10

Multi-Configurational Reference Perturbation Theory with a CASSCF Reference Function

Roland Lindh and Ignacio Fdez. Galván

Department of Chemistry – BMC, Uppsala University, SE-751 23 Uppsala, Sweden

Abstract

The purpose of the present chapter is to give students a detailed introduction to time-independent multi-configurational reference perturbation theory (MRPT), in particular the CASPT2 method, one of the standard tools for the study of excited states of molecular systems of small to intermediate size. To achieve this, in as closed a form as possible, we believe that the chapter should begin with the very basics of Rayleigh–Schrödinger (RS) and Møller–Plesset (MP) perturbation theory (PT) before we present the MRPT. The multi-configurational version of perturbation theory has many features and flaws which are either a direct consequence of the original Rayleigh–Schrödinger formulation, or due to the choice of the zeroth order Hamiltonian. These flaws or peculiarities have to be clearly identified and understood before we proceed toward the more elaborate multi-configurational reference perturbation theories. This is then followed by a section dedicated to the most popular versions of single-state multi-configurational reference perturbation theory. Here we again present the formulation of these methods and also address the various problems they encounter. Multi-state versions of perturbation theory are today based on effective Hamiltonian approximations. Hence, these deserve their own treatment due to the fact that this approach introduces its own kind of features and problems, which need to be addressed separately. At the very end of this chapter we summarize and present an outlook on the subject of MRPT.

Let us briefly, before we start, address the reason behind the need of a perturbation theory based on "multi-configurational reference" functions. Single-configuration reference methods, such as standard coupled-cluster methods and Møller–Plesset perturbation theory, have been successful in modeling molecular systems close to their equilibrium structure. However, at molecular geometries at which such reference functions are qualitatively wrong, as in the case of bond breaking, for example, single-configuration reference perturbation or coupled-cluster theory has shown convergence problems. These approaches frequently produce significant artifacts at such instances. This effect has been demonstrated to be associated with the qualitative incorrectness of the reference functions, from which coupled-cluster or perturbation theory will not be able to recover. Hence, the need for a theory based on reference functions which are qualitatively correct – multi-configurational reference functions. It is not uncommon to find the term "multi-reference" perturbation theory used to mean the same kind of theory. However, it should not be confused with quasi-degenerate perturbation theory, that simultaneously treats

several functions (which can be single- or multi-configurational) as their reference, and we avoid this usage, except when it is part of a specific method's name.

10.1 Rayleigh–Schrödinger Perturbation Theory

This section will cover a brief introduction to perturbation theory as presented originally by Schrödinger [1], who based it on previous work by Lord Rayleigh [2]. We will develop the theory explicitly up to second order, extensions to higher order are trivial. We will introduce the standard equation for the perturbed wave functions and the corresponding energies, but we will will also introduce the so-called $2n + 1$ rule for the efficient computation of energies. This section will be split into two sections: one for the basics of the single-state theory and a section on the issue of convergence properties and intruder states. The former is consistent with the conventional presentation of the subject, while the latter is often not discussed in the depth that students would appreciate. The second section deals with an important problem which arises in both single- and multi-configurational reference PTs. The discussion of the foundations will be resumed with quasi-degenerate perturbation theory after the most significant single-state multi-configurational reference perturbation methods have been presented.

10.1.1 The Single-State Theory

We would first like to point out that the formulation introduced by Schrödinger does not make any explicit reference to the model used for describing wave functions – single configuration, multiple configurations, or something else – but is basically a thought experiment based on the notion that exact solutions (energies and wave functions) are known for a particular Hamiltonian

$$\hat{H}_0 \Psi_i^{(0)} = E_i^{(0)} \Psi_i^{(0)}, \tag{10.1}$$

these energies and wave functions are denoted as unperturbed reference energies and wave functions of state i, the "(0)" is a notation that we will use to indicate the unperturbed entities. We also note that, due to the Hamiltonian being a Hermitian operator, the corresponding eigenfunctions form a set of complete functions, i.e., any other function can be expressed as a linear combination of these eigenfunctions. We will then try to find out to what extent this information can be used to solve the Schrödinger equation for any other Hamiltonian – a perturbed Hamiltonian – which can be expressed as

$$\hat{H} = \hat{H}_0 + \lambda V, \tag{10.2}$$

provided that V, the perturbation potential, is "small". In the context of this chapter, and of RS theory, V is a time-independent perturbation. λ is here introduced as the perturbation parameter, and when it takes the values 0 and 1 this Hamiltonian is equal to the unperturbed Hamiltonian and to the fully perturbed or "target" Hamiltonian, respectively. At this point we note that the theory is independent of what the actual nature of the partitioning and the unperturbed reference functions is, so these are completely arbitrary, apart from the desire that the resulting perturbation expansion be rapidly converging such that the approach will be useful in practice. There will now follow two different ways to derive the theory – the conventional and the variational approach.

10.1.1.1 The Conventional Projectional Derivation

The perturbation parameter is instrumental in our pursuit to make any sense of how the information of the unperturbed system can be of any use. In passing we note that RS perturbation theory is based on a *global* partitioning of the Hamiltonian, something it does not share with

subsequent implementations of perturbation theory. The theory then proceeds with introducing a Taylor expansions of the energy and the wave functions in terms of the perturbation parameter and with the unperturbed energies and wave functions as the reference points of the expansions. That is,

$$E_i = E_i^{(0)} + E_i^{(1)}\lambda + E_i^{(2)}\lambda^2 + \dots \tag{10.3}$$

for the energies and

$$\Psi_i = \Psi_i^{(0)} + \Psi_i^{(1)}\lambda + \Psi_i^{(2)}\lambda^2 + \dots \tag{10.4}$$

for the wave function, where $E_i^{(k)}$ and $\Psi_i^{(k)}$ correspond to the kth-order correction to the perturbed energy and wave function, respectively, as a function of the perturbation. At this time we do not need to discuss what these corrections are, apart the obvious fact that the $E_i^{(k)}$ are the kth-order derivatives of the energy with respect to the perturbation parameter λ. In the following we will make a difference between the nth-order corrected energy and the nth-order correction of the energy. While the latter is $E^{(n)}$, as found in Eq. (10.3), the former is the sum of the unperturbed energy and all corrections up to the nth order:

$$E_i^{[n]} = \sum_{i=0}^{n} E_i^{(i)}. \tag{10.5}$$

The RS theory now proceeds with expressing the Schrödinger equation of the perturbed system in terms of the unperturbed Hamiltonian, energies and wave functions, that is,

$$(\hat{H}_0 + \lambda V)(\Psi_i^{(0)} + \Psi_i^{(1)}\lambda + \Psi_i^{(2)}\lambda^2 + \dots) $$
$$= (E_i^{(0)} + E_i^{(1)}\lambda + E_i^{(2)}\lambda^2 + \dots)(\Psi_i^{(0)} + \Psi_i^{(1)}\lambda + \Psi_i^{(2)}\lambda^2 + \dots) . \tag{10.6}$$

After reordering the terms in the equation we find that

$$(\hat{H}_0\Psi_i^{(0)} - E_i^{(0)}\Psi_i^{(0)})\lambda^0$$
$$+ (\hat{H}_0\Psi_i^{(1)} + V\Psi_i^{(0)} - E_i^{(0)}\Psi_i^{(1)} - E_i^{(1)}\Psi_i^{(0)})\lambda^1$$
$$+ (\hat{H}_0\Psi_i^{(2)} + V\Psi_i^{(1)} - E_i^{(0)}\Psi_i^{(2)} - E_i^{(1)}\Psi_i^{(1)} - E_i^{(2)}\Psi_i^{(0)})\lambda^2$$
$$+ \dots = 0. \tag{10.7}$$

The Schrödinger equation is fulfilled for any value of the perturbation parameter, λ. Hence, this implies that each row of the equation above is independently identical to zero, and we have the following equations explicitly up to second order of the perturbation,

$$\hat{H}_0\Psi_i^{(0)} = E_i^{(0)}\Psi_i^{(0)} \tag{10.8}$$

$$\hat{H}_0\Psi_i^{(1)} + V\Psi_i^{(0)} = E_i^{(0)}\Psi_i^{(1)} + E_i^{(1)}\Psi_i^{(0)} \tag{10.9}$$

$$\hat{H}_0\Psi_i^{(2)} + V\Psi_i^{(1)} = E_i^{(0)}\Psi_i^{(2)} + E_i^{(1)}\Psi_i^{(1)} + E_i^{(2)}\Psi_i^{(0)}, \tag{10.10}$$

and then in general,

$$(\hat{H}_0 - E_i^{(0)})\Psi_i^{(n)} = -V\Psi_i^{(n-1)} + \sum_{k=1}^{n} E_i^{(k)}\Psi_i^{(n-k)}. \tag{10.11}$$

We identify the first equation as the Schrödinger equation of the unperturbed system. The subsequent equations, however, leave us far from our goal – expressions of the corrections to the perturbed energies and wave functions defined completely in terms of entities of the unperturbed system. We note that Eq. (10.11) is invariant to any admixture of the reference state, $\Psi_i^{(0)}$, into the

nth-order correction, $\Psi_i^{(n)}$, since this will leave the the left-hand side (LHS) unchanged. We relieve this ambiguity by the introduction of the *intermediate normalization*:

$$\langle \Psi_i^{(0)} | \Psi_i^{(k)} \rangle = \delta_{0,k} \ . \tag{10.12}$$

The combination of Eqs. (10.11) and (10.12) leads to two important results. First, Eq. (10.11) completely defines the nth-order correction of the wave function from lower-order corrections. Here we note a recursive nature of Eq. (10.11), a non-recursive variant would lead to an expression in terms of the Bloch equation – more about this later. Second, the exact energy can be computed as

$$E_i = \langle \Psi_i^{(0)} | \hat{H} | \Psi_i \rangle \ . \tag{10.13}$$

Equations (10.9) to (10.11), and so forth, can now be projected from the LHS and integrated with the unperturbed wave functions. We start by projecting Eq. (10.11) with $\Psi_i^{(0)}$ this gives

$$\langle \Psi_i^{(0)} | (\hat{H}_0 - E_i^{(0)}) | \Psi_i^{(n)} \rangle = -\langle \Psi_i^{(0)} | V | \Psi_i^{(n-1)} \rangle + \sum_{k=1}^{n} \langle \Psi_i^{(0)} | E_i^{(k)} | \Psi_i^{(n-k)} \rangle \ . \tag{10.14}$$

Using the Hermiticity $\hat{H}_0$ and the intermediate normalization this expression reduces to

$$E_i^{(n)} = \langle \Psi_i^{(0)} | V | \Psi_i^{(n-1)} \rangle \ . \tag{10.15}$$

This seems to mean that we need the $(n-1)$th-order correction to the wave function in order to compute the nth-order correction to the energy. This is, however, far from true and will be discussed below.

Let us now return to the corrections of the wave function and Eq. (10.11). We will rearrange the equation to form our general vehicle for the computation of corrections to the wave function. We first note that the equation is devoid of any component of $\Psi_i^{(0)}$. This property was actually used when going from Eq. 10.14 to Eq. 10.15, since

$$\begin{aligned} \langle \Psi_i^{(0)} | (\hat{H}_0 - E_i^{(0)}) | \Psi_i^{(n)} \rangle &= \langle \Psi_i^{(0)} | \hat{H}_0 | \Psi_i^{(n)} \rangle - \langle \Psi_i^{(0)} | E_i^{(0)} | \Psi_i^{(n)} \rangle \\ &= E_i^{(0)} \delta_{0,n} - E_i^{(0)} \delta_{0,n} = 0. \end{aligned} \tag{10.16}$$

This means that we can now multiply Eq. (10.11) with the inverse of $(\hat{H}_0 - E_i^{(0)})$ avoiding any singularity problems. We get

$$\Psi_i^{(n)} = (\hat{H}_0 - E_i^{(0)})^{-1} \left(-V \Psi_i^{(n-1)} + \sum_{k=1}^{n} E_i^{(k)} \Psi_i^{(n-k)} \right) \ . \tag{10.17}$$

Knowing that the correction and any term of Eq. (10.15) are devoid of $\Psi_i^{(0)}$, we introduce the projection operator

$$\hat{Q} = 1 - |\Psi_i^{(0)}\rangle\langle\Psi_i^{(0)}| \ . \tag{10.18}$$

We can now simplify Eq. (10.17) without any approximation and get

$$\Psi_i^{(n)} = \hat{Q}(\hat{H}_0 - E_i^{(0)})^{-1} \hat{Q} \left(-V \Psi_i^{(n-1)} + \sum_{k=1}^{n-1} E_i^{(k)} \Psi_i^{(n-k)} \right) \ . \tag{10.19}$$

Note that we have reduced the summation range, the projection operator will now, whenever needed, cancel any appearance of the reference wave function. Let us pause at this point and see what we have accomplished. We have first derived our master equation (10.11). From this equation we have derived a general expression for the correction to the energy, see Eq. (10.15), in which the nth-order correction to the energy requires the $(n-1)$th-order correction to the wave function.

Finally, in Eq. (10.19) we now have a recursive approach in which corrections to the wave function can be computed. We also note that in RSPT the corrected energy is neither an eigenvalue nor an expectation value, and the corrected wave function is not an eigenfunction to the Hamiltonian.

Before we complete this section let us return to our dangling question: "To which order do we need to know the correction of the wave function if we want to compute the *n*th-order correction to the energy?" From what we have seen so far the answer seems to be $(n - 1)$. However, as noted before, this is wrong. Consider that the first- and second-order corrections to the wave function, according to Eq. (10.19), can be expressed as

$$\Psi_i^{(1)} = -\hat{Q}(\hat{H}_0 - E_i^{(0)})^{-1}\hat{Q}V\Psi_i^{(0)}, \tag{10.20}$$

and

$$\Psi_i^{(2)} = \hat{Q}(\hat{H}_0 - E_i^{(0)})^{-1}\hat{Q}(-V + E_i^{(1)})\Psi_i^{(1)}. \tag{10.21}$$

We can then, for example, write the third-order correction of the energy as

$$\begin{aligned} E_i^{(3)} &= \langle \Psi_i^{(0)} | V | \Psi_i^{(2)} \rangle \\ &= \langle \Psi_i^{(0)} | V\hat{Q}(\hat{H}_0 - E_i^{(0)})^{-1}\hat{Q}(-V + E_i^{(1)}) | \Psi_i^{(1)} \rangle \\ &= \langle \Psi_i^{(1)} | V - E_i^{(1)} | \Psi_i^{(1)} \rangle \ . \end{aligned} \tag{10.22}$$

This clearly demonstrates that the third-order energy can be computed from the first-order corrections to the wave function, a first demonstration of Wigner's $2n + 1$ rule. To develop this in more depth is beyond the scope of this chapter. For those students interested in this we recommend the excellent book by Helgaker and co-workers [3], where this particular issue is analyzed with a variational Lagrangian approach.

Up to this point we have not been specific about the corrections to the wave function, other than the fact that we use the intermediate normalization. At this point we will make an ansatz based on our observation that the corrections are orthogonal to the reference functions, that is, we will take the set of unperturbed eigenfunctions as a basis set,

$$\Psi_i^{(1)} = \sum_{j \neq i} c_{ji} \Psi_j^{(0)}. \tag{10.23}$$

Using our master equation, Eq. (10.11), replacing the *n*th-order correction with the previous ansatz, left-multiplying with $\langle \Psi_l^{(0)} |$ and finally integrating we get

$$\sum_{j \neq i} c_{ji} \langle \Psi_l^{(0)} | (\hat{H}_0 - E_i^{(0)}) | \Psi_j^{(0)} \rangle = -\langle \Psi_l^{(0)} | V | \Psi_i^{(0)} \rangle \ , \tag{10.24}$$

which simplifies to

$$c_{li} = \frac{\langle \Psi_l^{(0)} | V | \Psi_i^{(0)} \rangle}{E_i^{(0)} - E_l^{(0)}} \ . \tag{10.25}$$

Following Eq. (10.15) and substituting the first-order correction as described above, we can formulate the second-order correction to the energy as

$$E_i^{(2)} = \sum_{j \neq i} \frac{\langle \Psi_i^{(0)} | V | \Psi_j^{(0)} \rangle \langle \Psi_j^{(0)} | V | \Psi_i^{(0)} \rangle}{E_i^{(0)} - E_j^{(0)}} \ . \tag{10.26}$$

At this point we stop any further analysis. There are several reasons for this. Two dominating ones are that the expressions get more and more complicated and cumbersome, but also, and maybe more importantly, perturbation theory beyond second-order corrections to the energy has not been

found to constitute a significant enough improvement to compensate the additional complexity and computational expense. The curious readers can either derive the next set of equations themselves or find them published elsewhere.

10.1.1.2 The Bi-Variational Approach

In the above conventional approach we projected the Schrödinger equation from the left with the reference wave function, $\Psi_i^{(0)}$. This introduces an asymmetry which is in particular manifested by the shape of the equation in which we compute corrections to the energy (see Eq. 10.15). In this section we will explore an alternative derivation which will be in accordance with Wigner's $2n + 1$ rule. To explore this we start from an expression which is originally symmetric with respect to its bra and ket side. For convenience, in the rest of this section we will omit the subindex i in the wave functions and energies, since all expressions refer to a single state. In order to maintain the symmetry we will use a bi-functional expression of a generalization of the expectation value for the energy, that is

$$\mathcal{E}(\lambda, \bar{\Psi}, \tilde{\Psi}) = \langle \bar{\Psi} | \hat{H}(\lambda) | \tilde{\Psi} \rangle \,, \tag{10.27}$$

which we note is equivalent to the variational expression of the exact energy,

$$\mathcal{E}(\lambda, \Psi, \Psi) = E(\lambda, \Psi),$$

given an arbitrary actual value of the perturbational parameter, and if the normalized wave functions $\bar{\Psi}$ and $\tilde{\Psi}$ are equivalent to the exact eigenfunction, Ψ, of the Hamiltonian $\hat{H}(\lambda)$, in which case we can also write

$$\langle \Psi | \hat{H}(\lambda) - \mathcal{E}(\lambda, \Psi, \Psi) | \Psi \rangle = 0. \tag{10.28}$$

We note that, $\mathcal{E}(\lambda, \Psi, \Psi)$, the exact energy, is at the global minimum and the gradient, with respect to any variation of the wave function, is zero:

$$\nabla_{\bar{\Psi}} \mathcal{E} |_{\bar{\Psi}=\Psi} = 0$$
$$\nabla_{\tilde{\Psi}} \mathcal{E} |_{\tilde{\Psi}=\Psi} = 0. \tag{10.29}$$

Since this holds for any value of the perturbational parameter, we also have that the derivatives must vanish order by order, for $k > 0$:

$$\nabla_{\bar{\Psi}} \mathcal{E}^{(n)} |_{\bar{\Psi}=\Psi} = 0 \qquad \nabla_{\bar{\Psi}^{(k)}} \mathcal{E}^{(n)} |_{\bar{\Psi}=\Psi} = 0$$
$$\nabla_{\tilde{\Psi}} \mathcal{E}^{(n)} |_{\tilde{\Psi}=\Psi} = 0 \qquad \nabla_{\tilde{\Psi}^{(k)}} \mathcal{E}^{(n)} |_{\tilde{\Psi}=\Psi} = 0. \tag{10.30}$$

These conditions of vanishing gradients for the variational parameters will help us later in simplifying the equations we derive. From now on, we will consider the case where $\bar{\Psi} = \tilde{\Psi} = \Psi$, but we will continue to make a difference between the two functions, $\bar{\Psi}$ and $\tilde{\Psi}$, although we set them to be identical, Ψ.

We proceed from Eq. (10.28) and replace Ψ with $\Psi^{(0)} + \hat{Q}\Psi$

$$\langle \Psi^{(0)} | \hat{H} - \mathcal{E} | \Psi^{(0)} \rangle + \langle \Psi^{(0)} | \hat{H} - \mathcal{E} | \hat{Q}\Psi \rangle$$
$$+ \langle \hat{Q}\Psi | \hat{H} - \mathcal{E} | \Psi^{(0)} \rangle + \langle \hat{Q}\Psi | \hat{H} - \mathcal{E} | \hat{Q}\Psi \rangle = 0. \tag{10.31}$$

After expanding $\hat{H} = \hat{H}_0 + \lambda V$ and some rearrangement, noting that $\langle \Psi^{(0)} | Q\Psi \rangle = 0$, we have

$$\mathcal{E}(\lambda, \Psi, \Psi) = \langle \Psi^{(0)} | \hat{H}_0 | \Psi^{(0)} \rangle + \lambda \langle \Psi^{(0)} | V | \Psi^{(0)} \rangle$$
$$+ \lambda \langle \Psi^{(0)} | V | \hat{Q}\Psi(\lambda) \rangle + \lambda \langle \hat{Q}\Psi(\lambda) | V | \Psi^{(0)} \rangle$$
$$+ \langle \hat{Q}\Psi(\lambda) | \hat{H}_0 + \lambda V - \mathcal{E}(\lambda, \dots) | \hat{Q}\Psi(\lambda) \rangle \,, \tag{10.32}$$

where the third and fourth terms have not been merged together to reflect the difference of the bra and ket wave functions. It could be argued that nothing has been gained by the introduction of this equation, compared with a conventional approach using a projection with the reference wave function on the Schrödinger equation. In particular, the exact energy appears on both sides of the equation and it is more complicated. However, there is a significant important advantage with this approach as will become evident soon. Moreover, an analysis order by order shows that the contribution from the energy on the right-hand side (RHS) will only depend on lower-order corrections to the energy than the LHS and will not involve the actual value of the exact energy. That is, at the stationary point at which $\mathcal{E} = E_{\text{exact}}$, we have that

$$
\begin{aligned}
\mathcal{E}^{(n)} = E^{(n)} = {} & \delta_{n0}E^{(0)} + \sum_{k=0}^{n-1} \langle \Psi^{(k)}|V|\Psi^{(n-1-k)} \rangle \\
& + \sum_{k=1}^{n-1} \langle \Psi^{(k)}|\hat{H}_0 - E^{(0)}|\Psi^{(n-k)} \rangle \\
& - \sum_{k=1}^{n-2} E^{(k)} \sum_{l=1}^{n-k-1} \langle \Psi^{(l)}|\Psi^{(n-k-l)} \rangle \,.
\end{aligned}
\tag{10.33}
$$

In particular, we note that this formulation is completely symmetric with respect to the treatment of the bra and ket sides. At this point we stop using the explicit reference to $\mathcal{E}$ and will just use E. However, we should keep in mind that when we differentiate we do that on $\mathcal{E}$! To continue, Eq. (10.33) now trivially provides us with expressions for the corrected energy up to first order which are identical to those obtained from the projectional derivation,

$$
E^{(0)} = \langle \Psi^{(0)}|\hat{H}_0|\Psi^{(0)} \rangle
\tag{10.34}
$$

$$
E^{(1)} = \langle \Psi^{(0)}|V|\Psi^{(0)} \rangle \,,
\tag{10.35}
$$

However, at second order things take a different direction, we get

$$
E^{(2)} = \langle \Psi^{(0)}|V|\Psi^{(1)} \rangle + \langle \Psi^{(1)}|V|\Psi^{(0)} \rangle + \langle \Psi^{(1)}|\hat{H}_0 - E^{(0)}|\Psi^{(1)} \rangle
\tag{10.36}
$$

instead of (cf. Eq. 10.15) just $E^{(2)} = \langle \Psi^{(0)}|V|\Psi^{(1)} \rangle$! For the third-order correction we do not get $E^{(3)} = \langle \Psi^{(0)}|V|\Psi^{(2)} \rangle$, but

$$
\begin{aligned}
E^{(3)} = {} & \langle \Psi^{(0)}|V|\Psi^{(2)} \rangle + \langle \Psi^{(1)}|V|\Psi^{(1)} \rangle \\
& + \langle \Psi^{(2)}|V|\Psi^{(0)} \rangle + \langle \Psi^{(1)}|\hat{H}_0 - E^{(0)}|\Psi^{(2)} \rangle \\
& + \langle \Psi^{(2)}|\hat{H}_0 - E^{(0)}|\Psi^{(1)} \rangle - E^{(1)}\langle \Psi^{(1)}|\Psi^{(1)} \rangle \,.
\end{aligned}
\tag{10.37}
$$

What is going on? Well, remember that these expressions are variational. Taking the second-order energy correction, for example, differentiating with respect to $\bar{\Psi}^{(1)}$, and equating to zero we find that

$$
\nabla_{\bar{\Psi}^{(1)}} E^{(2)} = V|\Psi^{(0)} \rangle + (\hat{H}_0 - E^{(0)})|\Psi^{(1)} \rangle = 0,
\tag{10.38}
$$

and similarly for $\tilde{\Psi}^{(1)}$. That allows us to reduce Eq. (10.36) in two different ways getting,

$$
E^{(2)} = \langle \Psi^{(0)}|V|\Psi^{(1)} \rangle = \langle \Psi^{(1)}|V|\Psi^{(0)} \rangle \,.
\tag{10.39}
$$

To continue, for the third-order correction to the energy, we find that the gradient reveals the following conditions,

$$
\nabla_{\bar{\Psi}^{(1)}} E^{(3)} = (\hat{H}_0 - E_0)|\Psi^{(2)} \rangle + (V - E^{(1)})|\Psi^{(1)} \rangle = 0
\tag{10.40}
$$

$$
\nabla_{\bar{\Psi}^{(2)}} E^{(3)} = V|\Psi^{(0)} \rangle + (\hat{H}_0 - E^{(0)})|\Psi^{(1)} \rangle = 0,
\tag{10.41}
$$

and the equivalent equations for the bra side. Here we are left with some ambiguity: Which of the relations should we use to eliminate redundancies and in which order? We will, of course, use the conditions that eliminate the higher-order corrections first, and continue until we have exhausted all possibilities. For the third-order energy correction, for example, we note that if we use one of the last conditions (e.g., for the bra) first and then the corresponding one of the first conditions, we will gain nothing – we will simply regenerate the Eq. (10.15) expression, albeit in a very clumsy way. However, using the two last conditions we compile the third-order correction to the energy as

$$E^{(3)} = \langle \Psi^{(1)} | V - E^{(1)} | \Psi^{(1)} \rangle \,, \tag{10.42}$$

where we have completely eliminated $\Psi^{(2)}$ from the expression!

We will now generalize this procedure to reduce Eq. (10.33), for which we examine all elements of the form

$$\langle \Psi^{(k)} | \hat{O} | \Psi^{(l)} \rangle \,,$$

that is,

$$\langle \Psi^{(k)} | V | \Psi^{(n-1-k)} \rangle$$
$$\langle \Psi^{(k)} | \hat{H}_0 - E^{(0)} | \Psi^{(n-k)} \rangle$$
$$\langle \Psi^{(m)} | \Psi^{(n-k-m)} \rangle \,.$$

To tackle this reduction let us rewrite Eq. (10.33) in a grid form, representing the $\langle \Psi^{(k)} | \hat{O} | \Psi^{(l)} \rangle$ terms. The indices k and l will be used to label columns and rows, and each cell contains the associated operator $\hat{O}$ (empty cells are zero). We will display cells in green if they survive the reduction process.

Starting with $n = 1$, we get

l \ k	0
0	V

where the only cell represents $\langle \Psi^{(0)} | V | \Psi^{(0)} \rangle$. To proceed with $n = 2$, we have

l \ k	0	1
0		V
1	V	$\hat{H}_0 - E^{(0)}$

The gradient conditions are now seen as full rows or columns, and we can eliminate one of them, indicated with a zigzag strike-out. The result corresponds to $\langle \Psi^{(0)} | V | \Psi^{(1)} \rangle$. For $n = 3$ we again see the gradient conditions as rows and columns, and we can strike out $k = 2$ and $l = 2$:

l \ k	0	1	2
0			V
1		$V - E^{(1)}$	$\hat{H}_0 - E^{(0)}$
2	V	$\hat{H}_0 - E^{(0)}$	

It is only allowed to strike out a row or column if it contains no non-empty cell that has already been stricken out. We could have stricken out $k = 2$ and $k = 1$ to generate $\langle \Psi^{(0)} | V | \Psi^{(2)} \rangle$, but instead we have here obtained $\langle \Psi^{(1)} | V - E^{(1)} | \Psi^{(1)} \rangle$, which requires only first-order entities (cf. Eq. 10.22). Similarly, for $n = 4$

l \ k	0	1	2	3
0				V
1		$-E^{(2)}$	$V - E^{(1)}$	$\hat{H}_0 - E^{(0)}$
2		$V - E^{(1)}$	$\hat{H}_0 - E^{(0)}$	
3		V	$\hat{H}_0 - E^{(0)}$	

and for $n = 5$

l \ k	0	1	2	3	4
0					V
1		$-E^{(3)}$	$-E^{(2)}$	$V - E^{(1)}$	$\hat{H}_0 - E^{(0)}$
2		$-E^{(2)}$	$V - E^{(1)}$	$\hat{H}_0 - E^{(0)}$	
3		$V - E^{(1)}$	$\hat{H}_0 - E^{(0)}$		
4	V	$\hat{H}_0 - E^{(0)}$			

As for $n = 3$, we could eliminate all columns $k > 0$ (or rows $l > 0$), and we would end up with Eq. (10.15). But in order to eliminate higher-order entities, we proceed by eliminating the rows and columns with the highest k and l indices until elimination is impossible. By inspection we find that

$$E^{(4)} = \langle \Psi^{(2)} | V - E^{(1)} | \Psi^{(1)} \rangle - E^{(2)} \langle \Psi^{(1)} | \Psi^{(1)} \rangle \tag{10.43}$$

$$E^{(5)} = \langle \Psi^{(2)} | V - E^{(1)} | \Psi^{(2)} \rangle - 2E^{(2)} \langle \Psi^{(1)} | \Psi^{(2)} \rangle - E^{(3)} \langle \Psi^{(1)} | \Psi^{(1)} \rangle . \tag{10.44}$$

Thus, a general grid representation gives us the following table.

l \ k	0	1	2	$\cdots$	i	$\cdots$	$n-3$	$n-2$	$n-1$
0									V
1		$-E^{(n-2)}$	$-E^{(n-3)}$		$-E^{(n-i-1)}$		$-E^{(2)}$	$V - E^{(1)}$	$\hat{H}_0 - E^{(0)}$
2		$-E^{(n-3)}$	$-E^{(n-4)}$		$-E^{(n-i-2)}$		$V - E^{(1)}$	$\hat{H}_0 - E^{(0)}$	
$\vdots$									
i		$-E^{(n-i-1)}$	$-E^{(n-i-2)}$		$V - E^{(1)}$				
$\vdots$									
$n-3$		$-E^{(2)}$	$V - E^{(1)}$						
$n-2$		$V - E^{(1)}$	$\hat{H}_0 - E^{(0)}$						
$n-1$	V	$\hat{H}_0 - E^{(0)}$							

First, it is obvious from the elimination process that all rows and columns that contain the operator $\hat{H}_0 - E^{(0)}$ can be ignored. At the end of the elimination process we will be left with a block of elements of the size $m \times (m-1)$ if $n = 2m$, the even case, or $m \times m$ if $n = 2m + 1$, the odd case – the green blocks in all table representations above. We note that in all terms but one the operator is of the nature $-E^{(o)}$ and that o has a value such that, together with the order indices of the wave function corrections, adds up to n $(k + l + o = n)$. The remaining element is

$$\langle \Psi^{(m)} | V - E^{(1)} | \Psi^{(m+i-1)} \rangle$$

if we generalize the odd and even case to $n = 2m + i$, where i is 0 or 1 for the even and odd cases, respectively. Summing up the remaining elements row-/column-wise we derive the following general expression:

$$n = 2m + i, \quad i \in \{0, 1\}$$

$$E^{(n)} = \langle \Psi^{(m)} | V | \Psi^{(m+i-1)} \rangle - \sum_{k=1}^{m} \sum_{l=1}^{m+i-1} E^{(n-k-l)} \langle \Psi^{(k)} | \Psi^{(l)} \rangle \,. \tag{10.45}$$

This expression can be found in ref. [3] – in a derivation based on a variational analysis of a Lagrange multiplier approach – albeit a different order of summation. To conclude this section, the primary benefit of this formulation was the derivation of a variational expression for the second-order correction to the total energy – a feature some developers use for finding the correction in an efficient way. Additionally, if higher-order corrections are to be computed this formalism offers an optimal procedure for that purpose.

10.1.2 Convergence Properties and Intruder States

The use of perturbation theory seems at first glance to be a rock-solid approach that would never fail. However, this is unfortunately not the case. Here we will discuss the convergence properties of perturbation theory to some extent and the associated issue of intruder states. This short section will not only clarify the issue but will also suggest a remedy to correct the problem. The presentation here, based on Kato's work [4], will be simplified and brief. A longer and more detailed presentation can be found in refs. [5] and [3].

In the text that follows we will address the issue of the convergence properties of RSPT. We will do the core analysis based on a simple two-state case – a study of a multi-state case would make the analysis complicated and messy. We will initially do that on the assumption that the perturbation strength parameter λ takes only real values. This will give us insight into how the relative values of the energy gap between the two states and the coupling element play a fundamental role in determining whether or not the perturbation series of a particular partitioning will converge. We will then introduce a generalization of this by allowing the perturbation strength to take complex values – this will improve and clarify the analysis. In combination with this we will introduce the notion of degeneracies, which can be characterized as the origin of the so-called *front-* or *back-door intruder states*. Finally, we will, in the two-state model, introduce a *gap-shift parameter* which will let us model and modify the conditions of perturbation expansion convergence and the appearance of intruder states. The gap-shift technique will also be the inspiration for techniques to improve the convergence and remove intruder states.

Assume that we, in a two-state model, partition the Hamiltonian as follows:

$$\boldsymbol{H} = \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix} + \lambda \begin{pmatrix} 0 & \delta \\ \delta & 0 \end{pmatrix} \,, \tag{10.46}$$

where α and β represent the eigenvalues of the zeroth-order states, δ corresponds to the interaction term between the reference states over the perturbation potential, and λ is the standard perturbational strength parameter. For this trivial case we can compute the exact energies of the final states as

$$E_\pm = \frac{\alpha + \beta}{2} \pm \frac{\sqrt{(\beta - \alpha)^2 + 4\lambda^2\delta^2}}{2} \ . \tag{10.47}$$

We rewrite this for the ground state as

$$E = \frac{\alpha + \beta}{2} - \frac{|\beta - \alpha|}{2}\sqrt{1 + \frac{4\lambda^2\delta^2}{(\beta - \alpha)^2}}, \tag{10.48}$$

and note that the way forward to reshape this into an RS summation order by order goes via the Taylor expansion of $\sqrt{1 + x}$ for which we know that the expansion converges if $|x| < 1$. This leads, as a requirement for convergence, to the inequality

$$\frac{4\delta^2}{(\beta - \alpha)^2} < 1 \tag{10.49}$$

and we get the following final convergence condition:

$$|\beta - \alpha| > 2|\delta|, \tag{10.50}$$

that is, the energy gap between the states should be larger than twice the interaction between the unperturbed states over the perturbation potential to guarantee convergence in a perturbational approach. This is now a tool that will assist us in using information from the unperturbed states. Clearly, it suggests that we have a serious problem if two unperturbed states are close to degenerate.

To proceed, we will also have to analyze what happens when we have a degeneracy for the perturbed states. To do this we will use the technique of Kato [4]. That is, we follow the same procedure as above with the two new ingredients, (i) λ is allowed to take complex values, $\lambda = \lambda_R + i\lambda_I$, and (ii) we identify values for λ at which the two states are degenerate, $E_i(\lambda_R + i\lambda_I) = E_j(\lambda_R + i\lambda_I)$. In passing it is noted that these degeneracies occur in conjugate pairs (λ, λ^*).

In this analysis it is found that the convergence of the perturbation series depends on the behavior of the energies $E_i(\lambda)$ inside the unit circle $|\lambda| \leq 1$ in the complex plane. That is, if the value of λ that causes a degeneracy in the perturbed states represents a point inside the unit circle, we have a convergence issue and the state that becomes degenerate with the reference state is called an intruder state. The intruder states are divided into two classes, (i) front-door intruders ($\lambda_R > 0$), which in real life tend to correspond to low-excitation states, and (ii) back-door intruders ($\lambda_R < 0$), corresponding to high-excitation states. The difference between the front- and back-door intruders is manifested in different convergence patterns of the energy (oscillating versus alternating), change in reference weight (slow versus abrupt), etc.

To get a better understanding of this situation an alternative partitioning is utilized,

$$\boldsymbol{H} = \begin{pmatrix} \alpha & 0 \\ 0 & \beta + \gamma \end{pmatrix} + \lambda \begin{pmatrix} 0 & \delta \\ \delta & -\gamma \end{pmatrix} \ , \tag{10.51}$$

where $\gamma \in \mathbb{R}$ is the gap-shift parameter. The solutions in this case are given by

$$E_\pm = \frac{\alpha + \beta + (1 - \lambda)\gamma}{2} \pm \frac{\sqrt{[|\beta - \alpha| + (1 - \lambda)\gamma]^2 + 4\lambda^2\delta^2}}{2} \ . \tag{10.52}$$

We note that this particular partitioning is such that at $\lambda = 1$ the energy does not depend on the gap-shift parameter. Equation (10.52) has a degeneracy (as can be found by setting the expression

in the square root to zero) for a value of the perturbation strength parameter of

$$\lambda_{\pm} = \frac{|\beta - \alpha| + \gamma}{4\delta^2 + \gamma^2}(\gamma \pm \mathrm{i}2\delta),$$

(10.53)

and if $|\lambda_{\pm}| < 1$ we have a convergence problem and a corresponding intruder state. We can twist this analysis around and reformulate it to answer the question "What is the minimal gap-shift parameter which will make the perturbation expansion converge?", arriving at the result

$$\gamma_c = \frac{4\delta^2 - (\beta - \alpha)^2}{2|\beta - \alpha|} \ .$$

(10.54)

Alternatively, we can paraphrase Eq. (10.50) to see the influence of the gap-shift on the original convergence condition, that is

$$2|\delta| < |\beta - \alpha|\sqrt{1 + \frac{2\gamma}{|\beta - \alpha|}}.$$

(10.55)

To conclude the analysis, convergence is not an obvious property of the perturbation expansion. The relative sizes of energy differences and interactions over the perturbation are of significance for the convergence condition. We find that when convergence breaks down this is associated with front- or back-door intruder states. It was finally demonstrated that the convergence issue could be managed by introducing a gap-shift parameter which, with the appropriate value, will guarantee convergence of the perturbation expansion. We will now continue with a presentation of two methods that will solve convergence issues in perturbation theory (as implemented in the CASPT2 method): the real and imaginary shift methods – both draw their inspiration from the gap-shift parameter trick.

10.1.2.1 Real and Imaginary Shift Techniques

In the gap-shift technique suggested in association with multi-state situations the partitioning of the Hamiltonian is established as

$$\hat{H} = (\hat{H}_0 + \epsilon\hat{Q}) + (V - \epsilon\hat{Q}),$$

(10.56)

where ϵ is the gap-shift (also referred to by some as a level-shift) parameter – a small real number. With this partitioning we get that the energy corrected up to first order is invariant to the gap-shift parameter,

$$\begin{aligned}
E^{[1]} &= E^{(0)} + E^{(1)} \\
&= \langle\Psi^{(0)}|\hat{H}_0 + \epsilon\hat{Q}|\Psi^{(0)}\rangle + \langle\Psi^{(0)}|V - \epsilon\hat{Q}|\Psi^{(0)}\rangle \\
&= \langle\Psi^{(0)}|\hat{H}_0|\Psi^{(0)}\rangle + \langle\Psi^{(0)}|V|\Psi^{(0)}\rangle \ .
\end{aligned}$$

(10.57)

However, after this we will start to see effects from the gap-shift parameter, for example, the coefficients of the first-order correction to the wave function under the influence of the gap-shift parameter are computed as (see Eq. 10.25)

$$\tilde{c}_{ji} = -\frac{\langle\Psi_j^{(0)}|V|\Psi_i^{(0)}\rangle}{E_j^{(0)} - E_i^{(0)} + \epsilon} = \frac{\langle\Psi_j^{(0)}|V|\Psi_i^{(0)}\rangle}{E_i^{(0)} - E_j^{(0)} - \epsilon} \ ,$$

(10.58)

and the second-order energy correction, which also depends on the gap-shift parameter, is expressed as

$$\tilde{E}_i^{(2)} = \sum_{j\neq i} \frac{\langle\Psi_i^{(0)}|V|\Psi_j^{(0)}\rangle\langle\Psi_j^{(0)}|V|\Psi_i^{(0)}\rangle}{E_i^{(0)} - E_j^{(0)} - \epsilon} \ ,$$

(10.59)

where we indicate with a tilde that these quantities are obtained in the presence of the gap shift.

We will try to remove the influence of the gap-shift parameter as if there was no intruder-state problem. This dependence is removed by the following procedure and approximation. First we note that we can, after some struggle, rewrite the level-shifted second-order correction to the energy as

$$\tilde{E}_i^{(2)} = E_i^{(2)} + \epsilon \sum_{j \neq i} \tilde{c}_{ji}^2 \left(1 - \frac{\epsilon}{E_i^{(0)} - E_j^{(0)}} \right) , \qquad (10.60)$$

which under the assumption $|E_j^{(0)} - E_i^{(0)}| \gg |\epsilon|$ (equivalent to not having an intruder-state problem) gives us the approximation

$$E_i^{(2)} \approx \tilde{E}_i^{(2)} - \epsilon \left(\frac{1}{\tilde{\omega}_i} - 1 \right) \equiv E_i^{(2),\text{LS}} , \qquad (10.61)$$

where $\tilde{\omega}_i = \left(\sum_{j \neq i} \tilde{c}_{ji}^2 + 1 \right)^{-1}$ is the weight of the reference function in the normalized expansion of the wave function including the first-order correction under the influence of the gap-shift parameter. As can be seen from the equation, in the absence of any intruder state $\tilde{E}_i^{(2)}$ is linear with respect to the gap-shift parameter, and $E_i^{(2),\text{LS}}$, the level-shift (LS) corrected second-order energy, is, to first order in ϵ, equal to the result of the original approach. Hence, the technique of gap shifting can be used in general, since it will not significantly affect the result when there is no convergence problem. For a detailed description of this type of a real-shift technique and results of using it can be found in the work by Roos and Andersson [6]. In particular, they have demonstrated that the potential energy surface away from the molecular structures at which singularities arise is unaffected by the gap-shift approach and the use of the level-shifted corrected values for the second-order energy corrections gives results which are much more independent of the size of the gap-shift parameter. See fig. 10.1 for a case in which the method is instrumental in avoiding intruder states.

This work inspired Forsberg and Malmqvist [7] to find alternative ways to get a second-order corrected energy which was even more independent of the size of the gap-shift parameter. Forsberg and Malmqvist introduced the imaginary gap-shift parameter, $i\epsilon$, to investigate this. It should be mentioned that the idea of an imaginary shift had been explored just a year earlier [8]. Similar to the case of a real-valued gap-shift parameter we will have that the energy up to first order is unaffected by the perturbation, the parameter shows up again in the second-order correction to the energy and elimination will have to be made in a fashion different from what was demonstrated above. We get trivially that the second-order energy correction is

$$\tilde{E}_i^{(2)} = \sum_{j \neq i} \frac{\langle \Psi_i^{(0)}|V|\Psi_j^{(0)} \rangle \langle \Psi_j^{(0)}|V|\Psi_i^{(0)} \rangle (E_i^{(0)} - E_j^{(0)} - i\epsilon)}{(E_i^{(0)} - E_j^{(0)})^2 + \epsilon^2} , \qquad (10.62)$$

where we are interested in the real part and want to explicitly eliminate the dependence of the strength of the gap-shift parameter. This would be trivial if the calculation was performed with complex arithmetic, but that is seldom the case and something which we try to avoid. To our assistance comes the formalism we derived using the bi-functional variational technique to derive the energy equations order by order. In particular, we have that the second-order energy correction (Eq. 10.36), can be expressed as (now with the modified $\hat{H}_0$ and V terms)

$$\begin{aligned}
\mathcal{E}_i^{(2)}(\bar{\Psi}_i = \Psi_i, \tilde{\Psi}_i = \Psi_i) \\
= \langle \Psi_i^{(0)}|V|\Psi_i^{(1)} \rangle + \langle \Psi_i^{(1)}|V|\Psi_i^{(0)} \rangle + \langle \Psi_i^{(1)}|\hat{H}_0 + i\epsilon - E_i^{(0)}|\Psi_i^{(1)} \rangle ,
\end{aligned} \qquad (10.63)$$

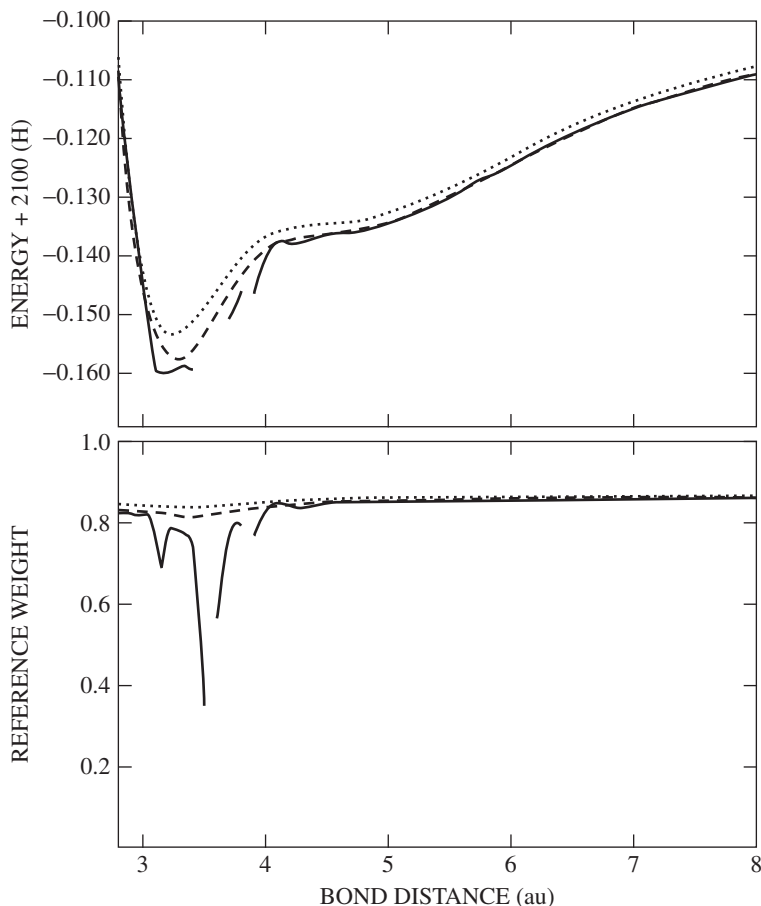**Figure 10.1** The potential energy curve for the ground state of Cr_2 for three values of the level shift, 0.05 E_h (solid line), 0.10 E_h (dashed line), and 0.20 E_h (dotted line). The LS correction of Eq. (10.61) has been applied. The lower diagram gives the corresponding weight of the CASSCF reference function. Reprinted with permission from ref. [6]. Copyright 1995, Elsevier.

where the gap-shift parameter associated with V disappears due to the orthogonality of $\Psi^{(0)}$ and $\Psi^{(1)}$ under intermediate normalization. The derivative of the expression above is

$$\nabla_{\bar{\Psi}_i^{(1)}} \mathcal{E}_i^{(2)}|_{\bar{\Psi}_i=\Psi_i, \bar{\Psi}_i=\Psi_i} = V|\Psi_i^{(0)}\rangle + (\hat{H}_0 + i\epsilon - E_i^{(0)})|\Psi_i^{(1)}\rangle = 0, \qquad (10.64)$$

which we can solve in an iterative manner using a preconditioned conjugate gradient (PCG) procedure (it should be of no surprise that this equation becomes equivalent to Eq. (10.11) if we project with $\hat{Q}$ – we are only interested in the part of the LHS which is perpendicular to the reference function). In the implementation of the PCG a particular routine and the preconditioner are slightly modified to incorporate the imaginary shift. At convergence the trial vector is used in a last call to the routine, but this time for the unshifted $\hat{H}_0$ and the corrected energy is computed using Eq. (10.63). In the closing of their report, Forsberg and Malmqvist list the advantages of the new method compared to the real-shift approach. First, the effect of the imaginary-shift approach on points away from the singularity is smaller than that of the real-shift method (compare figs. 10.1 and 10.2). Second, the imaginary-shift method has the advantage that it does not create new singularities – the real-valued shift technique does – but removes the singularities altogether.
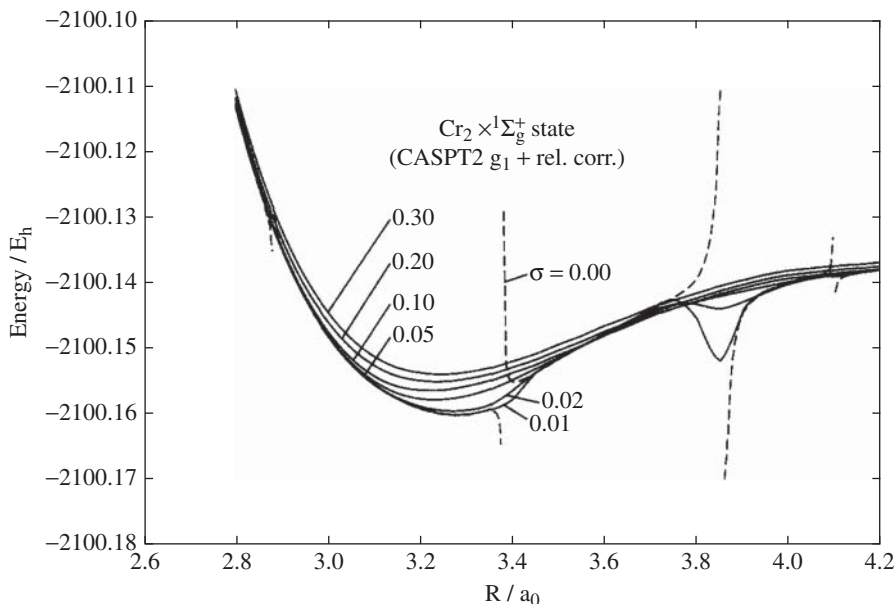
Figure 10.2 The CASPT2 potential energy curves for the ($^1\Sigma_g^+$) ground state of Cr_2, unshifted and with five values of the imaginary shift (called σ here). Reprinted with permission from ref. [7]. Copyright 1997, Elsevier.

This is particularly important for the purpose of studies of excited states. To summarize their conclusions, the imaginary-shift technique includes no calculations using complex arithmetic, the computed energy is real, and the procedure seems to have an edge over real-valued shift techniques in resolving cases with close to perturbation expansion divergence and the associated intruder states. We finally note that the iterative procedure mentioned above is a normal procedure in perturbation theory, although for some cases such as the Møller–Plesset partitioning in association with single determinant wave functions a non-iterative approach is accessible.

We will later see other examples of shift techniques, in particular the sometimes controversial IPEA shift (see section 10.3.3.5). We will have examples in which the shift is not universal but actually specific with respect to the state in the first-order interacting space. This is trivially achieved with a state-specific shift parameter designed as follows

$$\sum_{j \neq i} \epsilon_j |\Psi_j^{(0)}\rangle\langle\Psi_j^{(0)}| \,, \tag{10.65}$$

where the sum runs over the space of the complementary states. We will return to this as we discuss the IPEA shift.

This ends the discussion of single-state perturbation theory from the perspective of RSPT. In the subsequent sections we will see examples of specifics with respect to the partitioning of the Hamiltonian, how to deal with the fact that we do not have a complete reference space, what happens if the partitioning is state dependent as in the case of a multi-state formalism, etc.

10.2 Møller–Plesset Perturbation Theory

Møller–Plesset perturbation theory (MPPT), in which the reference functions are single determinants, is discussed here for several reasons. One, it deals with the issue of what to do if there

is no set of known reference functions which are eigenfunctions to $\hat{H}_0$. Two, CASPT2 theory is designed to coincide with MPPT2 results in the case the complete active space SCF collapses to a single-determinant SCF wave function. Hence, a special study of MPPT has its merits. We will also introduce the use of the second-quantization formalism here, where it is simpler, such that the student will be familiar with the technique once it is used later in the chapter. There exist alternative single-determinant partitioning schemes, such as, for example, the Epstein–Nesbet partitioning [9, 10]. However, the use of these in combination with multi-configurational reference functions is limited and will not be discussed in this chapter.

10.2.1 The Reference Function

The preceding sections leave us in the position of being able to discuss the Møller–Plesset PT without any new twists or approximations as compared to the Rayleigh–Schrödinger approach. However, let us first inspect the implications of our reference function's being a single Slater determinant (SD). This is at this time a bit tedious, but it will help us in understanding the procedure and decisions that will be made at the CASPT2 level of theory. Due to the single-determinant nature of the self-consistent field (SCF) wave function the one- and two-electron particle density matrices, expressed in the basis of the SCF canonical orbitals, reduce to

$$D_{ij} = 2\delta_{ij} \tag{10.66}$$

and

$$\Gamma_{ijkl} = 4\delta_{ij}\delta_{kl} - 2\delta_{il}\delta_{jk} , \tag{10.67}$$

and correspondingly the closed-shell SCF energy (an *expectation value*, not an eigenvalue), expressed in spatial doubly occupied orbitals, is compiled as

$$E_{\text{SCF}} = 2\sum_i h_{ii} + \sum_{ij}[2(ii|jj) - (ij|ji)] , \tag{10.68}$$

note that we have not added the constant corresponding to the nucleus–nucleus repulsion, V_{NN}, as is conventionally added to the electronic SCF energy. The corresponding orbital Fock operator, is written as

$$\hat{f}_{pq}(\Psi^{(0)}) = \hat{E}_{pq}\left(h_{pq} + \sum_k[2(pq|kk) - (pk|kq)]\right) \tag{10.69}$$

$$= \hat{E}_{pq}(h_{pq} + v_{pq}(\Psi^{(0)})) , \tag{10.70}$$

note that the summation over the orbitals k is in terms of the occupied orbitals of the reference SCF wave function $\Psi^{(0)}$ – this part of the equation expresses the interaction of a single electron with all other electrons. Hence, the recursive nature of the operator. This operator will yield the Fock matrix, which on diagonalization produces the canonical orbitals and the eigenvalues ϵ_i – the orbital energy of canonical orbital i. At convergence, the orbitals in the summation, expressing the electron density, are identical to the canonical orbitals – self-consistency is reached.

Let us define a molecular Fock operator as

$$\hat{F}(\Psi^{(0)}) = \sum_{pq}\hat{f}_{pq}(\Psi^{(0)}), \tag{10.71}$$

which has the following expectation value for the restricted closed-shell SCF wave function:

$$\langle\Psi^{(0)}|\hat{F}(\Psi^{(0)})|\Psi^{(0)}\rangle = 2\sum_i \epsilon_i , \tag{10.72}$$

where the summation runs over the n doubly occupied spatial orbitals and their associated orbital energy (whereas in Eq. (10.71) the double summation runs over *all* spatial orbitals, note the index usage according to Section 1.5). The orbitals energies, ϵ_i, in the canonical SCF basis is expressed as

$$\epsilon_i = h_{ii} + \sum_j [2(ii|jj) - (ij|ij)]. \tag{10.73}$$

This allows us to write the restricted closed-shell SCF electronic energy as

$$E_{\mathrm{SCF}} = 2\sum_i \epsilon_i - \sum_{ij} [2(ii|jj) - (ij|ji)] \ . \tag{10.74}$$

In the sum of the orbital energies we sum the interaction energies of one electron with all other electrons. This summation will double-count the electron–electron interaction, hence the subtraction of the latter term in the expression above to remove the double counting. Alternatively, we can write the SCF energy without making explicit use of the two-electron integrals:

$$E_{\mathrm{SCF}} = \sum_i (h_{ii} + \epsilon_i). \tag{10.75}$$

To further elaborate on the implications, a given molecular Fock operator – i.e., the interaction field defined by the set of occupied canonical orbitals – has eigenfunctions which are determinants of n orbitals selected out of the full set of occupied and virtual canonical orbitals. We note that there is one determinant which corresponds to the determinant of only the occupied canonical orbitals – the self-consistent solution – while the other determinants, although they certainly are valid eigenfunctions of the given molecular Fock operator, they do not correspond to self-consistent excited states. This is, however, not a requirement or a problem for perturbation theory. We only require that any Hamiltonian operator we use has a set of eigenfunctions which span the full Hilbert space of the system and that we can express them in a systematic and efficient way.

10.2.2 The Partitioning of the Hamiltonian

Now we naturally follow Møller and Plesset, who suggested using as a reference Hamiltonian the molecular Fock operator, $\hat{H}_0 = \hat{F}$, that is, the total Hamiltonian is partitioned as follows,

$$\hat{H} = \hat{h}^{(1)} + \hat{V}$$

$$= \hat{F} + \lambda \left(\hat{V} - \sum_{pq} \hat{E}_{pq} v_{pq} \right) \ . \tag{10.76}$$

By defining $\hat{\mathcal{V}}$ as the term in parentheses and introducing a constant C (to be determined later), we get,

$$\hat{H} = (\hat{F} + C) + \lambda(\hat{\mathcal{V}} - C). \tag{10.77}$$

Note that, while the RHS expression in Eq. (10.76) is general, this partitioning depends on the reference function – the Fock operator is a function of the electron density of the reference state. This trivially gives us the reference energy,

$$E_{\mathrm{MP}}^{(0)} = \langle \Psi^{(0)} | \hat{F} + C | \Psi^{(0)} \rangle = 2\sum_i \epsilon_i + C, \tag{10.78}$$

and the first-order correction of the energy,

$$E_{\mathrm{MP}}^{(1)} = \left\langle \Psi^{(0)} \Big| \hat{V} - \sum_{pq} \hat{E}_{pq} v_{pq} - C \Big| \Psi^{(0)} \right\rangle = \langle \Psi^{(0)} | \hat{\mathcal{V}} | \Psi^{(0)} \rangle - C. \tag{10.79}$$

The energy corrected to first order (recall Eq. 10.5) is identical to the SCF energy,

$$E_{\text{MP}}^{[1]} = E_{\text{MP1}} = \langle \Psi^{(0)} | \hat{F} | \Psi^{(0)} \rangle + C + \langle \Psi^{(0)} | \hat{\mathcal{V}} | \Psi^{(0)} \rangle - C$$
$$= \langle \Psi^{(0)} | \hat{F} + \hat{\mathcal{V}} | \Psi^{(0)} \rangle = E_{\text{SCF}} \ . \tag{10.80}$$

That is, the SCF total energy of the determinant is in general recovered only after the first-order correction and actual corrections to the SCF energy appear first at second order in the perturbation expansion. An alternative approach is to set $C = \langle \Psi^{(0)} | \hat{\mathcal{V}} | \Psi^{(0)} \rangle$, using a shifted-$\hat{H}_0$ operator. In this approach we have that $E^{(0)} = E_{SCF}$ and $E^{(1)} = 0$, and for higher orders of the correction the parameter disappears altogether due to the orthonormality of the reference states – in all, the modification could be considered semantic. In fact, the only significant difference is that the SCF determinants are now eigenfunctions of the shifted-$\hat{H}_0$ operator, something which we will meet in association the description of some of the multi-configurational reference function approaches discussed below.

To continue, the second-order correction to the energy can be computed as

$$E_{\text{MP}}^{(2)} = \langle \Psi^{(0)} | \hat{\mathcal{V}} | \Psi^{(1)} \rangle = \langle \Psi^{(0)} | \hat{H} - \hat{H}_0 | \Psi^{(1)} \rangle \ , \tag{10.81}$$

which suggests that we need to explicitly resolve $\Psi^{(1)}$, something we will do in the following section.

10.2.3 The First-Order Interacting Space and Second-Order Energy Correction

In order to proceed, we need to find the states with which, to first order, the reference function will interact. Before that, however, let us contemplate the impact of this statement. As we explore the effect of the excited states of the unperturbed system on the first-order correction to the wave function, we note that there is a hidden simplification in the equations. In our general ansatz, Eq. (10.23), we imply a summation over *all* excited states. However, upon examination of the equation for the corresponding coefficients of the first-order corrections (see Eq. 10.25) we find that only excited states which interact with the reference state over the perturbation potential have coefficients which are non-zero. We call this space the *first-order interacting space*, a space much smaller than the full Hilbert space. For the SCF reference determinant this translates to the fact that the first-order interacting space contains determinants which at most correspond to double replacements of orbitals from the reference SD.

For a given Fock operator, that is, once we have defined the reference SD which generates the reference electronic density and the operator, we can generate all excited states (eigenfunctions of the unperturbed system) for that particular operator by doing replacements of the occupied orbitals with the virtual (unoccupied) orbitals, starting with one, then two, and so on. As noted above, these "excited states" do not correspond to the true self-consistent excited states of the SCF procedure – that would require us to iteratively redefine the electronic density of the Fock operator to be self-consistent. One would then assume that the first-order correction to $\Psi^{(0)}$ would correspond to such generated eigenfunctions by doing a single excitation – from an occupied to a virtual orbital – that is, to form states as

$$\Psi_i^{a(0)} = \hat{E}_{ai} \Psi^{(0)} \ . \tag{10.82}$$

That would require us to form matrix elements such as

$$\langle \Psi^{(0)} | \hat{H} - \hat{H}_0 | \Psi_i^{a(0)} \rangle = \langle \Psi^{(0)} | \hat{H} | \Psi_i^{a(0)} \rangle = f_{ai} = 0 \tag{10.83}$$

when we determine the coefficients of those states in the first-order correction to the wave function (see Eq. 10.23). However, as seen above, the matrix element is identical to zero: the second term

vanishes since the reference state is an eigenstate to $\hat{H}_0$ while the first term is zero due to the variational minimization of the SCF energy – the Fock matrix is diagonalized such that these elements are identical to zero (Brillouin's theorem).

Let us now (*i*) construct the excited states, (*ii*) obtain their coefficients in the first-order corrected wave function, and (*iii*) subsequently compute the second-order correction of the energy. We do this using both spin-specific and spin-averaged operators, the latter used in the second quantization formalism, as they operate on the reference closed-shell single determinant. In the first case the procedure will be trivial, while in the second case the maths will not be kind to us. The latter will train us for the ordeal that will come when generating excited states for multi-configurational reference functions, a procedure which is always done in the spin-averaged formalism. Here we will have to introduce a new strategy to work around the problem.

We start with the spin-specific case where the replacement operator is expressed as $\hat{a}_{a_\sigma}^\dagger \hat{a}_{i_\sigma}$. In it we will have an operator that annihilates an electron with spin σ in occupied orbital i and places an electron with the same spin in virtual orbital a. For the rest of this subsection, on the use of spin-specific replacement operator, we will use spin-orbitals and index accordingly – that is, no explicit index for the spin. As discussed above the ground-state reference function will not interact with states that correspond to a single replacement – we need double replacements. To proceed, we will represent the doubly excited states as

$$\Psi_{ij}^{ab} = \hat{a}_b^\dagger \hat{a}_j \hat{a}_a^\dagger \hat{a}_i \Psi^{(0)} . \tag{10.84}$$

In a setting of using normalized molecular spin-orbitals the replacement operators will generate a new normalized determinant. Moreover, we note that these generated states are orthogonal – a desirable property. Finally, we have to take the four-fold permutational symmetry, $\Psi_{ij}^{ab} = \Psi_{ji}^{ba} = \Psi_{ji}^{ab} = \Psi_{ij}^{ba}$ into account. That is, the first-order wave function is expressed as

$$\Psi^{(1)} = \sum_{\substack{a>b \\ i>j}} c_{ijab} \Psi_{ij}^{ab}, \tag{10.85}$$

where the restricted summation over the spin-orbitals is introduced to include only the unique excited states. The individual coefficients are computed as

$$c_{ijab} = \frac{\langle \Psi^{(0)} | \hat{V} | \Psi_{ij}^{ab} \rangle}{E^{(0)} - E_{ij}^{ab}} . \tag{10.86}$$

The denominator trivially reduces to $\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b$ since $E_{ij}^{ab} = E^{(0)} - \epsilon_i - \epsilon_j + \epsilon_a + \epsilon_b$. The numerator is computed to be

$$\langle \Psi^{(0)} | \hat{V} | \Psi_{ij}^{ab} \rangle = (ai|bj) - (aj|bi). \tag{10.87}$$

Thus, the first-order correction to the wave function is expressed as

$$\Psi^{(1)} = \sum_{\substack{a>b \\ i>j}} \frac{(ai|bj) - (aj|bi)}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} \Psi_{ij}^{ab}. \tag{10.88}$$

We are now in a position to compute the second-order energy correction in terms of spin-orbitals as

$$E^{(2)} = \langle \Psi^{(0)} | \hat{V} | \Psi^{(1)} \rangle = \sum_{\substack{a>b \\ i>j}} \frac{[(ai|bj) - (aj|bi)]^2}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} . \tag{10.89}$$

If we unrestrict the summation, expand the square expression and finally sum over the spin we obtain the second-order correction of the energy in spatial orbitals as

$$
\begin{aligned}
E^{(2)} &= \frac{1}{4} \sum_{abij} \frac{[(ai|bj) - (aj|bi)]^2}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} \\
&= \frac{1}{2} \sum_{abij} \frac{(ai|bj)^2}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} - \frac{1}{2} \sum_{abij} \frac{(ai|bj)(aj|bi)}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b}
\end{aligned}
\tag{10.90}
$$

where we use the fact that the $\sum (ai|bj)^2$ and $\sum (aj|bi)^2$ terms are equal, since they are independent of the order of summation, now we switch to spatial orbital indexation

$$
\begin{aligned}
E^{(2)} &= 2 \sum_{abij} \frac{(ai|bj)^2}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} - \sum_{abij} \frac{(ai|bj)(aj|bi)}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} \\
&= \sum_{abij} \frac{(ai|bj)[2(ai|bj) - (aj|bi)]}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} \ .
\end{aligned}
\tag{10.91}
$$

We have obtained the appropriate equations for the computation of the second-order correction to the energy expressed in both spin-orbitals (see Eq. 10.90) and spatial orbitals (see Eq. 10.91). We did so by using a technique expressed in spin-orbitals, which provided us with (*i*) an easy way to generate orthonormal excited states and (*ii*) trivial equations to compute the corresponding coefficients for the first-order correction to the wave function. In closing this passage, we note that the computation of the second-order correction of the energy in the closed-shell case of MP can be solved in a single step. We do not need to explicitly store the coefficients of the first-order corrected wave function – these are evaluated from analytical integrals which are computed on the fly as we need them. This is a luxury we will not find in the multi-configurational reference function version of the theory – here the coefficients have to be computed in an iterative procedure and need to be explicitly stored before we compile the second-order correction to the energy. It should also be pointed out that the summations in the above equations run over *all* orbitals in the system. It is common in practical calculations to reduce the ranges by ignoring the lowest-lying occupied orbitals and/or the highest-lying virtual orbitals. In the first case – what is usually known as *frozen core* approximation – the inner occupied orbitals are not considered when generating the first-order interacting space, but their Coulomb potential must still be taken into account.

We will now see how we would proceed if we had to repeat these steps using spin-averaged replacement operators in association with a second-quantized Hamiltonian. The procedure will, to some extent, introduce a formalism that is not obviously simpler, however, it will be a general recipe that will guarantee that the generated states have the same eigenvalues of the spin operators $\hat{S}^2$ and $\hat{S}_z$ as the reference functions, since $\hat{E}_{pq}$ commutes with these operators. We start now with the process of defining the first-order interacting space. Here we note that the operator $\hat{E}_{ai}$ will, when it operates on a single closed-shell SD, generate two determinants, one in which an α electron has been replaced and another for the β electron. Thus, we have that an orthonormalized singly excited state is expressed as,

$$
\Psi_i^a = \frac{1}{\sqrt{2}} \hat{E}_{ai} \Psi^{(0)} \ ,
\tag{10.92}
$$

since

$$
\langle \Psi_j^b | \Psi_i^a \rangle = \langle \Psi^{(0)} | \hat{E}_{jb} \hat{E}_{ai} | \Psi^{(0)} \rangle = 2 \delta_{ij} \delta_{ab} \ .
$$

Again, for the reasons discussed above, we need excited states that are doubly excited relative to the reference function if we are to form the first-order correction to the wave function. We now explore

the normalization using a pair of replacement operators. We define doubly excited states as

$$\Psi_{ij}^{ab} = \hat{E}_{bj}\hat{E}_{ai}\Psi^{(0)} \; , \tag{10.93}$$

for which we have the squared norm

$$\langle \Psi_{ij}^{ab}|\Psi_{ij}^{ab}\rangle = 4 + 4\delta_{ij}\delta_{ab} - 2\delta_{ij} - 2\delta_{ab} \; . \tag{10.94}$$

This is consistent with the fact that each replacement operator generates two determinants when operating on a closed-shell determinant. In the trivial case of no equal indices, $i \neq j$ and $a \neq b$, applying the two replacement operators we should get four determinants. This leads us to conclude that we need to incorporate the normalization of the excited states as we use them in our equations. Let us now explore the orthogonality within the set under a permutation of the indices. A double permutation, which translates to a change of order in the application of the replacement operators, does not cause any difference in the generated excited states, since

$$\hat{E}_{bj}\hat{E}_{ai} = \hat{E}_{ai}\hat{E}_{bj} + \delta_{aj}\hat{E}_{bj} - \delta_{bj}\hat{E}_{aj} \; , \tag{10.95}$$

where the two latter terms for obvious reasons disappear, that is $\Psi_{ij}^{ab} = \Psi_{ji}^{ba}$. However, under the condition of a single permutation of the virtual or the occupied indices, we find that the generated determinants are not identical, but they obey the relation

$$\langle \Psi_{ij}^{ba}|\Psi_{ij}^{ab}\rangle = \langle \Psi_{ji}^{ab}|\Psi_{ij}^{ab}\rangle = -2 - 2\delta_{ij}\delta_{ab} + 4\delta_{ij} + 4\delta_{ab} \; , \tag{10.96}$$

which implies that we not only have to normalize the states generated by Eq. (10.93): we have to orthonormalize them. Given the above result in the spin-specific approach we are guaranteed that the generated excited states are not redundant, except, of course, when they are identical by not properly constraining the indices in Eq. (10.93). Hence, we should not have to worry about redundancy and its elimination – this has to be considered for the case of a multi-configurational reference function. We note that some, if not all, of the non-orthogonality can be removed by forming the appropriate linear combinations as $\Psi_{ij}^{ab,\pm} = \Psi_{ij}^{ab} \pm \Psi_{ij}^{ba}$, corresponding to a singlet and triplet coupling of the virtual orbitals. This is achieved in a similar manner to how configurations of two electrons in two orbitals are linearly combined to produce singlet and triplet spin eigenstates with $M_s = 0$.

Let us, for the sake of convenience, simply index the doubly excited states generated in Eq. (10.93) with a compound index $I = I(i, j, a, b)$, such that $\Psi_{ij}^{ab} = \Psi_I$. To resolve the issue of orthonormalization we form the overlap matrix

$$S_{IJ} = \langle \Psi_I|\Psi_J\rangle \; , \tag{10.97}$$

which we subsequently diagonalize

$$\boldsymbol{s} = \boldsymbol{U}^{\dagger}\boldsymbol{S}\boldsymbol{U}, \tag{10.98}$$

where $\boldsymbol{s}$ is a diagonal $(L \times L)$ matrix with the eigenvalues and $\boldsymbol{U}$ is the eigenvector matrix. Since in closed-shell MPPT L can be limited to run over just the non-redundant space of double excitations, $\boldsymbol{U}$ is guaranteed to be square. However, in the general case the size of $\boldsymbol{U}$ will be $(M \times L)$, where M is the size of the original partially redundant double excitation space. However, we proceed by combining the eigenvectors with the eigenvalues according to

$$\tilde{\boldsymbol{U}} = \boldsymbol{U}\boldsymbol{s}^{-\frac{1}{2}} \; . \tag{10.99}$$

We now have a set of orthonormalized doubly excited states which are computed as

$$\tilde{\Psi}_I = \sum_J \tilde{U}_{JI}\Psi_J \; . \tag{10.100}$$

This will then render the equations for the computation of the coefficients in the first-order correction to the wave function and the second-order correction to the energy to be, starting with the former,

$$\Psi^{(1)} = \sum_I c_I \tilde{\Psi}_I \tag{10.101}$$

with the coefficients as

$$c_I = \frac{\langle \Psi^{(0)} | \hat{V} | \tilde{\Psi}_I \rangle}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} \ , \tag{10.102}$$

where the indices of the eigenvalues of the Fock operator are resolved for the compound index I. The second-order energy correction is now computed as

$$E^{(2)} = \sum_I \frac{\langle \Psi^{(0)} | \hat{V} | \tilde{\Psi}_I \rangle^2}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} \ . \tag{10.103}$$

Before we are finished here, let us break this down one step further and use a matrix representation all along. First we define the following matrices and vectors,

$$(\boldsymbol{H}_0)_{IJ} = \langle \Psi_I | \hat{H}_0 | \Psi_J \rangle \tag{10.104}$$

and

$$\boldsymbol{V}_I = \langle \Psi_I | V | \Psi^{(0)} \rangle \ . \tag{10.105}$$

We transform these matrices to the space of the orthonormalized excited states,

$$\tilde{\boldsymbol{H}}_0 = \tilde{\boldsymbol{U}}^\dagger \boldsymbol{H}_0 \tilde{\boldsymbol{U}} \tag{10.106}$$

and

$$\tilde{\boldsymbol{V}} = \tilde{\boldsymbol{U}}^\dagger \boldsymbol{V}. \tag{10.107}$$

The coefficients $\tilde{\boldsymbol{C}}$ in the expression for the first-order correction to the wave function, are found by solving

$$(\tilde{\boldsymbol{H}}_0 - E^{(0)}\boldsymbol{1})\tilde{\boldsymbol{C}} = -\tilde{\boldsymbol{V}}, \tag{10.108}$$

and the second-order correction to the energy is expressed as

$$E^{(2)} = \tilde{\boldsymbol{C}}^\dagger \tilde{\boldsymbol{V}}. \tag{10.109}$$

A final note here, in this approach we will be forced to use a two-step approach – first compute the coefficients and then the energy. With this we are now ready to discuss some particularities of the CASPT2 and some other related methods.

10.3 State-Specific Multi-Configurational Reference Perturbation Methods

After this rather lengthy introduction to perturbation theory, in this section we will, in chronological order, present the four most popular versions of multi-configurational reference perturbation theory, namely the complete active space second-order Møller–Plesset (CAS-MP2) [11], the complete active space second-order perturbation theory (CASPT2) [12, 13], the multi-reference second-order Møller–Plesset (MRMP2) [14–17] and the *n*-electron valence state second-order

perturbation theory (NEVPT2) [18–20]. In particular, we will devote special attention to the CASPT2 method and present it in some detail, while the three other approaches are presented more briefly and in order to mention the fundamental differences between the approaches. Briefly, the major differences between these four methods are associated with the initial selection of the reference functions $\Psi^{(0)}$ and the partitioning of the Hamiltonian into the reference Hamiltonian, $\hat{H}_0$, and the perturbation potential, V. We would also like to point out that the three latter methods CASPT2, MRMP2 and NEVPT2 have in common a *diagonalize-then-perturb* approach. That is, the CASSCF equations are effectively solved via a diagonalization-like approach and the dynamic correlation is introduced using perturbation technique. The CAS-MP2 method, on the other hand, is a hybrid somewhere in between a *diagonalize-then-perturb* and a *perturb-then-diagonalize* approach (see section 10.4 et seq.). The drawbacks and advantages of the four different approximations are discussed. In addition to these four versions of multi-configurational reference perturbation theory, a number of other flavors exist. They will not, however, be discussed in this chapter, mostly because they have not, to date, made any significant impact on standard applications. Those desiring a complete list of other such methods are recommended to consult the review paper by Lischka and co-workers [21] and specifically the subsection on "Mutireference [sic] Perturbation Theory".

10.3.1 The Generation of the Reference Hamiltonian

Before we proceed, however, let us address an issue relevant to most versions of multi-configurational reference perturbation theory beyond single determinant RSPT: What to do if (*i*) our reference function is not an eigenfunction to $\hat{H}_0$, and (*ii*) we have only exact knowledge about the reference functions, and not about the complementary functions that extend to the full Hilbert space. In this context there would be a problem to apply RSPT to known wave function models. Let us address the first problem here by constructing a new operator to solve the problem as follows. Let us assume that we have an operator, $\hat{O}$, for which our reference and complementary states are not eigenfunctions, but which we for some practical reason want to use. Then an initial operator can be constructed for the model space – the space spanned by all our reference functions – as

$$\hat{\bar{O}} = \sum_i |\Psi_i^{(0)}\rangle\langle\Psi_i^{(0)}|\hat{O}|\Psi_i^{(0)}\rangle\langle\Psi_i^{(0)}| \,, \tag{10.110}$$

for which now all the functions $\Psi_i^{(0)}$ of the model space are eigenfunctions. This has now solved the first issue. The second problem is related to the fact that in most quantum chemistry simulations one does not have explicit access to the eigenfunctions of the complementary space. Hence, let us further assume that we can construct a complementary space to our model space, with functions that are all orthonormalized. To proceed we note that to the general expression above we are at liberty to add any complementary space we can construct, as long as these functions are mutually orthonormal. In that respect we can rewrite the expression above as

$$\hat{\bar{O}} = \sum_{i=1}^{d} |\Psi_i^{(0)}\rangle\langle\Psi_i^{(0)}|\hat{O}|\Psi_i^{(0)}\rangle\langle\Psi_i^{(0)}| + \sum_{i=d+1}^{\infty} |\psi_i\rangle\langle\psi_i|\hat{O}|\psi_i\rangle\langle\psi_i| \,, \tag{10.111}$$

where $\Psi_i^{(0)}$ are our d reference functions, which we know exactly, and ψ_i are artificial functions with the properties $\langle\psi_i|\psi_j\rangle = \delta_{ij}$ and $\langle\Psi_i^{(0)}|\psi_j\rangle = 0$. This now solves the second problem – the lack of a complete knowledge of all eigenfunctions. The remaining obstacle is the construction of the artificial complementary space. Here a standard procedure is used, applying sets of spin-averaged electron replacement operators to our known reference wave functions – in MPPT this procedure actually leads to true eigenfunctions of the molecular Fock operator – combined with the operator

$\hat{Q} = 1 - \hat{P}$ (see Eq. 10.18) and the orthonormality conditions, to generate arbitrary complementary spaces.

10.3.2 CAS-MP2 Theory

In 1988 McDouall and co-workers published the first practical solution for extending normal single-configuration MPPT to the case of multi-configuration reference functions, in particular using a CASSCF reference function [11]. Their procedure is based on a Löwdin-type partitioning [22] of the full Hamiltonian into the subspace of the electronic configurations (Slater determinants or configuration state functions), and the complementary space. The resulting equations are subsequently solved using simple MPPT. Let us outline, in some detail, the approach here. We define a model space that includes all the CASSCF states. In this aspect we have that the projection operators $\hat{P}$ and $\hat{Q}$ play the same role as in the section on the multi-state case (see section 10.4). In this formalism we divide the Hamiltonian by blocks into P and Q spaces. In particular we note that $\hat{P} = \sum_i |\Psi_i^{(0)}\rangle\langle\Psi_i^{(0)}|$, where $\Psi_i^{(0)}$ are solutions to the CASSCF equations. We then introduce a transformation, U, which transforms the Hamiltonian into a representation in which the P and Q spaces are decoupled, that is

$$\begin{pmatrix} H^{PP} & H^{PQ} \\ H^{QP} & H^{QQ} \end{pmatrix} U = U \begin{pmatrix} H^{\text{eff}} & 0 \\ 0 & W^* \end{pmatrix} . \tag{10.112}$$

In this matrix representation $H^{PP} = PHP$ and so on. For the model space we have the normal CASSCF solutions to the time-independent Schrödinger equation,

$$\hat{P}\hat{H}\hat{P}\Psi_i^{(0)} = E_i^{(0)}\Psi_i^{(0)}. \tag{10.113}$$

The transformation matrix, U, will have diagonal blocks in which the transformation is arbitrary, for the sake of convenience we set the matrix to have the form

$$U \equiv \begin{pmatrix} 1 & -C^{\text{T}} \\ C & 1 \end{pmatrix} , \tag{10.114}$$

where C is a set of column vectors, one for each function in the model space. This now generates two working equations,

$$(H^{\text{eff}})_{ij} = (H^{PP})_{ij} + (H^{QP})_i^{\text{T}} C_j \tag{10.115}$$

and

$$(H^{QP})_i + H^{QQ}C_i = \sum_j (H^{\text{eff}})_{ji} C_j , \tag{10.116}$$

where two subindices indicate a matrix element, and one subindex indicates a whole column. The total energy for state i is then approximated as

$$E_i \approx \langle\Psi_i^{(0)}|\hat{H}^{\text{eff}}|\Psi_i^{(0)}\rangle . \tag{10.117}$$

McDouall and co-workers point out that this approach – a hybrid approach between a *diagonalize-then-perturb* and a *perturb-then-diagonalize* scheme – comes with two advantages. First, single-excitation contributions are zero, and second, the off-diagonal correction of Eq. (10.115) is expected to be small and can be ignored. The energy is now expressed as

$$E_i \approx E_i^{(0)} + \langle\Psi_i^{(0)}|(\hat{H}^{QP})_i^{\text{T}}\hat{C}_i|\Psi_i^{(0)}\rangle , \tag{10.118}$$

where $\hat{X}_i$ is the operator representation of the ith column of matrix $\boldsymbol{X}$. What remains now is to compute $\boldsymbol{C}$ using MPPT theory. The authors note that one can approximate Eq. (10.116) by restricting the summation on the RHS to only include the same $\boldsymbol{C}$-vector as on the LHS – i.e., taking only $j = i$, an approximation which affects the results only at the third order of the perturbation. This leaves us the following equation,

$$\boldsymbol{C}_i = -(\boldsymbol{H}^{\mathrm{QQ}} - (\boldsymbol{H}^{\mathrm{eff}})_{ii}\boldsymbol{1})^{-1}(\boldsymbol{H}^{\mathrm{QP}})_i \; . \tag{10.119}$$

We can now proceed solving this by applying the follow procedure. First, the inverse problem in Eq. (10.119) is approached with the use of a power series expansion. Second, sums over states are replaced with sums over orbitals. Third and finally, the diagonal elements of the matrix to invert are replaced with orbital energy differences from the diagonal elements of some reference one-electron Hamiltonian $\hat{h}_i$. To be explicit, the authors used the power series expansion

$$(\boldsymbol{A} - \boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1} + \dots, \tag{10.120}$$

where $\boldsymbol{A}$ is the matrix of the diagonal elements of $\boldsymbol{H}^{\mathrm{QQ}} - (\boldsymbol{H}^{\mathrm{eff}})_{ii}\boldsymbol{1}$. Moreover, for the reference one-electron Hamiltonian the authors used different $\hat{h}_i$ for each reference state i. This makes sense since in some reference state (read determinant or possibly configuration state function) an active orbital might be occupied while in another state it is empty. Using a unified $\hat{h}$ operator under these conditions would render the perturbation expansion to diverge. However, the rules for the definition of the $\hat{h}_i$ are to some extent *ad hoc* and sensitive to the orbital representation, McDouall and co-workers have selected to use a localization of the active orbitals leading to an orthogonal valence-bond expansion. As for the performance with respect to computer time the authors state that the method will cost n times the cost of a single-reference MP2, where n is the number of functions in the reference space. This fact means that the approach is not often used. Furthermore, the literature lacks any benchmarks comparing this approach to any of the other MRPT methods presented in this section. From a theoretical ground there is no reason for the CAS-MP2 approach to be worse than the methods that will follow, rather it might actually have a slight edge due to the use of a hybrid approach. However, the poor scaling with the size of the CI expansion will, in general, render the approach not useful.

10.3.3 CASPT2 Theory

Here we present the mathematical equations which govern the complete active space second-order perturbation theory (CASPT2) method, a second-order perturbation method.[1] More specifically, we will present the state-specific CASPT2 method (SS-CASPT2), in which each state is treated independently. In this approach we note that, if we have a perturbation series which is convergent, each of the states we consider will, as we go towards infinite order, get the energy and wave function correct. We will be able to compute the ground and any excited states which are spanned by the active space of the reference CASSCF wave function. In general this will work very well if the original CASSCF states are good qualitative representatives of the final states. However, convergence will be slow if the reference states are poor and the successive iterations over the order of the perturbational strength parameter are desperately needed for the corrected wave function to mirror the exact wave function of the state under study in a qualitative way. Examples of this can be numerous, for instance, the lack of dynamic electron correlation in the CASSCF model can cause

---

1 A review of the state-specific CASPT2 method was presented by Andersson and Roos in 1995 [23]. This section is an extension of that review and we recommend students to consult that material in parallel.

the reference functions to artificially mix states beyond the one they are trying to represent to be used in a perturbational approach. We will present below approaches that will fix these problems, but more on that later.

The CASPT2 method is based on (*i*) a multi-configurational reference wave function, the CASSCF wave function, as described in detail in Chapter 6, and (*ii*) a partitioning of the Hamiltonian, such that it reduces down to Møller–Plesset partitioning, that is, the definition of $\hat{H}_0$ is the same – a one-electron operator – if the CASSCF wave function is reduced down to a single-determinant wave function. The selection of $\hat{H}_0$ is for reasons of simplicity required to be a one-electron operator, alternative but similar partitionings have been introduced [24] but have over the years been abandoned due to suboptimal performance. As we proceed we face two issues. First, we need a generalization of the Fock operator to wave functions of arbitrary spin and number of open shells. Second, we need to formulate a $\hat{H}_0$ for which the CASSCF wave function is an eigenfunction.

10.3.3.1 The Partitioning of the Hamiltonian

Let us start with the mathematical description of the SS-CASPT2 method, as it is applied to a general CASSCF state. The generalized Fock operator, the $\hat{H}_0$ operator of the SS-CASPT2 method, is defined as

$$\hat{F}(D) = \sum_{pq} f_{pq}(D)\hat{E}_{pq} \ , \tag{10.121}$$

where D is the CASSCF one-particle reduced density matrix of the state of interest, represented by the state $\Psi^{(0)}$. The scalar value f_{pq} is computed as

$$f_{pq}(D) = \frac{1}{2} \sum_{\sigma=\alpha}^{\beta} \langle \Psi^{(0)} | [\hat{a}_{p_\sigma}^\dagger, [\hat{a}_{q_\sigma}, \hat{H}]] | \Psi^{(0)} \rangle \tag{10.122}$$

$$= h_{pq} + \sum_{rs} D_{rs} \left[(pq|rs) - \frac{1}{2}(pr|qs) \right] \ , \tag{10.123}$$

which in the case of the standard closed-shell single-determinant case will boil down to the conventional matrix elements of the molecular orbital Fock operator. In passing we note that the matrix f formed by the f_{pq} values can be broken down into nine blocks based on the division of the orbitals into inactive, active and virtual orbitals,

orbital class	inactive	active	virtual
inactive	ϵ_i	f_{it}	0
active	f_{ti}	ϵ_t	f_{ta}
virtual	0	f_{at}	ϵ_a

The diagonal block of this matrix (green boxes above) can be diagonalized independently resulting in four remaining non-zero blocks (red boxes) (the inactive–active, active–inactive, active–virtual, and virtual–active blocks) and two zero blocks (the inactive–virtual and virtual–inactive blocks). In the initial implementation of CASPT2 the Fock operator was approximated to be diagonal to simplify the calculation and compute the second-order energy and coefficients of the first-order wave function simultaneously [12]. However, this approximation, called D-CASPT2, proved to be too severe and later an iterative implementation was presented in which the non-diagonal nature

of the Fock matrix is kept [13], the N-CASPT2 methods. This prefix has now been dropped, the D-CASPT2 results are very poor and all CASPT2 results now refer to the use of the full Fock matrix.

10.3.3.2 The First-Order Interacting Space

We have now solved our first problem – we have a $\hat{H}_0$ operator, however, in contrast to MP theory our CASSCF reference wave functions are not eigenfunctions to the generalized Fock operator. In this respect we will have to follow the recipe described a couple of sections above, that is we introduce the identity as defined by the projection operators

$$1 = |\Psi^{(0)}\rangle\langle\Psi^{(0)}| + \hat{P}_K + \hat{P}_{SD} + \hat{P}_{TQ} + \dots, \tag{10.124}$$

in which the projection operators $\hat{P}_K$, $\hat{P}_{SD}$, etc., are defined as

$$\hat{P}_X = \sum_{i\in X} |\Psi_i\rangle\langle\Psi_i| \,,$$

where Ψ_i is a orthonormal state of the X space ($X \in \{K, SD, TQ, \dots\}$). More specifically, $\hat{P}_K$ projects into the complementary space, wave functions corresponding to all other possible wave functions that are a solution in the CAS space, but are not in the reference space. $\hat{P}_{SD}$ corresponds to states which are single or double replacements (SD)[2], relative to the reference function, generated as $\hat{E}_{pq}\hat{E}_{rs}|\Psi_i^{(0)}\rangle$. Note that, as in the case of MPPT, single replacements can be safely ignored, since for a variationally optimized CASSCF wave function their interaction with $\Psi^{(0)}$ is zero (Brillouin–Levy–Berthier theorem). We selected this approach before for a reason. Let us visit the SD space in a bit more detail, since if we are only to compute second-order corrected energies we need only worry about the design and structure of this space. In the SD space the electron replacements (see below) are complete or partial replacements out of the inactive–active space into the active–virtual space. This includes the so-called semi-internal replacements, involving the coupling of an internal replacement (inactive–active replacement) with an external replacement (inactive–virtual or active–virtual promotion). To present all the SD space, it is divided into eight classes of operators acting on the reference wave function ($V_A, \dots, V_H$), they are presented as:

$$\text{internal}\begin{cases} V_A\,, & k = 1 : \hat{E}_{ti}\hat{E}_{uv} \\ V_B\,, & k = 2 : \hat{E}_{ti}\hat{E}_{uj} \end{cases}$$

$$\text{semi-internal}\begin{cases} V_C\,, & k = -1 : \hat{E}_{at}\hat{E}_{uv} \\ V_D\,, & k = 0 : \hat{E}_{ai}\hat{E}_{tu}\,,\ \hat{E}_{ti}\hat{E}_{au} \\ V_E\,, & k = 1 : \hat{E}_{ti}\hat{E}_{aj} \end{cases}$$

$$\text{external}\begin{cases} V_F\,, & k = -2 : \hat{E}_{at}\hat{E}_{bu} \\ V_G\,, & k = -1 : \hat{E}_{ai}\hat{E}_{bt} \\ V_H\,, & k = 0 : \hat{E}_{ai}\hat{E}_{bj}\,, \end{cases}$$

where i and j are inactive orbital, a and b are virtual orbital, and t, u, and v are active orbital indices of the CASSCF wave function. The classification as internal, semi-internal or external corresponds to the number of virtual indices, 0, 1 or 2, respectively. The index k denotes the change of number of electrons in the active space under the application of the replacement operator – this is used in the NEVPT2 terminology to divide the first-order interacting space into subclasses

---

2 It is a confusing fact that both Slater determinants and the single and double replacement space of the first-order interacting space in the literature are abbreviated SD. However, we hope that it will be clear from the context what the abbreviation refers to.

(see section 10.3.5 below). We note that the last case corresponds to the normal set of parameters found in conventional MP2 theory and they are the ones that are the largest set. As in the case of MPPT, non-orthogonalities in the first-order interacting space can be removed by taking linear combinations of operators, as in $\hat{E}_{ai}\hat{E}_{bj} \pm \hat{E}_{bi}\hat{E}_{aj}$.

Despite the block nature of $\hat{F}(D)$, we will not get a block-diagonal Hamiltonian – this would translate to the correction vectors spanning the full Hilbert space. To overcome this we construct a projected $\hat{H}_0$ operator as

$$
\begin{aligned}
\hat{H}_0 = {}& |\Psi^{(0)}\rangle\langle\Psi^{(0)}|\hat{F}(D)|\Psi^{(0)}\rangle\langle\Psi^{(0)}| \\
& + \hat{P}_{\mathrm{K}}\hat{F}(D)\hat{P}_{\mathrm{K}} + \hat{P}_{\mathrm{SD}}\hat{F}(D)\hat{P}_{\mathrm{SD}} \\
& + \hat{P}_{\mathrm{TQ}}\hat{F}(D)\hat{P}_{\mathrm{TQ}} + \dots \, .
\end{aligned}
\tag{10.125}
$$

Thus, this $\hat{H}_0$ operator will be block-diagonal as follows

space	0	K	SD	TQ	$\cdots$
0	$\boldsymbol{H}_0^{(0,0)}$	**0**	**0**	**0**	**0**
K	**0**	$\boldsymbol{H}_0^{(\mathrm{K,K})}$	**0**	**0**	**0**
SD	**0**	**0**	$\boldsymbol{H}_0^{(\mathrm{SD,SD})}$	**0**	**0**
TQ	**0**	**0**	**0**	$\boldsymbol{H}_0^{(\mathrm{TQ,TQ})}$	**0**
$\vdots$	**0**	**0**	**0**	**0**	$\ddots$

which will assure us that the first-order correction to the wave function is completely defined by the SD subspace.

Proceeding with a brief analysis of the SD space we first note that we will have a set of excited states which are neither normalized nor orthogonal. Even worse, they are not necessarily linearly independent. Hence, orthonormalization and elimination will have to be included in the process in which the overlap matrix $\boldsymbol{S}_{IJ}$ in Eq. (10.97) is processed (block by block) to generate a non-redundant set of orthonormalized excited states. This is now used to make the appropriate transformations to generate the matrix $\tilde{\boldsymbol{H}}_0$ and the vector $\tilde{\boldsymbol{V}}$. All this is used in the equation to derive the coefficients of the first-order corrections to the wave function (see Eq. 10.108). Second, different to standard MPPT, with the use of a single closed-shell determinant as a reference wave function, the matrix of the unperturbed Hamiltonian will not be trivially diagonal, while still block-diagonal. A brute force approach would be to diagonalize the matrix on the LHS, however, that would limit the size of the calculations we could do – the space of the doubly excited states would be too large for the explicit storage of the $\tilde{\boldsymbol{H}}_0$ matrix. Rather the Equation will have to be solved in an iterative fashion using a preconditioned conjugate gradient approach computing vectors (known as σ vectors) of the form

$$
\sigma = (\tilde{\boldsymbol{H}}_0 - E^{(0)}\mathbf{1})\boldsymbol{C}^{(1)} \, ,
\tag{10.126}
$$

where $\boldsymbol{C}^{(1)}$ is a trial vector of the coefficients of the excited states in the SD subspace. In this context we will evaluate the matrix on the fly as we make the contraction (product) with the trial vector. That is, the matrix $\tilde{\boldsymbol{H}}_0 - E^{(0)}\mathbf{1}$ is not stored explicitly. Rather, as an element of the σ vector is generated only those parts of the $\tilde{\boldsymbol{H}}_0 - E^{(0)}\mathbf{1}$ matrix which are required are generated on the fly.

Moreover, as the projection operator, $\hat{P}$, is extended further to sets of two further replacements – the TQ manifold, and so on – we achieve a continued block diagonalization of $\hat{H}_0$ over the blocks of the reference function, the complementary CAS space, the SD space, the TQ space, etc.

We are now at the point at which we need to form the matrix elements (I, J are compound indices for the states in the SD space)

$$(\boldsymbol{H}_0)_{IJ} = \langle \Psi_I | \hat{H}_0 | \Psi_J \rangle \ ,$$
$$\boldsymbol{S}_{IJ} = \langle \Psi_I | \Psi_J \rangle \ , \text{ and}$$
$$\boldsymbol{V}_I = \langle \Psi_I | \hat{V} | \Psi^{(0)} \rangle \ ,$$

where

$$\hat{V} = \sum_{pqrs} \hat{e}_{pqrs}(pq|rs) - \hat{E}_{pq} \sum_{rs} D_{rs} \left[(pq|rs) - \frac{1}{2}(pr|qs) \right] \ . \tag{10.127}$$

We will use these matrices and vectors to (*i*) eliminate linear dependence, (*ii*) introduce a non-redundant orthonormal space of states in the SD space, (*iii*) commence the iterative procedure to compute the coefficients of the first-order correction to the wave function, and (*iv*) compute $E^{(2)}$.

Before we continue let us examine, in some detail, the matrix elements which have to be computed and what this is associated with (a complete list of all required equations for the compilation of the CASPT2 energy are collected in the Appendix). We have that the wave functions Ψ_I and Ψ_J contain the product of two one-electron replacement operators, and so does $\hat{V}$, while $\hat{H}_0$ is a one-electron operator and contains single replacement operators. Therefore, the worst case could be expected to be the computation of matrix elements for $\hat{H}_0$. For example, for one of the external cases (V_{H}), we need to evaluate expressions such as

$$\sum_{pq} \langle \Psi_i^{(0)} | \hat{E}_{kc} \hat{E}_{ld} \hat{E}_{pq} \hat{E}_{ai} \hat{E}_{bj} | \Psi_j^{(0)} \rangle \ . \tag{10.128}$$

This and other expressions seem to indicate that n-particle density matrices up to five-particle matrices would be needed, since there are five replacement operators. However, the general reduction (for a complete list of the equations up to the second-order correction of the energy see the Appendix) will be such that the Hamiltonian will eliminate the four virtual indices, that is, a CASPT2 implementation needs, in addition to the one- and two-particle density matrices, three- and partially contracted four-particle density matrices (effectively of the size of a three-particle density matrix). The actual storage and computation cost for all the n-particle density matrices (1-, 2-, 3- and partially contracted 4-particle density matrices) will scale in the range of m^6–m^8, where m is the number of active orbitals. This seems to be tractable right now but, as will be seen later in section 10.3.6, the density matrix renormalization group approach to CASSCF calculations will facilitate large active spaces, at which point, when m can easily be of the order 10^2, the computation and handling of the density matrices can become a serious bottleneck.

We are almost done. However, we should not forget to address the issue of size-extensivity of the CASPT2 method. The CASSCF and MP2 methods (the latter at least in the case of closed-shell systems) are size-extensive. Does this hold for the CASPT2 method too? Helgaker and co-workers [3] show that since $\hat{H}_0$ is not additively separable (due to the projection operators present in the $\hat{H}_0$ definition), the CASPT2 model is not a size-extensive perturbation model. Further, they note that as the reference wave function increases the multi-configurational character, the extent of the non-size-extensivity increases.

10.3.3.3 Other Active Space References

The CASPT2 method, as should be clear from the preceding and following sections, is not free of issues. To the just mentioned non-size-extensivity, we can add the possibility of intruder states (10.1.2), the overstabilization of open-shell states (10.3.3.5) and the non-orthogonality of the first-order wave functions (10.5). It should be noted that these problems are not specific to CASPT2, and other multi-configurational reference perturbation methods can also show some or all of them to some extent. There is, however, a simple – and not particularly useful – recipe to solve or mitigate these problems: increase the active space size. From a purely theoretical point of view, by increasing the active space size we approach the full CI solution, at which point our reference CASSCF functions *are* the exact solutions, the zeroth-order Hamiltonian $\hat{H}_0$ is the exact Hamiltonian and the perturbation V vanishes. It "solves" the problems by making the PT treatment superfluous. From a practical standpoint, the exponential increase in the number of SDs with the size of the active space makes the required CASSCF calculations intractable for anything beyond about 20 orbitals, not to mention the PT2 step.

As described in Chapter 6, other approaches have been proposed to allow extending an active space beyond what is practically possible with CASSCF, while keeping the number of SDs manageable. Two of these approaches are the restricted active space SCF (RASSCF) and generalized active space SCF (GASSCF) methods. How can we apply a perturbative treatment to reference functions based on RASSCF or GASSCF, instead of CASSCF?

The corresponding extensions of CASPT2, RASPT2 [25] and GASPT2 [26], have been published relatively recently. In the case of RASPT2, the matrix representation of the Fock operator can be easily obtained from the one in CASPT2, noting that now the active space is split into three different subspaces:

		active			
orbital class	inactive	RAS1	RAS2	RAS3	virtual
inactive	ϵ_i		f_{it}		0
active RAS1		ϵ_t	f_{tt}	f_{tt}	
active RAS2	f_{ti}	f_{tt}	ϵ_t	f_{tt}	f_{ta}
active RAS3		f_{tt}	f_{tt}	ϵ_t	
virtual	0		f_{at}		ϵ_a

where we see that the active–active block cannot be fully diagonalized, because the different RAS spaces cannot be mixed without altering the reference wave function. The first-order interacting space for a RASPT2 wave function contains all the eight classes considered for CASPT2 plus a fully active class given by $\hat{E}_{tu}\hat{E}_{vx}$, where all indices refer to active orbitals. This class is absent in CASPT2 because all possible excitations within the active space are already included in the reference or complementary space, but in RASPT2 some of these excitations would violate the RASSCF restrictions and should therefore be included in the first-order interacting space. In GASPT2, for every pair of not fully connected GAS spaces (where not all possible interspace excitations are allowed), there exists a fully active class as in RASPT2.

The implementation of a rigorous PT2 treatment of such a structure poses some challenges. First, including active–active excitations would require sorting out those that still fulfill the RAS or GAS
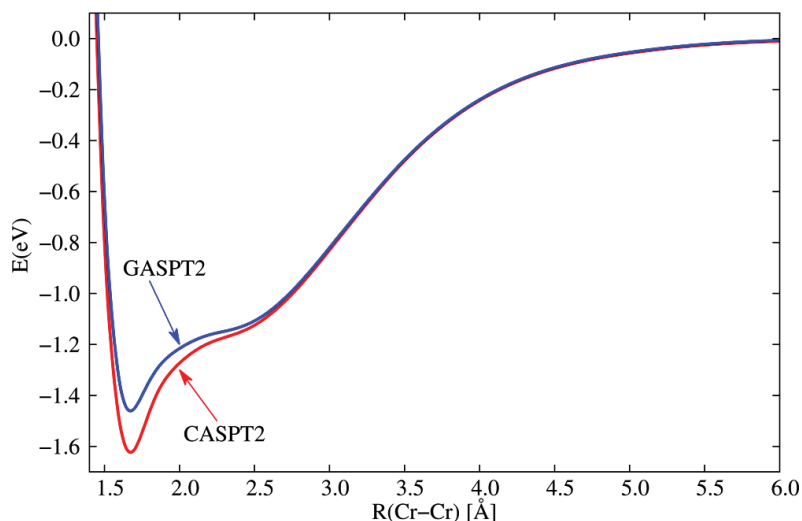
Figure 10.3 CASPT2(12,12) and GASPT2(12,12) potential energy curves for Cr_2. Reprinted with permission from ref. [26]. Copyright 2016, American Chemical Society.

restrictions from those that do not – the former should be removed, the latter should stay in the first-order interacting space. Furthermore, the computation of the matrix elements over $\hat{H}_0$ would require four-particle density matrices, which would be impractical for larger active spaces, and it is precisely for the use of large active spaces that we would like to use these methods. Therefore, in the current implementations of both RASPT2 and GASPT2, an approximation is employed in removing the active–active excitations from the first-order interacting space. The effect of this approximation is a degradation in the accuracy of the perturbative treatment, but it is expected that the possibility of using larger active spaces for the reference functions will compensate for it. As an example, fig. 10.3 shows a comparison of CASPT2 and GASPT2 curves for the difficult case of Cr_2, where a CASPT2 calculation is still affordable. The GASPT2 calculation was done by splitting the active space into six separate GAS spaces of two orbitals (bonding and antibonding) with two electrons, with no interspace excitations allowed. The difference between the two curves around the equilibrium geometry is attributed to the missing interspace excitations in the GASSCF treatment, which GASPT2 cannot recover due to the approximation employed.

10.3.3.4 Benchmark Results

A few words on the quality of the CASPT2 results are opportune at this point. The very first benchmark of the CASPT2 method was conducted by Andersson and Roos [27]. In this benchmark, molecular structures and binding energies were computed for some 27 molecules containing first- and second-row atoms (H–Ne). The results indicate an accuracy in bond distances and bond angles of 0.01 Å and 0–2 degrees, respectively. Furthermore, it is found that atomization energies are underestimated by 3–6 kcal mol^{-1} times the number of unpaired electrons created in the atomization process – the "correction" of this systematic error led to the so-called IPEA-shift technique, which is discussed below. Finally, the error in isogyric reactions – reactions which preserve the number of unpaired electrons – was in the range of −2.5–1 kcal mol^{-1}. In 2003 Guner and co-authors [28] reported benchmark results for 11 pericyclic reactions of unsaturated hydrocarbons. In particular they studied properties such as transition state structures, reaction enthalpies and entropies, and activation enthalpies and entropies. They reported, for example, mean absolute

deviations (MAD) for the reaction and activation enthalpies of 1.6 kcal mol^{-1} and 1.7 kcal mol^{-1}, respectively, at the CASPT2/6-31G*//CASSCF[3] level of approximation. Schreiber and co-workers [29] performed a comparative study of singlet (121 transitions) and triplet (71 transitions) vertical excitation energies computed at the CASPT2 and the CC3 levels. They reported MADs of 0.08 eV and 0.20 eV for the triplet and singlet transitions, respectively. These results, especially the organic spectroscopic data, were summarized by Serrano-Andrés in a special issue of IJQC [30] in honor of Björn O. Roos. In the same issue Pierloot [31] and Gagliardi [32] report on the state of the art of CASPT2 studies of transition metal compounds and actinide chemistry, respectively. Recently Budzák and collaborators established a database of ground-state and excited-state structures from some 35 molecules. Again, it was demonstrated that the CASPT2 method yields deviations of only 0.01 Å in any bond type when compared to the "gold standard" coupled-cluster methods [33].

10.3.3.5 IPEA Shift

As mentioned above, Andersson and Roos reported large deviations in atomization energies – up to 20 kcal mol^{-1}. Initially this caused much alarm, however, after some analysis it was established that this was a systematic error depending on the change of the number of unpaired electrons in the process under study. That is, an imbalance in the treatment of the reference Hamiltonian, $\hat{H}_0$, of closed-shell electrons with respect to unpaired electrons. In general this leads to too low bond and excitation energies. To remedy this artifact, Andersson introduced corrections to the Fock matrix, named G_1, G_2, and G_3 [24]. Although these corrections could relieve the problem in some cases, it was limited to situations where the energy gap between active and virtual orbitals is large. Ghigo and co-workers [34], however, suggested an alternative shift which seemed to reduce the systematic error significantly. In their analysis they identified that the systematic error could easily be resolved using a two-electron term, à la Dyall [35] (see section 10.3.5 on NEVPT2, where it is described in some detail), in the zeroth-order Hamiltonian. This would, however, lead to an increase in the computational expense, and a modification should preferably fit within the scope of a one-electron-operator-based formalism. Ghigo et al. further noted that for closed-shell orbitals the negative of the diagonal value of the Fock matrix corresponds to the ionization potential (IP), while for an empty orbital the value corresponds to an electron affinity (EA). This is trivially the case for the inactive and the virtual spaces, respectively, in the CASSCF model. For the active space, however, the situation is not that simple and the nature of the diagonal value is closely associated with the occupation number. To proceed, let us reformulate Eq. (10.123) as an interpolation between the two cases that the active orbital is doubly occupied ($-$IP) or empty ($-$EA),

$$f_{pp} = -\frac{1}{2}(D_{pp}(\text{IP})_p + (2 - D_{pp})(\text{EA})_p) \ . \tag{10.129}$$

Here we have an expression which is correct for $D_{pp} = 0$, 1, or 2 and in general mimics the behavior of the diagonal elements of the Fock matrix for the active orbitals. In particular this expression highlights that we will have a problem when an electron is attached to or removed from a singly occupied orbital. As a result, the denominators of the coefficients in the first-order correction of the wave function will be too small, the coefficients will be too large and thus contribute too much to the second-order correction to the energy. This is what had been empirically found – the perturbation energy is overestimated for open shells. Additionally, this will be yet another source for the manifestation of intruder states.

We would like to have a formalism such that we would use an electron affinity if we excite into an orbital and an ionization potential if we remove an electron from it. Ghigo and co-workers suggest

---

3 This acronym indicates that the study was conducted with a 6-31G* basis set, structures optimized at the CASSCF level, and finally the energies are computed with the CASPT2 method.

that this can be achieved by the introduction of a shift, the IPEA shift. In the case we move an electron into an orbital p, we would like to shift Eq. (10.129) by the value

$$\sigma^{\text{EA}} = \frac{1}{2} D_{pp}((\text{IP})_p - (\text{EA})_p) , \tag{10.130}$$

while in the case of the removal of an electron from the same orbital, we would like to shift by the value

$$\sigma^{\text{IP}} = -\frac{1}{2}(2 - D_{pp})((\text{IP})_p - (\text{EA})_p) . \tag{10.131}$$

This would be completely in line with the possibility of implementing a state-specific shift (see Eq. 10.65), where in this case the summation is limited to the states in the first-order interacting space and the shift is specific to each of the classes of states in this space. Subsequently Ghigo and co-workers took a leap of faith – considering that the definition of and the value of $(\text{IP})_p - (\text{EA})_p$ is not well defined nor is it in general known – they suggested the use of a single parameter σ, to be determined from accurate calculations or experimental results. They conclude their study by demonstrating, using a σ value of $0.25\,E_{\text{h}}$, that the computed dissociation energy errors for some 49 molecule can be reduced from more than 0.2 eV to less than 0.1 eV (see fig. 10.4). This value of the shift, $0.25\,E_{\text{h}}$, has since then been the default value in the MOLCAS quantum chemistry program package. [36] It should again be pointed out that this shift factor will be applied differently for each of the eight different classes of the first-order interacting space (see Eq. 10.65, and read refs. [37] and [38] for a detailed description of the implementation of the IPEA shift).

The IPEA-shift technique has, over the years, been an origin of irritation and many have criticized the approach. Some claim the shift makes the CASPT2 method semi-empirical, others state that the shift should be significantly larger, etc. [37, 39, 40]. To resolve the degree of "correctness" and validity of the IPEA shift, Zobel and co-workers [38] recently reported results from (*i*) an extensive study of literature of CASPT2 excitation energies, (*ii*) full configuration interaction benchmark results compared with CASPT2 results, and (*iii*) extensive studies using different values of σ. From the results Zobel and co-authors observed that the correction scales with the amount of dynamic correlation and is a function of the size of the basis set (see fig. 10.5). Thus they conclude that the notion of an universal IPEA shift does not have any support and that the practice should be abandoned.

10.3.4 MRMP2 Theory

About two years after the proposal and implementation of the CASPT2 theory, Hirao suggested a variation of the theory named multi-reference second-order Møller–Plesset perturbation theory (MRMP2) [14–17]. This proposal is similar in many respects to the CASPT2 theory and the differences will be briefly presented below. The major difference lies in the design of the $\hat{H}_0$ operator and the definition of the first-order interacting space.

10.3.4.1 The Partitioning of the Hamiltonian

The zeroth-order Hamiltonian is trivially defined as

$$\hat{H}_0 = \sum_i E_i^{(0)} |\Psi_i^{(0)}\rangle\langle\Psi_i^{(0)}| , \tag{10.132}$$

where $\Psi_i^{(0)}$ are CASSCF wave functions and $E_i^{(0)}$ are the reference energies defined as follows. First, the reference energy is defined as a sum over orbital-energy-like terms, generalized orbital energies,

$$E_i^{(0)} = \sum_p \epsilon_{pp} \langle\Psi_i^{(0)}|\hat{E}_{pp}|\Psi_i^{(0)}\rangle , \tag{10.133}$$
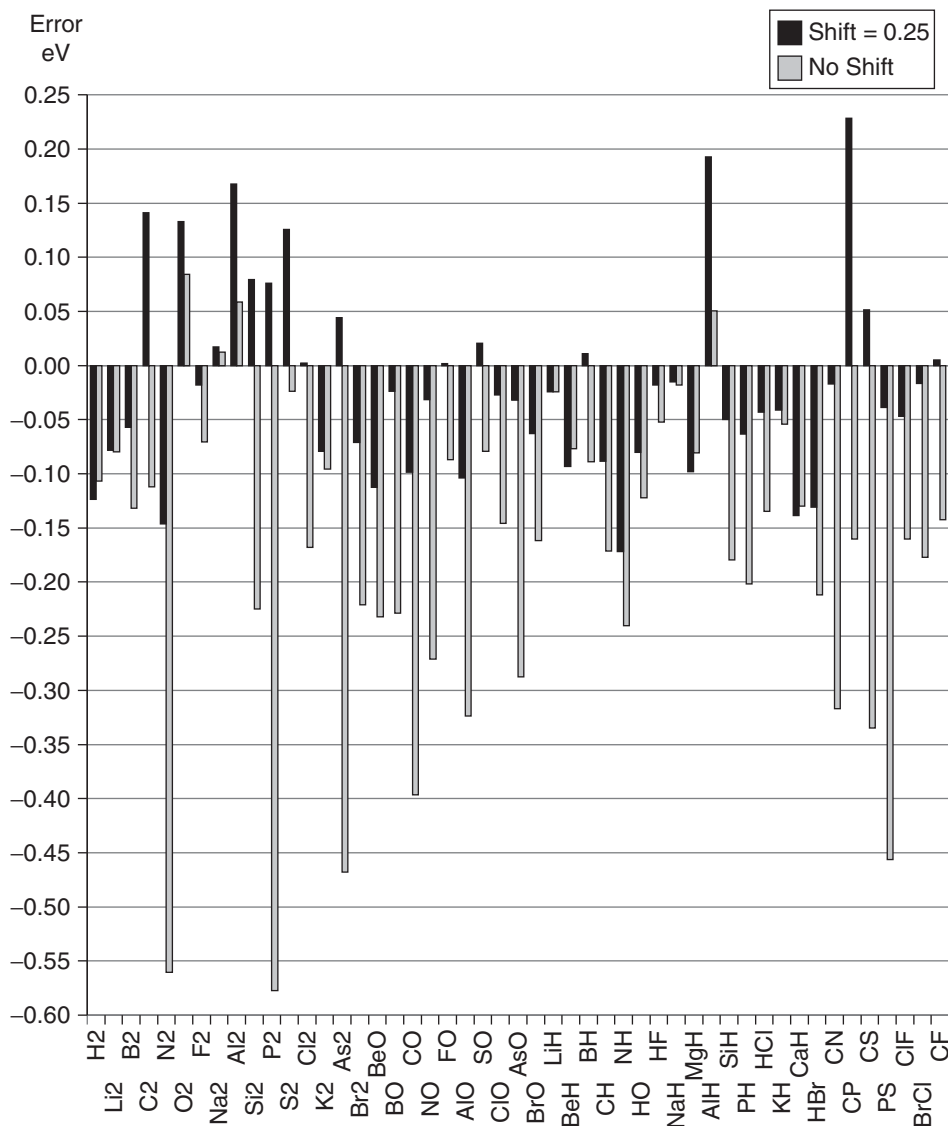
Figure 10.4 Errors in the dissociation energy for 49 diatomic molecules obtained with the IPEA shift parameter equal to 0.25 E_h and with no IPEA shift. Reprinted with permission from ref. [34]. Copyright 2004, Elsevier.

where $\hat{E}_{pp}$ is the number operator (it returns the occupation number for the pth orbital). It is important to note that these matrix elements are given in the orbital representation that diagonalizes the diagonal blocks of the Fock matrix. In this orbital representation the generalized orbital energies are computed as

$$\epsilon_{pq} = h_{pq} + \sum_{rs} D_{rs} \left[(pq|rs) - \frac{1}{2}(pr|qs) \right] . \tag{10.134}$$

10.3.4.2 The First-Order Interacting Space

While the CASPT2 and NEVPT2 methods (see section 10.3.5 below) define the first-order interacting space by the use of replacement operators applied to the contracted CASSCF wave function(s),
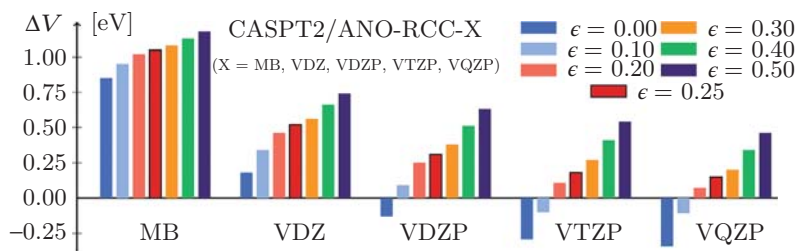
Figure 10.5 Mean signed errors (MSEE) in eV of CASPT2 vertical excitation energies compared to experimental reference data for different ANO-RCC basis sets and IPEA shift values ϵ (in E_h) Reprinted from ref. [38], published by The Royal Society of Chemistry, distributed under Attribution 3.0 Unported (CC BY 3.0) license.

the MRMP2 method defines the space by collecting all SDs that correspond to a single or double replacement relative to the SDs that span the CASSCF CI space. In that respect this space is larger than the corresponding space in, for example, the CASPT2 approach, however it comes with the benefit that the representation provides a diagonal representation of $\hat{H}_0$. This space has, of course, to be subjected to elimination of linear dependences.

We conclude this section by noting that the MRMP2 method is not used extensively and there is a lack of any significant amount of benchmarking in the literature, compared to other MRPT methods.

10.3.5 NEVPT2 Theory

The *n*-electron valence state for multi-reference perturbation theory (NEVPT2) was introduced some 10 years after the CASPT2 method [18]. At the time three different versions of the method were introduced: the totally uncontracted, the partially contracted (PC), and the strongly contracted (SC). Different to CASPT2, NEVPT2 uses an $\hat{H}_0$ operator that includes two-electron terms, and the method is attributed to avoid the problems with intruder states, which frequently cause issues in CASPT2 studies. Thus, two of the major advantages that the method claims are (*i*) exempt from the intruder state problem and (*ii*) the method is size consistent.

The totally uncontracted method entails a high computational cost and has not been used in practical applications (for some recent developments, see the end of section 10.3.6). The two latter versions, the PC-NEVPT2 and the SC-NEVPT2 versions, however, are in common use. Hence, we will discuss these methods briefly.

10.3.5.1 The Partitioning of the Hamiltonian
Below, we will describe the $\hat{H}_0$, however, in the case of NEVPT2 this requires the generation of the functions in the first-order interacting space – also known as *perturbers*, since they will perturb the wave function as we include corrections through the perturbation expansion. We will, at this point, assume that these perturbers are generated and denoted, $\Psi_l^{\{k\}}$ and $\Psi_{l\mu}^{\{k\}}$, for the strongly and the partially contracted NEVPT2 versions, respectively. Here k and l index the class and subclass, respectively, of the first-order interacting functions, and μ is a particular member of such a sub-class – this will become clearer in the next section. We can now write two different definitions of the $\hat{H}_0$ operator, namely

$$\hat{H}_0^{\text{SC}} = \sum_m^{\text{CAS}} |\Psi_m^{(0)}\rangle E_m^{(0)} \langle\Psi_m^{(0)}| + \sum_{kl} |\Psi_l^{\{k\}}\rangle E_l^{\{k\}} \langle\Psi_l^{\{k\}}| \tag{10.135}$$

and

$$\hat{H}_0^{\text{PC}} = \sum_m^{\text{CAS}} |\Psi_m^{(0)}\rangle E_m^{(0)} \langle \Psi_m^{(0)}| + \sum_{kl\mu} |\Psi_{l\mu}^{\{k\}}\rangle E_{l\mu}^{\{k\}} \langle \Psi_{l\mu}^{\{k\}}| \,, \tag{10.136}$$

for the strongly and partially contracted versions of NEVPT2, respectively. The first term involves all possible states of the CASSCF wave function, while the second term includes the normalized perturbers of the first-order interacting space. Since, after the next section, we have these perturbers, we only need to find a way to obtain the associated "eigenenergies" $E_{l/l\mu}^{\{k\}}$. To do this, we will employ Dyall's Hamiltonian, $\hat{H}^D$ [35] (see below), in which the associated eigenvalues are simply computed as

$$E_l^{\{k\}} = \langle \Psi_l^{\{k\}} | \hat{H}^D | \Psi_l^{\{k\}} \rangle \tag{10.137}$$

for the strongly contracted case. For the partially contracted case a more elaborate procedure is engaged. First a projected Hamiltonian, based on $\hat{H}^D$, is formed with eigenfunctions and eigenvalues given by

$$\hat{P}_l^{\{k\}} \hat{H}^D \hat{P}_l^{\{k\}} \Psi_{l\mu}^{\{k\}} = E_{l\mu}^{\{k\}} \Psi_{l\mu}^{\{k\}}, \tag{10.138}$$

where $\Psi_{l\mu}^{\{k\}}$ are spanned by the linearly independent functions $\Phi_{lm}^{\{k\}}$ of the space $S_l^{\{k\}}$.

Before we continue, let us mention Dyall's Hamiltonian – here presented in a form invariant to unitary rotations inside each orbital class (inactive, active, and virtual). This Hamiltonian is introduced as an approximation to the electronic Hamiltonian and is subdivided, for computational efficiency, into an inactive–virtual part (iv) and an active part (act) as

$$\hat{H}^D = \hat{H}_{\text{iv}} + \hat{H}_{\text{act}} \,, \tag{10.139}$$

where the inactive–virtual part is a simple one-electron diagonal operator,

$$\hat{H}_{\text{iv}} = \sum_{ij} f_{ij} \hat{E}_{ij} + \sum_{ab} f_{ab} \hat{E}_{ab} + C, \tag{10.140}$$

where we use the index definitions of Chapter 1, section 1.5. The constant C and the Fock matrix elements f_{pq} are defined below. For the active part of the operator, $\hat{H}_{\text{act}}$, we have the following expression,

$$\hat{H}_{\text{act}} = \sum_{tu} h_{tu}^{\text{eff}} \hat{E}_{tu} + \frac{1}{2} \sum_{tuvx} (tu|vx) \hat{e}_{tuvx} \,, \tag{10.141}$$

where $h_{tu}^{\text{eff}} = h_{tu} + \sum_i (2(tu|ii) - (ti|ui))$ are the effective one-electron integrals, introducing the field of the inactive electrons into the active part of the Hamiltonian. Furthermore, the constant C is chosen to be,

$$C = 2 \sum_i h_{ii} + \sum_{ij} [2(ii|jj) - (ij|ji)] - 2 \sum_k^{\text{frozen}} \epsilon_k \,, \tag{10.142}$$

with which the result is equivalent to the full Hamiltonian within the CAS space,

$$\hat{H}^D \Psi_m^{(0)} = E_m^{(0)} \Psi_m^{(0)}. \tag{10.143}$$

Finally the Fock matrix elements f_{pq} are defined as

$$f_{ij} = -\langle \Psi_m^{(0)} | \hat{a}_i^\dagger \hat{H} \hat{a}_j | \Psi_m^{(0)} \rangle + \delta_{ij} E_m^{(0)} \tag{10.144}$$

and

$$f_{ab} = \langle \Psi_m^{(0)} | \hat{a}_a \hat{H} \hat{a}_b^\dagger | \Psi_m^{(0)} \rangle - \delta_{ab} E_m^{(0)}. \tag{10.145}$$

In this fashion the Dyall Hamiltonian is equivalent to the Møller–Plesset $\hat{H}_0$ in the inactive and the virtual space while in the active space it is the complete active space Hamiltonian. The latter ensures that the NEVPT2 approach is a size-consistent method.

10.3.5.2 The First-Order Interacting Space

Here we will describe the first-order interacting space (also denoted the SD-space) for the strongly contracted and the partially contracted schemes. In the SC-NEVPT2 version we will use only one perturber from each class in the first-order interacting space, while in the PC-NEVPT2 method we will include all possible states of the SD-space (just as in the CASPT2 method). The classes are similar to the classification in the CASPT2 method (see section 10.3.3.2 above). The index k, ranging from -2 to 2, represents the change in the number of electrons in the active space. An additional index, l, is used to discriminate between subclasses when there are several types of operators belonging to the same k-class. We note that we have eight different types of subclasses, just as in the CASPT2 method, although the primary classification scheme there was based on internal, semi-internal, and external excitations, instead of the k index.

In the SC-NEVPT2 method a single perturber per subclass is generated as follows,

$$\Psi_l^{\{k\}} = \hat{P}_l^{\{k\}} \hat{H} \Psi_m^{(0)}, \tag{10.146}$$

where $\Psi_m^{(0)}$ is the reference function of state m, and $\hat{P}_l^{\{k\}}$ is the projector onto $S_l^{\{k\}}$, the subspace of all states in subclass l of class k. To some extent this is similar to the first iteration in a CI solver algorithm. Since only one perturber is picked from each type of class there will be no problem with linear dependences.

For the PC-NEVPT2 method, however, we will employ all possible states in each subclass, just as in the CASPT2 method, and as in the CASPT2 case we will have possible linear dependences. Thus a similar procedure is adopted to generate a set of linearly independent functions, $\Phi_{lm}^{\{k\}}$, of subspace $S_l^{\{k\}}$. The final functions, $\Psi_{l\mu}^{\{k\}}$, to be used in the $\hat{H}_0$ operator, are generated by finding the eigenvalues and eigenvectors of the projection of Dyall's Hamiltonian on $S_l^{\{k\}}$ Eq. (10.138).

To sum up the technical description of the NEVPT2 formalism and highlight significant differences: (*i*) NEVPT2 uses a two-electron $\hat{H}_0$ resulting in longer computation time and the need of up to explicit four-particle density matrices, on the other hand, in NEVPT2 the Hamiltonian is actually block-diagonal for each excitation class, which reduces the size of the problem, (*ii*) the NEVPT2 method is strictly size extensive – CASPT2 is not, and (*iii*) the NEVPT2 method does not, from a practical perspective, suffer from intruder-state problems. The latter property is due to the fact that with the NEVPT2 partitioning of the Hamiltonian the energies of the perturbing wave functions – the states in the first-order interacting space – are far from degenerate with the energy of the zeroth-order state. The worst case is associated with an excited state corresponding to a semi-internal excitation which will effectively generate an energy denominator corresponding to the sum of the ionization potential of an active electron and the energy of a virtual orbital. The former should be negative and can at worst be close to zero for a Rydberg orbital, while the latter is positive (for neutral species) and rarely close to zero. Therefore, calculations for electronic
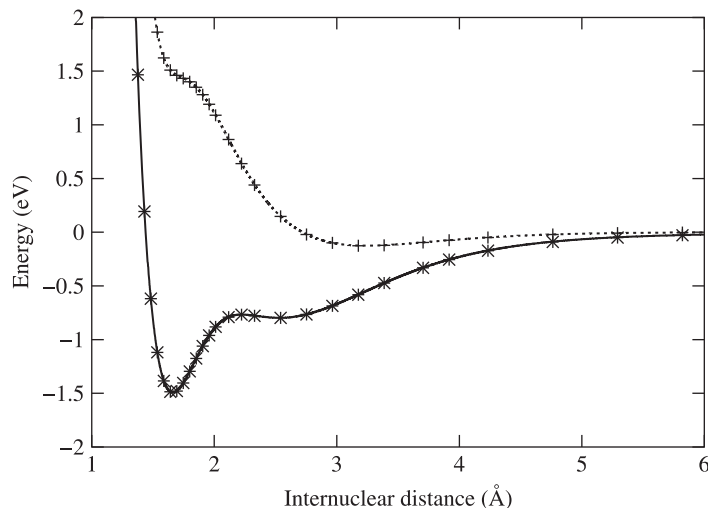
Figure 10.6 Potential energy curves for the $^1\Sigma^+$ states of Cr_2 (ANO basis set, (12/12) CAS-CI space). Dashed line (CASSCF) and full line (SC-NEVPT2). Reprinted with permission from ref. [19]. Copyright 2001, Elsevier.

excited states should be possible with NEVPT2 without experiencing any intruder state problem. For a detailed discussion on this we recommend the reader to consult ref. [41].

Finally, extensive benchmarking of the NEVPT2 method does not exist at present. However, let us again point out that different to CASPT2 the intruder state problem is effectively removed and the result is strictly size consistent. The former is explicitly demonstrated in fig. 10.6, in which the curve was generated without the need for any type of level-shifting technique. Preliminary benchmark studies by Angeli and co-workers [42] on valence and Rydberg transitions in formaldehyde and acetone indicated errors in the range of 0.1–0.2 eV and an absence of intruder states. A recent benchmark by Shapiro et al. [43] on vertical excitation energies of some 28 medium sized organic molecules, does not provide decisive arguments for any discrimination between the PC-NEVPT2, SC-NEVPT2, and SS-CASPT2 methods based on the accuracy of the predictions – if any, SS-CASPT2 has a small edge (see fig. 10.7). For example, the mean unsigned error is 0.28 eV, 0.23 eV, and 0.21 eV for the methods as ordered above. Rather than accuracy, it will be qualities such as computer efficiency and others that will, from case to case, be the determining factor in the selection of which method to use.

10.3.6 Performance Improvements

The use of the CASPT2 and NEVPT2 methods has been very successful in predicting and understanding a number of phenomena. However, the conventional implementations suffer from various bottlenecks such as the transformation of the integrals to the molecular orbital basis or the evaluation of the amplitudes. The former scales effectively as $\mathcal{O}(n^5)$ where n is the size of the basis set, while the latter suffers from the "exponential wall" typical of a conventional configuration-interaction treatment of the wave function. This limits the use of the methods to modest-sized systems with active spaces of up to about 18 electron in 18 orbitals.

For moderate active spaces, the integral transformation step tends to dominate. This integral-transformation bottleneck was effectively removed by the use of a CASPT2 implementation based on Cholesky decomposed (CD) integrals, the CD-CASPT2 approach [44]. This
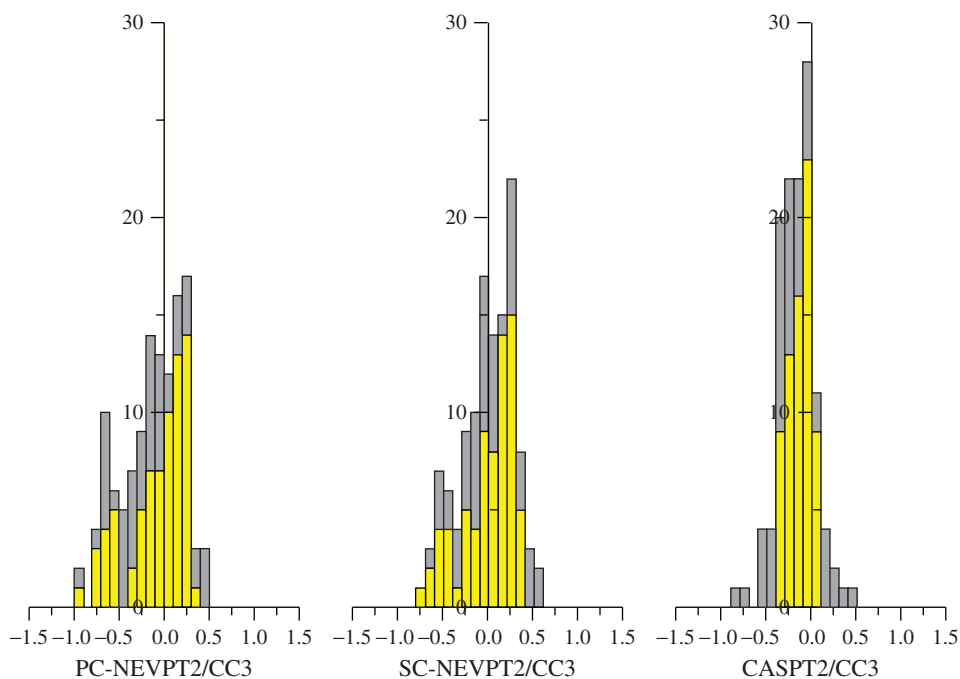
Figure 10.7 Histograms of the frequency of deviations in the singlet excitation energies in eV. Statistics of all excitation energies are shown in gray and those below 7 eV are represented by yellow bars. Reprinted with permission from ref. [43]. Copyright 2013, American Chemical Society.

reduced the scaling of the integral transformation by one order of magnitude while simultaneously reducing cache faults and I/O requests. Benchmarks demonstrate that CD in combination with CASSCF/CASPT2 reduces the computer time by a factor of four up to 1–2 orders of magnitude, with a loss of accuracy for excitation energies of organic molecules of about 0.001 eV [45]. Another early, but not as successful, approach is the frozen natural orbital CASPT2 (FNO-CASPT2) method [46, 47]. In this approach, an approximation to the one-particle density matrix corrected up to second order is diagonalized in the virtual–virtual block. Subsequently eigenvectors are eliminated if the corresponding eigenvalue is below a predefined threshold. It has been demonstrated that the use of this kind of reduced virtual space recovers some 95 % of the second-order correction to the energy while significantly reducing the computational expense. Some more recent and significant progress, however, has been made for both the CASPT2 and the NEVPT2 methods with respect to molecular-size scaling given a fixed active space – an ability which can be of significance in the study of transition metals with large ligands. In these approaches the effective number of amplitudes in the first-order interacting space is dramatically reduced without any significant loss of accuracy by a combined initial generation of projected atomic orbitals (PAO) [48] and a subsequent generation of pair natural orbital (PNO) [49]. Both the PNO-CASPT2 [50] and the domain-based local PNOs flavor of NEVPT2 (DBLPNO-NEVPT2) [51] methods have demonstrated, given a fixed active space, linear scaling with respect to molecular size in combination with insignificant loss of accuracy – recovering 99.9 % of the second-order correction to the energy. More recent developments using density matrix renormalization group (DMRG) approximations to treat large active spaces in CASPT2 [52–56] and NEVPT2 [57–59] – the DMRG-CASPT2 and DMRG-NEVPT2 methods – have been very successful in removing the

conventional limitation on the size of the active space. Finally, the progress in overcoming the bottleneck of the higher-order *n*-particle matrix evaluation should be mentioned [60–64]. In these studies, the efficiency of different approaches has been explored, with mixed results, namely: (*i*) approximating the cumulants of the higher-order reduced density matrices [60, 65], (*ii*) using a matrix product state representation (found, for example, in DMRG) in the computation of the first-order interacting space – called MPS-PT2 – in state-specific or quasi-degenerate (see below) PT [61, 62, 66], (*iii*) using a time-dependent formulation of MRPT [63], (*iv*) and a projection of the first-order interacting space in an SC-NEVPT2 setting onto a renormalized Hilbert space [64].

10.4 Quasi-Degenerate Perturbation Theory

In the course of using the SS-CASPT2 method, it was evident that the approach is only usable when the reference functions are close to qualitatively correct, and that the first-order correction to the wave function would be sufficient for a quantitatively correct energy only if the perturbation expansion terminates at second order. Anything else would be wishful thinking. In the literature there are ample cases in which the reference CASSCF wave functions are qualitatively wrong due to the lack of dynamic electron correlation – see, for example, the case of the *V*-state of ethylene, in which CASSCF introduces an artificial mixing of Rydberg and valence character, something the SS-CASPT2 approach never recovers from [67]. The trivial solution to this problem could have been to increase the active space or to go beyond second order in the perturbation series – which sounds trivial but is expensive, complicated, or not even feasible. A more pragmatic approach involves formulating a multi-state version of perturbation theory, something that is not straightforwardly done from the single-state theory. What is the problem? Well, the single-state approach has the problem that, if applied to several states at the same time, it does not allow the reference states to trivially mix with each other as the perturbation is introduced order by order. To avoid this, multi-state perturbation theories use an effective Hamiltonian procedure to produce corrected energies originating from corrected *model states*, which are linear combinations of the original *reference states*. This gives rise to quasi-degenerate perturbation theory (QDPT), the basics of which we will see in this section.

To proceed with a multi-state theory, however, some new tools have to be derived – the Bloch equations and the wave operator (see, for example, refs. [68–70]). We start by defining a model space – or reference space – commonly denoted *P*, spanned by some eigenfunctions of the Hamiltonian of the unperturbed system $\hat{H}_0$, the *reference states*,

$$\Psi_i^{(0)} \quad i = 1, \dots, d, \tag{10.147}$$

where *d* is the dimension of the model space. We also define the operators

$$\hat{P} = \sum_{i=1}^{d} |\Psi_i^{(0)}\rangle\langle\Psi_i^{(0)}| \text{ , and} \tag{10.148}$$

$$\hat{Q} = 1 - \hat{P}. \tag{10.149}$$

The operator $\hat{P}$ can now be used to project the exact wave function of state *i* onto the the model space as

$$\hat{P}\Psi_i = \Psi_i^P = \sum_{j=1}^{d} c_{ij}\Psi_j^{(0)}, \tag{10.150}$$

where Ψ_i^P is the *model state* of state Ψ_i, expressed by a linear combination of the reference states in the model space. These linear combinations are found by diagonalization of the multi-state Hamiltonian matrix. The wave operator, $\hat{\Omega}$, does the corresponding back-transformation, acting on the model states it generates the exact states,

$$\Psi_i = \hat{\Omega}\Psi_i^P \quad i = 1, \dots, d, \tag{10.151}$$

where we note the similarities to Eq. (10.19) for the single-state case – both provide a recipe to find the corrections of the wave function(s) to arbitrary order. For completeness, $\hat{\Omega}$ is defined as a null operator when operating on the complementary space, i.e., $\hat{\Omega}\hat{Q} = 0$ and therefore $\hat{\Omega}\hat{P} = \hat{\Omega}(1 - \hat{Q}) = \hat{\Omega}$.

We will here derive the Bloch equation – the master equation to express the wave operator – in which an explicit dependence on the exact energy is avoided. First, we start by left-multiplying the "exact" Schrödinger equation with $\hat{\Omega}\hat{P}$, that is,

$$\hat{\Omega}\hat{P}(\hat{H}_0 + V)\Psi_i = E_i\Psi_i . \tag{10.152}$$

The operator $\hat{\Omega}\hat{P}$ will project any exact wave function to the reference space and then subsequently regenerate the exact wave function. Hence, the operator will leave an exact wave function untouched, and the RHS of the equation is identical to the original expression. Using $[\hat{H}_0, \hat{P}] = 0$, and with the help of Eqs. (10.150) and (10.151), the following expression is found

$$(\hat{\Omega}\hat{H}_0 + \hat{\Omega}\hat{P}V\hat{\Omega})\Psi_i^P = E_i\Psi_i , \tag{10.153}$$

and by using Eq. (10.151), $\hat{\Omega}\hat{P} = \hat{\Omega}$, and the original Schrödinger equation to eliminate the energy, one gets

$$(\hat{\Omega}\hat{H}_0 + \hat{\Omega}V\hat{\Omega})\Psi_i^P = (\hat{H}_0\hat{\Omega} + V\hat{\Omega})\Psi_i^P, \tag{10.154}$$

which holds for the model space. Since the operators on both sides are null when operating on the complementary space, it holds for any state, and can be paraphrased into the generalized form of the Bloch equation,

$$[\hat{\Omega}, \hat{H}_0] = V\hat{\Omega} - \hat{\Omega}V\hat{\Omega}. \tag{10.155}$$

In the single-state case the equation can be rewritten as

$$(E_i^{(0)} - \hat{H}_0)\hat{\Omega} = V\hat{\Omega} - \hat{\Omega}V\hat{\Omega}. \tag{10.156}$$

Expanding the wave operator in powers of the perturbation with $\hat{\Omega}^{(0)} = \hat{P}$ [68] we have

$$\hat{\Omega} = \hat{P} + \hat{\Omega}^{(1)} + \hat{\Omega}^{(2)} + \dots, \tag{10.157}$$

and finally, due to the recursive nature of the Bloch equation we have

$$[\hat{\Omega}^{(n)}, \hat{H}_0] = \hat{Q}V\hat{\Omega}^{(n-1)} - \sum_{k=1}^{n-1} \hat{\Omega}^{(n-k)}V\hat{\Omega}^{(k-1)} . \tag{10.158}$$

Then we get the wave operator, up to third order, as

$$[\hat{\Omega}^{(1)}, \hat{H}_0] = \hat{Q}V\hat{P} = \hat{Q}\hat{H}\hat{P}, \tag{10.159}$$

$$[\hat{\Omega}^{(2)}, \hat{H}_0] = \hat{Q}V\hat{\Omega}^{(1)} - \hat{\Omega}^{(1)}V\hat{P}, \text{ and} \tag{10.160}$$

$$[\hat{\Omega}^{(3)}, \hat{H}_0] = \hat{Q}V\hat{\Omega}^{(2)} - \hat{\Omega}^{(2)}V\hat{P} - \hat{\Omega}^{(1)}V\hat{\Omega}^{(1)} . \tag{10.161}$$

In the single-state case the LHS is replaced with $(E_i^0 - \hat{H}_0)\hat{\Omega}_i^{(n)}$, where the wave operator is unique to the state i, and the general wave operator is now expressed as

$$\hat{\Omega}^{(n)} = \sum_{i=1}^d \hat{\Omega}_i^{(n)} |\Psi_i^{(0)}\rangle\langle\Psi_i^{(0)}| \ . \tag{10.162}$$

The effective Hamiltonian, by definition, is written as

$$\hat{H}^{\text{eff}} = \hat{P}\hat{H}\hat{\Omega}\hat{P}, \tag{10.163}$$

which we note only acts on the reference space. The model states, $\Psi_i^{(0)}$, $i = 1, \dots, d$, are eigenfunctions of the effective Hamiltonian,

$$\hat{H}^{\text{eff}}\Psi_i^{(0)} = E_i\Psi_i^{(0)}, \tag{10.164}$$

and the effective Hamiltonian up to second order is

$$\hat{H}^{\text{eff}[2]} = \hat{P}\hat{H}\hat{P} + \hat{P}\hat{H}\hat{\Omega}^{(1)}\hat{P}. \tag{10.165}$$

We note at once that the energy corrected up to first order is computed as

$$E_i^{[1]} = \langle\Psi_i^{(0)}|\hat{H}_0 + V|\Psi_i^{(0)}\rangle \ , \tag{10.166}$$

and the second-order correction to the energy is expressed as

$$E_i^{(2)} = \langle\Psi_i^{(0)}|\hat{H}|\Psi_i^{(1)}\rangle \ . \tag{10.167}$$

Let us look at the first-order corrected wave function in terms of a single- and multi-state formalism. The first-order correction to the wave function in a single-state formalism is expressed as

$$\Psi_i^{(1)} = \hat{\Omega}_i^{(1)}\Psi_i^{(0)} = \sum_{\substack{j=1 \\ j\neq i}}^{\infty} C_{ij}\Psi_j^{(0)}, \tag{10.168}$$

while in a multi-state approach we have (see Eqs. 10.150 and 10.162)

$$\Psi_i^{(1),P} = \hat{\Omega}^{(1)}\Psi_i^P \tag{10.169}$$

$$= \sum_{j=1}^d \hat{\Omega}_j^{(1)}|\Psi_j^{(0)}\rangle\langle\Psi_j^{(0)}|\Psi_i^P\rangle$$

$$= \sum_{j=1}^d c_{ji} \sum_{\substack{k=1 \\ k\neq j}}^{\infty} C_{jk}\Psi_k^{(0)}$$

$$= \sum_{j=1}^d c_{ji}\Psi_j^{(1)}. \tag{10.170}$$

Note that this does not imply that the first-order corrected model states are orthogonal, they will only be so at infinite order. The coefficients, C_{jk}, are computed from the general Bloch equation according to

$$\sum_{j\neq i} C_{ij}\langle\Psi_k^{(0)}|(E_i^{(0)} - \hat{H}_0)|\Psi_j^{(0)}\rangle = \langle\Psi_k^{(0)}|\hat{H}|\Psi_i^{(0)}\rangle \ , \tag{10.171}$$

which simplifies to

$$C_{ik} = \frac{\langle\Psi_k^{(0)}|V|\Psi_i^{(0)}\rangle}{E_i^{(0)} - E_k^{(0)}} \ , \tag{10.172}$$

and now allows the computation of the second-order energy as described above. The matrix elements of the effective Hamiltonian are now expressed as

$$\langle \Psi_i^{(0)} | \hat{H}^{\text{eff}[2]} | \Psi_i^{(0)} \rangle = E_i^{[2]} \tag{10.173}$$

and

$$\langle \Psi_i^{(0)} | \hat{H}^{\text{eff}[2]} | \Psi_j^{(0)} \rangle = \langle \Psi_i^{(0)} | \hat{H} | \Psi_j^{(0)} + \Psi_j^{(1)} \rangle \, . \tag{10.174}$$

Given these matrix elements we can now proceed with a diagonalization and derive the model states of the model space. To conclude, this approach can be viewed as *perturb-then-diagonalize*, in which we use perturbation theory to set up the effective Hamiltonian which is subsequently diagonalized.

10.5 Multi-State Multi-Configurational Reference Perturbation Methods

In the pursuit of practical methods to overcome single-state MRPT limitations, the tools of quasi-degenerate perturbation theory come in handy. Defining a relevant model space of several states provides a pathway in which (*i*) the computational effort only increases linearly with the number of states in the model space and (*ii*) artificial redundancies (due to deficiencies on the CASSCF treatment) are removed, in one single step, under the influence of the perturbation (that is, including electron–electron correlation). These methods will now be of the nature *diagonalize-then-perturb-then-diagonalize* approach, in the sense that we will use basis functions in the setup of the effective Hamiltonian which are based on a *diagonalize-then-perturb* technique. Specifically, below we will describe two such methods in some detail – the multi-state and the extended multi-state CASPT2 (MS-CASPT2 and XMS-CASPT2). Similar approaches have successfully been applied to both the MRMP2 (multi-configurational quasi-degenerate second-order perturbation theory, MCQDPT2) [71, 72] and NEVPT2 single-state theories (QD-NEVPT2) [73].

10.5.1 Multi-State CASPT2 Theory

As described above, QDPT requires that $\hat{H}_0$ and $\hat{H}_i^{\text{eff}[2]}$ be unique for all the reference states. However, much of the success of SS-CASPT2 lies in the use of an $\hat{H}_0$ that is specifically tailored to the state under study. In 1998 Finley and co-workers suggested a hybrid approach of SS-CASPT2 with quasi-degenerate perturbation theory [74] – the multi-state CASPT2 (MS-CASPT2) method – to improve the SS-CASPT2 method in those cases where the reference CASSCF wave function does not meet the qualitative requirements for a fast convergence of the perturbation expansion, additionally it should remove erroneous behavior of SS-CASPT2 at points of degeneracy. This approach is very similar to what was described above for standard quasi-degenerate perturbation theory, that is a *perturb-then-diagonalize* approach. However, in this case the starting reference functions have already been the subject of a diagonalization procedure – the CASSCF procedure – and in this aspect the MS-CASPT2 procedure is a *diagonalize-then-perturb-then-diagonalize*-like procedure. The suggestion of Finley et al. revolves entirely around Eqs. (10.173) and (10.174). How are those equations to be used in association with the SS-CASPT2 approach? In this respect the authors suggest the following. First, and maybe trivial, the diagonal elements in the effective Hamiltonian are simply the individual SS-CASPT2 energies,

$$\langle \Psi_i^{(0)} | \hat{H}^{\text{MS-CASPT2}} | \Psi_i^{(0)} \rangle = E_i^{\text{SS-CASPT2}} = \langle \Psi_i^{(0)} | \hat{H}_i^{\text{eff}[2]} | \Psi_i^{(0)} \rangle \, . \tag{10.175}$$

At this point we also recognize the significant difference to Eq. (10.173): the effective Hamiltonian operator and the corresponding implicit partitioning of the Hamiltonian is *state-specific* [75]. This raises a problem when we proceed to the computation of the off-diagonal elements of the effective Hamiltonian matrix, that is,

$$\langle\Psi_i^{(0)}|\hat{H}_j^{\text{eff}[2]}|\Psi_j^{(0)}\rangle = \langle\Psi_i^{(0)}|\hat{H}|\Psi_j^{(1)}\rangle \neq \langle\Psi_j^{(0)}|\hat{H}|\Psi_i^{(1)}\rangle = \langle\Psi_j^{(0)}|\hat{H}_i^{\text{eff}[2]}|\Psi_i^{(0)}\rangle \,, \tag{10.176}$$

which for obvious reason produces a non-symmetric Hamiltonian matrix. In the current implementation this problem is fixed with a simple symmetrization of the matrix with the limitation that results should be questioned if corresponding off-diagonal elements are large and different. The MS-CASPT2 results are generally in line with SS-CASPT2 results. In the original report by Finley and co-workers, the authors demonstrated explicitly how the MS-CASPT2 approach solves the problem with the artificial valence–Rydberg mixing in the *V*-state of ethane, and how MS-CASPT2, unlike SS-CASPT2, provides qualitatively correct results around the avoided crossing of the ionic–neutral states in the dissociation of LiF (see fig. 10.8). When MS-CASPT2 and SS-CASPT2 results differ, however, the prudent user will have to address the issue of whether or not the MS-CASPT2 results should be trusted. In particular, it has been demonstrated that the MS-CASPT2 method resolves cases with accidental degeneracy at the CASSCF level if it does not persist at the SS-CASPT2 level. On the other hand, the MS-CASPT2 is known to provide erroneous behavior around conical intersections [76, 77] and other cases of near-degeneracy. The latter problem is known to be a consequence of the use of state-specific effective operators, a procedure will be suggested below to remove these problems.

10.5.2 Extended MS-CASPT2 Theory

Granovsky analysed MCQDPT theory and put forward qualitative criteria which a quasi-degenerate perturbation theory should satisfy [77]. One of these is that the eigenvalues of the effective Hamiltonian should be invariant to the particular basis functions selected to span the model space, as long as they span the same subspace. In this respect Granovsky suggested that the $\hat{H}_0$ operator should be extended from

$$\hat{H}_0 = \sum_i |\Psi_i^{(0)}\rangle\langle\Psi_i^{(0)}|\hat{F}|\Psi_i^{(0)}\rangle\langle\Psi_i^{(0)}| \tag{10.177}$$

to

$$\hat{H}_0 = \sum_{ij} |\Psi_i^{(0)}\rangle\langle\Psi_i^{(0)}|\hat{F}|\Psi_j^{(0)}\rangle\langle\Psi_j^{(0)}| \,, \tag{10.178}$$

where $\hat{F}$ is a Fock-like operator, as in CASPT2. Thus, the $\hat{H}_0$ operator is explicitly independent of the selection of the basis functions spanning a particular model space. This new technique is denoted as extended MCQDPT (XMCQDPT). This extension has been applied to MS-CASPT2 theory yielding the method known as XMS-CASPT2 [78]. In this approach the $\hat{H}_0$ matrix is diagonalized. The "non-extended" procedure is then continued in the basis of these eigenfunctions $\Psi_i^{X(0)}$.

It should be noted that the XMS-CASPT2 approach requires the use of a state-specific partitioning to be abandoned, as used in the MS-CASPT2 method. The state-specific approach and the MS-CASPT2 method have been successful in predicting vertical excitation energies. The
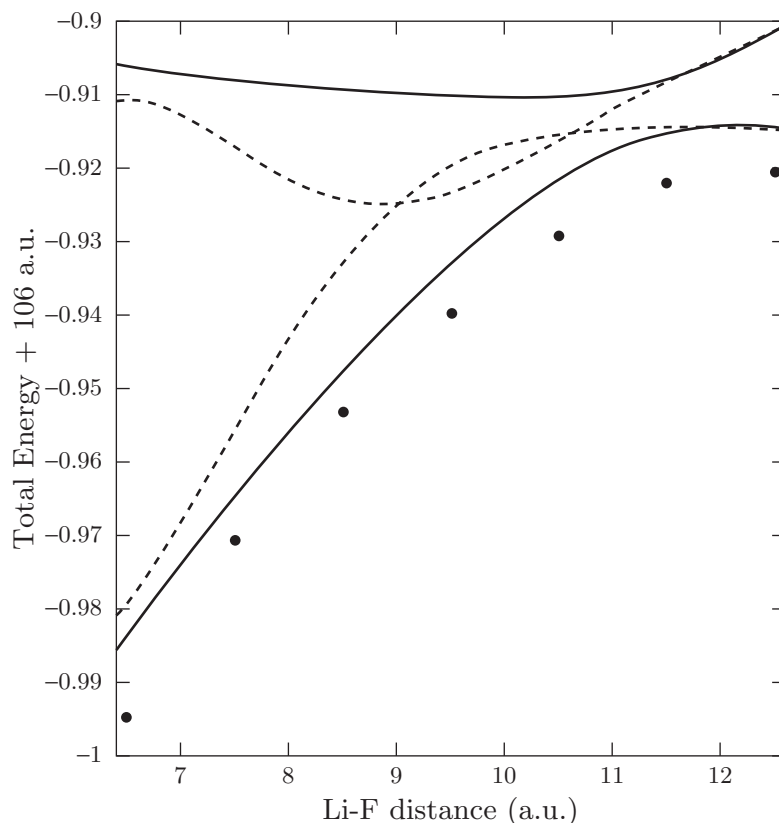
Figure 10.8 Potential energy curves for the two lowest $^1\Sigma^+$ states of LiF in the region of their neutral–ionic avoided crossing. The dotted lines represent the SS-CASPT2 calculations and the solid lines represent the MS-CASPT2 calculations. The points indicate the FCI calculations. Reprinted with permission from ref. [74]. Copyright 1998, Elsevier.

XMS-CASPT2 method certainly will resolve some of the flaws of the MS-CASPT2 theory in and close to conical intersections (see fig. 10.9). However, this will be at the expense of accuracy of predicting vertical excitation energies for well-separated states. Furthermore, in the XMS-CASPT2 method the eigenvalues of the model space are sometimes too sensitive to changes of the model space by including more or fewer states.

10.6 Summary and Outlook

In this chapter we have described the CASPT2, MS-CASPT2, and XMS-CASPT2 methods in some detail. We have also lightly touched on some related methods such as the CAS-MP2, MRMP2, and NEVPT2 approximations. This has been presented in a bottom-up order, starting with a thorough presentation of Rayleigh–Schrödinger perturbation theory. Here we also discussed the issue of the
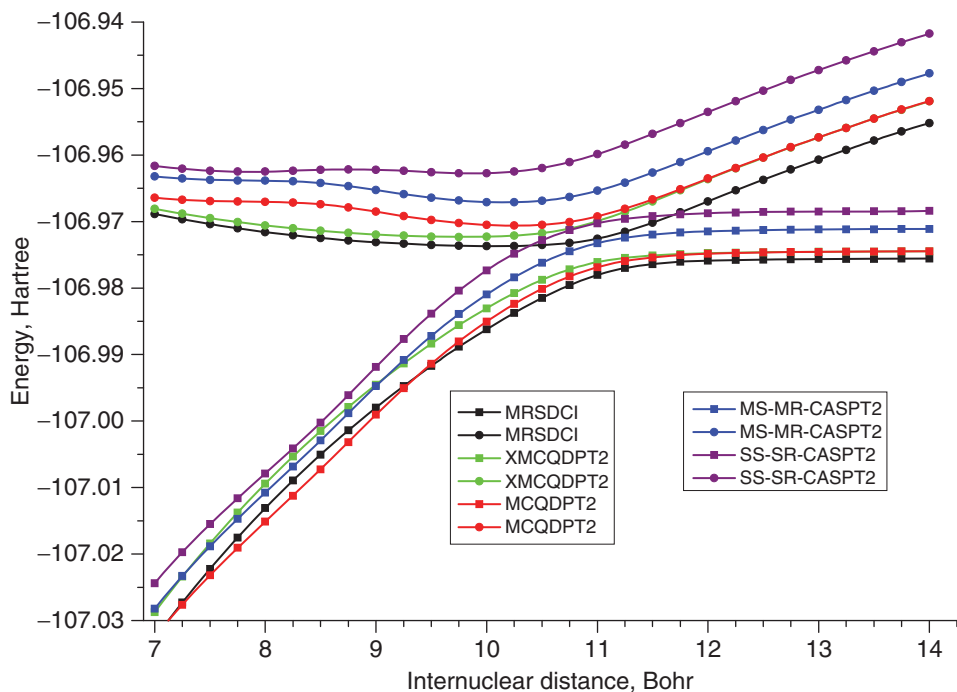
Figure 10.9 Potential energy curves (PECs) of the LiF molecule near the avoided crossing area computed using different MS-MRPT approaches. MRSDCI PECs are given as reference. PECs were intentionally left non-shifted to a common dissociation limit. Reprinted with permission from ref. [77]. Copyright 2011, American Institute of Physics.

convergence of the perturbation expansion and how divergence or near-divergence is manifested as the presence of intruder states. In the analysis of these we introduced the gap-shift technique. This tool later became a general technique to avoid divergence – real and imaginary shift techniques – or a way to compensate for systematic errors – the IPEA shift. This was followed up by a detailed presentation of the Møller–Plesset partitioning in association with the single determinant SCF wave function – the CASPT2 model is supposed to reduce to this as the active space is reduced to null. Here we also discussed numerical methods to compute the first-order correction to the wave function and subsequently the second-order correction to the energy, when analytical solutions do not exist. At this point the single-state multi-configurational perturbation schemes were introduced (CAS-MP2, CASPT2, MRMP2, and NEVPT2). The chapter discussed the benchmark results of these methods, which clearly exhibit explicit or implicit properties of the single-state approaches. These second-order perturbation approaches have their limitations and multi-state versions of perturbation theory have, sometimes, to be used to relieve these shortcomings. For that purpose we presented, in some detail, the basic steps of quasi-degenerate perturbation theory – this allowed us to pave the way for an understanding of the multi-state versions of perturbation theories mentioned above. Prepared by this introduction, we proceeded with a description of the multi-state version of CASPT2, which has some parallels with the corresponding MRMP2 and NEVPT2 versions. Finally, an extended version of MS-CASPT2 was introduced – based on extended quasi-degenerate perturbation theory – which will have an $\hat{H}_0$ operator whose eigenvalues are invariant to rotations of the

states in the model space. This, in particular, should resolve some problems of MS-CASPT2 with artificial surface crossings observed close to or at conical intersections.

The future of MRPT, its implementation and approximations, is closely connected with the partitioning of the Hamiltonian for the CASPT2 method. Here there is still room for significant improvement. While the NEVPT2 model does not suffer from the intruder-state problem to the degree that CASPT2 does, it derives this character at a larger computational cost. For the NEVPT2 approach to become tractable for larger active spaces, approximations will have to be developed to reduce this disadvantage compared to the CASPT2 method. Additionally, with the advent of the DMRG approach to enable CASSCF calculations with large active spaces, both the CASPT2 and the NEVPT2 methods have to be recast to overcome the issues that computing and storing the whole (or partially contracted) four-particle density matrix – at least if the notion of contracted reference functions is to survive. For recent developments in MRPT, we point the reader to the references mentioned at the end of section 10.3.6. Finally, the issue of how to select the reference functions needs to be readdressed. Is there maybe a golden path to be explored in which the reference function already – through some simple approach – carries a significant part of the dynamic electron correlation?

Acknowledgments

References

1 Schrödinger, E. (1926). Quantisierung als Eigenwertproblem. *Ann. Phys.* 385 (13): 437–490.

2 Strutt, B.R. and William, J. (1894). *The Theory of Sound*, 2e. London: MacMillan.

3 Helgaker, T., Jørgensen, P., and Olsen, J. (2000). *Molecular Electronic-Structure Theory*. Wiley.

4 Kato, T. (1995). *Perturbation Theory for Linear Operators*. Berlin Heidelberg: Springer.

5 Olsen, J., Jørgensen, P., Helgaker, T., and Christiansen, O. (2000). Divergence in Møller–Plesset theory: a simple explanation based on a two-state model. *J. Chem. Phys.* 112 (22): 9736–9748.

6 Roos, B.O. and Andersson, K. (1995). Multiconfigurational perturbation theory with level shift – the Cr_2 potential revisited. *Chem. Phys. Lett.* 245 (2–3): 215–223.

7 Forsberg, N. and Malmqvist, P.-Å. (1997). Multiconfiguration perturbation theory with imaginary level shift. *Chem. Phys. Lett.* 274 (1–3): 196–204.

8 Surján, P.R. and Szabados, Á. (1996). Damping of perturbation corrections in quasidegenerate situations. *J. Chem. Phys.* 104 (9): 3320–3324.

9 Epstein, P.S. (1926). The Stark effect from the point of view of Schroedinger's quantum theory. *Phys. Rev.* 28 (4): 695–710.

10 Nesbet, R.K. (1955). Configuration interaction in orbital theories. *Proc. Math. Phys. Eng. Sci.* 230 (1182): 312–321.

11 McDouall, J.J.W., Peasley, K., and Robb, M.A. (1988). A simple MC SCF perturbation theory: orthogonal valence bond Møller–Plesset 2 (OVB MP2). *Chem. Phys. Lett.* 148 (2–3): 183–189.

12 Andersson, K., Malmqvist, P.Å., Roos, B.O. et al. (1990). Second-order perturbation theory with a CASSCF reference function. *J. Phys. Chem.* 94 (14): 5483–5488.

13 Andersson, K., Malmqvist, P.-Å., and Roos, B.O. (1992). Second-order perturbation theory with a complete active space self-consistent field reference function. *J. Chem. Phys.* 96 (2): 1218–1226.

14 Hirao, K. (1992). Multireference Møller–Plesset method. *Chem. Phys. Lett.* 190 (3–4): 374–380.

15 Hirao, K. (1992). Multireference Møller–Plesset perturbation theory for high-spin open-shell systems. *Chem. Phys. Lett.* 196 (5): 397–403.

16 Hirao, K. (1992). Multireference Møller–Plesset perturbation treatment of potential energy curve of N_2. *Int. J. Quantum Chem.* 44 (S26): 517–526.

17 Hirao, K. (1993). State-specific multireference Møller–Plesset perturbation treatment for singlet and triplet excited states, ionized states and electron attached states of H_2O. *Chem. Phys. Lett.* 201 (1–4): 59–66.

18 Angeli, C., Cimiraglia, R., Evangelisti, S. et al. (2001). Introduction of *n*-electron valence states for multireference perturbation theory. *J. Chem. Phys.* 114 (23): 10252–10264.

19 Angeli, C., Cimiraglia, R., and Malrieu, J.-P. (2001). *n*-electron valence state perturbation theory: a fast implementation of the strongly contracted variant. *Chem. Phys. Lett.* 350 (3–4): 297–305.

20 Angeli, C., Cimiraglia, R., and Malrieu, J.-P. (2002). *n*-electron valence state perturbation theory: a spinless formulation and an efficient implementation of the strongly contracted and of the partially contracted variants. *J. Chem. Phys.* 117 (20): 9138–9153.

21 Lischka, H., Nachtigallová, D., Aquino, A.J.A. et al. (2018). Multireference approaches for excited states of molecules. *Chem. Rev.* 118 (15): 7293–7361.

22 Löwdin, P.-O. (1951). A note on the quantum-mechanical perturbation theory. *J. Chem. Phys.* 19 (11): 1396–1401.

23 Andersson, K. and Roos, B.O. (1995). Multiconfigurational second-order perturbation theory. In: *Modern Electronic Structure Theory*, 55–109. World Scientific Publishing Company.

24 Andersson, K. (1995). Different forms of the zeroth-order Hamiltonian in second-order perturbation theory with a complete active space self-consistent field reference function. *Theor. Chim. Acta* 91 (1–2): 31–46.

25 Malmqvist, P.Å., Pierloot, K., Shahi, A.R.M. et al. (2008). The restricted active space followed by second-order perturbation theory method: theory and application to the study of CuO_2 and Cu_2O_2 systems. *J. Chem. Phys.* 128 (20): 204109.

26 Ma, D., Manni, G.L., Olsen, J., and Gagliardi, L. (2016). Second-order perturbation theory for generalized active space self-consistent-field wave functions. *J. Chem. Theory Comput.* 12 (7): 3208–3213.

27 Andersson, K. and Roos, B.O. (1993). Multiconfigurational second-order perturbation theory: a test of geometries and binding energies. *Int. J. Quantum Chem.* 45 (6): 591–607.

28 Guner, V., Khuong, K.S., Leach, A.G. et al. (2003). A standard set of pericyclic reactions of hydrocarbons for the benchmarking of computational methods: the performance of ab initio, density functional, CASSCF, CASPT2, and CBS-QB3 methods for the prediction of activation barriers, reaction energetics, and transition state geometries. *Chem. A Eur. J.* 107 (51): 11445–11459.

29 Schreiber, M., Silva-Junior, M.R., Sauer, S.P.A., and Thiel, W. (2008). Benchmarks for electronically excited states: CASPT2, CC2, CCSD, and CC3. *J. Chem. Phys.* 128 (13): 134110.

30 Serrano-Andrés, L. (2011). Organic spectroscopy under Björn O. Roos. *Int. J. Quantum Chem.* 111 (13): 3284–3290.

31 Pierloot, K. (2011). Transition metals compounds: outstanding challenges for multiconfigurational methods. *Int. J. Quantum Chem.* 111 (13): 3291–3301.

32 Gagliardi, L. (2011). The study of actinide chemistry with multiconfigurational quantum chemical methods. *Int. J. Quantum Chem.* 111 (13): 3302–3306.

33 Budzák, i., Scalmani, G., and Jacquemin, D. (2017). Accurate excited-state geometries: a CASPT2 and coupled-cluster reference database for small molecules. *J. Chem. Theory Comput.* 13 (12): 6237–6252.

34 Ghigo, G., Roos, B.O., and Malmqvist, P.-Å. (2004). A modified definition of the zeroth-order Hamiltonian in multiconfigurational perturbation theory (CASPT2). *Chem. Phys. Lett.* 396 (1–3): 142–149.

35 Dyall, K.G. (1995). The choice of a zeroth-order Hamiltonian for second-order perturbation theory with a complete active space self-consistent-field reference function. *J. Chem. Phys.* 102 (12): 4909–4918.

36 Aquilante, F., Autschbach, J., Carlson, R.K. et al. (2015). Molcas8: new capabilities for multiconfigurational quantum chemical calculations across the periodic table. *J. Comput. Chem.* 37 (5): 506–541.

37 Kepenekian, M., Robert, V., and Le Guennic, B. (2009). What zeroth-order Hamiltonian for CASPT2 adiabatic energetics of $Fe(II)N_6$ architectures? *J. Chem. Phys.* 131 (11): 114702.

38 Zobel, J.P., Nogueira, J.J., and González, L. (2017). The IPEA dilemma in CASPT2. *Chem. Sci.* 8 (2): 1482–1499.

39 Vela, S., Fumanal, M., Ribas-Ariño, J., and Robert, V. (2015). On the zeroth-order Hamiltonian for CASPT2 calculations of spin crossover compounds. *J. Comput. Chem.* 37 (10): 947–953.

40 Daku Lawson, L.M., Aquilante, F., Robinson, T.W., and Hauser, A. (2012). Accurate spin-state energetics of transition metal complexes. 1. CCSD(T), CASPT2, and DFT study of $[M(NCH)_6]^{2+}$ (M = Fe, Co). *J. Chem. Theory Comput.* 8 (11): 4216–4231.

41 Angeli, C. and Cimiraglia, R. (2014). Some useful odds and ends from the *n*-electron valence state perturbation theory. *Chem. A Eur. J.* 118 (33): 6435–6439.

42 Angeli, C., Borini, S., and Cimiraglia, R. (2004). An application of second-order *n*-electron valence state perturbation theory to the calculation of excited states. *Theor. Chem. Acc.* 111 (2–6): 352–357.

43 Schapiro, I., Sivalingam, K., and Neese, F. (2013). Assessment of *n*-electron valence state perturbation theory for vertical excitation energies. *J. Chem. Theory Comput.* 9 (8): 3567–3580.

44 Aquilante, F., Malmqvist, P.-Å., Pedersen, T.B. et al. (2008). Cholesky decomposition-based multiconfiguration second-order perturbation theory (CD-CASPT2): application to the spin-state energetics of Co^{III}(diiminato)(NPh). *J. Chem. Theory Comput.* 4 (5): 694–702.

45 Boström, J., Delcey, M.G., Aquilante, F. et al. (2010). Calibration of Cholesky auxiliary basis sets for multiconfigurational perturbation theory calculations of excitation energies. *J. Chem. Theory Comput.* 6 (3): 747–754.

46 Aquilante, F., Todorova, T.K., Gagliardi, L. et al. (2009). Systematic truncation of the virtual space in multiconfigurational perturbation theory. *J. Chem. Phys.* 131 (3): 034113.

47 Segarra-Martí, J., Garavelli, M., and Aquilante, F. (2015). Multiconfigurational second-order perturbation theory with frozen natural orbitals extended to the treatment of photochemical problems. *J. Chem. Theory Comput.* 11 (8): 3772–3784.

48 Pulay, P. (1983). Localizability of dynamic electron correlation. *Chem. Phys. Lett.* 100 (2): 151–154.

49 Fink, R. and Staemmler, V. (1993). A multi-configuration reference CEPA method based on pair natural orbitals. *Theor. Chim. Acta* 87 (1–2): 129–145.

50 Menezes, F., Kats, D., and Werner, H.-J. (2016). Local complete active space second-order perturbation theory using pair natural orbitals (PNO-CASPT2). *J. Chem. Phys.* 145 (12): 124115.

51 Guo, Y., Sivalingam, K., Valeev, E.F., and Neese, F. (2016). SparseMaps–A systematic infrastructure for reduced-scaling electronic structure methods. III. Linear-scaling multireference domain-based pair natural orbital n-electron valence perturbation theory. *J. Chem. Phys.* 144 (9): 094111.

52 Kurashige, Y. and Yanai, T. (2011). Second-order perturbation theory with a density matrix renormalization group self-consistent field reference function: theory and application to the study of chromium dimer. *J. Chem. Phys.* 135 (9): 094104.

53 Liu, F., Kurashige, Y., Yanai, T., and Morokuma, K. (2013). Multireference ab initio density matrix renormalization group (DMRG)-CASSCF and DMRG-CASPT2 study on the photochromic ring opening of spiropyran. *J. Chem. Theory Comput.* 9 (10): 4462–4469.

54 Kurashige, Y. (2013). Multireference electron correlation methods with density matrix renormalisation group reference functions. *Mol. Phys.* 112 (11): 1485–1494.

55 Wouters, S., Van Speybroeck, V., and Van Neck, D. (2016). DMRG-CASPT2 study of the longitudinal static second hyperpolarizability of all-trans polyenes. *J. Chem. Phys.* 145 (5): 054120.

56 Nakatani, N. and Guo, S. (2017). Density matrix renormalization group (DMRG) method as a common tool for large active-space CASSCF/CASPT2 calculations. *J. Chem. Phys.* 146 (9): 094102.

57 Guo, S., Watson, M.A., Hu, W. et al. (2016). *n*-electron valence state perturbation theory based on a density matrix renormalization group reference function, with applications to the chromium dimer and a trimer model of poly(*p*-phenylenevinylene). *J. Chem. Theory Comput.* 12 (4): 1583–1591.

58 Knecht, S., Hedegård, E.D., Keller, S. et al. (2016). New approaches for *ab initio* calculations of molecules with strong electron correlation. *CHIMIA Int. J. Chem.* 70 (4): 244–251.

59 Freitag, L., Knecht, S., Angeli, C., and Reiher, M. (2017). Multireference perturbation theory with Cholesky decomposition for the density matrix renormalization group. *J. Chem. Theory Comput.* 13 (2): 451–459.

60 Zgid, D., Ghosh, D., Neuscamman, E., and Chan, G.K.-L. (2009). A study of cumulant approximations to *n*-electron valence multireference perturbation theory. *J. Chem. Phys.* 130 (19): 194107.

61 Sharma, S. and Chan, G.K.-L. (2014). Communication: a flexible multi-reference perturbation theory by minimizing the Hylleraas functional with matrix product states. *J. Chem. Phys.* 141 (11): 111101.

62 Sharma, S., Jeanmairet, G., and Alavi, A. (2016). Quasi-degenerate perturbation theory using matrix product states. *J. Chem. Phys.* 144 (3): 034103.

63 Sokolov, A.Y. and Chan, G.K.-L. (2016). A time-dependent formulation of multi-reference perturbation theory. *J. Chem. Phys.* 144 (6): 064102.

64 Roemelt, M., Guo, S., and Chan, G.K.-L. (2016). A projected approximation to strongly contracted *n*-electron valence perturbation theory for DMRG wave functions. *J. Chem. Phys.* 144 (20): 204113.

65 Phung, Q.M., Wouters, S., and Pierloot, K. (2016). Cumulant approximated second-order perturbation theory based on the density matrix renormalization group for transition metal complexes: a benchmark study. *J. Chem. Theory Comput.* 12 (9): 4352–4361.

66 Sokolov, A.Y., Guo, S., Ronca, E., and Chan, G.K.-L. (2017). Time-dependent *n*-electron valence perturbation theory with matrix product state reference wave functions for large active spaces and basis sets: applications to the chromium dimer and all-trans polyenes. *J. Chem. Phys.* 146 (24): 244102.

67 Lindh, R. and Roos, B.O. (1989). A theoretical study of the diffuseness of the $v(^1b_{1u})$ state of planar ethylene. *Int. J. Quantum Chem.* 35 (6): 813–825.

68 Lindgren, I. (1974). The Rayleigh–Schrodinger perturbation and the linked-diagram theorem for a multi-configurational model space. *J. Phys. B: At. Mol. Phys.* 7 (18): 2441–2470.

69 Lindgren, I. (1978). A coupled-cluster approach to the many-body perturbation theory for open-shell systems. *Int. J. Quantum Chem.* 14 (S12): 33–58.

70 Pittner, J. (2003). Continuous transition between Brillouin–Wigner and Rayleigh–Schrödinger perturbation theory, generalized Bloch equation, and Hilbert space multireference coupled cluster. *J. Chem. Phys.* 118 (24): 10876–10889.

71 Nakano, H. (1993). Quasidegenerate perturbation theory with multiconfigurational self-consistent-field reference functions. *J. Chem. Phys.* 99 (10): 7983–7992.

72 Nakano, H., Uchiyama, R., and Hirao, K. (2002). Quasi-degenerate perturbation theory with *general* multiconfiguration self-consistent field reference functions. *J. Comput. Chem.* 23 (12): 1166–1175.

73 Angeli, C., Borini, S., Cestari, M., and Cimiraglia, R. (2004). A quasidegenerate formulation of the second order *n*-electron valence state perturbation theory approach. *J. Chem. Phys.* 121 (9): 4043–4049.

74 Finley, J., Malmqvist, P.-Å., and Björn, O. (1998). Roos, and Luis Serrano-Andrés. The multi-state CASPT2 method. *Chem. Phys. Lett.* 288 (2–4): 299–306.

75 Zaitsevskii, A. and Malrieu, J.-P. (1995). Multi-partitioning quasidegenerate perturbation theory. A new approach to multireference Møller–Plesset perturbation theory. *Chem. Phys. Lett.* 233 (5–6): 597–604.

76 Serrano-Andrés, L., Merchán, M., and Lindh, R. (2005). Computation of conical intersections by using perturbation techniques. *J. Chem. Phys.* 122 (10): 104107.

77 Granovsky, A.A. (2011). Extended multi-configuration quasi-degenerate perturbation theory: the new approach to multi-state multi-reference perturbation theory. *J. Chem. Phys.* 134 (21): 214113.

78 Shiozaki, T., Győrffy, W., Celani, P., and Werner, H.-J. (2011). Communication: extended multi-state complete active space second-order perturbation theory: energy and nuclear gradients. *J. Chem. Phys.* 135 (8): 081106.

79 Andersson, K. (1992). *Multiconfigurational perturbation theory*. PhD thesis. Lund University, Lund, Sweden. https://urn.kb.se/resolve?urn=urn:nbn:se:kau:diva-21628.

Appendix

The following is an almost verbatim copy of appendix A in K. Andersson's thesis [79], reproduced here with her kind permission. Some adaptation has been applied in order to maintain consistency with the notation used in the rest of this chapter. Where different this will be pointed out.

This appendix is a collection of all matrix elements and vector components used in the first-order equation (see for example eqs. 10.24 and 10.108). We note that the orbital basis used in the appendix is the one in which the diagonal blocks of the Fock matrix (inactive, active and virtual), have been diagonalized. To emphasize this the Fock matrix elements are denoted f' so that there is a distinction with respect to a Fock matrix in a general orbital representation, f.

The following additional definitions and notations have been used (note that the two-particle density-matrix elements, Γ_{pqrs}, carry a factor $\frac{1}{2}$ in the work of Andersson, while here we stand by the definition in Chapter 1, section 1.6):

$$\hat{e}_{ut} = 2\delta_{ut} - \hat{E}_{ut} \ ,$$

$$d_{ut} = \langle \Psi^{(0)} | \hat{e}_{ut} | \Psi^{(0)} \rangle \ , \text{ and}$$

$$\gamma_{utxv} = \langle \Psi^{(0)} | \hat{e}_{ut}\hat{e}_{xv} - \delta_{uv}\hat{e}_{xt} | \Psi^{(0)} \rangle \ .$$

The index notation specified in Chapter 1, section 1.5, will be used strictly in this appendix except in the following formula:

$$f_{pq}^{\mathrm{I}} = h_{pq} + \sum_k [2(pq|kk) - (pk|kq)] \;,$$

where k runs over both the *frozen* and inactive subspaces, i.e., over all doubly-occupied orbitals.

In the evaluation of the matrix elements and vector components the following commutation relations have been useful:

$$[\hat{E}_{pq}, \hat{E}_{rs}] = \delta_{rq}\hat{E}_{ps} - \delta_{ps}\hat{E}_{rq} \;,$$
$$[\hat{e}_{pq}, \hat{e}_{rs}] = \delta_{ps}\hat{e}_{rq} - \delta_{rq}\hat{e}_{ps} \;.$$

Matrix elements and vector components, with respect to combinations stemming from interactions between states of the same class within the first-order interacting space, will be presented in the following section. Corresponding matrix elements between states of different classes of the first-order interacting space will be presented subsequently. We note also that we will use the following abbreviation: $|\Psi^{(0)}\rangle = |0\rangle$ and similarly for the bra.

The Diagonal Blocks

The following elements are given in this section:

$$\langle\sigma|\hat{H}_0 - E^{(0)}|\tau\rangle \;,\;\; \langle\sigma|V|0\rangle \;,\;\; |\sigma\rangle, |\tau\rangle \in V_X \;,$$

where $\hat{H}_0 = \sum_{pq} f'_{pq}\hat{E}_{pq}$ and $X \in \{A, \dots, H\}$. We note that, $f'_{pq} = \delta_{pq}\epsilon_p$, if both indices, p and q, belong to the same orbital class, i.e., inactive, active, and virtual. Non-zero elements are found in the four inactive–active and active–virtual blocks.

Internal excitations: $X \in \{A, B\}$

A	**B** $\qquad\qquad i \geq j, k \geq l, t \geq u, x \geq y$
$|ituv\rangle = \hat{E}_{ti}\hat{E}_{uv}|0\rangle$	$|ijtu\rangle^{\pm} = [\hat{E}_{tj}\hat{E}_{ui} \pm \hat{E}_{ti}\hat{E}_{uj}]|0\rangle$
$\langle ituv|\hat{H}_0 - E^{(0)}|kxyz\rangle$	$^{\pm}\langle ijtu|\hat{H}_0 - E^{(0)}|klxy\rangle^{\pm}$
$\quad = \delta_{ik}[\alpha^i S^A_{tuv,xyz} + B_{tuv,xyz}]$	$\quad = \delta_{ik}\delta_{jl}[1 \pm \delta_{ij}][\alpha^{ij}S^{\pm}_{tu,xy} + B^{\pm}_{tu,xy}]$
$\langle ituv|V|0\rangle = V^i_{tuv}$	$^{\pm}\langle ijtu|V|0\rangle = \frac{1}{2}[1 \pm \delta_{ij}](\boldsymbol{S}^{\pm}\boldsymbol{W}^{ij,\pm})_{tu}$
$\alpha^i = -\epsilon_i - \sum_w \epsilon_w D_{ww}$	$\alpha^{ij} = -\epsilon_i - \epsilon_j + \sum_w \epsilon_w d_{ww}$
$S^A_{tuv,xyz} = \langle 0|\hat{E}_{vu}\hat{e}_{xt}\hat{E}_{yz}|0\rangle$	$S^{\pm}_{tu,xy} = S_{tu,xy} \pm S_{tu,yx}, S_{tu,xy} = 2\gamma_{xtyu}$
$B_{tuv,xyz} = \epsilon_x\langle 0|\hat{E}_{vu}\hat{e}_{xt}\hat{E}_{yz}|0\rangle$	$B^{\pm}_{tu,xy} = B_{tu,xy} \pm B_{tu,yx}$
$\quad + \sum_w \epsilon_w\langle 0|\hat{E}_{vu}\hat{e}_{xt}\hat{E}_{ww}\hat{E}_{yz}|0\rangle$	$B_{tu,xy} = [\epsilon_t + \epsilon_y][2\gamma_{xtyu} + \delta_{ty}d_{xu}]$
$V^i_{tuv} = 2f^{\mathrm{I}}_{ti}D_{vu} - f^{\mathrm{I}}_{ui}D_{vt} - \sum_w \Gamma_{vuwt}f^{\mathrm{I}}_{wi}$	$\quad + 2\sum_w \epsilon_w[\delta_{ty}\gamma_{xuww} - \langle 0|\hat{e}_{yu}\hat{e}_{ww}\hat{e}_{xt}|0\rangle]$
$\quad + \sum_{wxy}\langle 0|\hat{E}_{vu}\hat{e}_{wt}\hat{E}_{xy}|0\rangle(wi|xy)$	$W^{ij,\pm}_{tu} = \frac{1}{1+\delta_{ij}}\left[1 - \frac{1}{2}\delta_{tu}\right][(ti|uj) \pm (ui|tj)]$

Semi-internal excitations: $X \in \{C, D, E\}$

C

$|atuv\rangle = \hat{E}_{at}\hat{E}_{uv}|0\rangle$

$\langle atuv|\hat{H}_0 - E^{(0)}|cxyz\rangle$

$\quad = \delta_{ac}[\alpha^a S^C_{tuv,xyz} + B_{tuv,xyz}]$

$\langle atuv|V|0\rangle = V^a_{tuv}$

$\alpha^a = \epsilon_a - \sum_w \epsilon_w D_{ww}$

$S^C_{tuv,xyz} = \langle 0|\hat{E}_{vu}\hat{E}_{tx}\hat{E}_{yz}|0\rangle$

$B_{tuv,xyz} = -\epsilon_x \langle 0|\hat{E}_{vu}\hat{E}_{tx}\hat{E}_{yz}|0\rangle$

$\quad + \sum_w \epsilon_w \langle 0|\hat{E}_{vu}\hat{E}_{tx}\hat{E}_{ww}\hat{E}_{yz}|0\rangle$

$V^a_{tuv} = \sum_{wxy}\langle 0|\hat{E}_{vu}\hat{E}_{tw}\hat{E}_{xy}|0\rangle(aw|xy)$

$\quad + \sum_w [\Gamma_{vutw} + \delta_{ut}D_{vw}]\left[f^I_{aw} - \sum_y(ay|yw)\right]$

E $\qquad\qquad i \geq j, k \geq l$

$|ijat\rangle^{\pm} = [\hat{E}_{tj}\hat{E}_{ai} \pm \hat{E}_{ti}\hat{E}_{aj}]|0\rangle$

$^{\pm}\langle ijat|\hat{H}_0 - E^{(0)}|klcx\rangle^{\pm}$

$\quad = 2[2 \mp 1]\delta_{ac}\delta_{ik}\delta_{jl}[1 + \delta_{ij}][\alpha^{ija}S_{tx} + B_{tx}]$

$^{\pm}\langle ijat|V|0\rangle = [2 \mp 1][1 \pm \delta_{ij}](\boldsymbol{SW}^{ija,\pm})_t$

$\alpha^{ija} = \epsilon_a - \epsilon_i - \epsilon_j + \sum_w \epsilon_w d_{ww}$

$S_{tx} = d_{xt}$

$B_{tx} = -\sum_w \epsilon_w \gamma_{wwxt}$

$W^{ija,\pm}_t = \frac{1}{1+\delta_{ij}}[(ai|tj) \pm (aj|ti)]$

D

$|iatu\rangle^1 = \hat{E}_{ai}\hat{E}_{tu}|0\rangle, \ |iatu\rangle^2 = \hat{E}_{ti}\hat{E}_{au}|0\rangle$

$^Q\langle iatu|\hat{H}_0 - E^{(0)}|kcxy\rangle^P$

$\quad = \delta_{ac}\delta_{ik}[\alpha^{ia}S^{Q,P}_{tu,xy} + B^{Q,P}_{tu,xy}]$

$^Q\langle iatu|V|0\rangle = V^{ia,Q}_{tu} \qquad\qquad P, Q \in \{1, 2\}$

$\alpha^{ia} = \epsilon_a - \epsilon_i - \sum_w \epsilon_w D_{ww}$

$S = \begin{bmatrix} S^{1,1} & S^{1,2} \\ S^{2,1} & S^{2,2} \end{bmatrix}, B = \begin{bmatrix} B^{1,1} & B^{1,2} \\ B^{2,1} & B^{2,2} \end{bmatrix}$

$S^{1,1}_{tu,xy} = -2S^{2,1}_{tu,xy} = 2[\Gamma_{utxy} + \delta_{tx}D_{uy}]$

$S^{2,2}_{tu,xy} = -[\Gamma_{xtuy} - 2\delta_{tx}D_{uy}]$

$B^{1,1}_{tu,xy} = -2B^{2,1}_{tu,xy}$

$\quad = 2\sum_w \epsilon_w \langle 0|\hat{E}_{ut}\hat{E}_{ww}\hat{E}_{xy}|0\rangle$

$B^{2,2}_{tu,xy} = [\epsilon_u - \epsilon_x]\Gamma_{xtuy} + \epsilon_u \delta_{tu}D_{xy}$

$\quad + 2\epsilon_x \delta_{tx}D_{uy} + \sum_w \epsilon_w [2\delta_{tx}\Gamma_{wwuy}$

$\quad + \delta_{tu}\Gamma_{wwxy} - \langle 0|\hat{E}_{xt}\hat{E}_{ww}\hat{E}_{uy}|0\rangle]$

$\boldsymbol{V}^{ia} = f^I_{ai}\boldsymbol{U} + \boldsymbol{SW}^{ia}$

$\boldsymbol{V}^{ia} = \begin{bmatrix} V^{ia,1} \\ V^{ia,2} \end{bmatrix}, \boldsymbol{U} = \begin{bmatrix} U^1 \\ U^2 \end{bmatrix}, \boldsymbol{W}^{ia} = \begin{bmatrix} W^{ia,1} \\ W^{ia,2} \end{bmatrix}$

$U^1_{tu} = 2D_{ut}, U^2_{tu} = -D_{ut}$

$W^{ia,1}_{tu} = (ai|tu), W^{ia,2}_{tu} = (ti|au)$

External excitations: $X \in \{F, G, H\}$

F $\qquad a \geq b, c \geq d, t \geq u, x \geq y$

$|abtu\rangle^{\pm} = [\hat{E}_{bt}\hat{E}_{au} \pm \hat{E}_{at}\hat{E}_{bu}]|0\rangle$

$^{\pm}\langle abtu|\hat{H}_0 - E^{(0)}|cdxy\rangle^{\pm}$

$\quad = \delta_{ac}\delta_{bd}[1 \pm \delta_{ab}][\alpha^{ab}S^{\pm}_{tu,xy} + B^{\pm}_{tu,xy}]$

$^{\pm}\langle abtu|V|0\rangle = \frac{1}{2}[1 \pm \delta_{ab}](\boldsymbol{S}^{\pm}\boldsymbol{W}^{ab,\pm})_{tu}$

$\alpha^{ab} = \epsilon_a + \epsilon_b - \sum_w \epsilon_w D_{ww}$

$S^{\pm}_{tu,xy} = S_{tu,xy} \pm S_{tu,yx}, S_{tu,xy} = 2\Gamma_{txuy}$

$B^{\pm}_{tu,xy} = B_{tu,xy} \pm B_{tu,yx}$

$B_{tu,xy} = -[\epsilon_t + \epsilon_y][2\Gamma_{txuy} + \delta_{ty}D_{ux}]$

$\quad -2\sum_w \epsilon_w [\delta_{ty}\Gamma_{uxww} - \langle 0|\hat{E}_{uy}\hat{E}_{ww}\hat{E}_{tx}|0\rangle]$

$W^{ab,\pm}_{tu} = \frac{1}{1+\delta_{ab}}\left[1 - \frac{1}{2}\delta_{tu}\right][(au|bt) \pm (at|bu)]$

G $\qquad a \geq b, c \geq d$

$|iabt\rangle^{\pm} = [\hat{E}_{bi}\hat{E}_{at} \pm \hat{E}_{ai}\hat{E}_{bt}]|0\rangle$

$^{\pm}\langle iabt|\hat{H}_0 - E^{(0)}|kcdx\rangle^{\pm}$

$\quad = 2[2 \mp 1]\delta_{ik}\delta_{ac}\delta_{bd}[1 \pm \delta_{ab}][\alpha^{iab}S_{tx} + B_{tx}]$

$^{\pm}\langle iabt|V|0\rangle = [2 \mp 1][1 \pm \delta_{ab}](\boldsymbol{SW}^{iab,\pm})_t$

$\alpha^{iab} = \epsilon_a + \epsilon_b - \epsilon_i - \sum_w \epsilon_w D_{ww}$

$S_{tx} = D_{tx}$

$B_{tx} = \sum_w \epsilon_w \Gamma_{wwtx}$

$W^{iab,\pm}_t = \frac{1}{1+\delta_{ab}}[(at|bi) \pm (bt|ai)]$

H $\qquad\qquad i \geq j, k \geq l, a \geq b, c \geq d$

$|ijab\rangle^{\pm} = [\hat{E}_{bj}\hat{E}_{ai} \pm \hat{E}_{aj}\hat{E}_{bi}]|0\rangle$

$^{\pm}\langle ijab|\hat{H}_0 - E^{(0)}|klcd\rangle^{\pm} = 4[2 \mp 1]\delta_{ik}\delta_{jl}\delta_{ac}\delta_{bd}[1 \pm \delta_{ij}][1 \pm \delta_{ab}][\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j]$

$^{\pm}\langle ijab|V|0\rangle = 2[2 \mp 1][(ai|bj) \pm (aj|bi)]$

The Nondiagonal Blocks

The following elements are given in this section:

$$\langle \sigma | \hat{H}_0 | \tau \rangle \ , \ |\sigma\rangle \in V_X \ , \ |\tau\rangle \in V_Y \ , \ X \neq Y \in \{A, \dots, H\} \ .$$

AB $\hspace{8cm} k \geq l, x \geq y$

$$\langle ituv | \hat{H}_0 | klxy \rangle^{\pm} = \sum_w [\delta_{il} f'_{wk} \pm \delta_{ik} f'_{wl}] \{ -[S^A_{tuv,xyw} \pm S^A_{tuv,yxw}]$$

$$+ [2 \mp 1][\delta_{yw}(2\delta_{xt} D_{vu} - \delta_{ux} D_{vt} - \Gamma_{vuxt}) \pm \delta_{xw}(2\delta_{yt} D_{vu} - \delta_{uy} D_{vt} - \Gamma_{vuyt})]\}$$

CF $\hspace{8cm} c \geq d, x \geq y$

$$\langle atuv | \hat{H}_0 | cdxy \rangle^{\pm} = \sum_w [\delta_{ad} f'_{wc} \pm \delta_{ac} f'_{wd}]$$

$$\times \{ [S^C_{tuv,xwy} - \delta_{ux}(\Gamma_{vuty} + \delta_{ut} D_{vy})] \pm [S^C_{tuv,ywx} - \delta_{uy}(\Gamma_{vutx} - \delta_{ut} D_{vx})]\}$$

AD $\hspace{5cm}$ **CD**

$$\langle ituv | \hat{H}_0 | kcxy \rangle^1 = \delta_{ik} \sum_w f'_{wc} S^A_{tuv,wxy}$$
$$\langle ituv | \hat{H}_0 | kcxy \rangle^2 = \delta_{ik} \sum_w f'_{wc} S^A_{tuv,xwy}$$

$$\langle atuv | \hat{H}_0 | kcxy \rangle^1 = -\delta_{ac} \sum_w f'_{wk} S^C_{tuv,wxy}$$
$$\langle atuv | \hat{H}_0 | kcxy \rangle^2 = -\delta_{ac} \sum_w f'_{wk} [S^C_{tuv,yxw}$$
$$- 2\delta_{ux}(\Gamma_{vuty} + \delta_{ut} D_{vy}) - \delta_{xy}(\Gamma_{vutw} + \delta_{ut} D_{vw})]$$

BE $\hspace{7cm} i \geq j, k \geq l, t \geq u$

$$^{\pm}\langle ijtu | \hat{H}_0 | klcx \rangle^{\pm} = 2\delta_{ik}\delta_{jl}[1 \pm \delta_{ij}] \sum_u f'_{wc} [\gamma_{xtwu} \pm \gamma_{xuwt}]$$

DE $\hspace{8cm} k \geq l$

$$^1\langle iatu | \hat{H}_0 | klcx \rangle^{\pm} = -[2 \mp 1]\delta_{ac} \sum_w [\delta_{il} f'_{wl} \pm \delta_{il} f'_{wk}][\Gamma_{utxw} + \delta_{tx} D_{uw} - 2\delta_{xw} D_{ut}]$$

$$^2\langle iatu | \hat{H}_0 | klcx \rangle^{\pm} = \delta_{ac} \sum_w [\delta_{il} f'_{wk} \pm \delta_{ik} f'_{wl}]$$

$$\times \{ [\Gamma_{uwxt} + \delta_{xw} D_{ut} - 2\delta_{tx} D_{uw}] \pm [\Gamma_{utxw} + \delta_{tx} D_{uw} - 2\delta_{xw} D_{ut}]\}$$

FG $\hspace{7cm} a \geq b, c \geq d, t \geq u$

$$^{\pm}\langle abtu | \hat{H}_0 | kcdx \rangle^{\pm} = -2\delta_{ac}\delta_{bd}[1 \pm \delta_{ab}] \sum_u f'_{wk} [\Gamma_{uxtw} \pm \Gamma_{uwtx}]$$

DG $\hspace{8cm} c \geq d$

$$^1\langle iatu | \hat{H}_0 | kcdx \rangle^{\pm} = [2 \mp 1]\delta_{ik} \sum_w [\delta_{ad} f'_{wc} \pm \delta_{ac} f'_{wd}][\Gamma_{utux} + \delta_{tw} D_{ux}]$$

$$^2\langle iatu | \hat{H}_0 | kcdx \rangle^{\pm} = \delta_{ik} \sum_w [\delta_{ad} f'_{wc} \pm \delta_{ac} f'_{wd}] \{ [-\Gamma_{utux} - \delta_{tw} D_{ux}] \pm [-\Gamma_{uxwt} + 2\delta_{tw} D_{ux}]\}$$

EH $\hspace{7cm} i \geq j, k \geq l, c \geq d$

$$^{\pm}\langle ijat | \hat{H}_0 | klcd \rangle^{\pm} = 2[2 \mp 1]\delta_{ik}\delta_{jl}[1 \pm \delta_{ij}] \sum_w [\delta_{ac} f'_{wd} \pm \delta_{ad} f'_{wc}] d_{wt}$$

GH $\hspace{7cm} a \geq b, k \geq l, c \geq d$

$$^{\pm}\langle iabt | \hat{H}_0 | klcd \rangle^{\pm} = -2[2 \mp 1]\delta_{ac}\delta_{bd}[1 \pm \delta_{ab}] \sum_w [\delta_{il} f'_{wk} \pm \delta_{ik} f'_{wl}] D_{tw}$$

The remaining upper-triangular blocks are equal to zero and the lower-triangular blocks are given by symmetry, since $\hat{H}_0$ is a Hermitian operator.