optimize the orbital spaces needed to provide accurate descriptions of electronic transitions. For a given XOS of orthogonal occupied and virtual orbitals provided by the LoFEx scheme, any single reference CC code can be applied. We could, for instance, imagine combinations with PNO algorithms to further reduce the cost of the solution of the CC equations.\(^8\)

Another possibility is to combine LoFEx with hybrid quantum mechanics and molecular mechanics (QM/MM) techniques, like the polarizable-embedding approach,\(^43,44\) where a large quantum mechanical part could be treated by LoFEx, while the environment effects would be described at the molecular mechanics level.

ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Research Council under the European Union’s Seventh Framework Programme (No. FP/2007-2013)/ERC Grant Agreement No. 291371.

The numerical results presented in this work were performed at the Centre for Scientific Computing, Aarhus (http://phys.au.dk/forskning/cscaa/).

This research also used resources of the Oak Ridge Leadership Computing Facility at Oak Ridge National Laboratory, which is supported by the Office of Science of the Department of Energy under Contract No. DE-AC05-00OR22725.

(2011).
(2012).
26 P. Ettenhuber, P. Baudin, T. Kjærgaard, P. Jørgensen, and K. Kristensen,
Phys. 144, 054102 (2016).
32 H. Koch, O. Christiansen, P. Jørgensen, and J. Olsen, Chem. Phys. Lett. 244,
104108 (2011).
34 T. Schwabe, K. Sneskov, J. M. Haugaard Olsen, J. Kongsted, O. Christiansen,