

Analytic Energy Gradients for Constrained DFT–Configuration Interaction

Benjamin Kaduk, Takashi Tsuchimochi, and Troy Van Voorhis*

Massachusetts Institute of Technology

77 Massachusetts Avenue

Cambridge, MA 02139

Abstract

The constrained density functional theory–configuration (CDFT-CI) method has previously been used to calculate ground-state energies and barrier heights, and to describe electronic excited states, in particular conical intersections. However, the method has been limited to evaluating the electronic energy at just a single nuclear configuration, with the gradient of the energy being available only via finite difference. In this paper, we present analytic gradients of the CDFT-CI energy with respect to nuclear coordinates, which allows the potential for accurate geometry optimization and molecular dynamics on both the ground and excited electronic states, a realm which is currently quite challenging for electronic structure theory. We report the performance of CDFT-CI geometry optimization for representative reaction transition states as well as molecules in an excited state. It turns out that the overall accuracy of CDFT-CI in the barrier heights of the tested reactions is almost unchanged with the geometries optimized through QCISD and CDFT-CI, meaning CDFT-CI produces very good reaction transition states. It also proves the capability of opening a frontier of excited state calculations.

* tvan@mit.edu

I. INTRODUCTION

Electronic excited states are of interest in a great many chemical systems, being of relevance to photochemistry,[1–8] photodamage to DNA,[9–18] organic semiconductors,[19–23] and more.[24–26] Of particular interest is the *dynamics* on the excited state, after an excitation event has occurred. In order to study the geometric relaxation of electronic excited states, one requires the force experienced by the nuclei on the excited-state PES. This requirement limits the spectrum of electronic structure methods which are usable, with the field being limited to time-dependent density functional theory (TD-DFT),[27–35] configuration–interaction singles (CIS),[7, 31, 36–41] complete active space self-consistent field (CASSCF) and its second order perturbation theory (CASPT2)[42–48] (note that analytic gradients for CASPT2 has become available only recently[49]) equation-of-motion coupled-cluster singles and doubles (EOM-CCSD)[50–54] and its approximate form (EOM-CC2),[31, 55–61] and sometimes multi-reference configuration interaction (MRCI).[62–65] Even for EOM-CCSD, CASPT2, and MRCI, the computational expense will vary with the implementation and application, and such calculations become impractical for systems with more than 10 or 15 atoms.

On the other hand, DFT methods gain a significant advantage of practicality as the system size increases. TD-DFT has seen broad use for electronic excited states in general, and excited-state dynamics and geometry optimization are no exception.[30–35] However, it still suffers from the deficiencies in describing multiple excitations and charge-transfer excitations which render it a less-than-general solution for vertical excitation energy calculations,[66–75] though recent developments show the state of affairs may be improving.[76–79] Restricted open-shell Kohn-Sham (ROKS) [80–82] is a state specific method, and its gradients are easily available compared to TD-DFT.[83] However, ROKS consists of a two-determinant wave function, and therefore cannot describe electronic structures with multiple excitations. Constrained DFT (CDFT) is designed to directly construct charge- and spin-constrained states and as such can find charge-transfer states directly, using self-consistent ground state techniques.[84–86] The self-consistent nature of the solution means that nonlinear response of the density is included, and hence in principle permits the treatment of multiple excitations from the ground state. However, CDFT has limitations of its own; it is still a single-reference method (and thus suffers from the limitations of DFT in the face of strong static correlation),

and it has no effective prescription for describing valence excitations. CDFT-configuration interaction (CDFT-CI) explicitly introduces multiple configurations to the electronic structure treatment, which greatly improves results for situations where static correlation is strongly present, such as dissociation curves and reaction transition states.[87, 88] Additionally, it can treat the ground state and excited states on the same footing, which is necessary for describing conical intersections qualitatively.[67] However, applications of CDFT-CI have heretofore remained somewhat limited due to the unavailability of gradients of the electronic energy. In this work, we present the theory and implementation of analytic energy gradients for CDFT-CI. These forces are used to optimize the transition-state geometries for a standard set of reaction barriers.

In most cases, the energy does not change noticeably from the reference transition-state geometry to the optimized geometry, indicating that CDFT-CI-optimized geometries for transition states are of comparable quality to the reference geometries. Geometry optimization on the excited state using CDFT-CI energy gradients also converges to geometries of acceptable quality, promising the capability of handling the hard cases of, for example, conical intersections and states with significant charge-transfer and valence mixing.

This paper is organized as follows. In Sec. II A, our general strategy for obtaining the CDFT-CI energy gradients is outlined. Sec. II B shows the derivatives of matrix elements and couplings as well as how to avoid those of molecular orbitals (MO) with respect to nuclear positions. Sec. II C presents promolecule contribution to the gradients, which is a special feature of CDFT-CI. Constraint potential contribution to the gradient is given in II D. Sec. II E summarizes the overall expressions derived. In Sec. III we evaluate the performance of the CDFT-CI gradients with the HTBH38/04 and NHTBH38 sets for benchmark calculations of transition states, and excited states of small molecules. Finally, we draw our conclusion and perspective on the present study in Sec. IV.

II. METHODS

We briefly summarize the equations of CDFT and CDFT-CI before proceeding to the derivation of expressions for the gradient of the energy. CDFT takes as input a density functional giving the energy $E[\rho]$ and adds a constraint Lagrange multiplier term to yield a

new functional

$$E[\rho, V_k] = E[\rho] + \sum_k V_k \left(\int \hat{w}_k \rho(\mathbf{r}) d\mathbf{r} - N_k \right), \quad (1)$$

where \hat{w}_k is a “weight” operator that probes the number of electrons in some particular region of space and N_k is a target value for that operator. Minimizing E with respect to ρ under the constraints imposed yields a state with the desired constrained charge and spin properties. The effect of the constraint terms in the energy expression are equivalent to adding an additional “constraint” potential $\sum_k V_k \hat{w}_k(\mathbf{r}) = \sum_k \hat{V}_k(\mathbf{r}) = \hat{V}$ acting on the electrons in Kohn-Sham theory.

CDFT-CI requires the user to specify a collection of different constrained states which are used as a basis/active space for constructing a configuration-interaction matrix; the basis states are specified as charge and spin constraints on particular fragments of the system in question. This CI matrix then gives the CDFT-CI eigenvalue equation,

$$\begin{pmatrix} \mathcal{H}_{11} & \mathcal{H}_{12} \\ \mathcal{H}_{21} & \mathcal{H}_{22} \end{pmatrix} \begin{pmatrix} C_1 \\ C_2 \end{pmatrix} = \mathcal{E} \begin{pmatrix} \mathcal{S}_{11} & \mathcal{S}_{12} \\ \mathcal{S}_{21} & \mathcal{S}_{22} \end{pmatrix} \begin{pmatrix} C_1 \\ C_2 \end{pmatrix}$$

We show only the two-state case, but the generalization to N states is easily made. The diagonal elements of \mathbf{H} are just the energies of the constrained states that form the basis for the active space; the off-diagonal elements are constructed as:[87, 88]

$$\mathcal{H}_{IJ} = \mathcal{H}_{JI} = \frac{F_I + F_J}{2} \mathcal{S}_{IJ} - \left\langle \Phi_I \left| \frac{\hat{V}_I + \hat{V}_J}{2} \right| \Phi_J \right\rangle, \quad (2)$$

where Φ_I is the Kohn-Sham determinant for the I th CDFT state, F_I is the energy of the I th CDFT state in the presence of the constraining potential \hat{V}_I for state I , and \mathcal{S}_{IJ} is just $\langle \Phi_I | \Phi_J \rangle$. For the rest of this work, we will assume the two-state form, using capital letters I and J to index the different states; extension to the N state case is straightforward. We also adhere to the convention that i and j index occupied orbitals, p and q index all orbitals, and Greek letters atomic orbitals (AO).

If we write $\mathbf{HC} = \mathcal{E}\mathbf{SC}$, then we can easily take the derivative with respect to a nuclear coordinate x and write

$$\mathbf{H}^x \mathbf{C} + \mathbf{HC}^x = \mathcal{E}^x \mathbf{SC} + \mathcal{E} \mathbf{S}^x \mathbf{C} + \mathcal{E} \mathbf{SC}^x. \quad (3)$$

Bracketing on the left with \mathbf{C}^\dagger and rearranging lets us solve for the gradient \mathcal{E}^x :

$$\mathcal{E}^x = \mathbf{C}^\dagger (\mathbf{H}^x - \mathcal{E} \mathbf{S}^x) \mathbf{C}. \quad (4)$$

\mathcal{E} and \mathbf{C} are already known from the single-point energy evaluation, so the only new terms required for the gradient expression are \mathbf{H}^x and \mathbf{S}^x .

A. Overview

In order to actually use Eq. (4) to obtain \mathcal{E}^x , we must consider both diagonal terms of the form \mathbf{H}_{II}^x and off-diagonal terms \mathbf{H}_{IJ}^x and \mathbf{S}_{IJ}^x . (There are no diagonal overlap terms, since normalized states will always have unit self-overlap.) The diagonal terms \mathbf{H}_{II}^x are just the energy gradient of the constrained states, so we focus on the off-diagonal elements \mathbf{H}_{IJ}^x and \mathbf{S}_{IJ}^x . However, since we only use these two quantities in combination, it proves convenient to define an auxiliary quantity $\mathbf{W} = \mathbf{H} - \mathcal{E}\mathbf{S}$ where \mathcal{E} is treated as a *constant* and does not vary with changes in any other parameters. We then seek to compute $\mathbf{W}^x = (\mathbf{H} - \mathcal{E}\mathbf{S})^x = \mathbf{H}^x - \mathcal{E}\mathbf{S}^x$.

At this point, the high-level view is no longer sufficient and we must expand the expression to include the Kohn-Sham determinants of our CI basis states.

$$W_{II} = H_{II} \tag{5}$$

$$W_{II}^x = H_{II}^x \tag{6}$$

$$W_{IJ} = \left(\frac{F_I + F_J}{2} - \mathcal{E} \right) \langle \Phi_I | \Phi_J \rangle - \left\langle \Phi_I \left| \frac{\hat{V}_I + \hat{V}_J}{2} \right| \Phi_J \right\rangle \tag{7}$$

$$W_{IJ}^x = \frac{F_I^x + F_J^x}{2} \mathcal{S}_{IJ} - \left\langle \Phi_I \left| \frac{\hat{V}_I^x + \hat{V}_J^x}{2} \right| \Phi_J \right\rangle + \langle \Phi_I^x | \hat{O} | \Phi_J \rangle + \langle \Phi_I | \hat{O} | \Phi_J^x \rangle \tag{8}$$

$$\hat{O} = \left(\frac{F_I + F_J}{2} - \mathcal{E} \right) - \frac{\hat{V}_I + \hat{V}_J}{2} \tag{9}$$

However, the terms in $\langle \Phi_I^x |$ and $| \Phi_J^x \rangle$ are unreasonable to compute, given that the wavefunction gradient requires $O(N^3)$ space to store and $O(N^5)$ time to compute. As such, we seek an alternate route to W_{IJ}^x which does not involve the gradient of the wavefunction; we will adopt the standard framework of making the matrix element variational.[49]

As we will see, the procedure to evaluate Eq. (8) is quite complex and does not have a linear logical order of execution. To assist in understanding the steps involved in the computation, a flow chart of the relevant expressions is presented in Figure 1. The labeled boxes in the flowchart correspond roughly to the subsections that follow, though we start off with the general computation of the gradient of a matrix element in Sec II B, used

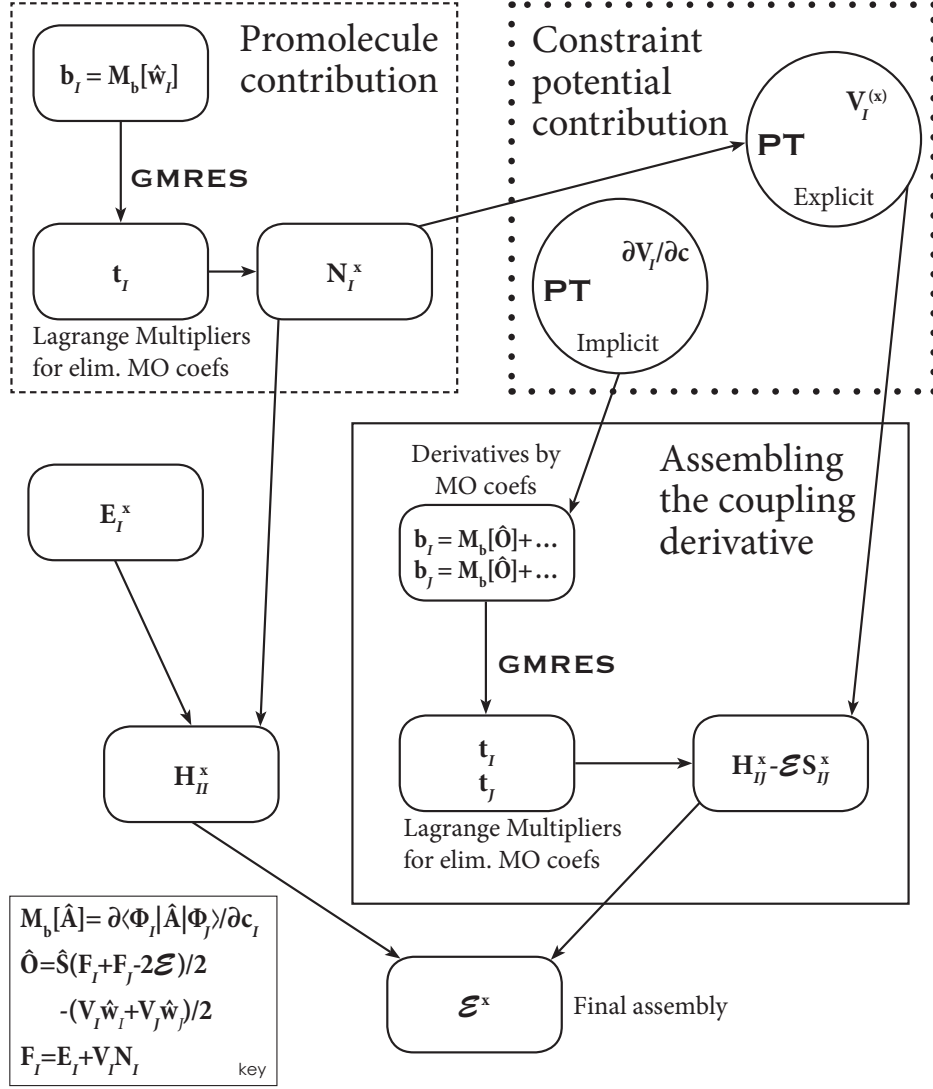


FIG. 1. The flowchart for evaluating the gradient of the CDFT-CI energy. The gradient of the promolecule-adjusted constraint values is computed as a variational matrix element (dashed box), using the GMRES linear solver to obtain the Lagrange multipliers needed for variationality. These N_I^x are combined with the CDFT energy gradient to yield the diagonal elements of the Hamiltonian (bottom left), and also used to determine the explicit dependence of the constraint potential Lagrange multipliers on the nuclear coordinates, using a perturbation theory expression (upper right). A separate perturbation theory expression (also in the dotted box) gives the dependence of the constraint potential Lagrange multipliers on the MO coefficients (and thus the implicit dependence on nuclear position), which enters into the variationality of the coupling derivative (solid box), computed in a similar fashion to the promolecule contribution. With both diagonal and off-diagonal matrix elements available, the energy gradient is evaluated per Eq. (4) (bottom).

for both the promolecule contribution (Sec II C), which is required to carry out CDFT-CI calculations[87], and the coupling element derivative. We then discuss the explicit and implicit contributions of constraint potential to the energy gradient in Sec II D.

B. Assembling a matrix element/coupling derivative

Looking forward, we will need to evaluate several expressions of similar form, so we step back from W_{IJ}^x and consider the general case of the matrix element of a one-electron (or zero-electron) operator \hat{O} between two states $|\Phi_I\rangle$ and $|\Phi_J\rangle$, which has been normalized for possible variations of the orbitals which underlie the states.

$$M[\hat{O}] = \frac{\langle \Phi_I | \hat{O} | \Phi_J \rangle}{\sqrt{\langle \Phi_I | \Phi_I \rangle \langle \Phi_J | \Phi_J \rangle}} \quad (10)$$

$$= \frac{\text{Tr} \left[\mathbf{c}_I^\dagger \mathbf{O} \mathbf{c}_J \left(\mathbf{c}_I^\dagger \mathbf{S} \mathbf{c}_J \right)^{-1} \right] \det(\mathbf{c}_I^\dagger \mathbf{S} \mathbf{c}_J)}{\sqrt{\det(\mathbf{c}_I^\dagger \mathbf{S} \mathbf{c}_I) \det(\mathbf{c}_J^\dagger \mathbf{S} \mathbf{c}_J)}}. \quad (11)$$

We will primarily be concerned with the case of $\hat{O} = (\hat{V}_I + \hat{V}_J)/2$ and also the case where there is no operator (so the matrix element becomes just an overlap), but will also need this machinery for $\hat{O} = \hat{w}$ in a special case, per Sec. II C. To avoid the need for the expensive wavefunction gradient (or, equivalently, the gradient of the MO coefficients), we introduce an auxiliary function which has the same value as M at all points but is variational with respect to the MO coefficients; this allows the gradient of the MO coefficients to be ignored via the chain rule. We thus write

$$M^{\text{var}} = M[\hat{O}] - L(\mathbf{c}_I, \mathbf{t}_I) - L(\mathbf{c}_J, \mathbf{t}_J). \quad (12)$$

1. Lagrange multipliers to eliminate dependence of M on the MO coefficients

This term L depends on Lagrange multipliers \mathbf{t} which enforce variationality; there are $N_{\text{occ}} \times N_{\text{basis}}$ relevant MO coefficients, so there are $N_{\text{occ}} \times N_{\text{basis}}$ Lagrange multipliers \mathbf{t} , appearing as

$$L(\mathbf{c}, \mathbf{t}) = \text{Tr} \left[\mathbf{t}^\dagger \cdot \left(\mathbf{F}[\tilde{\mathbf{P}}] \mathbf{c} - \mathbf{S} \mathbf{c} \epsilon \right) \right], \quad (13)$$

where $\tilde{\mathbf{P}} = 3\mathbf{PSP} - 2\mathbf{PSPSP}$ is the density matrix after McWeeny's purification transformation, [89] $\mathbf{P} = \mathbf{c}\mathbf{c}^\dagger$ is the density matrix, and \mathbf{S} remains the atomic orbital overlap matrix. ϵ is a diagonal matrix of MO energies, which we define to be evaluated as $\epsilon_i = \frac{\mathbf{c}_i^\dagger \mathbf{F} \mathbf{c}_i}{\mathbf{c}_i^\dagger \mathbf{S} \mathbf{c}_i}$ so as to remain normalized when the orbitals themselves become unnormalized, per Eq. (11). We also introduce the Fock matrix \mathbf{F} , which has dependence on both the MO coefficients and the nuclear position, but we do not need to enforce that $\partial M^{\text{var}}/\partial \mathbf{F} = 0$. Eq. (13) is constructed such that the quantity in parentheses will always evaluate to zero when the system is at SCF convergence. Accordingly, L will also always be zero at convergence, so M^{var} will have the same value as M . Furthermore, $\partial M^{\text{var}}/\partial \mathbf{t}$ will also be zero by the self-consistency condition, which removes any need for gradients of \mathbf{t} in evaluating chain-rule terms. (Note that the actual values of \mathbf{t} are as-yet unspecified.) The McWeeny purified density matrix is required so that changes in the MO coefficients which do not preserve normalization do not affect the resulting Fock matrix; changes in normalization of the MO coefficients at first order will only affect the purified density matrix at the second order. No correction is needed for the MO coefficients that appear directly in $\mathbf{F}\mathbf{c} - \mathbf{S}\mathbf{c}\epsilon$, as that expression is merely enforcing that the orbitals remain eigenvectors; a change in normalization does not affect that condition.

To enforce the variationality of M^{var} , we require

$$\frac{\partial M^{\text{var}}}{\partial \mathbf{c}} = 0 \quad (14)$$

$$\Rightarrow \frac{\partial M}{\partial \mathbf{c}} = \frac{\partial L(\mathbf{c}, \mathbf{t})}{\partial \mathbf{c}}, \quad (15)$$

for derivatives with respect to both \mathbf{c}_I and \mathbf{c}_J . Both sides of this equation expand out to sizable expressions, which we treat singly.

$$\begin{aligned} \frac{\partial M}{\partial \mathbf{c}_I} &= \det(\mathbf{c}_I^\dagger \mathbf{S} \mathbf{c}_J) \mathbf{O} \mathbf{c}_J (\mathbf{c}_I^\dagger \mathbf{S} \mathbf{c}_J)^{-1} - \mathbf{S} \mathbf{c}_J (\mathbf{c}_I^\dagger \mathbf{S} \mathbf{c}_J)^{-1} \mathbf{c}_I^\dagger \mathbf{O} \mathbf{c}_J (\mathbf{c}_I^\dagger \mathbf{S} \mathbf{c}_J)^{-1} \\ &\quad + M[\hat{O}] \mathbf{S} \mathbf{c}_J (\mathbf{c}_I^\dagger \mathbf{S} \mathbf{c}_J)^{-1} - M[\hat{O}] \mathbf{S} \mathbf{c}_I \end{aligned} \quad (16)$$

$$\begin{aligned} \frac{\partial L(\mathbf{c}, \mathbf{t})}{\partial c_{\mu i}} &= t_{\nu j} \frac{\partial F_{\nu\lambda}[\tilde{\mathbf{P}}]}{\partial c_{\mu i}} c_{\lambda j} + t_{\nu i} F_{\nu\mu}[\tilde{\mathbf{P}}] - t_{\nu i} S_{\nu\mu} \epsilon_i - t_{\nu j} S_{\nu\delta} c_{\lambda j} \frac{\partial F_{\lambda\sigma}[\tilde{\mathbf{P}}]}{\partial c_{\mu i}} c_{\sigma j} c_{\delta j} \\ &\quad - 2t_{\nu i} S_{\nu\delta} F_{\mu\sigma}[\tilde{\mathbf{P}}] c_{\sigma i} c_{\delta i} + 2t_{\nu i} S_{\nu\delta} c_{\lambda i} F_{\lambda\sigma}[\tilde{\mathbf{P}}] c_{\sigma i} S_{\mu\gamma} c_{\gamma i} c_{\delta i}. \end{aligned} \quad (17)$$

However, we can see that the overall structure of the expression is really

$$M_b[\hat{O}] \equiv \frac{\partial M}{\partial c_{\mu i}} = b_{\mu i} = A_{\mu i}^{\nu j} t_{\nu j}, \quad (18)$$

where we have adopted explicit indices and the Einstein summation convention to make clear the complicated structure of some of these terms, and defined a new quantity $M_b[\hat{O}]$ for future use.[90] The final line, however, makes it clear that this is essentially a linear system which can be solved for \mathbf{t} in a conceptually straightforward fashion, regardless of the complexity of some of the terms. The \mathbf{A} matrix that defines this linear system, however, is of size $(N_{occ} \cdot N_{virt}) \times (N_{occ} \cdot N_{virt})$ which requires $O(N^4)$ storage and $O(N^6)$ time for a direct inversion, a step backwards from Eq. (9). However, we can solve the linear system in Eq. (18) without constructing \mathbf{A} , by using an iterative linear solver. This allows us to leverage the fact that the product $A_{\mu i}^{\nu j} t_{\nu j}$ may be evaluated efficiently without computing \mathbf{A} . The explicit expression for the contraction is given by

$$A_{\mu i}^{\nu j} t_{\nu j} = A_{0 \mu i}^{\nu j} t_{\nu j} + \frac{\partial F_{\nu \lambda}[\tilde{\mathbf{P}}]}{\partial c_{\mu i}} Q_{\lambda \nu}[\mathbf{t}], \quad (19)$$

$$A_{0 \mu i}^{\nu j} t_{\nu j} = t_{\nu j} \delta_{ij} \left(F_{\nu \mu}[\tilde{\mathbf{P}}] - S_{\nu \mu} \epsilon_j - 2S_{\nu \delta} F_{\mu \sigma}[\tilde{\mathbf{P}}] c_{\sigma j} c_{\delta j} + S_{\nu \delta} c_{\lambda j} F_{\lambda \sigma}[\tilde{\mathbf{P}}] c_{\sigma j} S_{\mu \gamma} c_{\gamma j} c_{\delta j} \right), \quad (20)$$

where we have explicitly constructed the pseudo-density matrix

$$Q_{\lambda \nu} = c_{\lambda i} t_{\nu i} - c_{\lambda i} t_{\alpha i} S_{\alpha \beta} c_{\beta i} c_{\nu i}. \quad (21)$$

We have also separated \mathbf{A} into a term \mathbf{A}_0 and terms dependent on $\partial \mathbf{F} / \partial \mathbf{c}$. It is worth noting that, if considered as a matrix, \mathbf{A}_0 is block-diagonal — each orbital only interacts with the corresponding “orbital” from t . In other words, all of the terms in A_0 include a Kronecker delta δ_{ij} , so there is only $O(N^3)$ work to be done in the overall multiplication.

We have conveniently left unexpanded the expression $\partial F_{\lambda \sigma}[\tilde{\mathbf{P}}] / \partial c_{\mu i}$, a quantity whose determination is complicated by the use of the purified density $\tilde{\mathbf{P}}$. Performing this computation requires a breakdown of the different contributions to \mathbf{F} , with the Coulomb integrals and Hartree-Fock exchange being treated differently from DFT exchange and correlation (XC) functionals. (The one-electron Hamiltonian of course has no dependence on the MO coefficients.) Bearing in mind our need to compute only the product $\mathbf{A}t$ and not \mathbf{A} itself, we examine the contraction of some pseudo-density-matrix $Q_{\lambda \sigma}$ against $\partial F_{\lambda \sigma}[\tilde{\mathbf{P}}] / \partial c_{\delta i}$, looking at each of these terms in turn.

$$\begin{aligned} \frac{\partial J_{\lambda \sigma}[\tilde{\mathbf{P}}]}{\partial c_{\delta i}} Q_{\lambda \sigma} &= 2J_{\mu \delta}[\mathbf{Q}] P_{\mu \alpha} S_{\alpha \beta} c_{\beta i} + 2S_{\delta \beta} P_{\beta \nu} J_{\mu \nu}[\mathbf{Q}] c_{\mu i} \\ &\quad - 4S_{\delta \eta} P_{\eta \nu} J_{\mu \nu}[\mathbf{Q}] P_{\mu \alpha} S_{\alpha \beta} c_{\beta i}, \end{aligned} \quad (22)$$

where we use the fact that $\mathbf{PSP} = \mathbf{P}$ at convergence. This requires only a single Coulomb build from the pseudo-density \mathbf{Q} , and matrix multiplications with \mathbf{S} and \mathbf{P} .

In a similar fashion, the expression for the exchange derivative becomes

$$\begin{aligned} \frac{\partial K_{\lambda\sigma}[\tilde{\mathbf{P}}]}{\partial c_{\delta i}} Q_{\lambda\sigma} &= 2K_{\mu\delta}[\mathbf{Q}]P_{\mu\alpha}S_{\alpha\beta}c_{\beta i} + 2S_{\delta\beta}P_{\beta\nu}K_{\mu\nu}[\mathbf{Q}]c_{\mu i} \\ &\quad - 4S_{\delta\eta}P_{\eta\nu}K_{\mu\nu}[\mathbf{Q}]P_{\mu\alpha}S_{\alpha\beta}c_{\beta i} \end{aligned} \quad (23)$$

The DFT contributions are not quite as straightforward, in that a standard calculation of the XC matrix from the pseudodensity matrix \mathbf{Q} is not the correct approach to this derivative. In fact, the XC matrix (for a pure functional) is more properly written like

$$v_{xc} = v_{xc}[\rho(\tilde{\mathbf{P}})]$$

From the chain rule,

$$\frac{\partial v_{xc}}{\partial c_{\delta i}} = \frac{\partial v_{xc}}{\partial \rho} \frac{\partial \rho}{\partial \tilde{\mathbf{P}}} \frac{\partial \tilde{\mathbf{P}}}{\partial c_{\delta i}} \quad (24)$$

$$= \left(\frac{\partial v_{xc}}{\partial \rho} \frac{\partial \rho}{\partial \tilde{\mathbf{P}}} \right) \frac{\partial \tilde{\mathbf{P}}}{\partial c_{\delta i}}. \quad (25)$$

The quantity in parentheses is the implicit first derivative of the XC matrix, which is generally only used by being contracted against a ‘‘trial density’’, as we are doing here as we contract against \mathbf{Q} . Thus,

$$\frac{\partial (v_{xc})_{\lambda\sigma}}{\partial c_{\delta i}} Q_{\lambda\sigma} = \left(Q_{\lambda\sigma} \frac{\partial (v_{xc})_{\lambda\sigma}}{\partial \rho} \frac{\partial \rho}{\partial \tilde{P}_{\alpha\beta}} \right) \frac{\partial \tilde{P}_{\alpha\beta}}{\partial c_{\delta i}} \quad (26)$$

$$= X_{\alpha\beta} \frac{\partial \tilde{P}_{\alpha\beta}}{\partial c_{\delta i}} \quad (27)$$

$$= 2X_{\delta\mu}P_{\mu\nu}S_{\nu\lambda}c_{\lambda i} + 2S_{\delta\mu}P_{\mu\nu}X_{\nu\lambda}c_{\lambda i} - 4S_{\delta\mu}P_{\mu\nu}X_{\nu\lambda}P_{\lambda\sigma}S_{\sigma\eta}c_{\eta i}, \quad (28)$$

in an analogous fashion to the coulomb and exchange terms. We implicitly define the quantity $X_{\alpha\beta}$ as the contraction of \mathbf{Q} against the implicit first derivative of the XC matrix. All of these terms (Coulomb, exchange, and DFT) may be efficiently evaluated in $O(N^3)$ time or less and $O(N^2)$ space.

2. Iterative linear solver (GMRES)

Now that the matrix-vector product is available, we can proceed to the iterative linear solver. We have implemented the GMRES (Generalized Minimum Residual) algorithm in

Q-CHEM; the algorithm is covered extensively elsewhere,[91, 92] but we give a brief summary here and expound further in Appendix A. The goal is to construct an approximate solution to the linear system

$$\mathbf{A} \cdot \mathbf{x} = \mathbf{b}$$

without explicitly operating on the matrix \mathbf{A} , instead only evaluating matrix-vector products $\mathbf{A} \cdot \mathbf{t}_i$. GMRES requires the matrix \mathbf{A} to be square. The outline of the algorithm is to find an approximate solution in a projected subspace at each step; by preconditioning the system appropriately, an approximate solution with sufficiently small residual can be obtained in a constant number of iterations, essentially independent of the dimension of \mathbf{A} . For the systems we consider here, that constant is around twenty GMRES iterations.

The GMRES method is most effective for diagonal-dominant matrices (or rather, those whose eigenvalues are clustered near each other in the complex plane, which is not quite the same), frequently exhibiting quadratic convergence for such systems.[91] It is therefore quite common to introduce a (left) preconditioner, an approximation $\mathbf{A}_0 \approx \mathbf{A}$ which may be easily inverted, and then applying the GMRES algorithm to the preconditioned system:

$$(\mathbf{A}_0^{-1} \mathbf{A}) \cdot \mathbf{t} = \mathbf{A}_0^{-1} \cdot \mathbf{b}$$

Our linear system of Eq. (18) is not diagonal dominant, so we will need some nontrivial preconditioning strategy for GMRES to converge efficiently. However, there is a block-diagonal component to our \mathbf{A} , which in general is much larger in magnitude than the contribution coming from contractions against the derivative of the Fock matrix. This block-diagonal term, already referred to as \mathbf{A}_0 above, could be explicitly constructed with $O(N^3)$ effort and blockwise inverted for $O(N^4)$ effort, but it proves more convenient to only consider the first two terms of \mathbf{A}_0 ,

$$A'_{0\mu i}{}^{\nu j} = \delta_{ij} (F_{\mu\nu} - S_{\mu\nu} \epsilon_j), \quad (29)$$

since in this formulation computing $\mathbf{A}'_0{}^{-1} x$ will only require $O(N^3)$ work.

This inversion is equivalent to solving the systems

$$(\mathbf{F} - \mathbf{S} \epsilon_i) \mathbf{t}_i = \mathbf{b}_i, \quad (30)$$

which does not necessarily involve explicitly constructing $\mathbf{A}'_0{}^{-1}$ in matrix form. Conceptually, this is effected by transforming to the MO basis, where \mathbf{F} and \mathbf{S} are diagonal, so the

inversion is trivial. However, since $(\mathbf{F} - \mathbf{S}\epsilon_i)^{-1}$ is singular at ϵ_i , we use the pseudo-inverse which can be justified below. Then after a simple algebra, it is easy to show that

$$t_{\mu i} = \frac{1}{\epsilon_p - \epsilon_i} c_{\mu p} c_{\nu p} b_{\nu i} \quad (p \neq i). \quad (31)$$

The actual determination of \mathbf{x} involves just matrix-matrix products, yielding the desired $O(N^3)$ time. Since the energy denominator is only nonzero when $p \neq i$, this preconditioner will not treat components of $b_{\mu i}$ which are proportional to actual orbitals $c_{\mu i}$; however, these components are zero by construction, being eliminated by the normalization denominator in Eq. (11).

3. Assembling the coupling derivative

Having amassed a great deal of machinery to determine the Lagrange multipliers \mathbf{t} which make the function M^{var} of Eq. (12) variational with respect to the MO coefficients, we now return to the chain rule and actually make use of them. The terms from $\partial M^{\text{var}}/\partial \mathbf{t}$ have already been shown to be zero, but it turns out that our definition of L also introduced a dependence on \mathbf{F} to the full M^{var} which must be included in the chain rule terms so that

$$\frac{dM^{\text{var}}}{dx} = \frac{\partial M^{\text{var}}}{\partial \mathbf{O}} \frac{d\mathbf{O}}{dx} + \frac{\partial M^{\text{var}}}{\partial \mathbf{S}} \frac{d\mathbf{S}}{dx} + \frac{\partial M^{\text{var}}}{\partial \mathbf{F}} \frac{d\mathbf{F}}{dx}. \quad (32)$$

In evaluating these chain-rule terms, it proves convenient to repartition this expression as

$$\frac{dM}{dx} = \frac{dM^{\text{var}}}{dx} = \tilde{M}_x[\hat{O}] - \tilde{L}_x(\mathbf{c}_I, \mathbf{t}_I) - \tilde{L}_x(\mathbf{c}_J, \mathbf{t}_J), \quad (33)$$

where the tilde indicates to exclude the terms involving $\partial \mathbf{c}/\partial x$. $M[\hat{O}]$ depends on \mathbf{O} and \mathbf{S} but not \mathbf{F} , whereas L does not depend on \mathbf{O} but only on \mathbf{S} and \mathbf{F} , so this repartitioning simplifies the formal parameters somewhat. We can write

$$\begin{aligned} \tilde{M}_x[\hat{O}] = & \text{Tr} \left[\left\{ \mathbf{c}_I^\dagger \mathbf{O}^x \mathbf{c}_J + \mathbf{c}_I^\dagger \mathbf{O} \mathbf{c}_J \left(\mathbf{c}_I^\dagger \mathbf{S} \mathbf{c}_J \right)^{-1} \left(\mathbf{c}_I^\dagger \mathbf{S}^x \mathbf{c}_J \right) \right\} \left(\mathbf{c}_I^\dagger \mathbf{S} \mathbf{c}_J \right)^{-1} \right] \det(\mathbf{c}_I^\dagger \mathbf{S} \mathbf{c}_J) \\ & + M[\hat{O}] \cdot \text{Tr} \left[\left(\mathbf{c}_I^\dagger \mathbf{S} \mathbf{c}_J \right)^{-1} \left(\mathbf{c}_I^\dagger \mathbf{S}^x \mathbf{c}_J \right) \right] \\ & - \frac{1}{2} M[\hat{O}] \text{Tr} \left[\mathbf{c}_I^\dagger \mathbf{S}^x \mathbf{c}_I + \mathbf{c}_J^\dagger \mathbf{S}^x \mathbf{c}_J \right]. \end{aligned} \quad (34)$$

The L contribution does not need the explicit normalization term n ;

$$\tilde{L}_x(\mathbf{c}, \mathbf{t}) = \text{Tr} \left[\mathbf{t}^\dagger \cdot \left(\mathbf{F}^{(x)}[\tilde{\mathbf{P}}] \mathbf{c} - \mathbf{S}^x \mathbf{c} \epsilon - \mathbf{S} \mathbf{c} \epsilon^{(x)} \right) \right], \quad (35)$$

where

$$\begin{aligned}
\epsilon_p^{(x)} &= \frac{c_{\mu p} F_{\mu\nu}^{(x)}[\tilde{\mathbf{P}}] c_{\nu p}}{c_{\alpha p} S_{\alpha\beta} c_{\beta p}} - \frac{c_{\mu p} F_{\mu\nu}[\tilde{\mathbf{P}}] c_{\nu p}}{(c_{\alpha p} S_{\alpha\beta} c_{\beta p})^2} c_{\lambda p} S_{\lambda\sigma}^x c_{\sigma p} \\
&= c_{\mu p} F_{\mu\nu}^{(x)}[\tilde{\mathbf{P}}] c_{\nu p} - \epsilon_p c_{\mu p} S_{\mu\nu}^x c_{\nu p} \\
&= c_{\mu p} \left(F_{\mu\nu}^{(x)}[\tilde{\mathbf{P}}] - S_{\mu\nu}^x \epsilon_p \right) c_{\nu p}
\end{aligned} \tag{36}$$

and $\mathbf{F}^{(x)}[\tilde{\mathbf{P}}]$ is the partial derivative of the Fock matrix including the McWeeny purification of the density matrix (but excluding the position dependence of the MO coefficients).

That is,

$$\mathbf{F}^{(x)}[\tilde{\mathbf{P}}] = \mathbf{F}^{(x)}[\mathbf{P}] + \mathbf{J}[\tilde{\mathbf{P}}^{(x)}] + c_K \mathbf{K}[\tilde{\mathbf{P}}^{(x)}] + \left(\frac{\partial(v_{xc})}{\partial\rho} \frac{\partial\rho}{\partial\tilde{\mathbf{P}}} \right) \tilde{\mathbf{P}}^{(x)}, \tag{37}$$

$$\tilde{\mathbf{P}}^{(x)} = 3\mathbf{P}\mathbf{S}^x\mathbf{P} - 2\mathbf{P}\mathbf{S}^x\mathbf{P}\mathbf{S}\mathbf{P} - 2\mathbf{P}\mathbf{S}\mathbf{P}\mathbf{S}^x\mathbf{P} \tag{38}$$

$$= -\mathbf{P}\mathbf{S}^x\mathbf{P}. \tag{39}$$

The DFT contribution is again an instance of the implicit first derivative of the XC matrix used in Eq. (28). $\mathbf{F}^{(x)}[\mathbf{P}]$ is just the standard partial derivative of the Fock matrix with respect to the nuclear position, and c_K represents the coefficient of exact exchange in the density functional.

Since explicitly computing and storing $\mathbf{F}^{(x)}[\tilde{\mathbf{P}}]$ can be expensive, we contract pseudo density matrices \mathbf{X} (defined below) for \mathbf{J} and \mathbf{K} instead of $\tilde{\mathbf{P}}^{(x)}$, changing the indices as well as the order of summation. Eq.(35) can then be written as

$$\begin{aligned}
\tilde{L}_x(\mathbf{c}, \mathbf{t}) &= t_{\mu p} F_{\mu\nu}^{(x)}[\tilde{\mathbf{P}}] c_{\nu p} - t_{\mu p} S_{\mu\nu}^x c_{\nu p} \epsilon_p - t_{\mu p} S_{\mu\nu} c_{\nu p} c_{\lambda p} F_{\lambda\sigma}^{(x)}[\tilde{\mathbf{P}}] c_{\sigma p} + t_{\mu p} S_{\mu\nu} c_{\nu p} c_{\lambda p} S_{\lambda\sigma}^x c_{\sigma p} \epsilon_p \\
&= t_{\mu p} F_{\mu\nu}^{(x)}[\tilde{\mathbf{P}}] c_{\nu p} - t_{\lambda p} S_{\lambda\sigma} c_{\sigma p} c_{\mu p} F_{\mu\nu}^{(x)}[\tilde{\mathbf{P}}] c_{\nu p} - t_{\mu p} S_{\mu\nu}^x c_{\nu p} \epsilon_p + t_{\lambda p} S_{\lambda\sigma} c_{\sigma p} c_{\mu p} S_{\mu\nu}^x c_{\nu p} \epsilon_p \\
&= F_{\mu\nu}^{(x)}[\tilde{\mathbf{P}}] (t_{\mu p} c_{\nu p} - t_{\lambda p} S_{\lambda\sigma} c_{\sigma p} c_{\mu p} c_{\nu p}) - S_{\mu\nu}^x (t_{\mu p} c_{\nu p} + t_{\lambda p} S_{\lambda\sigma} c_{\sigma p} c_{\mu p} c_{\nu p}) \epsilon_p \\
&= F_{\mu\nu}^{(x)}[\tilde{\mathbf{P}}] X_{\nu\mu} - S_{\mu\nu}^x Y_{\nu\mu} \\
&= \text{Tr} \left[\mathbf{F}^{(x)}[\mathbf{P}] \mathbf{X} + \left\{ \mathbf{J}[\mathbf{X}] + c_k \mathbf{K}[\mathbf{X}] + \left(\frac{\partial(v_{xc})}{\partial\rho} \frac{\partial\rho}{\partial\tilde{\mathbf{P}}} \mathbf{X} \right) \right\} \tilde{\mathbf{P}}^{(x)} - \mathbf{S}^x \mathbf{Y} \right]
\end{aligned} \tag{40}$$

where we have defined

$$X_{\nu\mu} = t_{\mu p} c_{\nu p} - t_{\lambda p} S_{\lambda\sigma} c_{\sigma p} c_{\mu p} c_{\nu p} \tag{41}$$

$$Y_{\nu\mu} = (t_{\mu p} c_{\nu p} - t_{\lambda p} S_{\lambda\sigma} c_{\sigma p} c_{\mu p} c_{\nu p}) \epsilon_p \tag{42}$$

The structure of Eqs.(16-18) suggests \mathbf{t}_I really depends on J state. Therefore we have to compute $\mathbf{X}_I^{(J)}$ and the corresponding Fock-like matrix for each pair of states (note $\mathbf{X}_I^{(J)} \neq$

$\mathbf{X}_J^{(I)}$). This task, however, is much more efficient than working on $\mathbf{F}^{(x)}[\tilde{\mathbf{P}}^{(x)}]$ because in most cases the number of states is less than the number of atoms in a CDFTI-CI calculation. More importantly, it also enables us to keep the $O(N^2)$ storage.

It should be also noted that, in general, $\mathbf{X}_I^{(J)}$ is asymmetric, resulting in asymmetric $\mathbf{J}[\mathbf{X}]$, etc. However, it proves computationally useful to symmetrize $\mathbf{X}_I^{(J)}$, since it is easy to show that $\text{Tr}[\mathbf{J}[\mathbf{X}]\tilde{\mathbf{P}}^{(x)}] = \text{Tr}[\mathbf{J}[(\mathbf{X} + \mathbf{X}^\dagger)/2]\tilde{\mathbf{P}}^{(x)}]$ with symmetric $\tilde{\mathbf{P}}^{(x)}$.

The procedure to obtain $dM/dx = dM^{\text{var}}/dx$ then is to determine \mathbf{b}_I and \mathbf{b}_J using the $M_b[\hat{O}]$ formula, and perform two GMRES calculations (using the appropriate \mathbf{A} matrix for state I or J) to determine the Lagrange multipliers \mathbf{t}_I and \mathbf{t}_J . Then, $\tilde{M}_x[\hat{O}]$ from Eq. (34) and $\tilde{L}_x(\mathbf{c}, \mathbf{t})$ from Eq. (40) are substituted into Eq. (33) to obtain the final gradient.

It bears reiterating that $\tilde{L}_x \neq dL/dx$, since it omits the $\partial L/\partial \mathbf{c}$ terms which only cancel when the full quantity dM/dx is being evaluated. This is why \mathbf{t} must be redetermined for each operator and for each state.

C. Promolecule contribution

Having established the general form for evaluating the gradient of a matrix element of a one-electron operator between two distinct states, we now step back and note a particular issue with the formulation of CDFT-CI which makes the computation of \mathcal{E}^x (Eq. (4)) more complicated. Recall that the CDFT equations involve minimizing the value of $E[\rho] + \sum_k V_k (\int \hat{w}_k \rho d^3r - N_k)$ with respect to ρ and maximizing with respect to V_k for fixed \hat{w}_k and N_k (Eq. 1). However, when adapting CDFT for use in CDFT-CI, the concept of a “promolecule density” was introduced which produced modified values of N_k for a given system.[87] This used the converged density from independent calculations on noninteracting fragments, in conjunction with \hat{w}_k , to produce new values of N_k . Therefore, N_k also depends on the nuclear position since it is clear that the converged density of non-interacting fragments has a dependence on nuclear position (when there is more than one atom in a fragment) This dependence will in turn trickle through to affect the other properties of the system, such as the free energy of state I

$$F_I = E_I + \sum_k V_{I,k} N_{I,k}, \quad (43)$$

in the presence of constraints. Considering the above fact, the free energy gradient becomes

$$F_I^x = E_I^x + \sum_k V_{I,k} N_{I,k}^x. \quad (44)$$

Thus, in equation (6) when we said that H_{II}^x is the gradient of the CDFT state free energy, it is the gradient of that energy provided that the constraint values are also changing according to the promolecule formalism, i.e., it is the gradient including this correction of Eq. (44).

Recalling the definition of N_k :

$$N_k = \langle \Phi_I | \hat{w}_k | \Phi_I \rangle, \quad (45)$$

we note that the expression is precisely the matrix element of a one-electron operator, so we can reuse wholesale the algebraic machinery developed for $M[\hat{O}]$ in Sec. II B 3 to obtain N_k^x , with $\mathbf{b}_k = 2M_b[\hat{w}_k]$ yielding Lagrange multipliers \mathbf{t}_k via GMRES, and $N_k^x = \tilde{M}_x[\hat{w}_k] - \tilde{L}_x(\mathbf{c}, \mathbf{t}_k)$. That N_k is a matrix element between two identical states serves only to simplify the algebra in that only one set of Lagrange multipliers is needed and $\partial N_k / \partial \mathbf{c} = 2M_b[\hat{w}_k]$ due to the symmetry. It should be pointed out that evaluation of $\partial M[\hat{w}_k] / \partial \mathbf{c}$ requires to retain isolation between the independent fragments.

The gradient of the weight operator (for the full system, not the isolated fragments) is needed to evaluate $\tilde{M}_x[\hat{w}_k]$; for the Becke weights used in this implementation of CDFT-CI, such gradient terms have been computed in Reference [93].

D. Constraint potential contribution

While the single-state values of N_k^x were sufficient to correct W_{II}^x as in Eq. (44), the off-diagonal elements W_{IJ}^x include the matrix element of the constraint potential \hat{V}_k between two different states, and those quantities cannot be related to single-state values of N_k^x .

Given that the overall gradient of the constraint potential is

$$\hat{V}_k^x = V_k \hat{w}_k^x + V_k^x \hat{w}_k, \quad (46)$$

and the gradients of the weight matrices, \mathbf{w}_k^x , were used in evaluating $\tilde{M}_x[\hat{w}_k]$, it remains only to determine the gradient of the constraint potential Lagrange multipliers, V_k^x . These are intrinsically linked to the constraint value gradients N_k^x , changing in lockstep to maintain the SCF solution of the inner SCF procedure. As such, we require a mechanism to go

from the promolecule-derived constraint value gradients N_k^x to a gradient of the constraint potential Lagrange multipliers, V_k^x . It becomes clear that over the course of the entire double SCF calculation, the Lagrange multipliers V_k depend on nuclear coordinates both “directly”, through the explicit dependence of the Fock matrix (without constraint potentials) and AO overlap on the nuclear coordinates, and also indirectly, through the dependence of the Fock matrix on the MO coefficients (which in turn depends on the nuclear coordinates). It proves convenient to separate these dependencies, pushing the implicit contribution back into the \mathbf{b} vector of Eq. (18) for the quantity of interest, and only treating the explicit dependence at this junction.

1. *Explicit contribution*

Treating just the explicit contribution requires holding the MO coefficients used to build the Fock matrix fixed (while varying the nuclear geometry to $x + \delta x$), essentially just limiting the calculation to a single cycle of the outer SCF loop. The resulting change in the constraint potential Lagrange multipliers δV_k are the quantities we need for the explicit contribution. Because we consider only explicit changes in the Fock matrix (ignoring its nonlinear dependence on the MO coefficients), the resulting changes to the orbitals can be determined solely via perturbation theory.

$$N_k^x - \langle \Phi | \hat{w}_k^x | \Phi \rangle + \sum_i \langle \phi_i | w_k | \phi_i \rangle \langle \phi_i | \hat{S}^x | \phi_i \rangle - 2 \sum_{i \neq p} \langle \phi_i | \hat{w}_k | \phi_p \rangle \frac{\langle \phi_p | \hat{F}^{(x)} - \epsilon_i \hat{S}^x + \sum_l V_l \hat{w}_l^x | \phi_i \rangle}{\epsilon_i - \epsilon_p} = 2 \sum_l \sum_{i \neq p} \langle \phi_i | \hat{w}_k | \phi_p \rangle \frac{\langle \phi_p | \hat{w}_l | \phi_i \rangle}{\epsilon_i - \epsilon_p} V_l^x \quad (47)$$

$$b_k = A_k^l V_l^x, \quad (48)$$

Since the promolecule specification requires both charge and spin constraints on a given fragment, there will in general be at least two constraints, and thus a linear system to be solved. Fortunately, the \mathbf{A} matrix does not have any position dependence and can be precomputed and inverted just once, with b (and thus V_l^x) being computed for a single nuclear coordinate at a time. The b vector contains quantities which have already been computed; the gradient of the weight matrix is easy to compute, and the gradient of the overlap. On the other hand, since we have avoided to build $\mathbf{F}^{(x)}[\tilde{\mathbf{P}}]$ explicitly in Eq.(40), we again need to work on the fourth term in the left hand side of Eq.(47) using some pseudo

density matrix. Noting that this expression is written in the MO basis, the Fock contribution of this term is given by

$$\begin{aligned} -2 \sum_{i \neq p} w_{ip}^k \frac{F_{pi}^{(x)}[\tilde{\mathbf{P}}]}{\epsilon_i - \epsilon_p} &= -2 \sum_{i \neq p} w_{ip}^k \frac{c_{\mu p} c_{\nu i} F_{\mu\nu}^{(x)}[\tilde{\mathbf{P}}]}{\epsilon_i - \epsilon_p} \\ &= -2 F_{\mu\nu}^{(x)}[\tilde{\mathbf{P}}] \tilde{w}_{\nu\mu}^k \end{aligned} \quad (49)$$

in the AO basis, where we have defined

$$\tilde{w}_{\nu\mu}^k = \sum_p \sum_{i \neq p} c_{\nu i} \frac{w_{ip}^k}{\epsilon_i - \epsilon_p} c_{\mu p}. \quad (50)$$

Hence, we can use exactly the same machinery as we did in Eq.(40).

2. Implicit dependence

Having constructed an expression for the explicit dependence of the constraint potential on the nuclear position, in the form of an expression $(\delta V)\hat{w} + V(\delta\hat{w})$ which is added to the gradient of the Fock matrix, it remains to treat the implicit dependence, through the MO coefficients used to build the Fock matrix. As previously indicated, this will be included through the Lagrange multipliers in $L(\mathbf{c}, \mathbf{t})$, by including an extra contribution from $\partial V/\partial \mathbf{c}$ to the \mathbf{b} vector in Eq. (18). Obtaining this $\partial V/\partial \mathbf{c}$ contribution actually requires a very similar perturbation-theory structure to that of the $\partial V/\partial x$ contribution above, including the need for a linear system in the various constraints. In this case the perturbation is now $\delta\hat{F} = \partial\hat{F}/\partial \mathbf{c}$ and $\delta V_l = \partial V_l/\partial \mathbf{c}$. The form of the equations is identical to Eq. (47) with simplifications that δN_k is zero (the target constraint value does not depend on the MO coefficients passed to the Fock matrix), and $\delta\hat{S}$ and $\delta\hat{w}$ are also zero. The orbital overlap then becomes just

$$\langle \phi_p | \delta \phi_i \rangle = \frac{\langle \phi_p | \delta F + \sum_l (\delta V_l) \hat{w}_l | \phi_i \rangle}{\epsilon_i - \epsilon_p}. \quad (51)$$

And the linear system to be solved:

$$- \sum_{i \neq p} \langle \phi_i | \hat{w}_k | \phi_p \rangle \frac{\langle \phi_p | \delta \hat{F} | \phi_i \rangle}{\epsilon_i - \epsilon_p} = \sum_l \sum_{i \neq p} \langle \phi_i | \hat{w}_k | \phi_p \rangle \frac{\langle \phi_p | \hat{w}_l | \phi_i \rangle}{\epsilon_i - \epsilon_p} (\delta V_l), \quad (52)$$

$$b_k = A_k^l (\delta V_l). \quad (53)$$

An astute reader will note that the \mathbf{A} matrix is identical to the one in Eq. (47). All of the b vectors may be generated at once as contractions against $\partial F/\partial \mathbf{c}$, repeated for the number of constraints applied to the system. Such contractions against $\partial F/\partial \mathbf{c}$ were described in Eqs. (22) through (28).

E. Final assembly

At this point, all the pieces are in place to compute $(\mathbf{H} - \mathcal{E}\mathbf{S})_{IJ}^x = W_{IJ}^x$, the last remaining piece before Eq. (4) may be applied to obtain \mathcal{E}^x . In the now-familiar procedure, we construct

$$W_{IJ}^{\text{var}} = \frac{W_{IJ}}{\sqrt{\langle \Phi_I | \Phi_I \rangle \langle \Phi_J | \Phi_J \rangle}} - L(\mathbf{c}_I, \mathbf{t}_I) - L(\mathbf{c}_J, \mathbf{t}_J), \quad (54)$$

and solve for the Lagrange multipliers \mathbf{t}_I and \mathbf{t}_J which make W_{IJ}^{var} variational with respect to the MO coefficients. To do so, we need vectors \mathbf{b}_I and \mathbf{b}_J as input for GMRES calculations to determine \mathbf{t}_I and \mathbf{t}_J ; in our formalism the contribution from \mathbf{H} is split out into terms arising from the constrained states, so this really looks like

$$\begin{aligned} & \frac{\partial (\mathbf{H} - \mathcal{E}\mathbf{S})_{IJ}}{\partial \mathbf{c}_I} \\ &= \frac{\partial}{\partial \mathbf{c}_I} \left[\frac{\left(\frac{F_I + F_J}{2} S_{IJ} - \langle \Phi_I | \frac{\hat{V}_I + \hat{V}_J}{2} | \Phi_J \rangle - \mathcal{E} S_{IJ} \right)}{\sqrt{\langle \Phi_I | \Phi_I \rangle \langle \Phi_J | \Phi_J \rangle}} \right] \\ &= - \frac{\partial}{\partial \mathbf{c}_I} \left[\frac{\langle \Phi_I | \frac{\hat{V}_I + \hat{V}_J}{2} | \Phi_J \rangle}{\sqrt{\langle \Phi_I | \Phi_I \rangle \langle \Phi_J | \Phi_J \rangle}} \right] + \left(\frac{F_I + F_J}{2} - \mathcal{E} \right) \frac{\partial}{\partial \mathbf{c}_I} \left[\frac{S_{IJ}}{\sqrt{\langle \Phi_I | \Phi_I \rangle \langle \Phi_J | \Phi_J \rangle}} \right]. \end{aligned} \quad (55)$$

In the notation we have developed, we can then write

$$\mathbf{b}_I = -M_b \left[\left(\hat{V}_I + \hat{V}_J \right) / 2 \right] + \left(\frac{F_I + F_J}{2} - \mathcal{E} \right) M_b[\hat{S}] + \frac{1}{2} \sum_l M_b[\hat{w}_{I,l}], \quad (56)$$

where $\partial V_{I,l}/\partial \mathbf{c}_I$ are determined from Eq. (53); \mathbf{b}_J is determined similarly. A pair of GMRES calculations then yields \mathbf{t}_I and \mathbf{t}_J , which we can then assemble into the final

$$\frac{dW_{IJ}}{dx} = -\tilde{M}_x \left[\left(\hat{V}_I + \hat{V}_J \right) / 2 \right] + \left(\frac{F_I + F_J}{2} - \mathcal{E} \right) \tilde{M}_x[\hat{S}] - \tilde{L}_x(\mathbf{c}_I, \mathbf{t}_I) - \tilde{L}_x(\mathbf{c}_J, \mathbf{t}_J), \quad (57)$$

noting that the $\mathbf{F}^{(x)}$ matrices which are used in computing the \tilde{L}_x must include the contributions from $\hat{V}^{(x)} = \frac{\partial V}{\partial x} \hat{w} + V \hat{w}^x$ (where $\partial V/\partial x$ come from Eq. (47)).

This completes the construction of W_{IJ}^x and returns the gradient calculation of \mathcal{E}^x to Eq. (4), recalling that W_{II}^x comes from Eq. (44). We note that the CI vector C used in Eq. (4) is an eigenvector of the generalized eigenvector problem, in contrast to the coefficient vector produced by many CDFT-CI calculations, which is in the orthogonalized diabatic basis; a factor of $\mathbf{S}^{-1/2}$ allows interconversion.

III. RESULTS

We have implemented CDFT-CI gradients in a development version of Q-CHEM, and confirmed that our analytic gradients are correct by testing them against finite difference results. Since we do not have the benefit of the Hellmann-Feynman theorem for gradients, we expect that the CPU time for a CDFT-CI gradient evaluation should be comparable to that needed for a Hessian evaluation using regular DFT. As some indication of the qualitative similarity, we note that for the $\text{OH} + \text{C}_2\text{H}_6 \leftrightarrow \text{H}_2\text{O} + \text{C}_2\text{H}_5$ system presented below (with the 6-311++G** basis, and 100 radial and 302 angular grid points), some timings are presented in Table I. The gradient evaluation is within roughly a factor of one and half of a DFT

	BLYP	B3LYP
$\partial V/\partial \mathbf{c}$	57	60
W^x	475	474
$S^x, F^{(x)}$	559	561
N^x	1468	1651
V^x	44	48
$H^x - \mathcal{E}S^x$	739	1009
CDFT-CI gradient	3342	3804
DFT Hessian	2272	2432

TABLE I. Execution time for each term, in seconds of CPU time.

hessian evaluation, which is reasonable for our comparatively unoptimized code. We have endeavored to retain the $O(N^3)$ scaling behavior of DFT, albeit with a rather large pre-factor. (Each GMRES iteration requires some number of $O(N^3)$ matrix manipulations, and it is not atypical for 20 GMRES iterations to be required for convergence.)

A. Transition State Optimization

The present work found it opportune to return to the set of reactions previously used to evaluate CDFT-CI[88], taken from the HTBH38/04 and NHTBH38 databases of Zhao et al.[94, 95] The newly implemented analytic gradients allow us to locate optimized transition-state geometries at the CDFT-CI level of theory, to compare against the reference geometries which were optimized at a QCISD/MG3 level of theory.[94] Since the CI vector should be strongly spread over both configurations at the transition-state, these geometry optimizations represent a stringent test of the CDFT-CI coupling gradient computation — any inaccuracies would be highlighted by the delocalized CI vector. Furthermore, the change between the CDFT-CI energy calculated at the reference geometry and at the CDFT-CI optimized geometry presents a measure of the “goodness” of the CDFT-CI geometry; systems with small energy change are expected to have a converged CDFT-CI geometry close to the reference geometry. It also presents an opportunity to once again examine the overall quality of CDFT-CI for barrier heights. The present calculations are performed using a 6-311++G** basis set, as opposed to the 6-311+G(3df,2p) basis set used in Reference [88]. For performing gradient evaluations, we also found it necessary to increase the quality of the DFT integration grid (though this is less critical when a smaller basis set is used) — SG-1[96], the default grid in Q-CHEM, is insufficient to yield sufficiently converged SCF calculations. The present work uses a Lebedev grid with 100 radial points and 302 angular points. The exchange-correlation functional used for CDFT-CI is B3LYP (need a citation?). We also deem it sufficient to present results using a single functional, given the overall robust performance with multiple functionals in the previous work.[88] Table II shows the results for the 32 reactions considered, with forward (and backward, when distinct) reaction barrier heights for CDFT-CI at the reference geometry, CDFT-CI at the optimized geometry, and the reference barrier heights. It also presents geometrical RMSD’s of CDFT-CI and B3LYP compared against the reference transition state geometries by using GROMACS[97]. There are some reactions where CDFT-CI and B3LYP were unable to determine a reaction transition state, either because GMRES did not converge or because there was no barrier predicted.

Reaction	Barrier heights			Geom. RMSD	
	QCISD at	CDFT-CI at	CDFT-CI at	CDFT-CI	B3LYP

	QCISD geom.	QCISD geom.	CDFT-CI geom.		
1 $\text{H} + \text{HCl} \leftrightarrow \text{Cl} + \text{H}_2$	5.7	4.71	5.66	0.042	0.060
	8.7	11.09	12.05		
2 $\text{OH} + \text{H}_2 \leftrightarrow \text{H} + \text{H}_2\text{O}$	5.1	7.86	6.35	0.010	0.032
	21.2	19.66	18.15		
3 $\text{CH}_3 + \text{H}_2 \leftrightarrow \text{H} + \text{CH}_4$	12.1	12.60	12.69	0.010	0.009
	15.3	13.51	13.60		
4 $\text{OH} + \text{CH}_4 \leftrightarrow \text{CH}_3 + \text{H}_2\text{O}$	6.7	14.54	-	-	0.025
	19.6	25.43	-		
5 $\text{H} + \text{H}_2 \leftrightarrow \text{H}_2 + \text{H}$	9.6	7.64	7.64	0.003	0.002
6 $\text{OH} + \text{NH}_3 \leftrightarrow \text{H}_2\text{O} + \text{NH}_2$	3.2	2.80	2.89	0.105	0.210
	12.7	11.47	11.56		
7 $\text{HCl} + \text{CH}_3 \leftrightarrow \text{Cl} + \text{CH}_4$	1.7	6.27	5.81	0.055	0.053
	7.9	13.57	13.11		
8 $\text{OH} + \text{C}_2\text{H}_6 \leftrightarrow \text{H}_2\text{O} + \text{C}_2\text{H}_5$	3.4	6.83	-	-	0.361
	19.9	22.28	-		
9 $\text{F} + \text{H}_2 \leftrightarrow \text{H} + \text{HF}$	1.8	1.72	1.32	0.051	-
	33.4	29.64	29.24		
10 $\text{O} + \text{CH}_4 \leftrightarrow \text{OH} + \text{CH}_3$	13.7	23.72	17.93	0.075	0.007
	8.1	19.81	14.02		
11 $\text{H} + \text{PH}_3 \leftrightarrow \text{PH}_2 + \text{H}_2$	3.1	3.21	3.29	0.015	0.065
	23.2	27.63	27.70		
12 $\text{H} + \text{HO} \leftrightarrow \text{H}_2 + \text{O}$	10.7	9.23	11.41	0.053	0.003
	13.1	12.23	14.41		
13 $\text{H} + \text{H}_2\text{S} \leftrightarrow \text{H}_2 + \text{HS}$	3.5	5.20	4.78	0.032	0.045
	17.3	22.56	22.14		
14 $\text{O} + \text{HCl} \leftrightarrow \text{OH} + \text{Cl}$	9.8	5.10	5.86	0.028	0.044
	10.4	8.49	9.25		
15 $\text{NH}_2 + \text{CH}_3 \leftrightarrow \text{CH}_4 + \text{NH}$	8.0	9.70	9.69	0.119	0.054
	22.4	21.92	21.91		
16 $\text{NH}_2 + \text{C}_2\text{H}_5 \leftrightarrow \text{C}_2\text{H}_6 + \text{NH}$	7.5	11.68	11.10	0.37	0.034
	18.3	19.34	18.77		
17 $\text{C}_2\text{H}_6 + \text{NH}_2 \leftrightarrow \text{NH}_3 + \text{C}_2\text{H}_5$	10.4	12.97	12.83	0.038	0.108
	17.4	19.75	19.60		
18 $\text{NH}_2 + \text{CH}_4 \leftrightarrow \text{CH}_3 + \text{NH}_3$	14.5	15.54	15.39	0.111	0.174
	17.8	17.76	17.60		
19 $\text{H} + \text{N}_2\text{O} \leftrightarrow \text{OH} + \text{N}_2$	18.14	15.63	16.21	0.031	0.022

	83.22	79.40	79.98		
20 $H + FH \leftrightarrow FH + H$	42.18	37.29	37.12	0.007	0.004
21 $H + ClH \leftrightarrow HCl + H$	18.0	22.39	22.00	0.009	0.003
22 $H + FCH_3 \leftrightarrow HF + CH_3$	30.38	26.03	25.93		
	57.02	53.61	53.51	0.013	0.009
23 $H + F_2 \leftrightarrow HF + F$	2.27	-1.34	-0.23		
	106.18	104.69	105.80	0.040	-
24 $CH_3 + FCl \leftrightarrow CH_3F + Cl$	7.43	0.94	1.321		
	60.17	58.49	58.86	0.073	-
25 $F^- + CH_3F \leftrightarrow FCH_3 + F^-$	-0.34	0.55	0.25	0.032	0.033
26 $F^- \cdots CH_3F \leftrightarrow FCH_3 \cdots F^-$	13.38	14.76	14.61	0.026	0.013
27 $Cl^- + CH_3Cl \leftrightarrow ClCH_3 + Cl^-$	3.10	1.56	1.43	0.026	0.342
28 $Cl^- \cdots CH_3Cl \leftrightarrow ClCH_3 \cdots Cl^-$	13.61	11.28	11.36		
29 $F^- + CH_3Cl \leftrightarrow FCH_3 + Cl^-$	-12.54	-13.28	-13.45		
	20.11	20.94	20.77	0.024	0.014
30 $F^- \cdots CH_3Cl \leftrightarrow FCH_3 \cdots Cl^-$	2.89	2.53	2.56		
	29.62	29.75	29.75		
31 $OH^- + CH_3F \leftrightarrow HOCH_3 + F^-$	-2.78	-3.23	-3.24		
	17.33	18.65	18.64	0.116	0.116
32 $OH^- \cdots CH_3F \leftrightarrow HOCH_3 \cdots F^-$	10.96	11.28	11.42		
	47.20	46.91	47.25		
Average				0.044	0.060

TABLE II: Reaction barrier heights from various methods in kcal/mol; forward and backward reaction barrier heights are shown when distinct. Reference data of QCISD/MG3 from References [94] and [95], the CDFT-CI energy at the reference geometry, and the CDFT-CI energy at the optimized CDFT-CI geometry are shown. CDFT-CI and regular B3LYP data used the 6-311++G** basis set. For further comparison, also shown are geometrical RMSD's (Å) of CDFT-CI and B3LYP optimized geometries compared against the reference.

Given that CDFT-CI with the reactant/product constrained states as its basis degenerates into ordinary DFT calculations on the reactant or product fragments at infinite separation, in some cases there is conflict between accurate forward and backward barrier heights,

when DFT does not treat the reactant and product states equally well. Over the entire set of reactions (modulo those with no CDFT-CI transition state yet converged), CDFT-CI at the reference geometry has a mean error (ME) of 0.56 kcal/mol and a mean absolute error of 2.63 kcal/mol; using energies from the optimized geometries performs essentially the same, with ME of 0.09 kcal/mol and mean absolute error 2.07 kcal/mol. Splitting the reactions by class, average results are shown in Table III. There appear to be no substantial differences

	ME, initial	ME, optimized	MAE, initial	MAE, optimized
Hydrogen transfer	1.72	0.98	2.68	2.17
Heavy atom transfer	-2.78	-2.45	3.66	3.25
Nucleophilic substitution	-0.07	-0.10	0.88	0.84
All reactions	0.56	0.09	2.63	2.07

TABLE III. Deviation of CDFT-CI barrier heights from the reference values. MEs and mean absolute errors (MAE) are given, broken down by the type of reaction, at the initial (reference) geometry, and at the final optimized geometry.

between the reference geometry results and the optimized geometry results at a per-category level, with the optimized geometries consistently performing slightly better. On the other hand, Table II shows an ME of 0.044 Å for the metric of geometrical RMSD (excluded are the reactions where a transition state was not located by CDFT-CI). CDFT-CI outperforms B3LYP (ME 0.060 Å) in getting the correct geometries, but there are some reactions where CDFT-CI and B3LYP can have a rather large RMSD. It should be mentioned that these results might suffer somewhat from a few distorted or rotated atoms, which are energetically ineffective, if the number of atoms in the reaction is relatively small.

Overall, CDFT-CI seems to produce very good reaction transition states, being essentially statistically indistinguishable from the reference geometries in barrier heights, although actual geometries can be slightly different. We should note that the results might be slightly better if symmetrized (e.g., spherically average) promolecule densities were used.

B. Excited-state optimizations

Gradients on the ground-state allow for geometry optimization of critical points, both minima and saddle points (transition states). It is less common to have gradients of the

excited state energy available, which enable optimization of minima on excited-state electronic PESs. Our CDFT-CI gradient implementation can produce forces on both the ground and excited states, treating them on an equal footing. As a simple example, we optimize the first singlet excited state of H_2 . This molecule has been exhaustively studied, and it is known that the first ${}^1\Sigma_g^+$ excited state has two minima, with the lower-energy minimum at a separation of 1.0 Å and a second minimum at 2.3 Å.[98] The two minima correspond, qualitatively, to a $1s \rightarrow 2s$ Rydberg excitation and an linear combination of ionic states, respectively. Since CDFT is unable to describe valence excitations, it would be surprising if CDFT-CI could reproduce this double-minimum in the excited state. In fact, with the cc-pVDZ basis set and the B3LYP functional, using the standard four-state CDFT-CI active space for diatomics (H^+H^- , H^-H^+ , $\text{H}^\uparrow\text{H}^\downarrow$, and $\text{H}^\downarrow\text{H}^\uparrow$), we find only a single minimum at a separation of 2.235 Å (from an optimization starting at the ground-state equilibrium geometry). Throughout the optimization, the state in question is dominated by contributions from the ionic configurations, as expected given the unavailability of valence-excitation states. Nonetheless, the ionic minimum seems to be treated correctly, given that our modest basis set is not intended to yield quantitatively accurate results. It remains telling, though, that only one minimum is found — a reminder that we must always be conscious of the composition of the active space. If the active space does not include the proper states to describe a portion of configuration space, then the CDFT-CI energy will be unreliable; this active space dependence must be kept in mind when applying CDFT-CI to new systems.

H_2 is well studied in part because it is a very small test system, and it functions as a first test for new theoretical methods. However, the double-minimum in the excited state makes it less useful for assessing the validity of CDFT-CI given that we know it will fail for the valence excitations which comprise half of the double well. It is therefore useful to consider another simple molecule with well-known structure, but which has only a single minimum in the excited state. The diatomic Li_2 meets these criteria; it also has more than two electrons, presenting a somewhat more stringent test on the applicability of theoretical methods. A similar four-state geometry optimization on the first singlet excited state of Li_2 (also using cc-pVDZ and B3LYP) locates a geometry minimum (ungerade) at a separation of 3.187 Å. Furche and coworkers[29, 30] have compiled a benchmark suite of reference adiabatic excitation energies (including relaxation in the excited state), including excited-state geometries for for more than twenty molecules.[29] They give the dilithium

${}^1\Sigma_u^+$ minimum to be at 3.11 Å (experiment[99]), with all the tabulated TD-DFT methods underpredicting the minimum except for TDHF. Given the proper active space, CDFT-CI successfully optimizes geometries for HOMO \rightarrow LUMO excitations in these simple systems.

These simple diatomic systems do not present a compelling case for the necessity of analytic gradients with their single degree of freedom; moving to the ethylene cation adds more degrees of freedom while retaining a chemically simple system. It presents theoretical interest even in the ground state, in particular with the non-planar nature of the equilibrium state.[100–102] Consensus has been reached that the dihedral angle is around 25°,[103–106] but it is difficult to confidently state a more precise value. The doublet nature of this molecule allows for only a four-state CDFT-CI active space to be used once again, splitting the molecule in half through the carbon-carbon bond, so that the two fragments are both CH₂ units. Between those fragments, there are two possible splits for each of the charge and the spin; the product space gives four possible constraints: CH₂⁺CH₂[↑], CH₂⁺CH₂[↓], CH₂[↑]CH₂⁺, and CH₂[↓]CH₂⁺. A CDFT-CI ground-state optimization with that four-CDFT-state basis and again cc-pVDZ/B3LYP finds the dihedral angle at convergence to be 35.55°, even larger than the stock DFT state at 28.33°. However, this angle is known to be sensitive to the quality of the basis set,[106] so we do not necessarily seek quantitative accuracy. Of more interest to us at present is the behavior in the excited state; the minimum on the excited state is known to be at a conical intersection with the ground state, at a perpendicular geometry.[101] A CDFT-CI geometry optimization in the first excited state (starting from the equilibrium ground state structure) proceeds to a perpendicular geometry which is degenerate with the ground state. Unfortunately, Q-CHEM’s geometry optimizer does not treat conical intersections, so there is little more that may be said about this system at present. We are currently working along this line and will present our results on conical intersection in a future article.

The general pattern of a four-state CDFT-CI with simple charge/spin-constrained states has been successful in these previous applications, so it is natural to apply it to another polyatomic molecule (which does not have its excited-state minimum at a conical intersection). Ethane (C₂H₆) can be partitioned similarly to a diatomic (into two methyl groups) but has additional nuclear degrees of freedom, giving a more clear advantage to analytical gradients for geometry optimization. The four CDFT-CI states are now CH₃⁺CH₃⁻, CH₃⁻CH₃⁺, CH₃[↑]CH₃[↓], and CH₃[↓]CH₃[↑]. At the ground-state equilibrium geometry, the C – C distance is

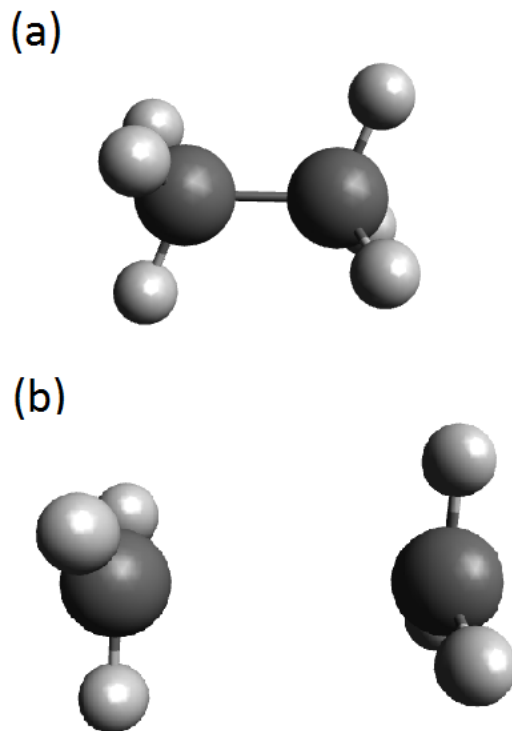


FIG. 2. Structures of ethane optimized by CDFT-CI: (a) ground state, (b) excited state.

1.53 Å with B3LYP/cc-pVDZ; the excited-state minimum for the ionic-like configuration has the carbons some 3.06 Å apart from each other (Figure 2). This is similar to the diatomics previously studied, a little more than twice the ground-state separation, indicating a commonality amongst the ionic-like minima. Here, the methyl groups have both become essentially planar and are parallel to each other, though they retain the staggered rotational conformation. The substantial geometry change is consistent with the mostly-continuous optical spectrum of ethane, given the minimal overlap with the ground state.[107] Again, CDFT-CI successfully locates the excited-state geometry of HOMO \rightarrow LUMO excitations, given a sufficient active space.

As can be seen, with the CDFT-CI energy gradient for excited states, we are now able to study more about excited state chemistry by locating the geometries of conical intersection, energy minima of states with significant charge transfer and valence mixing, all of which the regular DFT cannot even describe.

IV. CONCLUSIONS

We have derived and implemented the equations necessary to obtain analytical gradients of the CDFT-CI energy. The resulting implementation has been used to validate previously investigated reaction barrier heights at self-consistently optimized transition-state geometries, which have good accuracy as compared against reference values computed by high-level theory. Gradients are available equally for the ground state and electronic excited states, allowing for optimization of excited state geometries. As a density-functional method, CDFT-CI has potential application to large systems, with gradients allowing for excited-state dynamics on organic photoelectronic systems at the donor/acceptor interface, even with QM/MM embedding. There is also great potential in CDFT-CI as an economical method for tracking the decay of optically excited systems, including decay to conical intersections. However, that potential comes with a caveat, namely that the user must choose the active space for the calculation. Finding active spaces which remain valid over the entire area of the PES in question may prove to be challenging. Thus, it remains something of an open question what constitutes a “good” or “sufficient” active space for CDFT-CI calculations. Diatomics of low bond order seem well-understood, and the reactant/product split for the set of reaction transition states examined in this work produced good results, but no study has been made of whether increasing the active space would produce further improvement in transition states or elsewhere. Perhaps including configurations with charge-transfer character would shift the location of reaction transition states; the ability to optimize transition-state geometries allows any such effects to be studied, and the results used to give guidance for the selection of active spaces in general.

The availability of diabatic couplings and coupling gradients make possible another investigation of key interest to chemists: studying the Condon approximation that the electronic coupling is relatively invariant to changes in nuclear position. Now that we have implemented the gradient of the coupling element between states, we can proceed to throw it away (set $S_{12}^x = H_{12}^x = 0$ so that $\mathcal{E}^x \approx C_1^2 H_{11}^x + C_2^2 H_{22}^x$) and see how the omission changes the resulting nuclear dynamics. If the changes are small, then the Condon approximation can be safely applied for substantial computational speedup. Having the coupling derivative available allows the validity of the Condon approximation to be assessed on a system-by-system basis, giving greater confidence in the ensuing results.

Additionally, electronic excited states remain ever-tantalizing; to further assess CDFT-CI’s usability in this space, it would be fruitful to study simple photoisomerization systems. With only one bond changing, the difficulty of selecting a CDFT-CI active space is reduced, making isomerization studies feasible. Such studies would help bring insight into how to choose CDFT-CI active spaces for effective description of electronic excited states, helping to bring the DFT toolbox into scope for studying the photochemistry of more generic large molecules. Finally, with the efficiently available CDFT-CI energy gradient, an algorithm to locate minimum energy conical intersections is in need to be developed, which will enable us to study photochemistry with CDFT-CI more thoroughly.

V. ACKNOWLEDGEMENTS

This work was supported by an NSF-CAREER Award (CHE-0547877). T.V. gratefully acknowledges a Packard Fellowship.

Appendix A: GMRES

The general class of iterative linear solvers endeavor to obtain an approximate solution to the linear system

$$\mathbf{A} \cdot x = b$$

without explicitly operating on the matrix \mathbf{A} , instead only evaluating the matrix-vector product $\mathbf{A} \cdot v$. In the case of GMRES, \mathbf{A} must be square, though there exist iterative solvers which can act on non-square systems. The iterative solver gets its name by successively applying the matrix \mathbf{A} to an initial vector and then iterating to generate a series of vectors which is used to produce an active space. After each iteration, the algorithm produces a trial solution vector for that cycle of the algorithm and determines whether this approximate solution x_n is a “good enough” solution to the system in question, that is, whether the residual $|\mathbf{A}x_n - b|$ is smaller than some convergence threshold. In the absence of other information about the system, the only vector available to use as a seed for the iterations is b , making the iterative series $S_b = \{b, \mathbf{A}b, \mathbf{A}^2b, \dots, \mathbf{A}^nb, \dots\}$. The span of the truncated series at each step is called the n th Krylov space, $K_n = \text{span}\{b, \mathbf{A}b, \mathbf{A}^2b, \dots, \mathbf{A}^nb\}$. The approximate solution vector x_n is chosen as the best solution vector within the Krylov space

K_n , as detailed below. However, in order to work with the Krylov space, a useful basis is needed, given that S_b will quickly become numerically unstable as the largest eigenvalue rapidly becomes dominant with successive iterations. Instead, Arnoldi iterations[91] are used to construct an orthogonal basis Q_n for the Krylov space K_n ; the nature of the construction of K_n allows these basis sets to be a nested collection $Q_n = \{q_0, q_1 \cdots q_n\}$ with $Q_n \subset Q_{n+1}$. Thus, at step n it is only necessary to determine q_n , and the rest of the set is reused from the previous iteration. Accordingly, q_{n-1} contains all the information gained from what would formally be written as $\mathbf{A}^{n-1}b$, so likewise $y_n = \mathbf{A}q_{n-1}$ contains all the new information for Q_n . Decomposing it into components $y_n = y_n^\parallel + y_n^\perp$ with $y_n^\parallel \in K_{n-1}$ and $y_n^\perp \notin K_{n-1}$, it is clear then that $q_n = y_n^\perp$, which may be evaluated using a Gram-Schmidt orthogonalization. ($q_0 = \frac{b}{|b|}$.) Having constructed a useful basis Q_n for the Krylov space K_n , the next step is find the approximate solution vector $x_n \in K_n = \text{span}[Q_n]$ which is the best approximate solution $\mathbf{A}x_n \approx b$ in the least-squares sense.

In order to do so, we use the projection of \mathbf{A} into the basis Q_n . That is, we form a matrix \mathbf{H} of dimension $(n+1) \times n$ where the (i, j) -th element is the overlap of $\mathbf{A}q_j$ with q_i ; alternately, the j th column is the result of $\mathbf{A}q_j$ in the basis Q_n . In summary, this means that

$$\mathbf{A}Q_n = Q_{n+1}\mathbf{H} \tag{A1}$$

and

$$Q_{n+1}^\dagger \mathbf{A} = \mathbf{H}Q_n^\dagger \tag{A2}$$

which we will use later on. Not only is the product $\mathbf{A}q_j$ guaranteed to be in Q_n (by the sequential construction of Q_n and the extra row in \mathbf{H}), but we also have the property that $H_{ij} = 0$ when $i > j + 1$. That is, \mathbf{H} has Hessenberg structure (its lower-1 triangle is all zeros). Since $q_0 = \frac{b}{|b|}$, the target vector b is just $|b|e_0$ when projected into the basis Q_n . In this basis, then, our original equation $\mathbf{A}x = b$ becomes

$$\mathbf{H}y = |b|e_0 \tag{A3}$$

This new equation has a structure which lends itself to direct solution, and is of much smaller dimensionality than the original equation, making it feasible to solve directly. Unfortunately, the linear system in equation (A3) is overdetermined, and thus can only be solved in an approximate sense; here it proves convenient to use the least-squares residual as the metric for solution quality (hence “minimal residual” in GMRES).

However, it is not immediately clear that the vector y which produces the minimum residual $|\mathbf{H}y - |b|e_0|$ corresponds to a solution which will minimize the residual of interest, namely $|\mathbf{A}x - b|$. Fortunately, each of the basis sets Q_n is by construction orthonormal, so Q_n (or $Q_n^{-1} = Q_n^\dagger$) acting on a vector in the appropriate space will not change its norm. As such, we note that

$$|\mathbf{A}x - b| = |\mathbf{Q}_{n+1}^\dagger (\mathbf{A}x - b)| \quad (\text{A4})$$

$$= |\mathbf{H}\mathbf{Q}_n^\dagger x - \mathbf{Q}_{n+1}^\dagger b| \quad (\text{A5})$$

$$= |\mathbf{H}y - |b|e_0| \quad (\text{A6})$$

so a minimum-residual solution to the linear equation in the small basis will have the same residual when transformed back to the full basis. (This is of course not a minimum residual solution in the full basis, since the system has an exact solution in the full basis; it is, however, the best residual from a solution vector in the current Krylov space.)

In fact, another (distinct) orthogonal matrix is needed to actually determine the vector y which instantiates the minimal residual vector; this matrix \mathbf{Q} (again, distinct from any of the \mathbf{Q}_n) stems from the QR factorization of \mathbf{H} , partitioning as $\mathbf{H} = \mathbf{Q}\mathbf{R}$ with \mathbf{Q} orthogonal and \mathbf{R} upper-triangular.[91] It then becomes

$$|\mathbf{A}x - b| = |\mathbf{Q}\mathbf{R}y - |b|e_0| \quad (\text{A7})$$

$$= |\mathbf{R}y - |b|\mathbf{Q}^\dagger e_0| \quad (\text{A8})$$

At this point we note that although \mathbf{R} is upper-triangular, it is still of dimension $(n+1) \times n$, and as such this system remains overdetermined. However, it is clear that since the last row of \mathbf{R} is entirely zero, then the product $\mathbf{R}y$ will always have zero as its last element, and thus the norm of the whole expression will be bounded below by the last element of the vector $|b|\mathbf{Q}^\dagger e_0$. In the general case, the actual upper triangle of \mathbf{R} is nonsingular, so the remaining rows may be solved exactly, making the lower bound sharp.

Determination of the vector y producing the minimum residual then proceeds by normal solution of the triangular system

$$\mathbf{R}y = |b|\mathbf{Q}^\dagger e_0 \quad (\text{A9})$$

after projection to the $n \times n$ subspace.

However, the special structure of the GMRES problem means that we can effect a more efficient solution than using a standard QR partitioning routine and subsequent direct solution and substitution on the upper triangular system. This is because the original matrix \mathbf{H} has Hessenberg structure — we can progress down the lower subdiagonal eliminating element by element with a series of 2×2 rotation blocks, bringing the factorization step down to $O(n^2)$ from the general-case $O(n^3)$. However, since we expect n to never be more than twenty or perhaps thirty in our application, and the dimensions of \mathbf{A} scale as the square of the number of electrons in the system, this speedup probably does not contribute very much to the overall runtime of a CDFT-CI calculation.

In determining the factorization, the matrix \mathbf{Q} need not be maintained explicitly and only $|b\rangle\mathbf{Q}^{-1}e_0$ need be stored; this can be updated in-place if the factorization is effected via pairwise rotations as described above. The solution of the upper-triangular system for y is then straightforward; given this y_n , at cycle n of the GMRES algorithm, the approximate solution vector in the full basis is given by $x_n = \sum_{i=0}^n [y_n]_i q_i = \mathbf{Q}_n y_n$ through a standard change of basis.

Amusingly, GMRES is a special case of the Direct Inversion in the Iterative Subspace (DIIS) method which is almost universally used as a convergence accelerator for electronic structure SCF calculations.

Appendix B: Evaluating V_k^x

We start off with knowledge that the constraints must be satisfied initially

$$N_k = \langle \Phi | \hat{w}_k | \Phi \rangle. \quad (\text{B1})$$

A change in the constraint potential will manifest as a change in the Fock matrix, δF , which will induce a change in the wavefunction, $|\delta\Phi\rangle$. This change in the wavefunction will then contribute to the change in N_k , δN_k . Therefore,

$$\delta N_k = \langle \Phi | \delta \hat{w}_k | \Phi \rangle + 2 \langle \Phi | \hat{w}_k | \delta \Phi \rangle \quad (\text{B2})$$

$$= \langle \Phi | \delta \hat{w}_k | \Phi \rangle + 2 \sum_i \langle \phi_i | \hat{w}_k | \delta \phi_i \rangle \quad (\text{B3})$$

$$= \langle \Phi | \delta \hat{w}_k | \Phi \rangle + 2 \sum_{ip} \langle \phi_i | \hat{w}_k | \phi_p \rangle \langle \phi_p | \delta \phi_i \rangle. \quad (\text{B4})$$

The orbital overlaps $\langle \phi_p | \delta \phi_i \rangle$ are computed primarily from perturbation theory, taking care to include corrections from the nonorthogonal basis;

$$\langle \phi_p | \delta \phi_i \rangle = \frac{\langle \phi_p | \delta \hat{F} - \epsilon_i \delta \hat{S} | \phi_i \rangle}{\epsilon_i - \epsilon_p} \quad (\text{B5})$$

$$= \frac{\langle \phi_p | \delta \hat{F} - \epsilon_i \delta \hat{S} + \sum_l \delta \hat{V}_l | \phi_i \rangle}{\epsilon_i - \epsilon_p} \quad (\text{B6})$$

$$= \frac{\langle \phi_p | \delta \hat{F} - \epsilon_i \delta \hat{S} + \sum_l \{(\delta V_l) \hat{w}_l + V_l \delta \hat{w}_l\} | \phi_i \rangle}{\epsilon_i - \epsilon_p} \quad (\text{B7})$$

This expression only holds for $i \neq p$; the contribution from $p = i$ can be determined from the normalization constraint on the orbitals;

$$\langle \phi_i | \delta \phi_i \rangle = -\frac{1}{2} \langle \phi_i | \delta \hat{S} | \phi_i \rangle. \quad (\text{B8})$$

There is no $\delta \epsilon_i$ contribution in Eqs. (B5) through (B8) because only variations of the orbitals which preserve normalization are produced.

Substituting Eqs. (B7) and (B8) into Eq. (B4) yields:

$$\begin{aligned} \delta N_k &= \langle \Phi | \delta \hat{w}_k | \Phi \rangle \\ &+ 2 \sum_{i \neq p} \langle \phi_i | \hat{w}_k | \phi_p \rangle \frac{\langle \phi_p | \delta \hat{F} - \epsilon_i \delta \hat{S} + \sum_l \{(\delta V_l) \hat{w}_l + V_l \delta \hat{w}_l\} | \phi_i \rangle}{\epsilon_i - \epsilon_p} \\ &+ 2 \sum_i \langle \phi_i | \hat{w}_k | \phi_i \rangle \langle \phi_i | \delta \phi_i \rangle, \end{aligned} \quad (\text{B9})$$

$$\begin{aligned} \delta N_k - \langle \Phi | \delta \hat{w}_k | \Phi \rangle &= 2 \sum_{i \neq p} \langle \phi_i | \hat{w}_k | \phi_p \rangle \frac{\langle \phi_p | \sum_l (\delta V_l) \hat{w}_l | \phi_i \rangle}{\epsilon_i - \epsilon_p} \\ &+ 2 \sum_{i \neq p} \langle \phi_i | \hat{w}_k | \phi_p \rangle \frac{\langle \phi_p | \delta \hat{F} - \epsilon_i \delta \hat{S} + \sum_l V_l \delta \hat{w}_l | \phi_i \rangle}{\epsilon_i - \epsilon_p} \\ &- \sum_i \langle \phi_i | \hat{w}_k | \phi_i \rangle \langle \phi_i | \delta \hat{S} | \phi_i \rangle, \end{aligned} \quad (\text{B10})$$

Finally,

$$\begin{aligned} \delta N_k - \langle \Phi | \delta \hat{w}_k | \Phi \rangle &+ \sum_i \langle \phi_i | \hat{w}_k | \phi_i \rangle \langle \phi_i | \delta \hat{S} | \phi_i \rangle \\ - 2 \sum_{i \neq p} \langle \phi_i | \hat{w}_k | \phi_p \rangle \frac{\langle \phi_p | \delta \hat{F} - \epsilon_i \delta \hat{S} + \sum_l V_l \delta \hat{w}_l | \phi_i \rangle}{\epsilon_i - \epsilon_p} &= 2 \sum_l \sum_{i \neq p} \langle \phi_i | \hat{w}_k | \phi_p \rangle \frac{\langle \phi_p | \hat{w}_l | \phi_i \rangle}{\epsilon_i - \epsilon_p} (\delta V_l), \end{aligned} \quad (\text{B11})$$

$$b_k = A_k^l (\delta V_l), \quad (\text{B12})$$

which is a system of linear equations corresponding to the various constraints being applied.

- [1] I. S. K. Kerkines, I. D. Petsalakis, P. Argitis, and G. Theodorakopoulos, *Phys. Chem. Chem. Phys.* **13**, 21273 (2011).
- [2] D. Jacquemin et al., *J. Chem. Phys.* **125**, 164324 (2006).
- [3] M. Hartmann, J. Pittner, and V. Bonačić-Koutecký, *J. Chem. Phys.* **114**, 2123 (2001).
- [4] A. L. Kaledin and K. Morokuma, *J. Chem. Phys.* **113**, 5750 (2000).
- [5] U. Müller and G. Stock, *J. Chem. Phys.* **107**, 6230 (1997).
- [6] D. Schemmel and M. Schütz, *J. Chem. Phys.* **133**, 134307 (2010).
- [7] A. L. Sobolewski and W. Domcke, *J. Phys. Chem. A* **111**, 11725 (2007).
- [8] J. D. Coe and T. J. Martínez, *Mol. Phys.* **106**, 537 (2008).
- [9] M. Etinski and C. M. Marian, *Phys. Chem. Chem. Phys.* **12**, 4915 (2010).
- [10] A. W. Lange and J. M. Herbert, *J. Am. Chem. Soc.* **131**, 3913 (2009).
- [11] H. R. Hudock and T. J. Matrínez, *ChemPhysChem* **9**, 2486 (2008).
- [12] G. Groenhof et al., *J. Am. Chem. Soc.* **129**, 6812 (2007).
- [13] W. J. Schreier et al., *Science* **315**, 625 (2007).
- [14] C. E. Crespo-Hernández, B. Cohen, and B. Kohler, *Nature* **436**, 1141 (2005).
- [15] T. Gustavsson et al., *J. Am. Chem. Soc.* **128**, 607 (2006).
- [16] J.-M. L. Pecourt, J. Peon, and B. Kohler, *J. Am. Chem. Soc.* **123**, 10370 (2001).
- [17] S. Matsika, *J. Phys. Chem. A* **108**, 7584 (2004).
- [18] H. R. Hudock, B. G. Levine, A. L. Thompson, and T. J. Martinez, *AIP Conf. Proc.* **963**, 219 (2007).
- [19] T. Kowalczyk, S. R. Yost, and T. Van Voorhis, *J. Chem. Phys.* **134**, 054128 (2011).
- [20] J. Cornil, D. Beljonne, J.-P. Calbert, and J.-L. Brédas, *Adv. Mater.* **13**, 1053 (2001).
- [21] S. Difley and T. Van Voorhis, *J. Chem. Theory Comput.* **7**, 594 (2011).
- [22] R. H. Friend et al., *Nature* **397**, 121 (1999).
- [23] J.-L. Brédas, J. E. Norton, J. Cornil, and V. Coropceanu, *Acc. Chem. Res.* **42**, 1691 (2009).
- [24] A. J. A. Aquino, M. Barbatti, and H. Lischka, *ChemPhysChem* **7**, 2089 (2006).
- [25] O. Lehtonen, D. Sundholm, and T. Vänskä, *Phys. Chem. Chem. Phys.* **10**, 4535 (2008).
- [26] L. Lapinski, M. J. Nowak, J. Nowacki, M. F. Rode, and A. L. Sobolewski, *ChemPhysChem*

- 10**, 2290 (2009).
- [27] C. V. Caillie and R. D. Amos, *Chem. Phys. Lett.* **308**, 249 (1999).
- [28] C. V. Caillie and R. D. Amos, *Chem. Phys. Lett.* **317**, 159 (2000).
- [29] F. Furche and R. Ahlrichs, *J. Chem. Phys.* **117**, 7433 (2002).
- [30] R. Send, M. Kuhn, and F. Furche, *J. Chem. Theory Comput.* **7**, 2376 (2011).
- [31] M. Barbatti et al., *J. Photochem. Photobiol. A* **190**, 228 (2007).
- [32] J. Plötner and A. Dreuw, *Chem. Phys.* **347**, 472 (2007).
- [33] P. Wiggins, J. A. G. Williams, and D. J. Tozer, *J. Chem. Phys.* **131**, 091101 (2009).
- [34] A. Dreuw and M. Head-Gordon, *Chem. Rev.* **105**, 4009 (2005).
- [35] M. Wormit and A. Dreuw, *J. Phys. Chem. B* **110**, 24200 (2006).
- [36] J. B. Foresman, M. Head-Gordon, J. A. Pople, and M. J. Frisch, *J. Phys. Chem.* **96**, 135 (1992).
- [37] D. Maurice and M. Head-Gordon, *Mol. Phys.* **96**, 1533 (1999).
- [38] C. ik Song and Y. M. Rhee, *J. Am. Chem. Soc.* **133**, 12040 (2011).
- [39] R. Nithya, N. Santhanamoorthi, P. Kolandaivel, and K. Senthilkumar, *J. Phys. Chem. A* **115**, 6594 (2011).
- [40] W. J. Schreier et al., *J. Phys. Chem. B* **115**, 3689 (2011).
- [41] G. Bocchinfuso, C. Mazzuca, A. Palleschi, R. Pizzoferrato, and P. Tagliatesta, *J. Phys. Chem. A* **113**, 14887 (2009).
- [42] K. Andersson and B. O. Roos, *Int. J. Quantum Chem.* **45**, 591 (1993).
- [43] M. Merchán and B. O. Roos, *Theor. Chim. Acta* **92**, 227 (1995).
- [44] A. Rauk, D. Yu, P. Borowski, and B. Roos, *Chem. Phys.* **197**, 73 (1995).
- [45] C. S. Page and M. Olivucci, *J. Comput. Chem.* **24**, 289 (2003).
- [46] S. Olsen, K. Lamonthe, and T. J. Martinez, *J. Am. Chem. Soc.* **132**, 1192 (2010).
- [47] X. fang Yu, S. Yamazaki, and T. Taketsugu, *J. Chem. Theory Comput.* **7**, 1006 (2011).
- [48] P. B. Coto, A. Strambi, and M. Olivucci, *Chem. Phys.* **347**, 483 (2008).
- [49] P. Celani and H.-J. Werner, *J. Chem. Phys.* **119**, 5044 (2003).
- [50] J. F. Stanton, *J. Chem. Phys.* **99**, 8840 (1993).
- [51] J. F. Stanton and J. Gauss, *J. Chem. Phys.* **100**, 4695 (1994).
- [52] J. F. Stanton and J. Gauss, *Theor. Chim. Acta* **91**, 267 (1995).
- [53] L. A. Burns, D. Murdock, and P. H. Vaccaro, *J. Chem. Phys.* **130**, 144304 (2009).

- [54] D. Gerbig et al., *J. Am. Chem. Soc.* **132**, 7273 (2010).
- [55] O. Christiansen, J. F. Stanton, and J. Gauss, *J. Chem. Phys.* **108**, 3987 (1998).
- [56] A. Köhn and C. Hättig, *J. Chem. Phys.* **119**, 5021 (2003).
- [57] R. Brause, M. Santa, M. Schmitt, and K. Kleinermanns, *ChemPhysChem* **8**, 1394 (2007).
- [58] Y. Zhang et al., *J. Am. Chem. Soc.* **131**, 13784 (2009).
- [59] E. V. Gromov, I. Burghardt, H. Köppel, and L. S. Cederbaum, *J. Phys. Chem. A* **115**, 9237 (2011).
- [60] S. Smolarek et al., *J. Phys. Chem. A* **115**, 9399 (2011).
- [61] M. F. Rode and A. L. Sobolewski, *J. Phys. Chem. A* **114**, 11879 (2010).
- [62] H.-J. Werner and E.-A. Reinsch, *J. Chem. Phys.* **76**, 3144 (1982).
- [63] R. Shepard et al., *Int. J. Quantum Chem.* **34**, 149 (1988).
- [64] H. Lischka et al., *Phys. Chem. Chem. Phys.* **3**, 664 (2001).
- [65] H. Lischka, M. Dallos, and R. Shepard, *Mol. Phys.* **100**, 1647 (2002).
- [66] B. G. Levine, C. Ko, J. Quenneville, and T. J. Martínez, *Mol. Phys.* **104**, 1039 (2006).
- [67] B. Kaduk and T. Van Voorhis, *J. Chem. Phys.* **133**, 061102 (2010).
- [68] K. Giesbertz and E. Baerends, *Chem. Phys. Lett.* **461**, 338 (2008).
- [69] A. Dreuw and M. Head-Gordon, *J. Am. Chem. Soc.* **126**, 4007 (2004).
- [70] A. Dreuw, J. L. Weisman, and M. Head-Gordon, *J. Chem. Phys.* **119**, 2943 (2003).
- [71] M. J. G. Peach, P. Benfield, T. Helgaker, and D. J. Tozer, *J. Chem. Phys.* **128**, 044118 (2008).
- [72] S. Arulmozhiraja and M. L. Coote, *J. Chem. Theory Comput.* **8**, 575 (2012).
- [73] A. V. Kityk, *J. Phys. Chem. A* **116**, 3048 (2012).
- [74] R. M. Richard and J. M. Herbert, *J. Chem. Theory Comput.* **7**, 1296 (2011).
- [75] G. Sini, J. S. Sears, and J.-L. Brédas, *J. Chem. Theory Comput.* **7**, 602 (2011).
- [76] P. Romaniello et al., *J. Chem. Phys.* **130**, 044108 (2009).
- [77] D. Hofmann, T. Körzdörfer, and S. Kümmel, *Phys. Rev. Lett.* **108**, 146401 (2012).
- [78] N. Kuritz, T. Stein, R. Baer, and L. Kronik, *J. Chem. Theory Comput.* **7**, 2408 (2011).
- [79] P. Elliott, S. Goldson, C. Canahui, and N. T. Maitra, *Chem. Phys.* **391**, 110 (2011).
- [80] M. Filatov and S. Shaik, *Chem. Phys. Lett.* **304**, 429 (1999).
- [81] I. Frank, J. Hutter, D. Marx, and M. Parrinell, *J. Chem. Phys.* **108**, 4060 (1998).
- [82] I. Okazaki, F. Sato, T. Yoshihiro, T. Ueno, and H. Kashiwagi, *J. Mol. Struct. Theochem*

- 451, 109 (1998).
- [83] T. Kowalczyk, T. Tsuchimochi, P.-T. Chen, L. Top, and T. V. Voorhis, *J. Chem. Phys.* **XXX**, XXX (2013).
- [84] Q. Wu and T. Van Voorhis, *Phys. Rev. A* **72**, 024502 (2005).
- [85] Q. Wu and T. Van Voorhis, *J. Chem. Theory Comput.* **2**, 765 (2006).
- [86] Q. Wu and T. Van Voorhis, *J. Phys. Chem. A* **110**, 9212 (2006).
- [87] Q. Wu, C.-L. Cheng, and T. Van Voorhis, *J. Chem. Phys.* **127**, 164119 (2007).
- [88] Q. Wu, B. Kaduk, and T. Van Voorhis, *J. Chem. Phys.* **130**, 034109 (2009).
- [89] R. McWeeny, *Rev. Mod. Phys.* **32**, 335 (1960).
- [90] This definition does give preference to \mathbf{c}_I over \mathbf{c}_J , but this is immaterial given that the structure of M is symmetric with respect to the two states.
- [91] L. N. Trefethen and D. B. III, *Numerical Linear Algebra*, Society for Industrial and Applied Mathematics, Philadelphia, 1997.
- [92] Y. Saad and M. H. Schultz, *SIAM J. Sci and Stat. Comput.* **7**, 856 (1986).
- [93] B. G. Johnson, P. M. W. Gill, and J. A. Pople, *J. Chem. Phys.* **98**, 5612 (1993).
- [94] Y. Zhao, N. González-García, and D. G. Truhlar, *J. Phys. Chem. A* **109**, 2012 (2005).
- [95] Y. Zhao, B. J. Lynch, and D. G. Truhlar, *Phys. Chem. Chem. Phys.* **7**, 43 (2005).
- [96] P. M. W. Gill, B. G. Johnson, and J. A. Pople, *Chem. Phys. Lett.* **209**, 506 (1993).
- [97] B. Hess, C. Kutzner, D. van der Spoel, , and E. Lindahl, *J. Chem. Theory Comput.* **4**, 435 (2008).
- [98] E. R. Davidson, *J. Chem. Phys.* **35**, 1189 (1961).
- [99] K.-P. Huber and G. Herzberg, *Molecular Spectra and Molecular Structure. IV. constants of diatomic molecules*, Van Nostrand Reinhold Co., New York, 1979.
- [100] C. Sannen, G. Rašeev, C. Galloy, G. Fauville, and J. C. Lorquet, *J. Chem. Phys.* **74**, 2402 (1981).
- [101] H. Köppel, L. S. Cederbaum, and W. Domcke, *J. Chem. Phys.* **77**, 2014 (1982).
- [102] K. Takeshita, *J. Chem. Phys.* **95**, 1838 (1991).
- [103] H. Köppel, W. Domcke, L. S. Cederbaum, and W. von Niessen, *J. Chem. Phys.* **69**, 4252 (1978).
- [104] P. M. Dehmer and J. L. Dehmer, *J. Chem. Phys.* **70**, 4574 (1979).
- [105] J. E. Pollard, D. J. Trevor, J. E. Reutt, Y. T. Lee, and D. A. Shirley, *J. Chem. Phys.* **81**,

5302 (1984).

[106] N. C. Handy, R. H. Nobes, and H.-J. Werner, *Chem. Phys. Lett.* **110**, 459 (1984).

[107] G. Herzberg, *Molecular Spectra and Molecular Structure. III. Electronic spectra and electronic structure of polyatomic molecules*, Van Nostrand Co., Princeton, NJ, 1967.