Toward a Reduced-Scaling Method for Calculating Coupled Cluster Response Properties

Ashutosh Kumar

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Literature and Technology

T. Daniel Crawford, Chair

Eduard Valeyev

Diego Troya

Alan Esker

May 9, 2018

Blacksburg, Virginia

Keywords: Coupled Cluster, Reduced-Scaling, Response Properties

Toward a Reduced-Scaling Method For Calculating Coupled Cluster Response Properties

Ashutosh Kumar

(ABSTRACT)

One of the central problems limiting the application of accurate *ab initio* methods to large molecular systems is their high computational costs, i.e., their computing and storage requirements exhibit polynomial scaling with the size of the system. For example, the coupled cluster method — the "gold standard" of quantum chemistry — scales as $\mathcal{O}(N^6)$, where N refers to the number of molecular orbitals (MOs).

This work was supported by a grant (CHE-1465149) from the U.S. National Science Foundation. Advanced Research Computing Center at Virginia Tech provided the necessary computational resources and technical support for all the calculations reported here.

Acknowledgments

Contents

List of Figures

List of Tables

Chapter 1

Introduction

One of the central problems limiting the application of accurate *ab initio* methods to large molecular systems is their high computational costs, i.e., their computing and storage requirements exhibit polynomial scaling with the size of the system. For example, the coupled cluster method — the "gold standard" of quantum chemistry — scales as $\mathcal{O}(N^6)$, where N refers to the number of molecular orbitals (MOs). Over the last half century, numerous techniques for the reduction of the size of the MO space have been introduced to to overcome this scaling obstacle and Löwdins introduction[1] of "natural orbitals" (NOs) as eigenvectors of the one-electron reduced density matrix (1RDM) in 1955 stands as one of the earliest works in this area. The frozen virtual NO (FVNO) scheme has been successfully applied within coupled cluster theory for calculating correlation energies, ionization potentials etc., where truncations of even up to 50% of the virtual space has been shown to introduce minimal errors.[2, **?**] In the FVNO approach, the NOs are obtained from the 1RDM of a less expen-

sive correlated method such as second-order Møller-Plesset (MP2) theory, and the virtual space is truncated based on the orbitals' corresponding occupation numbers.

My research in the Crawford group at Virginia Tech has primarily focused on the first successful extension of the FVNO scheme to calculate higher-order response properties like dynamic polarizabilities and specific optical rotations within the coupled cluster linear-response formalism. The conventional FVNO method performs poorly for these properties, yielding large errors which increase linearly (at best) with the number of frozen virtual orbitals.[**?**] My investigation revealed that the source of these errors is the removal of diffuse virtual orbitals — NOs that contribute little to the correlation energy (and thus have low occupation numbers), but are vital to the description of the wave function responses associated with polarizabilities and related properties. Indeed, applying the FVNO procedure only to the non-diffuse virtual space while retaining the diffuse MOs in the canonical basis itself greatly minimizes these errors. However, the number of such orbitals that must be retained could be very high for larger basis sets, thus severaly limiting this approach for large molecules.

Based on these observations, I have developed a technique which I call FVNO++, where instead of the ground state MP2-1RDM, a second-order perturbed density based on a limited number of iterations of the second-order coupled cluster (CC2) method is used to obtain perturbed "natural orbitals". A proof of concept study where we chose the leading contribution to the full second-order perturbed CCSD density as our 1RDM, produced less than 1% errors even after removing close to 50% of the virtual space in most of the test cases, thus validating this approach. The initial results obtained using guess densities based on

CC2 used in conjunction with corrections for the external truncated space have been very promising and more results are forthcoming. I will now extend this approach to the local pair natural orbital (LPNO) methods, where a separate 1RDM is defined for every occupied pair of MOs leading to separate non-orthogonal virtual spaces. The usual LPNO approach suffers from the same deficiencies as the FVNO method described above and hence performs poorly for such properties.[3]

Chapter 2

Modeling Optical Rotation

Chapter 3

Coupled Cluster Theory

3.0.1 Ground State

3.0.2 Response Theory

Sum of States

Linear Response

Chapter 4

Frozen Virtual Natural Orbitals for Coupled-Cluster Linear-Response Theory

4.1 Introduction

In the construction of many-body electronic wave functions, the scaling of a given method with the number of molecular orbitals (MOs) plays a pivotal role in the ultimate cost of the calculation. For many-body methods such as coupled cluster (CC),[4, 5, 6] which, in its canonical formulation, displays a higher-order polynomial dependence on the number of MOs, numerous mechanisms have been explored over the last half century for reducing

the size of the virtual-MO space. Among the earliest of these was Löwdin's[1] introduction in 1955 of the concept of "natural orbitals" (NOs) — orbitals for which the one-electron density matrix is diagonal. Löwdin demonstrated that NOs yield faster convergence of the configuration interaction (CI) wave function expansion than Hartree-Fock MOs. Some years later, Bender and Davidson[7] used natural orbitals in conjunction with CI (NO-CI) calculations to construct and analyze the most important configurations contributing to the correlated wave functions for a series of closed- and open-shell first-row diatomic hydrides. This work motivated Barr and Davidson a year later[8] to utilize only the virtual natural orbitals, obtained by diagonalization of the virtual-virtual block of the one-electron density matrix, for NO-CI calculations on the Ne atom.

The concept of pair natural orbitals (PNOs – originally called "pseudonatural orbitals") was developed by Edmiston and Krauss,[9] by Meyer[10], and by Ahlrichs and co-workers.[11] In the PNO approach, the virtual-virtual MO block of the one-electron density is constructed and diagonalized independently for each occupied MO pair, leading to separate non-orthogonal virtual spaces. Although this approach leads to rapid convergence of the correlation energy with respect to the size of the virtual space, it was little used following initial investigations until it was resurrected in recent years by Neese and co-workers with great success in the context of reduced-scaling electron correlation methods.[12, 13]

Following these pioneering efforts, NOs have been exploited in numerous applications to construct compact CI,[14, 15, 16] multiconfigurational self-consistent-field (MCSCF),[17] and CC wave functions.[18, 19, 20, 2, 21, 22] In many of these studies, the virtual-MO block

of the one-electron density is first obtained from a simpler model, such as second-order many-body perturbation theory (MBPT) calculation, and then diagonalized to yield the virtual-NO space. The space is then truncated based on an occupation-number-related criterion and fixed for the subsequent correlated-wave-function calculation. In addition, the final energy is commonly corrected using the second-order Møller-Plesset perturbation theory (MP2) correlation energy contributions arising from the external virtual space. These studies have indicated that, for energetics and related properties, even aggressive truncation of the virtual-NO space often has only a small impact on the resulting property as compared to full-space computations. For example, Landau *et al.*[2] found that, for NOs combined with the equation-of-motion coupled-cluster method for ionized states (EOM-IP-CC), reduction of the virtual space by up to 70% yielded truncation errors within ca. 1 kcal/mol for ionization energies of organic compounds and non-parallelity errors in potential surfaces of weakly bound complexes.

While the above studies have demonstrated clearly the usefulness of NOs for aggressively truncating the virtual space when computing correlation energies, more complex properties have yet to be considered. As shown in a number of recent reports,[23, 24, 25, 26, 27, 3] properties that are related to the linear- or higher-order response of the wave function to external electric and magnetic fields, for example, exhibit much greater sensitivity to the quality of the wave function than simple energetics. In particular, local correlation methods have been shown[23, 24, 25, 26, 3] to require significantly larger domains for properties such as polarizabilities than for ground-state energies. Furthermore, the many-body expansion

— which has yielded impressive convergence for energetics and dipole moments for clusters of weakly interacting molecules (such as a solute embedded in an explicit solvent) — converges erratically, at best, for spectroscopic response properties due to its strong basis-set dependence.[28] In order to account for this, the ability to reduce the dimensionality of the virtual space becomes paramount. Thus, the focus of the present work is on the extension of the NO approach to linear-response properties, especially the case of frequency-dependent dipole polarizabilities.

4.2 Theoretical Background

4.2.1 Frozen Virtual Natural Orbitals

The MP2 unrelaxed one-electron density matrix can be written in terms of spin orbitals as:

$$\gamma_{pq} = \langle \Psi^{(1)} | \{ a_p^\dagger a_q \} | \Psi^{(1)} \rangle, \tag{4.1}$$

where $|\Psi^{(1)}\rangle$ is the first order correction to the Hartree-Fock wave function,

$$|\Psi^{(1)}\rangle = \frac{1}{4} \sum_{ijab} t_{ij}^{ab} |\Phi_{ij}^{ab}\rangle \tag{4.2}$$

and

$$t_{ij}^{ab} = \frac{\langle ij||ab\rangle}{.pdfilon_i + \epsilon_j - .pdfilon_a - \epsilon_b}. \tag{4.3}$$

Here, $\langle ij||ab\rangle$ is an antisymmetrized two-electron integral in Dirac's notation and $.pdfilon_i, \epsilon_a ...$ refer to the Hartree-Fock molecular orbital energies. We use the indices $i, j, k, ...$ to indicate

occupied orbitals while $a, b, c, ...$ denote virtual orbitals. $|\Phi_{ij}^{ab}\rangle$ refers to a doubly-excited determinant where occupied orbitals i and j are replaced by virtuals a and b respectively. The brackets around the second-quantized operators in Eq. (6.1) indicate normal ordering with respect to the reference wave function.

In the MP2 based NO method, the virtual-virtual block of γ_{pq} is constructed,

$$\gamma_{ab} = \frac{1}{2} \sum_{ijc} t_{ij}^{ac} t_{ij}^{bc}, \tag{4.4}$$

and then diagonalized,

$$\boldsymbol{\gamma V} = \boldsymbol{nV}. \tag{4.5}$$

The eigenvectors, $\boldsymbol{V}$, are the virtual natural orbitals (NOs), and the eigenvalues, $\boldsymbol{n}$, are the associated occupation numbers. As noted earlier, the wave function amplitudes contain significantly greater sparsity when represented in the NO basis than the original canonical MO basis; orbitals with lower occupation numbers yield $\hat{T}_2$ amplitudes with smaller magnitudes and concomitantly smaller contributions to the correlation energy. Thus, orbitals with occupation numbers below a selected threshold can be removed without introduction of significant errors, leading to reduced computational cost. The Hartree-Fock virtual molecular orbitals and associated integrals are then transformed to this truncated NO basis, followed by semicanonicalization of the virtual-virtual block of the Fock matrix, for subsequent computations using higher-order correlation methods such as CC theory. In most cases, the final correlation energy in the truncated virtual space is corrected using the MP2 energy in the external (non-truncated) NO space to minimize the resulting errors, as described below.

4.2.2 Coupled Cluster Response Theory

Dynamic response functions representing higher-order properties such as polarizabilities and hyperpolarizabilities, optical activity tensors, magnetizabilities, etc. can be obtained by expanding the expectation value of an appropriate time-independent operator in perturbational orders with respect to a time-dependent external field. The CC form of response theory has been routinely used for many years for accurate calculations of such properties.[29] The CC linear response function for property operators $\boldsymbol{A}$ and $\boldsymbol{B}$, for example, can be written as

$$\langle\langle \boldsymbol{A}; \boldsymbol{B} \rangle\rangle_\omega = \frac{1}{2}\hat{C}^{\pm\omega}\hat{P}[A(-\omega), B(+\omega)] \left[\langle \Psi_0 | (1 + \hat{\Lambda}) \left([\bar{A}, \hat{X}^B_\omega] + \frac{1}{2}[[\bar{H}, \hat{X}^A_{-\omega}], \hat{X}^B_\omega] \right) | \Psi_0 \rangle \right] \tag{4.6}$$

where Ψ_0 is the reference wavefunction, $\hat{\Lambda}$ is a de-excitation operator used to parametrize the CC left hand wavefunction, ω is the frequency of the external field, $\hat{C}$ is a symmetrizer that simultaneously interchanges the sign of the field frequency and takes the complex conjugate of the expression, and $\hat{P}$ symmetrizes the expression with respect to the operators $\boldsymbol{A}$ and $\boldsymbol{B}$. Operators with an overbar have been similarity transformed with the ground state cluster operators, $\hat{T}$, e.g. $\bar{H} = e^{-\hat{T}}\hat{H}e^{\hat{T}}$. The first-order (right-hand) perturbed wave function is represented by $\hat{X}^B_\omega$, whose amplitudes can be obtained by solving a set of appropriate linear equations, e.g.,

$$\langle \Phi^{ab\cdots}_{ij\cdots} | (\bar{H} - \omega) | \hat{X}^B_\omega | \Psi_0 \rangle = -\langle \Phi^{ab\cdots}_{ij\cdots} | \bar{B} | \Psi_0 \rangle. \tag{4.7}$$

In the case of dynamic polarizabilities, for example, both $\boldsymbol{A}$ and $\boldsymbol{B}$ are cartesian components of the electric dipole operator, $\boldsymbol{\mu} = -\boldsymbol{r}$, and the isotropic polarizability α_ω is related to the

trace of the polarizability tensor such that Eq. (6.6) reduces to the following form (for real wavefunctions),

$$\alpha_\omega = \frac{1}{3}\mathrm{Tr}\left[\langle\Psi_0|(1+\hat{\Lambda})\left(\left[\bar{\mu},(\hat{X}^\mu_\omega+\hat{X}^\mu_{-\omega})\right]+\left[\left[\bar{H},\hat{X}^\mu_{-\omega}\right],\hat{X}^\mu_\omega\right]\right)|\Psi_0\rangle\right] \tag{4.8}$$

For dynamic polarizabilities computed using the coupled cluster singles and doubles (CCSD) method, for example, six sets of perturbed amplitudes, $\hat{X}_1(X_i^a)$ and $\hat{X}_2(X_{ij}^{ab})$, must be computed, one for each cartesian component of $\boldsymbol{\mu}$ at both positive and negative frequencies.

4.3 Computational Details

The primary molecular test cases of this work is hydrogen peroxide, H_2O_2, though additional tests are reported for related species such as (*P*)-dimethylallene and (*S*)-methyloxirane, as well as those same molecules interacting with one or more water molecules. All molecular structures were optimized using the B3LYP functional[30, 31, 32] in the aug-cc-pVDZ (aDZ) basis.[33, 34, 35] (Coordinates of all structures are provided in Tables S1-S5 of the Supporting Information.) Frequency-dependent polarizabilities were computed at the coupled cluster singles and doubles (CCSD) level of theory[36] using a linear-response formulation.[37] While the aDZ basis set was used for most test calculations, the larger aug-cc-pVTZ (aTZ) and aug-cc-pVQZ (aQZ) basis sets were also employed for selected analyses.[34] All orbitals were active in the correlated models utilized here, and all coupled-cluster calculations were carried out using the PSI4 open-source quantum chemistry package.[38] In all calculations in which virtual orbitals are truncated, the same orbital space is used in the solution of the

unperturbed and perturbed amplitudes equations, just as is required for locally correlated property calculations.[39] We note that the size-extensivity of the polarizability is unaffected by the truncation of the virtual space.

4.4 Results and Discussion

It is well known that deletion of higher-energy canonical Hartree-Fock MOs (CMOs) typically leads to significant errors in the recovery of electron correlation energies, whereas, by design, the truncation of virtual NOs with low occupation numbers results in little loss of accuracy.[2, 19, 20] For example, Fig. 4.1 plots the error in the CCSD correlation energy of H_2O_2 in the

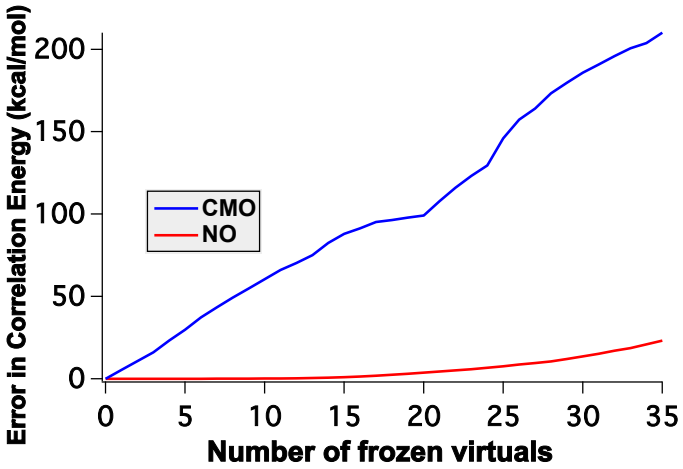

Figure 4.1: Error in the CCSD energy of H_2O_2 in kcal/mol as a function of the number of frozen virtual orbitals in both CMO and NO bases.

aDZ basis as a function of the number of frozen virtual CMOs (removed starting from the highest energy orbitals) or NOs (starting from the lowest occupation numbers). Clearly the correlation energy is very sensitive towards the removal of CMOs, with the error increasing by more than 4 kcal/mol after the deletion of even one virtual orbital. On the other hand, removal of low occupation-number NOs introduces errors of only ca. 2.5 kcal/mol even when up to 33% of the virtual space (18 of 55 orbitals) is truncated. The errors in the NO basis can be further minimized by employing an MP2 energy correction,

$$E_{\text{corr}}^{\text{MP2}} = \frac{1}{4}\sum_{ij}\sum_{ab\in\text{ext}}\frac{|\langle ij||ab\rangle|^2}{.pdfilon_i + \epsilon_j - \epsilon_a - \epsilon_b}. \tag{4.9}$$

where the summation over virtual orbitals is limited to NOs in the external, truncated space. Fig. 4.2 plots the error in the CCSD correlation energy for the same system as above in the

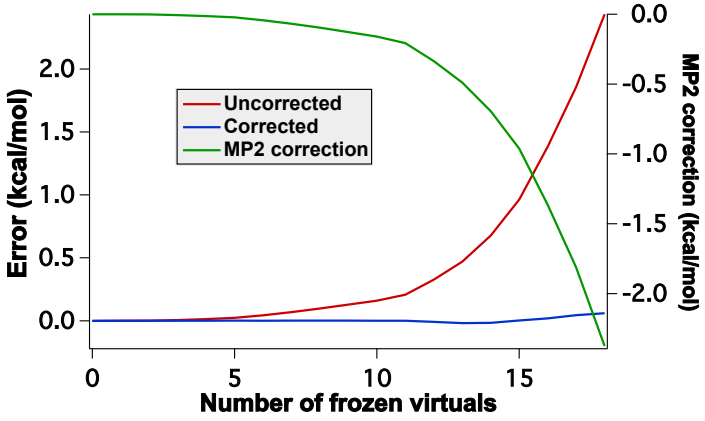

Figure 4.2: Error in CCSD energy of H_2O_2 in the NO bases, with and without MP2 corrections and MP2 correction as a function of the number of frozen virtual orbitals.

NO basis, with and without this correction (left-hand vertical axis), as well as the correction itself (right-hand axis). After employing the correction, the error in the correlation energy falls to less than 0.1 kcal/mol when 1/3 of the virtual space is eliminated. Similar results are obtained for aTZ and aQZ basis sets, where the truncation errors are less than 1 kcal/mol even after the removal of 50% of the virtual space. We note that the MP2 energy correction is significant (on the order of mE_h or several kcal/mol) even for small compounds such as H_2O_2. The correction scales linearly with the number of electrons and thus is a critical component for the success of the frozen virtual NO approach for larger molecules.

4.4.1 Frozen Virtual Orbitals and Response Properties

What is the impact of freezing virtual orbitals — whether CMO or NO — on higher-order properties? Fig. 4.3 plots errors in dynamic polarizabilites (computed at a wavelength of 589 nm) as a function of the number of virtual orbitals deleted at the CCSD/aDZ level of theory for the same H_2O_2 test case as above. Comparison to Fig. 4.1 reveals precisely the opposite behavior for polarizabilities as for correlation energies, *viz.*, truncation of the CMO virtual space induces much smaller errors than does that of the NO virtual space. For the latter, errors increase approximately linearly with the number of frozen virtual NOs. On the other hand, for the CMO basis, the error increases slowly to a maximum of 1.9% when 13 virtual orbitals are removed and then decreases to only -0.3% when as many as 27 orbitals (ca. 50% of the virtual space) are frozen. This trend is not unique to H_2O_2 or the aDZ basis set. As shown in Fig. S1 of the Supporting Information, the same behavior is observed

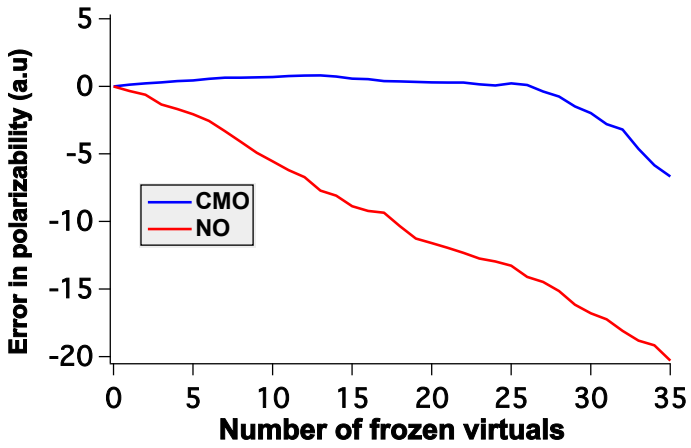

Figure 4.3: Errors in the CCSD/aDZ dynamic polarizability (589 nm) of H_2O_2 in in both CMO and NO bases as a function of number of virtual orbitals removed .

for other molecules such as dimethylallene, methyloxirane, and such compounds interacting with explicit solvent molecules. In addition, Figs. S2 and S3 report the same trends for H_2O_2 with the larger aTZ and aQZ basis sets.

What is the source of this unexpected behavior? It is well known that diffuse basis sets are essential for the accurate descriptions of a variety of response properties, such as dipole polarizabilities.[35] Fig. 4.4 plots the spatial extent — $\langle r^2 \rangle$ — for each virtual CMO or NO in the same ordering as they are deleted in Fig. 4.3. The Figure clearly shows that the earliest NOs to be removed (the ones with the lowest occupation numbers) are the most diffuse, *i.e.*, they should contribute substantially to the description of the dynamic polarizability. On the other hand, the first CMOs to be frozen (those with the highest orbital energies) are also the

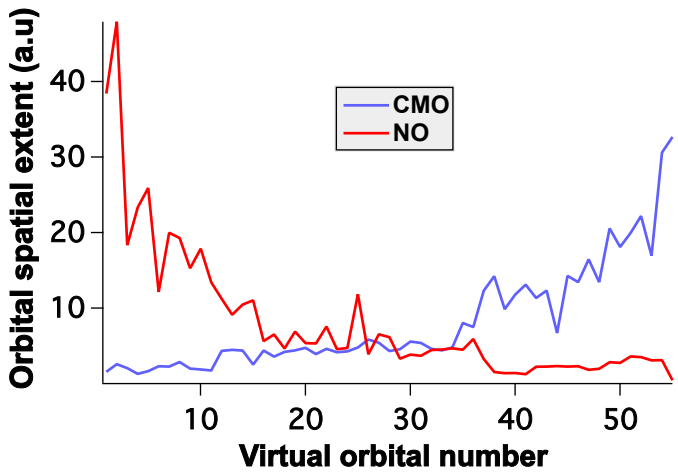

Figure 4.4: Spatial extent ($\langle r^2 \rangle$) of virtual orbitals of H_2O_2 in both CMO and NO bases. Orbitals are ordered left-to-right by decreasing energy (CMOs) or increasing occupation number (NOs).

most compact and thus contribute the least to this property. Given that highly diffuse basis functions typically contribute primarily to CMOs with low orbital energies — often below that of what is normally considered the true anti-bonding "LUMO" — the latter result is not surprising; these diffuse CMOs appear to the far right of Fig. 4.4 and are thus never deleted, leading to the good behavior of the CMO truncation in Fig. 4.3. These same diffuse basis functions, however, contribute little to the description of dynamical correlation effects, and thus exhibit very low occupation numbers upon transformation to the NO virtual space. Thus, they are truncated first in the NO basis, yielding the large errors in the polarizability depicted in Fig. 4.3.

The above observations suggest that, for computing response properties such as dynamic

polarizabilities, an alternative approach to truncation of the virtual space is to order the orbitals by increasing values of $\langle r^2 \rangle$ rather than by decreasing orbital energies (as is done for CMOs) or increasing occupation numbers (for NOs). Fig. 4.5 plots errors in the CCSD/aDZ

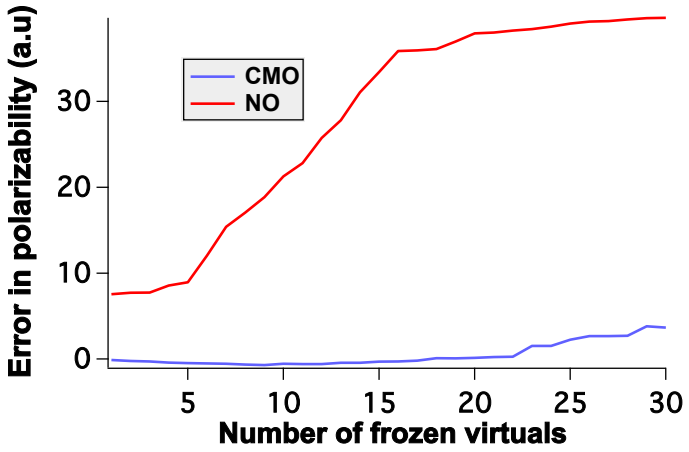

Figure 4.5: Spatial extent ($\langle r^2 \rangle$) of virtual orbitals of H_2O_2 in both CMO and NO bases. Orbitals are ordered left-to-right by decreasing energy (CMOs) or increasing occupation number (NOs).

dynamic polarizability of H_2O_2 as the virtual CMOs or NOs are removed in order of increasing spatial extent. While the errors are comparable to that observed in Fig. 4.3 for the CMO truncation, the behavior associated with removal of virtual NOs is significantly different. First, the polarizability errors exhibit two plateaus: one associated with diffuse NOs 2-5 and another with NOs 16-30, and removal of these NOs has little impact on the observed error. However, deletion of the virtual NO with the *smallest* spatial extent unexpectedly leads to a large initial error (ca. 7.5 a.u.), followed later by a linearly increasing error as virtual NOs

6-15 are removed. Clearly, spatial extent is not the only criterion by which we may predict the importance of a given virtual NO to the polarizability.

Another possible source of error is the lack of orbital response in the chosen formulation of the coupled cluster linear response function.[40] In the infinite-lifetime approximation, frequency-dependent properties such as dipole polarizabilities exhibit first-order poles at the excitation frequencies. In the coupled cluster formulation described above, the orbital response to the external field is typically neglected so that these poles correspond solely to the response for the correlated wave function, and additional, spurious poles arising due to the Hartree-Fock reference determinant will not appear. (This approximation is typically justified based on the fact that much of the orbital-response effects are accounted for by the singles amplitudes.[41]) In order to test whether the orbital relaxation significantly impacts the behavior of the computed polarizability as the virtual space is reduced, we have computed *static* ($\omega = 0.0$) polarizabilities using finite-differences (with a a central-difference formula with a differential field strength of 0.001 a.u.) The errors in the CCSD/aDZ static polarizability of H_2O_2 for both CMO and NO virtual spaces are reported in Fig. 4.6. Interestingly, with orbital relaxation included, the truncation of the NO space now becomes better behaved than the CMO space over a large domain of orbitals removed. Unfortunately, we cannot take advantage of this improvement in conjunction with Hartree-Fock orbitals without corrupting the pole structure of the response function. A Brueckner or orbital-optimized approach may prove superior in this regard, though further investigation is warranted.[40, 42]

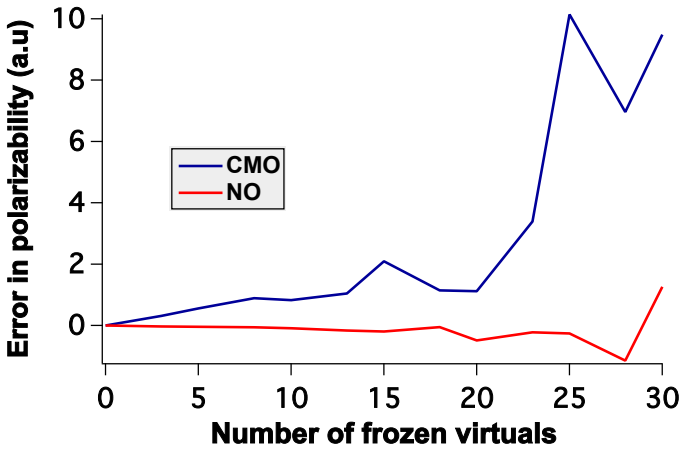

Figure 4.6: Errors in the CCSD/aDZ static polarizability (including orbital relaxation effects) of H_2O_2 in in both CMO and NO bases as a function of number of virtual orbitals removed.

4.4.2 Wave Function Truncation in the Virtual-Orbital Space

For additional insight into the above observations, we examine errors arising in dynamic polarizabilities as a function of truncation of specific wavefunction parameters in either the CMO or NO basis. Fig. 4.7 plots the errors in CCSD/aDZ dynamic polarizabilities of H_2O_2 as a result of truncations of the unperturbed ground-state cluster amplitudes $\hat{T}$ and $\hat{\Lambda}$, as well as perturbed amplitudes, $\hat{X}^{\mu}_{\omega}$, represented in the CMO basis. Note that, in this analysis, only the specified amplitudes associated with the selected CMOs are forced to zero; the CMOs remain active for all other wave function components. From the Figure, it can be clearly seen that removing $\hat{T}_1$ alone does not introduce any significant error in the polarizability, whereas truncating $\hat{T}_2$ amplitudes results in substantial positive errors which increase almost linearly

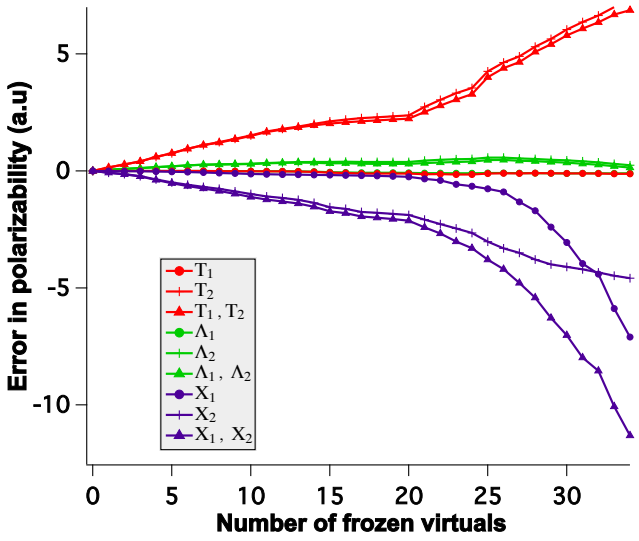

Figure 4.7: Errors introduced in CCSD/aDZ polarizabilities of H_2O_2 in the virtual CMO bases by the truncation of different classes of wave function amplitudes.

with the number of virtual CMOs. (Not surprisingly, freezing both $\hat{T}_1$ and $\hat{T}_2$ amplitudes together have essentially the same effect as freezing $\hat{T}_2$ amplitudes alone.) Alternatively, for the left-hand wave function, removing $\hat{\Lambda}_1$ and $\hat{\Lambda}_2$ amplitudes either separately or pairwise seem to have negligible impact. In the case of the perturbed amplitudes, only small (negative) errors are introduced even after freezing all $\hat{X}_1$ amplitudes involving almost 23 virtual CMOs, but the error then rises sharply with further truncation. On the other hand, the negative errors due to truncation of $\hat{X}_2$ amplitudes are significant from the beginning and increase almost linearly. Thus the error due to removal of both $\hat{X}_1$ and $\hat{X}_2$ amplitudes belonging to the first 23 CMOs is due to elimination of $\hat{X}_2$ amplitudes alone, whereas beyond that limit,

the total error corresponds to the sum of errors from $\hat{X}_1$ and $\hat{X}_2$ truncation.

A key observation is that, within the domain of the first 23 virtual CMOs, the positive errors introduced by truncation of $\hat{T}_2$ is cancelled almost exactly by the negative errors arising from the truncation of $\hat{X}_2$. This is further illustrated in Fig. **??**, which plots on a narrow range the

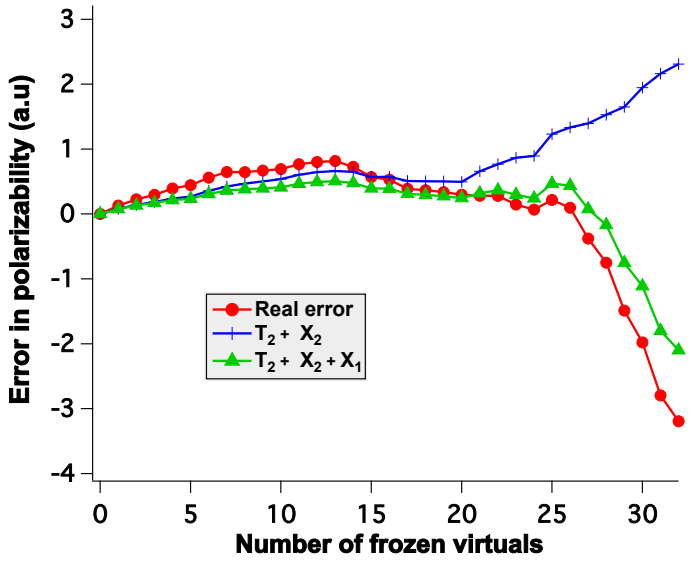

Figure 4.8: Errors introduced in CCSD/aDZ polarizabilities of H_2O_2 in the virtual CMO bases by the truncation of specific classes of wave function amplitudes as compared to the total errors obtained by freezing of virtual CMOs.

errors in the polarizability associated with truncating specific classes of $\hat{T}_2$ and $\hat{X}_2$ amplitudes against the total errors obtained by freezing CMOs entirely (for all amplitudes). Outside of this domain, errors associated with neglect of $\hat{X}_1$ amplitudes become dominant, leading to

the accumulation of negative total errors observed in Fig. 4.3. Thus, the apparently robust performance of the truncation of the virtual CMO space arises, in fact, from offsetting errors. Similarly to the above analysis for the virtual CMO space, Fig. 4.9 reports errors in the

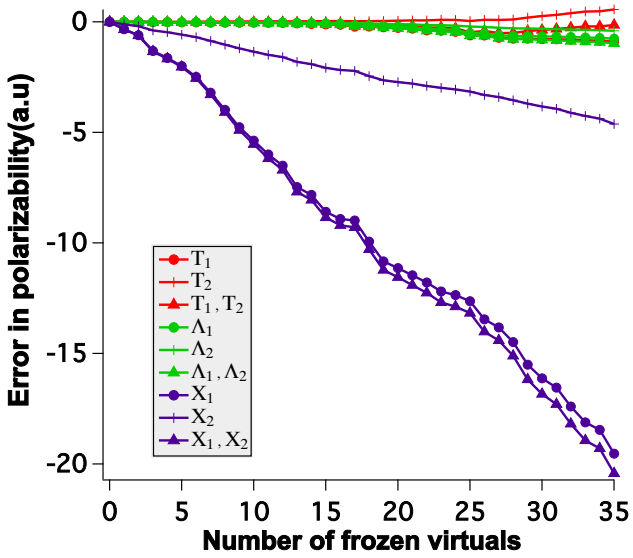

Figure 4.9: Errors introduced in CCSD/aDZ polarizabilities of H_2O_2 in the virtual NO bases by the truncation of different classes of wave function amplitudes.

CCSD/aDZ polarizability of H_2O_2 introduced by the neglect of various classes of wave function amplitudes associated with selected virtual NOs. We observe first that, unlike the CMO case, neglecting $\hat{T}_2$ amplitudes associated with particular virtual NOs has no significant effect on the error. This behavior is expected, because the $\hat{T}_2$ amplitudes are, by construction, sparse in the virtual NO basis such that orbitals with low occupation numbers are associated

with $\hat{T}_2$ amplitudes of smaller magnitude. Furthermore, just as in the CMO case, the removal of $\hat{T}_1$, $\hat{\Lambda}_1$ and $\hat{\Lambda}_2$ amplitudes introduces only small errors, while the removal of selected $\hat{X}_2$ amplitudes based on NOs yields significant negative errors that increase linearly with the number of virtual NOs truncated.

However, unlike the virtual CMO case, neglecting $\hat{X}_1$ amplitudes corresponding to specific virtual NOs introduces large negative errors in the polarizability even from the first NO removed, and the total error obtained by truncating both $\hat{X}_1$ and $\hat{X}_2$ amplitudes is almost the same as the error due to truncation of the $\hat{X}_1$ amplitudes alone. Indeed, the greatest contribution ($> 90\%$) to the total polarizability errors arises from the perturbed singles amplitudes.

The significance of the $\hat{X}_1$ amplitudes is evident upon analysis of their leading-order contribution to the polarizability [cf. Eq. (6.8)],

$$\alpha_\omega \leftarrow \frac{1}{3}\sum_{ia}\sum_{x}\mu_{ia}^{x}(X_{ia}^{x}(\omega)+X_{ia}^{x}(-\omega)), \tag{4.10}$$

where μ_{ia}^{x} is an element of the occupied-virtual block of the Cartesian component, x, of the electric-dipole moment integral matrix, and the inner sum runs over all such components. The singles themselves are obtained from the corresponding form of Eq. (6.7),

$$\langle\Psi_i^a|X_\omega^\mu|\Psi_0\rangle = -\sum_\nu \langle\Psi_i^a|(\bar{H}-\omega)^{-1}|\nu\rangle\langle\nu|\bar{\mu}|\Psi_0\rangle, \quad \nu\in\{\Psi_j^b,\Psi_{jk}^{cd}\}, \tag{4.11}$$

where $\bar{\mu}$ is the similarity transformed electric-dipole operator,

$$\bar{\mu} = \hat{\mu} + \Big[\hat{\mu},\hat{T}\Big] + \frac{1}{2}\Big[\Big[\hat{\mu},\hat{T}\Big],\hat{T}\Big]. \tag{4.12}$$

The corresponding leading-order contribution to the $\hat{X}_1$ amplitudes themselves is

$$X_{ia}^{\mu}(\omega) \leftarrow \frac{\mu_{ia}}{\overline{H}_{ii} - \overline{H}_{aa} + \omega}, \tag{4.13}$$

where the largest contribution to the diagonal elements of the similarity transformed Hamiltonian, $\bar{H}$, arises from the orbital energies (more precisely, the diagonal Fock matrix elements expressed in the CMO or NO basis), which are plotted in Fig. 4.10. While the values for the

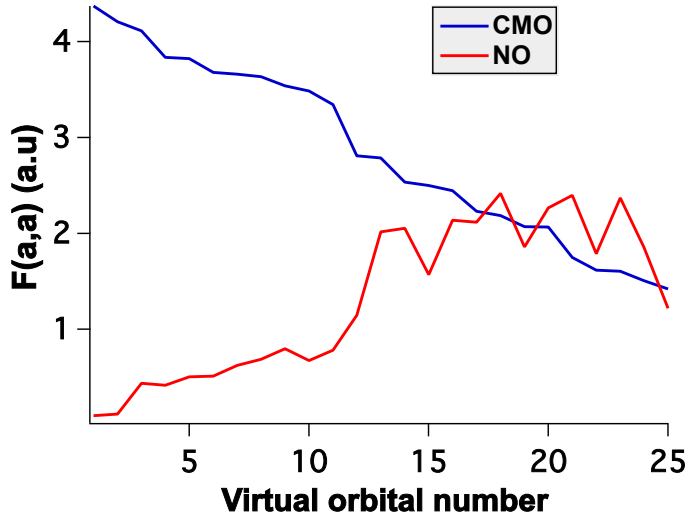

Figure 4.10: Virtual diagonal elements (a.u.) of the Fock matrix in the CMO and NO bases.

virtual CMOs decrease steadily, their NO counterparts actually *increase* and display greater oscillation. Clearly, the diagonal elements of the fock matrix for virtual NOs are significantly smaller in magnitude than the corresponding CMOs, which concomitantly increases the values of the $\hat{X}_1$ amplitudes associated with such NOs. This effect can be seen in Fig. 4.11 which plots the sum of the absolute values of $\hat{X}_1$ amplitudes for a given virtual orbital, i.e.

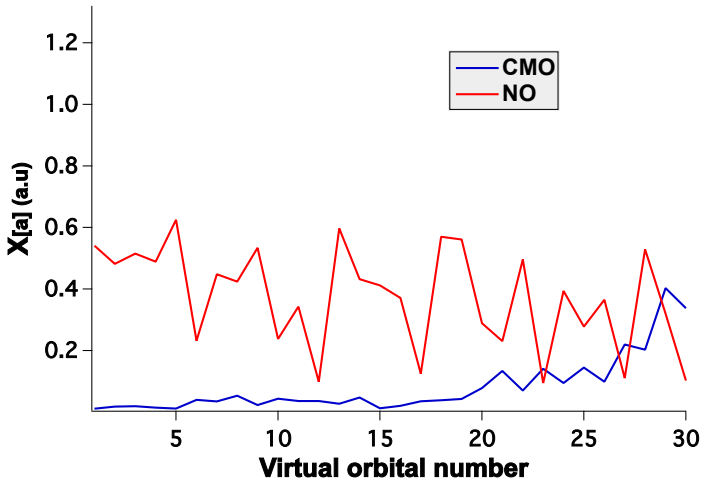

Figure 4.11: Sum of the absolute values of $\hat{X}_1$ amplitudes for a given virtual, $\sum_i |X_i^a|$, for perturbation μ_x and frequency 589 nm, plotted for each virtual NO or CMO.

$\sum_i |X_i^a|$, for both virtual CMOs and NOs. Thus, the sparsity of the $\hat{X}_1$ amplitudes present in the CMO basis is almost completely lost in the NO basis, which leads to the large errors in dynamic polarizabilities due to the truncation of $\hat{X}_1$ amplitudes in the NO basis observed earlier.

Furthermore, the dependence of the $\hat{X}_1$ amplitudes on both $\hat{T}_2$ and $\hat{X}_2$ plays a significant role in the overall *sign* of the polarizability error. This point is clearly illustrated in Fig. 4.12, which plots the 2-norm of the $\hat{X}_1$ vector as $\hat{T}_2$ or $\hat{X}_2$ amplitudes associated with a given virtual CMO is neglected (in the same manner as used in Figs. 4.7 and 4.8). As the $\hat{T}_2$ amplitudes associated with a given virtual orbital are removed, the norm of the $\hat{X}_1$ vector

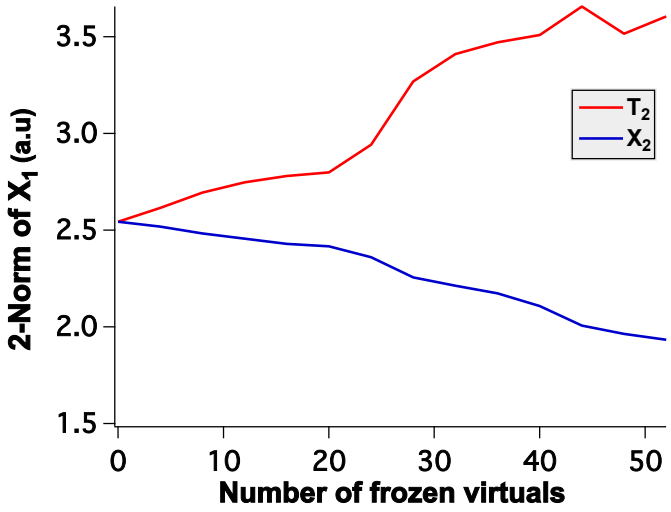

Figure 4.12: The 2-norm of the $\hat{X}_1$ amplitude vector in the CMO bases as a function of the truncation of classes of unperturbed $\hat{T}_2$ and perturbed $\hat{X}_2$ amplitudes.

increases, leading to the positive errors in the polarizability observed in Fig. 4.7. On the other hand, removal of $\hat{X}_2$ amplitudes leads to a decrease in the norm of the $\hat{X}_1$ vector and the negative errors in the polarizability appearing in Fig. 4.7.

4.4.3 External-Space Corrections

As noted earlier, a key aspect of the strong performance of virtual NOs for correlation energies is the use of MP2-based corrections for the contributions of the truncated or "external" virtual space, as given in Eq. (6.9). When considering a similar correction for properties, however, we have explored three options: time-dependent Hartree-Fock (TDHF) and both

static and dynamic CC2[41] corrections. In each case, we have used the same MP2-based virtual NO space and computed the correction as the difference between the full-virtual space polarizability and the truncated virtual-NO polarizability, with the CCSD/aDZ results for polarizabilties of H_2O_2 shown in Fig. 4.13. The CC2-based corrections recover nearly all of

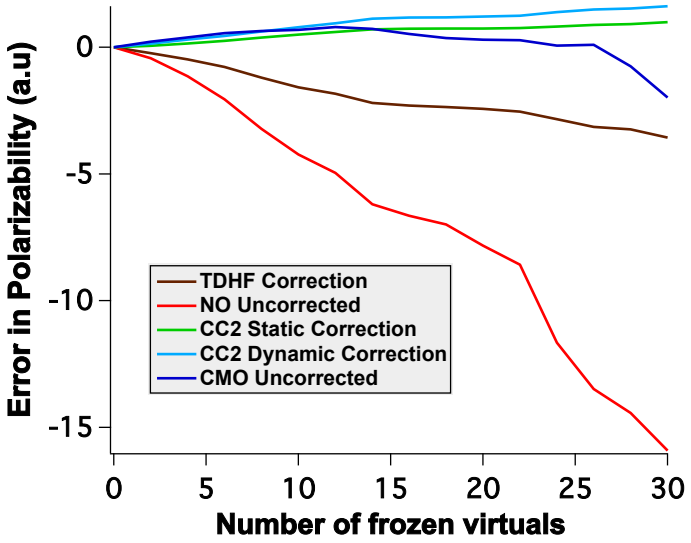

Figure 4.13: Correction schemes for the external truncated NO space for the CCSD/aDZ polarizabilities of H_2O_2.

the error associated with the uncorrected virtual NO space across a wide range of truncation, with the static correction actually yielding slightly smaller errors than its dynamic counterpart. The less expensive TDHF corrections offer significant improvement over the original virtual NO results, and recover most (ca. 85%) of the error until about 40% (21 orbitals) of the virtual space has been deleted, but it clearly produces overall larger errors than the

correlated methods. In addition, the TDHF corrections are potentially problematic because the pole structure of the polarizability naturally follows that of the underlying Hartree-Fock perturbed orbitals rather than that of the correlated wave functions,[43, 44, 45] a criticism that would also hold for purely MP2-based corrections. On the other hand, the CC2-based corrections are significantly more expensive than TDHF, in part because of their iterative nature and the need to transform two-electron integrals involving three virtual orbitals (with the latter criticism again holding for an MP2-based correction).

4.4.4 Perturbed Natural Orbitals

Given that the purpose of the NOs is to "focus" the important components of the basis for the description of electron correlation effects into a compact space, without consideration of the importance of the basis set for other properties, an alternative approach might be to build a virtual space that explicitly takes such properties into account, such as the inclusion of the perturbed one-electron density in the definition of the space. Instead of diagonalizing the ground-state MP2 density, we diagonalize its gradient with respect to the external electric field, thereby incorporating the effects of the external perturbations. The resulting "occupation numbers" obtained thus carry information about orbital occupancies in the perturbed states. The x^{th} Cartesian component of the perturbed density matrix can be written in terms of spin orbitals as

$$\gamma^x_{ab} = \frac{1}{2}\sum_{ijc} P_+(a,b)\left[\frac{t^{bc}_{ij}}{D^{ac}_{ij}}\left(P_-(a,c)\sum_d t^{ad}_{ij}\mu^x_{cd} - P_-(i,j)\sum_m t^{ac}_{im}\mu^x_{mj}\right)\right] \tag{4.14}$$

where P_+ and P_- are symmetric and anti-symmetric permutation operators, respectively, and the orbital-energy denominator $D_{ij}^{ac} = .pdfilon_i + \epsilon_j - \epsilon_a - \epsilon_c$. (For computational convenience, orbital relaxation terms have been neglected.) To obtain the perturbed NOs and their corresponding eigenvalues, we take an average of the three cartesian components of the density and diagonalize the result. The eigenvalues consist of both positive and negative values as the perturbed density matrix is not positive definite, and thus we truncate the orbitals based on the absolute values of these eigenvalues. Fig. 4.14 compares the

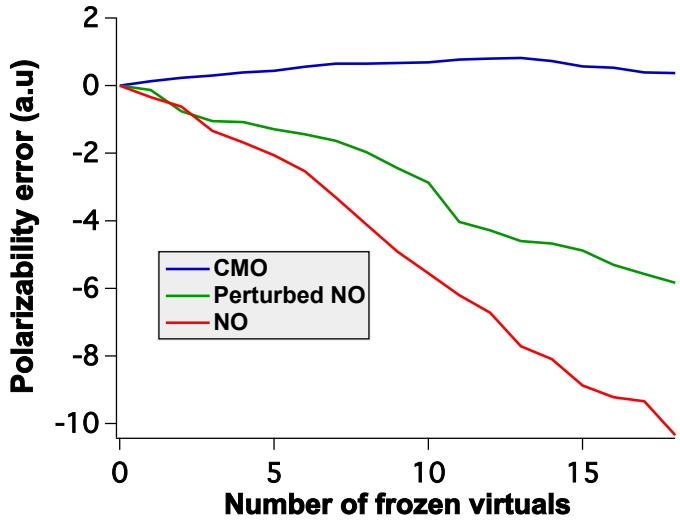

Figure 4.14: Errors introduced in CCSD/aDZ polarizabilities of H_2O_2 in the virtual CMO and NO bases, as well as the perturbed virtual NO basis as a function of number of virtual orbitals removed.

performance of these perturbed NOs with that of the CMOs and NOs, where the error in the CCSD/aDZ dynamic polarizability of H_2O_2 is plotted as a function of the number of

virtual orbitals removed. While the perturbed NOs lower the truncation errors associated with conventional NOs, they still introduce significantly higher errors than the corresponding CMOs. The reason for this underperformance is related to the definition of the perturbed density, which naturally bears strong similarity to its unperturbed counterpart [cf. Eqs. (6.4) and (6.14).] As a result, similar sparsity and orbital energy issues illustrated in Figs. 4.10 and 4.11 arise for the perturbed density just as for the original NOs, indicating that such an approach does not resolve the problem.

4.4.5 The Dipole-Amplitude Criterion

The above results demonstrate that the CMO basis provides the best performance among the various virtual spaces considered here, albeit based on cancellation of errors, but what criterion should be used to determine the truncation level that provides optimal balance between computational cost (most compact virtual space) and accuracy? The CMO orbital energies are one possibility, but they have no direct connection to the properties in question. Another alternative is the "dipole amplitude", d_a, which is defined for each virtual CMO as

$$d_a \equiv \sum_x \sum_i \frac{(\mu_{ia}^x)^2}{.pdfilon_i - \epsilon_a}. \tag{4.15}$$

This expression, which is trivially computed post-Hartree-Fock, is based on Eqs. (6.10) and (6.13), using the fact that the leading contributions to the diagonal elements of the similarty-transformed Hamiltonian are the orbital energies. As we are constructing a phenomenological truncation criterion, we have chosen to neglect the dependence on the field

frequency. Fig. 4.15 plots the values of the dipole amplitude for each CMO, as well as the

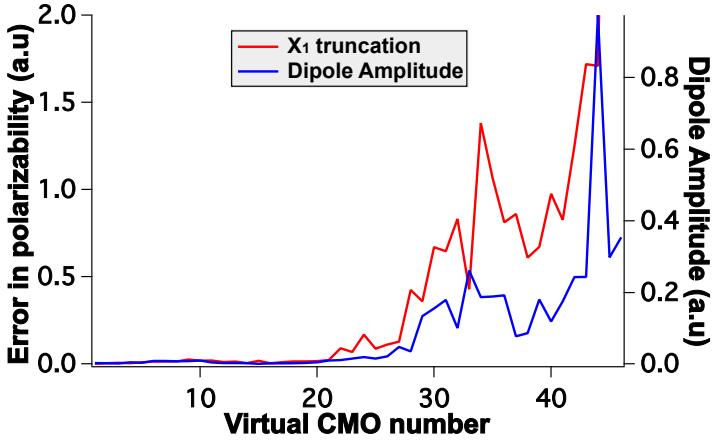

Figure 4.15: Absolute errors introduced in CCSD/aDZ polarizabilities of H_2O_2 due to truncation of $\hat{X}_1$ amplitudes and dipole amplitudes plotted as a function of different virtual CMOs.

corresponding error in the CCSD/aDZ dipole-polarizability of H_2O_2 introduced by deleting the $\hat{X}_1$ amplitude associated with a given virtual CMO. There is a clear correlation between the two functions, and one can see that as both the error due to truncation of $\hat{X}_1$ and the value of the dipole length increase sharply once we reach CMOs 27-28 — ca. 50% of the virtual space for this test case. Accordingly, one should stop the truncation of the virtual CMO space once the dipole length values start rising sharply. While the optimal choice of such a threshold will be addressed systematically in subsequent work, our preliminary analyses using H_2O_2, methyloxirane, dimethylallene, and related compounds suggest a cutoff of ca. 3-3.5% yields minimal errors in the polarizability.

4.5 Conclusions

On the basis of the above findings, we conclude that, in the absence of orbital relaxation, virtual NOs are not suited for higher-order property calculations such as dynamic polarizabilities, and that the occupation number is not an acceptable criterion for estimating the importance of a virtual orbital for such calculations. Although the use of external space corrections based on CC2 polarizabilities reduces the observed truncation errors, they are too relatively costly to be practical for large molecular systems. Furthermore, the use of perturbed virtual NOs offers only slight improvement as compared to unperturbed virtual NOs. CMOs, on the other hand, provide a much more stable mechanism for reducing the size of the virtual space — with truncation of up to 50% of the orbitals yielding shifts of less than 2% as compared to full-space calculation – but the source of their success lies in a significant cancellation of errors. Although further systematic studies are needed, the dipole amplitude appears to provide a useful threshold for an *a priori* truncation of the CMO virtual space.

4.6 Acknowledgements

This research was supported by a grant (CHE-1465149) from the U.S. National Science Foundation. The authors acknowledge Advanced Research Computing at Virginia Tech for providing computational resources and technical support that have contributed to the results reported within this paper.

Chapter 5

Frozen Virtual Natural Orbitals++ for Coupled-Cluster Linear-Response Theory

5.1 Introduction

In the construction of many-body electronic wave functions, the scaling of a given method with the number of molecular orbitals (MOs) plays a pivotal role in the ultimate cost of the calculation. For many-body methods such as coupled cluster (CC),[4, 5, 6] which, in its canonical formulation, displays a higher-order polynomial dependence on the number of MOs, numerous mechanisms have been explored over the last half century for reducing

the size of the virtual-MO space. Among the earliest of these was Löwdin's[1] introduction in 1955 of the concept of "natural orbitals" (NOs) — orbitals for which the one-electron density matrix is diagonal. Löwdin demonstrated that NOs yield faster convergence of the configuration interaction (CI) wave function expansion than Hartree-Fock MOs. Some years later, Bender and Davidson[7] used natural orbitals in conjunction with CI (NO-CI) calculations to construct and analyze the most important configurations contributing to the correlated wave functions for a series of closed- and open-shell first-row diatomic hydrides. This work motivated Barr and Davidson a year later[8] to utilize only the virtual natural orbitals, obtained by diagonalization of the virtual-virtual block of the one-electron density matrix, for NO-CI calculations on the Ne atom.

The concept of pair natural orbitals (PNOs – originally called "pseudonatural orbitals") was developed by Edmiston and Krauss,[9] by Meyer[10], and by Ahlrichs and co-workers.[11] In the PNO approach, the virtual-virtual MO block of the one-electron density is constructed and diagonalized independently for each occupied MO pair, leading to separate non-orthogonal virtual spaces. Although this approach leads to rapid convergence of the correlation energy with respect to the size of the virtual space, it was little used following initial investigations until it was resurrected in recent years by Neese and co-workers with great success in the context of reduced-scaling electron correlation methods.[12, 13]

Following these pioneering efforts, NOs have been exploited in numerous applications to construct compact CI,[14, 15, 16] multiconfigurational self-consistent-field (MCSCF),[17] and CC wave functions.[18, 19, 20, 2, 21, 22] In many of these studies, the virtual-MO block

of the one-electron density is first obtained from a simpler model, such as second-order many-body perturbation theory (MBPT) calculation, and then diagonalized to yield the virtual-NO space. The space is then truncated based on an occupation-number-related criterion and fixed for the subsequent correlated-wave-function calculation. In addition, the final energy is commonly corrected using the second-order Møller-Plesset perturbation theory (MP2) correlation energy contributions arising from the external virtual space. These studies have indicated that, for energetics and related properties, even aggressive truncation of the virtual-NO space often has only a small impact on the resulting property as compared to full-space computations. For example, Landau *et al.*[2] found that, for NOs combined with the equation-of-motion coupled-cluster method for ionized states (EOM-IP-CC), reduction of the virtual space by up to 70% yielded truncation errors within ca. 1 kcal/mol for ionization energies of organic compounds and non-parallelity errors in potential surfaces of weakly bound complexes.

While the above studies have demonstrated clearly the usefulness of NOs for aggressively truncating the virtual space when computing correlation energies, more complex properties have yet to be considered. As shown in a number of recent reports,[23, 24, 25, 26, 27, 3] properties that are related to the linear- or higher-order response of the wave function to external electric and magnetic fields, for example, exhibit much greater sensitivity to the quality of the wave function than simple energetics. In particular, local correlation methods have been shown[23, 24, 25, 26, 3] to require significantly larger domains for properties such as polarizabilities than for ground-state energies. Furthermore, the many-body expansion

— which has yielded impressive convergence for energetics and dipole moments for clusters of weakly interacting molecules (such as a solute embedded in an explicit solvent) — converges erratically, at best, for spectroscopic response properties due to its strong basis-set dependence.[28] In order to account for this, the ability to reduce the dimensionality of the virtual space becomes paramount. Thus, the focus of the present work is on the extension of the NO approach to linear-response properties, especially the case of frequency-dependent dipole polarizabilities.

5.2 Theoretical Background

5.2.1 Frozen Virtual Natural Orbitals

The MP2 unrelaxed one-electron density matrix can be written in terms of spin orbitals as:

$$\gamma_{pq} = \langle \Psi^{(1)} | \{ a_p^\dagger a_q \} | \Psi^{(1)} \rangle, \tag{5.1}$$

where $|\Psi^{(1)}\rangle$ is the first order correction to the Hartree-Fock wave function,

$$|\Psi^{(1)}\rangle = \frac{1}{4} \sum_{ijab} t_{ij}^{ab} |\Phi_{ij}^{ab}\rangle \tag{5.2}$$

and

$$t_{ij}^{ab} = \frac{\langle ij||ab\rangle}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b}. \tag{5.3}$$

Here, $\langle ij||ab\rangle$ is an antisymmetrized two-electron integral in Dirac's notation and $\epsilon_i, \epsilon_a ...$ refer to the Hartree-Fock molecular orbital energies. We use the indices $i, j, k, ...$ to indicate

occupied orbitals while $a, b, c, ...$ denote virtual orbitals. $|\Phi_{ij}^{ab}\rangle$ refers to a doubly-excited determinant where occupied orbitals i and j are replaced by virtuals a and b respectively. The brackets around the second-quantized operators in Eq. (6.1) indicate normal ordering with respect to the reference wave function.

In the MP2 based NO method, the virtual-virtual block of γ_{pq} is constructed,

$$\gamma_{ab} = \frac{1}{2} \sum_{ijc} t_{ij}^{ac} t_{ij}^{bc}, \tag{5.4}$$

and then diagonalized,

$$\boldsymbol{\gamma V} = \boldsymbol{nV}. \tag{5.5}$$

The eigenvectors, $\boldsymbol{V}$, are the virtual natural orbitals (NOs), and the eigenvalues, $\boldsymbol{n}$, are the associated occupation numbers. As noted earlier, the wave function amplitudes contain significantly greater sparsity when represented in the NO basis than the original canonical MO basis; orbitals with lower occupation numbers yield $\hat{T}_2$ amplitudes with smaller magnitudes and concomitantly smaller contributions to the correlation energy. Thus, orbitals with occupation numbers below a selected threshold can be removed without introduction of significant errors, leading to reduced computational cost. The Hartree-Fock virtual molecular orbitals and associated integrals are then transformed to this truncated NO basis, followed by semicanonicalization of the virtual-virtual block of the Fock matrix, for subsequent computations using higher-order correlation methods such as CC theory. In most cases, the final correlation energy in the truncated virtual space is corrected using the MP2 energy in the external (non-truncated) NO space to minimize the resulting errors, as described below.

5.2.2 Coupled Cluster Response Theory

Dynamic response functions representing higher-order properties such as polarizabilities and hyperpolarizabilities, optical activity tensors, magnetizabilities, etc. can be obtained by expanding the expectation value of an appropriate time-independent operator in perturbational orders with respect to a time-dependent external field. The CC form of response theory has been routinely used for many years for accurate calculations of such properties.[29] The CC linear response function for property operators $\boldsymbol{A}$ and $\boldsymbol{B}$, for example, can be written as

$$\langle\langle \boldsymbol{A}; \boldsymbol{B} \rangle\rangle_\omega = \frac{1}{2}\hat{C}^{\pm\omega}\hat{P}[A(-\omega), B(+\omega)] \left[\langle \Psi_0 | (1 + \hat{\Lambda}) \left([\bar{A}, \hat{X}^B_\omega] + \frac{1}{2}[[\bar{H}, \hat{X}^A_{-\omega}], \hat{X}^B_\omega] \right) | \Psi_0 \rangle \right] \tag{5.6}$$

where Ψ_0 is the reference wavefunction, $\hat{\Lambda}$ is a de-excitation operator used to parametrize the CC left hand wavefunction, ω is the frequency of the external field, $\hat{C}$ is a symmetrizer that simultaneously interchanges the sign of the field frequency and takes the complex conjugate of the expression, and $\hat{P}$ symmetrizes the expression with respect to the operators $\boldsymbol{A}$ and $\boldsymbol{B}$. Operators with an overbar have been similarity transformed with the ground state cluster operators, $\hat{T}$, e.g. $\bar{H} = e^{-\hat{T}}\hat{H}e^{\hat{T}}$. The first-order (right-hand) perturbed wave function is represented by $\hat{X}^B_\omega$, whose amplitudes can be obtained by solving a set of appropriate linear equations, e.g.,

$$\langle \Phi^{ab\cdots}_{ij\cdots} | (\bar{H} - \omega) | \hat{X}^B_\omega | \Psi_0 \rangle = -\langle \Phi^{ab\cdots}_{ij\cdots} | \bar{B} | \Psi_0 \rangle. \tag{5.7}$$

In the case of dynamic polarizabilities, for example, both $\boldsymbol{A}$ and $\boldsymbol{B}$ are cartesian components of the electric dipole operator, $\boldsymbol{\mu} = -\boldsymbol{r}$, and the isotropic polarizability α_ω is related to the

trace of the polarizability tensor such that Eq. (6.6) reduces to the following form (for real wavefunctions),

$$\alpha_\omega = \frac{1}{3}\mathrm{Tr}\left[\langle\Psi_0|(1+\hat{\Lambda})\left(\left[\bar{\mu},(\hat{X}^\mu_\omega+\hat{X}^\mu_{-\omega})\right]+\left[\left[\bar{H},\hat{X}^\mu_{-\omega}\right],\hat{X}^\mu_\omega\right]\right)|\Psi_0\rangle\right] \tag{5.8}$$

For dynamic polarizabilities computed using the coupled cluster singles and doubles (CCSD) method, for example, six sets of perturbed amplitudes, $\hat{X}_1(X_i^a)$ and $\hat{X}_2(X_{ij}^{ab})$, must be computed, one for each cartesian component of $\boldsymbol{\mu}$ at both positive and negative frequencies.

5.3 Computational Details

The primary molecular test cases of this work is hydrogen peroxide, H_2O_2, though additional tests are reported for related species such as (*P*)-dimethylallene and (*S*)-methyloxirane, as well as those same molecules interacting with one or more water molecules. All molecular structures were optimized using the B3LYP functional[30, 31, 32] in the aug-cc-pVDZ (aDZ) basis.[33, 34, 35] (Coordinates of all structures are provided in Tables S1-S5 of the Supporting Information.) Frequency-dependent polarizabilities were computed at the coupled cluster singles and doubles (CCSD) level of theory[36] using a linear-response formulation.[37] While the aDZ basis set was used for most test calculations, the larger aug-cc-pVTZ (aTZ) and aug-cc-pVQZ (aQZ) basis sets were also employed for selected analyses.[34] All orbitals were active in the correlated models utilized here, and all coupled-cluster calculations were carried out using the PSI4 open-source quantum chemistry package.[38] In all calculations in which virtual orbitals are truncated, the same orbital space is used in the solution of the

unperturbed and perturbed amplitudes equations, just as is required for locally correlated property calculations.[39] We note that the size-extensivity of the polarizability is unaffected by the truncation of the virtual space.

5.4 Results and Discussion

It is well known that deletion of higher-energy canonical Hartree-Fock MOs (CMOs) typically leads to significant errors in the recovery of electron correlation energies, whereas, by design, the truncation of virtual NOs with low occupation numbers results in little loss of accuracy.[2, 19, 20] For example, Fig. 4.1 plots the error in the CCSD correlation energy of H_2O_2 in the aDZ basis as a function of the number of frozen virtual CMOs (removed starting from the highest energy orbitals) or NOs (starting from the lowest occupation numbers). Clearly the correlation energy is very sensitive towards the removal of CMOs, with the error increasing by more than 4 kcal/mol after the deletion of even one virtual orbital. On the other hand, removal of low occupation-number NOs introduces errors of only ca. 2.5 kcal/mol even when up to 33% of the virtual space (18 of 55 orbitals) is truncated. The errors in the NO basis can be further minimized by employing an MP2 energy correction,

$$E_{\text{corr}}^{\text{MP2}} = \frac{1}{4} \sum_{ij} \sum_{ab \in \text{ext}} \frac{|\langle ij||ab\rangle|^2}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b}. \tag{5.9}$$

where the summation over virtual orbitals is limited to NOs in the external, truncated space. Fig. 4.2 plots the error in the CCSD correlation energy for the same system as above in the NO basis, with and without this correction (left-hand vertical axis), as well as the correction

itself (right-hand axis). After employing the correction, the error in the correlation energy falls to less than 0.1 kcal/mol when 1/3 of the virtual space is eliminated. Similar results are obtained for aTZ and aQZ basis sets, where the truncation errors are less than 1 kcal/mol even after the removal of 50% of the virtual space. We note that the MP2 energy correction is significant (on the order of mE_h or several kcal/mol) even for small compounds such as H_2O_2. The correction scales linearly with the number of electrons and thus is a critical component for the success of the frozen virtual NO approach for larger molecules.

5.4.1 Frozen Virtual Orbitals and Response Properties

What is the impact of freezing virtual orbitals — whether CMO or NO — on higher-order properties? Fig. 4.3 plots errors in dynamic polarizabilites (computed at a wavelength of 589 nm) as a function of the number of virtual orbitals deleted at the CCSD/aDZ level of theory for the same H_2O_2 test case as above. Comparison to Fig. 4.1 reveals precisely the opposite behavior for polarizabilities as for correlation energies, *viz.*, truncation of the CMO virtual space induces much smaller errors than does that of the NO virtual space. For the latter, errors increase approximately linearly with the number of frozen virtual NOs. On the other hand, for the CMO basis, the error increases slowly to a maximum of 1.9% when 13 virtual orbitals are removed and then decreases to only -0.3% when as many as 27 orbitals (ca. 50% of the virtual space) are frozen. This trend is not unique to H_2O_2 or the aDZ basis set. As shown in Fig. S1 of the Supporting Information, the same behavior is observed for other molecules such as dimethylallene, methyloxirane, and such compounds interacting

with explicit solvent molecules. In addition, Figs. S2 and S3 report the same trends for H_2O_2 with the larger aTZ and aQZ basis sets.

What is the source of this unexpected behavior? It is well known that diffuse basis sets are essential for the accurate descriptions of a variety of response properties, such as dipole polarizabilities.[35] Fig. 4.4 plots the spatial extent — $\langle r^2 \rangle$ — for each virtual CMO or NO in the same ordering as they are deleted in Fig. 4.3. The Figure clearly shows that the earliest NOs to be removed (the ones with the lowest occupation numbers) are the most diffuse, *i.e.*, they should contribute substantially to the description of the dynamic polarizability. On the other hand, the first CMOs to be frozen (those with the highest orbital energies) are also the most compact and thus contribute the least to this property. Given that highly diffuse basis functions typically contribute primarily to CMOs with low orbital energies — often below that of what is normally considered the true anti-bonding "LUMO" — the latter result is not surprising; these diffuse CMOs appear to the far right of Fig. 4.4 and are thus never deleted, leading to the good behavior of the CMO truncation in Fig. 4.3. These same diffuse basis functions, however, contribute little to the description of dynamical correlation effects, and thus exhibit very low occupation numbers upon transformation to the NO virtual space. Thus, they are truncated first in the NO basis, yielding the large errors in the polarizability depicted in Fig. 4.3.

The above observations suggest that, for computing response properties such as dynamic polarizabilities, an alternative approach to truncation of the virtual space is to order the orbitals by increasing values of $\langle r^2 \rangle$ rather than by decreasing orbital energies (as is done for

CMOs) or increasing occupation numbers (for NOs). Fig. 4.5 plots errors in the CCSD/aDZ dynamic polarizability of H_2O_2 as the virtual CMOs or NOs are removed in order of increasing spatial extent. While the errors are comparable to that observed in Fig. 4.3 for the CMO truncation, the behavior associated with removal of virtual NOs is significantly different. First, the polarizability errors exhibit two plateaus: one associated with diffuse NOs 2-5 and another with NOs 16-30, and removal of these NOs has little impact on the observed error. However, deletion of the virtual NO with the *smallest* spatial extent unexpectedly leads to a large initial error (ca. 7.5 a.u.), followed later by a linearly increasing error as virtual NOs 6-15 are removed. Clearly, spatial extent is not the only criterion by which we may predict the importance of a given virtual NO to the polarizability.

Another possible source of error is the lack of orbital response in the chosen formulation of the coupled cluster linear response function.[40] In the infinite-lifetime approximation, frequency-dependent properties such as dipole polarizabilities exhibit first-order poles at the excitation frequencies. In the coupled cluster formulation described above, the orbital response to the external field is typically neglected so that these poles correspond solely to the response for the correlated wave function, and additional, spurious poles arising due to the Hartree-Fock reference determinant will not appear. (This approximation is typically justified based on the fact that much of the orbital-response effects are accounted for by the singles amplitudes.[41]) In order to test whether the orbital relaxation significantly impacts the behavior of the computed polarizability as the virtual space is reduced, we have computed *static* ($\omega = 0.0$) polarizabilities using finite-differences (with a a central-difference formula

with a differential field strength of 0.001 a.u.) The errors in the CCSD/aDZ static polarizability of H_2O_2 for both CMO and NO virtual spaces are reported in Fig. 4.6. Interestingly, with orbital relaxation included, the truncation of the NO space now becomes better behaved than the CMO space over a large domain of orbitals removed. Unfortunately, we cannot take advantage of this improvement in conjunction with Hartree-Fock orbitals without corrupting the pole structure of the response function. A Brueckner or orbital-optimized approach may prove superior in this regard, though further investigation is warranted.[40, 42]

5.4.2 Wave Function Truncation in the Virtual-Orbital Space

For additional insight into the above observations, we examine errors arising in dynamic polarizabilities as a function of truncation of specific wavefunction parameters in either the CMO or NO basis. Fig. 4.7 plots the errors in CCSD/aDZ dynamic polarizabilities of H_2O_2 as a result of truncations of the unperturbed ground-state cluster amplitudes $\hat{T}$ and $\hat{\Lambda}$, as well as perturbed amplitudes, $\hat{X}^{\mu}_{\omega}$, represented in the CMO basis. Note that, in this analysis, only the specified amplitudes associated with the selected CMOs are forced to zero; the CMOs remain active for all other wave function components. From the Figure, it can be clearly seen that removing $\hat{T}_1$ alone does not introduce any significant error in the polarizability, whereas truncating $\hat{T}_2$ amplitudes results in substantial positive errors which increase almost linearly with the number of virtual CMOs. (Not surprisingly, freezing both $\hat{T}_1$ and $\hat{T}_2$ amplitudes together have essentially the same effect as freezing $\hat{T}_2$ amplitudes alone.) Alternatively, for the left-hand wave function, removing $\hat{\Lambda}_1$ and $\hat{\Lambda}_2$ amplitudes either separately or pairwise

seem to have negligible impact. In the case of the perturbed amplitudes, only small (negative) errors are introduced even after freezing all $\hat{X}_1$ amplitudes involving almost 23 virtual CMOs, but the error then rises sharply with further truncation. On the other hand, the negative errors due to truncation of $\hat{X}_2$ amplitudes are significant from the beginning and increase almost linearly. Thus the error due to removal of both $\hat{X}_1$ and $\hat{X}_2$ amplitudes belonging to the first 23 CMOs is due to elimination of $\hat{X}_2$ amplitudes alone, whereas beyond that limit, the total error corresponds to the sum of errors from $\hat{X}_1$ and $\hat{X}_2$ truncation.

A key observation is that, within the domain of the first 23 virtual CMOs, the positive errors introduced by truncation of $\hat{T}_2$ is cancelled almost exactly by the negative errors arising from the truncation of $\hat{X}_2$. This is further illustrated in Fig. **??**, which plots on a narrow range the errors in the polarizability associated with truncating specific classes of $\hat{T}_2$ and $\hat{X}_2$ amplitudes against the total errors obtained by freezing CMOs entirely (for all amplitudes). Outside of this domain, errors associated with neglect of $\hat{X}_1$ amplitudes become dominant, leading to the accumulation of negative total errors observed in Fig. 4.3. Thus, the apparently robust performance of the truncation of the virtual CMO space arises, in fact, from offsetting errors.

Similarly to the above analysis for the virtual CMO space, Fig. 4.9 reports errors in the CCSD/aDZ polarizability of H_2O_2 introduced by the neglect of various classes of wave function amplitudes associated with selected virtual NOs. We observe first that, unlike the CMO case, neglecting $\hat{T}_2$ amplitudes associated with particular virtual NOs has no significant effect on the error. This behavior is expected, because the $\hat{T}_2$ amplitudes are, by construction, sparse in the virtual NO basis such that orbitals with low occupation numbers are associated

with $\hat{T}_2$ amplitudes of smaller magnitude. Furthermore, just as in the CMO case, the removal of $\hat{T}_1$, $\hat{\Lambda}_1$ and $\hat{\Lambda}_2$ amplitudes introduces only small errors, while the removal of selected $\hat{X}_2$ amplitudes based on NOs yields significant negative errors that increase linearly with the number of virtual NOs truncated.

However, unlike the virtual CMO case, neglecting $\hat{X}_1$ amplitudes corresponding to specific virtual NOs introduces large negative errors in the polarizability even from the first NO removed, and the total error obtained by truncating both $\hat{X}_1$ and $\hat{X}_2$ amplitudes is almost the same as the error due to truncation of the $\hat{X}_1$ amplitudes alone. Indeed, the greatest contribution ($> 90\%$) to the total polarizability errors arises from the perturbed singles amplitudes.

The significance of the $\hat{X}_1$ amplitudes is evident upon analysis of their leading-order contribution to the polarizability [cf. Eq. (6.8)],

$$\alpha_\omega \leftarrow \frac{1}{3}\sum_{ia}\sum_{x}\mu_{ia}^{x}(X_{ia}^{x}(\omega)+X_{ia}^{x}(-\omega)), \tag{5.10}$$

where μ_{ia}^{x} is an element of the occupied-virtual block of the Cartesian component, x, of the electric-dipole moment integral matrix, and the inner sum runs over all such components. The singles themselves are obtained from the corresponding form of Eq. (6.7),

$$\langle\Psi_i^a|X_\omega^\mu|\Psi_0\rangle = -\sum_\nu \langle\Psi_i^a|(\bar{H}-\omega)^{-1}|\nu\rangle\langle\nu|\bar{\mu}|\Psi_0\rangle, \quad \nu\in\{\Psi_j^b,\Psi_{jk}^{cd}\}, \tag{5.11}$$

where $\bar{\mu}$ is the similarity transformed electric-dipole operator,

$$\bar{\mu} = \hat{\mu} + \Big[\hat{\mu},\hat{T}\Big] + \frac{1}{2}\Big[\Big[\hat{\mu},\hat{T}\Big],\hat{T}\Big]. \tag{5.12}$$

The corresponding leading-order contribution to the $\hat{X}_1$ amplitudes themselves is

$$X_{ia}^{\mu}(\omega) \leftarrow \frac{\mu_{ia}}{\overline{H}_{ii} - \overline{H}_{aa} + \omega}, \tag{5.13}$$

where the largest contribution to the diagonal elements of the similarity transformed Hamiltonian, $\bar{H}$, arises from the orbital energies (more precisely, the diagonal Fock matrix elements expressed in the CMO or NO basis), which are plotted in Fig. 4.10. While the values for the virtual CMOs decrease steadily, their NO counterparts actually *increase* and display greater oscillation. Clearly, the diagonal elements of the fock matrix for virtual NOs are significantly smaller in magnitude than the corresponding CMOs, which concomitantly increases the values of the $\hat{X}_1$ amplitudes associated with such NOs. This effect can be seen in Fig. 4.11 which plots the sum of the absolute values of $\hat{X}_1$ amplitudes for a given virtual orbital, i.e. $\sum_i |X_i^a|$, for both virtual CMOs and NOs. Thus, the sparsity of the $\hat{X}_1$ amplitudes present in the CMO basis is almost completely lost in the NO basis, which leads to the large errors in dynamic polarizabilities due to the truncation of $\hat{X}_1$ amplitudes in the NO basis observed earlier.

Furthermore, the dependence of the $\hat{X}_1$ amplitudes on both $\hat{T}_2$ and $\hat{X}_2$ plays a significant role in the overall *sign* of the polarizability error. This point is clearly illustrated in Fig. 4.12, which plots the 2-norm of the $\hat{X}_1$ vector as $\hat{T}_2$ or $\hat{X}_2$ amplitudes associated with a given virtual CMO is neglected (in the same manner as used in Figs. 4.7 and **??**). As the $\hat{T}_2$ amplitudes associated with a given virtual orbital are removed, the norm of the $\hat{X}_1$ vector increases, leading to the positive errors in the polarizability observed in Fig. 4.7. On the other hand, removal of $\hat{X}_2$ amplitudes leads to a decrease in the norm of the $\hat{X}_1$ vector and

the negative errors in the polarizability appearing in Fig. 4.7.

5.4.3 External-Space Corrections

As noted earlier, a key aspect of the strong performance of virtual NOs for correlation energies is the use of MP2-based corrections for the contributions of the truncated or "external" virtual space, as given in Eq. (6.9). When considering a similar correction for properties, however, we have explored three options: time-dependent Hartree-Fock (TDHF) and both static and dynamic CC2[41] corrections. In each case, we have used the same MP2-based virtual NO space and computed the correction as the difference between the full-virtual space polarizability and the truncated virtual-NO polarizability, with the CCSD/aDZ results for polarizabilties of H_2O_2 shown in Fig. 4.13.

The CC2-based corrections recover nearly all of the error associated with the uncorrected virtual NO space across a wide range of truncation, with the static correction actually yielding slightly smaller errors than its dynamic counterpart. The less expensive TDHF corrections offer significant improvement over the original virtual NO results, and recover most (ca. 85%) of the error until about 40% (21 orbitals) of the virtual space has been deleted, but it clearly produces overall larger errors than the correlated methods. In addition, the TDHF corrections are potentially problematic because the pole structure of the polarizability naturally follows that of the underlying Hartree-Fock perturbed orbitals rather than that of the correlated wave functions,[43, 44, 45] a criticism that would also hold for purely MP2-

based corrections. On the other hand, the CC2-based corrections are significantly more expensive than TDHF, in part because of their iterative nature and the need to transform two-electron integrals involving three virtual orbitals (with the latter criticism again holding for an MP2-based correction).

5.4.4 Perturbed Natural Orbitals

Given that the purpose of the NOs is to "focus" the important components of the basis for the description of electron correlation effects into a compact space, without consideration of the importance of the basis set for other properties, an alternative approach might be to build a virtual space that explicitly takes such properties into account, such as the inclusion of the perturbed one-electron density in the definition of the space. Instead of diagonalizing the ground-state MP2 density, we diagonalize its gradient with respect to the external electric field, thereby incorporating the effects of the external perturbations. The resulting "occupation numbers" obtained thus carry information about orbital occupancies in the perturbed states. The x^{th} Cartesian component of the perturbed density matrix can be written in terms of spin orbitals as

$$\gamma_{ab}^{x} = \frac{1}{2}\sum_{ijc} P_{+}(a,b)\left[\frac{t_{ij}^{bc}}{D_{ij}^{ac}}\left(P_{-}(a,c)\sum_{d} t_{ij}^{ad}\mu_{cd}^{x} - P_{-}(i,j)\sum_{m} t_{im}^{ac}\mu_{mj}^{x}\right)\right] \tag{5.14}$$

where P_{+} and P_{-} are symmetric and anti-symmetric permutation operators, respectively, and the orbital-energy denominator $D_{ij}^{ac} = \epsilon_i + \epsilon_j - \epsilon_a - \epsilon_c$. (For computational convenience, orbital relaxation terms have been neglected.) To obtain the perturbed NOs and their

corresponding eigenvalues, we take an average of the three cartesian components of the density and diagonalize the result. The eigenvalues consist of both positive and negative values as the perturbed density matrix is not positive definite, and thus we truncate the orbitals based on the absolute values of these eigenvalues.

Fig. 4.14 compares the performance of these perturbed NOs with that of the CMOs and NOs, where the error in the CCSD/aDZ dynamic polarizability of H_2O_2 is plotted as a function of the number of virtual orbitals removed. While the perturbed NOs lower the truncation errors associated with conventional NOs, they still introduce significantly higher errors than the corresponding CMOs. The reason for this underperformance is related to the definition of the perturbed density, which naturally bears strong similarity to its unperturbed counterpart [cf. Eqs. (6.4) and (6.14).] As a result, similar sparsity and orbital energy issues illustrated in Figs. 4.10 and 4.11 arise for the perturbed density just as for the original NOs, indicating that such an approach does not resolve the problem.

5.4.5 The Dipole-Amplitude Criterion

The above results demonstrate that the CMO basis provides the best performance among the various virtual spaces considered here, albeit based on cancellation of errors, but what criterion should be used to determine the truncation level that provides optimal balance between computational cost (most compact virtual space) and accuracy? The CMO orbital energies are one possibility, but they have no direct connection to the properties in question.

Another alternative is the "dipole amplitude", d_a, which is defined for each virtual CMO as

$$d_a \equiv \sum_x \sum_i \frac{(\mu_{ia}^x)^2}{\epsilon_i - \epsilon_a}. \tag{5.15}$$

This expression, which is trivially computed post-Hartree-Fock, is based on Eqs. (6.10) and (6.13), using the fact that the leading contributions to the diagonal elements of the similarty-transformed Hamiltonian are the orbital energies. As we are constructing a phenomenological truncation criterion, we have chosen to neglect the dependence on the field frequency.

Fig. 4.15 plots the values of the dipole amplitude for each CMO, as well as the corresponding error in the CCSD/aDZ dipole-polarizability of H_2O_2 introduced by deleting the $\hat{X}_1$ amplitude associated with a given virtual CMO. There is a clear correlation between the two functions, and one can see that as both the error due to truncation of $\hat{X}_1$ and the value of the dipole length increase sharply once we reach CMOs 27-28 — ca. 50% of the virtual space for this test case. Accordingly, one should stop the truncation of the virtual CMO space once the dipole length values start rising sharply. While the optimal choice of such a threshold will be addressed systematically in subsequent work, our preliminary analyses using H_2O_2, methyloxirane, dimethylallene, and related compounds suggest a cutoff of ca. 3-3.5% yields minimal errors in the polarizability.

5.5 Conclusions

On the basis of the above findings, we conclude that, in the absence of orbital relaxation, virtual NOs are not suited for higher-order property calculations such as dynamic polariz-

abilities, and that the occupation number is not an acceptable criterion for estimating the importance of a virtual orbital for such calculations. Although the use of external space corrections based on CC2 polarizabilities reduces the observed truncation errors, they are too relatively costly to be practical for large molecular systems. Furthermore, the use of perturbed virtual NOs offers only slight improvement as compared to unperturbed virtual NOs. CMOs, on the other hand, provide a much more stable mechanism for reducing the size of the virtual space — with truncation of up to 50% of the orbitals yielding shifts of less than 2% as compared to full-space calculation – but the source of their success lies in a significant cancellation of errors. Although further systematic studies are needed, the dipole amplitude appears to provide a useful threshold for an *a priori* truncation of the CMO virtual space.

5.6 Acknowledgements

This research was supported by a grant (CHE-1465149) from the U.S. National Science Foundation. The authors acknowledge Advanced Research Computing at Virginia Tech for providing computational resources and technical support that have contributed to the results reported within this paper.

Chapter 6

PNO++ approach for Coupled-Cluster Linear-Response Theory

6.1 Introduction

In the construction of many-body electronic wave functions, the scaling of a given method with the number of molecular orbitals (MOs) plays a pivotal role in the ultimate cost of the calculation. For many-body methods such as coupled cluster (CC),[4, 5, 6] which, in its canonical formulation, displays a higher-order polynomial dependence on the number of MOs, numerous mechanisms have been explored over the last half century for reducing

the size of the virtual-MO space. Among the earliest of these was Löwdin's[1] introduction in 1955 of the concept of "natural orbitals" (NOs) — orbitals for which the one-electron density matrix is diagonal. Löwdin demonstrated that NOs yield faster convergence of the configuration interaction (CI) wave function expansion than Hartree-Fock MOs. Some years later, Bender and Davidson[7] used natural orbitals in conjunction with CI (NO-CI) calculations to construct and analyze the most important configurations contributing to the correlated wave functions for a series of closed- and open-shell first-row diatomic hydrides. This work motivated Barr and Davidson a year later[8] to utilize only the virtual natural orbitals, obtained by diagonalization of the virtual-virtual block of the one-electron density matrix, for NO-CI calculations on the Ne atom.

The concept of pair natural orbitals (PNOs – originally called "pseudonatural orbitals") was developed by Edmiston and Krauss,[9] by Meyer[10], and by Ahlrichs and co-workers.[11] In the PNO approach, the virtual-virtual MO block of the one-electron density is constructed and diagonalized independently for each occupied MO pair, leading to separate non-orthogonal virtual spaces. Although this approach leads to rapid convergence of the correlation energy with respect to the size of the virtual space, it was little used following initial investigations until it was resurrected in recent years by Neese and co-workers with great success in the context of reduced-scaling electron correlation methods.[12, 13]

Following these pioneering efforts, NOs have been exploited in numerous applications to construct compact CI,[14, 15, 16] multiconfigurational self-consistent-field (MCSCF),[17] and CC wave functions.[18, 19, 20, 2, 21, 22] In many of these studies, the virtual-MO block

of the one-electron density is first obtained from a simpler model, such as second-order many-body perturbation theory (MBPT) calculation, and then diagonalized to yield the virtual-NO space. The space is then truncated based on an occupation-number-related criterion and fixed for the subsequent correlated-wave-function calculation. In addition, the final energy is commonly corrected using the second-order Møller-Plesset perturbation theory (MP2) correlation energy contributions arising from the external virtual space. These studies have indicated that, for energetics and related properties, even aggressive truncation of the virtual-NO space often has only a small impact on the resulting property as compared to full-space computations. For example, Landau *et al.*[2] found that, for NOs combined with the equation-of-motion coupled-cluster method for ionized states (EOM-IP-CC), reduction of the virtual space by up to 70% yielded truncation errors within ca. 1 kcal/mol for ionization energies of organic compounds and non-parallelity errors in potential surfaces of weakly bound complexes.

While the above studies have demonstrated clearly the usefulness of NOs for aggressively truncating the virtual space when computing correlation energies, more complex properties have yet to be considered. As shown in a number of recent reports,[23, 24, 25, 26, 27, 3] properties that are related to the linear- or higher-order response of the wave function to external electric and magnetic fields, for example, exhibit much greater sensitivity to the quality of the wave function than simple energetics. In particular, local correlation methods have been shown[23, 24, 25, 26, 3] to require significantly larger domains for properties such as polarizabilities than for ground-state energies. Furthermore, the many-body expansion

— which has yielded impressive convergence for energetics and dipole moments for clusters of weakly interacting molecules (such as a solute embedded in an explicit solvent) — converges erratically, at best, for spectroscopic response properties due to its strong basis-set dependence.[28] In order to account for this, the ability to reduce the dimensionality of the virtual space becomes paramount. Thus, the focus of the present work is on the extension of the NO approach to linear-response properties, especially the case of frequency-dependent dipole polarizabilities.

6.2 Theoretical Background

6.2.1 Frozen Virtual Natural Orbitals

The MP2 unrelaxed one-electron density matrix can be written in terms of spin orbitals as:

$$\gamma_{pq} = \langle \Psi^{(1)} | \{ a_p^\dagger a_q \} | \Psi^{(1)} \rangle, \tag{6.1}$$

where $|\Psi^{(1)}\rangle$ is the first order correction to the Hartree-Fock wave function,

$$|\Psi^{(1)}\rangle = \frac{1}{4} \sum_{ijab} t_{ij}^{ab} |\Phi_{ij}^{ab}\rangle \tag{6.2}$$

and

$$t_{ij}^{ab} = \frac{\langle ij||ab\rangle}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b}. \tag{6.3}$$

Here, $\langle ij||ab\rangle$ is an antisymmetrized two-electron integral in Dirac's notation and $\epsilon_i, \epsilon_a ...$ refer to the Hartree-Fock molecular orbital energies. We use the indices $i, j, k, ...$ to indicate

occupied orbitals while $a, b, c, ...$ denote virtual orbitals. $|\Phi_{ij}^{ab}\rangle$ refers to a doubly-excited determinant where occupied orbitals i and j are replaced by virtuals a and b respectively. The brackets around the second-quantized operators in Eq. (6.1) indicate normal ordering with respect to the reference wave function.

In the MP2 based NO method, the virtual-virtual block of γ_{pq} is constructed,

$$\gamma_{ab} = \frac{1}{2} \sum_{ijc} t_{ij}^{ac} t_{ij}^{bc}, \tag{6.4}$$

and then diagonalized,

$$\boldsymbol{\gamma V} = \boldsymbol{nV}. \tag{6.5}$$

The eigenvectors, $\boldsymbol{V}$, are the virtual natural orbitals (NOs), and the eigenvalues, $\boldsymbol{n}$, are the associated occupation numbers. As noted earlier, the wave function amplitudes contain significantly greater sparsity when represented in the NO basis than the original canonical MO basis; orbitals with lower occupation numbers yield $\hat{T}_2$ amplitudes with smaller magnitudes and concomitantly smaller contributions to the correlation energy. Thus, orbitals with occupation numbers below a selected threshold can be removed without introduction of significant errors, leading to reduced computational cost. The Hartree-Fock virtual molecular orbitals and associated integrals are then transformed to this truncated NO basis, followed by semicanonicalization of the virtual-virtual block of the Fock matrix, for subsequent computations using higher-order correlation methods such as CC theory. In most cases, the final correlation energy in the truncated virtual space is corrected using the MP2 energy in the external (non-truncated) NO space to minimize the resulting errors, as described below.

6.2.2 Coupled Cluster Response Theory

Dynamic response functions representing higher-order properties such as polarizabilities and hyperpolarizabilities, optical activity tensors, magnetizabilities, etc. can be obtained by expanding the expectation value of an appropriate time-independent operator in perturbational orders with respect to a time-dependent external field. The CC form of response theory has been routinely used for many years for accurate calculations of such properties.[29] The CC linear response function for property operators $\boldsymbol{A}$ and $\boldsymbol{B}$, for example, can be written as

$$\langle\langle \boldsymbol{A}; \boldsymbol{B} \rangle\rangle_\omega = \frac{1}{2}\hat{C}^{\pm\omega}\hat{P}[A(-\omega), B(+\omega)] \left[\langle \Psi_0 | (1 + \hat{\Lambda}) \left([\bar{A}, \hat{X}^B_\omega] + \frac{1}{2}[[\bar{H}, \hat{X}^A_{-\omega}], \hat{X}^B_\omega] \right) | \Psi_0 \rangle \right] \tag{6.6}$$

where Ψ_0 is the reference wavefunction, $\hat{\Lambda}$ is a de-excitation operator used to parametrize the CC left hand wavefunction, ω is the frequency of the external field, $\hat{C}$ is a symmetrizer that simultaneously interchanges the sign of the field frequency and takes the complex conjugate of the expression, and $\hat{P}$ symmetrizes the expression with respect to the operators $\boldsymbol{A}$ and $\boldsymbol{B}$. Operators with an overbar have been similarity transformed with the ground state cluster operators, $\hat{T}$, e.g. $\bar{H} = e^{-\hat{T}}\hat{H}e^{\hat{T}}$. The first-order (right-hand) perturbed wave function is represented by $\hat{X}^B_\omega$, whose amplitudes can be obtained by solving a set of appropriate linear equations, e.g.,

$$\langle \Phi^{ab\cdots}_{ij\cdots} | (\bar{H} - \omega) | \hat{X}^B_\omega | \Psi_0 \rangle = -\langle \Phi^{ab\cdots}_{ij\cdots} | \bar{B} | \Psi_0 \rangle. \tag{6.7}$$

In the case of dynamic polarizabilities, for example, both $\boldsymbol{A}$ and $\boldsymbol{B}$ are cartesian components of the electric dipole operator, $\boldsymbol{\mu} = -\boldsymbol{r}$, and the isotropic polarizability α_ω is related to the

trace of the polarizability tensor such that Eq. (6.6) reduces to the following form (for real wavefunctions),

$$\alpha_\omega = \frac{1}{3}\mathrm{Tr}\left[\langle\Psi_0|(1+\hat{\Lambda})\left(\left[\bar{\mu},(\hat{X}^\mu_\omega+\hat{X}^\mu_{-\omega})\right]+\left[\left[\bar{H},\hat{X}^\mu_{-\omega}\right],\hat{X}^\mu_\omega\right]\right)|\Psi_0\rangle\right] \tag{6.8}$$

For dynamic polarizabilities computed using the coupled cluster singles and doubles (CCSD) method, for example, six sets of perturbed amplitudes, $\hat{X}_1(X_i^a)$ and $\hat{X}_2(X_{ij}^{ab})$, must be computed, one for each cartesian component of $\boldsymbol{\mu}$ at both positive and negative frequencies.

6.3 Computational Details

The primary molecular test cases of this work is hydrogen peroxide, H_2O_2, though additional tests are reported for related species such as (*P*)-dimethylallene and (*S*)-methyloxirane, as well as those same molecules interacting with one or more water molecules. All molecular structures were optimized using the B3LYP functional[30, 31, 32] in the aug-cc-pVDZ (aDZ) basis.[33, 34, 35] (Coordinates of all structures are provided in Tables S1-S5 of the Supporting Information.) Frequency-dependent polarizabilities were computed at the coupled cluster singles and doubles (CCSD) level of theory[36] using a linear-response formulation.[37] While the aDZ basis set was used for most test calculations, the larger aug-cc-pVTZ (aTZ) and aug-cc-pVQZ (aQZ) basis sets were also employed for selected analyses.[34] All orbitals were active in the correlated models utilized here, and all coupled-cluster calculations were carried out using the PSI4 open-source quantum chemistry package.[38] In all calculations in which virtual orbitals are truncated, the same orbital space is used in the solution of the

unperturbed and perturbed amplitudes equations, just as is required for locally correlated property calculations.[39] We note that the size-extensivity of the polarizability is unaffected by the truncation of the virtual space.

6.4 Results and Discussion

It is well known that deletion of higher-energy canonical Hartree-Fock MOs (CMOs) typically leads to significant errors in the recovery of electron correlation energies, whereas, by design, the truncation of virtual NOs with low occupation numbers results in little loss of accuracy.[2, 19, 20] For example, Fig. 4.1 plots the error in the CCSD correlation energy of H_2O_2 in the aDZ basis as a function of the number of frozen virtual CMOs (removed starting from the highest energy orbitals) or NOs (starting from the lowest occupation numbers). Clearly the correlation energy is very sensitive towards the removal of CMOs, with the error increasing by more than 4 kcal/mol after the deletion of even one virtual orbital. On the other hand, removal of low occupation-number NOs introduces errors of only ca. 2.5 kcal/mol even when up to 33% of the virtual space (18 of 55 orbitals) is truncated. The errors in the NO basis can be further minimized by employing an MP2 energy correction,

$$E_{\text{corr}}^{\text{MP2}} = \frac{1}{4} \sum_{ij} \sum_{ab \in \text{ext}} \frac{|\langle ij||ab\rangle|^2}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b}. \tag{6.9}$$

where the summation over virtual orbitals is limited to NOs in the external, truncated space. Fig. 4.2 plots the error in the CCSD correlation energy for the same system as above in the NO basis, with and without this correction (left-hand vertical axis), as well as the correction

itself (right-hand axis). After employing the correction, the error in the correlation energy falls to less than 0.1 kcal/mol when 1/3 of the virtual space is eliminated. Similar results are obtained for aTZ and aQZ basis sets, where the truncation errors are less than 1 kcal/mol even after the removal of 50% of the virtual space. We note that the MP2 energy correction is significant (on the order of mE_h or several kcal/mol) even for small compounds such as H_2O_2. The correction scales linearly with the number of electrons and thus is a critical component for the success of the frozen virtual NO approach for larger molecules.

6.4.1 Frozen Virtual Orbitals and Response Properties

What is the impact of freezing virtual orbitals — whether CMO or NO — on higher-order properties? Fig. 4.3 plots errors in dynamic polarizabilites (computed at a wavelength of 589 nm) as a function of the number of virtual orbitals deleted at the CCSD/aDZ level of theory for the same H_2O_2 test case as above. Comparison to Fig. 4.1 reveals precisely the opposite behavior for polarizabilities as for correlation energies, *viz.*, truncation of the CMO virtual space induces much smaller errors than does that of the NO virtual space. For the latter, errors increase approximately linearly with the number of frozen virtual NOs. On the other hand, for the CMO basis, the error increases slowly to a maximum of 1.9% when 13 virtual orbitals are removed and then decreases to only -0.3% when as many as 27 orbitals (ca. 50% of the virtual space) are frozen. This trend is not unique to H_2O_2 or the aDZ basis set. As shown in Fig. S1 of the Supporting Information, the same behavior is observed for other molecules such as dimethylallene, methyloxirane, and such compounds interacting

with explicit solvent molecules. In addition, Figs. S2 and S3 report the same trends for H_2O_2 with the larger aTZ and aQZ basis sets.

What is the source of this unexpected behavior? It is well known that diffuse basis sets are essential for the accurate descriptions of a variety of response properties, such as dipole polarizabilities.[35] Fig. 4.4 plots the spatial extent — $\langle r^2 \rangle$ — for each virtual CMO or NO in the same ordering as they are deleted in Fig. 4.3. The Figure clearly shows that the earliest NOs to be removed (the ones with the lowest occupation numbers) are the most diffuse, *i.e.*, they should contribute substantially to the description of the dynamic polarizability. On the other hand, the first CMOs to be frozen (those with the highest orbital energies) are also the most compact and thus contribute the least to this property. Given that highly diffuse basis functions typically contribute primarily to CMOs with low orbital energies — often below that of what is normally considered the true anti-bonding "LUMO" — the latter result is not surprising; these diffuse CMOs appear to the far right of Fig. 4.4 and are thus never deleted, leading to the good behavior of the CMO truncation in Fig. 4.3. These same diffuse basis functions, however, contribute little to the description of dynamical correlation effects, and thus exhibit very low occupation numbers upon transformation to the NO virtual space. Thus, they are truncated first in the NO basis, yielding the large errors in the polarizability depicted in Fig. 4.3.

The above observations suggest that, for computing response properties such as dynamic polarizabilities, an alternative approach to truncation of the virtual space is to order the orbitals by increasing values of $\langle r^2 \rangle$ rather than by decreasing orbital energies (as is done for

CMOs) or increasing occupation numbers (for NOs). Fig. 4.5 plots errors in the CCSD/aDZ dynamic polarizability of H_2O_2 as the virtual CMOs or NOs are removed in order of increasing spatial extent. While the errors are comparable to that observed in Fig. 4.3 for the CMO truncation, the behavior associated with removal of virtual NOs is significantly different. First, the polarizability errors exhibit two plateaus: one associated with diffuse NOs 2-5 and another with NOs 16-30, and removal of these NOs has little impact on the observed error. However, deletion of the virtual NO with the *smallest* spatial extent unexpectedly leads to a large initial error (ca. 7.5 a.u.), followed later by a linearly increasing error as virtual NOs 6-15 are removed. Clearly, spatial extent is not the only criterion by which we may predict the importance of a given virtual NO to the polarizability.

Another possible source of error is the lack of orbital response in the chosen formulation of the coupled cluster linear response function.[40] In the infinite-lifetime approximation, frequency-dependent properties such as dipole polarizabilities exhibit first-order poles at the excitation frequencies. In the coupled cluster formulation described above, the orbital response to the external field is typically neglected so that these poles correspond solely to the response for the correlated wave function, and additional, spurious poles arising due to the Hartree-Fock reference determinant will not appear. (This approximation is typically justified based on the fact that much of the orbital-response effects are accounted for by the singles amplitudes.[41]) In order to test whether the orbital relaxation significantly impacts the behavior of the computed polarizability as the virtual space is reduced, we have computed *static* ($\omega = 0.0$) polarizabilities using finite-differences (with a a central-difference formula

with a differential field strength of 0.001 a.u.) The errors in the CCSD/aDZ static polarizability of H_2O_2 for both CMO and NO virtual spaces are reported in Fig. 4.6. Interestingly, with orbital relaxation included, the truncation of the NO space now becomes better behaved than the CMO space over a large domain of orbitals removed. Unfortunately, we cannot take advantage of this improvement in conjunction with Hartree-Fock orbitals without corrupting the pole structure of the response function. A Brueckner or orbital-optimized approach may prove superior in this regard, though further investigation is warranted.[40, 42]

6.4.2 Wave Function Truncation in the Virtual-Orbital Space

For additional insight into the above observations, we examine errors arising in dynamic polarizabilities as a function of truncation of specific wavefunction parameters in either the CMO or NO basis. Fig. 4.7 plots the errors in CCSD/aDZ dynamic polarizabilities of H_2O_2 as a result of truncations of the unperturbed ground-state cluster amplitudes $\hat{T}$ and $\hat{\Lambda}$, as well as perturbed amplitudes, $\hat{X}^{\mu}_{\omega}$, represented in the CMO basis. Note that, in this analysis, only the specified amplitudes associated with the selected CMOs are forced to zero; the CMOs remain active for all other wave function components. From the Figure, it can be clearly seen that removing $\hat{T}_1$ alone does not introduce any significant error in the polarizability, whereas truncating $\hat{T}_2$ amplitudes results in substantial positive errors which increase almost linearly with the number of virtual CMOs. (Not surprisingly, freezing both $\hat{T}_1$ and $\hat{T}_2$ amplitudes together have essentially the same effect as freezing $\hat{T}_2$ amplitudes alone.) Alternatively, for the left-hand wave function, removing $\hat{\Lambda}_1$ and $\hat{\Lambda}_2$ amplitudes either separately or pairwise

seem to have negligible impact. In the case of the perturbed amplitudes, only small (negative) errors are introduced even after freezing all $\hat{X}_1$ amplitudes involving almost 23 virtual CMOs, but the error then rises sharply with further truncation. On the other hand, the negative errors due to truncation of $\hat{X}_2$ amplitudes are significant from the beginning and increase almost linearly. Thus the error due to removal of both $\hat{X}_1$ and $\hat{X}_2$ amplitudes belonging to the first 23 CMOs is due to elimination of $\hat{X}_2$ amplitudes alone, whereas beyond that limit, the total error corresponds to the sum of errors from $\hat{X}_1$ and $\hat{X}_2$ truncation.

A key observation is that, within the domain of the first 23 virtual CMOs, the positive errors introduced by truncation of $\hat{T}_2$ is cancelled almost exactly by the negative errors arising from the truncation of $\hat{X}_2$. This is further illustrated in Fig. **??**, which plots on a narrow range the errors in the polarizability associated with truncating specific classes of $\hat{T}_2$ and $\hat{X}_2$ amplitudes against the total errors obtained by freezing CMOs entirely (for all amplitudes). Outside of this domain, errors associated with neglect of $\hat{X}_1$ amplitudes become dominant, leading to the accumulation of negative total errors observed in Fig. 4.3. Thus, the apparently robust performance of the truncation of the virtual CMO space arises, in fact, from offsetting errors.

Similarly to the above analysis for the virtual CMO space, Fig. 4.9 reports errors in the CCSD/aDZ polarizability of H_2O_2 introduced by the neglect of various classes of wave function amplitudes associated with selected virtual NOs. We observe first that, unlike the CMO case, neglecting $\hat{T}_2$ amplitudes associated with particular virtual NOs has no significant effect on the error. This behavior is expected, because the $\hat{T}_2$ amplitudes are, by construction, sparse in the virtual NO basis such that orbitals with low occupation numbers are associated

with $\hat{T}_2$ amplitudes of smaller magnitude. Furthermore, just as in the CMO case, the removal of $\hat{T}_1$, $\hat{\Lambda}_1$ and $\hat{\Lambda}_2$ amplitudes introduces only small errors, while the removal of selected $\hat{X}_2$ amplitudes based on NOs yields significant negative errors that increase linearly with the number of virtual NOs truncated.

However, unlike the virtual CMO case, neglecting $\hat{X}_1$ amplitudes corresponding to specific virtual NOs introduces large negative errors in the polarizability even from the first NO removed, and the total error obtained by truncating both $\hat{X}_1$ and $\hat{X}_2$ amplitudes is almost the same as the error due to truncation of the $\hat{X}_1$ amplitudes alone. Indeed, the greatest contribution ($> 90\%$) to the total polarizability errors arises from the perturbed singles amplitudes.

The significance of the $\hat{X}_1$ amplitudes is evident upon analysis of their leading-order contribution to the polarizability [cf. Eq. (6.8)],

$$\alpha_\omega \leftarrow \frac{1}{3}\sum_{ia}\sum_{x}\mu_{ia}^{x}(X_{ia}^{x}(\omega)+X_{ia}^{x}(-\omega)), \tag{6.10}$$

where μ_{ia}^{x} is an element of the occupied-virtual block of the Cartesian component, x, of the electric-dipole moment integral matrix, and the inner sum runs over all such components. The singles themselves are obtained from the corresponding form of Eq. (6.7),

$$\langle\Psi_i^a|X_\omega^\mu|\Psi_0\rangle = -\sum_\nu \langle\Psi_i^a|(\bar{H}-\omega)^{-1}|\nu\rangle\langle\nu|\bar{\mu}|\Psi_0\rangle, \quad \nu\in\{\Psi_j^b,\Psi_{jk}^{cd}\}, \tag{6.11}$$

where $\bar{\mu}$ is the similarity transformed electric-dipole operator,

$$\bar{\mu} = \hat{\mu} + \Big[\hat{\mu},\hat{T}\Big] + \frac{1}{2}\Big[\Big[\hat{\mu},\hat{T}\Big],\hat{T}\Big]. \tag{6.12}$$

The corresponding leading-order contribution to the $\hat{X}_1$ amplitudes themselves is

$$X_{ia}^{\mu}(\omega) \leftarrow \frac{\mu_{ia}}{\overline{H}_{ii} - \overline{H}_{aa} + \omega}, \tag{6.13}$$

where the largest contribution to the diagonal elements of the similarity transformed Hamiltonian, $\bar{H}$, arises from the orbital energies (more precisely, the diagonal Fock matrix elements expressed in the CMO or NO basis), which are plotted in Fig. 4.10. While the values for the virtual CMOs decrease steadily, their NO counterparts actually *increase* and display greater oscillation. Clearly, the diagonal elements of the fock matrix for virtual NOs are significantly smaller in magnitude than the corresponding CMOs, which concomitantly increases the values of the $\hat{X}_1$ amplitudes associated with such NOs. This effect can be seen in Fig. 4.11 which plots the sum of the absolute values of $\hat{X}_1$ amplitudes for a given virtual orbital, i.e. $\sum_i |X_i^a|$, for both virtual CMOs and NOs. Thus, the sparsity of the $\hat{X}_1$ amplitudes present in the CMO basis is almost completely lost in the NO basis, which leads to the large errors in dynamic polarizabilities due to the truncation of $\hat{X}_1$ amplitudes in the NO basis observed earlier.

Furthermore, the dependence of the $\hat{X}_1$ amplitudes on both $\hat{T}_2$ and $\hat{X}_2$ plays a significant role in the overall *sign* of the polarizability error. This point is clearly illustrated in Fig. 4.12, which plots the 2-norm of the $\hat{X}_1$ vector as $\hat{T}_2$ or $\hat{X}_2$ amplitudes associated with a given virtual CMO is neglected (in the same manner as used in Figs. 4.7 and **??**). As the $\hat{T}_2$ amplitudes associated with a given virtual orbital are removed, the norm of the $\hat{X}_1$ vector increases, leading to the positive errors in the polarizability observed in Fig. 4.7. On the other hand, removal of $\hat{X}_2$ amplitudes leads to a decrease in the norm of the $\hat{X}_1$ vector and

the negative errors in the polarizability appearing in Fig. 4.7.

6.4.3 External-Space Corrections

As noted earlier, a key aspect of the strong performance of virtual NOs for correlation energies is the use of MP2-based corrections for the contributions of the truncated or "external" virtual space, as given in Eq. (6.9). When considering a similar correction for properties, however, we have explored three options: time-dependent Hartree-Fock (TDHF) and both static and dynamic CC2[41] corrections. In each case, we have used the same MP2-based virtual NO space and computed the correction as the difference between the full-virtual space polarizability and the truncated virtual-NO polarizability, with the CCSD/aDZ results for polarizabilties of H_2O_2 shown in Fig. 4.13.

The CC2-based corrections recover nearly all of the error associated with the uncorrected virtual NO space across a wide range of truncation, with the static correction actually yielding slightly smaller errors than its dynamic counterpart. The less expensive TDHF corrections offer significant improvement over the original virtual NO results, and recover most (ca. 85%) of the error until about 40% (21 orbitals) of the virtual space has been deleted, but it clearly produces overall larger errors than the correlated methods. In addition, the TDHF corrections are potentially problematic because the pole structure of the polarizability naturally follows that of the underlying Hartree-Fock perturbed orbitals rather than that of the correlated wave functions,[43, 44, 45] a criticism that would also hold for purely MP2-

based corrections. On the other hand, the CC2-based corrections are significantly more expensive than TDHF, in part because of their iterative nature and the need to transform two-electron integrals involving three virtual orbitals (with the latter criticism again holding for an MP2-based correction).

6.4.4 Perturbed Natural Orbitals

Given that the purpose of the NOs is to "focus" the important components of the basis for the description of electron correlation effects into a compact space, without consideration of the importance of the basis set for other properties, an alternative approach might be to build a virtual space that explicitly takes such properties into account, such as the inclusion of the perturbed one-electron density in the definition of the space. Instead of diagonalizing the ground-state MP2 density, we diagonalize its gradient with respect to the external electric field, thereby incorporating the effects of the external perturbations. The resulting "occupation numbers" obtained thus carry information about orbital occupancies in the perturbed states. The x^{th} Cartesian component of the perturbed density matrix can be written in terms of spin orbitals as

$$\gamma_{ab}^{x} = \frac{1}{2}\sum_{ijc} P_{+}(a,b)\left[\frac{t_{ij}^{bc}}{D_{ij}^{ac}}\left(P_{-}(a,c)\sum_{d} t_{ij}^{ad}\mu_{cd}^{x} - P_{-}(i,j)\sum_{m} t_{im}^{ac}\mu_{mj}^{x}\right)\right] \tag{6.14}$$

where P_{+} and P_{-} are symmetric and anti-symmetric permutation operators, respectively, and the orbital-energy denominator $D_{ij}^{ac} = \epsilon_i + \epsilon_j - \epsilon_a - \epsilon_c$. (For computational convenience, orbital relaxation terms have been neglected.) To obtain the perturbed NOs and their

corresponding eigenvalues, we take an average of the three cartesian components of the density and diagonalize the result. The eigenvalues consist of both positive and negative values as the perturbed density matrix is not positive definite, and thus we truncate the orbitals based on the absolute values of these eigenvalues.

Fig. 4.14 compares the performance of these perturbed NOs with that of the CMOs and NOs, where the error in the CCSD/aDZ dynamic polarizability of H_2O_2 is plotted as a function of the number of virtual orbitals removed. While the perturbed NOs lower the truncation errors associated with conventional NOs, they still introduce significantly higher errors than the corresponding CMOs. The reason for this underperformance is related to the definition of the perturbed density, which naturally bears strong similarity to its unperturbed counterpart [cf. Eqs. (6.4) and (6.14).] As a result, similar sparsity and orbital energy issues illustrated in Figs. 4.10 and 4.11 arise for the perturbed density just as for the original NOs, indicating that such an approach does not resolve the problem.

6.4.5 The Dipole-Amplitude Criterion

The above results demonstrate that the CMO basis provides the best performance among the various virtual spaces considered here, albeit based on cancellation of errors, but what criterion should be used to determine the truncation level that provides optimal balance between computational cost (most compact virtual space) and accuracy? The CMO orbital energies are one possibility, but they have no direct connection to the properties in question.

Another alternative is the "dipole amplitude", d_a, which is defined for each virtual CMO as

$$d_a \equiv \sum_x \sum_i \frac{(\mu_{ia}^x)^2}{\epsilon_i - \epsilon_a}. \tag{6.15}$$

This expression, which is trivially computed post-Hartree-Fock, is based on Eqs. (6.10) and (6.13), using the fact that the leading contributions to the diagonal elements of the similarty-transformed Hamiltonian are the orbital energies. As we are constructing a phenomenological truncation criterion, we have chosen to neglect the dependence on the field frequency.

Fig. 4.15 plots the values of the dipole amplitude for each CMO, as well as the corresponding error in the CCSD/aDZ dipole-polarizability of H_2O_2 introduced by deleting the $\hat{X}_1$ amplitude associated with a given virtual CMO. There is a clear correlation between the two functions, and one can see that as both the error due to truncation of $\hat{X}_1$ and the value of the dipole length increase sharply once we reach CMOs 27-28 — ca. 50% of the virtual space for this test case. Accordingly, one should stop the truncation of the virtual CMO space once the dipole length values start rising sharply. While the optimal choice of such a threshold will be addressed systematically in subsequent work, our preliminary analyses using H_2O_2, methyloxirane, dimethylallene, and related compounds suggest a cutoff of ca. 3-3.5% yields minimal errors in the polarizability.

6.5 Conclusions

On the basis of the above findings, we conclude that, in the absence of orbital relaxation, virtual NOs are not suited for higher-order property calculations such as dynamic polariz-

abilities, and that the occupation number is not an acceptable criterion for estimating the importance of a virtual orbital for such calculations. Although the use of external space corrections based on CC2 polarizabilities reduces the observed truncation errors, they are too relatively costly to be practical for large molecular systems. Furthermore, the use of perturbed virtual NOs offers only slight improvement as compared to unperturbed virtual NOs. CMOs, on the other hand, provide a much more stable mechanism for reducing the size of the virtual space — with truncation of up to 50% of the orbitals yielding shifts of less than 2% as compared to full-space calculation – but the source of their success lies in a significant cancellation of errors. Although further systematic studies are needed, the dipole amplitude appears to provide a useful threshold for an *a priori* truncation of the CMO virtual space.

6.6 Acknowledgements

This research was supported by a grant (CHE-1465149) from the U.S. National Science Foundation. The authors acknowledge Advanced Research Computing at Virginia Tech for providing computational resources and technical support that have contributed to the results reported within this paper.

Bibliography

[1] P.-O. Löwdin. Quantum theory of many-particle systems. i. physical interpretations by means of density matrices, natural spin-orbitals, and convergence problems in the method of configurational interaction. *Phys. Rev.*, 97(6):1474–1489, 1955.

[2] A. Landau, K. Khistyaev, S. Dolgikh, and A. I. Krylov. Frozen natural orbitals for ionized states within equation-of-motion coupled-cluster formalism. *J. Chem. Phys.*, 132:014109, 2010.

[3] H. R. McAlexander and T. D. Crawford. A comparison of three approaches to the reduced-scaling coupled cluster treatment of non-resonant molecular response properties. *J. Chem. Theory Comp.*, 12(1):209–222, 2016.

[4] I. Shavitt and R. J. Bartlett. *Many-Body Methods in Chemistry and Physics: MBPT and Coupled-Cluster Theory*. Cambridge University Press, Cambridge, 2009.

[5] J. Gauss. The coupled-cluster method. In P.v.R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer III, and P. R. Schreiner, editors, *Encyclopedia of Computational Chemistry*, pages 615–636. John Wiley and Sons, Chichester, 1998.

[6] T. D. Crawford and H. F. Schaefer. An introduction to coupled cluster theory for computational chemists. In K. B. Lipkowitz and D. B. Boyd, editors, *Reviews in Computational Chemistry*, volume 14, chapter 2, pages 33–136. VCH Publishers, New York, 2000.

[7] C. F. Bender and E. R. Davidson. Studies in configuration interaction: The first-row diatomic hydrides. *Phys. Rev.*, 183(1):23–30, 1969.

[8] T. L. Barr and E. R. Davidson. Nature of the configuration-interaction method in *ab initio* calculations. i. ne ground state. *Phys. Rev. A*, 1(3):644–658, 1970.

[9] C. Edmiston and M. Krauss. Pseudonatural orbitals as a basis for the superposition of configurations. i. he_2^+. *J. Chem. Phys.*, 45(5):1833–1839, 1966.

[10] W. Meyer. PNO-CI studies of electron correlation effects. I. Configuration expansion by means of nonorthogonal orbitals, and application of the ground state and ionized states of methane. *J. Chem. Phys.*, 58:1017, 1973.

[11] R. Ahlrichs, H. Lischka, V. Staemmler, and W. Kutzelnigg. Pno-ci (pair natural orbital configuration interaction) and cepa-pno (coupled cluster pair approxiomation with pair natural orbitals) calculations of molecular systems. i. outline of the method for closed-shell states. *J. Chem. Phys.*, 62(4):1225–1234, 1975.

[12] F. Neese, A. Hansen, and D. G. Liakos. Efficient and accurate approximations to the local coupled cluster singles doubles method using a truncated pair natural orbital basis. *J. Chem. Phys.*, 131:064103, 2009.

[13] C. Riplinger, P. Pinski, U. Becker, E. F. Valeev, and F. Neese. Sparse maps — a systematic infrastructure for reduced-scaling electronic structure methods. ii. linear scaling domain based pair natural orbital coupled cluster theory. *J. Chem. Phys.*, 144:024109, 2016.

[14] J. T. Fermann, C. D¿ Sherrill, T. D. Crawford, and H. F. Schaefer. Benchmark studies of electron correlation in six-electron systems. *J. Chem. Phys.*, 100:8132–8139, 1994.

[15] C. D. Sherrill and H. F. Schaefer. The configuration interaction method: Advances in highly correlated approaches. *Adv. Quantum Chem.*, 34:143–269, 1999.

[16] M. L. Abrams and C. D. Sherrill. Natural orbitals as substitutes for optimized orbitals in complete active space wavefunctions. *Chem. Phys. Lett.*, 395:227–232, 2004.

[17] H. Jørgen Aa. Jensen, P. Jørgensen, H. Ågren, and J. Olsen. Second-order møller–plesset perturbation theory as a configuration and orbital generator in multiconfiguration self-consistent field calculations. *J. Chem. Phys.*, 88(6):3834–3839, 1988.

[18] C. Sosa, J. Geertsen, G. W. Trucks, R. J. Bartlett, and J. A. Franz. Selection of the reduced virtual space for correlated calculations. an application to the energy and dipole moment of h_2o. *Chem. Phys. Lett.*, 159(2,3):148–154, 1989.

[19] A. G. Taube and R. J. Bartlett. Frozen natural orbitals: Systematic basis set truncation for coupled-cluster theory. *Coll. Czech. Chem. Commun.*, 70(6):837–850, 2005.

[20] A. G. Taube and R. J. Bartlett. Frozen natural orbital coupled-cluster theory: Forces and application to decomposition of nitroethane. *J. Chem. Phys.*, 128:164101, 2008.

[21] A. E. DePrince and C. D. Sherrill. Accurate noncovalent interaction energies using truncated basis sets based on frozen natural orbitals. *J. Chem. Theory Comput.*, 9:293–299, 2013.

[22] A. E. DePrince and C. D. Sherrill. Accuracy and efficiency of coupled-cluster theory using density fitting/cholesky decomposition, frozen natural orbitals, and a t_1-transformed hamiltonian. *J. Chem. Theory. Comp.*, 9:2687–2696, 2013.

[23] T. Korona, K. Pflüger, and H.-J. Werner. The effect of local approximations in coupled-cluster wave functions on dipole moments and static dipole polarizabilities. *Phys. Chem. Chem. Phys*, 6:2059–2065, 2004.

[24] N. J. Russ and T. D. Crawford. Local correlation in coupled cluster calculations of molecular response properties. *Chem. Phys. Lett.*, 400:104–111, 2004.

[25] N. J. Russ and T. D. Crawford. Local correlation domains for coupled cluster theory: Optical rotation and magnetic field perturbations. *Phys. Chem. Chem. Phys.*, 10:3345–3352, 2008.

[26] H. R. McAlexander, T. J. Mach, and T. D. Crawford. Localized optimized orbitals, coupled cluster theory, and chiroptical response properties. *Phys. Chem. Chem. Phys.*, 14(21):7830–7836, 2012.

[27] J. Friedrich, H. R. McAlexander, A. Kumar, and T. D. Crawford. Incremental evaluation of coupled cluster dipole polarizabilities. *Phys. Chem. Chem. Phys.*, 17:14284–14296, 2015.

[28] T. J. Mach and T. D. Crawford. Computing optical rotation via an n-body approach. *Theor. Chem. Acc.*, 133:1449, 2014.

[29] T. Helgaker, S. Coriani, P. Jørgensen, K. Kristensen, J. Olsen, and K. Ruud. Recent advances in wave function-based methods of molecular -property calculations. *Chem. Rev.*, 112:543–631, 2012.

[30] A. D. Becke. Density-functional thermochemistry. iii. the role of exact exchange. *J. Chem. Phys.*, 98(7):5648–5652, 1993.

[31] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch. *Ab initio* calculation of vibrational absorption and circular dichroism spectra using density functional theory. *J. Phys. Chem.*, 98(45):11623–11627, 1994.

[32] C. Lee, W. Yang, and R. G. Parr. Development of the colle-salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B.*, 37:785–789, 1988.

[33] T. H. Dunning. Gaussian basis sets for use in correlated molecular calculations. i. the atoms boron through neon. *J. Chem. Phys.*, 90:1007, 1989.

[34] R. A. Kendall, T. H. Dunning, and R. J. Harrison. Electron affinities of the first-row atoms revisited. systematic basis sets and wave functions. *J. Chem. Phys.*, 96(9):6796–6806, 1992.

[35] D. E. Woon and T. H. Dunning. Gaussian basis sets for use in correlated molecular calculations. iv. calculation of static electrical response properties. *J. Chem. Phys.*, 100(4):2975–2988, 1994.

[36] G. D. Purvis and R. J. Bartlett. A full coupled-cluster singles and doubles model: The inclusion of disconnected triples. *J. Chem. Phys.*, 76:1910–1918, 1982.

[37] O. Christiansen, P. Jørgensen, and C. Hättig. Response functions from fourier component variational perturbation theory applied to a time-averaged quasienergy. *Int. J. Quantum Chem.*, 68:1–52, 1998.

[38] Justin M. Turney, Andrew C. Simmonett, Robert M. Parrish, Edward G. Hohenstein, Francesco A. Evangelista, J. T. Fermann, Benjamin J. Mintz, Lori A. Burns, Jeremiah J. Wilke, Micah L. Abrams, Nicholas J. Russ, Matthew L. Leininger, Curtis L. Janssen, Edward T. Seidl, Wesley D. Allen, Henry F. Schaefer, Rollin A. King, Edward. F. Valeev, C. David Sherrill, and T. Daniel Crawford. Psi4: An open-source ab initio electronic structure program. *WIREs. Comput. Mol. Sci.*, 2:556–565, 2012.

[39] T. D. Crawford. Reduced-scaling coupled-cluster theory for response properties of large molecules. In P. Carsky, J. Pittner, and J. Paldus, editors, *Recent Progress in Coupled*

Cluster Methods: Theory and Applications, volume 11 of *Challenges and Advances in Computational Chemistry and Physics*, chapter 2, pages 37–55. Springer, Berlin, 2010.

[40] H. Koch, R. Kobayashi, and P. Jørgensen. Brueckner coupled cluster response functions. *Int. J. Quantum Chem.*, 49:835, 1994.

[41] O. Christiansen, H. Koch, and Poul Jørgensen. The second-order approximate coupled cluster singles and doubles model cc2. *Chem. Phys. Lett.*, 243(5-6):409–418, September 1995.

[42] T. B. Pedersen, B. Fernández, and H. Koch. Gauge invariant coupled cluster response theory using optimized nonorthogonal orbitals. *J. Chem. Phys.*, 114(16):6983–6993, 2001.

[43] J. E. Rice and N. C. Handy. The calculation of frequency-dependent polarizabilities and pseudo-energy derivatives. *J. Chem. Phys.*, 94(7):4959–4971, 1991.

[44] C. Hättig and B. A. Heß. Correlated frequency-dependent polarizabilities and dispersion coefficients in the time-dependent second-order møller-plesset approximation. *Chem. Phys. Lett.*, 233:359–370, 1995.

[45] F. Aiga and R. Itoh. Calculation of frequency-dependent polarizability and hyperpolarizabilities by the second-order møller-plesset perturbation theory. *Chem. Phys. Lett.*, 251:372–380, 1996.