Reduced scaling formulation of CASPT2 analytical gradients using the supporting subspace method

Chenchen Song^{1,2}, Jeffrey B. Neaton^{1,2,3}, and Todd J. Martínez^{4,5}

¹*Department of Physics, University of California Berkeley, Berkeley, CA 94720, USA*

²*Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*

³*Kavli Energy NanoScience Institute of Berkeley, Berkeley, CA 94720, USA*

⁴*Department of Chemistry and the PULSE Institute,*

Stanford University, Stanford, CA 94305, USA

⁵*SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA*

Abstract

We present a reduced scaling and exact reformulation of state specific CASPT2 analytical gradients in terms of the MP2 and Fock derivatives using the supporting subspace method. This work follows naturally from the supporting subspace formulation of the CASPT2 energy in terms of the MP2 energy using dressed orbitals and Fock builds. For a given active space configuration, the terms corresponding to the MP2-gradient can be evaluated with $O(N^5)$ operations, while the rest of the calculations can be computed with $O(N^3)$ operations using Fock builds, Fock gradients, and linear algebra. When tensor-hypercontraction is applied simultaneously, the computational cost can be further reduced to $O(N^4)$ for a fixed active space size. The new formulation enables efficient implementation of CASPT2 analytical gradients by leveraging existing GPU-based MP2 and Fock routines. We present benchmark results that demonstrate the accuracy and performance of the new method. Example applications of the new method in *ab initio* molecular dynamics simulation and constrained geometry optimization are given.

1. Introduction

Ab initio molecular dynamics (AIMD) simulations¹⁻² and geometry optimizations³⁻⁴ have become indispensable tools in computational chemistry. For example, studies of reaction mechanisms often involve the combination of AIMD simulations⁵ to find possible pathways and geometry optimizations (e.g. nudged-elastic band methods⁶) to locate transition states. The prerequisites for all of these calculations are electronic structure methods that provide accurate descriptions of both the potential energy surfaces and forces (i.e. analytic nuclear potential gradients).

Effectively capturing electron correlation is essential for the accurate description of potential energy surfaces, and these correlation effects can be qualitatively separated into static and dynamic contributions. Multireference second-order perturbation theory⁷ uses a multiconfigurational wavefunction reference to capture static correlation and second-order perturbation theory to capture dynamic correlation, providing an efficient way to capture both effects simultaneously. Among the different variants of multireference perturbation theory, the complete active space second order perturbation (CASPT2) method,⁸⁻¹¹ where the multireference wavefunction is obtained from complete active space self-consistent field (CASSCF) calculations,¹²⁻¹³ is one of the most successful. CASPT2 has been widely applied to molecules with multi-reference effects, such as biradicals¹⁴ and transition metal complexes.¹⁵

The first-order wavefunction of CASPT2 is spanned by all possible double excitations from the reference.¹⁶ In practice, the choice of basis functions representing these double excitations may differ in the contraction level, ranging from uncontracted configuration state functions (CSFs),¹⁷⁻¹⁹ partially contracted configurations,²⁰⁻²² and fully internally contracted configurations (ICCs).^{23, 10} Correspondingly, analytical gradients were also developed for different contraction levels. Early developments often used either uncontracted²⁴⁻²⁶ or partially contracted²⁷ bases due to their relative simplicity, but analytical gradients for the fully internally contracted variant have also been recently developed with the help of automated code generation.²⁸ In addition, the density-fitting approximation²⁹ to the two-electron repulsion integrals (ERIs) has been used to improve the efficiency of analytical gradient calculations.³⁰ Analytical gradients for other variants of multireference perturbation theory have also been reported, including n-electron valence

state perturbation theory (NEVPT),³¹⁻³² and driven similarity renormalization group perturbation theory.³³

When the reference space is spanned by a single CASSCF wavefunction, the method is referred to as state-specific CASPT2. The intruder state problem and near-degeneracies are two major challenges for state-specific CASPT2 methods. To address the former problem, the simplest solution is to use level shifting³⁴⁻³⁵ by applying a small constant shift to the zeroth-order Hamiltonian. To address the latter problem, one successful approach incorporates multiple CASSCF states into the reference space, leading to the multi-state CASPT2 method (MS-CASPT2).³⁶ By further imposing invariance with respect to rotations among the reference states,³⁷ this becomes the extended multi-state CASPT2 (XMS-CASPT2).³⁸ Although the level-shift and multi-state approaches effectively solve the aforementioned difficulties, they unavoidably complicate the analytical gradients. In state-specific CASPT2, the second order perturbation energy is invariant with respect to the first order wavefunction amplitude.²⁷ However, both the level-shift and multi-state methods break this invariance. As a result, the nuclear dependence of the first-order wavefunction amplitude has to be taken into account. The Lagrangian formulation of analytical gradients³⁹ provides a simple way to account for this dependence, and has enabled the development of both analytical gradients^{38, 40} and non-adiabatic coupling derivatives⁴¹ of MS- or XMS-CASPT2 with level-shifts applied.

We recently introduced the supporting subspace method⁴² that formulates the state specific CASPT2 and MS or XMS-CASPT2 energies in terms of an MP2 energy and Fock builds. The supporting subspace method establishes a connection between the multireference and single-reference perturbation methods, and makes CASPT2 calculations feasible for system sizes where MP2 can be applied. One would naturally expect that in the supporting subspace formulation of CASPT2, the analytical gradients should be computed with the corresponding MP2 and Fock gradients. In this work, we will focus on reformulating analytical gradients of state-specific CASPT2 with level-shifts using the supporting subspace method.

The structure of this paper is as follows. We will first summarize all the notations and terminologies used throughout the paper. In Section 2.1, we will give a brief review of the main ideas of how CASPT2 is formulated with the supporting subspace. In Section 2.2,

we will then review the definition of the CASPT2 Lagrangian, which is the starting point for deriving analytical gradients. In Section 2.3, we will discuss how the supporting subspace method can be used to solve the Λ -equation efficiently, which determines the Lagrange multiplier for the CASPT2 first-order wavefunction amplitudes. In Section 2.4, we will go through every term in the CASPT2 Lagrangian, and discuss how each one contributes to the total analytical gradients. In Section 2.5 and Section 2.6, by further introducing Laplace quadrature for energy denominators, we show how the quantities derived in Section 2.4 can be formulated with an MP2 gradient ($O(N^5)$ with conventional or density-fitting MP2, or $O(N^4)$ with THC-MP2), Fock gradients ($O(N^3)$) and linear algebra operations ($O(N^3)$). In Section 3, we will present results demonstrating the performance and computational scaling of the new method. We will also discuss results from *ab initio* molecular dynamics simulations and constrained geometry optimizations.

2. Method

We first summarize the notations used in this paper, which are mostly consistent with our previous work on the CASPT2 energy⁴² and we will refer to our prior work as Paper I in the following.

(1) Symbols related to types of orbitals or excitations. Note that all orbital indices are *spin-orbitals*.

- $\mu, \nu, \gamma, \lambda$ label atomic orbitals
- Subclasses of semi-canonical molecular orbitals (obtained from CASSCF). The corresponding orbital coefficients will be denoted as C (e.g. $C_{r\mu}$):
 - i, j label one of the N_{closed} closed orbitals (doubly-occupied in the CASSCF wavefunction)
 - a, b label one of the N_{virt} virtual orbitals (unoccupied in the CASSCF wavefunction)
 - f, g, x, y label closed or virtual orbitals
 - T, U, V, W label one of the N_{active} active orbitals (variable occupation in the CASSCF wavefunction)
 - r, s label arbitrary molecular orbitals (closed, virtual, or active)

- Hole-type and particle-type orbitals:
 - K : hole-type active orbital index. There are at most N_{active} of these orbitals (see Section 2.2 of SI on H1P2 subspace).
 - C : particle-type active orbital index. There are at most N_{active} of these orbitals (see Section 2.3 of SI on H2P1 subspace).
 - h : a hole-type orbital, i.e. an orbital from the union of closed orbitals and hole-type active orbitals. The corresponding coefficients are denoted as $\{\mathbf{C}_{h\mu}\} = \{\mathbf{C}_{i\mu}\} \cup \{\mathbf{C}_{K\mu}\}$.
 - p : a particle-type orbital, i.e. an orbital from the union of virtual orbitals and particle-type active orbitals. The corresponding coefficients are denoted as $\{\bar{\mathbf{C}}_{p\mu}\} = \{\bar{\mathbf{C}}_{a\mu}\} \cup \{\bar{\mathbf{C}}_{C\mu}\}$
- Dressed hole and particle orbitals:
 - η : dressed hole orbitals. The number of such orbitals is same as the hole-type orbitals. The corresponding orbital coefficients are denoted as $\mathbb{C}_{\eta\mu}$.
 - π : dressed particle orbitals. The number of such orbitals is same as the particle-type orbitals. The corresponding orbital coefficients are denoted as $\bar{\mathbb{C}}_{\pi\mu}$.

- D : index for orthonormalized multiple excitations within the active space.

(2) Symbols denoting configuration spaces of different subsets of possible double excitations:

- $H(n_H)P(n_P)$: represents excitations that involve n_H closed (doubly occupied in all reference configurations) orbitals and n_P virtual (unoccupied in all reference configurations) orbitals, where $0 \leq n_H + n_P \leq 4$. The number of excitations involving active orbitals is then $4 - n_H - n_P$.
- L : Leading space; union of H2P2, H1P2 and H2P1
- T : Trailing space, union of H0P2, H1P1, H2P0, H0P1, H1P0 and H0P0
- S : Supporting space, which is complementary to the leading space and has two or more indices corresponding to hole-type (K) or particle-type (C) active orbitals.

- LS: Supported leading space, union of the leading and supporting spaces. The corresponding tensor is indexed as $t_{hp,h'p'}$.

(3) Symbols for other quantities:

- κ : Laplace quadrature point index
- Φ : index for determinants in the CASSCF wavefunction
- m, n : indices for basis functions that span the first order interacting space

(4) Accents:

- Prime (e.g. p and p'): represent different indices of the same type
- Overbar: when appearing in pairs (e.g. $C_{h\mu}$ and $\bar{C}_{p\mu}$), the former represents quantities related to holes, and the latter represents quantities related to particles.
- Underbar: when appearing in pairs (e.g. $C_{h\mu}$ and $\underline{C}_{h\mu}$), the former appears in the energy expression, and the latter represents the corresponding partial derivative (e.g. $\underline{C}_{h\mu} = \frac{\partial E}{\partial C_{h\mu}}$)
- Overarch: when appearing in pairs (e.g. θ_{rs} and $\hat{\theta}_{rs}$), the former appears in the energy expression, and latter is the corresponding Lagrange multiplier in the Lagrangian that enforces constraints on θ_{rs} .
- Left superscript (e.g. $^{LS}b_{hp,h'p'}$): denotes the space that the tensor belongs to
- Right superscript with bracket (e.g. $L^{[CAS]}, L^{[PT2]}$): denotes the type of the contributions. In tables and figures, they are also represented as $L[CAS]$, $L[PT2]$.
- Parenthesis (e.g. $F(d)$): represents a function that takes the quantities in the parenthesis as input parameters.

2.1 Review of CASPT2 and supporting subspace method

In this section, we will review the main ideas of the supporting subspace method, and how it reformulates CASPT2 in terms of Fock-builds given a one-electron density matrix d :

$$\mathbf{F}_{rs}(d) = h_{rs} + \sum_{r's'} (rs || r's') d_{r's'} = h_{rs} + \mathbf{G}_{rs}(d) \quad (1)$$

and the MP2-energy given the orbital energies $\{\varepsilon_i, \varepsilon_a\}$ and coefficients $\{C_{i\mu}, C_{a\mu}\}$

$$E_{MP2}(\varepsilon_i, C_{i\mu}, \varepsilon_a, C_{a\mu}) = -\frac{1}{4} \sum_{iajb} \frac{(ia \parallel jb)(ia \parallel jb)}{-\varepsilon_i + \varepsilon_a - \varepsilon_j + \varepsilon_b} \quad (2)$$

In the above equations, h_{rs} denotes one-electron integrals (i.e. kinetic and nuclear attraction integrals), and $(rs \parallel r's')$ represents the antisymmetrized two-electron repulsion integrals (ERIs) $(rs \parallel r's') = (rs | r's') - (rs' | r's)$. The Fock builds are formally quartic scaling, but in practice often exhibit quadratic scaling when prescreening is applied based on the sizes of the integrals and density matrix elements. The scaling of the MP2 energy will depend on the type of ERIs being used. If either exact ERIs (i.e. conventional MP2) or density-fitting ERIs (i.e. DF-MP2)⁴³ are used, then the formal scaling will be $O(N^5)$. If tensor hyper-contraction is used (i.e. THC-MP2),⁴⁴⁻⁴⁵ then the formal scaling will be $O(N^4)$. To emphasize that the ERIs used in MP2 (referred to as MP2-ERIs) may be different from exact ERIs, we will denote the exact ERIs as $\mathbf{I}_{\mu\nu\lambda\gamma} = (\mu\nu | \lambda\gamma)$, while the MP2-ERIs will be denoted as $\mathbb{I}_{\mu\nu\lambda\gamma} = (\mu\nu | \lambda\gamma)^{[I]}$. Fock-builds with $\mathbb{I}_{\mu\nu\lambda\gamma}$ will then be denoted as $\mathbb{G}_{rs}(d)$ and $\mathbb{F}_{rs}(d)$ respectively.

We now show how to reformulate CASPT2 with the above ingredients, i.e. Fock-builds with exact ERI (i.e. $\mathbf{F}$ and $\mathbf{G}$), MP2 energies with $\mathbb{I}_{\mu\nu\lambda\gamma}$, and Fock-builds with MP2-ERI (i.e. $\mathbb{F}$ and $\mathbb{G}$). The level-shifted single state CASPT2 equation is defined as

$$\sum_n \langle \chi_n | \hat{H}_0 - E^{[0]} + \varepsilon_{shift} | \chi_{n'} \rangle t_{n'} = \langle \chi_n | \hat{H} | 0 \rangle \quad (3)$$

where $E^{[0]} = \langle 0 | \hat{H}_0 | 0 \rangle$ is the zeroth order energy, $|\chi_n\rangle$ represents the fully internally contracted configurations (ICCs), and ε_{shift} is the level-shifting parameter. The second order perturbation energy (including the correction to remove the level-shift) can be computed from the first-order wavefunction amplitude as

$$E^{[2]} = -\sum_n t_n \langle \chi_n | \hat{H} | 0 \rangle - \varepsilon_{shift} \sum_n t_n^2 \quad (4)$$

The zeroth order Hamiltonian is often chosen as

$$\hat{H}_0 = \sum_{rs} \mathbf{F}_{rs} \left(d^{[0]} \right) \cdot \hat{E}_{rs} \quad (5)$$

where $d_{rs}^{[0]} = \langle 0 | \hat{E}_{rs} | 0 \rangle$ is the one-electron density matrix of the reference state. By separating the excitations into the leading and trailing space,⁴² we can represent Eq.(3) as

$$\begin{pmatrix} {}^{L \leftarrow L} \mathbf{\Omega} & {}^{L_3 \leftarrow T_2} \mathbf{\Gamma}^\dagger \\ {}^{T_2 \leftarrow L_3} \mathbf{\Gamma} & {}^{T \leftarrow T} \mathbf{\Theta} \end{pmatrix} \begin{pmatrix} {}^L x \\ {}^T y \end{pmatrix} = \begin{pmatrix} {}^L b_0 \\ {}^T f_0 \end{pmatrix} \quad (6)$$

For a fixed active space size, the dimension of the leading space (containing H2P2, H1P2 and H2P1) scales as $O(N^4)$ with system size, while the dimension of the trailing space scales as $O(N^2)$. We review the details of Eq. (6) in the supporting information (section 2).

Two types of ERIs are involved in this equation: ${}^L b_0$ is computed with the MP2-ERIs $\mathbb{I}_{\mu\nu\lambda\gamma}$, and takes the form

$$[{}^L b_0]_{hp,h'p'} = (hp \| h'p')^{[\mathbb{I}]} \quad (7)$$

${}^T f_0$ is computed from Fock builds using exact ERIs, i.e. $\mathbf{F}_{rs} \left(d^{[c]} \right), d_{\mu\nu}^{[c]} = \sum_i C_{i\mu} C_{i\nu}$ and $\mathbf{G}_{rs} \left(d^{[UV]} \right), d_{r's'}^{[UV]} = \delta_{Ur'} \delta_{Vs'}$. Because the density matrix $d^{[c]}$ is quite dense, we find that using Fock-builds with exact ERIs to compute $\mathbf{F}_{rs} \left(d^{[c]} \right)$ is numerically much more accurate (compared to the use of DF or THC ERIs) with little performance penalty. Note that U and V are active orbital indices whose range is generally smaller than 20. In line with the ideas that reformulate CASSCF with Fock-builds,⁴⁶ the number of Fock-builds required is quadratic with respect to the active orbitals, and constant with respect to system size.

By using Löwdin partitioning⁴⁷ (i.e. block Gaussian elimination), Eq. (6) is equivalent to the reduced linear system below, with dimension equal to the dimension of the trailing space (i.e. $O(N^2)$):

$$\left({}^{T \leftarrow T} \mathbf{\Theta} - {}^{T_2 \leftarrow L_3} \mathbf{\Gamma} \cdot {}^{L_3 \leftarrow L_3} \mathbf{\Omega}^{-1} \cdot {}^{L_3 \leftarrow T_2} \mathbf{\Gamma}^\dagger \right) \cdot {}^T y = {}^T f_0 - {}^{T_2 \leftarrow L_3} \mathbf{\Gamma} \cdot {}^{L_3 \leftarrow L} \mathbf{\Omega}^{-1} \cdot {}^L b_0 \quad (8)$$

and

$${}^L x = {}^{L \leftarrow L} \mathbf{\Omega}^{-1} \cdot \left({}^L b_0 - {}^{L_3 \leftarrow T_2} \mathbf{\Gamma}^\dagger \cdot {}^T y \right) \quad (9)$$

In order to solve Eqs. (8) and (9), the key is to find an efficient way to compute the product ${}^L_Z = {}^{L\leftarrow L}\Omega^{-1} \cdot {}^L_V$, which is the main motivation for introducing the supporting subspace method.

The supporting subspaces are defined to expand ${}^{L\leftarrow L}\Omega$ into the “supported leading space” as

$${}^{LS\leftarrow LS}\Xi = \begin{pmatrix} {}^{L\leftarrow L}\Omega & {}^{L_3\leftarrow S_2}\mathcal{G}^\dagger \\ {}^{S_2\leftarrow L_3}\mathcal{G} & {}^{T\leftarrow T}\mathcal{Q} \end{pmatrix} \quad (10)$$

such that Ξ can be factorized in the form of Kronecker sum as

$$\Xi_{h_1p_1h_2p_2,h_3p_3h_4p_4} = M_{h_1h_3} \oplus_K \bar{M}_{p_1p_3} \oplus_K M_{h_2h_4} \oplus_K \bar{M}_{p_2p_4} \quad (11)$$

Details on Eqs. (10) and (11) can be found in Paper I. By diagonalizing $M_{hh'}$ and $\bar{M}_{pp'}$ separately, each with cost of $O(N^3)$, we obtain

$$M_{hh'} = -\sum_{\eta} U_{h'\eta}^\dagger \omega_{\eta} U_{\eta h} \quad (12)$$

and

$$\bar{M}_{pp'} = \sum_{\pi} \bar{U}_{p'\pi}^\dagger \bar{\omega}_{\pi} \bar{U}_{\pi p} \quad (13)$$

The inversion of Ξ^{-1} can be easily computed from Eqs. (12) and (13) by using the properties of Kronecker sums:

$$[\Xi^{-1}]_{h_1p_1h_2p_2,h_3p_3h_4p_4} = \sum_{\eta\pi\eta'\pi'} U_{h_1\eta}^\dagger \bar{U}_{p_1\pi}^\dagger U_{h_2\eta'}^\dagger \bar{U}_{p_2\pi'}^\dagger \cdot \left(-\omega_{\eta} + \bar{\omega}_{\pi} - \omega_{\eta'} + \bar{\omega}_{\pi'} \right)^{-1} U_{\eta h_3} \bar{U}_{\pi p_3} U_{\eta' h_4} \bar{U}_{\pi' p_4} \quad (14)$$

By further introducing Laplace quadrature for the energy denominators,⁴⁸⁻⁴⁹ we obtain

$$\frac{1}{\left(-\omega_{\eta} + \bar{\omega}_{\pi} - \omega_{\eta'} + \bar{\omega}_{\pi'} \right)} = \sum_{\kappa} \tau_{\eta}^{(\kappa)} \bar{\tau}_{\pi}^{(\kappa)} \tau_{\eta'}^{(\kappa)} \bar{\tau}_{\pi'}^{(\kappa)} \quad (15)$$

where given the Laplace quadrature point $t^{(\kappa)}$, $\tau_{\eta}^{(\kappa)} = \exp(\omega_{\eta} \cdot t^{(\kappa)})$, $\bar{\tau}_{\pi}^{(\kappa)} = \exp(-\bar{\omega}_{\pi} t^{(\kappa)})$,

and Eq. (14) can finally be factorized as

$$\Xi^{-1}_{h_1p_1h_2p_2,h_3p_3h_4p_4} = \sum_{\kappa} T_{h_1h_3}^{(\kappa)} \bar{T}_{p_1p_3}^{(\kappa)} T_{h_2h_4}^{(\kappa)} \bar{T}_{p_2p_4}^{(\kappa)} \quad (16)$$

where

$$T_{h'h}^{(\kappa)} = \sum_{\eta} U_{h'\eta}^{\dagger} \tau_{\eta}^{(\kappa)} U_{\eta h} \quad (17)$$

and

$$\bar{T}_{p'p}^{(\kappa)} = \sum_{\pi} \bar{U}_{p'\pi}^{\dagger} \bar{\tau}_{\pi}^{(\kappa)} \bar{U}_{\pi p} \quad (18)$$

The above analysis shows that Ξ^{-1} is simple to evaluate.

In order to replace the challenging to compute Ω^{-1} with the simpler-to-compute Ξ^{-1} , we search for the supporting tensor δ_v such that

$${}^L z = {}^{L \leftarrow L} \Omega^{-1} \cdot {}^L v = {}^{L \leftarrow LS} \Xi \cdot ({}^L v + {}^S \delta_v) \quad (19)$$

Eq. (19) indicates that the supporting tensor δ_v should satisfy $0 = {}^{S \leftarrow LS} \Xi \cdot ({}^L v + {}^S \delta_v)$.

Therefore, δ_v can be solved from the linear system

$${}^{S \leftarrow S} \Xi \cdot {}^S \delta_v = -{}^{S \leftarrow LS} \Xi \cdot {}^L v \quad (20)$$

Details on how Eq. (20) is solved are described in Paper I, and we will represent the solver as

$${}^S \delta_v = \text{GetSupport}({}^S \zeta_v = {}^{S \leftarrow LS} \Xi \cdot {}^L v) \quad (21)$$

Summarizing the computation of ${}^L z = {}^{L \leftarrow L} \Omega^{-1} \cdot {}^L v$ with the supporting subspace method given a vector ${}^L v$: (1) compute ${}^S \zeta_v$ from ${}^L v$ (see parenthesis of Eq. (21)); (2) call $\text{GetSupport}({}^S \zeta_v)$ to find the corresponding supporting tensor ${}^S \delta_v$; and (3) use Eq. (19) to compute ${}^L z$.

Applying the supporting subspace method to ${}^{L_3 \leftarrow L} \Omega^{-1} \cdot {}^L b_0$, we can then evaluate the right hand side of Eq.(8) as

$${}^T f = {}^T f_0 - {}^T f_b - {}^{T_2 \leftarrow L_3} \Gamma \cdot {}^{L_3 \leftarrow S} \Xi^{-1} \cdot {}^S \delta_b \quad (22)$$

where

$${}^T f_b = {}^{T_2 \leftarrow L_3} \Gamma \cdot {}^{L_3 \leftarrow LS} \Xi^{-1} \cdot {}^{LS} b \quad (23)$$

$${}^S \delta_b = \text{GetSupport}({}^S \zeta_b = {}^{S \leftarrow LS} \Xi \cdot {}^{LS} b) \quad (24)$$

Our previous work on energy has shown that Eqs. (23) and (24) can both be computed with Fock build using the MP2-ERIs. This is also described in Sections 2.6.2 and 2.6.3, which show that the number of Fock-builds required is quadratic with respect to the number of active orbitals and constant with respect to the system size. Once Eq. (22) is computed, the reduced linear system in Eq. (8) can be solved with the supporting subspace method as described in Paper I. We will represent the solver to the reduced CASPT2 linear system (i.e. Eq. (8)) as

$$\begin{pmatrix} {}^T y, {}^S \delta_{(x)} \end{pmatrix} = \text{SolveReducedSystem} \left({}^T f_0, {}^T f_b, {}^S \zeta_b \right) \quad (25)$$

The input $\begin{pmatrix} {}^T f_0, {}^T f_b, {}^S \zeta_b \end{pmatrix}$ are used to construct the right-hand side of the reduced system (see Eq.(22)). The output $\begin{pmatrix} {}^T y, {}^S \delta_{(x)} \end{pmatrix}$ provides the solutions to the wavefunction amplitude, where ${}^S \delta_{(x)}$ is the supporting tensor corresponding to Eq. (9) such that ${}^L x$ can be factorized as

$${}^L x = {}^{L \leftarrow LS} \Xi^{-1} \cdot \left({}^{LS} b - {}^{L_3 \leftarrow T_2} \Gamma^\dagger \cdot {}^T y + {}^S \delta_{(x)} \right) \quad (26)$$

Finally, the CASPT2 correlation energy can be computed with

$$E^{[2]} = - {}^{LS} b \cdot {}^{LS \leftarrow LS} \Xi^{-1} \cdot {}^{LS} b - {}^T y \cdot {}^T f - \varepsilon_{shift} \cdot t^2 \quad (27)$$

where the first term is computed with the MP2 energy in the dressed orbitals as

$$- {}^{LS} b \cdot {}^{LS \leftarrow LS} \Xi^{-1} \cdot {}^{LS} b = E_{MP2} \left(\omega_\eta, \mathbb{C}_{\eta\mu}, \bar{\omega}_\pi, \bar{\mathbb{C}}_{\pi\mu} \right) \quad (28)$$

The norm of the wavefunction amplitude t^2 in Eq. (27) can be computed from the intermediates during the evaluation of Eq. (28), with details described in Paper I.

To summarize the implementation of CASPT2 using the supporting subspace method:

- (1) MP2 energy with integrals $\mathbb{I}_{\mu\nu\lambda\gamma}$ is used to compute Eq. (28);
- (2) Fock-builds with integrals $\mathbb{I}_{\mu\nu\lambda\gamma}$ are used during the transformation of ${}^{LS} b$ to the trailing space (Eq. (23)) and to the supporting space (Eq. (24))
- (3) Fock-builds with the exact integrals $\mathbf{I}_{\mu\nu\lambda\gamma}$ are used to set up $\hat{H}_0$ (Eq.(5)) as well as the right hand side of the trailing space ${}^T f_0$;

- (4) All other operations can be formulated as linear algebra with at most $O(N^3)$ scaling (details are described in Paper I).

2.2 CASPT2 Lagrangian

The starting point for computing the analytical gradient of the CASPT2 energy is the CASPT2 Lagrangian,^{50, 40} which can be separated as

$$L^{[\text{CASPT2}]} = L^{[\text{CAS}]}(E^{[0]}, \hat{\kappa}, \hat{\theta}) + L^{[\text{PT2}]}(t, \hat{t}) \quad (29)$$

The term $L^{[\text{CAS}]}$ contains the energy of the reference state $E^{[0]}$, constraints on the unitary rotations of the molecular orbitals $\frac{\partial}{\partial \kappa_{rs}} = \sum_{\mu} C_{r\mu} \frac{\partial}{\partial C_{s\mu}} - \sum_{\mu} \frac{\partial}{\partial C_{r\mu}} C_{s\mu}$ (with Lagrange multipliers $\hat{\kappa}_{rs}$), and constraints on the unitary rotations of CI coefficients

$\frac{\partial}{\partial \theta_{RS}} = \sum_{\Phi} c_{R\Phi} \frac{\partial}{\partial c_{S\Phi}} - \sum_{\Phi} \frac{\partial}{\partial c_{R\Phi}} c_{S\Phi}$ (with Lagrange multipliers $\hat{\theta}_{RS}$). The exact form of

$L^{[\text{CAS}]}$ will depend on how the reference is computed, e.g. from SA-CASSCF,⁴⁶ FOMO-CASCI⁵¹ or another multi-reference method that provides a CAS-type reference wavefunction. For example, when SA-CASSCF is used to compute the reference, the term $L^{[\text{CAS}]}$ can be written as

$$L^{[\text{CAS}]} = E^{[0]} + \sum_{r < s} \hat{\kappa}_{rs} \frac{\partial E^{SA}}{\partial \kappa_{rs}} + \sum_{R < S} \hat{\theta}_{RS} \frac{\partial E^{SA}}{\partial \theta_{RS}} \quad (30)$$

where the second and terms indicate that the state-averaged energy E^{SA} is stationary with respect to the unitary rotations among MO coefficients and CI coefficients. The term $L^{[\text{PT2}]}$ contains the second-order perturbation energy as well as constraints on how the first-order amplitude is determined, and is composed of the following four terms

$$L^{[\text{PT2}, t]} = - \sum_n t_n \langle \chi_n | \hat{H} | 0 \rangle \quad (31)$$

$$L^{[\text{PT2}, tt]} = - \epsilon_{\text{shift}} \sum_n t_n \langle \chi_n | \chi_{n'} \rangle t_{n'} \quad (32)$$

$$L^{[\text{PT2}, \hat{t}]} = - \sum_n \hat{t}_n \langle \chi_n | \hat{H} | 0 \rangle \quad (33)$$

and

$$L^{[\text{PT2}, \hat{t}]} = \sum_{n, n'} \hat{t}_n \langle \chi_n | \hat{H}_0 - E_0 + \varepsilon_{\text{shift}} | \chi_{n'} \rangle t_n \quad (34)$$

Eqs. (31) and (32) together form the second-order perturbation energy (see Eq. (4)), while Eqs. (33) and (34) guarantee that the t amplitudes satisfy Eq. (3).

The Lagrange multipliers $\hat{\kappa}, \hat{\theta}, \hat{t}$ are chosen such that the Lagrangian $L^{[\text{CASPT2}]}$ is stationary with respect to all parameters, i.e.

$$\frac{\partial L^{[\text{CASPT2}]}}{\partial \kappa_{rs}} = 0, \frac{\partial L^{[\text{CASPT2}]}}{\partial \theta_{RS}} = 0, \frac{\partial L^{[\text{CASPT2}]}}{\partial t_n} = 0 \quad (35)$$

Once the Lagrange multipliers $\hat{\kappa}, \hat{\theta}, \hat{t}$ are solved from the above conditions, the analytical gradients for the CASPT2 total energy $E = E^{[0]} + E^{[2]}$ can be computed as

$$\frac{dE}{d\mathbf{R}} = \frac{\partial L^{[\text{CAS}]}(E^{[0]}, \hat{\kappa}, \hat{\theta})}{\partial \mathbf{R}} + \frac{\partial L^{[\text{PT2}]}(t, \hat{t})}{\partial \mathbf{R}} \quad (36)$$

The gradient contributions from the reference (i.e. the first term $\frac{\partial L^{[\text{CAS}]}}{\partial \mathbf{R}}$) will be treated as a black box through the following interface

$$\frac{\partial L^{[\text{CAS}]}}{\partial \mathbf{R}} = \text{CASCoupledPerturbed} \left(\{c_{0\Phi}\}, \left\{ C_{r\mu}^{[\text{PT2}]} = \frac{\partial L^{[\text{PT2}]}}{\partial C_{r\mu}} \right\}, \left\{ c_{0\Phi}^{[\text{PT2}]} = \frac{\partial L^{[\text{PT2}]}}{\partial c_{0\Phi}} \right\} \right) \quad (37)$$

where $\{c_{0\Phi}\}$ is the CI coefficient of the reference state. The codes encapsulated in the interface should first solve for the Lagrange multipliers $\hat{\kappa}, \hat{\theta}$ from the coupled-perturbed equation corresponding to the chosen reference method, and then compute and return the nuclear derivatives of $L^{[\text{CAS}]}$. In this work, we use SA-CASSCF⁴⁶ to obtain the reference wavefunctions, and our interface is primarily built on top of the works by Snyder et.al.,⁵²⁻⁵³ who have developed efficient analytic gradients of SA-CASSCF states using GPU-accelerated Fock builds and Fock gradients. However, the energies of the SA-CASSCF state satisfy properties (e.g. $\frac{\partial E^{[0]}}{\partial c_{0\Phi}} = 0, \frac{\partial E^{[0]}}{\partial \kappa_{UV}} = 0$) that are generally not possessed by the CASPT2 state. Therefore, further generalization is required to suit the needs of CASPT2. This generalization is discussed in detail in the SI (Section 1), with both working equations as well as pseudo-code.

The calculation of the CASPT2 gradient in Eq. (36) follows three major steps:

(1) Determine the Lagrange multiplier $\hat{t}$ from the condition $\frac{\partial L}{\partial t_n} = 0$. We follow previous

work⁵⁰ in calling the equation that solves $\hat{t}$ as the “ Λ -equation.”

(2) Construct the following partial derivatives:

$$h_{\mu\nu}^{[\text{PT2}]} = \frac{\partial L^{[\text{PT2}]}}{\partial h_{\mu\nu}}, \quad \mathbf{I}_{\mu\nu\lambda\gamma}^{[\text{PT2}]} = \frac{\partial L^{[\text{PT2}]}}{\partial(\mu\nu|\lambda\gamma)}, \quad \mathbb{I}_{\mu\nu\lambda\gamma}^{[\text{PT2}]} = \frac{\partial L^{[\text{PT2}]}}{\partial \mathbb{I}_{\mu\nu\lambda\gamma}}, \quad C_{r\mu}^{[\text{PT2}]} = \frac{\partial L_{PT2}}{\partial C_{r\mu}}, \quad c_{0\Phi}^{[\text{PT2}]} = \frac{\partial L_{PT2}}{\partial c_{0\Phi}} \quad (38)$$

In practice, instead of computing the four-index two-electron integral density $\mathbf{I}_{\mu\nu\lambda\gamma}^{[\text{PT2}]}$, they are computed and stored in the forms of $D_{\mu\nu}^{[J]} \cdot D_{\lambda\gamma}^{[J]}$ for J-Fock gradients, or $D_{\mu\gamma}^{[K]} \cdot D_{\lambda\nu}^{[K]}$ for K-Fock gradients. Similarly, the four-index $\mathbb{I}_{\mu\nu\lambda\gamma}^{[\text{PT2}]}$ is decomposed based on the type of MP2-ERI. For example, if THC is used, then $\mathbb{I}_{\mu\nu\lambda\gamma}^{[\text{PT2}]}$ will be stored as $X_{p\mu}^{[\text{PT2}]} = \frac{\partial L^{[\text{PT2}]}}{\partial X_{\mu}^p}$ and

$Z_{pQ}^{[\text{PT2}]} = \frac{\partial L^{[\text{PT2}]}}{\partial Z^{pQ}}$, where X_{μ}^p and Z^{pQ} are the THC tensors.⁵⁴ We will continue using the four-index notation $\mathbf{I}_{\mu\nu\lambda\gamma}^{[\text{PT2}]}$ and $\mathbb{I}_{\mu\nu\lambda\gamma}^{[\text{PT2}]}$, but readers should understand that these are stored in factorized form.

(3) Given the partial derivatives from Eq. (38), the final CASPT2 analytical gradients can be computed as

$$\frac{dE}{d\mathbf{R}} = \sum_{\mu\nu} h_{\mu\nu}^{[\text{PT2}]} \frac{\partial h_{\mu\nu}}{\partial \mathbf{R}} + \sum_{\mu\nu\lambda\gamma} \mathbf{I}_{\mu\nu\lambda\gamma}^{[\text{PT2}]} \frac{\partial \mathbf{I}_{\mu\nu\lambda\gamma}}{\partial \mathbf{R}} + \sum_{\mu\nu\lambda\gamma} \mathbb{I}_{\mu\nu\lambda\gamma}^{[\text{PT2}]} \cdot \frac{\partial \mathbb{I}_{\mu\nu\lambda\gamma}}{\partial \mathbf{R}} + \frac{\partial L^{[\text{CAS}]}}{\partial \mathbf{R}} (c_{0\Phi}, C_{r\mu}^{[\text{PT2}]}, c_{0\Phi}^{[\text{PT2}]}) \quad (39)$$

where the last term is computed using Eq. (37), and the first three terms require interfacing with codes that evaluate the nuclear derivatives of integrals.

Based on the above outline, in Section 2.3, we will first discuss how the Lagrange multipliers $\{\hat{t}_n\}$ are solved from the Λ -equation using the supporting subspace method. In Section 2.4 we will discuss how the quantities in Eq. (38) are evaluated using the supporting subspace method. In particular, since $L^{[\text{PT2}]}$ is composed of four terms, i.e.

$L^{[\text{PT2},t]}$, $L^{[\text{PT2},tt]}$, $L^{[\text{PT2},\hat{t}]}$ and $L^{[\text{PT2},\hat{t}]}$ (Eqs. (31)-(34)), their contributions will be discussed in Section 2.4.1 through Section 2.4.4 respectively.

2.3 Solving the Λ -equation with the supporting subspace method

From the condition that $\frac{\partial L^{[\text{CASPT2}]}}{\partial t} = 0$, we obtain the linear equation for solving the Lagrange multiplier $\hat{t}$ (i.e. the Λ -equation):

$$\sum_{n,n'} \langle \chi_n | \hat{H}_0 - E_0 + \varepsilon_{\text{shift}} | \chi_{n'} \rangle \hat{t}_{n'} = \langle \chi_n | \hat{H} | 0 \rangle + 2\varepsilon_{\text{shift}} t_n \quad (40)$$

When $\varepsilon_{\text{shift}} = 0$, the Λ -equation in Eq. (40) is identical with Eq.(3), in which case we will simply have $\hat{t} = t$. When $\varepsilon_{\text{shift}} \neq 0$, we can write $\hat{t} = t + \hat{t}^{[\text{shift}]}$, where $\hat{t}^{[\text{shift}]}$ satisfies

$$\sum_{n,n'} \langle \chi_n | \hat{H}_0 - E_0 + \varepsilon_{\text{shift}} | \chi_{n'} \rangle \hat{t}_{n'}^{[\text{shift}]} = 2\varepsilon_{\text{shift}} t_n \quad (41)$$

If the Lagrange multiplier $\hat{t}^{[\text{shift}]}$ is separated into the leading space and trailing space components as $\hat{t}^{[\text{shift}]} = \begin{pmatrix} {}^L \Lambda, {}^T \lambda \end{pmatrix}$, then following Eq. (6), Eq. (41) can be expressed as

$$\begin{pmatrix} {}^{L \leftarrow L} \Omega & {}^{L_3 \leftarrow T_2} \Gamma^\dagger \\ {}^{T_2 \leftarrow L_3} \Gamma & {}^{T \leftarrow T} \Theta \end{pmatrix} \begin{pmatrix} {}^L \Lambda \\ {}^T \lambda \end{pmatrix} = \begin{pmatrix} 2\varepsilon_{\text{shift}} \cdot {}^L x \\ 2\varepsilon_{\text{shift}} \cdot {}^T y \end{pmatrix} \quad (42)$$

Since the left hand side of the Λ -equation is identical to the CASPT2 equation, it can be solved following the same procedure described in Section 1. In order to use the reduced system solver in Eq. (25), we first need to prepare the inputs by following the examples in Eqs. (23)-(24):

$$\begin{aligned} {}^S \zeta_{2ex} &= 2\varepsilon_{\text{shift}} \cdot {}^{S \leftarrow LS} \Xi \cdot {}^L x \\ &= 2\varepsilon_{\text{shift}} \cdot {}^{S \leftarrow LS} \Xi^{-2} \cdot \left({}^{LS} b - {}^{L_3 \leftarrow T_2} \Gamma^\dagger \cdot {}^{T_2} y + {}^S \delta_{(x)} \right) \end{aligned} \quad (43)$$

and

$$\begin{aligned} {}^T f_{2ex} &= 2\varepsilon_{\text{shift}} \cdot {}^{T_2 \leftarrow L_3} \Gamma^\dagger \cdot {}^{L_3 \leftarrow LS} \Xi^{-1} \cdot {}^L x \\ &= 2\varepsilon_{\text{shift}} \cdot {}^{T_2 \leftarrow L_3} \Gamma^\dagger \cdot {}^{L_3 \leftarrow LS} \Xi^{-2} \cdot \left({}^{LS} b - {}^{L_3 \leftarrow T_2} \Gamma^\dagger \cdot {}^{T_2} y + {}^S \delta_{(x)} \right) \end{aligned} \quad (44)$$

where we have used the factorization of Lx in Eq. (26). We can then call the solver to obtain the solution $\hat{t}^{[\text{shift}]} = \left({}^L\Lambda, {}^T\lambda \right)$

$$\left({}^T\lambda, {}^S\delta_{(\lambda)} \right) = \text{SolveReducedSystem} \left(2\epsilon_{\text{shift}} {}^Ty, {}^Tf_{2\epsilon x}, {}^S\zeta_{2\epsilon x} \right) \quad (45)$$

and the leading space component ${}^L\hat{\Lambda}$ can be factorized as

$${}^L\hat{\Lambda} = {}^{L\leftarrow LS} \Xi^{-1} \cdot \left(2\epsilon_{\text{shift}} {}^Lx - {}^{L_3\leftarrow T_2} \Gamma^\dagger \cdot {}^{T_2} \lambda + {}^S\delta_{(\lambda)} \right) \quad (46)$$

Therefore, the Lagrange multiplier $\hat{t} = t + \hat{t}^{[\text{shift}]}$ is factorized as:

$${}^T\hat{t} = {}^Ty + {}^T\hat{\lambda} \quad (47)$$

and

$$\begin{aligned} {}^L\hat{t} &= {}^Lx + {}^L\hat{\Lambda} \\ &= {}^{L\leftarrow LS} \left(\Xi^{-1} + 2\epsilon_{\text{shift}} \Xi^{-2} \right) \left({}^{LS}b - {}^{L_3\leftarrow T_2} \Gamma^\dagger \cdot {}^{T_2}y + {}^S\delta_{(x)} \right) \\ &\quad + {}^{L\leftarrow LS} \Xi^{-1} \cdot \left(-{}^{L_3\leftarrow T_2} \Gamma^\dagger \cdot {}^{T_2} \lambda + {}^S\delta_{(\lambda)} \right) \end{aligned} \quad (48)$$

Eqs. (47)-(48) will be used to compute the quantities required in Eq. (38).

2.4 Computing partial derivatives of $L^{[\text{PT2}]}$ with the supporting subspace method

The four components of $L^{[\text{PT2}]}$ (Eqs. (31)-(34)) depend on two quantities: the wavefunction amplitudes $t = \left({}^Lx, {}^Ty \right)$ with Lx factorized as per Eq. (26), and the corresponding Lagrange multiplier $\hat{t}$ factorized as in Eqs. (47)-(48). The factorizations of t and $\hat{t}$ altogether involve five different tensors: the supported-leading space tensor ${}^{LS}b$, the trailing space tensors Ty and ${}^T\lambda$, and the supporting space tensor ${}^S\delta_{(x)}$ and ${}^S\delta_{(\lambda)}$. To keep the equations concise, we will henceforth omit the left superscript and simply denote these tensors as $\hat{b}$, y , λ , $\delta_{(x)}$ and $\delta_{(\lambda)}$ respectively. In this section, we show how the partial derivatives of $L^{[\text{PT2}]}$ are computed from the above five tensors using the supporting subspace method. Due to the complexity of the method, in the main text, we will only highlight the important results related to the supporting subspace method, i.e. quantities that involve the leading space. For detailed derivations and mathematical proofs, as well as the treatment for the trailing space that is the same as conventional CASPT2, readers are referred to the supporting information (Sections 3-5).

2.4.1 Partial derivatives of $L^{[\text{PT2},t]}$

Detailed analysis of $L^{[\text{PT2},t]}$ and corresponding pseudocode is given in section 3 of the SI. From the expression for $L^{[\text{PT2},t]}$ in Eq. (31) and the factorization of x in Eq. (26), $L^{[\text{PT2},t]}$ can be represented as

$$L^{[\text{PT2},t]} = -y^\dagger \cdot f_0 - \left(\hat{b} - \Gamma^\dagger \cdot y + \delta_{(x)} \right)^\dagger \cdot \Xi^{-1} \cdot \hat{b} \quad (49)$$

Because f_0 is computed with Fock-builds using exact ERIs, its partial derivatives (i.e. $y^\dagger \cdot \frac{\partial f_0}{\partial Q}$, where Q can be $h_{\mu\nu}$, $\mathbb{I}_{\mu\nu\lambda\gamma}$, $C_{r\mu}$ or $c_{0\Phi}$) can all be computed from the responses of Fock-builds (see Sections 1.1 and 3.2 of SI).

To compute responses from the leading space, note that ${}^{LS}b = (hp \parallel h'p')$ depends on both the MP2-ERIs $\mathbb{I}_{\mu\nu\lambda\gamma}$ as well as the hole-type and particle-type molecular orbitals $\{C_{h\mu}, \bar{C}_{p\mu}\}$. The latter further depend on the semi-canonical molecular orbitals $\{C_{r\mu}\}$ and the CI coefficients of the reference $\{c_{0\Phi}\}$. Therefore, we first compute

$$\underline{Q}^{[t]} = -x^\dagger \cdot \frac{\partial \hat{b}}{\partial Q} = -\left(\hat{b} - \Gamma^\dagger \cdot y + \delta_{(x)} \right)^\dagger \cdot \Xi^{-1} \cdot \frac{\partial \hat{b}}{\partial Q} \quad (50)$$

where Q can be one of $Q = \{\mathbb{I}_{\mu\nu\lambda\gamma}, C_{h\mu}, \bar{C}_{p\mu}\}$. We can then compute $\underline{C}_{r\mu}^{[t]}$ as

$$\underline{C}_{i\mu}^{[t]} = y^\dagger \cdot \frac{\partial f_0}{\partial C_{i\mu}} + \underline{C}_{i\mu}^{[t]} \quad (51)$$

$$\underline{C}_{a\mu}^{[t]} = y^\dagger \cdot \frac{\partial f_0}{\partial C_{a\mu}} + \bar{C}_{a\mu}^{[t]} \quad (52)$$

and

$$\underline{C}_{w\mu}^{[t]} = y^\dagger \cdot \frac{\partial f_0}{\partial C_{w\mu}} + \sum_K \underline{C}_{K\mu}^{[t]} \frac{\partial C_{K\mu}}{\partial C_{w\mu}} + \sum_C \bar{C}_{C\mu}^{[t]} \frac{\partial \bar{C}_{C\mu}}{\partial C_{w\mu}} \quad (53)$$

and compute $\underline{C}_{0\Phi}^{[t]}$ as

$$\underline{C}_{0\Phi}^{[t]} = y^\dagger \cdot \frac{\partial f_0}{\partial c_{0\Phi}} + \sum_{K\mu} \underline{C}_{K\mu}^{[t]} \frac{\partial C_{K\mu}}{\partial c_{0\Phi}} + \sum_{C\mu} \bar{C}_{C\mu}^{[t]} \frac{\partial \bar{C}_{C\mu}}{\partial c_{0\Phi}} \quad (54)$$

The partial derivatives of $\{\mathbf{C}_{K\mu}, \bar{\mathbf{C}}_{C\mu}\}$ are provided in Section 3.1 of the SI.

2.4.2 Partial derivatives of $L^{[\text{PT2}, \hat{t}]}$

Using the factorization of $\hat{t}$ in Eqs. (47)-(48), and noting the similarity to Eqs. (31)-(33), the contributions from $L^{[\text{PT2}, \hat{t}]}$ can be computed in highly similar ways to $L^{[\text{PT2}, t]}$. We first compute

$$\begin{aligned} \underline{Q}^{[\hat{t}]} &= -(\mathbf{x} + \Lambda)^\dagger \cdot \frac{\partial \hat{b}}{\partial Q} \\ &= -(\hat{b} - \Gamma^\dagger \cdot \mathbf{y} + \delta_{(x)})^\dagger \cdot \Xi^{-2} \cdot \frac{\partial \hat{b}}{\partial Q} - (-\Gamma^\dagger \cdot \lambda + \delta_{(\lambda)})^\dagger \cdot \Xi^{-1} \cdot \frac{\partial \hat{b}}{\partial Q} \end{aligned} \quad (55)$$

where Q is any of $Q = \{\mathbb{I}_{\mu\nu\lambda\gamma}, \mathbf{C}_{h\mu}, \bar{\mathbf{C}}_{p\mu}\}$. The responses of $\underline{h}_{\mu\nu}^{[\hat{t}]}, \underline{I}_{\mu\nu\lambda\gamma}^{[\hat{t}]}, \underline{\mathbb{I}}_{\mu\nu\lambda\gamma}^{[\hat{t}]}, \underline{C}_{r\mu}^{[\hat{t}]}, \underline{c}_{0\Phi}^{[\hat{t}]}$ can then be computed from the Eq. (55) in the same way as $L^{[\text{PT2}, t]}$.

2.4.3 Partial derivatives of $L^{[\text{PT2}, t]}$

The only partial derivative that $L^{[\text{PT2}, t]}$ contributes to is $\underline{c}_{0\Phi}^{[t]}$, namely

$$\underline{c}_{0\Phi}^{[t]} = -\epsilon_{\text{shift}} \sum_n t_n \cdot \frac{\partial \langle \chi_n | \chi_{n'} \rangle}{\partial c_{0\Phi}} \cdot t_{n'} \quad (56)$$

Following the derivation in Section 4 of the SI, by separating Eq. (56) into different subspaces, it can be computed as

$$\underline{c}_{0\Phi}^{[t]} = -\epsilon_{\text{shift}} \cdot \left[\sum_n y_n^\dagger \cdot \frac{\partial \langle \chi_n | \chi_{n'} \rangle}{\partial c_{0\Phi}} \cdot y_{n'} + \sum_{KK'} D_{KK'}^{[t]} \frac{\partial S_{KK'}}{\partial c_{0\Phi}} + \sum_{CC'} \bar{D}_{CC'}^{[t]} \frac{\partial \bar{S}_{CC'}}{\partial c_{0\Phi}} \right] \quad (57)$$

Contributions of the trailing space (i.e. first term in Eq. (57)) are computed in the same way as conventional CASPT2, and are described in Section 4.2 of the SI. $S_{KK'}$ and $\bar{S}_{CC'}$ are the overlap matrices for H1P2 and H2P1 subspaces respectively, with $\frac{\partial S_{KK'}}{\partial c_{0\Phi}}, \frac{\partial \bar{S}_{CC'}}{\partial c_{0\Phi}}$

defined in Sections 4.1.2 and 4.1.3 of the SI, respectively.

In order to compute $D_{KK'}^{[t]}$ and $\bar{D}_{CC'}^{[t]}$ in Eq.(57), following the proof in Section 4.1.4 of the SI, we define the operator $\odot_{h_1 h_2}$ between two tensors of the supported-leading space as

$${}^{LS}a \odot_{{h_1 h_2}} {}^{LS}\tilde{a} = \frac{1}{4} \sum_{{p h' p'}} {}^{LS}a_{{h_1 p h' p'}} \cdot {}^{LS}\tilde{a}_{{h_2 p h' p'}} + \frac{1}{4} \sum_{{h p p'}} {}^{LS}a_{{h p h_1 p'}} \cdot {}^{LS}\tilde{a}_{{h p h_2 p'}} \quad (58)$$

and we define the operator $\bar{\odot}_{{p_1 p_2}}$ as

$${}^{LS}a \bar{\odot}_{{p_1 p_2}} {}^{LS}\tilde{a} = \frac{1}{4} \sum_{{h h' p'}} {}^{LS}a_{{h p_1 h' p'}} \cdot {}^{LS}\tilde{a}_{{h p_2 h' p'}} + \frac{1}{4} \sum_{{h p h'}} {}^{LS}a_{{h p h' p_1}} \cdot {}^{LS}\tilde{a}_{{h p h' p_2}} \quad (59)$$

In Section 4.1.4 of the SI, we have proved that $D_{K'K}^{[tt]}, \bar{D}_{C'C}^{[tt]}$ can be computed as

$$\begin{aligned} D_{K'K}^{[tt]} &= {}^{LS}x \odot_{{K'K}} {}^{LS}x \\ &= \left(\hat{b} - \Gamma^\dagger \cdot y + \delta_{(x)} \right)^\dagger \cdot \Xi^{-1} \odot_{{K'K}} \Xi^{-1} \cdot \left(\hat{b} - \Gamma^\dagger \cdot y + \delta_{(x)} \right) \end{aligned} \quad (60)$$

$$\begin{aligned} \bar{D}_{C'C}^{[tt]} &= {}^{LS}x \bar{\odot}_{{C'C}} {}^{LS}x \\ &= \left(\hat{b} - \Gamma^\dagger \cdot y + \delta_{(x)} \right)^\dagger \cdot \Xi^{-1} \bar{\odot}_{{C'C}} \Xi^{-1} \cdot \left(\hat{b} - \Gamma^\dagger \cdot y + \delta_{(x)} \right) \end{aligned} \quad (61)$$

In Section 3, we will discuss how tensor products of type $D_{K'K} = (z\Xi^{-n}) \odot_{{K'K}} (\Xi^{-m}z')$ and $\bar{D}_{C'C} = (z\Xi^{-n}) \bar{\odot}_{{C'C}} (\Xi^{-m}z')$ are treated with Laplace quadrature.

2.4.4 Partial derivatives of $L^{[PT2, \hat{t}]}$

The contributions from $L^{[PT2, \hat{t}]}$ are the most complicated among the four. Following earlier work on CASPT2 analytical gradients,⁵⁰ we first rewrite $L^{[PT2, \hat{t}]}$ as

$$L^{[PT2, \hat{t}]} = \sum_{rs} \mathbf{F}_{rs} [d^{[0]}] \cdot d_{rs}^{[\hat{t}]} + \varepsilon_{shift} \sum_n \hat{t}_n \langle \chi_n | \chi_{n'} \rangle t_{n'} \quad (62)$$

where

$$d_{rs}^{[\hat{t}]} = \sum_{nn'} \hat{t}_n \langle \chi_n | [\hat{E}_{rs} - \langle 0 | \hat{E}_{rs} | 0 \rangle] | \chi_{n'} \rangle t_{n'} \quad (63)$$

Denote the first term in Eq. (62) as $E_F(d^{[\hat{t}]}, d^{[0]})$. Since the first term in Eq. (62) contains the Fock matrix, $\underline{h}_{\mu\nu}^{[\hat{t}]}$, $\underline{\mathbf{I}}_{\mu\nu\lambda\gamma}^{[\hat{t}]}$ and $\underline{C}_{r\nu}^{[\hat{t}]}$ can all be directly computed from $d^{[\hat{t}]}$ and $d^{[0]}$ by using the functions that compute the responses of Fock builds as described in Section 1.1 of the SI,

$$\underline{Q}^{[\hat{t}]} = \frac{\partial E_F(d^{[\hat{t}]}, d^{[0]})}{\partial Q} \quad (64)$$

where Q can be $Q = \{h_{\mu\nu}, \mathbf{I}_{\mu\nu\lambda\gamma}, C_{r\mu}\}$. The partial derivative with respect to the CI coefficient $c_{0\Phi}^{[\tilde{u}]}$ can be computed as follows (see derivations in Section 5 of the SI)

$$\begin{aligned} c_{0\Phi}^{[\tilde{u}]} &= \sum_{rs} \mathbf{G}_{rs} [d^{[\tilde{u}]}] \cdot \frac{\partial d_{rs}^{[0]}}{\partial c_{0\Phi}} \\ &+ \sum_{nn'} \hat{t}_n \left\langle \frac{\partial \chi_n}{\partial c_{0\Phi}} \right| \hat{H}_0 - E^{[0]} + \varepsilon_{shift} |\chi_{n'}\rangle t_{n'} + \sum_{nn'} \hat{t}_n \langle \chi_n | \hat{H}_0 - E^{[0]} + \varepsilon_{shift} \left| \frac{\partial \chi_{n'}}{\partial c_{0\Phi}} \right\rangle t_{n'} \quad (65) \\ &+ \left(\sum_n \hat{t}_n t_n \right) \sum_{rs} \mathbf{F}_{rs} [d^{(0)}] \cdot \frac{\partial \langle 0 | E_{rs} | 0 \rangle}{\partial c_{0\Phi}} \end{aligned}$$

The above discussion shows that $d^{[\tilde{u}]}$ defined in Eq. (63) and $c_{0\Phi}^{[\tilde{u},2]}$ defined in Eq. (65) are sufficient to obtain all the partial derivatives of $L^{[PT2,\tilde{u}]}$. We will now discuss how to compute these two quantities using the supporting subspace method. Because the derivations are quite complex, we again highlight the main results in the main text, leaving detailed proofs to the supporting information (Section 5).

In order to compute $d_{rs}^{[\tilde{u}]}$, we first need to compute the following intermediates

$$\begin{aligned} D_{h'h}^{[\tilde{u}]} &= (x + \Lambda) \odot_{h'h} x \\ &= (\hat{b} - \Gamma^\dagger \cdot y + \delta_{(x)})^\dagger \cdot (\Xi^{-1} + 2\varepsilon_{shift} \Xi^{-2}) \odot_{h'h} \Xi^{-1} (\hat{b} - \Gamma^\dagger \cdot y + \delta_{(x)}) \quad (66) \\ &+ (-\Gamma^\dagger \cdot \lambda + \delta_{(\lambda)})^\dagger \cdot \Xi^{-1} \odot_{h'h} \Xi^{-1} (\hat{b} - \Gamma^\dagger \cdot y + \delta_{(x)}) \end{aligned}$$

$$\begin{aligned} \bar{D}_{p'p}^{[\tilde{u}]} &= (x + \Lambda) \bar{\odot}_{p'p} x \\ &= (\hat{b} - \Gamma^\dagger \cdot y + \delta_{(x)})^\dagger \cdot (\Xi^{-1} + 2\varepsilon_{shift} \Xi^{-2}) \bar{\odot}_{p'p} \Xi^{-1} (\hat{b} - \Gamma^\dagger \cdot y + \delta_{(x)}) \quad (67) \\ &+ (-\Gamma^\dagger \cdot \lambda + \delta_{(\lambda)})^\dagger \cdot \Xi^{-1} \bar{\odot}_{p'p} \Xi^{-1} (\hat{b} - \Gamma^\dagger \cdot y + \delta_{(x)}) \end{aligned}$$

and

$$\begin{aligned} \underline{\Gamma}_{D,Ng}^{(\tilde{u})} &= (x + \Lambda) \circ_{D,Ng} y + (y + \lambda) \circ_{D,Ng} x \\ &= (\hat{b} - \Gamma y + \delta_{(x)})^\dagger (2\Xi^{-1} + 2\varepsilon_{shift} \Xi^{-2}) \circ_{D,Ng} y \quad (68) \\ &+ (-\Gamma \lambda + \delta_{(\lambda)})^\dagger \Xi^{-1} \circ_{D,Ng} y + (\hat{b} - \Gamma y + \delta_{(x)})^\dagger \Xi^{-1} \circ_{D,Ng} \lambda \end{aligned}$$

where the operator $\circ_{D,Ng}$ between a leading space tensor z and a trailing space tensor v is defined as

$$z \circ_{D,Ng} v = z \cdot \frac{\partial \Gamma^\dagger}{\partial \Gamma_{D,Ng}} v \quad (69)$$

The density matrix $d_{rs}^{[\hat{u}]}$ is then computed from Eqs. (66)-(68) as

$$d_{ij}^{[\hat{u}]} = d_{ij}^{[\hat{u},T]} - D_{ij}^{[\hat{u}]} \quad (70)$$

$$d_{ab}^{[\hat{u}]} = d_{ab}^{[\hat{u},T]} + \bar{D}_{ab}^{[\hat{u}]} \quad (71)$$

$$d_{UV}^{[\hat{u}]} = -\sum_{K'K} D_{K'K}^{[\hat{u}]} \cdot \mathbf{T}_{K'K,UV}^0 + \sum_{C'C} \bar{D}_{C'C}^{[\hat{u}]} \cdot \bar{\mathbf{T}}_{C'C,UV}^0 - \left(\sum_K D_{KK}^{[\hat{u}]} + \sum_C \bar{D}_{CC}^{[\hat{u}]} \right) \langle 0 | \hat{E}_{UV} | 0 \rangle \quad (72)$$

$$d_{iU}^{[\hat{u}]} + d_{Ui}^{[\hat{u}]} = d_{iU}^{[\hat{u},T]} + d_{Ui}^{[\hat{u},T]} + \sum_K \left(D_{Ki}^{[\hat{u}]} + D_{iK}^{[\hat{u}]} \right) \cdot \varpi_{KU} + \sum_{DN} \Gamma_{D, Ni}^{[\hat{u}]} \cdot \gamma_{D, NW} \quad (73)$$

and

$$d_{aU}^{[\hat{u}]} + d_{Ua}^{[\hat{u}]} = d_{aU}^{[\hat{u},T]} + d_{Ua}^{[\hat{u},T]} + \sum_C \left(\bar{D}_{Ca}^{[\hat{u}]} + \bar{D}_{aC}^{[\hat{u}]} \right) \cdot \bar{\varpi}_{CU} + \sum_{DN} \Gamma_{D, Na}^{[\hat{u}]} \cdot \gamma_{D, NW} \quad (74)$$

The quantities with superscript T (e.g. $d_{ij}^{[\hat{u},T]}$) only depend on the trailing space properties, and can be computed in the same way as conventional CASPT2 (also discussed in Section 5.3 of the SI). $\mathbf{T}_{K'K,UV}^0$ (defined in Section 2.2 of the SI) and $\bar{\mathbf{T}}_{C'C,UV}^0$ (defined Section 2.3 of the SI) in Eq. (72) are related to the H1P2 and H2P1 subspaces respectively. ϖ_{KU} (defined in Section 2.5 of the SI) and $\bar{\varpi}_{CU}$ (defined in Section 2.6 of the SI) are related to the coupling of H1P2 to H2P2 and coupling of H2P1 to H2P2 respectively. $\gamma_{D, NW}$ is related to the coupling between the trailing and leading space in general (see Section 2.7 of the SI).

The calculation of $\underline{c}_{0\Phi}^{[\hat{u},2]}$ requires a few other intermediates in addition to the quantities that are already computed from Eqs. (66)-(68). From the proof in Section 5.1 of the SI, we first define operator $\square_{h'h}$ between two tensors of the supported leading space as

$$\begin{aligned} {}^{LS} a_{\square_{h'h}} {}^{LS} \tilde{a} &= \frac{1}{4} \sum_{p_1 h_2 p_2, p_3 h_4 p_4} {}^{LS} a_{h' p_1 h_2 p_2} \cdot \left(\bar{M}_{p_1 p_3} \oplus_{\mathbf{K}} M_{h_2 h_4} \oplus_{\mathbf{K}} \bar{M}_{p_2 p_4} \right) \cdot {}^{LS} \tilde{a}_{h p_3 h_4 p_4} \\ &+ \frac{1}{4} \sum_{h_1 p_1 p_2, h_3 p_3 p_4} {}^{LS} a_{h_1 p_1 h' p_2} \cdot \left(M_{h_1 h_3} \oplus_{\mathbf{K}} \bar{M}_{p_1 p_3} \oplus_{\mathbf{K}} \bar{M}_{p_2 p_4} \right) \cdot {}^{LS} \tilde{a}_{h_3 p_3 h p_4} \end{aligned} \quad (75)$$

and also the related operator $\bar{\square}_{p'p}$ as

$$\begin{aligned}
{}^{LS} a_{\bar{p}'p} {}^{LS} \tilde{a} &= \frac{1}{4} \sum_{h_1 h_2 p_2, h_3 h_4 p_4} {}^{LS} a_{h_1 p' h_2 p_2} \cdot \left(M_{h_1 h_3} \oplus_{\mathbf{K}} M_{h_2 h_4} \oplus_{\mathbf{K}} \bar{M}_{p_2 p_4} \right) \cdot {}^{LS} \tilde{a}_{h_3 p h_4 p_4} \\
&+ \frac{1}{4} \sum_{h_1 p_1 h_2, h_3 p_3 h_4} {}^{LS} a_{h_1 p_1 h_2 p'} \cdot \left(M_{h_1 h_3} \oplus_{\mathbf{K}} \bar{M}_{p_1 p_3} \oplus_{\mathbf{K}} M_{h_2 h_4} \right) \cdot {}^{LS} \tilde{a}_{h_3 p_3 h_4 p}
\end{aligned} \tag{76}$$

We then compute the following intermediates

$$\begin{aligned}
W_{h'h}^{[\tilde{u}]} &= (x + \Lambda) \square_{h'h} x \\
&= \left(\hat{b} - \Gamma^\dagger \cdot y + \delta_{(x)} \right)^\dagger \cdot \left(\Xi^{-1} + 2\varepsilon_{shift} \Xi^{-2} \right) \square_{h'h} \Xi^{-1} \left(\hat{b} - \Gamma^\dagger \cdot y + \delta_{(x)} \right) \\
&+ \left(-\Gamma^\dagger \cdot \lambda + \delta_{(\lambda)} \right)^\dagger \cdot \Xi^{-1} \square_{h'h} \Xi^{-1} \left(\hat{b} - \Gamma^\dagger \cdot y + \delta_{(x)} \right)
\end{aligned} \tag{77}$$

and

$$\begin{aligned}
\bar{W}_{p'p}^{[\tilde{u}]} &= (x + \Lambda) \bar{\square}_{p'p} x \\
&= \left(\hat{b} - \Gamma^\dagger \cdot y + \delta_{(x)} \right)^\dagger \cdot \left(\Xi^{-1} + 2\varepsilon_{shift} \Xi^{-2} \right) \bar{\square}_{p'p} \Xi^{-1} \left(\hat{b} - \Gamma^\dagger \cdot y + \delta_{(x)} \right) \\
&+ \left(-\Gamma^\dagger \cdot \lambda + \delta_{(\lambda)} \right)^\dagger \cdot \Xi^{-1} \bar{\square}_{p'p} \Xi^{-1} \left(\hat{b} - \Gamma^\dagger \cdot y + \delta_{(x)} \right)
\end{aligned} \tag{78}$$

The partial derivative with respect to the CI coefficient as defined in Eq. (65) can finally be computed as

$$\underline{\mathcal{C}}_{\Phi}^{[\tilde{u}]} = \sum_{rs} \mathbf{G}_{rs} \left(d^{[\tilde{u}]} \right) \cdot \frac{\partial d_{rs}^{[0]}}{\partial c_{0\Phi}} + \underline{\mathcal{C}}_{\Phi}^{[\tilde{u}, T]} + \underline{\mathcal{C}}_{\Phi}^{[\tilde{u}, \text{H1P2}]} + \underline{\mathcal{C}}_{\Phi}^{[\tilde{u}, \text{H2P1}]} + \underline{\mathcal{C}}_{\Phi}^{[\tilde{u}, V]} \tag{79}$$

where $\underline{\mathcal{C}}_{\Phi}^{[\tilde{u}, T]}$ represents quantities that only depend on the trailing space (see Section 5.3 of the SI) and

$$\begin{aligned}
\underline{\mathcal{C}}_{\Phi}^{[\tilde{u}, \text{H1P2}]} &= \sum_{K'K} \left[W_{K'K}^{[\tilde{u}]} + \left(\frac{1}{4} \varepsilon_{shift} - E^{[0, act]} \right) \cdot D_{K'K}^{[\tilde{u}]} \right] \cdot \frac{\partial S_{K'K}}{\partial c_{0\Phi}} \\
&+ \sum_{K'KW} D_{K'K}^{[\tilde{u}]} \varepsilon_W \left(-\frac{\partial \mathbf{T}_{WW, K'K}^0}{\partial c_{0\Phi}} - \delta_{K'K} \frac{\partial \langle 0 | \hat{E}_{WW} | 0 \rangle}{\partial c_{0\Phi}} \right)
\end{aligned} \tag{80}$$

$$\begin{aligned}
\underline{\mathcal{C}}_{\Phi}^{[\tilde{u}, \text{H2P1}]} &= \sum_{C'C} \left[\bar{W}_{C'C}^{[\tilde{u}]} + \left(\frac{1}{4} \varepsilon_{shift} - E^{[0, act]} \right) \cdot \bar{D}_{C'C}^{[\tilde{u}]} \right] \cdot \frac{\partial \bar{S}_{C'C}}{\partial c_{0\Phi}} \\
&+ \sum_{C'CW} \bar{D}_{C'C}^{[\tilde{u}]} \varepsilon_W \left(\frac{\partial \bar{\mathbf{T}}_{WW, C'C}^0}{\partial c_{0\Phi}} - \delta_{C'C} \frac{\partial \langle 0 | \hat{E}_{WW} | 0 \rangle}{\partial c_{0\Phi}} \right)
\end{aligned} \tag{81}$$

and

$$\mathcal{C}_{0\Phi}^{[\hat{u},V]} = \sum_{KjW} \left(D_{Kj}^{[\hat{u}]} + D_{jK}^{[\hat{u}]} \right) \cdot \frac{\partial \bar{\omega}_{KW}}{\partial \mathcal{C}_{0\Phi}} F_{Wj} + \sum_{CbW} \left(\bar{D}_{Cb}^{[\hat{u}]} + \bar{D}_{bC}^{[\hat{u}]} \right) \cdot \frac{\partial \bar{\omega}_{CW}}{\partial \mathcal{C}_{0\Phi}} F_{Wb} + \sum_{DNWg} \Gamma_{D,Ng} \cdot \frac{\partial \gamma_{D,NW}}{\partial \mathcal{C}_{0\Phi}} F_{Wg} \quad (82)$$

In Eqs. (80)-(81), $E^{[0,act]} = \sum_U \epsilon_U \langle 0 | \hat{E}_{UU} | 0 \rangle$. In Eq. (82), $F_{Wg} = \mathbf{F}_{Wg} (d^{[0]})$. Partial derivatives of $\mathbf{T}_{KK,UV}^0$, $\bar{\mathbf{T}}_{CC,UV}^0$, and $\bar{\omega}_{CU}$ are all defined in Section 5.1 of the SI. The partial derivative $\gamma_{D,NW}$ is defined in Section 5.2 of the SI.

2.4.5 Summary of intermediate quantities depending on Ξ

In the list below, we summarize the important intermediates related to the leading space that are required to compute the partial derivatives of $L^{[\text{PT2}]}$:

- (1) $L^{[t]}: \mathbb{I}_{\mu\nu\lambda\gamma}^{[t]}, \mathbf{C}_{h\mu}^{[t]}, \bar{\mathbf{C}}_{p\mu}^{[t]}$; Eq. (50);
- (2) $L^{[\hat{t}]}: \mathbb{I}_{\mu\nu\lambda\gamma}^{[\hat{t}]}, \mathbf{C}_{h\mu}^{[\hat{t}]}, \bar{\mathbf{C}}_{p\mu}^{[\hat{t}]}$; Eq. (55);
- (3) $L^{[u]}: D_{KK}^{[u]}$ and $\bar{D}_{CC}^{[u]}$; Eqs. (60) and (61), respectively;
- (4) $L^{[\hat{u}]}: D_{h'h}^{[\hat{u}]}, \bar{D}_{p'p}^{[\hat{u}]}, \Gamma_{D,Ng}^{(\hat{u})}, W_{h'h}^{[\hat{u}]}$, and $\bar{W}_{p'p}^{[\hat{u}]}$; Eqs. (66)-(68) and (77)-(78)

All of the quantities listed above depend on Ξ .

2.5 Laplace quadrature factorization of Ξ^{-n} and its products

As all of the quantities in Section 2.4.5 depend on Ξ , it is crucial to factorize this efficiently. In this section, we show how Laplace quadrature can be used to achieve this goal.

2.5.1 Factorization of Ξ^{-n}

To facilitate the discussion in this section, we first define the function $E_{\Xi}^{(\kappa)}(z, z')$ between two tensors of the supported leading space as

$$E_{\Xi}^{(\kappa)} \left({}^{LS} z, {}^{LS} z' \right) = \frac{1}{4} \sum_{h_1 p_1, h_2 p_2, h_3 p_3, h_4 p_4} z_{h_1 p_1, h_2 p_2} \cdot T_{h_1 h_3}^{(\kappa)} \bar{T}_{p_1 p_3}^{(\kappa)} T_{h_2 h_4}^{(\kappa)} \bar{T}_{p_2 p_4}^{(\kappa)} \cdot z'_{h_3 p_3, h_4 p_4} \quad (83)$$

For example, with the new notation we have $\hat{b} \Xi^{-1} \hat{b} = \sum_{\kappa} E_{\Xi}^{(\kappa)}(\hat{b}, \hat{b})$, as can be seen by referring back to Eq. (16).

We first consider operations of the form $\underline{Q}(z, \underline{z}') = z \Xi^{-n} \cdot \frac{\partial z'}{\partial Q}$, which require efficient factorization of Ξ^{-n} . Such operations are needed to compute responses of $L^{[t]}$ or $L^{[\bar{t}]}$, as well as $\underline{\Gamma}_{D, N_g}$ for the response of $L^{[\bar{t}]}$. From Laplace transformation

$$\frac{1}{s^n} = \int_0^\infty \frac{t^{n-1}}{(n-1)!} e^{-st} dt \quad (84)$$

Eq. (15) can thus be generalized to

$$\frac{1}{(-\omega_\eta + \bar{\omega}_\pi - \omega_{\eta'} + \bar{\omega}_{\pi'})^n} = \sum_\kappa \frac{1}{(n-1)!} \left(t^{(\kappa)} \right)^{n-1} \cdot \tau_\eta^{(\kappa)} \bar{\tau}_\pi^{(\kappa)} \tau_{\eta'}^{(\kappa)} \bar{\tau}_{\pi'}^{(\kappa)} \quad (85)$$

From eigendecomposition of Ξ^{-1} in Eq. (14), we then have

$$\Xi_{h_1 p_1 h_2 p_2, h_3 p_3 h_4 p_4}^{-n} = \sum_\kappa \sigma^{(n, \kappa)} T_{h_1 h_3}^{(\kappa)} \bar{T}_{p_1 p_3}^{(\kappa)} T_{h_2 h_4}^{(\kappa)} \bar{T}_{p_2 p_4}^{(\kappa)} \quad (86)$$

with

$$\sigma^{(n, \kappa)} = \frac{1}{(n-1)!} \left(t^{(\kappa)} \right)^{n-1} \quad (87)$$

Therefore, in order to compute $\underline{Q}$, we define the function

$$Q^{(\kappa)}(z, \underline{z}') = \frac{\partial E_\Xi^{(\kappa)}(z, z')}{\partial z'} \frac{\partial z'}{\partial Q} = \frac{1}{4} \sum_{h_1 p_1 h_2 p_2, h_3 p_3 h_4 p_4} z_{h_1 p_1 h_2 p_2} \cdot T_{h_1 h_3}^{(\kappa)} \bar{T}_{p_1 p_3}^{(\kappa)} T_{h_2 h_4}^{(\kappa)} \bar{T}_{p_2 p_4}^{(\kappa)} \cdot \frac{\partial z'_{h_3 p_3 h_4 p_4}}{\partial Q} \quad (88)$$

We then have

$$\underline{Q} = z \Xi^{-n} \frac{\partial z'}{\partial Q} = \sum_\kappa \sigma^{(n, \kappa)} \underline{Q}^{(\kappa)}(z, \underline{z}') \quad (89)$$

In Eq. (88), we take the derivative with respect to z' . Similarly, we define the following function by taking the derivative with respect to z

$$Q^{(\kappa)}(\underline{z}, z') = \frac{\partial E_\Xi^{(\kappa)}(\underline{z}, z')}{\partial z} \frac{\partial z}{\partial Q} = \frac{1}{4} \sum_{h_1 p_1 h_2 p_2, h_3 p_3 h_4 p_4} \frac{\partial z_{h_1 p_1 h_2 p_2}}{\partial Q} \cdot T_{h_1 h_3}^{(\kappa)} \bar{T}_{p_1 p_3}^{(\kappa)} T_{h_2 h_4}^{(\kappa)} \bar{T}_{p_2 p_4}^{(\kappa)} \cdot z'_{h_3 p_3 h_4 p_4} \quad (90)$$

This shows that when partial derivatives containing different powers of Ξ are required, we only need to compute $\underline{Q}^{(\kappa)}(z, \underline{z}')$ or $\underline{Q}^{(\kappa)}(\underline{z}, z')$, and then scale the components from each Laplace quadrature with the appropriate prefactor before summation.

2.5.2 Factorization of $\Xi^{-n} \odot \Xi^{-m}$ and $\Xi^{-n} \bar{\odot} \Xi^{-m}$

In order to compute $D_{hh}^{[u]}, \bar{D}_{p'p}^{[u]}, D_{hh}^{[\bar{u}],} \bar{D}_{p'p}^{[\bar{u}]}$, we need operations of the form $(z\Xi^{-n}) \odot_{h_1 h_2} (\Xi^{-m} z')$ and $(z\Xi^{-n}) \bar{\odot}_{p_1 p_2} (\Xi^{-m} z')$, as defined in Eqs. (58) and (59) respectively. Using the eigendecomposition of Ξ^{-1} in Eq. (14), and also using the fact that the eigenvectors are orthogonal, i.e.

$$\sum_h U_{\eta_1 h} \cdot U_{h \eta_2}^\dagger = \delta_{\eta_1 \eta_2}, \sum_p \bar{U}_{\pi_1 p} \cdot \bar{U}_{p \pi_2}^\dagger = \delta_{\pi_1 \pi_2} \quad (91)$$

one can prove that

$$\begin{aligned} & \sum_{ph'p'} (z\Xi^{-n})_{h_1 p h' p'} \cdot (\Xi^{-m} z')_{h_2 p h' p'} \\ &= \sum_{\substack{\pi \eta' \pi' \\ \eta_1 \eta_2}} \left\{ (Uz)_{\eta_1 \pi \eta' \pi'} \frac{U_{\eta_1 h_1}}{(-\omega_{\eta_1} + \bar{\omega}_\pi - \omega_{\eta'} + \bar{\omega}_{\pi'})^n} \frac{U_{h_2 \eta_2}^\dagger}{(-\omega_{\eta_2} + \bar{\omega}_\pi - \omega_{\eta'} + \bar{\omega}_{\pi'})^m} (Uz')_{\eta_2 \pi, \eta' \pi'} \right\} \end{aligned} \quad (92)$$

where $(Uz)_{\eta_1 \pi \eta' \pi'} = \sum_{h_5 p_5 h_6 p_6} U_{\eta_1 h_5} \bar{U}_{\pi p_5} U_{\eta' h_6} \bar{U}_{\pi' p_6} z'_{h_5 p_5 h_6 p_6}$. The evaluation of Eq. (92) requires efficient factorization of the energy denominator product

$$\Delta_{\eta_1 \eta_2, \pi \eta' \pi'}^{(n, m, 0)} = \frac{1}{(-\omega_{\eta_1} + \bar{\omega}_\pi - \omega_{\eta'} + \bar{\omega}_{\pi'})^n} \frac{1}{(-\omega_{\eta_2} + \bar{\omega}_\pi - \omega_{\eta'} + \bar{\omega}_{\pi'})^m} \quad (93)$$

By using partial fraction expansion, we have proved⁵⁵ that Eq. (93) can be evaluated with Laplace quadrature for any integer (n, m) as

$$\Delta_{\eta_1 \eta_2, \pi \eta' \pi'}^{(n, m, 0)} = \sum_{\kappa} \alpha_{\eta_1 \eta_2}^{(n, m, \kappa)} \bar{\tau}_{\pi}^{(\kappa)} \tau_{\eta'}^{(\kappa)} \bar{\tau}_{\pi'}^{(\kappa)} \quad (94)$$

For CASPT2 analytical gradients, we need $(n=1, m=1)$, $(n=1, m=2)$ and $(n=2, m=1)$ where

$$\alpha_{\eta' \eta}^{(1, 1, \kappa)} = \begin{cases} t^{(\kappa)} \tau_{\eta}^{(\kappa)} & \eta' = \eta \\ \frac{\tau_{\eta'}^{(\kappa)} - \tau_{\eta}^{(\kappa)}}{\omega_{\eta'} - \omega_{\eta}} & \eta' \neq \eta \end{cases} \quad (95)$$

and

$$\alpha_{\eta'\eta}^{(1,2,\kappa)} = \begin{cases} \frac{1}{2} \left(t^{(\kappa)} \right)^2 \tau_{\eta}^{(\kappa)} & \eta' = \eta \\ \frac{\tau_{\eta'}^{(\kappa)} - \tau_{\eta}^{(\kappa)}}{\left(\omega_{\eta'} - \omega_{\eta} \right)^2} - \frac{t^{(\kappa)} \tau_{\eta}^{(\kappa)}}{\left(\omega_{\eta'} - \omega_{\eta} \right)} & \eta' \neq \eta \end{cases} \quad (96)$$

and $\alpha_{\eta'\eta}^{(2,1,\kappa)} = \alpha_{\eta\eta'}^{(1,2,\kappa)}$. Therefore, $\left(z \Xi^{-n} \right) \odot_{h_1 h_2} \left(\Xi^{-m} z' \right)$ can be computed as

$$\left(z \Xi^{-n} \right) \odot_{h_1 h_2} \left(\Xi^{-m} z' \right) = \sum_{\kappa} \sum_{\eta_1 \eta_2} U_{h_1 \eta_1}^{\dagger} U_{h_2 \eta_2}^{\dagger} \alpha_{\eta_1 \eta_2}^{(n,m,\kappa)} \sum_{h' h} U_{\eta_1 h'} U_{\eta_2 h} \cdot \underline{T}_{h' h}^{(\kappa)}(z, z') \quad (97)$$

where we have defined the function

$$\begin{aligned} \underline{T}_{h' h}^{(\kappa)}(z, z') &= \frac{\partial E_{\Xi}^{(\kappa)}(z, z')}{\partial T_{h' h}^{(\kappa)}} = \frac{1}{4} \sum_{p_1 h_2 p_2, p_3 h_4 p_4} z_{h' p_1 h_2 p_2} \cdot \bar{T}_{p_1 p_3}^{(\kappa)} T_{h_2 h_4}^{(\kappa)} \bar{T}_{p_2 p_4}^{(\kappa)} \cdot z'_{h p_3 h_4 p_4} \\ &\quad + \frac{1}{4} \sum_{h_1 p_1 p_2, h_3 p_3 p_4} z_{h_1 p_1 h' p_2} \cdot T_{h_1 h_3}^{(\kappa)} \bar{T}_{p_1 p_3}^{(\kappa)} \bar{T}_{p_2 p_4}^{(\kappa)} \cdot z'_{h_3 p_3 h p_4} \end{aligned} \quad (98)$$

Similarly, $\left(z \Xi^{-n} \right) \bar{\odot}_{p_1 p_2} \left(\Xi^{-m} z' \right)$ can be computed as

$$\left(z \Xi^{-n} \right) \bar{\odot}_{p_1 p_2} \left(\Xi^{-m} z' \right) = \sum_{\kappa} \sum_{\pi_1 \pi_2} \bar{U}_{p_1 \pi_1}^{\dagger} \bar{U}_{p_2 \pi_2}^{\dagger} \bar{\alpha}_{\pi_1 \pi_2}^{(n,m,\kappa)} \sum_{p p'} \bar{U}_{\pi_1 p'} \bar{U}_{\pi_2 p} \cdot \bar{\underline{T}}_{p' p}^{(\kappa)}(z, z') \quad (99)$$

where we have defined the function

$$\begin{aligned} \bar{\underline{T}}_{p' p}^{(\kappa)}(z, z') &= \frac{\partial E_{\Xi}^{(\kappa)}(z, z')}{\partial \bar{T}_{p' p}^{(\kappa)}} = \frac{1}{4} \sum_{h_1 h_2 p_2, h_3 h_4 p_4} z_{h_1 p' h_2 p_2} \cdot T_{h_1 h_3}^{(\kappa)} T_{h_2 h_4}^{(\kappa)} \bar{T}_{p_2 p_4}^{(\kappa)} \cdot z'_{h_3 p_3 h_4 p_4} \\ &\quad + \frac{1}{4} \sum_{h_1 p_1 h_2, h_3 p_3 h_4} z_{h_1 p_1 h_2 p'} \cdot T_{h_1 h_3}^{(\kappa)} \bar{T}_{p_1 p_3}^{(\kappa)} T_{h_2 h_4}^{(\kappa)} \cdot z'_{h_3 p_3 h_4 p} \end{aligned} \quad (100)$$

$\bar{\alpha}_{\pi_1 \pi_2}^{(n,m,\kappa)}$ in Eq. (99) can be computed in similar way to $\alpha_{\eta_1 \eta_2}^{(n,m,\kappa)}$ by replacing $\left\{ \omega_{\eta_1}, \omega_{\eta_2} \right\}$ with $\left\{ -\bar{\omega}_{\pi_1}, -\bar{\omega}_{\pi_2} \right\}$, and replacing $\left\{ \tau_{\eta_1}^{(\kappa)}, \tau_{\eta_2}^{(\kappa)} \right\}$ with $\left\{ \bar{\tau}_{\pi_1}^{(\kappa)}, \bar{\tau}_{\pi_2}^{(\kappa)} \right\}$. Later in our discussion, we will denote the transformation of $\underline{T}_{h' h}^{(\kappa)}(z, z')$ in Eq. (97) as $\alpha_{h_1 h_2 \leftarrow h' h}^{(m,n,\kappa)}$, and denote the transformation of $\bar{\underline{T}}_{p' p}^{(\kappa)}(z, z')$ in Eq. (99) as $\bar{\alpha}_{p_1 p_2 \leftarrow p' p}^{(m,n,\kappa)}$; both correspond to operations that scale cubically with respect to system size.

2.5.3 Factorization of $\Xi^{-n} \square \Xi^{-m}$ and $\Xi^{-n} \square \Xi^{-m}$

In order to compute $W_{h'h}^{[\hat{u}]}$ and $\bar{W}_{p'p}^{[\hat{u}]}$, we need operations of the form $(z\Xi^{-n})_{\square_{h'h}}(\Xi^{-m}z')$ and $(z\Xi^{-n})_{\square_{p'p}}(\Xi^{-m}z')$, as defined in Eqs. (75) and (76) respectively. Similar to the discussion in Section 2.5.2, by using the eigendecomposition of Ξ and Eq. (91), one can prove that

$$\begin{aligned} & \sum_{\substack{p_1 h_2 p_2, \\ p_3 h_4 p_4}} (z\Xi^{-n})_{h'p_1 h_2 p_2} \cdot (\bar{M}_{p_1 p_3} \oplus_{\mathbf{K}} M_{h_2 h_4} \oplus_{\mathbf{K}} \bar{M}_{p_2 p_4}) \cdot (\Xi^{-m}z')_{h p_3 h_4 p_4} \\ &= \sum_{\substack{\pi \eta' \pi' \\ \eta_1 \eta_2}} \left\{ (\mathbf{U}z)_{\eta_1 \pi \eta' \pi'} \frac{U_{\eta_1 h_1}}{(-\omega_{\eta_1} + \bar{\omega}_{\pi} - \omega_{\eta'} + \bar{\omega}_{\pi'})^n} \cdot (\bar{\omega}_{\pi} - \omega_{\eta'} + \bar{\omega}_{\pi'}) \cdot \frac{U_{h_2 \eta_2}^{\dagger}}{(-\omega_{\eta_2} + \bar{\omega}_{\pi} - \omega_{\eta'} + \bar{\omega}_{\pi'})^m} (\mathbf{U}z')_{\eta_2 \pi \eta' \pi'} \right\} \end{aligned} \quad (101)$$

which requires efficient factorization of

$$\Delta_{\eta_1 \eta_2, \pi \eta' \pi'}^{(n, m, 1)} = \frac{(\bar{\omega}_{\pi} - \omega_{\eta'} + \bar{\omega}_{\pi'})}{(-\omega_{\eta_1} + \bar{\omega}_{\pi} - \omega_{\eta'} + \bar{\omega}_{\pi'})^n (-\omega_{\eta_2} + \bar{\omega}_{\pi} - \omega_{\eta'} + \bar{\omega}_{\pi'})^m} \quad (102)$$

Using partial fraction expansion, we have proved that⁵⁵ Eq. (93) can be evaluated with Laplace quadrature as

$$\Delta_{\eta_1 \eta_2, \pi \eta' \pi'}^{(n, m, 1)} = \sum_{\kappa} \beta_{\eta_1 \eta_2}^{(n, m, \kappa)} \bar{\tau}_{\pi}^{(\kappa)} \tau_{\eta'}^{(\kappa)} \bar{\tau}_{\pi'}^{(\kappa)} \quad (103)$$

For CASPT2 analytical gradients, we need $(n=1, m=1)$, $(n=1, m=2)$ and $(n=2, m=1)$ where

$$\beta_{\eta' \eta}^{(1, 1, \kappa)} = \begin{cases} \tau_{\eta}^{(\kappa)} + t^{(\kappa)} \tau_{\eta}^{(\kappa)} \cdot \omega_{\eta} & \eta' = \eta \\ \frac{\omega_{\eta'} \tau_{\eta'}^{(\kappa)} - \omega_{\eta} \tau_{\eta}^{(\kappa)}}{\omega_{\eta'} - \omega_{\eta}} & \eta' \neq \eta \end{cases} \quad (104)$$

$$\beta_{\eta' \eta}^{(1, 2, \kappa)} = \begin{cases} t^{(\kappa)} \tau_{\eta}^{(\kappa)} + \frac{1}{2} (t^{(\kappa)})^2 \tau_{\eta}^{(\kappa)} \cdot \omega_{\eta} & \eta' = \eta \\ \frac{\omega_{\eta'}}{(\omega_{\eta'} - \omega_{\eta})^2} (\tau_{\eta'}^{(\kappa)} - \tau_{\eta}^{(\kappa)}) - \frac{\omega_{\eta}}{\omega_{\eta'} - \omega_{\eta}} t^{(\kappa)} \tau_{\eta}^{(\kappa)} & \eta' \neq \eta \end{cases} \quad (105)$$

and $\beta_{\eta'\eta}^{(2,1,\kappa)} = \beta_{\eta\eta'}^{(1,2,\kappa)}$. Therefore, $(z\Xi^{-n})_{\square_{h_1h_2}}(\Xi^{-m}z')$ can be computed as

$$(z\Xi^{-n})_{\square_{h_1h_2}}(\Xi^{-m}z') = \sum_{\kappa} \sum_{\eta_1\eta_2} U_{h_1\eta_1}^\dagger U_{h_2\eta_2}^\dagger \beta_{\eta_1\eta_2}^{(n,m,\kappa)} \sum_{h'h} U_{\eta_1h'} U_{\eta_2h} \cdot \underline{T}_{h'h}^{(\kappa)}(z, z') \quad (106)$$

Similarly $(z\Xi^{-n})_{\bar{\square}_{p_1p_2}}(\Xi^{-m}z')$ can be computed as

$$(z\Xi^{-n})_{\bar{\square}_{p_1p_2}}(\Xi^{-m}z') = \sum_{\kappa} \sum_{\pi_1\pi_2} \bar{U}_{p_1\pi_1}^\dagger \bar{U}_{p_2\pi_2}^\dagger \bar{\beta}_{\pi_1\pi_2}^{(n,m,\kappa)} \sum_{pp'} \bar{U}_{\pi_1p'} \bar{U}_{\pi_2p} \cdot \bar{\underline{T}}_{p'p}^{(\kappa)}(z, z') \quad (107)$$

where $\bar{\beta}_{\pi_1\pi_2}^{(n,m,\kappa)}$ can be computed in similar way to $\alpha_{\eta_1\eta_2}^{(n,m,\kappa)}$ by replacing $\{\omega_{\eta_1}, \omega_{\eta_2}\}$ with $\{-\bar{\omega}_{\pi_1}, -\bar{\omega}_{\pi_2}\}$, and replacing $\{\tau_{\eta_1}^{(\kappa)}, \tau_{\eta_2}^{(\kappa)}\}$ with $\{\bar{\tau}_{\pi_1}^{(\kappa)}, \bar{\tau}_{\pi_2}^{(\kappa)}\}$. Later in our discussion, we will denote the transformation in Eq. (106) as $\beta_{h_1h_2 \leftarrow h'h}^{(m,n,\kappa)}$, and denote the transformation in Eq. (107) as $\bar{\beta}_{p_1p_2 \leftarrow p'p}^{(m,n,\kappa)}$. Note that both Eqs. (97) and (106) depend on $\underline{T}_{h'h}^{(\kappa)}$, while Eqs. (98) and (107) both depend on $\bar{\underline{T}}_{p'p}^{(\kappa)}$. Therefore, in order to compute the various products with different powers of Ξ , we only need to evaluate $\underline{T}_{h'h}^{(\kappa)}$ or $\bar{\underline{T}}_{p'p}^{(\kappa)}$, and then apply the proper transformations.

2.5.4 Summary of Laplace quadrature factorization

We now apply the results from Section 2.5.1 through 2.5.3 to the intermediate quantities listed in Section 2.4.5.

(1) Quantities required by $L^{[\text{PT2}, t]}$:

$$\underline{Q}^{[t]} = -\sum_{\kappa} \left(\underline{Q}^{(\kappa)}(\hat{b}, \hat{\underline{b}}) - \underline{Q}^{(\kappa)}(\Gamma y, \hat{\underline{b}}) + \underline{Q}^{(\kappa)}(\delta_{(x)}, \hat{\underline{b}}) \right) \quad (108)$$

where $\underline{Q}$ can be one of $\underline{Q} = \{\mathbb{I}_{\mu\nu\lambda\gamma}, \mathbf{C}_{h\mu}, \bar{\mathbf{C}}_{p\mu}\}$. Function $\underline{Q}^{(\kappa)}(z, z')$ is defined in Eq. (88).

(2) Quantities required by $L^{[\text{PT2}, \bar{t}]}$:

$$\begin{aligned} \underline{Q}^{[\bar{t}]} = & -\sum_{\kappa} \left(1 + 2\varepsilon_{\text{shift}} \sigma^{(2,\kappa)} \right) \cdot \left(\underline{Q}^{(\kappa)}(\hat{b}, \hat{\underline{b}}) - \underline{Q}^{(\kappa)}(\Gamma y, \hat{\underline{b}}) + \underline{Q}^{(\kappa)}(\delta_{(x)}, \hat{\underline{b}}) \right) \\ & - \sum_{\kappa} \left[-\underline{Q}^{(\kappa)}(\Gamma \lambda, \hat{\underline{b}}) + \underline{Q}^{(\kappa)}(\delta_{(\lambda)}, \hat{\underline{b}}) \right] \end{aligned} \quad (109)$$

where Q can be one of $Q = \{\mathbb{I}_{\mu\nu\lambda\gamma}, \mathbf{C}_{h\mu}, \bar{\mathbf{C}}_{p\mu}\}$.

(3) Quantities required by $L^{[\text{PT2}, tt]}$:

$$\hat{D}_{MM}^{[tt]} = \sum_{\kappa} \hat{\theta}_{MM \leftarrow r'r}^{(1,1,\kappa)} \cdot \left[\begin{aligned} &\hat{T}_{r'r}^{(\kappa)}(\hat{b}, \hat{b}) - 2\hat{T}_{r'r}^{(\kappa)}(\Gamma y, \hat{b}) + 2\hat{T}_{r'r}^{(\kappa)}(\delta_{(x)}, \hat{b}) \\ &+ \hat{T}_{r'r}^{(\kappa)}(\Gamma y, \Gamma y) - 2\hat{T}_{r'r}^{(\kappa)}(\Gamma y, \delta_{(x)}) + \hat{T}_{r'r}^{(\kappa)}(\delta_{(x)}, \delta_{(x)}) \end{aligned} \right], \quad (110)$$

$\hat{D}_{MM}^{[tt]}$ can either be $D_{KK}^{[tt]}$ or $\bar{D}_{CC}^{[tt]}$. In the former case, $(\hat{D}_{MM}^{[tt]}, \hat{\theta}, \hat{T}) = (D_{KK}^{[tt]}, \alpha, \underline{T})$ where the function $\underline{T}^{(\kappa)}(z, z')$ is defined in Eq. (98). In the latter case, $(\hat{D}_{MM}^{[tt]}, \hat{\theta}, \hat{T}) = (\bar{D}_{CC}^{[tt]}, \bar{\alpha}, \bar{T})$ where the function $\bar{T}^{(\kappa)}(z, z')$ is defined in Eq. (100). For abbreviation, these two choices will together be represented as $(\hat{D}_{MM}^{[tt]}, \hat{\alpha}, \hat{T}) = \{(D_{KK}^{[tt]}, \alpha, \underline{T}), (\bar{D}_{CC}^{[tt]}, \bar{\alpha}, \bar{T})\}$.

(3) Quantities required by $L^{[\text{PT2}, \bar{t}\bar{t}]}$:

$$\begin{aligned} \hat{D}_{s's}^{[\bar{t}\bar{t}]} &= \sum_{\kappa} \left(\hat{\theta}_{s's \leftarrow r'r}^{(1,1,\kappa)} + 2\varepsilon_{\text{shift}} \hat{\theta}_{s's \leftarrow r'r}^{(2,1,\kappa)} \right) \left[\hat{T}_{r'r}^{(\kappa)}(\hat{b}, \hat{b}) + \hat{T}_{r'r}^{(\kappa)}(y, y) + \hat{T}_{r'r}^{(\kappa)}(\delta_{(x)}, \delta_{(x)}) \right] \\ &+ \sum_{\kappa} \left(2\hat{\theta}_{s's \leftarrow r'r}^{(1,1,\kappa)} + 2\varepsilon_{\text{shift}} \hat{\theta}_{s's \leftarrow r'r}^{(2,1,\kappa)} \right) \left[-\hat{T}_{r'r}^{(\kappa)}(y, \hat{b}) + \hat{T}_{r'r}^{(\kappa)}(\delta_{(x)}, \hat{b}) - \hat{T}_{r'r}^{(\kappa)}(\delta_{(x)}, y) \right] \\ &+ \sum_{\kappa} \hat{\theta}_{s's \leftarrow r'r}^{(1,1,\kappa)} \left[\begin{aligned} &-\hat{T}_{r'r}^{(\kappa)}(\lambda, \hat{b}) + \hat{T}_{r'r}^{(\kappa)}(y, \lambda) - \hat{T}_{r'r}^{(\kappa)}(\delta_{(x)}, \lambda) \\ &+ \hat{T}_{r'r}^{(\kappa)}(\delta_{(\lambda)}, \hat{b}) - \hat{T}_{r'r}^{(\kappa)}(\delta_{(\lambda)}, \lambda) + \hat{T}_{r'r}^{(\kappa)}(\delta_{(\lambda)}, \delta_{(x)}) \end{aligned} \right], \quad (111) \end{aligned}$$

The symbol $\hat{D}_{s's}^{[\bar{t}\bar{t}]}$ can be either $D_{h'h}^{[\bar{t}\bar{t}]}$, $D_{h'h}^{[\bar{t}\bar{t}]}$ or $D_{h'h}^{[\bar{t}\bar{t}]}$. Other symbols will be chosen correspondingly from $(\hat{D}_{s's}^{[\bar{t}\bar{t}]}, \hat{\theta}, \hat{T}) = \{(D_{h'h}^{[\bar{t}\bar{t}]}, \alpha, \underline{T}), (\bar{D}_{p'p}^{[\bar{t}\bar{t}]}, \bar{\alpha}, \bar{T}), (W_{h'h}^{[\bar{t}\bar{t}]}, \beta, \underline{T}), (\bar{W}_{p'p}^{[\bar{t}\bar{t}]}, \bar{\beta}, \bar{T})\}$.

Another quantity required is

$$\begin{aligned} \underline{\Gamma}_{D, Ng}^{[\bar{t}\bar{t}]} &= \sum_{\kappa} \left(2 + 2\varepsilon_{\text{shift}} \sigma^{(2,\kappa)} \right) \left[\underline{\Gamma}^{(\kappa)}(\underline{\Gamma} y, b) - \underline{\Gamma}^{(\kappa)}(\underline{\Gamma} y, \underline{\Gamma} y) + \underline{\Gamma}^{(\kappa)}(\delta_{(x)}, \underline{\Gamma} y) \right] \\ &+ \sum_{\kappa} \left[\begin{aligned} &-\underline{\Gamma}^{(\kappa)}(\underline{\Gamma} \lambda, \underline{\Gamma} y) - \underline{\Gamma}^{(\kappa)}(\underline{\Gamma} \lambda, \underline{\Gamma} y) + \underline{\Gamma}^{(\kappa)}(\delta_{(\lambda)}, \underline{\Gamma} y) \\ &+ \underline{\Gamma}^{(\kappa)}(\underline{\Gamma} \lambda, b) + \underline{\Gamma}^{(\kappa)}(\delta_{(x)}, \underline{\Gamma} \lambda) \end{aligned} \right] \quad (112) \end{aligned}$$

Note that in Eq. (112), when both tensors are from the trailing space (e.g. $\underline{\Gamma}^{(\kappa)}(\underline{\Gamma} y, \underline{\Gamma} y)$), the underline in the parenthesis indicates which Γ is being differentiated.

2.6. Tensor operations

In Section 2.5.4, we have shown that all the quantities that depend on Ξ can be computed

from operations of the type $\frac{\partial E^{(\kappa)}(z, z')}{\partial Q}$, where Q can be $Q = \{\mathbb{I}_{\mu\nu\lambda\gamma}, \mathbf{C}_{h\mu}, \bar{\mathbf{C}}_{p\mu}, T_{h'h}^{(\kappa)}, \bar{T}_{p'p}^{(\kappa)}\}$.

The results in Section 2.54 are organized according to the output quantity. In the list below, we reorganize the results based on the type of tensors (z, z') involved in the calculation:

(1) MP2-ERIs contracted with MP2-ERIs: the input is $\hat{b}$, the outputs include: $T_{h'h}^{(\kappa)}(\hat{b}, \hat{b})$,

$\bar{T}_{p'p}^{(\kappa)}(\hat{b}, \hat{b})$, $\mathbf{C}_{h\mu}^{(\kappa)}(\hat{b}, \hat{b})$, $\bar{\mathbf{C}}_{p\mu}^{(\kappa)}(\hat{b}, \hat{b})$ and $\mathbb{I}_{\mu\nu\lambda\gamma}^{(\kappa)}(\hat{b}, \hat{b})$;

(2) Trailing space tensor contracted with MP2-ERIs: the inputs are $(v, \hat{b})$ where $v = \{y, \lambda\}$

, the outputs include $T_{h'h}^{(\kappa)}(\Gamma v, \hat{b})$, $\bar{T}_{p'p}^{(\kappa)}(\Gamma v, \hat{b})$, $\Gamma_{D, Ng}^{(\kappa)}(\Gamma v, \hat{b})$, $\mathbf{C}_{h\mu}^{(\kappa)}(\Gamma v, \hat{b})$, $\bar{\mathbf{C}}_{p\mu}^{(\kappa)}(v, \hat{b})$ and

$\mathbb{I}_{\mu\nu\lambda\gamma}^{(\kappa)}(v, \hat{b})$;

(3) Supporting space tensor contracted with MP2-ERIs: the inputs are $(\zeta, \hat{b})$ where

$\zeta = \{\delta_{(x)}, \delta_{(\lambda)}\}$, the outputs include $T_{h'h}^{(\kappa)}(\zeta, \hat{b})$, $\bar{T}_{p'p}^{(\kappa)}(\zeta, \hat{b})$, $\mathbf{C}_{h\mu}^{(\kappa)}(\zeta, \hat{b})$, $\bar{\mathbf{C}}_{p\mu}^{(\kappa)}(\zeta, \hat{b})$ and

$\mathbb{I}_{\mu\nu\lambda\gamma}^{(\kappa)}(\zeta, \hat{b})$;

(4) Trailing space tensor contracted with trailing space tensor: the inputs are (u, v) which

can be (λ, y) or (y, y) , the outputs include $T_{h'h}^{(\kappa)}(\Gamma u, \Gamma v)$, $\bar{T}_{p'p}^{(\kappa)}(\Gamma u, \Gamma v)$, $\Gamma_{D, Ng}^{(\kappa)}(\Gamma u, \Gamma v)$ and

$\Gamma_{D, Ng}^{(\kappa)}(\Gamma u, \Gamma v)$;

(5) Supporting space tensor contracted with trailing space tensor: the inputs are (ζ, v)

which can be $(\delta_{(x)}, y)$, $(\delta_{(\lambda)}, y)$ or $(\delta_{(x)}, \lambda)$, the outputs include $T_{h'h}^{(\kappa)}(\zeta, \Gamma v)$, $\bar{T}_{p'p}^{(\kappa)}(\zeta, \Gamma v)$,

$\Gamma_{D, Ng}^{(\kappa)}(\zeta, \Gamma v)$.

(6) Supporting space tensor contracted with supporting space tensor: the inputs are (ζ, ξ)

which can be $(\delta_{(x)}, \delta_{(x)})$ or $(\delta_{(\lambda)}, \delta_{(x)})$, the output includes $T_{h'h}^{(\kappa)}(\zeta, \xi)$, $\bar{T}_{p'p}^{(\kappa)}(\zeta, \xi)$.

For all pairs of tensors appearing in the list above, $E^{(\kappa)}(z, z')$ is computed by linearly contracting a set of tensors together. We have previously shown⁵⁵ how such partial derivatives can be derived in a systematic and automated way. In particular, we have applied the idea to contractions among trailing or supporting space tensors, and we have demonstrated how the tensor operations for computing the quantities from (4) through (6) can be formulated as linear algebra that are at most cubic scaling. The same work also detailed the implementation of steps (1) through (3) using the THC form for the MP2-ERIs. More generally, the quantities from (1) through (3) can be computed from other types of MP2 gradients and Fock-gradients, which are not limited to THC, as discussed Section 2.6.1 through 2.6.3 respectively.

2.6.1 MP2 Gradient

The contraction between two MP2-ERI (i.e. $E = \hat{b}\Xi^{-1}\hat{b}$) can be computed with an MP2 energy as in Eq. (28). Interfacing with MP2 gradient⁵⁶⁻⁵⁷ codes requires extracting a few intermediates. We also require the Laplace quadrature factorization to be used during the MP2 calculations, such that the factorization developed in Section 2.5 can be applied. In order to compute the outputs in item (1) in the list above, we require the following intermediates from MP2-gradient codes

$$\left[\partial_{\mathbb{I}} E_{\text{MP2}}^{(\kappa)} \right]_{\mu\nu\lambda\gamma} = -\frac{1}{4} \sum_{\eta\pi\eta'\pi'} (\eta\pi \parallel \eta'\pi') \tau_{\eta}^{(\kappa)} \bar{\tau}_{\pi}^{(\kappa)} \tau_{\eta'}^{(\kappa)} \tau_{\pi'}^{(\kappa)} \frac{\partial(\eta\pi \parallel \eta'\pi')}{\partial \mathbb{I}_{\mu\nu\lambda\gamma}} \quad (113)$$

$$\left[P_{\text{MP2}}^{(\kappa)} \right]_{\mu\nu} = -\frac{1}{2} \sum_{\pi\eta\pi'} (\mu\pi \parallel \eta'\pi') \bar{\tau}_{\pi}^{(\kappa)} \tau_{\eta'}^{(\kappa)} \bar{\tau}_{\pi'}^{(\kappa)} (\nu\pi \parallel \eta'\pi') \quad (114)$$

and

$$\left[\bar{P}_{\text{MP2}}^{(\kappa)} \right]_{\mu\nu} = -\frac{1}{2} \sum_{\eta\pi\eta'} (\eta\pi \parallel \eta'\mu) \tau_{\eta}^{(\kappa)} \bar{\tau}_{\pi}^{(\kappa)} \tau_{\eta'}^{(\kappa)} (\eta\pi \parallel \eta'\nu) \quad (115)$$

Eq. (113) is the intermediate when computing the MP2 two-electron density matrix, while Eqs. (114)-(115) are intermediates formed during the computation of MP2 second order density matrices. The output $\mathbb{I}_{\mu\nu\lambda\gamma}^{(\kappa)}(\hat{b}, \hat{b})$ will be equal to Eq. (113), while the other outputs can be computed as

$$\underline{T}_{h'h}^{(\kappa)}(\hat{b}, \hat{b}) = \sum_{\mu\nu} \mathbf{C}_{h'\mu} \left[P_{\text{MP2}}^{(\kappa)} \right]_{\mu\nu} \mathbf{C}_{h\nu} \quad (116)$$

and

$$\underline{\mathbf{C}}_{hv}^{(\kappa)}(\hat{b}, \hat{b}) = \sum_{h'\mu} T_{h'h}^{(\kappa)} \cdot \mathbf{C}_{h'\mu} \left[P_{\text{MP2}}^{(\kappa)} \right]_{\mu\nu} \quad (117)$$

$\bar{T}_{p'p}^{(\kappa)}(\hat{b}, \hat{b})$ and $\bar{\mathbf{C}}_{p\mu}^{(\kappa)}(\hat{b}, \hat{b})$ can be computed from $\bar{P}_{\text{MP2}}^{(\kappa)}$ in similar ways by replacing $T_{h'h}^{(\kappa)}$ with $\bar{T}_{p'p}^{(\kappa)}$, and replacing $\mathbf{C}_{h\mu}$ with $\bar{\mathbf{C}}_{p\mu}$.

2.6.2 Contraction of trailing space tensor with ERI using Fock Builds

We first review how Fock builds can be used to effect the transformation from MP2-ERIs to trailing space tensor as in Eq. (23). We take ${}^T f_{ia,D}$ as an example, while the other subensors (e.g. ${}^T f_{ij,D}$, ${}^T f_{ab,D}$) can be computed in similar ways. By using the factorization of Ξ^{-1} from Eq. (16), we have

$$\begin{aligned} {}^T f_{ia,D} &= \sum_{\kappa, hph'p', jC} \Gamma_{jC,D}^{(0)} T_{ih}^{(\kappa)} \bar{T}_{ap}^{(\kappa)} T_{jh'}^{(\kappa)} \bar{T}_{Cp'}^{(\kappa)} \cdot (hp \parallel h'p')^{[\text{I}]} \\ &\quad + \sum_{\kappa, hph'p', Kb} \Gamma_{Kb,D}^{(1)} T_{ih}^{(\kappa)} \bar{T}_{ap}^{(\kappa)} T_{Kh'}^{(\kappa)} \bar{T}_{bp'}^{(\kappa)} \cdot (hp \parallel h'p')^{[\text{I}]} \\ &= \sum_{\kappa} \sum_{hp\mu\nu} \left(T_{ih}^{(\kappa)} \mathbf{C}_{h\mu} \right) \left(\bar{T}_{ap}^{(\kappa)} \bar{\mathbf{C}}_{p\nu} \right) \mathbb{G}_{\mu\nu} \left(d^{(\kappa,D)} \right) \end{aligned} \quad (118)$$

where

$$d_{\lambda\gamma}^{[\kappa,D]} = \sum_{h'p'} \left(\sum_{jC} \Gamma_{jC,D}^{(0)} T_{jh'}^{(\kappa)} \bar{T}_{Cp'}^{(\kappa)} + \sum_{Kb} \Gamma_{Kb,D}^{(1)} T_{Kh'}^{(\kappa)} \bar{T}_{bp'}^{(\kappa)} \right) \mathbf{C}_{h'\lambda} \bar{\mathbf{C}}_{p'\gamma} \quad (119)$$

Details on $\Gamma^{(0)}$ and $\Gamma^{(1)}$ are provided in Section 2.7 of the SI. The range of the index D is no greater than N_{active}^2 , thus the total number of Fock-builds with MP2-ERIs required in Eq. (118) is quadratic with respect to the active space and constant with respect to system size.

From Eq. (118), the contraction between the trailing space tensor and the MP2-ERIs can be represented as

$$E_{\Xi}^{(\kappa)}(\Gamma y, \hat{b}) = \sum_D \sum_{\mu\nu\lambda\gamma} \tilde{d}_{\mu\nu}^{[\kappa,D]}(\mu\nu \parallel \lambda\gamma)^{[\mathbb{I}]} \cdot d_{\lambda\gamma}^{[\kappa,D]} \quad (120)$$

with

$$\tilde{d}_{\mu\nu}^{[\kappa,D]} = \sum_{hp,ia} y_{ia,D} \left(T_{ih}^{(\kappa)} \mathbf{C}_{h\mu} \right) \left(\bar{T}_{ap}^{(\kappa)} \bar{\mathbf{C}}_{p\nu} \right) \quad (121)$$

The output $\mathbb{I}_{\mu\nu\lambda\gamma}^{(\kappa)}(\Gamma y, \hat{b})$ can then be constructed from $\tilde{d}_{\mu\nu}^{[\kappa,D]}$ and $d_{\lambda\gamma}^{[\kappa,D]}$ following Eq. (120). For all other outputs, we use the fact that

$$\frac{\partial E_{\Xi}^{(\kappa)}(\Gamma y, \hat{b})}{\partial Q} = \sum_D \sum_{\mu\nu} \left(\frac{\partial \tilde{d}_{\mu\nu}^{[\kappa,D]}}{\partial Q} \mathbb{G}_{\mu\nu}(d^{[\kappa,D]}) + \mathbb{G}_{\mu\nu}(\tilde{d}^{[\kappa,D]}) \frac{\partial d_{\mu\nu}^{[\kappa,Q]}}{\partial Q} \right) \quad (122)$$

The other outputs including $\underline{T}_{h'h}^{(\kappa)}(\Gamma v, \hat{b})$, $\bar{\underline{T}}_{p'p}^{(\kappa)}(\Gamma v, \hat{b})$, $\underline{\Gamma}_{D,Ng}^{(\kappa)}(\Gamma v, \hat{b})$, $\underline{\mathbf{C}}_{h\mu}^{(\kappa)}(\Gamma v, \hat{b})$, $\bar{\underline{\mathbf{C}}}_{p\mu}^{(\kappa)}(v, \hat{b})$ can then be computed using Eq. (122), where partial derivatives of $\tilde{d}_{\mu\nu}^{[\kappa,D]}$ and $d_{\mu\nu}^{[\kappa,Q]}$ can be easily defined from their definition (Eqs. (119) and (121)).

2.6.3 Contraction of supporting space tensor with ERI using Fock-Builds

The transformation from MP2-ERI to the supporting space tensor (Eq. (24)) can be computed with Fock-builds in a way very similar to Section 2.6.2. Here we take subvector ${}^S\zeta_{ia,KC}$ as an example, while other subvectors (e.g. ${}^S\zeta_{iC',jC}$, ${}^S\zeta_{K'a,Kb}$) can all be computed in a similar manner. By using the factorization of Ξ^{-1} from Eq. (16),

$$\begin{aligned} {}^S\zeta_{ia,KC} &= \sum_{\kappa, hph'p'} T_{ih}^{(\kappa)} \bar{T}_{ap}^{(\kappa)} T_{Kh'}^{(\kappa)} \bar{T}_{Cp'}^{(\kappa)} \cdot (hp \parallel h'p')^{[\mathbb{I}]} \\ &= \sum_{\kappa} \sum_{hp\mu\nu} \left(T_{ih}^{(\kappa)} \mathbf{C}_{h\mu} \right) \left(\bar{T}_{ap}^{(\kappa)} \bar{\mathbf{C}}_{p\nu} \right) \mathbb{G}_{\mu\nu}(d^{(\kappa,KC)}) \end{aligned} \quad (123)$$

with

$$d_{\lambda\gamma}^{[\kappa,KC]} = \left(\sum_{h'} T_{Kh'}^{(\kappa)} \mathbf{C}_{h'\lambda} \right) \left(\sum_{p'} \bar{T}_{Cp'}^{(\kappa)} \bar{\mathbf{C}}_{p'\gamma} \right) \quad (124)$$

Because the number of hole-type active orbitals (indexed with K) and particle-type active orbitals (indexed with C) are both no greater than the number of active orbitals, the total number of Fock-builds with MP2-ERIs required in Eq. (123) is at most N_{active}^2 , which is

quadratic with respect to the active space and constant with respect to system size. From Eq. (123), $E_{\Xi}^{(\kappa)}(\zeta, \hat{b})$ can then be represented as

$$E_{\Xi}^{(\kappa)}(\zeta, \hat{b}) = \sum_{KC} \sum_{\mu\nu\lambda\gamma} \tilde{d}_{\mu\nu}^{[\kappa, KC]} (\mu\nu \parallel \lambda\gamma)^{[\mathbb{I}]} \cdot d_{\lambda\gamma}^{[\kappa, KC]} \quad (125)$$

with

$$\tilde{d}_{\mu\nu}^{[\kappa, KC]} = \sum_{iahp} \zeta_{ia, KC} \left(T_{ih}^{(\kappa)} \mathbf{C}_{h\mu} \right) \left(\bar{T}_{ap}^{(\kappa)} \bar{\mathbf{C}}_{p\nu} \right) \quad (126)$$

Since Eq. (125) and Eq. (120) take a similar form, the corresponding outputs can be computed in the same way.

2.7 Summary of supporting subspace formulation of CASPT2 gradient

We now summarize the entire workflow of computing CASPT2 gradient using the supporting subspace method. The workflow is illustrated in Figure 1.

The entire workflow can be divided into four stages. The first stage computes the data required, i.e. $\hat{b}, y, \delta_{(x)}, \lambda$ and $\delta_{(\lambda)}$. The initialization sets up the matrix Ξ as well as various types of molecular orbital coefficients. The hole-type and particle-type orbital coefficients together with the MP2-ERI provide the definition of $\hat{b}$. Solving the CASPT2 linear system (Eq.(6)) generates y and $\delta_{(x)}$, which together with $\hat{b}$ define the wavefunction amplitudes t . Solution to the Λ -equation (Eq.(42)) generates λ and $\delta_{(\lambda)}$, which together with $\hat{b}, y$ and $\delta_{(x)}$ define the Lagrange multiplier $\hat{t}$.

Once the above five tensors are available, the next stage is the tensor operations corresponding to the list at the beginning of Section 2.6. Based on the types of tensors involved (i.e. supporting space tensor, trailing space tensor, or MP2-ERI), the corresponding operations will be called to compute the required quantities of partial derivatives in the list. Note that the tensor operations between different pairs of tensors are completely independent from each other. In addition, the computation of quantities corresponding to different Laplace quadrature is also independent from each other. Therefore, this stage is well suited for parallelization. In our current implementation, the

parallelization is only across different Laplace quadrature points. The parallelization over different tensors will be explored in future work.

The third stage is post processing, which accumulates the quantities obtained from the tensor operation stage to different response quantities. We further separate the post processing stage into two steps. In the first step, the outputs obtained from the tensor operation stage are applied with the proper transformations based on the corresponding Laplace quadrature index, and accumulated to the quantities as listed in Section 2.5.4 and also detailed in Table 1. These quantities contain all the information needed to compute contributions from the leading space. In the second step, outputs from step 1 are passed into the pseudocode listed in Table 2. The outputs from these pseudocodes will add up to the total density matrices of different types of integrals as well as partial derivatives with respect to the MO and CI coefficients, as required in Eq. (38).

The last stage is to evaluate the total CASPT2 analytical gradient. The density matrices for each type of integrals are passed into the corresponding codes to evaluate the analytical gradient contributions from the integrals. The partial derivatives with respect to the MO and CI coefficients are passed to the coupled-perturbed solver of the reference state. These contributions will together add up to the total CASPT2 ground state analytical gradients in Eq. (39).

3. Results and discussion

The method described in this work is implemented in the TERACHEM⁵⁸⁻⁶⁰ quantum chemistry package (currently supporting atom-centered spd Gaussian basis sets and spdf Gaussian auxiliary basis sets). The cc-pVDZ basis set⁶¹ is used unless otherwise specified. Tensor hyper-contraction is used to obtain the MP2-ERIs. The THC tensors and the corresponding analytical gradients are obtained with the GPU implementation^{62, 54} in TERACHEM using the cc-pVDZ-RI auxiliary basis sets⁶³ and THC grids optimized for cc-pVDZ basis sets.⁶⁴ No frozen core approximation is used, i.e. all molecular orbitals are included in the perturbation calculations. A level shift parameter of 0.3 Hartree is applied for all calculations unless otherwise specified. All calculations are performed on computing nodes with NVIDIA GeForce GTX 1080Ti GPUs and Intel Xeon E5-2637 CPUs.

We start by testing the accuracy and performance of the new method. Detailed comparisons of analytical and numerical gradients for ethylene using CASPT2(2,2) and butadiene using CASPT2(4,4) are provided in SI (Section 6). All numerical gradients are computed with 7-point central differences with displacement of 0.001 Bohr. The elements with largest absolute value in the difference vectors are $2.83\text{e-}7$ Hartree/Bohr and $3.41\text{e-}7$ Hartree/Bohr for ethylene and butadiene respectively. The root mean square of the elements in the difference vectors are $1.35\text{e-}7$ Hartree/Bohr and $1.72\text{e-}7$ Hartree/Bohr respectively. This agreement shows that the equations are both correctly derived and implemented.

To test the performance of our new method, we compute the ground state analytical gradients of butadiene or hexatriene solvated by an increasing number of methanol molecules. CASPT2(4,4) and CASPT2(6,6) are used for solvated butadiene and hexatriene respectively. These test systems have also been used for benchmarking in Paper I.⁴² In Figure 2a, we compare the wall time for single point gradient calculations with the corresponding energy calculations, where the timings were collected using a single CPU core plus a single GPU. The four lines in Figure 2a are almost parallel, indicating that the energy and gradient calculations exhibit the same sub-cubic scaling for both choices of active spaces. In terms of the prefactors, the total wall time of the gradient calculations is 2.5~3 times the wall time of the corresponding energy calculations. Our current implementation also supports parallelization over the Laplace quadrature points, and Figure S3 demonstrates the current parallel efficiency where almost 2x speedup is observed with 2 GPUs and 5.2x speedup is observed using 8 GPUs.

Because the MP2-gradient function has distinct scaling from the rest of the calculations, we further divide the wall time of CASPT2 gradient calculations into two parts, i.e. the “MP2-grad” part that uses THC-MP2 gradient codes with formal scaling of $O(N^4)$, and “the remainder” part corresponding to the rest of the calculations that formally scale as $O(N^3)$. The differences in the scalings of the two parts are clearly shown in Figure 2b. In addition, Figure 2b highlights several practical aspects that we also observed in energy calculations. On one hand, the THC-MP2 gradient part exhibits higher scaling with system size (around $O(N^{3.5})$) but its cost is independent of the active space size. This is because the prefactor of the THC-MP2 gradient depends linearly on the number of hole-

type (i.e. closed and active) orbitals, which is often far greater than the number of active orbitals. On the other hand, the remainder part exhibits a lower and close to quadratic scaling, but the prefactor is sensitive to the choice of the active space. This is because the prefactor of the remainder part (e.g. the number of Fock-builds required in Section 2.6.2 and 2.6.3, and the dimensions of the trailing or supporting space tensors) depends quadratically on the number of active orbitals. In Figure 2b, the remainder part consumes about 2.5x more time for (6,6) compared to (4,4) active spaces. This is consistent with the corresponding ratio of N_{active}^2 , i.e. $(6 \times 6)/(4 \times 4) = 2.25$. A detailed breakdown of the walltime for butadiene and hexatriene solvate by ten methanol molecules is given in the SI (section 7). Based on the workflow in Figure 1, we decompose the walltime into the MP2-gradient, solving the CASPT2 equation, solving the Λ -equation, tensor operations other than MP2-gradient, and the rest (i.e. post processing, CASSCF coupled perturbed and integral gradients). The absolute timing of each part is shown in Figure S1, while the corresponding pie-chart is shown in Figure S2. Even though the MP2-gradient is the component with the highest scaling, Figure S1 and S2 show that solving the CASPT2 and Λ -equations consumes a large portion of the walltime. The reduced system solver (Eq.(25)) involves transformations among tensors of the trailing and supporting space, and such transformations are called many times during conjugate gradient iterations. Optimization of these transformations will improve the performance of both energy and gradient calculations, and will be pursued in future work.

The development of analytical gradients enables *ab initio* molecular dynamics simulations with CASPT2. We tested the energy conservation of microcanonical AIMD simulations on one butene molecule for 10 picoseconds, and the results are shown in Figure 3. CASPT2(2,2)/cc-pVDZ with a level-shift of 0.5 a.u. is used for the simulations, and velocity Verlet with a 0.5 fs time step is used to integrate Hamilton's equations. The initial structure has the double bond twisted by 90 degrees. Initial velocities were drawn from a Maxwell-Boltzmann distribution (478K), and the instantaneous kinetic temperature fluctuates between 500K to over 2000K during the course of the simulations. As shown in Figure 3, the trajectory remains stable for the entire 10 picoseconds, and the fluctuations in total energy are very small compared to the scale of the fluctuations in the kinetic and potential energy. The drift of the total energy is 1.041e-2 kcal/mol per picosecond,

corresponding to an energy drift of $2.89\text{e-}4$ kcal/mol per picosecond per degree of freedom. These results demonstrate that our gradient implementation gives the correct nuclear gradients corresponding to the underlying CASPT2 potential energy surface.

The CASPT2 gradients can also be used for geometry optimizations. As a demonstration, we performed constrained optimizations for a doubly bonded diamondoid dimer⁶⁵ ($\text{C}_{28}\text{H}_{36}$). The trans (crystal structure) and cis (rotated by 180°) conformers are fully relaxed with unconstrained optimizations, and the energy of the cis conformer is higher than trans conformer by 2.97 kcal/mol. For the points between 15 degrees to 165 degrees, the dihedral angles highlighted in blue and yellow in the inset of Figure 4 are constrained to the corresponding values while all other degrees of freedom are relaxed. Constrained optimizations of this type are often used to study the energy profiles of reactions along a particular reaction coordinate of interest. The constrained geometry optimizations were performed using the geomeTRIC optimizer,⁶⁶ and the convergence criteria were set as specified in the figure caption. As the dihedral angle is twisted, the reference wavefunction changes from strongly single-reference to multireference as evidenced by the square of the two largest CI coefficients of the CASSCF reference wavefunctions (0.959, 0.040 at 0° vs. 0.505, 0.496 at 90°). At 90° the energy profile reaches a plateau rather than a cusp, which is also evidenced by a sharp drop in the root-mean-squared nuclear gradient mostly along the constraint degree of freedom. The observation of the energy plateau is consistent with the well-known energy profile of twisted ethylene, indicating the proper description of the multi-reference effects.

4. Conclusions

By using the supporting subspace method, we have developed a reduced scaling formulation of state-specific CASPT2 analytical gradients in terms of MP2 gradients and Fock gradients. With the new formulation, the formal scaling of the CASPT2 analytical gradients for a given active space is the same as that of MP2 gradients. By exploiting existing optimized MP2 and Fock routines, the new formulation enables the application of CASPT2 gradients to any molecule where MP2 gradients are feasible.

The active space configurations used in the calculations presented are all relatively small. The supporting subspace formulation of the CASPT2 analytical gradients imposes

no restrictions on the active space size, and applications to large active space are certainly possible. At the moment, there are two major practical obstacles to using large active spaces. First, CASPT2 analytical gradients require differentiating the three and four-particle density matrices in the active space, but our current implementation for computing these involves nested loops on the CPU and is very inefficient. One possible way to address this issue is to develop a more efficient GPU kernel for computing these quantities. Secondly, the prefactor for computing the remainder part (see discussion of Figure 2b) increases quadratically with respect to the active space size. Therefore, optimizing the tensor operations, especially those transformations between trailing or supporting subspace tensors that are called repeatedly in the preconditioned conjugate gradient, will become even more important for large active spaces.

The workflow illustrated in Figure 1 clearly shows that the new formulation of CASPT2 gradient has multiple levels of independent operations, especially during the first and second stages. For example, the “MP2-gradient” block is independent from all other blocks in the first two stages, thus in principle it can be computed at the same time as “Solve CASPT2” and “Solve Lambda”. This observation suggests that in addition to the parallelization of Laplace quadrature that is already used in our current implementation, there is still much room for improving the performance through massive parallelization. This possibility will be explored in our future work.

In previous work, we have shown that the supporting subspace formulation can also be applied to multi-state CASPT2 (MS-CASPT2), which enables descriptions of excited states and conical intersections. Therefore, the next step will be to generalize the formulation of CASPT2 gradient developed in this work to the analytical gradients and non-adiabatic couplings of MS-CASPT2. We will introduce these ideas in our upcoming paper, which will then enable many exciting applications including searches of minimum-energy conical intersections (MECIs) as well as non-adiabatic excited state molecular dynamics simulations.

Supplemental Material

See supplemental material for details about mathematical derivations and proofs, additional test results for accuracy and performance.

Acknowledgements

This work was supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing, Scientific Discovery through Advanced Computing (SciDAC) program. This work was also supported by the Chemical Sciences, Geosciences, and Biosciences Division in the Office of Basic Energy Sciences of the U.S. Department of Energy under contract number under Contract No. DE-AC02-05CH11231.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Figure 1. Workflow of the supporting subspace formulation for CASPT2 gradients. Boxes with black borders and no fill represent data. Arrows along with boxes filled with colors represent functions and operations, and point from the input data to the output data. The solvers to “Solve CASPT2” and “Solve Lambda” are described in Ref. 42. The operations for “Trail-Trail”, “Trail-Support” and “Support-Support” only require linear algebra and no integrals. “Code 1.&2.” are provided in Section 1 of the SI.

$L[PT2]$			$L[t]$			$L[\hat{t}]$			$L[tt]$		$L[\hat{t}\hat{t}]$				
Output			$\mathbb{I}_{\mu\nu\lambda\gamma}^{[t]}$	$\mathbb{C}_{h\mu}^{[t]}$	$\bar{\mathbb{C}}_{p\mu}^{[t]}$	$\mathbb{I}_{\mu\nu\lambda\gamma}^{[\hat{t}]}$	$\mathbb{C}_{h\mu}^{[\hat{t}]}$	$\bar{\mathbb{C}}_{p\mu}^{[\hat{t}]}$	$D_{K'K}^{[tt]}$	$\bar{D}_{C'C}^{[tt]}$	$\bar{D}_{h'h}^{[\hat{t}\hat{t}]}$	$\bar{D}_{p'p}^{[\hat{t}\hat{t}]}$	$\mathcal{W}_{h'h}^{[\hat{t}\hat{t}]}$	$\bar{\mathcal{W}}_{p'p}^{[\hat{t}\hat{t}]}$	$\Gamma_{D,Ng}^{[\hat{t}\hat{t}]}$
Input			$\mathbb{I}_{\mu\nu\lambda\gamma}^{(\kappa)}$	$\mathbb{C}_{h\mu}^{(\kappa)}$	$\bar{\mathbb{C}}_{p\mu}^{(\kappa)}$	$\mathbb{I}_{\mu\nu\lambda\gamma}^{(\kappa)}$	$\mathbb{C}_{h\mu}^{(\kappa)}$	$\bar{\mathbb{C}}_{p\mu}^{(\kappa)}$	$T_{h'h}^{(\kappa)}$	$\bar{T}_{p'p}^{(\kappa)}$	$T_{h'h}^{(\kappa)}$	$\bar{T}_{p'p}^{(\kappa)}$	$T_{h'h}^{(\kappa)}$	$\bar{T}_{p'p}^{(\kappa)}$	$\Gamma_{D,Ng}^{(\kappa)}$
Laplace Factor			σ			σ			α	$\bar{\alpha}$	α	$\bar{\alpha}$	β	$\bar{\beta}$	σ
Equation			Eq.(108)			Eq.(109)			Eq.(110)		Eq.(111)				Eq.(112)
	$\hat{b}$	$\hat{b}$	-1			$-1 - 2\varepsilon_s\sigma(2,\kappa)$			$\hat{\theta}(1,1,\kappa)$		$\hat{\theta}(1,1,\kappa) + 2\varepsilon_s\hat{\theta}(2,1,\kappa)$				
After Solving CASPT2	Γ_y	$\hat{b}$	+1			$1 + 2\varepsilon_s\sigma(2,\kappa)$			$-2\hat{\theta}(1,1,\kappa)$		$\frac{-2\hat{\theta}(1,1,\kappa)}{-2\varepsilon_s\hat{\theta}(2,1,\kappa) - 2\varepsilon_s\hat{\theta}(1,2,\kappa)}$				$2 + 2\varepsilon_s\sigma(2,\kappa)$
	$\delta_{(x)}$	$\hat{b}$	-1			$-1 - 2\varepsilon_s\sigma(2,\kappa)$			$2\hat{\theta}(1,1,\kappa)$		$\frac{2\hat{\theta}(1,1,\kappa)}{+2\varepsilon_s\hat{\theta}(2,1,\kappa) + 2\varepsilon_s\hat{\theta}(1,2,\kappa)}$				
	Γ_y	Γ_y							$\hat{\theta}(1,1,\kappa)$		$\hat{\theta}(1,1,\kappa) + 2\varepsilon_s\hat{\theta}(2,1,\kappa)$				$0/-2 - 2\varepsilon_s\sigma(2,\kappa)$
	$\delta_{(x)}$	Γ_y							$-2\hat{\theta}(1,1,\kappa)$		$\frac{-2\hat{\theta}(1,1,\kappa)}{-2\varepsilon_s\hat{\theta}(2,1,\kappa) - 2\varepsilon_s\hat{\theta}(1,2,\kappa)}$				$2 + 2\varepsilon_s\sigma(2,\kappa)$
	$\delta_{(x)}$	$\delta_{(x)}$							$\hat{\theta}(1,1,\kappa)$		$\hat{\theta}(1,1,\kappa) + 2\varepsilon_s\hat{\theta}(2,1,\kappa)$				
After Solving Lambda	Γ_λ	$\hat{b}$				+1					$-\hat{\theta}(1,1,\kappa)$				+1
	$\delta_{(\lambda)}$	$\hat{b}$				-1					$\hat{\theta}(1,1,\kappa)$				
	Γ_λ	Γ_y									$\hat{\theta}(1,1,\kappa)$				-1/-1
	Γ_λ	$\delta_{(x)}$									$-\hat{\theta}(1,1,\kappa)$				+1
	$\delta_{(\lambda)}$	Γ_y									$-\hat{\theta}(1,1,\kappa)$				+1
	$\delta_{(\lambda)}$	$\delta_{(\lambda)}$									$\hat{\theta}(1,1,\kappa)$				

Table 1. Post processing of data coming from tensor operations between different pairs of tensors. The outputs are computed by transforming and accumulating the inputs based on the corresponding Laplace quadrature points. ε_s stands for ε_{shift} . The Laplace factor $\sigma^{(n,\kappa)}$ is a scalar as defined in Eq. (86). Depending on the type of output, $\hat{\theta}$ represents one of the following transformations: $\alpha_{h'h \leftarrow h_1 h_2}^{(n,m,\kappa)}$, $\bar{\alpha}_{p'p \leftarrow p_1 p_2}^{(n,m,\kappa)}$ (both discussed in Section 2.5.2), $\beta_{h'h \leftarrow h_1 h_2}^{(n,m,\kappa)}$ or $\bar{\beta}_{p'p \leftarrow p_1 p_2}^{(n,m,\kappa)}$ (both discussed in Section 2.5.3). In the last column, if both tensors are trailing

space tensors (e.g. $\Gamma\lambda$ or Γy), the notation “ c_1 / c_2 ” is used where c_1 and c_2 are the Laplace factors applied to the derivatives with respect to the first and second instance of Γ , respectively.

Term	Input	Post-process	Output				
			$\mathbb{I}_{\mu\nu\lambda\gamma}[\text{PT2}]$	$\underline{h}_{\mu\nu}[\text{PT2}]$	$\mathbf{I}_{\mu\nu\lambda\gamma}[\text{PT2}]$	$\underline{C}_{r\mu}[\text{PT2}]$	$\underline{c}_{0\Phi}[\text{PT2}]$
$L[t]$	L: $\begin{cases} \mathbb{I}_{\mu\nu\lambda\gamma}[t] \\ \underline{C}_{h\mu}[t], \bar{\underline{C}}_{p\mu}[t] \end{cases}$ T: y	Code.3	$\mathbb{I}_{\mu\nu\lambda\gamma}[t]$	$\underline{h}_{\mu\nu}[t]$	$\mathbf{I}_{\mu\nu\lambda\gamma}[t]$	$\underline{C}_{r\mu}[t]$ Eq.(51)-(53)	$\underline{c}_{0\Phi}[t]$ Eq.(54)
$L[tt]$	L: $D_{KK}[tt], \bar{D}_{CC}[tt]$ T: y	Code.4					$\underline{c}_{0\Phi}[tt]$ Eq.(57)
$L[\hat{t}]$	L: $\begin{cases} \mathbb{I}_{\mu\nu\lambda\gamma}[\hat{t}] \\ \underline{C}_{h\mu}[\hat{t}], \bar{\underline{C}}_{p\mu}[\hat{t}] \end{cases}$ T: λ	Code.3	$\mathbb{I}_{\mu\nu\lambda\gamma}[\hat{t}]$	$\underline{h}_{\mu\nu}[\hat{t}]$	$\mathbf{I}_{\mu\nu\lambda\gamma}[\hat{t}]$	$\underline{C}_{r\mu}[\hat{t}]$ Eq.(51)-(53)	$\underline{c}_{0\Phi}[\hat{t}]$ Eq.(54)
$L[\hat{t}\hat{t}]$	LL: $\begin{cases} D_{KK}[\hat{t}\hat{t}], \bar{D}_{CC}[\hat{t}\hat{t}] \\ W_{KK}[\hat{t}\hat{t}], \bar{W}_{CC}[\hat{t}\hat{t}] \end{cases}$ LT: $\Gamma_{D,Ng}[\hat{t}\hat{t}]$ TT: λ, y	Code.5		$\underline{h}_{\mu\nu}[\hat{t}\hat{t}]$ Eq.(64)	$\mathbf{I}_{\mu\nu\lambda\gamma}[\hat{t}\hat{t}]$ Eq.(64)	$\underline{C}[\hat{t}\hat{t}]$ Eq.(64)	$\underline{c}_{0\Phi}[\hat{t}\hat{t}]$ Eq.(79)

Table 2. Post processing of data coming from Table 1 to form the partial derivatives of $L^{[\text{PT2}]}$. Codes 3,4,5 are provided in Sections 3,4,5 of the SI, respectively. Because $\underline{h}_{\mu\nu}$ and $\mathbf{I}_{\mu\nu\lambda\gamma}$ only depend on the trailing space and are not discussed in the main text, they do not have equation numbers. Each output (e.g. $\underline{c}_{0\Phi}[\text{PT2}]$) is computed by summing over the quantities in the corresponding column.

Figure 2. Performance and scaling of the supporting subspace formulation of CASPT2 gradients using THC for the MP2-ERIs, showing (a) comparison with the corresponding energy-only calculations from Ref. 42 and (b) breakdown of timings for the gradient calculations. “MP2-gradient” corresponds to computing the list of quantities in Section 2.6.1 using THC-MP2 gradient routines. “Remainder” corresponds to the rest of the gradient calculations. All calculations used the cc-pVDZ basis set. The upper inset geometry illustrates hexatriene solvated with an increasing number of methanol molecules, which are the test systems for the (6,6) active space. The lower inset geometry illustrates butadiene solvated with increasing number of methanol molecules, which are the test systems for the (4,4) active space. All timings were performed with TERCHEM using one Intel Xeon E5-2637 CPU core and one GeForce GTX 1080Ti GPU. All coordinates are provided in the supporting information.

Figure 3. Energy as a function of time in a microcanonical *ab initio* molecular dynamics simulation of one butene molecule (C_4H_8) using CASPT2(2,2) with the supporting subspace formulation and THC. The regularization threshold for the THC metric matrix is 1.0×10^{-4} a.u. The velocity Verlet integrator with a 0.5 fs time step is used. The linear regression fits the total energy (in kcal/mol) as a function of time (in picoseconds), which shows a total energy drift of 1.041×10^{-2} kcal/mol per picosecond, corresponding to a drift of 2.89×10^{-4} kcal/mol per picosecond per degree of freedom. The three snapshots from left to right are taken at 0 ps, 5 ps, and 10 ps respectively. Hydrogen, and carbon atoms are colored as white and grey respectively.

Figure 4. CASPT2 potential energy profile from constrained optimizations for a double bonded diamondoid dimer ($C_{28}H_{36}$, shown in inset). The orange line shows the relative energy (left Y axis), and the blue line shows the root mean square (RMS) of nuclear gradients (right Y axis). Initial structure is taken from Cambridge structural database, CSD entry GISSEI01. The trans (0° , crystal structure) and cis (180°) conformers are fully relaxed with unconstrained optimizations. For other points, the dihedral angles highlighted in blue and yellow are constrained from $15\text{--}75^\circ$ in 15 degree increments, followed by $80\text{--}100^\circ$ in 5 degree increments, then $105\text{--}165^\circ$ in 15 degree increments. Structures were optimized using the geomeTRIC software package. The optimization is deemed converged when the following five criteria are satisfied simultaneously: 1) RMSD from the previous step is less than $1.2 \times 10^{-3} \text{ \AA}$; 2) maximum atomic displacement from the previous step is less than $1.8 \times 10^{-3} \text{ \AA}$; 3) RMS of the nuclear gradient with constraint direction projected out is less than 3.0×10^{-4} hartree/bohr; 4) maximum projected nuclear gradient for any atom is less than 4.5×10^{-4} hartree/bohr; 5) energy change from the previous step is less than 1.0×10^{-6} hartree. Hydrogen and carbon atoms are colored as white and grey respectively.

References

1. Payne, M. C.; Teter, M. P.; Allan, D. C.; Arias, T. A.; Joannopoulos, J. D., Iterative minimization techniques for ab initio total-energy calculations: molecular dynamics and conjugate gradients. *Rev. Mod. Phys.* **1992**, *64*, 1045-1097.
2. Tuckerman, M. E., Ab initio molecular dynamics: basic concepts, current trends and novel applications. *J. Phys.: Condens. Matter* **2002**, *14*, R1297-R1355.
3. Baker, J.; Pulay, P., Geometry optimization of atomic microclusters using inverse - power distance coordinates. *J. Chem. Phys.* **1996**, *105*, 11100-11107.
4. Billeter, S. R.; Turner, A. J.; Thiel, W., Linear scaling geometry optimisation and transition state search in hybrid delocalised internal coordinates. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2177-2186.
5. Wang, L.-P.; Titov, A.; McGibbon, R.; Liu, F.; Pande, V. S.; Martínez, T. J., Discovering chemistry with an ab initio nanoreactor. *Nature Chem.* **2014**, *6*, 1044-1048.
6. Henkelman, G.; Uberuaga, B. P.; Jónsson, H., A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J. Chem. Phys.* **2000**, *113*, 9901-9904.
7. Chattopadhyay, S.; Chaudhuri, R. K.; Mahapatra, U. S.; Ghosh, A.; Ray, S. S., State-specific multireference perturbation theory: development and present status. *WIREs: Comp. Mol. Sci.* **2016**, *6*, 266-291.
8. Roos, B. O.; Linse, P.; Siegbahn, P. E. M.; Blomberg, M. R. A., A simple method for the evaluation of the 2nd-order perturbation energy from external double-excitations with a CASSCF reference wavefunction. *Chem. Phys.* **1982**, *66*, 197-207.
9. Andersson, K.; Malmqvist, P. A.; Roos, B. O.; Sadlej, A. J.; Wolinski, K., 2nd-order perturbation theory with a CASSCF reference function. *J. Phys. Chem.* **1990**, *94*, 5483-5488.
10. Andersson, K.; Malmqvist, P. A.; Roos, B. O., 2nd-order perturbation theory with a complete active space self-consistent field reference function. *J. Chem. Phys.* **1992**, *96*, 1218-1226.
11. Pulay, P., A Perspective on the CASPT2 Method. *Int. J. Quantum Chem.* **2011**, *111*, 3273-3279.
12. Roos, B. O.; Taylor, P. R.; Siegbahn, P. E. M., A Complete Active Space SCF Method (CASSCF) using a Density Matrix Formulated Super-CI approach. *Chem. Phys.* **1980**, *48*, 157-173.
13. Roos, B. O., The Complete Active Space Self-Consistent Field Method and its Applications in Electronic Structure Calculations. *Adv. Chem. Phys.* **2007**, *69*, 399-445.
14. Ning, H.; Gong, C.; Li, Z.; Li, X., Pressure-Dependent Kinetics of Initial Reactions in Iso-octane Pyrolysis. *J. Phys. Chem. A* **2015**, *119*, 4093-4107.
15. Nakagaki, M.; Sakaki, S., CASPT2 study of inverse sandwich-type dinuclear 3d transition metal complexes of ethylene and dinitrogen molecules: similarities and differences in geometry, electronic structure, and spin multiplicity. *Phys. Chem. Chem. Phys.* **2015**, *17*, 16294-16305.
16. McLean, A. D.; Liu, B., Classification of configurations and determination of interacting and noninteracting spaces in configuration interaction. *J. Chem. Phys.* **1973**, *58*, 1066-1078.
17. Murphy, R. B.; Messmer, R. P., Generalized Moller-Plesset perturbation theory applied to general MCSCF reference wave-functions. *Chem. Phys. Lett.* **1991**, *183*, 443-448.

18. Hirao, K., Multireference Moller-Plesset method. *Chem. Phys. Lett.* **1992**, *190*, 374-380.
19. Kozłowski, P. M.; Davidson, E. R., Construction of open-shell perturbation-theory invariant with respect to orbital degeneracy. *Chem. Phys. Lett.* **1994**, *226*, 440-446.
20. Knowles, P. J.; Werner, H. J., An efficient method for the evaluation of coupling coefficients in configuration interaction calculations. *Chem. Phys. Lett.* **1988**, *145*, 514-522.
21. Werner, H. J.; Knowles, P. J., An efficient internally-contracted multiconfiguration reference configuration interaction method. *J. Chem. Phys.* **1988**, *89*, 5803-5814.
22. Celani, P.; Werner, H. J., Multireference perturbation theory for large restricted and selected active space reference wave functions. *J. Chem. Phys.* **2000**, *112*, 5546-5557.
23. Siegbahn, P. E. M., Direct configuration interaction with a reference state composed of many reference configurations. *Int. J. Quantum Chem.* **1980**, *18*, 1229-1242.
24. Nakano, H.; Hirao, K.; Gordon, M. S., Analytic energy gradients for multiconfigurational self-consistent field second-order quasidegenerate perturbation theory (MC-QDPT). *J. Chem. Phys.* **1998**, *108*, 5660-5669.
25. Dudley, T. J.; Khait, Y. G.; Hoffmann, M. R., Molecular gradients for the second-order generalized Van Vleck variant of multireference perturbation theory. *J. Chem. Phys.* **2003**, *119*, 651-660.
26. Theis, D.; Khait, Y. G.; Hoffmann, M. R., GVVPT2 energy gradient using a Lagrangian formulation. *J. Chem. Phys.* **2011**, *135*, 044117.
27. Celani, P.; Werner, H.-J., Analytical energy gradients for internally contracted second-order multireference perturbation theory. *J. Chem. Phys.* **2003**, *119*, 5044-5057.
28. MacLeod, M. K.; Shiozaki, T., Communication: Automatic code generation enables nuclear gradient computations for fully internally contracted multireference theory. *J. Chem. Phys.* **2015**, *142*, 051103.
29. Whitten, J. L., Coulombic potential energy integrals and their approximations. *J. Chem. Phys.* **1973**, *58*, 4496-4501.
30. Györfy, W.; Shiozaki, T.; Knizia, G.; Werner, H.-J., Analytical energy gradients for second-order multireference perturbation theory using density fitting. *J. Chem. Phys.* **2013**, *138*, 104104.
31. Park, J. W., Analytical Gradient Theory for Strongly Contracted (SC) and Partially Contracted (PC) N-Electron Valence State Perturbation Theory (NEVPT2). *J. Chem. Theory Comput.* **2019**, *15*, 5417-5425.
32. Park, J. W., Analytical Gradient Theory for Quasidegenerate N-Electron Valence State Perturbation Theory (QD-NEVPT2). *J. Chem. Theory Comput.* **2020**, *16*, 326-339.
33. Wang, S.; Li, C.; Evangelista, F. A., Analytic gradients for the single-reference driven similarity renormalization group second-order perturbation theory. *J. Chem. Phys.* **2019**, *151*, 044118.
34. Roos, B. O.; Andersson, K., Multiconfigurational perturbation theory with level shift: The Cr₂ potential revisited. *Chem. Phys. Lett.* **1995**, *245*, 215-223.
35. Ghigo, G.; Roos, B. O.; Malmqvist, P. A., A modified definition of the zeroth-order Hamiltonian in multiconfigurational perturbation theory (CASPT2). *Chem. Phys. Lett.* **2004**, *396*, 142-149.
36. Finley, J.; Malmqvist, P.-Å.; Roos, B. O.; Serrano-Andrés, L., The multi-state CASPT2 method. *Chem. Phys. Lett.* **1998**, *288*, 299-306.

37. Granovsky, A. A., Extended multi-configuration quasi-degenerate perturbation theory: The new approach to multi-state multi-reference perturbation theory. *J. Chem. Phys.* **2011**, *134*, 214113.
38. Shiozaki, T.; Gyorffy, W.; Celani, P.; Werner, H. J., Communication: Extended multi-state complete active space second-order perturbation theory: Energy and nuclear gradients. *J. Chem. Phys.* **2011**, *135*.
39. Szalay, P. G., Analytic energy derivatives for coupled-cluster methods describing excited states: General formulas and comparison of computational costs. *Int. J. Quantum Chem.* **1995**, *55*, 151-163.
40. Vlaisavljevich, B.; Shiozaki, T., Nuclear Energy Gradients for Internally Contracted Complete Active Space Second-Order Perturbation Theory: Multistate Extensions. *J. Chem. Theory Comput.* **2016**, *12*, 3781-3787.
41. Park, J. W.; Shiozaki, T., Analytical Derivative Coupling for Multistate CASPT2 Theory. *J. Chem. Theory Comput.* **2017**, *13*, 2561-2570.
42. Song, C.; Martínez, T. J., Reduced scaling CASPT2 using supporting subspaces and tensor hyper-contraction. *J. Chem. Phys.* **2018**, *149*, 044108.
43. Feyereisen, M.; Fitzgerald, G.; Komornicki, A., Use of approximate integrals in ab initio theory: An application in MP2 energy calculations. *Chem. Phys. Lett.* **1993**, *208*, 359-363.
44. Hohenstein, E. G.; Parrish, R. M.; Martínez, T. J., Tensor hypercontraction density fitting. I. Quartic scaling second- and third-order Møller-Plesset perturbation theory. *J. Chem. Phys.* **2012**, *137*, 044103.
45. Parrish, R. M.; Hohenstein, E. G.; Martínez, T. J.; Sherrill, C. D., Tensor hypercontraction. II. Least-squares renormalization. *J. Chem. Phys.* **2012**, *137*, 224106.
46. Hohenstein, E. G.; Luehr, N.; Ufimtsev, I. S.; Martínez, T. J., An atomic orbital-based formulation of the complete active space self-consistent field method on graphical processing units. *J. Chem. Phys.* **2015**, *142*, 224103.
47. Lowdin, P. O., Partitioning Technique, Perturbation Theory, and Rational Approximations. *Int. J. Quantum Chem.* **1982**, *21*, 69-92.
48. Almlöf, J., Elimination of energy denominators in Møller-Plesset perturbation theory by a Laplace transform approach. *Chem. Phys. Lett.* **1991**, *181*, 319-320.
49. Haser, M.; Almlöf, J., Laplace transform techniques in Møller-Plesset perturbation theory. *J. Chem. Phys.* **1992**, *96*, 489-494.
50. Shiozaki, T.; Györffy, W.; Celani, P.; Werner, H.-J., Communication: Extended multi-state complete active space second-order perturbation theory: Energy and nuclear gradients. *J. Chem. Phys.* **2011**, *135*, 081106.
51. Hohenstein, E. G.; Bouduban, M. E. F.; Song, C.; Luehr, N.; Ufimtsev, I. S.; Martínez, T. J., Analytic first derivatives of floating occupation molecular orbital-complete active space configuration interaction on graphical processing units. *J. Chem. Phys.* **2015**, *143*, 014111.
52. Snyder, J. W.; Hohenstein, E. G.; Luehr, N.; Martínez, T. J., An atomic orbital-based formulation of analytical gradients and nonadiabatic coupling vector elements for the state-averaged complete active space self-consistent field method on graphical processing units. *J. Chem. Phys.* **2015**, *143*, 154107.

53. Snyder, J. W.; Fales, B. S.; Hohenstein, E. G.; Levine, B. G.; Martínez, T. J., A direct-compatible formulation of the coupled perturbed complete active space self-consistent field equations on graphical processing units. *J. Chem. Phys.* **2017**, *146*, 174113.
54. Song, C.; Martínez, T. J., Analytical gradients for tensor hyper-contracted MP2 and SOS-MP2 on graphical processing units. *J. Chem. Phys.* **2017**, *147*, 161723.
55. Song, C.; Martínez, T. J.; Neaton, J. B., An Automatic Differentiation and Diagrammatic Notation Approach for Developing Analytical Gradients of Tensor Hyper-Contracted Electronic Structure Methods. doi.org/10.26434/chemrxiv.13002965.v1 **2020**.
56. Frisch, M. J.; Head-Gordon, M.; Pople, J. A., A direct MP2 gradient method. *Chem. Phys. Lett.* **1990**, *166*, 275-280.
57. Frisch, M. J.; Head-Gordon, M.; Pople, J. A., Semi-direct algorithms for the MP2 energy and gradient. *Chem. Phys. Lett.* **1990**, *166*, 281-289.
58. Ufimtsev, I. S.; Martinez, T. J., Quantum chemistry on graphical processing units. 1. Strategies for two-electron integral evaluation. *J. Chem. Theory Comput.* **2008**, *4*, 222-231.
59. Ufimtsev, I. S.; Martinez, T. J., Quantum Chemistry on Graphical Processing Units. 3. Analytical Energy Gradients, Geometry Optimization, and First Principles Molecular Dynamics. *J. Chem. Theory Comput.* **2009**, *5*, 2619-2628.
60. Ufimtsev, I. S.; Martinez, T. J., Quantum Chemistry on Graphical Processing Units. 2. Direct Self-Consistent-Field Implementation. *J. Chem. Theory Comput.* **2009**, *5*, 1004-1015.
61. Dunning, T. H., Gaussian basis sets for use in correlated molecular calculations. 1. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007-1023.
62. Song, C.; Martínez, T. J., Atomic orbital-based SOS-MP2 with tensor hypercontraction. I. GPU-based tensor construction and exploiting sparsity. *J. Chem. Phys.* **2016**, *144*, 174111.
63. Weigend, F., A fully direct RI-HF algorithm: Implementation, optimised auxiliary basis sets, demonstration of accuracy and efficiency. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4285-4291.
64. Schumacher, S.; Hohenstein, E. G.; Parrish, R. M.; Wang, L. P.; Martinez, T. J., Tensor Hypercontraction Second-Order Moller-Plesset Perturbation Theory: Grid Optimization and Reaction Energies. *J. Chem. Theory Comput.* **2015**, *11*, 3042-3052.
65. Zhuk, T. S.; Koso, T.; Pashenko, A. E.; Hoc, N. T.; Rodionov, V. N.; Serafin, M.; Schreiner, P. R.; Fokin, A. A., Toward an Understanding of Diamond sp²-Defects with Unsaturated Diamondoid Oligomer Models. *J. Am. Chem. Soc.* **2015**, *137*, 6577-6586.
66. Wang, L.-P.; Song, C., Geometry optimization made simple with translation and rotation coordinates. *J. Chem. Phys.* **2016**, *144*, 214108.