UC Berkeley
UC Berkeley Previously Published Works

Authors
Tran, Lan Nguyen
Shea, Jacqueline AR
Neuscamman, Eric

Tracking excited states in wave function optimization using density matrices and variational principles

Lan Nguyen Tran,[1, 2, *] Jacqueline A. R. Shea,[1] and Eric Neuscamman[1, 3, †]

[1]*Department of Chemistry, University of California, Berkeley, California, 94720, USA*
[2]*Ho Chi Minh City Institute of Physics, VAST, Ho Chi Minh City, 700000, Vietnam*
[3]*Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA*
(Dated: April 12, 2019)

We present a method for finding individual excited states' energy stationary points in complete active space self-consistent field theory that is compatible with standard optimization methods and highly effective at overcoming difficulties due to root flipping and near-degeneracies. Inspired by both the maximum overlap method and recent progress in excited state variational principles, our approach combines these ideas in order to track individual excited states throughout the orbital optimization process. In a series of tests involving root flipping, near-degeneracies, charge transfers, and double excitations, we show that this approach is more effective for state-specific optimization than either the naive selection of roots based on energy ordering or a more direct generalization of the maximum overlap method. Furthermore, we provide evidence that this state-specific approach improves the performance of complete active space perturbation theory. With a simple implementation, a low cost, and compatibility with large active space methods, the approach is designed to be useful in a wide range of excited state investigations.

I. INTRODUCTION

While linear response theory has indisputably been the most widely successful paradigm for making predictions about electronic excitations in chemistry, there remain important situations in which its use is either difficult or entirely inadvisable. The most affordable methods in this category,[1] such as time-dependent density functional theory (TD-DFT) and configuration interaction singles (CIS), are well known to face challenges when an excitation significantly alters a molecule's charge density, as occurs in charge transfer, Rydberg, and core excitations. Although TD-DFT has additional concerns,[2] one major issue in these cases comes from the fact that significant charge density changes induce orbital relaxation effects that are not captured in the linear response of a Slater determinant.[3] Even equation of motion coupled cluster theory with singles and doubles excitations (EOM-CCSD),[4] which is responding around a much more sophisticated wave function, has difficulty in capturing these relaxation effects in double excitations[4–6] and some core excitations.[7] Orbital relaxation effects aside, these linear response methods — as well as many other excited state methods including GW theory,[8] the Bethe-Salpeter equation,[8] and most variants of the algebraic diagrammatic construction[9] — rely on the assumption that the ground state is single-reference in character. When this assumption fails, as for example at conical intersections where at least two states mix strongly, one is typically forced to abandon the realm of weakly correlated methods altogether.

For decades now, complete active space self-consistent field theory (CASSCF)[10–13] and its generalizations[14–16] have been the go-to starting points for strongly correlated treatments of molecular ground and excited states alike. Just as Hartree-Fock (HF) theory[17] provides a reference state in which the molecular orbitals are relaxed in the presence of strong Pauli correlations, CASSCF relaxes the molecular orbitals in the presence of both Pauli correlations and any other strong correlations that exist within an active set of orbitals and electrons, typically chosen as a small set near the Fermi level. While expanding the size of the active space that can be treated remains a high priority and has been the focus of much recent work,[18–24] an additional important concern in excited state modeling is the issue of root flipping[25–29]. In the same way that HF and ΔSCF methods[30–33] relax orbitals by finding stationary points on the single-Slater-determinant energy surface, one would ideally like to prepare a CASSCF reference state by finding the CASSCF energy stationary point corresponding to the excited state in question. However, the common approach of separating the configuration interaction (CI) coefficient equations and the orbital rotation equations can create problems when two CI stationary points (typically called roots) cross each other in energy upon updating the orbitals in pursuit of the (initially) higher energy root's overall stationary point. Simple root selection (SRS) — in which one naively seeks to make the nth root's energy stationary during each orbital rotation step — will fail in this scenario, because the nth root and the root below it exchange ordering indefinitely during the "tick-tock" cycle of CI and orbital relaxation steps.

Although early methods in the coupled nonlinear optimization of CI coefficients and orbitals showed promise in bypassing this issue,[34] the most common remedy in use today is the state averaged (SA) approach[35], in which one simply minimizes the average energy of multiple states. The SA approach has the benefit of being compatible with modern, highly efficient tick-tock optimization schemes and is largely immune to root flipping by virtue of being insensitive to the ordering of the states within the average. However, it achieves these advan-

tages by abandoning the quest to locate individual excited states' stationary points. This lack of stationarity can be inconvenient when evaluating gradients,[36] but the more troubling issue is the potential loss of accuracy incurred by producing reference states whose orbitals are not fully relaxed. Whether this matters in practice of course depends on the system and whether or not post-CASSCF correlation methods can correct for this shortcoming alongside their treatment of out-of-active-space correlation effects. In cases where different states require drastically different orbital relaxations, such as in the presence of the large dipole shifts associated with charge transfer excitations[37], it would be particularly appealing to avoid state averaging while at the same time avoiding root flipping and retaining compatibility with tick-tock optimization.

Recently, fully state-specific orbital relaxations have been achieved for weakly correlated excited states[38] by exploiting simple approximations to excited state variational principles.[39–42] This progress came in the form of a nonlinear optimization in which the approximated variational principle was minimized under the constraint that the energy end up stationary with respect to both CI coefficients and orbital rotations. While one can imagine a straightforward application of these ideas to the CASSCF wave function, this would lead to a coupled nonlinear optimization of CI and orbital variables that would not necessarily be cost-competitive with modern tick-tock optimization schemes. In this study, we will instead explore whether affordable approximations to excited state variational principles can aid CASSCF in successfully tracking individual excited states in the presence of root flipping. The general idea is similar in spirit to the maximum overlap method (MOM)[43,44] that works to prevent variational collapse during ΔSCF optimizations, and indeed we find that state-tracking is most effective when we combine approximate variational principles with a MOM-inspired matching condition based on reduced density matrices. In particular, we demonstrate that this combination is more effective at finding CASSCF excited state stationary points than either the naive SRS approach or an approach based on approximate wave function overlaps as estimated via the CI vectors. In doing so, we also provide evidence that, although the improvement is sometimes modest, state specific CASSCF (SS-CASSCF) is a better reference for post-CASSCF methods than SA-CASSCF. Finally, as the CASSCF step is rarely the bottleneck when post-CASSCF methods are in use, our approach does not significantly increase overall computational cost, and so we are able to recommend its use in general. In particular, the approach requires only one- and two-body reduced density matrices, and so should be immediately compatible with the rapidly expanding collection of large active space methods that have come on to the scene in recent years.[18–23]

II. THEORY

A. Excited state variational principles

It has long been recognized[41,42] that the variational principle

$$W = \frac{\langle \Psi | (\omega - \hat{H})^2 | \Psi \rangle}{\langle \Psi | \Psi \rangle}, \tag{1}$$

has the Hamiltonian eigenstate with energy closest to ω as its global minimum. Recently, this and similar forms involving $\hat{H}^2$ have found tractable applications in spite of the challenges posed by the squared Hamiltonian operator, both in single-reference excited state approaches[38,45] and in quantum Monte Carlo,[37,46–52] where one can separate the two powers of $\hat{H}$ using a resolution of the identity $\sum_I |I\rangle \langle I|$ over which statistical sampling may be performed,

$$W = \frac{1}{\langle \Psi | \Psi \rangle} \sum_I \langle \Psi | (\omega - \hat{H}) | I \rangle \langle I | (\omega - \hat{H}) | \Psi \rangle. \tag{2}$$

In the context of CASSCF, one can simplify this resolution of the identity significantly by exploiting the fact that the space of wave functions that $\hat{H}$ can connect the CASSCF reference to is exactly spanned by the active space itself and all internally contracted singles and doubles excitations out of it. For ease of discussion, we will consider the case in which there are no closed shell orbitals, noting that the generalization to the full theory with a closed shell is straightforward. With active spin orbitals labeled by i and j and virtual spin orbitals labeled by a and b and the trivial assumption that the CASSCF wave function Φ is normalized, the resolution of the identity form of the variational principle can thus be organized as

$$W = W_0 + W_1 + W_2, \tag{3}$$

$$W_0 = (\omega - E)^2, \tag{4}$$

$$W_1 = \sum_{ia} \left| \langle \Phi | \hat{H} \, \hat{a}_a^+ \, \hat{a}_i | \Phi \rangle \right|^2, \tag{5}$$

$$W_2 = \sum_{ijab} \left| \langle \Phi | \hat{H} \, \hat{a}_a^+ \, \hat{a}_b^+ \, \hat{a}_j \, \hat{a}_i | \Phi \rangle \right|^2. \tag{6}$$

Of particular note is that, by virtue of being limited to internally contracted single and double excitations, a full evaluation of W, should one wish to pursue it, should be similar in complexity to constructing the right hand side of the first order wave function equation in complete active space second order perturbation theory (CASPT2).[53,54]

Inspecting the three components of W, we find that they have simple interpretations if we take Φ to be a root of the complete active space CI problem. First, W_0 is simply stating that the eigenstate we are after should have an energy close to ω. Second, noting that CI roots'

energies are already stationary with respect to the CI coefficients, we see that W_1 is simply a measure of how close the wave function is to being an overall energy stationary point as each of its terms is proportional to the energy derivative with respect to an active-to-virtual orbital rotation. Indeed, if Φ is a CI root and $W_1 = 0$, then the CASSCF energy is stationary. Finally, noting that W_1 and W_2 are unaffected if we make the replacement $\hat{H} \to \hat{H} - E$, we see that W_2 contains all the terms that would need to be zero in addition to those in W_1 in order for the energy variance $\sigma^2 = \langle (\hat{H} - E)^2 \rangle$ to be zero. As an exact eigenstate of $\hat{H}$ will be both an energy stationary point and a zero variance state, we see that, together, W_0, W_1, and W_2 are simply a least-squares way of saying that we want the CASSCF wave function that is closest to the exact energy eigenstate near ω.

In principle, we could use these expressions to follow the approach of excited state mean field (ESMF) theory[38] and minimize the Lagrangian

$$L = W - \vec{\mu} \cdot \frac{\partial E}{\partial \vec{\nu}}, \tag{7}$$

with respect to a variable set $\vec{\nu}$ that contained both orbital rotations and the CI coefficients. This approach uses the variational principle W, or an approximatin to it, to guide an optimization to the desired energy stationary point via constrained Lagrangian minimization. However, making such an approach cost-competitive with CASSCF tick-tock optimization methods would not be trivial, and in this study we seek to exploit approximations to W in a much simpler context.

Within the standard tick-tock approach of switching back and forth between Davidson CI diagonalizations and orbital optimization steps, it is typically the CI step that dominates the cost when the active space gets large. We will therefore leave the Davidson step unchanged for now and consider using approximations to W in the remainder of the optimization. In our initial testing, we have found that while an L-BFGS minimization of $|\nabla L|^2$ with respect to orbital rotation variables is effective when we set $W \approx W_0 + W_1$, it is in most cases equally effective to relax the orbitals by a simple L-BFGS minimization of $|\nabla E|^2$. This observation suggests that simple generalizations of standard Newton-Raphson style orbital optimizations in which the line search is set to minimize $|\nabla E|^2$ rather than E itself are likely adequate in many cases (and much faster than L-based quasi-Newton methods), even if they lack the strong resistance to ground state collapse offered by L. However, orbital relaxation for a particular CI root is only part of the process of converging to a desired stationary point during a CASSCF tick-tock optimization. The method used for selecting which CI root to relax the orbitals for is equally important and in general quite challenging. As we now discuss, it is in this area that we find excited state variational ideas to be most helpful.

B. Maximum overlap analogues

First introduced by Gill and coworkers,[43] the maximum overlap method (MOM) helps prevent variational collapse to the ground state when attempting to locate excited state solutions to the Roothan equations in Hartree Fock or density functional theory. The idea is to choose the excited state orbital occupation for the the orbitals generated by a newly diagonalized Fock matrix by selecting the orbitals that give the largest combined overlap with the molecular orbitals from the previous iteration of the method, or some target set of molecular orbitals believed to be similar to those of the desired excited state. In essence, MOM is an attempt to follow the trail of a particular excited state through the sequence of discreet (and sometimes large) orbital relaxations that occur over the course of the self consistent field iterations. This goal is very similar to what we desire when faced with a root flipping problem in CASSCF: we wish to track a particular excited state through the sequence of discreet (and sometimes large) changes to the Davidson CI roots that occur over the course of a tick-tock CASSCF optimization.

If one wished to pursue a strategy similar in spirit to that of the orbital-overlap-based MOM, a simple strategy would be to hope that changes to the CI vector $\vec{c}$ between iterations were never too large and simply define a tracking function Q_{MOM} based on the approximate wave function overlap

$$Q_{\mathrm{MOM}}(\vec{c}) = \vec{c}_t \cdot \vec{c}, \tag{8}$$

between each of the current iteration's CI roots and some target CI vector $\vec{c}_t$, taken here to be the CI vector selected in the previous iteration. Of course, like MOM itself, this strategy, and any tracking strategy based on measuring wave function similarities across iterations, will not necessarily be robust in cases where the iterative method makes large changes to the wave function in a single iteration. Ideally, we would therefore like to augment this strategy with a component that is less dependent on iteration-to-iteration similarity. The central finding of this study is that, in combination with a measure of similarity that is more robust than CI vector dot products, the excited state variational principle W can help in this way.

In hopes of creating a more robust state-tracking function, we define the following quality measure for a newly generated Davidson root with CI vector $\vec{c}$.

$$Q_{W\Gamma}(\vec{c}) = W_0(\vec{c}) + W_1(\vec{c}) + D(\vec{c}), \tag{9}$$

$$D(\vec{c}) = \frac{|| \Gamma_t - \Gamma(\vec{c}) ||}{n_{\mathrm{CAS}}}. \tag{10}$$

Here Γ_t and $\Gamma(\vec{c})$ are the one-body reduced density matrices (1RDMs) for the target wave function and the current root in question, respectively, with Γ_t rotated into the current orbital basis in order to reduce sensitivity to orbital changes. While one would like to do something

similar for Q_{MOM}, it is not obvious how this would be done as active-virtual orbital rotations prevent $\vec{c}_t$ from being expressible within the new orbitals' active space. Note that we divide the Frobenius norm of the 1RDM difference by the number of active orbitals n_{CAS} in order to make the relative importance of W_0, W_1, and D less sensitive to the size of the chosen active space. This choice is built on the idea that the out-of-active-space parts of the density matrix difference are small, as the occupation vector is fixed for these orbitals and we do not expect qualitative changes in the closed-shell orbital shapes. While it may be that a different balance between W_0, W_1, and the density matrix difference is optimal, in practice we have found that this choice is effective in most cases. While one could of course also include W_2 in this quality measure, we have chosen to omit it due to its relatively high cost of evaluation and our observation that $Q_{W\Gamma}$ is quite effective even without it.

As for why we chose a density matrix difference for our measure of iteration-to-iteration similarity, the logic is that we could have measured similarity via a combination of any number of wave function properties (e.g. dipole moment), but a large number of properties are themselves determined via the 1RDM. Of course, one could also consider 2RDM differences, but for simplicity's sake we have for now limited our investigation to differences of 1RDMs. As our results will demonstrate, the quality measure $Q_{W\Gamma}$, although not perfect and certainly less sophisticated than it could be, is far more effective at dealing with root flipping when attempting to track a specific excited state through a CASSCF ticktock optimization than either SRS or Q_{MOM}.

C. Optimization procedure

To summarize, the overall optimization procedure that we test here involves the following steps.

1. Choose an orbital basis as an initial guess, perform an initial CASCI calculation, and select the root that will be targeted for state-specific convergence.

2. Relax the orbital coefficient matrix $= C_0 e^X$ via an L-BFGS minimization of either $|\nabla_X L|^2 + |\nabla_\mu L|^2$ (with $W \approx W_0 + W_1$) or the even simpler objective function $|\nabla_X E|^2$. While the latter is in principle more prone to variational collapse, we find that in the cases studied here the two orbital rotation objective functions lead to the same results. Either way, we implement the gradients needed for L-BFGS within the TensorFlow automatic differentiation framework.[55]

3. Perform a new CASCI calculation (via the pySCF package[56]) to obtain the low-lying roots of the appropriate space and spin symmetry in new orbital basis.

4. Select the root with the lowest value for Q_{MOM} or $Q_{W\Gamma}$, depending on which quality measure is being employed. If instead one is following the SRS approach, then simply select the root based on its position in the energy ordering.

5. Return to step 2 and continue until an overall CASSCF energy stationary point is found.

III. RESULTS AND DISCUSSION

We now turn to a collection of molecular examples with which we seek to gain insight into three key questions. First, how effective is the $W\Gamma$ approach at overcoming root flipping in comparison to the MOM and SRS methods? Second, can the $W\Gamma$ approach succeed in cases where there are nearly degenerate states that cannot be distinguished by W alone? Finally, do the SS-CASSCF solutions that this approach helps us find outperform their SA-CASSCF counterparts as reference functions for post-CASSCF methods like CASPT2 and Davidson-corrected multi-reference configuration interaction with singles and doubles (MRCI+Q)?

These questions are studied in the molecular systems summarized in Table I. The cc-pVDZ basis set was used throughout, as were CASSCF energy, orbital gradient, and density matrix convergence thresholds of 10^{-7}, 10^{-4}, and 10^{-4}, respectively. In most cases, we began our optimizations in the HF orbital basis, but in MgO we began in the LDA basis instead because both the MOM and $W\Gamma$ approaches converged more rapidly from an LDA starting point. All CASPT2 and MRCI+Q calculations were carried out using Molpro,[57] while EOM-CCSD calculations were performed using QChem.[58] Note that all post-CASSCF methods employed the usual frozen-core approximation, but for CASSCF itself the core was frozen or not as described in Table I. For MRCI+Q, we used Molpro's default convergence thresholds for the energy and density matrix in all systems except for MgO, where we found it necessary to set them both to 10^{-5} in order to avoid unstable oscillations in the energies.

TABLE I. Summary of frozen and active orbitals used for all systems.

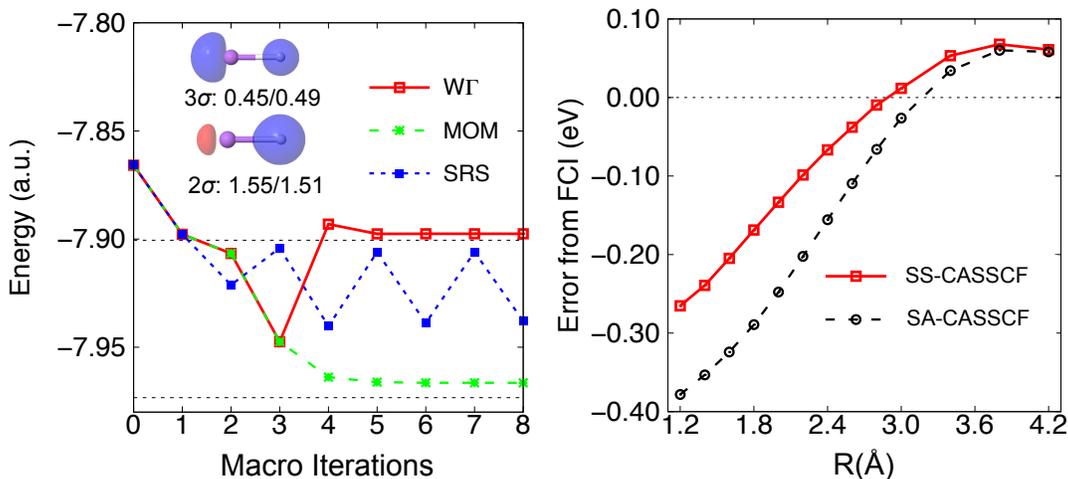| Molecules	frozen orbitals	active orbitals
LiH	none	4e,4o: Li $1s2s2p_z$ H $1s$
O_3	O $1s2s$	12e,9o: O $2p$
CH_2O	C $1s2s$ O $1s2s$	8e,8o: C $2p$ O $2p$ H $1s$
MgO	none	8e,8o: Mg $3s3p$ O $2s2p$

FIG. 1. Left panel: The change in energy during CASSCF optimizations of the first excited state ($A^1\Sigma^+$) of LiH at a bond distance of 2.6 Å. The SRS, MOM, and $W\Gamma$ approaches all start with the Hartree-Fock orbitals as the initial guess. For reference, the horizontal lines show the FCI energies for the ground and excited states. The inset shows the 2σ and 3σ natural orbitals and occupation numbers for the excited state as produced by the $W\Gamma$ approach. The occupation numbers from FCI (the second number) are given for comparison. Li and H atoms are in purple and white, respectively. Right panel: Excitation energy errors relative to FCI at different bond distances R.

A. LiH

Let us begin with the well-known example of root flipping that occurs in LiH. Near the equilibrium bond length ($\sim$1.6Å), this molecule's ground state ($X^1\Sigma^+$) is basically ionic, while the first excited state ($A^1\Sigma^+$) is predominantly neutral due to a charge transfer excitation. As the bond length is stretched, however, the ground state becomes increasingly neutral and the excited state increasingly ionic. At intermediate distances, the system shows an avoided crossing between these states that causes a well-recognized example of root flipping for SRS, as seen in the left panel of Figure 1. As the orbitals are optimized for the excited state, the energy of the ground state CASCI root rises while that of the excited state root falls. Soon the two roots flip in the energy ordering, at which point SRS is now effectively trying to optimize the orbitals for the ground state, which in turn causes the energy ordering to flip back. This process continues indefinitely, preventing the SRS approach from converging at all.

After the first orbital relaxation, both the MOM and $W\Gamma$ approaches recognize that the ordering of the roots has changed and select the lower root, thus diverging from the optimization path of SRS. Their two quality measures then select the same root for one more iteration, at which point orbital relaxations clearly work to push the state towards the ground state, as can be seen by the large energy lowering between macro iterations 2 and 3. It is here that the two methods diverge, with MOM's quality measure selecting the root that ultimately becomes the ground state. The $W\Gamma$ measure, on the other hand, successfully keeps track of the excited state root

and ultimately converges to a CASSCF stationary point that clearly corresponds to the desired excited state. As seen in the right panel of Figure 1 and in the more detailed tabulation within the Supporting Information (SI), this success is repeated at all bond distances, and we find that the excitation energies based on SS-CASSCF energy differences tend to be a bit more accurate than those based on equally-weighted SA-CASSCF energies.

B. Asymmetric O_3

We now turn to asymmetrically stretched ozone, in which we focus our attention on two nearly degenerate excitations at the somewhat arbitrary asymmetric geometry $R_{O_1O_2} = 1.3$ Å, $R_{O_2O_3} = 1.8$ Å, $\angle O_1O_2O_3 = 120°$. These states correspond to the fourth and fifth SS-CASSCF excitation energies (relative to the ground $1^1A'$ state) in this C_s geometry's A'' representation, and so we will label them as $4^1A''$ and $5^1A''$ even though we will see that when out-of-active-space correlation is included it is no longer obvious which actually has the lower energy. The $4^1A''$ state's primary component is the configuration with four singly occupied orbitals resulting from a double excitation that moves one electron each from the two highest-occupied A' orbitals into the lowest unoccupied A' and A'' orbitals, whereas the $5^1A''$ state's primary component is the single excitation from the highest-occupied A'' orbital into the lowest unoccupied A' orbital. That said, both of these states have nontrivial coefficients on the other state's primary component, meaning that neither state can be described as a simple single excitation. To add to the confusion, these
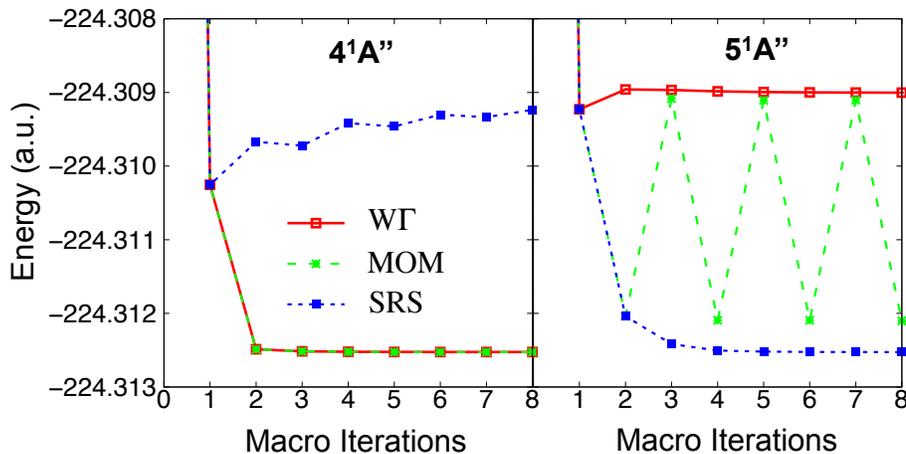
FIG. 2. The energy of the selected root after each CASCI calculation during SS-CASSCF optimizations of asymmetric O_3. The left and right panels show results when the SRS, MOM, and $W\Gamma$ methods are initialized with the $4^1A''$ and $5^1A''$ CASCI roots, respectively, in the initial Hartree-Fock orbital basis.

states show up in the reverse order ($5^1A'$ before $4^1A''$) in the energy-ordered CASCI roots in the initial HF orbital basis, and so root flipping is essentially guaranteed during state-specific optimization.

Unlike the lowest three A'' states — all of which converge without trouble using SRS — this pair of states proves a nontrivial challenge for state specific optimization. As shown in Figure 2, the MOM finds the $4^1A''$ stationary point without difficulty when starting from the corresponding (fifth) A'' root of the HF-orbital CASCI, but becomes trapped in a limit cycle when we target the $5^1A''$ by starting from the corresponding (fourth) root. SRS succeeds in finding both stationary points, but it does so by making a root-flipping-induced qualitative swap in the state being tracked. This (or perhaps collapse to a lower state) is to be expected, as the energy ordering of these states is reversed at their individual stationary points as compared to the initial HF-orbital CASCI. The practical results are twofold. First, the state the user gets is not the same excited state as was initially targeted. Second, the convergence of the $4^1A''$ state proves to be very slow with the gradient still not quite converged even after 100 macro iterations.

The $W\Gamma$ approach, in contrast, converges rapidly for both the $4^1A''$ and $5^1A''$ states when it is initiated with the corresponding root from the initial HF-orbital CASCI. Unlike SRS, these optimizations do not involve any switching between qualitatively different states, despite the fact that the energy ordering does change. We explicitly verify that the desired state is tracked at every iteration of the optimization in Figure 3, which shows both how little the relevant active space natural orbital occupation numbers change as well as how clearly qualitatively different the $4^1A''$ and $5^1A''$ states are. At convergence, we see that the stationary points found in the $W\Gamma$ approach correspond closely to both the initial CASCI wave functions and the results from equal-weighted 3-state ($1^1A'$, $4^1A''$, $5^1A''$) and 6-state ($1^1A'$, $1^1A''$, $2^1A''$, $3^1A''$, $4^1A''$, $5^1A''$) SA-CASSCF calculations.

Of course, when comparing to SA-CASSCF, the more pressing question is whether the ability to find the correct stationary points in SS-CASSCF offers any advantages in final accuracy. To investigate this question, we have performed both CASPT2 and MRCI+Q calculations based on the different SS and SA wave functions, the results of which are displayed in Figure 4. We find that MRCI+Q predicts the states to be nearly degenerate with excitation energies within 0.02 eV of each other regardless of whether we base it on either of the SA-CASSCF calculations' orbitals or on the SS-CASSCF orbitals.

Note that the latter SS-MRCI+Q case amounts to three different MRCI+Q calculations, one each in the ground, $4^1A''$, and $5^1A''$ state's orbital basis, with the appropriate MRCI+Q root's energy extracted in each case. CASPT2, on the other hand, only predicts the degeneracy to be within 0.02 eV when based on the SS-CASSCF wave functions, showing gaps of more than 0.07 eV in both the 3-state and 6-state SA cases (note that we used single-state zero order Hamiltonians for CASPT2 both when starting from SA-CASSCF and SS-CASSCF). One way to explain these findings is to note that the out-of-active-space single excitations that provide state-specific orbital relaxations are treated perturbatively in CASPT2, and so the fact that SS-CASSCF has already provided state-specific orbitals puts us in a regime where the perturbative assumption of small singles (and doubles) coefficients is more likely to be satisfied in practice. Although the improvement is modest, the SS-CASPT2 is in closer agreement with MRCI+Q than is SA-CASPT2, which is encouraging as it suggests that SS-CASSCF may help the lower-cost CASPT2 method agree better with the higher-cost and typically higher-accuracy MRCI+Q method. As we will see, this same pattern plays out again
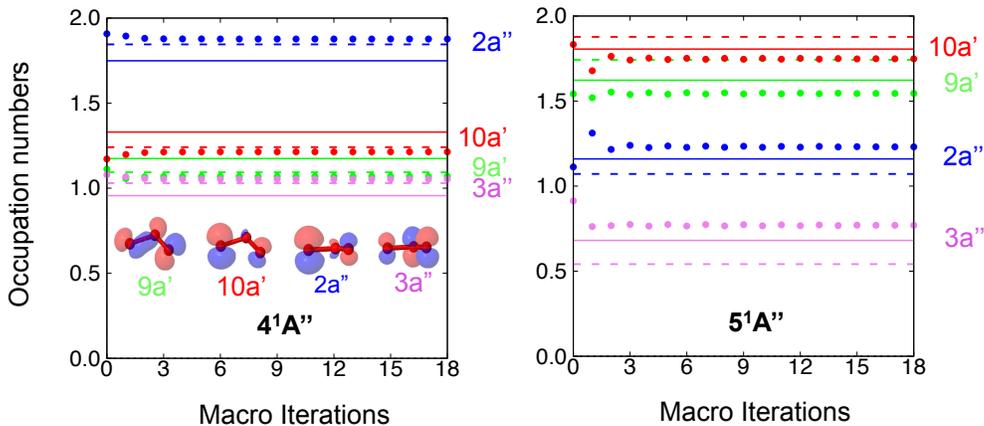
FIG. 3. The natural orbital occupation numbers after each CASCI calculation during the $W\Gamma$-guided CASSCF optimizations of the $4^1A''$ (left panel) and $5^1A''$ (right panel) states in asymmetric O_3, with the 0th iteration corresponding to the initial HF-orbital CASCI. For comparison, the solid and dashed horizontal lines give the corresponding final occupation numbers from 3-state and 6-state SA-CASSCF calculations, respectively. The inset shows the final shapes of these four natural orbitals for the SS-CASSCF $4^1A''$ case. Note that the shapes for the SS-CASSCF $5^1A''$ natural orbitals look very similar even though the natural occupation numbers largely differ.
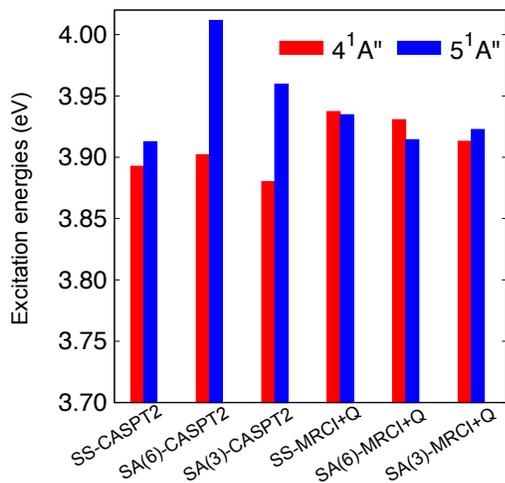

FIG. 4. CASPT2 (with no level shift) and MRCI+Q excitation energies for the $4^1A''$ and $5^1A''$ states of asymmetric O_3 when starting from SS-CASSCF and 3- or 6-state SA-CASSCF wave functions.

and again in the states studied below.

C. CH$_2$O

We now turn to formaldehyde — with C_{2V} geometry $R_{CO} = 1.2\mathring{A}$, $R_{CH} = 1.1\mathring{A}$, $\angle(H,C,H) = 116.43°$ — to investigate a case of two excited states with very similar components. According to EOM-CCSD, there are two excited states in the $^1A_1$ symmetry sector whose principal components are superpositions of the $1b_1 \rightarrow 2b_1$ and

$2b_2 \rightarrow 3b_2$ single-electron transitions (orbitals are plotted in Figure 5). These two states, which in the HF-orbital CASCI are the 2^1A_1 and 4^1A_1 states, are dominated by the sum of or difference between these components and turn out to have very similar natural orbital occupation numbers (see Figure 5). One might worry that such similarity could confuse a density matrix based approach, whereas it looks at first glance like the MOM approach should be effective as the CI vectors are clearly quite different, at least according to EOM-CCSD. As we will see, however, the more difficult of these two states confounds both the MOM and $W\Gamma$ approaches as we defined them in Section II, and we were only able to succeed with SS-CASSCF by *increasing* the weighting of the density matrix difference within our $Q_{W\Gamma}$ measure.

Starting all CASSCF optimizations with HF orbitals as the initial guess, we find that the SRS, MOM, and $W\Gamma$ approaches all succeed at finding the energy stationary point associated with the 2^1A_1 state. However, even though we are using a nearly full-valence active space, none of these approaches succeeded for the 4^1A_1 state when applied as described in the theory section. For example, Figure 5 shows how the MOM approach collapses to the ground state after just a few macro iterations. Interestingly, this 4^1A_1 state proves to be an example of a case in which the relative weighting of the W_1, W_2, and D components matters when employing the $W\Gamma$ quality measure of Eq. (9). When we double the weighting of the density matrix difference, i.e. $Q_{W\Gamma} \rightarrow W_0 + W_1 + 2D$, the $W\Gamma$ approach converges successfully to the 4^1A_1 state, as seen in Figure 5. Thus, despite this being a case where it is not obvious a priori that root flipping will occur and despite the states in question having easily distinguished CI vectors at the outset of the optimization, we find that the approach that incorporates density matrix differences
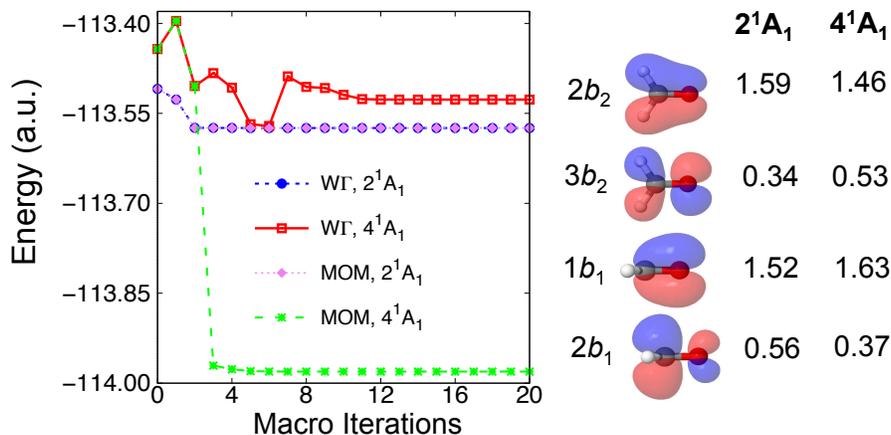
FIG. 5. Left panel: changes in the energy during MOM and reweighted $W\Gamma$ CASSCF optimizations of two CH_2O excited states. Right panel: plots of the four most important orbitals contributing to these states (we are showing SS-CASSCF natural orbitals for 4^1A_1, the corresponding 2^1A_1 orbitals are quite similar) as well as the SS-CASSCF natural orbital occupation numbers.

is more effective at converging to the relevant stationary points than either the SRS or MOM approaches.

To ascertain how much the ability to find these states' stationary points matters at the end of the day, we compare CASPT2 and MRCI+Q results based on SS- and SA-CASSCF in Table II. Here we have employed two SA schemes, one that gives equal weighting to the first four $^1A_1$ states (w1) and one that uses weightings of $(0.6, 0.1, 0.1, 0.2)$ and $(0.2, 0.1, 0.1, 0.6)$ when modeling the ground and 4^1A_1 states, respectively (w2). As CASPT2 was found to suffer from intruder states, we have employed two different level shifts in each case. One difference that is noticed immediately is that, compared

TABLE II. Excitation energies (eV) for the 2^1A_1 and 4^1A_1 states of CH_2O along with the IPEA level shifts (a.u.) used for CASPT2.

Method	Shift	2^1A_1	4^1A_1
SA(w1)-CASPT2	0.1	9.71	10.94
SA(w1)-CASPT2	0.2	9.83	11.15
SA(w1)-MRCI+Q	N/A	9.86	11.10
SA(w2)-CASPT2	0.1	N/A	11.03
SA(w2)-CASPT2	0.2	N/A	11.19
SA(w2)-MRCI+Q	N/A	N/A	11.21
SS-CASPT2	0.1	9.74	11.24
SS-CASPT2	0.2	9.83	11.35
SS-MRCI+Q	N/A	9.83	11.30
EOM-CCSD	N/A	10.08	11.38

to SA-CASPT2, the excitation energies for SS-CASPT2 are less sensitive to changes in the level shift, which is certainly a desirable property. For the 2^1A_1 state, we see that CASPT2 and MRCI+Q excitation energies are not so sensitive to the choice between SA-CASSCF and SS-CASSCF. Note that for this state, there is some question about how accurate EOM-CCSD is expected to be, as the CASSCF CI vectors all show a significant weight on the $(2b_2)^2 \to (2b_1)^2$ double excitation, and EOM-CCSD is known to overestimate the energies of double excitations due to its inability to relax their orbitals.[4–6] The 4^1A_1 state, on the other hand, does not have any significant doubly excited components, and so the coupled cluster result should be more reliable for it. Indeed, we find that as we go in order of increasing state specificity, SA(w1) $\to$ SA(w2) $\to$ SS, the CASPT2 and MRCI+Q results for this state move towards the coupled cluster number, suggesting that there is a modest accuracy improvement to be had for this state by achieving SS-CASSCF. Recalling that overall cost in large systems is dominated by post-CASSCF methods, we see that accuracy improvements of this type should be achievable at a negligible extra cost compared to a SA approach.

D. MgO

The final system we investigate is MgO at a bond distance of 1.8 Å. Like LiH, the ground state at this near-equilibrium geometry has ionic character, but in MgO the closed-shell determinant that dominates the ground state wave function is somewhat doubly ionic, with both of Mg's 3s electrons moving into a bonding σ orbital with substantial O 2p character. We therefore expect to find a challenging assortment of low-lying excitations including double charge transfers that return the system

to more neutral states as well as excitations that retain the ground state's ionic nature. Indeed, Table III shows that the eight lowest-lying $^1A_1$ states in an initial CASCI calculation in the LDA orbital basis (LDA-CASCI) contain three excited states whose dipoles suggest that they largely retain the ground state's ionic nature, as well as four charge transfer states whose electron densities have shifted significantly towards the Mg atom. The prevalence of double excitations and even double charge transfers among these states makes MgO a clear example where a multi-reference treatment is necessary, but the mixture of neutral and ionic states makes it hard to know a priori how to construct a SA scheme that treats all states fairly. Ideally, this concern could be bypassed via state-specific optimization, in which the CASSCF energy stationary point corresponding to each of these states was located. However, as we now discuss, MgO proves to be especially difficult in this regard, with SRS failing to converge to the targeted excited state in every case and the MOM and $W\Gamma$ approaches succeeding in only 4 and 6 out of the 7 cases, respectively.

To begin, we emphasize that the SRS approach fails to converge to the initially targeted state in every one of the seven MgO excited states studied here. That said, Figure 6 shows that although there are cases where SRS does not converge at all, it often finds a stationary point corresponding to a different state than the one originally targeted. This is true even when targeting state V2, which in Figure 6 looks like a success. However, inspecting the converged wave function's natural orbital occupation numbers and comparing to those in Figure 7 (and Table S2 in the SI), we find that when targeting state V2, SRS converges to the nearly-degenerate stationary point belonging to state CT1. Similarly, when targeting CT1, CT2, and CT3, SRS converges to the stationary points for CT2, CT3, and a state outside our set of seven, respectively.

Compared to SRS, both the MOM and $W\Gamma$ approaches prove substantially more effective. For the six states shown in Figure 6, they at least both converge to a stationary point in every case. Careful inspection of the final wave functions, including the natural orbital occupation analysis of Figure 7, shows that $W\Gamma$ in fact converges to the intended stationary point in all six cases. In contrast, MOM succeeds in only four cases: when attempting to find the stationary point for the CT1 and CT4 states, it collapses to the V1 stationary point. We therefore see that, as in other systems in this study, the $W\Gamma$ approach proves more effective for SS-CASSCF than the CI-vector-based MOM, while both of these greatly outperform SRS. That said, we note that none of these methods succeeded in converging to the M1 stationary point, and indeed we have not been able to locate this point by any means. Thus, while the $W\Gamma$ approach is promising, it would clearly benefit from further improvements in future work.

To investigate whether these SS-CASSCF solutions offer advantages when using post-CASSCF methods, we must first decide on SA schemes to compare them against. There are of course an infinite number of possible SA weighting schemes, and it can be difficult to predict a priori what will be most effective. While this difficulty is especially concerning in a case like MgO where different states have significantly different charge distributions, we have used two SA schemes here that each try to achieve simple forms of balance. First, we take an equally-weighted 8-state SA approach, which we denote as SA(8). Second, we attempt to minimize the number of states involved in the average by only including enough states such that the desired state is present, and equally weight the states in the average. This SA(N) approach is much less straightforward, because the Nth initial LDA-CASCI root does not necessarily show up in the same order (or at all) in an N-state SA. States V1, CT1, CT2, and CT3, show up as the Nth root in equal-weight N-state SA calculations for $N =$ 2, 3, 5, and 8, respectively. However, the near-degeneracy between V2 and CT1 at the CASSCF level prevents V2 from showing up at all in a 4-state SA. The minimum number of states to include in order for V2 to show up turned out to be $N = 6$, in which case it appears as the 6th root. State CT4 was even more problematic, and indeed does not show up at all in equal-weight SA calculations for any $N \leq 8$, which is unsurprising in light of how much higher in energy
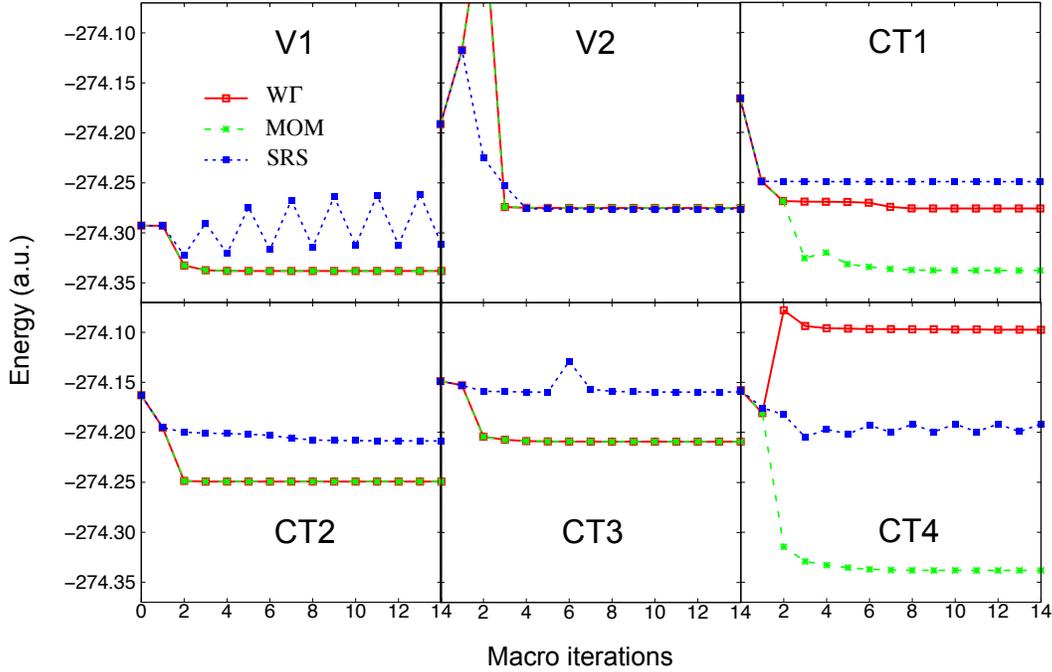
TABLE III. Information for the eight lowest MgO $^1A_1$ states in the initial LDA-CASCI calculation, including their dipole moments (μ, in Debye), dominant excitation characters (DECs) and primary active space configurations (PASCs). For states with two important configurations of nearly equal weight, we have listed both. See Figure S1 in the SI for the LDA orbital shapes. Note that we order the states' labels within a category (GS for ground state, M for missing, V for valence, CT for charge transfer) by their SS-CASSCF energies and that the ordering of CT3 and CT4 inverts during SS optimization.

State	Label	μ	DECs	PACSs
1^1A_1	GS	–3.95	N/A	$5\sigma^2 2\pi^4 6\sigma^2$
2^1A_1	M1	–5.39	$6\sigma \rightarrow 7\sigma$	$5\sigma^2 2\pi^4 6\sigma^1 7\sigma^1$
			$2\pi \rightarrow 3\pi$	$5\sigma^2 2\pi^3 6\sigma^2 3\pi^1$
3^1A_1	V1	–4.88	$2\pi \rightarrow 3\pi$	$5\sigma^2 2\pi^3 6\sigma^2 3\pi^1$
4^1A_1	V2	–5.93	$6\sigma \rightarrow 8\sigma$	$5\sigma^2 2\pi^4 6\sigma^1 8\sigma^1$
5^1A_1	CT1	3.84	$2\pi^2 \rightarrow 7\sigma^2$	$5\sigma^2 2\pi^2 6\sigma^2 7\sigma^2$
6^1A_1	CT2	3.93	$2\pi^2 \rightarrow 7\sigma^2$	$5\sigma^2 2\pi^2 6\sigma^2 7\sigma^2$
			$6\sigma^2 \rightarrow 7\sigma^2$	$5\sigma^2 2\pi^4 7\sigma^2$
7^1A_1	CT4	2.33	$2\pi 6\sigma \rightarrow 3\pi 7\sigma$	$5\sigma^2 2\pi^3 6\sigma^1 3\pi^1 7\sigma^1$
8^1A_1	CT3	3.66	$2\pi 6\sigma \rightarrow 3\pi 7\sigma$	$5\sigma^2 2\pi^3 6\sigma^1 3\pi^1 7\sigma^1$
			$2\pi^2 \rightarrow 7\sigma^2$	$5\sigma^2 2\pi^2 6\sigma^2 7\sigma^2$

FIG. 6. The selected CASCI root's energy at each macro iteration during SS-CASSCF optimizations of six MgO excited states under the the SRS, MOM, and $W\Gamma$ approaches to root selection.

its stationary point in Figure 6 is relative to the initial LDA-CASCI energy. Thus, in summary, CT4 was not included in the SA comparison, while the SA(N) approach for states V1, V2, CT1, CT2, and CT3 uses the values $N = 2, 6, 3, 5$, and 8, respectively.

The LDA-CASCI, CASSCF, CASPT2, and MRCI+Q excitation energies for the six $^1A_1$ excited states for which SS-CASSCF stationary points were found are shown in Table IV. In Figure 8, we show the difference from the corresponding MRCI+Q excitation energies for both CASSCF and CASPT2 when working in different methods' (LDA, SA(8), SA(N), SS-CASSCF) orbital bases. In every case, the SS-CASSCF orbital basis leads to the smallest difference between CASPT2 and MRCI+Q excitation energies, which supports the hypothesis that state-specific orbital relaxations should be beneficial when relying on the perturbative assumption that the CASSCF wave function is close to the exact wave function. Unsurprisingly, working in the LDA orbital basis led to the largest differences, while the SA approaches were in between these two extremes. While SS-CASSCF thus appears to offer improvements for CASPT2, it is important to note that its CASSCF excitation energies are often not closer to the corresponding MRCI+Q when compared to the situation in SA-CASSCF, as revealed by the left panel of Figure 8. This result should not be surprising, as CASSCF lacks all out-of-active-space weak correlation effects, the size of which is expected to differ significantly for different states. For example, neutral states have roughly 12 and 8 electrons located on the Mg and O atoms, respectively, while the doubly ionic ground state

has 10 and 10, and so these states have different numbers of electrons in close proximity to each other. As most of the energetic effects coming from weak electron correla-

TABLE IV. Excitation energies for six $^1A_1$ excited states of MgO. CASPT2 used an IPEA level shift of 0.2 a.u. to avoid intruder states. In some cases, states were not found (NF) in the SA approach, and in others, the MRCI+Q Davidson solver did not converge (NC).

| Method	V1	V2	CT1	CT2	CT3	CT4
LDA-CASCI	3.70	6.46	5.10	7.14	8.16	8.07
LDA-CASPT2	3.80	6.42	5.68	6.21	6.02	7.31
LDA-MRCI+Q	3.81	6.33	5.92	5.59	6.57	NC
SA(8)-CASSCF	3.97	7.24	5.10	5.67	6.94	NF
SA(8)-CASPT2	3.91	6.76	5.68	6.23	7.02	NF
SA(8)-MRCI+Q	3.81	6.31	5.92	6.46	6.66	NF
SA(N)-CASSCF	4.10	6.76	5.29	5.94	6.94	NF
SA(N)-CASPT2	3.89	6.57	5.66	6.32	7.02	NF
SA(N)-MRCI+Q	3.76	6.39	5.99	6.46	6.66	NF
SS-CASSCF	4.88	6.60	6.57	7.16	8.39	11.47
SS-CASPT2	3.76	6.55	5.73	6.32	6.91	10.44
SS-MRCI+Q	3.79	6.52	5.97	6.44	6.76	NC
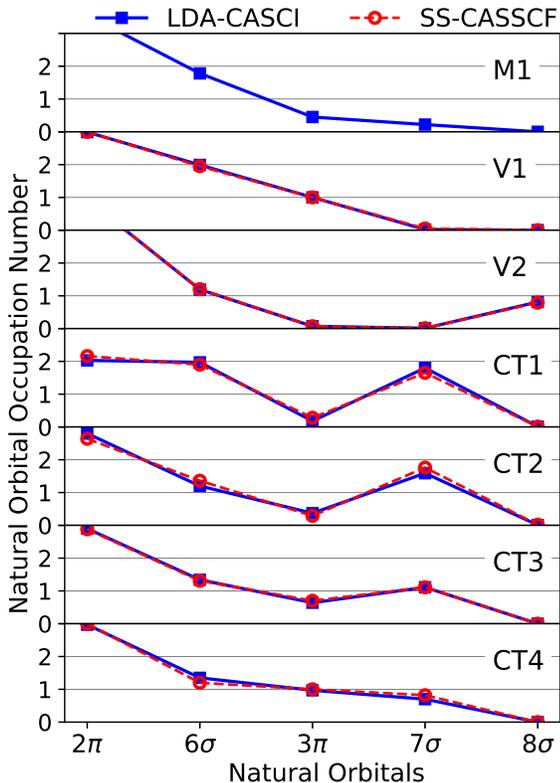
FIG. 7. MgO's excited states may be distinguished using natural orbital occupation numbers.

tions are local in nature, a crude accounting of how many local electron pairs can be enumerated on each atom (12 choose 2 and 8 choose 2 compared to twice 10 choose 2) suggests that the size of weak correlation effects should be different for different states. Thus, while SS-CASSCF does not and is not expected to bring the CASSCF energetics closer to those of MRCI+Q, it does serve as a better platform for CASPT2 than either of the two SA approaches.

IV. CONCLUSION

We have presented an approach for locating state-specific CASSCF energy stationary points that overcomes root flipping via a metric of quality based on an excited state variational principle and density matrix differences. This approach was inspired by both the maximum overlap method and recent progress in excited state variational methods, as well as the twin goals of controlling cost and maintaining compatibility with widely used tick-tock methods that treat orbital and CI coefficients in separate optimization steps. Our central finding is that this approach is highly effective in the face of both

root flipping and near-degeneracies, significantly outperforming both simple root selection and a CI-vector-based adaptation of the maximum overlap method. Although the improvements can be modest, we find that, compared to state averaging, state-specific CASSCF provides a superior starting point for CASPT2 in that it brings this method's excitation energies more closely in line with those of the more reliable but expensive MRCI+Q. As the additional cost of converging each CASSCF state individually is typically small compared to the cost of post-CASSCF methods, we are happy to recommend that applications of this approach be explored across a wider spectrum of chemical systems.

Aside from post-CASSCF accuracy improvements, the ability to locate individual excited states' energy stationary points delivers a number of important and useful properties. Unlike state averaged CASSCF, the excited state specific version retains the size extensivity and size consistency of its ground state counterpart, which implies that its CASSCF excitation energies will be size intensive. Furthermore, stationarity with respect to the energy makes the calculation of many properties, not least of which are the nuclear gradients, significantly more straightforward as it avoids the need for Lagrangian or Z-vector techniques. Finally, state-specific optimization avoids the need to decide on how many states to average and what weights to use, choices that are both difficult to justify a priori and which can have significant effects on CASSCF and post-CASSCF energetics.

While we have presented evidence in this study of how our approach to stationary points can be beneficial to CASPT2, other excited state methods are likely to benefit as well. For example, the most difficult optimization stage in the application of variational Monte Carlo to excited states tends to be the step in which orbital relaxations are enabled. A question that we are very eager to answer in future work is thus whether orbitals optimized for an individual excited state by CASSCF could serve as a good alternative to those optimized by Monte Carlo, and whether these two sources of optimized orbitals make any difference in practice when fed in to the diffusion Monte Carlo method. Back in the area of quantum chemistry, it is quite possible that the recently-introduced excited state mean field method could be accelerated by tick-tock optimization methods, at which point it too would come face to face with root flipping issues that the $W\Gamma$ approach presented here might alleviate. Finally, it will be exciting to pair this improved state-specific methodology with large active space methods, which already offer substantial advantages when dealing with root flipping.

Looking forward, there are numerous promising ways to extend and apply this new methodology. Although we have chosen to avoid the double excitation components of the variational principle in the present study to keep the approach simple and inexpensive, an excellent approximation to the effect of this term should be accessible via the application of modern selective CI integral
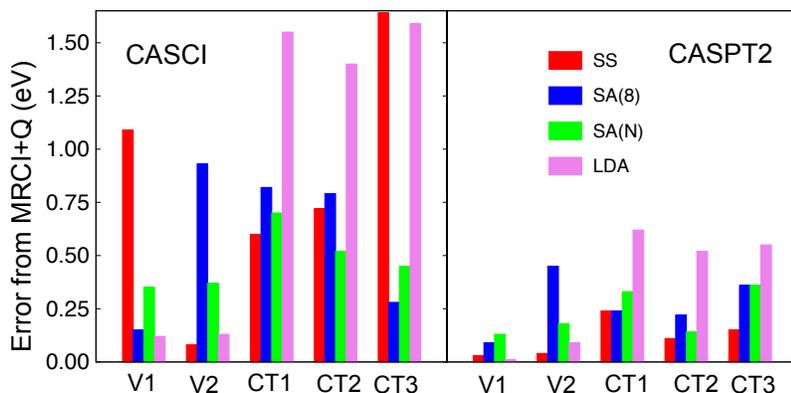
FIG. 8. For different sets of molecular orbitals (SS-CASSCF, SA(8)-CASSCF, SA(N)-CASSCF, and LDA) we compare CASCI and CASPT2 excitation energies with MRCI+Q. In each case, all three methods are carried out in the same basis and the difference taken between the excitation energy for the method in question (CASCI on the left panel and CASPT2 on the right) and the excitation energy for MRCI+Q. Note that CASPT2 used an IPEA level shift of 0.2 a.u. to avoid intruder states.

handling technology. Similarly, although the two-body reduced density matrix in the full orbital space may be impractical to employ for root comparisons due to high memory demands, one can imagine inspecting it in a second, much larger active space. As we have already identified at least one state in MgO that continues to defy all of the attempted state specific optimizations, we are eager to explore these extensions in order to make the methodology more reliable and robust. In terms of applications, the freedom to define the target state in whichever way the user chooses would appear to be a natural fit when trying to optimize the two (or more!) adiabatic states near a conical intersection. While these states mix freely at the intersection itself and so differentiating between them in terms of their diabatic character is less meaningful, away from but nearby the intersection the use of diabatic states as the targets for state-specific CASSCF optimization could help identify which adiabatic states are which and also help ensure that all

the states in question were found in the state specific optimization process. More generally, it will be interesting to investigate how a more aggressively state-specific approach to excited states in CASSCF can be applied to the ongoing challenge of handling the strong excited state orbital relaxations common to core and charge transfer excitations.

[*] lantrann@berkeley.edu

[†] eneuscamman@berkeley.edu

[1] A. Dreuw and M. Head-Gordon, Chem. Rev. **105**, 4009 (2005).

[2] A. Dreuw, J. L. Weisman, and M. Head-Gordon, J. Chem. Phys. **119**, 2943 (2003).

[3] J. E. Subotnik, J. Chem. Phys. **135**, 071104 (2011).

[4] A. I. Krylov, Annu. Rev. Phys. Chem. **59**, 433 (2008).

[5] J. D. Watts and R. J. Bartlett, Chem. Phys. Lett. **258**, 581 (1996).

[6] E. Neuscamman, J. Chem. Phys. **145**, 081103 (2016).

[7] J. Liu, D. Matthews, S. Coriani, and L. Cheng, J. Chem. Theory Comput. **15**, 1642 (2019).

[8] G. Onida, L. Reining, and A. Rubio, Rev. Mod. Phys. **74**, 601 (2002).

[9] A. Dreuw and M. Wormit, WIR: Comput. Mol. Sci. **5**, 82 (2015).

[10] K. Ruedenberg, M. W. Schmidt, M. M. Gilbert, and S. T. Elbert, Chem. Phys. **71**, 41 (1982).

[11] H.-J. Werner and P. J. Knowles, J. Chem. Phys. **82**, 5053 (1985).

[12] P. J. Knowles and H.-J. Werner, Chem. Phys. Lett. **115**, 259 (1985).

[13] B. O. Roos, Adv. Chem. Phys. **69**, 399 (1987).

[14] J. Olsen, B. O. Roos, P. Jørgensen, and H. J. A. Jensen, J. Chem. Phys. **89**, 2185 (1988).

[15] P.-Å. Malmqvist, K. Pierloot, A. R. M. Shahi, C. J. Cramer, and L. Gagliardi, J. Chem. Phys. **128**, 204109 (2008).

[16] G. Li Manni, D. Ma, F. Aquilante, J. Olsen, and L. Gagliardi, J. Chem. Theory Comput. **9**, 3375 (2013).

[17] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory* (Dover Publications, Mineola, N.Y., 1996).

[18] D. Ghosh, J. Hachmann, T. Yanai, and G. K.-L. Chan, J. Chem. Phys. **128**, 144117 (2008).

[19] D. Zgid and M. Nooijen, J. Chem. Phys. **128**, 144116 (2008).

[20] R. E. Thomas, Q. Sun, A. Alavi, and G. H. Booth, J. Chem. Theory Comput. **11**, 5316 (2015).

[21] G. Li Manni, S. D. Smart, and A. Alavi, J. Chem. Theory Comput. **12**, 1245 (2016).

[22] J. E. T. Smith, B. Mussard, A. A. Holmes, and S. Sharma, J. Chem. Theory Comput. **13**, 5468 (2017).

[23] L. N. Tran and D. Zgid, J. Phys. Chem. Lett. **8**, 2200 (2017).

[24] L. N. Tran, S. Iskakov, and D. Zgid, J. Phys. Chem. Lett. **9**, 4444 (2018).

[25] R. N. Diffenderfer and D. R. Yarkony, J. Phys. Chem. **86**, 5098 (1982).

[26] D. R. Yarkony, Acc. Chem. Res. **31**, 511 (1998).

[27] B. G. Levine, J. D. Coe, and T. J. Martínez, J. Phys. Chem. B **112**, 405 (2008).

[28] B. G. Levine, C. Ko, J. Quenneville, and T. J. Martínez, Mol. Phys. **104**, 1039 (2006).

[29] M. Garavelli, F. Bernardi, M. Olivucci, T. Vreven, S. Klein, P. Celani, and M. A. Robb, Faraday Discuss. **110**, 51 (1998).

[30] P. S. Bagus, Phys. Rev. **139**, A619 (1965).

[31] H.-l. Hsu, E. R. Davidson, and R. M. Pitzer, J. Chem. Phys. **65**, 609 (1976).

[32] A. Naves de Brito, N. Correia, S. Svensson, and H. Ågren, J. Chem. Phys. **95**, 2965 (1991).

[33] N. A. Besley, A. T. Gilbert, and P. M. Gill, J. Chem. Phys. **130**, 124308 (2009).

[34] H. J. A. Jensen, P. Jo/rgensen, and H. Ågren, J. Chem. Phys. **87**, 451 (1987).

[35] H.-J. Werner and W. Meyer, J. Chem. Phys. **74**, 5794 (1981).

[36] J. W. Snyder Jr, B. S. Fales, E. G. Hohenstein, B. G. Levine, and T. J. Martínez, J. Chem. Phys. **146**, 174113 (2017).

[37] S. D. Pineda Flores and E. Neuscamman, J. Phys. Chem. A **123**, 1487 (2018).

[38] J. A. Shea and E. Neuscamman, J. Chem. Phys. **149**, 081101 (2018).

[39] D. Weinstein, Proc. Nat. Acad. Sci. **20**, 529 (1934).

[40] J. MacDonald, Phys. Rev. **46**, 828 (1934).

[41] R. P. Messmer, Theor. Chim. Acta. **14**, 319 (1969).

[42] J. H. Choi, C. F. Lebeda, and R. P. Messmer, Chem. Phys. Lett. **5**, 503 (1970).

[43] A. T. B. Gilbert, N. A. Besley, and P. M. W. Gill, J. Phys. Chem. A **112**, 13164 (2008).

[44] G. M. Barca, A. T. Gilbert, and P. M. Gill, J. Chem. Theory and Comput. **14**, 1501 (2018).

[45] H.-Z. Ye, M. Welborn, N. D. Ricke, and T. Van Voorhis, J. Chem. Phys. **147**, 214104 (2017).

[46] L. Zhao and E. Neuscamman, J. Chem. Theory Comput. **12**, 3436 (2016).

[47] L. Zhao and E. Neuscamman, J. Chem. Theory Comput. **13**, 2604 (2017).

[48] J. A. R. Shea and E. Neuscamman, J. Chem. Theory Comput. **13**, 6078 (2017).

[49] N. S. Blunt and E. Neuscamman, J. Chem. Phys. **147**, 194101 (2017).

[50] P. J. Robinson, S. D. Pineda Flores, and E. Neuscamman, J. Chem. Phys. **147**, 164114 (2017).

[51] L. Zhao and E. Neuscamman, arXiv.org **1804.09663** (2018).

[52] N. S. Blunt and E. Neuscamman, J. Chem. Theory Comput. **15**, 178 (2019).

[53] K. Andersson, P. A. Malmqvist, B. O. Roos, A. J. Sadlej, and K. Wolinski, J. Phys. Chem. **94**, 5483 (1990).

[54] K. Andersson, P.-Å. Malmqvist, and B. O. Roos, J. Chem. Phys. **96**, 1218 (1992).

[55] https://www.tensorflow.org [accessed 22-March-2019].

[56] Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, *et al.*, WIREs Comput. Mol. Sci. **8**, e1340 (2018).

[57] H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, and M. Schütz, WIREs Comput. Mol. Sci. **2**, 242 (2012).

[58] Y. Shao, Z. Gan, E. Epifanovsky, A. T. Gilbert, M. Wormit, J. Kussmann, A. W. Lange, A. Behn, J. Deng, X. Feng, D. Ghosh, M. Goldey, P. R. Horn, L. D. Jacobson, I. Kaliman, R. Z. Khaliullin, T. Kuś, A. Landau, J. Liu, E. I. Proynov, Y. M. Rhee, R. M. Richard, M. A. Rohrdanz, R. P. Steele, E. J. Sundstrom, H. L. W. III, P. M. Zimmerman, D. Zuev, B. Albrecht, E. Alguire, B. Austin, G. J. O. Beran, Y. A. Bernard, E. Berquist, K. Brandhorst, K. B. Bravaya, S. T. Brown, D. Casanova, C.-M. Chang, Y. Chen, S. H. Chien, K. D. Closser, D. L. Crittenden, M. Diedenhofen, R. A. D. Jr., H. Do, A. D. Dutoi, R. G. Edgar, S. Fatehi, L. Fusti-Molnar, A. Ghysels, A. Golubeva-Zadorozhnaya, J. Gomes, M. W. Hanson-Heine, P. H. Harbach, A. W. Hauser, E. G. Hohenstein, Z. C. Holden, T.-C. Jagau, H. Ji, B. Kaduk, K. Khistyaev, J. Kim, J. Kim, R. A. King, P. Klunzinger, D. Kosenkov, T. Kowalczyk, C. M. Krauter, K. U. Lao, A. D. Laurent, K. V. Lawler, S. V. Levchenko, C. Y. Lin, F. Liu, E. Livshits, R. C. Lochan, A. Luenser, P. Manohar, S. F. Manzer, S.-P. Mao, N. Mardirossian, A. V. Marenich, S. A. Maurer, N. J. Mayhall, E. Neuscamman, C. M. Oana, R. Olivares-Amaya, D. P. O'Neill, J. A. Parkhill, T. M. Perrine, R. Peverati, A. Prociuk, D. R. Rehn, E. Rosta, N. J. Russ, S. M. Sharada, S. Sharma, D. W. Small, A. Sodt, T. Stein, D. Stück, Y.-C. Su, A. J. Thom, T. Tsuchimochi, V. Vanovschi, L. Vogt, O. Vydrov, T. Wang, M. A. Watson, J. Wenzel, A. White, C. F. Williams, J. Yang, S. Yeganeh, S. R. Yost, Z.-Q. You, I. Y. Zhang, X. Zhang, Y. Zhao, B. R. Brooks, G. K. Chan, D. M. Chipman, C. J. Cramer, W. A. G. III, M. S. Gordon, W. J. Hehre, A. Klamt, H. F. S. III, M. W. Schmidt, C. D. Sherrill, D. G. Truhlar, A. Warshel, X. Xu, A. Aspuru-Guzik, R. Baer, A. T. Bell, N. A. Besley, J.-D. Chai, A. Dreuw, B. D. Dunietz, T. R. Furlani, S. R. Gwaltney, C.-P. Hsu, Y. Jung, J. Kong, D. S. Lambrecht, W. Liang, C. Ochsenfeld, V. A. Rassolov, L. V. Slipchenko, J. E. Subotnik, T. V. Voorhis, J. M. Herbert, A. I. Krylov, P. M. Gill, and M. Head-Gordon, Mol. Phys. **113**, 184 (2015).

SUPPORTING INFORMATION

TRACKING EXCITED STATES IN WAVE FUNCTION OPTIMIZATION USING DENSITY MATRICES AND VARIATIONAL PRINCIPLES

Lan Nguyen Tran, Jacqueline A. R. Shea and Eric Neuscamman

TABLE S1. Excitation energies (in eV) for $A^1\Sigma^+$ of LiH from FCI, SA- and SS-CASSCF at different bond lengths (in Å).

R	FCI	SA-CASSCF	SS-CASSCF
1.2	3.85	3.47	3.58
1.4	3.71	3.36	3.47
1.6	3.47	3.14	3.26
1.8	3.18	2.88	3.01
2.0	2.86	2.61	2.73
2.2	2.55	2.34	2.45
2.4	2.25	2.09	2.18
2.6	1.98	1.87	1.94
2.8	1.75	1.68	1.74
3.0	1.56	1.54	1.58
3.4	1.36	1.39	1.41
3.8	1.35	1.41	1.42
4.2	1.45	1.51	1.51

TABLE S2. Occupation numbers and dipole moments μ (in Debye) of eight lowest $^1A_1$ states including the ground state from LDA-CASCI and SS-CASSCF calculations for MgO.

| CASCI ordering	Label	Method	Occupation numbers						μ
			5σ	2π	6σ	3π	7σ	8σ	
1^1A_1	GS	CASCI	2.00	3.92	1.63	0.07	0.36	0.00	−3.95
		SS-CASSCF	2.00	3.93	1.73	0.07	0.27	0.01	−4.62
2^1A_1	M1	CASCI	2.00	3.55	1.78	0.45	0.22	0.00	−5.39
3^1A_1	V1	CASCI	2.00	2.99	1.99	1.00	0.02	0.00	−4.88
		SS-CASSCF	2.00	3.00	1.95	1.00	0.05	0.00	−4.58
4^1A_1	V2	CASCI	2.00	3.93	1.19	0.07	0.01	0.81	−5.93
		SS-CASSCF	2.00	3.93	1.20	0.07	0.01	0.79	−5.48
5^1A_1	CT1	CASCI	2.00	2.03	1.97	0.18	1.80	0.01	3.84
		SS-CASSCF	2.00	2.16	1.90	0.28	1.65	0.01	3.14
6^1A_1	CT2	CASCI	2.00	2.80	1.20	0.37	1.59	0.01	3.93
		SS-CASSCF	2.00	2.64	1.36	0.29	1.76	0.02	3.48
7^1A_1	CT4	CASCI	2.00	2.97	1.35	0.97	0.70	0.00	3.66
		SS-CASSCF	2.00	3.00	1.20	1.00	0.82	0.00	3.16
8^1A_1	CT3	CASCI	2.00	2.90	1.34	0.64	1.11	0.00	2.33
		SS-CASSCF	2.00	2.88	1.31	0.70	1.11	0.00	2.18
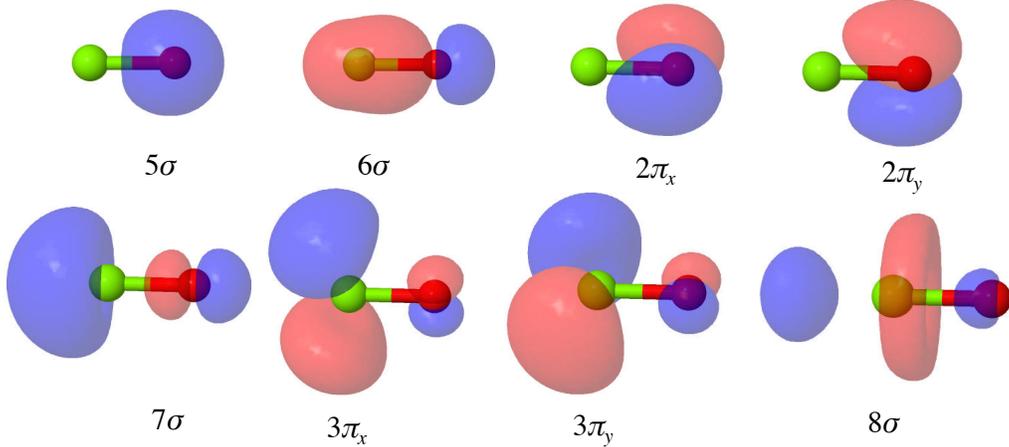
FIG. S1. Shapes of the eight LDA orbitals used to construct the active space for MgO. In each image, the Mg atom is on the left.