Efficient implementation of molecular CCSD gradients with Cholesky-decomposed electron repulsion integrals

Schnack-Petersen, Anna Kristina; Koch, Henrik; Coriani, Sonia; Kjønstad, Eirik F.

Published in:
Journal of Chemical Physics

Link to article, DOI:
10.1063/5.0087261

Publication date:
2022

Document Version
Peer reviewed version

Accepted to J. Chem. Phys. 10.1063/5.0087261

Efficient implementation of molecular CCSD gradients with Cholesky-decomposed electron repulsion integrals

Anna Kristina Schnack-Petersen,[1] Henrik Koch,[2,3] Sonia Coriani,[1,3] and Eirik F. Kjønstad[3, a)]
[1)] *Department of Chemistry, Technical University of Denmark, 2800 Kongens Lyngby, Denmark*
[2)] *Scuola Normale Superiore, Piazza dei Cavaleri 7, 56126 Pisa, Italy*
[3)] *Department of Chemistry, Norwegian University of Science and Technology, 7491 Trondheim, Norway*

(Dated: 6 May 2022)

We present an efficient implementation of ground and excited state CCSD gradients based on Cholesky-decomposed electron repulsion integrals. Cholesky decomposition, like density-fitting, is an inner projection method, and thus similar implementation schemes can be applied for both methods. One well-known advantage of inner projection methods, which we exploit in our implementation, is that one can avoid storing large V^3O and V^4 arrays by instead considering three-index intermediates. Furthermore, our implementation does not require the formation and storage of Cholesky vector derivatives. The new implementation is shown to perform well, with less than 10% of the time spent calculating the gradients in geometry optimizations. Furthermore, the computational time spent per optimization cycle is significantly lower compared to other implementations based on an inner projection method. We showcase the capabilities of the implementation by optimizing the geometry of the retinal molecule ($C_{20}H_{28}O$) at the CCSD/aug-cc-pVDZ level of theory.

I. INTRODUCTION

The gradient of the electronic energy with respect to the nuclear coordinates, known as the molecular gradient, is a particularly useful quantity in computational chemistry. It is essential for determining both local energy minima and equilibrium geometries[1] and thus for predicting the stability and structure of molecular systems, as well as molecular properties at equilibrium. In addition, the molecular gradient is essential for locating transition state geometries, which can aid in elucidating chemical reaction paths and in estimating reaction rates.[2] Moreover, molecular gradients are required for predicting the time evolution of molecular systems, since the gradient provides the forces that act on the atomic nuclei in the absence of sizeable non-adiabatic effects.[3,4]

Over the past three decades, coupled cluster (CC) methods have gained popularity[5–9] due to their high and systematically improvable accuracy. Today, they are widely considered the most efficient for describing dynamical correlation whenever the ground state is dominated by a single determinant.[9,10] Coupled cluster methods that include approximate triple excitations (such as CC3[11]) are generally regarded as the state-of-the-art in computational chemistry, but they are still too costly for molecular systems with more than about fifteen second-row elements.[12] Nonetheless, coupled cluster calculations are becoming increasingly feasible, particularly for methods that include double excitations either approximately, such as CC2,[13] or in full, that is, CCSD.[14]

It is therefore of considerable interest to develop efficient implementations of molecular gradients at the CCSD level of theory, both for the ground and the excited states. Implementations of such gradients already exist in a number of programs. Ground state gradients are available in commercial codes such as Q-Chem[15,16] and Gaussian[17] and in open-source programs such as Psi4[18,19] and Dalton,[20,21] as well as in free programs such as CFOUR[22,23] and MRCC.[24,25] However, out of the above-mentioned programs, only Q-Chem, CFOUR, and MRCC include excited state gradients at the CCSD level.

Cholesky decomposition (CD) of the electron repulsion integrals[26–34] has become a valuable tool for efficient implementations in quantum chemistry. Due to the rank-deficiency of the integral matrix, CD implies significantly reduced computational requirements, both in terms of storage and the number of floating-point operations. The CD method dates back to the 1970s,[26] but it has seen a resurgence of interest over the past decades due to improvements in algorithms[29,35,36] and computer hardware. These developments have made CD competitive with the prevailing inner projection method, the resolution-of-identity (RI) or density-fitting method.[32,34,37–46] As a result, there is currently a demand for efficient CD-based coupled cluster implementations (e.g., for molecular gradients). To date, however, such implementations are still rather scarce. Indeed, out of the programs mentioned above, only Q-Chem offers an implementation of CCSD gradients based on Cholesky-decomposed integrals.[16]

The factorized form of the electron repulsion integrals, obtained by inner projection, has an interesting implication for molecular gradient algorithms. Two- and three-index density intermediates naturally arise, allowing one to avoid storing the V^3O and V^4 blocks of the density matrix. This has been exploited both in CD (for CASSCF)[35] and in RI (for CCSD).[19] The equivalence between CD and RI means that RI implementations can

[a)] Electronic mail: eirik.kjonstad@ntnu.no

be adapted to the framework of CD, which is the focus of this work. This strategy was first employed by Aquilante *et al.*[31] for molecular gradients in density functional theory. An alternative algorithm for CD-CCSD gradients was recently suggested by Feng *et al.*[16] and implemented in Q-Chem.[15] This algorithm similarly employs a strategy where V^3O and V^4 terms are avoided,[16,33] but it relies on the calculation and temporary storage of Cholesky vector derivatives. This implies a relatively large $\mathrm{O}(N^4)$ storage requirement, thereby imposing a limitation on the system size. Feng *et al.* report that for a system with 418 basis functions, 0.9 GB and 91 GB RAM were required to compute the Cholesky vectors and their derivatives, respectively.[16] While these vectors are processed and released directly after decomposition, the procedure still needs that a large amount of memory is available. The requirement can, in principle, be reduced to $\mathrm{O}(N^3)$ by utilizing the sparsity of the Cholesky vectors in the atomic orbital (AO) basis. Another approach, which we employ here, is to calculate the required terms on the fly, eliminating any need to store the derivatives.

This paper describes a new and efficient implementation of the ground and excited state CCSD gradients, which we have incorporated into a development version of the open-source program e^T (version 1.6).[47] This implementation is partly based on the one reported by Bozkaya and Sherrill for RI-CCSD,[19] where the gradient is constructed from two- and three-index density intermediates; for these intermediates, see also the CD-CASSCF implementation by Delcey *et al.*[35] Our implementation makes use of the recent two-step CD algorithm by Folkestad *et al*[36] and is well-suited for large-scale applications. In particular, as already noted, derivative integrals involving the auxiliary basis are calculated on the fly, ensuring that no $\mathrm{O}(N^4)$ storage requirements are associated with the Cholesky vectors.

II. THEORY

A. Analytical expression for the molecular gradient

Using a closed-shell, spin-restricted formulation, the molecular CCSD gradient is conveniently derived from the Lagrangian

$$\begin{aligned}\mathcal{L} &= E + \sum_{\mu\neq 0} \bar{\zeta}_\mu \Omega_\mu + \sum_{p\leq q} \bar{\kappa}_{pq}(F_{pq} - \delta_{pq}\epsilon_p) + \bar{\lambda}\mathcal{O} \\ &= \sum_{pq} h_{pq} D_{pq} + \sum_{pqrs} g_{pqrs} d_{pqrs} + \sum_{\mu\neq 0} \bar{\zeta}_\mu \Omega_\mu \\ &\quad + \sum_{p\leq q} \bar{\kappa}_{pq}(F_{pq} - \delta_{pq}\epsilon_p) + \bar{\lambda}\mathcal{O}.\end{aligned} \tag{1}$$

In this expression, E is the energy, h_{pq} and g_{pqrs} are the one- and two-electron integrals associated with the Hamiltonian H, and D_{pq} and d_{pqrs} are the one- and two-electron densities. Lagrangian multipliers are denoted with a bar ($\bar{\zeta}_\mu, \bar{\kappa}_{pq}, \bar{\lambda}$). Furthermore,

$$\Omega_\mu = \langle \mu | \bar{H} | \mathrm{HF} \rangle, \tag{2}$$

where $|\mathrm{HF}\rangle$ is the Hartree-Fock state and

$$\bar{H} = \exp(-T)\exp(\kappa) H \exp(-\kappa)\exp(T). \tag{3}$$

Here κ denotes the orbital rotation operator, where it is implied that $\kappa = 0$ at the nuclear geometry where the derivative is to be evaluated.[21,48] The cluster operator is $T = \sum_\mu t_\mu \tau_\mu$, where τ_μ is an excitation operator and excited configurations are denoted as $|\mu\rangle = \tau_\mu|\mathrm{HF}\rangle$; in addition, $\tau_0 = \mathbb{I}$ and hence $|0\rangle = |\mathrm{HF}\rangle$. Moreover, the Fock matrix is given as

$$F_{pq} = h_{pq} + \sum_k (2g_{pqkk} - g_{pkkq}) \tag{4}$$

and ϵ_p denotes the energy of the p'th molecular orbital (MO). Finally, $\mathcal{O}$ is defined such that $\mathcal{O} = 0$ ensures normalization, i.e., that the left and right coupled cluster states are binormal. This term is described in more detail below.

The integrals in the Fock matrix (F_{pq}) are always expressed in the MO basis, whereas the integrals in the coupled cluster energy,

$$E = \sum_{pq} h_{pq} D_{pq} + \sum_{pqrs} g_{pqrs} d_{pqrs}, \tag{5}$$

are either expressed in the MO basis or in the T_1-transformed basis. Above and throughout, p, q, r, and s are used to denote generic MOs; i, j, k, and l denote occupied MOs; and a, b, c, and d denote virtual MOs.

The expressions for the energy E and the normalization condition $\mathcal{O}$ depend on whether we are considering the ground state or an excited state. In particular,

$$E^{\mathrm{GS}} = \langle \mathrm{HF} | \bar{H} | \mathrm{HF} \rangle \tag{6}$$

$$E^{\mathrm{ES}} = \sum_{\mu\nu} L_\mu \langle \mu | \bar{H} | \nu \rangle R_\nu \tag{7}$$

and

$$\mathcal{O}^{\mathrm{GS}} = 0 \tag{8}$$

$$\mathcal{O}^{\mathrm{ES}} = 1 - \sum_\mu L_\mu R_\mu, \tag{9}$$

where L_μ and R_μ denote the left and right excited state amplitudes. For the ground state, normalization is automatically fulfilled,[10] and can be ignored as $\mathcal{O}^{\mathrm{GS}} = 0$, which effectively removes the normalization condition from the Lagrangian.

The stationarity conditions with respect to $\bar{\zeta}_\mu$ and $\bar{\kappa}_{pq}$ are, respectively, the well-known amplitude and canonical Hartree-Fock equations, $\Omega_\mu = 0$ and $F_{pq} = \delta_{pq}\epsilon_p$. Furthermore, the stationarity condition with respect to $\bar{\lambda}$ enforces the biorthonormality constraint. The orbital

rotation multipliers $\bar{\kappa}_{pq}$ are determined from the stationarity condition with respect to the orbital rotation parameters κ_{pq},[21,48,49]

$$\frac{\partial \mathcal{L}}{\partial \kappa_{pq}} = 0 \iff \bar{\boldsymbol{\kappa}}\, {}^{\bar{\kappa}}\boldsymbol{A} = -\,{}^{\bar{\kappa}}\boldsymbol{\eta}. \tag{10}$$

Here,

$$\begin{aligned} {}^{\bar{\kappa}}A_{pqrs} = {} & 2\delta_{pr}\delta_{qs}(\epsilon_p - \epsilon_q) \\ & + (v_s - v_r)(8g_{pqrs} - 2g_{prqs} - 2g_{psrq}) \end{aligned} \tag{11}$$

is the Hartree-Fock Hessian, with v_r denoting the occupancy (0 or 1) of orbital r in the Hartree-Fock state.[50] In the case of CCSD, ${}^{\bar{\kappa}}\boldsymbol{\eta}$ is given as[21,50]

$$\begin{aligned} {}^{\bar{\kappa}}\eta_{pq} = {} & \sum_t h_{pt}(D_{tq} + D_{qt}) - \sum_t h_{qt}(D_{tp} + D_{pt}) \\ & + \sum_{rst} g_{ptrs}(d_{tqrs} + d_{qtrs}) - \sum_{rst} g_{qtrs}(d_{tprs} + d_{ptrs}). \end{aligned} \tag{12}$$

Similarly, the amplitude multipliers $\bar{\zeta}_\mu$ are determined from the stationarity condition with respect to the coupled cluster amplitudes t_μ,

$$\frac{\partial \mathcal{L}}{\partial t_\mu} = 0, \tag{13}$$

which yields different equations for the ground and the excited states. In particular,

$$\bar{\boldsymbol{\zeta}}^{\mathrm{GS}}\boldsymbol{A} = -\boldsymbol{\eta} \tag{14}$$

$$\bar{\boldsymbol{\zeta}}^{\mathrm{ES}}\boldsymbol{A} = -\boldsymbol{\eta} - \boldsymbol{J}\boldsymbol{A} - \boldsymbol{F}(\boldsymbol{L})\boldsymbol{R}, \tag{15}$$

with

$$A_{\mu\nu} = \langle \mu | [\bar{H}, \tau_\nu] | \mathrm{HF} \rangle \tag{16}$$

$$\eta_\mu = \langle \mathrm{HF} | [\bar{H}, \tau_\mu] | \mathrm{HF} \rangle \tag{17}$$

$$J_{ai} = \sum_{bj} L_{ij}^{ab} R_j^b \tag{18}$$

$$F(L)_{\mu\nu} = \sum_\lambda L_\lambda \langle \lambda | [[\bar{H}, \tau_\mu], \tau_\nu] | \mathrm{HF} \rangle. \tag{19}$$

For the ground state, we obtain the ground state multiplier equations, Eq. (14), and we will let $\bar{\boldsymbol{\zeta}}^{\mathrm{GS}} = \bar{\boldsymbol{t}}$, following the conventional notation for these multipliers.[10] For the excited states, Eq. (15) is obtained. The excited state multipliers, referred to below as the amplitude response, will similarly be denoted as $\bar{\boldsymbol{\zeta}}^{\mathrm{ES}} = \bar{\boldsymbol{t}}^{\mathrm{ES}}$.

With all orbital and wave function parameters variationally determined, the gradient can now be evaluated as the first partial derivative of the Lagrangian with respect to the nuclear coordinates. As is well known, this derivative can be written as[21,48]

$$\mathcal{L}^{(1)} = \sum_{pq} h_{pq}^{(1)} D_{pq} + \sum_{pqrs} g_{pqrs}^{(1)} d_{pqrs} + \sum_{p \le q} \bar{\kappa}_{pq} F_{pq}^{(1)}, \tag{20}$$

where $h_{pq}^{(1)}$ and $g_{pqrs}^{(1)}$ denote one- and two-electron derivative integrals and where $F_{pq}^{(1)}$ denotes the derivative Fock matrix. The one- and two-electron derivative integrals are here evaluated in an orthonormal MO (OMO) basis, i.e., a basis strictly orthonormal at all geometries.[51,52] These OMOs are obtained from the nonorthogonal unmodified MOs (UMOs), which are defined from the AOs at the displaced geometry and the MO coefficients of the unperturbed geometry. The orthonormalization matrix that transforms UMOs to OMOs defines an orbital connection and is known as the connection matrix.[52]

Here we will use the symmetric connection, for which the connection matrix is given as the inverse square root of the UMO overlap matrix.[51] For this connection, the derivative of the OMO Hamiltonian can be written as

$$H^{(1)} = H^{[1]} - \frac{1}{2}\{S^{[1]}, H\}, \tag{21}$$

where $H^{[1]}$ and $S^{[1]}$ denote derivatives of the Hamiltonian and of the overlap in the UMO basis. The second term of Eq. (21) is known as the reorthonormalization term. The UMO derivatives are evaluated by differentiating the AO integrals and then transforming them to the UMO basis. The notation $\{A, B\}$ means the one-index transformation of operator B by the transformation matrix A:[51]

$$\{A, B\}_{pq} = \sum_t (A_{pt}B_{tq} + A_{qt}B_{pt}) \tag{22}$$

$$\begin{aligned} \{A, B\}_{pqrs} = {} & \sum_t (A_{pt}B_{tqrs} + A_{qt}B_{ptrs} \\ & + A_{rt}B_{pqts} + A_{st}B_{pqrt}) \end{aligned} \tag{23}$$

The expression for $F^{(1)}$ is similar to that for $H^{(1)}$ in Eq. (21) and is omitted. See Refs. 51 and 52 for further details on orbital connections.

We will begin by deriving expressions for the UMO contributions to the gradient, that is, the contributions originating from the first term in Eq. (21). The reorthonormalization contributions are presented separately, see Section II D.

The first term in Eq. (20) has been thoroughly described in other works for the CCSD case, e.g. by Scheiner *et al.*[53] The second and third terms are not trivial and will be described in more detail. For the second term, the CCSD two-electron densities are required. Expressions for both ground and excited state densities have been rederived and are given in Appendix A. Evaluating this term also requires that we consider the electron repulsion integrals. Below, unless otherwise specified, these integrals are expressed in the T_1-transformed basis, and hence the densities have been made independent of T_1; see Eq. (A1). In the T_1-transformed basis, the Hamiltonian integrals can be written

$$h_{pq} = \sum_{rs} x_{pr} y_{qs} h_{rs}^{\mathrm{MO}} \tag{24}$$

$$g_{pqrs} = \sum_{tumn} x_{pt} y_{qu} x_{rm} y_{sn} g^{\mathrm{MO}}_{tumn}, \tag{25}$$

where $\boldsymbol{x} = \boldsymbol{I} - \boldsymbol{t}_1$ and $\boldsymbol{y} = \boldsymbol{I} + \boldsymbol{t}_1^T$, and where $\boldsymbol{h}^{\mathrm{MO}}$ and $\boldsymbol{g}^{\mathrm{MO}}$ denote integrals expressed in the MO basis.[10]

To find expressions for the derivatives of g_{pqrs}, we expand the integral matrix in terms of its Cholesky decomposition. That is, we write

$$\begin{aligned} g_{pqrs} = (pq|rs) &= \sum_{KL} (pq|K)(\boldsymbol{\mathcal{S}}^{-1})_{KL}(L|rs) \\ &= \sum_{J} L^J_{pq} L^J_{rs}, \end{aligned} \tag{26}$$

where

$$\mathcal{S}_{KL} = (K|L) \tag{27}$$

and where K and L are elements in the Cholesky basis.[26,30] The Cholesky decomposition of $\boldsymbol{\mathcal{S}}$ defines the $\boldsymbol{Q}$ matrix, from which we can evaluate the inverse of $\boldsymbol{\mathcal{S}}$:

$$\boldsymbol{\mathcal{S}} = \boldsymbol{Q}\boldsymbol{Q}^T \implies \boldsymbol{\mathcal{S}}^{-1} = \boldsymbol{Q}^{-T}\boldsymbol{Q}^{-1}, \quad \boldsymbol{Q}^{-T} = (\boldsymbol{Q}^{-1})^T \tag{28}$$

This yields the definition of the Cholesky vectors:

$$L^J_{pq} = \sum_K (pq|K) Q^{-T}_{KJ}. \tag{29}$$

The Cholesky vectors are also expressed in the T_1-transformed basis, where

$$L^J_{pq} = \sum_{rs} x_{pr} y_{qs} (L^J_{rs})^{\mathrm{MO}}. \tag{30}$$

Here $(L^J_{rs})^{\mathrm{MO}}$ denotes the Cholesky vectors in the MO basis. From the above definitions, we can write the derivative two-electron integrals as

$$\begin{aligned} (pq|rs)^{[1]} = &\sum_K (pq|K)^{[1]} Z^K_{rs} + \sum_L (rs|L)^{[1]} Z^L_{pq} \\ &- \sum_{MN} Z^M_{pq} \mathcal{S}^{[1]}_{MN} Z^N_{rs}, \end{aligned} \tag{31}$$

where we have defined

$$Z^K_{rs} = \sum_L (\boldsymbol{\mathcal{S}}^{-1})_{KL}(L|rs) \tag{32}$$

and used the identity

$$(\boldsymbol{\mathcal{S}}^{-1})^{[1]} = -\boldsymbol{\mathcal{S}}^{-1}\boldsymbol{\mathcal{S}}^{[1]}\boldsymbol{\mathcal{S}}^{-1} \tag{33}$$

Upon contraction with the two-electron density, the second term of Eq. (20) becomes

$$\sum_{pqrs} d_{pqrs}(pq|rs)^{[1]} = 2\sum_{pqK}(pq|K)^{[1]} W^K_{pq} - \sum_{MN} V_{MN}\mathcal{S}^{[1]}_{MN} \tag{34}$$

with

$$W^K_{pq} = \sum_{rs} d_{pqrs} Z^K_{rs} \tag{35}$$

$$V_{MN} = \sum_{pq} Z^M_{pq} W^N_{pq}. \tag{36}$$

The first term in Eq. (34) is more conveniently calculated in the non-transformed basis by transferring the T_1-terms back to W^K_{pq},

$$(W^K_{pq})^{\mathrm{MO}} = \sum_{rs} x_{rp} y_{sq} W^K_{rs}, \tag{37}$$

before contracting with the differentiated MO integrals, $(pq|K)^{[1],\mathrm{MO}}$.

B. Two-electron density intermediates

Expressions for the various blocks of the two-electron density are reported in Appendix A. In this section we describe in detail the two- and three-index density intermediates. All contributions to W^J_{pq} from the O^4, O^3V, and O^2V^2 density blocks are constructed straightforwardly by contracting the density block with Z^J_{pq}; hence, we will not discuss them further. To avoid storing the V^4 and OV^3 blocks of the density in memory, and, in addition, to avoid batching when constructing the OV^3 terms, we directly build their contributions to W^J_{pq} and store these instead. For improved readability, Einstein's implicit summation over repeated indices will be used in the remainder of this section. The contributions from the OV^3-density blocks to the gradient are

$$\begin{aligned} d_{abci}(ab|ci)^{[1]} = &(ab|K)^{[1]} W^K_{ab} \\ &+ (ci|L)^{[1]} W^L_{ci} - V_{MN}\mathcal{S}^{[1]}_{MN} \end{aligned} \tag{38}$$

$$\begin{aligned} d_{abic}(ab|ic)^{[1]} = &(ab|K)^{[1]} W^K_{ab} \\ &+ (ic|L)^{[1]} W^L_{ic} - V_{MN}\mathcal{S}^{[1]}_{MN}. \end{aligned} \tag{39}$$

We thus directly construct the contributions to W^K_{ci} and W^K_{ic}, as well as to W^K_{ab} and V_{MN}, from the two OV^3 blocks of the two-electron density. From d_{abci} we get the contributions for the excited state

$$W^K_{ab} = R^b_j X^K_{aj} \tag{40}$$

$$W^K_{ci} = L^{ac}_{ji} P^K_{aj} \tag{41}$$

$$V_{MN} = Z^M_{ab} W^N_{ab} + Z^M_{ci} W^N_{ci}, \tag{42}$$

where L^{ac}_{ji} and R^b_j denote excited state amplitudes, see Appendix A. For the ground state, all contributions from d_{abci} are zero. From d_{abic} we similarly obtain

$$W^K_{ab} = L^a_m \tilde{Y}^K_{bm}(T_2) R_0 \tag{43}$$

$$W^K_{ic} = \tilde{t}^{cb}_{im} R_0 O^K_{bm} \tag{44}$$

$$V_{MN} = Z^M_{ab} W^N_{ab} + Z^M_{ic} W^N_{ic} \tag{45}$$

for the ground state, and

$$\begin{aligned} W_{ab}^K &= L_m^a \tilde{Y}_{bm}^K(\tilde{R}_2) \\ &\quad + 2X_{ab}(L_2,T_2)P^K - O_{ai}^J R_i^b \\ &\quad - (V_{akbi}(L_2,T_2) + Y_{akbi}(L_2,T_2))P_{ik}^K + 2Q_{ak}^K R_k^b \\ &\quad + W_{am}\tilde{Y}_{bm}^K(T_2) - K_{icab}Z_{ic}^K \end{aligned} \tag{46}$$

$$\begin{aligned} W_{ic}^K &= (2\tilde{R}_{mi}^{bc} - \tilde{R}_{mi}^{cb})O_{bm}^K \\ &\quad + 2C^K R_i^c - X_{ac}(L_2,T_2)P_{ai}^K \\ &\quad - K_{ik}^K R_k^c + 2Y_{akci}(L_2,T_2)P_{ak}^K \\ &\quad + U_{bm}^K \tilde{t}_{mi}^{bc} - S_{iacb}Z_{ab}^K \end{aligned} \tag{47}$$

$$V_{MN} = Z_{ab}^M W_{ab}^N + Z_{ic}^M W_{ic}^N \tag{48}$$

for the excited state. The intermediates introduced in these contributions are given in Table I.

The V^4 density contribution to W_{ab}^J is also evaluated directly as a contraction between the density and Z_{cd}^K, albeit with batching. This contribution is the steepest-scaling term in the gradient and implies the calculation of the density contribution

$$O_{abcd}^{(2)} = L_{ij}^{ac} C_{ij}^{bd}. \tag{49}$$

To evaluate this term as written would have a cost of O^2V^4. However, it is possible to reduce the cost by a factor of four by adapting the well-known strategy for constructing the A2 term of $\mathbf{\Omega}$.[54] In the case of Eq. (49), we form the symmetric and anti-symmetric combinations of L_2 and C_2:

$$X_{ij}^{ab\pm} = X_{ij}^{ab} \pm X_{ji}^{ab}, \quad X = L_2, C_2. \tag{50}$$

Then, by contracting the symmetric and anti-symmetric terms separately and adding them together, we need not loop over all indices, but merely $i \geq j$, $a \geq c$, and $b \geq d$, leading to an eight-fold reduction in cost. However, since this must be done twice (for symmetric and anti-symmetric terms), the net reduction in cost is a factor of four.

TABLE I. Intermediates used in constructing the W_{ab}^K, W_{ci}^K, and W_{ic}^K density intermediates. Observe that C_2 denotes a set of double amplitudes, and $\tilde{C}_{ij}^{ab} = 2C_{ij}^{ab} - C_{ji}^{ab}$. Also note that $\tilde{R}_2$ is a redefinition of the R_2 amplitudes (see Appendix A). In the expressions given in this table, $C_2 = T_2$. The definition of C_2 may be different as C_2 is merely a placeholder; see also Appendix A. Summation over repeated indices is assumed.

$\tilde{t}_{ij}^{ab}$	$= 2t_{ij}^{ab} - t_{ji}^{ab}$
W_{ai}	$= L_{il}^{ad} R_{dl}$
$X_{ab}(L_2,C_2)$	$= L_{kl}^{ad} C_{kl}^{bd}$
$Y_{ajbi}(L_2,C_2)$	$= L_{jl}^{ad} C_{il}^{bd}$
$V_{ajbi}(L_2,C_2)$	$= L_{kj}^{ac} C_{ki}^{bc}$
S_{iacb}	$= B_{iamn} t_{mn}^{bc}$
B_{iamn}	$= L_{mn}^{ad} R_i^d$
K_{ik}^K	$= (V_{akbi}(L_2,T_2) + Y_{akbi}(L_2,T_2))Z_{ab}^K$
X_{aj}^K	$= L_{ji}^{ac} Z_{ci}^K$
P_{aj}^K	$= R_j^b Z_{ab}^K$
$\tilde{Y}_{bm}^K(C_2)$	$= \tilde{C}_{im}^{cb} Z_{ic}^K$
P^K	$= R_i^c Z_{ic}^K$
O_{ai}^J	$= X_{ac}(L_2,T_2) Z_{ic}^K$
P_{ik}^K	$= R_k^c Z_{ic}^K$
Q_{ak}^K	$= Y_{akci}(L_2,T_2) Z_{ic}^K$
K_{icab}	$= (L_{mn}^{ad} R_i^d) t_{mn}^{bc}$
Q^K	$= X_{ab}(L_2,T_2) Z_{ab}^K$
P_{ai}^K	$= R_i^b Z_{ab}^K$
U_{bm}^K	$= W_{am} Z_{ab}^K$

C. Orbital relaxation contributions

In order to obtain the gradient, all three terms of Eq. (20) must be evaluated. So far, we have not yet discussed the third term. First, the $\bar{\kappa}$ parameters are determined from the Z-vector equation given in Eq. (10). Since CCSD is orbital invariant, we only consider the VO block of the $\bar{\kappa}$ vector.[21] The UMO contribution to the orbital relaxation then reads

$$\sum_{ai} \bar{\kappa}_{ai} F_{ai}^{[1]} = \sum_{ai} \bar{\kappa}_{ai} \Big(h_{ai}^{[1]} + \sum_j (2g_{aijj}^{[1]} - g_{ajji}^{[1]}) \Big). \tag{51}$$

The integrals are here expressed in the MO basis. For efficiency, the second and third terms are rewritten by using the Cholesky decomposition:

$$\begin{aligned} \sum_{aij} \bar{\kappa}_{ai}(2g_{aijj}^{[1]} - g_{ajji}^{[1]}) &= \sum_{aiK} K_{ai}^K (ai \mid K)^{[1]} \\ &\quad + \sum_{jK} L^K (K \mid jj)^{[1]} + \sum_{ijK} N_{ji}^K (K \mid ji)^{[1]} \\ &\quad - \sum_{KL} M_{KL} S_{KL}^{[1]} + \frac{1}{2} \sum_{KL} O_{KL} S_{KL}^{[1]}. \end{aligned} \tag{52}$$

The introduced intermediates are given in Table II. The first three terms are added to the $W_{pq}^J g_{pqrs}^{[1]}$ term, while the remaining terms are added to the $V_{MN} S_{MN}^{[1]}$ term; see Eq. (34).

Accepted to J. Chem. Phys. 10.1063/5.0087261

TABLE II. Intermediates used in constructing the $\sum_{aij} \bar{\kappa}_{ai}(2g^{[1]}_{aijj} - g^{[1]}_{ajji})$. Here Z^K_{pq} are expressed in the MO basis.

$K^K_{ai} = \sum_j \bar{\kappa}_{ai}(2Z^K_{jj} - Z^K_{ji})$
$L^K = 2\sum_{ai} \bar{\kappa}_{ai} Z^K_{ai}$
$N^K_{ji} = -\sum_a \bar{\kappa}_{ai} Z^K_{aj}$
$M_{KL} = \sum_j L^K Z^L_{jj}$
$O_{KL} = \sum_{ij} N^K_{ji} Z^L_{ji}$

To determine $\bar{\kappa}_{ai}$, we also need the right-hand-side of the Z-vector equation. This vector is conveniently constructed from a three-index intermediate $\tilde{W}^J_{pq} = W^J_{pq}(L)$, defined as W^J_{pq}, but constructed using L^J_{pq} rather than Z^J_{pq}. In terms of this intermediate, and using integrals expressed in the MO basis, we have[21]

$$^{\bar{\kappa}}\eta_{ai} = (1 - P_{ai})\Big(\sum_t \mathcal{D}_{ti} h_{at} + \sum_{tJ} \tilde{\mathcal{W}}^J_{ti} L^J_{at}\Big), \tag{53}$$

where $P_{ai}X_{ai} = X_{ia}$ and

$$\mathcal{D}_{pq} = D_{pq} + D_{qp} \tag{54}$$
$$\tilde{\mathcal{W}}^J_{pq} = \tilde{W}^J_{pq} + \tilde{W}^J_{qp}. \tag{55}$$

D. Reorthonormalization contributions

The second term in Eq. (21) gives rise to reorthonormalization contributions to the gradient. These contributions can be expressed as $(-\sum_{pq} \mathcal{F}_{pq} S^{[1]}_{pq})$, where

$$\mathcal{F}_{pq} = \sum_i (\mathcal{D}_{pi} h_{qi} + \sum_J \tilde{\mathcal{W}}^J_{pi} L^J_{qi}) + \mathcal{F}^{\bar{\kappa}}_{pq}. \tag{56}$$

The orbital relaxation contribution $\mathcal{F}^{\bar{\kappa}}_{pq}$ is described in detail in Appendix B. Note that all terms in Eq. (56) are given in the MO basis.

III. COMPUTATIONAL DETAILS

The CD-CCSD gradient for ground and excited states has been implemented in a development version of e^T 1.6. We apply the gradient implementation to determine equilibrium geometries for thymine, azobenzene, and retinal (see Figure 1), where we consider the ground state in all three systems and the lowest singlet excited state in thymine and azobenzene. To perform these calculations, we have also implemented an optimizer that uses the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm with the rational function (RF)[55] level shift.

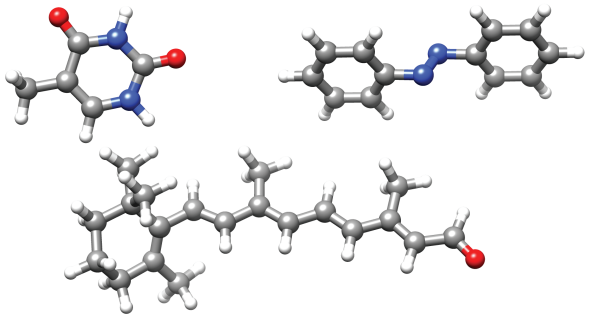

FIG. 1. Ground state geometries, optimized at the CCSD/aug-cc-pVDZ level, for thymine (top left), azobenzene (top right), and retinal (bottom).

This implementation makes use of the redundant internal coordinates introduced by Bakken and Helgaker,[56] along with the initial "simple model Hessian guess" proposed by the same authors.

For comparison, we carried out CD-CCSD calculations with Q-Chem 5.4[15] and RI-CCSD calculations with Psi4 1.3.[18] An aug-cc-pVDZ basis was used in all calculations. For consistency with our implementation, we disabled the frozen core approximation in Q-Chem and Psi4. Default thresholds were used in all calculations, except in the case of the CD convergence threshold. The reason for this is that a CD threshold of e.g. 10^{-3} can result in slow convergence because the Cholesky basis varies on the potential energy surface, causing discontinuities of the same order of magnitude as the CD threshold. This was also observed by both Aquilante *et al.*[31] and Feng *et al.*[16] Thus, in order to converge the gradient to 3×10^{-4} (the Baker convergence criterion), we employed a tighter CD threshold of 10^{-4} throughout. The one exception to this was for the large retinal molecule, where we instead applied a CD threshold of 10^{-3}.

The e^T program does not utilize point group symmetry, whereas this was enabled in Q-Chem and Psi4. However, the initial geometries does not possess point group symmetry, and this should therefore not affect the comparison of timings significantly. Initial and optimized molecular geometries can be found in Ref. 57. At the initial geometries, the excitation energies in thymine and azobenzene are 5.20 eV and 3.46 eV, respectively; at the optimized excited state geometries, the excitation energies are 3.98 eV and 2.35 eV. All calculations were performed on one node with two Intel Xeon E5-2699 v4 processors with 44 cores and given 1 TB of shared memory.

IV. RESULTS AND DISCUSSION

1. Timing comparisons of different implementations

To demonstrate the efficiency of our implementation, we compare calculation times for geometry optimizations

of a small system (thymine) and a medium-large system (azobenzene) with the Psi4 and Q-Chem programs. Excited state gradients are not available in Psi4 and thus excited state geometry optimizations have not been performed with this program. We report the percentage of the time spent determining the gradients for the e^T calculations. This is not reported for Q-Chem or Psi4, as this time is not directly available from their respective outputs.

Timings for e^T and Psi4 are given in Table III. From the results in the table we see that the calculation of the gradient amounts to only 7% of the total calculation time, illustrating the efficiency of the gradient implementation. Here, the time required to determine the multipliers is not included in the gradient time. This task may represent a significant part of the total time, but it is determined by the efficiency of the existing implementation of Jacobian transformations, rather than the construction of the gradient, which is the focus of this work. We observe that the calculation time required by e^T, per optimization cycle, is roughly half of that required by Psi4. This primarily reflects the difference in efficiency of the ground state implementation in the two programs.

When applying an inner projection method, the integral costs are proportional to the size of the auxiliary basis.[36] In the density-fitting scheme used in Psi4, the thymine calculation required 786 auxiliary basis functions, while 967 functions were used in our CD-based implementation. These two numbers are of the same order of magnitude; thus, the computational resources required should be comparable. This is also the case for the azobenzene calculations, in which Psi4 and e^T required 1238 and 1444 auxiliary basis functions, respectively. Hence, the two approaches are similar in terms of computational costs. However, note that the time spent on the Cholesky decomposition itself is negligible and that the resulting integral errors are strictly lower than the CD threshold (here 10^{-4}). Such strict error control is not possible with the density-fitting method.[36]

Timings for e^T and Q-Chem are given in Table IV. We now also consider excited state geometry optimizations, and observe once again that the calculation time is not dominated by the gradient. In fact, the calculation of the gradient amounts to 5% of the full calculation time for the excited state (compared to 7% for the ground state). Note that when reporting the time spent on determining the excited state gradients, we have again excluded the time needed to determine the amplitude response.

From Table IV it is furthermore observed that we can carry out ground state optimizations in roughly 50% of the time required by Q-Chem, or even less. The savings are significantly larger for ground state optimizations. This can be attributed to differences, between the two programs, in the efficiency of the ground and excited state implementations. The calculation time in Q-Chem can be significantly reduced with the utilization of the frozen core approximation (for timings, see SI), resulting in computation times for the excited state that are comparable to those observed in e^T without this approximation. For the ground state, however, our new implementation still offers significant time savings, despite frozen core calculations being inherently less computationally demanding.

The reported comparisons were carried out without enforcing point-group symmetry. As also noted above, further improvements to the Q-Chem and Psi4 timings could have been obtained by starting from geometries with point-group symmetry. Further gains are also possible with a single-precision execution.[58]

2. Comparisons of Obtained Geometries for Small Molecules

In order to investigate the accuracy of our CD-CCSD implementation compared to standard CCSD, optimized bond lengths and bond angles have been determined for 10 small molecules. It is observed from Table V that our CD-CCSD/cc-pCVQZ results (CD-threshold 10^{-4}) exactly reproduce previously reported results[10] using standard CCSD. Our results are thus well converged within the accuracy of the CCSD model, and hence we conclude that no accuracy is lost by employing this more efficient CD implementation scheme.

3. Convergence threshold effects

In addition to investigating the calculation times, we also studied the effect of changing the CD and gradient convergence thresholds on the final geometry of thymine—as measured by changes in redundant internal coordinates.[56] From Table VI we observe that a very small error is obtained in the final geometry, even at a CD threshold of 10^{-3}. This small error was also pointed out by Feng *et al.*[16] At our chosen CD threshold of 10^{-4}, the largest change in bond length is 0.01 pm; the changes in angles and dihedral angles are smaller than 0.002 radians, corresponding to about 0.01°. Thus, at the CCSD level of theory, the optimized geometry of thymine can be considered fully converged with our chosen CD threshold.

It must be noted, however, that even at this threshold, small differences in the CD basis can occur. This may cause changes in the number of required optimization cycles from run to run, but the final geometries remain unchanged to within the specified convergence thresholds.

4. Illustration of large-scale application

To showcase the applicability of our implementation, we have also performed geometry optimization on retinal, which contains 49 atoms and 150 electrons, corresponding to 735 basis functions with our chosen basis set (aug-cc-pVDZ). This optimization was only carried out

TABLE III. Ground state calculation times for thymine and azobenzene using Psi4 and our new implementation in e^T. The number of optimization cycles required, n_{cycles}, are reported, as well as the total calculation time, t_{total}, and the time per cycle, t_{cycle}. The average fraction of time spent calculating the gradient per optimization cycle, t_{gradient}, is reported for e^T.

	e^T				Psi4		
	n_{cycles}	t_{total}	t_{cycle}	t_{gradient}	n_{cycles}	t_{total}	t_{cycle}
Thymine	6	42 m	7m	6.8 %	7	2h 4m	18m
Azobenzene	10	7h 42m	46m	7.1%	7†	14h 27m	2h 4m

† The optimization failed after 7 cycles in Psi4.

TABLE IV. Ground and excited state calculation times for thymine and azobenzene using Q-Chem and our new implementation in e^T. The excited states are the lowest singlet states in all cases. The number of optimization cycles required, n_{cycles}, are reported, as well as the total calculation time, t_{total}, and the time per cycle, t_{cycle}. The average fraction of time spent calculating the gradient per optimization cycle, t_{gradient} is reported for e^T.

	e^T				Q-Chem		
	n_{cycles}	t_{total}	t_{cycle}	t_{gradient}	n_{cycles}	t_{total}	t_{cycle}
Thymine (GS)	6	42 m	7m	6.8 %	7	3h 42m	32m
Thymine (ES)	7	2h 3m	18m	4.7%	7	5h 42m	49m
Azobenzene (GS)	10	7h 42m	46m	7.1%	7	22h 53m	3h 16m
Azobenzene (ES)	9	23h 0m	2h 33m	4.2%	6	33h 43m	5h 37m

using e^T. While the calculation does indeed put a strain on the computational resources, requiring 17 hours per optimization cycle, it can in fact be performed. As in the previous examples, the time spent calculating the gradient amounts to only a small fraction of the total calculation time. Note that this calculation was performed using a development version of e^T 1.4. The version used in the other calculations (e^T 1.6) includes some optimizations of the left Jacobian transformation, potentially allowing for further reductions in computational cost.

V. CONCLUSIONS

We have presented an efficient implementation of CCSD gradients for ground and excited states based on Cholesky-decomposed electron repulsion integrals. Since CD is an inner projection scheme, we have chosen an implementation approach that follows earlier schemes[19,35] for CD and RI where one avoids the storage of OV^3 and V^4 arrays by constructing 3-index intermediates. We have furthermore chosen not to store the derivative Cholesky vectors; instead, the associated contributions are constructed on-the-fly. This allows us to significantly reduce the storage requirements.

Relative to Psi4 and Q-Chem, our implementation was shown to reduce the calculation time of geometry optimizations by roughly a factor of two. To a large extent, this reduction in time reflects the efficiency of the coupled cluster code in the e^T program. However, the calculation of gradients was found to require only a small fraction of the total calculation time, showcasing the efficiency of the gradient implementation.

The capabilities of our implementation was highlighted by showing that a geometry optimization of retinal could be carried out.

Further reduction in computational demands would be achieved by means of the frozen-core approximation. Work on this is currently in progress.

VI. SUPPLEMENTARY MATERIAL

Comparison of frozen-core and non-frozen-core calculations in Q-Chem as well as a breakdown of timings for ground and excited state geometry optimization calculations of azobenzene and thymine.

VII. DATA AVAILABILITY STATEMENT

Geometries can be found in Ref. 57. The code will be released in an upcoming version of the e^T program, which is open-source.[47]

ACKNOWLEDGMENTS

We thank Xintian Feng for taking the time to answer our questions about the Q-Chem gradient implementation. We acknowledge support from the DTU Partnership PhD programme (PhD grant to AKSP). AKSP acknowledges funding from the European Cooperation in Science and Technology, COST Action CA18222 *Attochem*. SC acknowledges support from the Independent Research Fund Denmark (DFF-RP2 Grant 7014-00258B). E.F.K, S.C., and H.K. acknowledge the Research Council of Norway through FRINATEK projects

TABLE V. Ground state optimized and experimental bond lengths, r/pm, and angles α/deg for 10 small molecules. The experimental results as well as the calculated CCSD/cc-pCVQZ results are taken from the compilation in Ref. 10

			CD-CCSD/cc-pCVQZ		CCSD/cc-pCVQZ		Experiment	
Molecule	Bond (r_{AB})	Angle (α)	r/pm	α/deg	r/pm	α/deg	r/pm	α/deg
H_2O	OH	∠HOH	95.4	104.5	95.4	104.5	96.6	104.5
HOF	OH	∠HOF	96.2	98.6	96.2	98.6	96.6	97.5
H_2O_2	OH	∠HOO	95.8	100.7	95.8	100.7	96.7	102.3
NH_3	NH	∠HNH	100.9	106.6	100.9	106.6	101.1	106.7
N_2H_2	NH	∠HNN	102.5	106.6	102.5	106.6	102.9	106.3
HNO	NH	∠HNO	104.8	108.3	104.8	108.3	106.2	108.5
C_2H_4	CH	∠HCH	107.9	117.0	107.9	117.0	108.1	117.4
CH_2O	CH	∠HCH	109.9	116.4	109.9	116.4	110.1	116.3
CH_2	CH	∠HCH	110.5	102.0	110.5	102.0	110.7	102.4
O_3	OO	∠OOO	124.1	117.8	124.1	117.8	127.2	116.8

TABLE VI. Deviations in the optimized geometry of thymine. The deviations are given relative to an optimized geometry obtained with a CD convergence threshold of 10^{-8} and a gradient convergence threshold of $3 \cdot 10^{-8}$. Changes in internal coordinates are measured by the largest change in a bond length (Δr), an angle ($\Delta\alpha$), and a dihedral angle ($\Delta\theta$).

CD threshold	Gradient threshold	Max Δr/pm	Max $\Delta\alpha$/rad	Max $\Delta\theta$/rad
10^{-3}	$3 \cdot 10^{-3}$	0.058	0.00068	0.0012
10^{-4}	$3 \cdot 10^{-4}$	0.012	0.00040	0.0012
10^{-5}	$3 \cdot 10^{-5}$	0.002	0.00007	0.0010

TABLE VII. Calculation time for optimizing the geometry of retinal in e^T. The number of optimization cycles required, n_{cycles}, are reported, as well as the total calculation time, t_{total}, the time spent per cycle, t_{cycles}, and the average time spent calculating the gradient per optimization cycle, t_{gradient}. The CD convergence threshold was here 10^{-3}.

	n_{cycles}	t_{total}	t_{cycle}	t_{gradient}
Retinal (GS)	24	17d 10h 23m	17h 26m	6.6%

263110 and 275506. Computing resources through UNINETT Sigma2—the National Infrastructure for High Performance Computing and Data Storage in Norway (Project No. NN2962k) are also acknowledged.

Appendix A: Two-electron densities

The two-electron density is here taken as

$$d_{pqrs} = L_\nu \langle \nu | e^{-T_2} e_{pqrs} e^{T_2} | \rho \rangle R_\rho, \tag{A1}$$

where

$$e_{pqrs} = E_{pq}E_{rs} - \delta_{qr}E_{ps}. \tag{A2}$$

Note that we are using a T_1-transformed basis, where the T_1-dependence has been moved into the derivative integrals, as they will be contracted with the density later on. Throughout, we assume a spin-adapted singlet basis, where the kets are expressed in the so-called elementary basis and bras are expressed in the basis biorthonormal to the kets.[10]

In the expression for the density, we can write, for the ground state,

$$L_0 = 1, \quad L_\mu = \bar{t}_\mu, \tag{A3}$$
$$R_0 = 1, \quad R_\mu = 0, \tag{A4}$$

and, for excited states,

$$L_0 = 0, \qquad L_\mu = L_\mu, \tag{A5}$$
$$R_0 = 0, \qquad R_\mu = R_\mu. \tag{A6}$$

Here $\boldsymbol{R}$ and $\boldsymbol{L}$ are defined for $\mu > 0$, and

$$\boldsymbol{A}\boldsymbol{R} = \omega\boldsymbol{R} \tag{A7}$$
$$\boldsymbol{A}^T\boldsymbol{L} = \omega\boldsymbol{L}. \tag{A8}$$

In addition, for the excited state, terms with

$$L_0 = 0, \quad L_\mu = \bar{t}_\mu^{\text{ES}} \tag{A9}$$
$$R_0 = 1, \quad R_\mu = 0 \tag{A10}$$

must be determined to account for the terms of the excited state Lagrangian containing the amplitude response $\bar{\zeta}_\mu$. These terms will however be formally identical to the terms of the ground state that do not include L_0, and these will therefore not be written out explicitly. This applies to both the one- and two-electron densities.

In the following, we shall utilize that R_2 can be written as $\tilde{R}_2$:

$$\begin{aligned} R_2 &= \frac{1}{2}\sum_{aibj} R_{aibj}(1+\delta_{ai,bj})E_{ai}E_{bj} \\ &\equiv \frac{1}{2}\sum_{aibj} \tilde{R}_{aibj}E_{ai}E_{bj} = \tilde{R}_2. \end{aligned} \tag{A11}$$

There are eight unique combinations of occupied and virtual indices, because of the symmetry

$$e_{pqrs} = e_{rspq} \implies d_{pqrs} = d_{rspq}. \tag{A12}$$

For each of these combinations, we derive below the corresponding CCSD two-electron density block. For improved readability, Einstein's implicit summation over repeated indices will be used.

For the ground state we obtain and implement

$$d^{\mathrm{gs}}_{ijkl} = (L_0 R_0)\Lambda_{ijkl} + O^{(2)}_{ijkl}(L_2, T_2)R_0 \tag{A13}$$
$$d^{\mathrm{gs}}_{aijk} = (2L^a_i \delta_{jk} - L^a_k \delta_{ji})R_0 \tag{A14}$$
$$d^{\mathrm{gs}}_{ijka} = O^{(M)}_{ijka}(L_1, T_2)R_0 \tag{A15}$$
$$d^{\mathrm{gs}}_{abij} = O^{(2)}_{abij}(L_2, T_2)R_0 \tag{A16}$$
$$d^{\mathrm{gs}}_{aibj} = L^{ab}_{ij} R_0 \tag{A17}$$
$$d^{\mathrm{gs}}_{iajb} = 2R_0 L_0 \tilde{t}^{ab}_{ij} + R_0 \frac{1}{2} O^{(2)}_{iajb}(L_2, T_2) \tag{A18}$$
$$d^{\mathrm{gs}}_{aijb} = O^{(M)}_{aijb}(L_2, T_2)R_0 \tag{A19}$$
$$d^{\mathrm{gs}}_{abci} = 0 \tag{A20}$$
$$d^{\mathrm{gs}}_{abic} = L^a_m \tilde{t}^{cb}_{im} R_0 \tag{A21}$$
$$d^{\mathrm{gs}}_{abcd} = O^{(2)}_{abcd}(L_2, T_2)R_0. \tag{A22}$$

and, for the excited state, we similarly obtain and implement

$$\begin{aligned} d^{\mathrm{es}}_{ijkl} = {} & O^{(2)}_{ijkl}(L_2, \tilde{R}_2) + (\boldsymbol{L}^T \boldsymbol{R})\Lambda_{ijkl} \\ & + O^{[1]}_{ijkl}(L_1, R_1) \end{aligned} \tag{A23}$$
$$d^{\mathrm{es}}_{aijk} = -L^{ad}_{ik} R_{dj} + 2W_{ai}\delta_{jk} - W_{ak}\delta_{ji} \tag{A24}$$
$$\begin{aligned} d^{\mathrm{es}}_{ijka} = {} & O^{(M)}_{ijka}(L_1, \tilde{R}_2) \\ & - 2X_{ij}(L_2, T_2)R^a_k \\ & + X_{kj}(L_2, T_2)R^a_i + X_{da}R^d_i \delta_{jk} \\ & + \Big(V_{djai}(L_2, T_2) + Y_{djai}(L_2, T_2) \\ & - 2\delta_{ij}X_{da}(L_2, T_2))\Big)R^d_k \\ & + \Big(Z_{ijkl}(L_2, T_2) + X_{il}(L_2, T_2)\delta_{jk} \\ & - 2\delta_{ij}X_{kl}(L_2, T_2))\Big)R^a_l \\ & - \tilde{t}^{ea}_{ik}W_{ej} - \Big(\tilde{t}^{ea}_{mi}\delta_{jk} - 2\delta_{ij}\tilde{t}^{ea}_{mk}\Big)W_{em} \\ & - \tilde{Y}_{djak}(L_2, T_2)R^d_i \end{aligned} \tag{A25}$$
$$d^{\mathrm{es}}_{abij} = O^{(2)}_{abij}(L_2, \tilde{R}_2) + O^{(1)}_{abij}(L_1, R_1) \tag{A26}$$
$$d^{\mathrm{es}}_{aibj} = 0 \tag{A27}$$
$$\begin{aligned} d^{\mathrm{es}}_{iajb} = {} & O^{(2)}_{iajb}(L_2, \tilde{R}_2) \\ & + 2L_{ck}\tilde{t}^{ac}_{ik}R_{bj} - L_{ck}\tilde{t}^{ac}_{jk}R_{bi} + 2L_{ck}\tilde{t}^{bc}_{jk}R_{ai} \\ & - L_{ck}\tilde{t}^{bc}_{ik}R_{aj} - Y_{cb}\tilde{t}^{ac}_{ij} - Y_{ca}\tilde{t}^{bc}_{ji} - Y_{jk}\tilde{t}^{ab}_{ik} \\ & - Y_{ik}\tilde{t}^{ab}_{kj} + 2L_{ck}\tilde{t}^{ab}_{ij}R_{ck} + L^{cd}_{kl}\tilde{R}^{cd}_{kl}\tilde{t}^{ab}_{ij} \end{aligned} \tag{A28}$$
$$d^{\mathrm{es}}_{aijb} = O^{(M)}_{aijb}(L_2, \tilde{R}_2) + 2L_{ai}R_{bj} - Y_{ab}\delta_{ij} \tag{A29}$$
$$d^{\mathrm{es}}_{abci} = L^{ac}_{ji}R^b_j \tag{A30}$$
$$\begin{aligned} d^{\mathrm{es}}_{abic} = {} & L^a_m(2\tilde{R}^{bc}_{mi} - \tilde{R}^{cb}_{mi}) \\ & + 2X_{ab}(L_2, T_2)R^c_i \\ & - (V_{akbi}(L_2, T_2) + Y_{akbi}(L_2, T_2))R^c_k \\ & + L^{ad}_{mk}\tilde{t}^{bc}_{mi}R^d_k - L^{ad}_{mn}t^{bc}_{mn}R^d_i \\ & + \tilde{Y}_{akci}(L_2, T_2)R^b_k - X_{ac}(L_2, T_2)R^b_i \end{aligned} \tag{A31}$$
$$d^{\mathrm{es}}_{abcd} = O^{(2)}_{abcd}(L_2, \tilde{R}_2). \tag{A32}$$

Here we have defined the following quantities:

$$\Lambda_{ijkl} = 4\delta_{ij}\delta_{kl} - 2\delta_{il}\delta_{kj} \tag{A33}$$
$$\begin{aligned} O^{(1)}_{ijkl}(L_1, R_1) = {} & -2Y_{ij}\delta_{kl} + Y_{il}\delta_{kj} \\ & - 2\delta_{ij}Y_{kl} + \delta_{il}Y_{kj} \end{aligned} \tag{A34}$$
$$Y_{kj} = R_{ek}L_{ej} = R^e_k L^e_j \tag{A35}$$
$$\begin{aligned} O^{(2)}_{ijkl}(L_2, C_2) = {} & -2\delta_{kl}X_{ij}(L_2, C_2) \\ & - 2\delta_{ij}X_{kl}(L_2, C_2) \\ & + \delta_{kj}X_{il}(L_2, C_2) + \delta_{il}X_{kj}(L_2, C_2) \\ & + Z_{ijkl}(L_2, C_2) \end{aligned} \tag{A36}$$
$$X_{ij}(L_2, C_2) = C^{ef}_{mi}L^{ef}_{mj} \tag{A37}$$
$$Z_{ijkl}(L_2, C_2) = C^{ef}_{ik}L^{ef}_{jl} \tag{A38}$$
$$\begin{aligned} O^{(M)}_{ijka}(L_1, T_2) = {} & L_{em}\left(2\delta_{ij}\tilde{t}^{ae}_{km} - \delta_{jk}\tilde{t}^{ea}_{mi}\right)R_0 \\ & - L_{ej}\tilde{t}^{ea}_{ik}R_0 \end{aligned} \tag{A39}$$
$$\tilde{Y}_{djak}(L_2, T_2) = L^{ed}_{mj}\tilde{t}^{ea}_{mk} \tag{A40}$$
$$Y_{ab} = L^a_k R^b_k \tag{A41}$$
$$O^{(1)}_{abij}(L_1, R_1) = 2\delta_{ij}Y_{ab}(L_1, R_1) - L^a_j R^b_i \tag{A42}$$
$$\begin{aligned} O^{(2)}_{abij}(L_2, C_2) = {} & 2\delta_{ij}X_{ab}(L_2, C_2) - Y_{ajbi}(L_2, C_2) \\ & - V_{ajbi}(L_2, C_2) \end{aligned} \tag{A43}$$

$$
\begin{aligned}
O^{(2)}_{iajb}(L_2,C_2) = &\tilde{Y}_{ckai}(L_2,T_2)\tilde{C}^{cb}_{kj} \\
&+ \tilde{Y}_{ckbj}(L_2,T_2)\tilde{C}^{ca}_{ki} \\
&+ Z_{ikjl}(L_2,T_2)C^{ab}_{kl} \\
&+ Z_{imjn}(L_2,C_2)t^{ab}_{mn} \\
&- X_{ca}(L_2,T_2)\tilde{C}^{cb}_{ij} \\
&- X_{ea}(L_2,C_2)\tilde{t}^{eb}_{ij} \\
&- X_{ik}(L_2,T_2)\tilde{C}^{ab}_{kj} \\
&- X_{im}(L_2,C_2)\tilde{t}^{ab}_{mj} \\
&- X_{jk}(L_2,T_2)\tilde{C}^{ab}_{ik} \\
&- X_{jm}(L_2,C_2)\tilde{t}^{ab}_{im} \\
&- X_{cb}(L_2,T_2)\tilde{C}^{ca}_{ji} \\
&- X_{eb}(L_2,C_2)\tilde{t}^{ea}_{ji} \\
&- \tilde{Y}_{embi}(L_2,C_2)t^{ea}_{mj} \\
&- Y_{emaj}(L_2,C_2)\tilde{t}^{eb}_{mi} \\
&+ Y_{ckbi}(L_2,C_2)t^{ac}_{kj} \\
&+ Y_{ckbi}C^{ac}_{kj} \\
&+ V_{ckaj}(L_2,T_2)C^{cb}_{ik} \\
&+ V_{ckbi}(L_2,T_2)C^{ac}_{kj}
\end{aligned}
\tag{A44}
$$

$$
\begin{aligned}
O^{(M)}_{aijb}(L_2,C_2) = &-L^{ae}_{im}C^{eb}_{jm}R_0 \\
&+ 2Y_{aibj}(L_2,C_2) - \delta_{ij}X_{ab}(L_2,C_2)
\end{aligned}
\tag{A45}
$$

$$
O^{(2)}_{abcd}(L_2,C_2) = L^{ac}_{ij}C^{bd}_{ij}, \tag{A46}
$$

where C_2 denotes a set of double amplitudes and $\tilde{C}_2$ is defined in an identical fashion to $\tilde{t}$.

All O^4, O^3V and O^2V^2 densities are directly constructed and stored in memory, while the OV^3 and V^4 densities are not explicitly constructed. As discussed in Section II B, we instead construct their contributions to the 3-index density intermediates W^J_{pq} and store these in memory.

Appendix B: Orbital relaxation in reorthonormalization

The relaxation contributions—which we derive by expanding $h^{(1)}_{pq}$ and $g^{(1)}_{pqrs}$ using Eq. (21) and inserting into the differentiated Fock matrix, see Eq. (4) and (20)—are

$$
\mathcal{F}^{\bar{\kappa}}_{aq} = \frac{1}{2}\bar{\kappa}_{ai}h_{qi} + (g_{qijj} - \frac{1}{2}g_{qjji})\bar{\kappa}_{ai} \tag{B1}
$$

$$
\begin{aligned}
\mathcal{F}^{\bar{\kappa}}_{iq} = &\frac{1}{2}\bar{\kappa}_{ai}h_{aq} + (g_{aqjj} - \frac{1}{2}g_{ajjq})\bar{\kappa}_{ai} \\
&+ 2g_{ajqi}\bar{\kappa}_{aj} - \frac{1}{2}g_{aqij}\bar{\kappa}_{aj} - \frac{1}{2}g_{aiqj}\bar{\kappa}_{aj}
\end{aligned}
\tag{B2}
$$

These terms scale as $O(N^4)$ once expressed in terms of Cholesky vectors, in terms of which

$$
\mathcal{F}^{\bar{\kappa}}_{aq} = \frac{1}{2}\bar{\kappa}_{ai}h_{qi} + \gamma^J D^J_{qa} - \frac{1}{2}L^J_{qj}D^J_{ja} \tag{B3}
$$

$$
\begin{aligned}
\mathcal{F}^{\bar{\kappa}}_{iq} = &\frac{1}{2}\bar{\kappa}_{ai}h_{aq} + E^J_{qi}\gamma^J \\
&- \frac{1}{2}M_{aq}\bar{\kappa}_{ai} + 2\delta^J L^J_{qi} - \frac{1}{2}E^J_{qj}L^J_{ij} \\
&- \frac{1}{2}E^J_{ij}L^J_{qj},
\end{aligned}
\tag{B4}
$$

where

$$
D^J_{qa} = L^J_{qi}\bar{\kappa}_{ai} \tag{B5}
$$

$$
\gamma^J = L^J_{jj} \tag{B6}
$$

$$
E^J_{qi} = L^J_{aq}\bar{\kappa}_{ai} \tag{B7}
$$

$$
M_{aq} = L^J_{aj}L^J_{jq} \tag{B8}
$$

$$
\delta^J = L^J_{ai}\bar{\kappa}_{ai}. \tag{B9}
$$

We have used the Einstein's implicit summation in all these expressions.

[1] H. B. Schlegel, WIREs Comput. Mol. Sci. **1**, 790 (2011), https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/wcms.34.

[2] S. J. Klippenstein, V. S. Pande, and D. G. Truhlar, J. Am. Chem. Soc. **136**, 528 (2014), https://doi.org/10.1021/ja408723a.

[3] D. Marx and J. Hutter, *Ab initio molecular dynamics: basic theory and advanced methods* (Cambridge University Press, 2009).

[4] B. F. E. Curchod and T. J. Martínez, Chem. Rev. **118**, 3305 (2018), https://doi.org/10.1021/acs.chemrev.7b00423.

[5] J. Čížek, J. Chem. Phys. **45**, 4256 (1966), https://doi.org/10.1063/1.1727484.

[6] R. J. Bartlett, Annu. Rev. Phys. Chem **32**, 359 (1981), https://doi.org/10.1146/annurev.pc.32.100181.002043.

[7] R. J. Bartlett, J. Phys. Chem. **93**, 1697 (1989), https://doi.org/10.1021/j100342a008.

[8] T. D. Crawford and H. F. Schaefer III, "An introduction to coupled cluster theory for computational chemists," in *Reviews in Computational Chemistry* (John Wiley & Sons, Ltd, 2000) pp. 33–136, https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470125915.ch2.

[9] R. J. Bartlett and M. Musiał, Rev. Mod. Phys. **79**, 291 (2007).

[10] T. Helgaker, P. Jørgensen, and J. Olsen, *Molecular electronic-structure theory* (John Wiley & Sons, 2014).

[11] H. Koch, O. Christiansen, P. Jørgensen, A. M. Sanchez de Merás, and T. Helgaker, J. Chem. Phys. **106**, 1808 (1997), https://doi.org/10.1063/1.473322.

[12] R. Izsák, Wiley Interdiscip. Rev. Comput. Mol. Sci **10**, e1445 (2020), https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcms.1445.

[13] O. Christiansen, H. Koch, and P. Jørgensen, Chem. Phys. Lett. **243**, 409 (1995).

[14] G. D. Purvis and R. J. Bartlett, J. Chem. Phys. **76**, 1910 (1982), https://doi.org/10.1063/1.443164.

[15] E. Epifanovsky, A. T. B. Gilbert, X. Feng, J. Lee, Y. Mao, N. Mardirossian, P. Pokhilko, A. F. White, M. P. Coons, A. L. Dempwolff, Z. Gan, D. Hait, P. R. Horn, L. D. Jacobson, I. Kaliman, J. Kussmann, A. W. Lange, K. U. Lao, D. S. Levine, J. Liu, S. C. McKenzie, A. F. Morrison, K. D. Nanda, F. Plasser, D. R. Rehn, M. L. Vidal, Z.-Q. You, Y. Zhu, B. Alam, B. J. Albrecht, A. Aldossary, E. Alguire, J. H. Andersen, V. Athavale, D. Barton, K. Begam, A. Behn, N. Bellonzi, Y. A. Bernard, E. J. Berquist, H. G. A. Burton, A. Carreras, K. Carter-Fenk, R. Chakraborty, A. D. Chien, K. D. Closser, V. Cofer-Shabica, S. Dasgupta, M. de Wergifosse, J. Deng, M. Diedenhofen, H. Do, S. Ehlert, P.-T. Fang, S. Fatehi, Q. Feng, T. Friedhoff, J. Gayvert, Q. Ge, G. Gidofalvi, M. Goldey, J. Gomes, C. E. González-Espinoza, S. Gulania, A. O. Gunina, M. W. D. Hanson-Heine, P. H. P. Harbach, A. Hauser, M. F. Herbst, M. Hernández Vera, M. Hodecker, Z. C. Holden, S. Houck,

X. Huang, K. Hui, B. C. Huynh, M. Ivanov, A. Jász, H. Ji, H. Jiang, B. Kaduk, S. Kähler, K. Khistyaev, J. Kim, G. Kis, P. Klunzinger, Z. Koczor-Benda, J. H. Koh, D. Kosenkov, L. Koulias, T. Kowalczyk, C. M. Krauter, K. Kue, A. Kunitsa, T. Kus, I. Ladjánszki, A. Landau, K. V. Lawler, D. Lefrancois, S. Lehtola, R. R. Li, Y.-P. Li, J. Liang, M. Liebenthal, H.-H. Lin, Y.-S. Lin, F. Liu, K.-Y. Liu, M. Loipersberger, A. Luenser, A. Manjanath, P. Manohar, E. Mansoor, S. F. Manzer, S.-P. Mao, A. V. Marenich, T. Markovich, S. Mason, S. A. Maurer, P. F. McLaughlin, M. F. S. J. Menger, J.-M. Mewes, S. A. Mewes, P. Morgante, J. W. Mullinax, K. J. Oosterbaan, G. Paran, A. C. Paul, S. K. Paul, F. Pavošević, Z. Pei, S. Prager, E. I. Proynov, Á. Rák, E. Ramos-Cordoba, B. Rana, A. E. Rask, A. Rettig, R. M. Richard, F. Rob, E. Rossomme, T. Scheele, M. Scheurer, M. Schneider, N. Sergueev, S. M. Sharada, W. Skomorowski, D. W. Small, C. J. Stein, Y.-C. Su, E. J. Sundstrom, Z. Tao, J. Thirman, G. J. Tornai, T. Tsuchimochi, N. M. Tubman, S. P. Veccham, O. Vydrov, J. Wenzel, J. Witte, A. Yamada, K. Yao, S. Yeganeh, S. R. Yost, A. Zech, I. Y. Zhang, X. Zhang, Y. Zhang, D. Zuev, A. Aspuru-Guzik, A. T. Bell, N. A. Besley, K. B. Bravaya, B. R. Brooks, D. Casanova, J.-D. Chai, S. Coriani, C. J. Cramer, G. Cserey, A. E. DePrince, R. A. DiStasio, A. Dreuw, B. D. Dunietz, T. R. Furlani, W. A. Goddard, S. Hammes-Schiffer, T. Head-Gordon, W. J. Hehre, C.-P. Hsu, T.-C. Jagau, Y. Jung, A. Klamt, J. Kong, D. S. Lambrecht, W. Liang, N. J. Mayhall, C. W. McCurdy, J. B. Neaton, C. Ochsenfeld, J. A. Parkhill, R. Peverati, V. A. Rassolov, Y. Shao, L. V. Slipchenko, T. Stauch, R. P. Steele, J. E. Subotnik, A. J. W. Thom, A. Tkatchenko, D. G. Truhlar, T. Van Voorhis, T. A. Wesolowski, K. B. Whaley, H. L. Woodcock, P. M. Zimmerman, S. Faraji, P. M. W. Gill, M. Head-Gordon, J. M. Herbert, and A. I. Krylov, J. Chem. Phys. **155**, 084801 (2021).

[16]X. Feng, E. Epifanovsky, J. Gauss, and A. I. Krylov, J. Chem. Phys. **151**, 014110 (2019), https://doi.org/10.1063/1.5100022.

[17]M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, "Gaussian˜16 Revision C.01," (2016), gaussian Inc. Wallingford CT.

[18]D. G. A. Smith, L. A. Burns, A. C. Simmonett, R. M. Parrish, M. C. Schieber, R. Galvelis, P. Kraus, H. Kruse, R. Di Remigio, A. Alenaizan, A. M. James, S. Lehtola, J. P. Misiewicz, M. Scheurer, R. A. Shaw, J. B. Schriber, Y. Xie, Z. L. Glick, D. A. Sirianni, J. S. O'Brien, J. M. Waldrop, A. Kumar, E. G. Hohenstein, B. P. Pritchard, B. R. Brooks, H. F. Schaefer, A. Y. Sokolov, K. Patkowski, A. E. DePrince, U. Bozkaya, R. A. King, F. A. Evangelista, J. M. Turney, T. D. Crawford, and C. D. Sherrill, J. Chem. Phys. **152**, 184108 (2020), https://doi.org/10.1063/5.0006002.

[19]U. Bozkaya and C. D. Sherrill, J. Chem. Phys. **144**, 174103 (2016), https://doi.org/10.1063/1.4948318.

[20]K. Aidas, C. Angeli, K. L. Bak, V. Bakken, R. Bast, L. Boman, O. Christiansen, R. Cimiraglia, S. Coriani, P. Dahle, E. K. Dalskov, U. Ekström, T. Enevoldsen, J. J. Eriksen, P. Ettenhuber, B. Fernández, L. Ferrighi, H. Fliegl, L. Frediani, K. Hald, A. Halkier, C. Hättig, H. Heiberg, T. Helgaker, A. C. Hennum, H. Hettema, E. Hjertenæs, S. Høst, I.-M. Høyvik, M. F. Iozzi, B. Jansík, H. J. Aa. Jensen, D. Jonsson, P. Jørgensen, J. Kauczor, S. Kirpekar, T. Kjærgaard, W. Klopper, S. Knecht, R. Kobayashi, H. Koch, J. Kongsted, A. Krapp, K. Kristensen, A. Ligabue, O. B. Lutnæs, J. I. Melo, K. V. Mikkelsen, R. H. Myhre, C. Neiss, C. B. Nielsen, P. Norman, J. Olsen, J. M. H. Olsen, A. Osted, M. J. Packer, F. Pawlowski, T. B. Pedersen, P. F. Provasi, S. Reine, Z. Rinkevicius, T. A. Ruden, K. Ruud, V. V. Rybkin, P. Sałek, C. C. M. Samson, A. S. de Merás, T. Saue, S. P. A. Sauer, B. Schimmelpfennig, K. Sneskov, A. H. Steindal, K. O. Sylvester-Hvid, P. R. Taylor, A. M. Teale, E. I. Tellgren, D. P. Tew, A. J. Thorvaldsen, L. Thøgersen, O. Vahtras, M. A. Watson, D. J. D. Wilson, M. Ziolkowski, and H. Ågren, WIREs Comput.Mol.Sci. **4**, 269 (2014).

[21]K. Hald, A. Halkier, P. Jørgensen, S. Coriani, C. Hättig, and T. Helgaker, J. Chem. Phys. **118**, 2985 (2003), https://doi.org/10.1063/1.1531106.

[22]D. A. Matthews, L. Cheng, M. E. Harding, F. Lipparini, S. Stopkowicz, T.-C. Jagau, P. G. Szalay, J. Gauss, and J. F. Stanton, J. Chem. Phys. **152**, 214108 (2020), https://doi.org/10.1063/5.0004837.

[23]J. Gauss, J. F. Stanton, and R. J. Bartlett, J. Chem. Phys. **95**, 2623 (1991), https://doi.org/10.1063/1.460915.

[24]M. Kállay, P. R. Nagy, D. Mester, Z. Rolik, G. Samu, J. Csontos, J. Csóka, P. B. Szabó, L. Gyevi-Nagy, B. Hégely, I. Ladjánszki, L. Szegedy, B. Ladóczki, K. Petrov, M. Farkas, P. D. Mezei, and Á. Ganyecz, J. Chem. Phys. **152**, 074107 (2020), https://doi.org/10.1063/1.5142048.

[25]M. Kállay, J. Gauss, and P. G. Szalay, J. Chem. Phys. **119**, 2991 (2003), https://doi.org/10.1063/1.1589003.

[26]N. H. F. Beebe and J. Linderberg, Int. J. Quantum Chem **12**, 683 (1977), https://onlinelibrary.wiley.com/doi/pdf/10.1002/qua.560120408.

[27]I. Røeggen and E. Wisløff-Nilssen, Chem. Phys. Lett. **132**, 154 (1986).

[28]D. W. O'neal and J. Simons, Int. J. Quant. Chem. **36**, 673 (1989), https://onlinelibrary.wiley.com/doi/pdf/10.1002/qua.560360602.

[29]H. Koch, A. Sánchez de Merás, and T. B. Pedersen, J. Chem. Phys. **118**, 9481 (2003), https://doi.org/10.1063/1.1578621.

[30]F. Aquilante, T. B. Pedersen, and R. Lindh, J. Chem. Phys. **126**, 194106 (2007), https://doi.org/10.1063/1.2736701.

[31]F. Aquilante, R. Lindh, and T. B. Pedersen, J. Chem. Phys. **129**, 034106 (2008), https://doi.org/10.1063/1.2955755.

[32]A. E. DePrince and C. D. Sherrill, J. Chem. Theory Comput **9**, 293 (2013), https://doi.org/10.1021/ct300780u.

[33]E. Epifanovsky, D. Zuev, X. Feng, K. Khistyaev, Y. Shao, and A. I. Krylov, J. Chem. Phys. **139**, 134105 (2013), https://doi.org/10.1063/1.4820484.

[34]U. Bozkaya, J. Chem. Theory Comput. **10**, 2371 (2014), https://doi.org/10.1021/ct500231c.

[35]M. G. Delcey, L. Freitag, T. B. Pedersen, F. Aquilante, R. Lindh, and L. González, J. Chem. Phys. **140**, 174103 (2014), https://doi.org/10.1063/1.4873349.

[36]S. D. Folkestad, E. F. Kjønstad, and H. Koch, J. Chem. Phys. **150**, 194112 (2019), https://doi.org/10.1063/1.5083802.

[37]J. L. Whitten, J. Chem. Phys. **58**, 4496 (1973), https://doi.org/10.1063/1.1679012.

[38]B. I. Dunlap, J. W. D. Connolly, and J. R. Sabin, J. Chem. Phys. **71**, 3396 (1979), https://doi.org/10.1063/1.438728.

[39]M. Feyereisen, G. Fitzgerald, and A. Komornicki, Chem. Phys. Lett. **208**, 359 (1993).

[40]O. Vahtras, J. Almlöf, and M. Feyereisen, Chem. Phys. Lett. **213**, 514 (1993).

[41]A. P. Rendell and T. J. Lee, J. Chem. Phys. **101**, 400 (1994), https://doi.org/10.1063/1.468148.

[42]F. Weigend, Phys. Chem. Chem. Phys. **4**, 4285 (2002).

[43]A. Sodt, J. E. Subotnik, and M. Head-Gordon, J. Chem. Phys. **125**, 194109 (2006), https://doi.org/10.1063/1.2370949.

[44]H.-J. Werner, F. R. Manby, and P. J. Knowles, J. Chem. Phys. **118**, 8149 (2003), https://doi.org/10.1063/1.1564816.

[45]M. Schütz and F. R. Manby, Phys. Chem. Chem. Phys. **5**, 3349 (2003).
[46]H.-J. Werner and M. Schütz, J. Chem. Phys. **135**, 144116 (2011), https://doi.org/10.1063/1.3641642.
[47]S. D. Folkestad, E. F. Kjønstad, R. H. Myhre, J. H. Andersen, A. Balbi, S. Coriani, T. Giovannini, L. Goletto, T. S. Haugland, A. Hutcheson, I.-M. Høyvik, T. Moitra, A. C. Paul, M. Scavino, A. S. Skeidsvoll, Å. H. Tveten, and H. Koch, J. Chem. Phys. **152**, 184103 (2020), https://doi.org/10.1063/5.0004713.
[48]T. Helgaker and P. Jørgensen, "Calculation of geometrical derivatives in molecular electronic structure theory," in *Methods in Computational Molecular Physics*, edited by S. Wilson and G. H. F. Diercksen (Springer US, Boston, MA, 1992) pp. 353–421.
[49]T. Helgaker and P. Jørgensen, Theoret. Chim. Acta **75**, 111 (1989), doi.org/10.1007/BF00527713.
[50]P. Jørgensen and T. Helgaker, J. Chem. Phys. **89**, 1560 (1988), https://doi.org/10.1063/1.455152.
[51]T. Helgaker and P. Jørgensen, Adv. Quant. Chem. **19**, 183 (1988).
[52]T. Helgaker, S. Coriani, P. Jørgensen, K. Kristensen, J. Olsen, and K. Ruud, Chem. Rev. **112**, 543 (2012).
[53]A. C. Scheiner, G. E. Scuseria, J. E. Rice, T. J. Lee, and H. F. Schaefer, J. Chem. Phys. **87**, 5361 (1987), https://doi.org/10.1063/1.453655.
[54]G. E. Scuseria, C. L. Janssen, and H. F. Schaefer, J. Chem. Phys. **89**, 7382 (1988), https://doi.org/10.1063/1.455269.
[55]A. Banerjee, N. Adams, J. Simons, and R. Shepard, J. Phys. Chem. **89**, 52 (1985).
[56]V. Bakken and T. Helgaker, J. Chem. Phys. **117**, 9160 (2002), https://doi.org/10.1063/1.1515483.
[57]A. K. Schnack-Petersen, H. Koch, S. Coriani, and E. F. Kjønstad, (2022), https://doi.org/10.5281/zenodo.5957852.
[58]P. Pokhilko, E. Epifanovsky, and A. I. Krylov, J. Chem. Theory Comput. **14**, 4088 (2018), https://doi.org/10.1021/acs.jctc.8b00321.