8 Orbital relaxation

Remark 8.1. ***Orbital relaxation.*** According to the Thouless theorem (appendix A), the effect of the singles CC operator is to transform the orbitals of the reference determinant into a new set $\{\widetilde{\psi}_i\}$ by "mixing in" some of the virtual orbitals.

$$\Psi_{\mathrm{CC}} = \exp(T_2 + T_3 + \cdots)\widetilde{\Phi} \qquad \widetilde{\Phi} \equiv \exp(T_1)\Phi = \tfrac{1}{\sqrt{n!}}\det(\widetilde{\psi}_1 \cdots \widetilde{\psi}_n) \qquad \widetilde{\psi}_i = \psi_i + \sum_a \psi_a t_a^i \tag{8.1}$$

This can be thought of as "relaxing" the orbitals in the presence of electron correlation. The size of this *orbital relaxation effect* can be monitored as the root mean square difference from the reference orbitals, which is known as the $\mathcal{T}_1$ *diagnostic*.

$$\mathcal{T}_1 \equiv \sqrt{\frac{1}{n}\sum_{i=1}^{n} \|\widetilde{\psi}_i - \psi_i\|^2} = \frac{\|\mathbf{t}_1\|}{\sqrt{n}} \tag{8.2}$$

Significant orbital relaxation generally indicates that the reference determinant forms a poor approximation to the wavefunction, which can lead to large errors for low-order truncated methods like CCSD or CCSD(T). In closed-shell systems, significant orbital relaxation is usually associated with an inherent *multireference character*, which means that no single determinant dominates the wavefunction with any choice of orbitals. Empirically, $\mathcal{T}_1 \geq 0.02$ is considered large in this context. In open-shell systems, mean-field methods like Hartree-Fock theory are often deficient even for non-multireference systems. In this case, orbital relaxation effects can generally be cured by choosing a new determinant which is optimized in the presence of dynamical[1] electron correlation.

Remark 8.2. ***Brueckner and orbital-optimized methods.*** The two most common ways of defining an ideal reference determinant for the correlated wavefunction are the *best overlap criterion* and the *best energy criterion*.[2]

$$\Phi_{\mathrm{B}} = \arg\max_{\Phi} \langle\Phi|\Psi\rangle \qquad \Phi_{\mathrm{O}} = \arg\min_{\Phi} \langle\Psi|H|\Psi\rangle, \quad \Psi = \Omega\,\Phi \tag{8.3}$$

The *best overlap* or *Brueckner determinant*, Φ_{B}, has maximum overlap with the wavefunction. The *best energy* or *orbital-optimized determinant*, Φ_{O}, yields the lowest energy for a given Ansatz. The spin-orbital sets from which these determinants are constructed will be denoted $\{\psi_p\}_{\mathrm{B}}$ and $\{\psi_p\}_{\mathrm{O}}$.

Remark 8.3. Appendix B shows that non-redundant transformations of the orbitals in Φ can be parametrized as

$$\Phi(\mathbf{x}) = \exp(X - X^\dagger)\Phi \tag{8.4}$$

where X has the form of a T_1 operator. This parametrization can be substituted into equation 8.3 to derive the explicit conditions satisfied by Brueckner orbitals and energy-optimized orbitals.[3]

$$\{\psi_p\}_{\mathrm{B}}: \ \frac{\partial}{\partial x_a^{i*}}\langle\Phi(\mathbf{x})|\Psi\rangle\bigg|_{\mathbf{x}=\mathbf{0}} = \langle\Phi_i^a|\Psi\rangle \stackrel{!}{=} 0 \qquad \{\psi_p\}_{\mathrm{O}}: \ \frac{\partial}{\partial x_a^{i*}}\langle\Psi(\boldsymbol{x})|H|\Psi(\boldsymbol{x})\rangle\bigg|_{\mathbf{x}=\mathbf{0}} = \langle\Psi|\,[a_a^i, H]\,|\Psi\rangle \stackrel{!}{=} 0$$

[1] As opposed to mean-field.

[2] See `https://en.wikipedia.org/wiki/Arg_max` for the notation used here.

[3] $|\Psi(\boldsymbol{x})\rangle = \exp(X - X^\dagger)|\Psi\rangle$ applies the orbital rotation to all of the determinants in the wavefunction expansion.

A The Thouless theorem

Notation A.1. Let $\boldsymbol{\psi}$ be a row vector of orthonormal spin-orbitals, $(\boldsymbol{\psi})_p = \psi_p$, which can be split into occupied and virtual blocks as $\boldsymbol{\psi} = [\boldsymbol{\psi}_\mathrm{o}\ \boldsymbol{\psi}_\mathrm{v}]$. Other sets of spin-orbitals are related to this one by a transformation $|\boldsymbol{\psi}'\rangle = |\boldsymbol{\psi}\rangle\,\mathbf{U}$ which is unitary if the primed orbitals are also orthonormal. Let Φ' be the *transformed reference determinant*, constructed from the first n orbitals in the transformed space. Then the occupied and virtual orbitals of the transformed space are given by

$$|\boldsymbol{\psi}'_\mathrm{o}\rangle = |\boldsymbol{\psi}_\mathrm{o}\rangle\,\mathbf{U}_\mathrm{oo} + |\boldsymbol{\psi}_\mathrm{v}\rangle\,\mathbf{U}_\mathrm{vo} \qquad\qquad |\boldsymbol{\psi}'_\mathrm{v}\rangle = |\boldsymbol{\psi}_\mathrm{o}\rangle\,\mathbf{U}_\mathrm{ov} + |\boldsymbol{\psi}_\mathrm{v}\rangle\,\mathbf{U}_\mathrm{vv} \tag{A.1}$$

in terms of the occupied and virtual blocks of the transformation. This kind of unitary transformation of the spin-orbital basis is sometimes referred to as an *orbital rotation*.

Theorem A.1. ***The Thouless theorem.***

1. *The function $e^{T_1}\Phi$ is an intermediately normalized determinant $\frac{1}{\sqrt{n!}}\det(\widetilde{\psi}_1 \cdots \widetilde{\psi}_n)$ with orbitals $\widetilde{\psi}_i = \psi_i + \sum_a \psi_a t^i_a$.*

   Proof: Intermediate normalization follows from $\langle\Phi|e^{T_1}\Phi\rangle = 1$. This function has the form of a determinant

   $$e^{T_1}|\Phi\rangle = e^{\sum_a t^1_a a^a_1 + \cdots + \sum_a t^n_a a^a_n} a^\dagger_1 \cdots a^\dagger_n|\mathrm{vac}\rangle = \widetilde{a}^\dagger_1 \cdots \widetilde{a}^\dagger_n|\mathrm{vac}\rangle = |\widetilde{\Phi}\rangle \qquad\qquad \widetilde{a}^\dagger_i \equiv \exp(\textstyle\sum_a t^i_a a^a_i)\, a^\dagger_i$$

   since $\sum_a t^i_a a^a_i$ commutes with all creation operators except $a^\dagger_i$. The transformed orbitals are given by

   $$|\widetilde{\psi}_i\rangle = \widetilde{a}^\dagger_i|\mathrm{vac}\rangle = \exp(\textstyle\sum_a t^i_a a^\dagger_a a_i)\, a^\dagger_i|\mathrm{vac}\rangle = (1 + \sum_a t^i_a a^\dagger_a a_i)\, a^\dagger_i|\mathrm{vac}\rangle = |\psi_i\rangle + \sum_a t^i_a|\psi_a\rangle$$

   using $a_i^2 = 0$ and $a_i a_i^\dagger|\mathrm{vac}\rangle = |\mathrm{vac}\rangle$.

2. *Any intermediately normalized determinant $\widetilde{\Phi} = \frac{1}{\sqrt{n!}}\det(\widetilde{\psi}_1 \cdots \widetilde{\psi}_n)$ can be written as $e^{T_1}\,\Phi$.*

   Proof: Intermediate normalization is only possible if $\widetilde{\Phi}$ has non-zero overlap with the reference determinant. Therefore, $\widetilde{\Phi}$ can be written as $\Phi'/\langle\Phi|\Phi'\rangle$ where Φ' is a Slater determinant. The normalization factor is given by

   $$\langle\Phi|\Phi'\rangle = \tfrac{1}{n!}\textstyle\sum_{\pi,\sigma}^{\mathrm{S}_n} \varepsilon_\pi\varepsilon_\sigma \langle\psi_{\pi(1)}|\psi'_{\sigma(1)}\rangle \cdots \langle\psi_{\pi(n)}|\psi'_{\sigma(n)}\rangle = \sum_\sigma^{\mathrm{S}_n} \varepsilon_\sigma \langle\psi_1|\psi'_{\sigma(1)}\rangle \cdots \langle\psi_n|\psi'_{\sigma(n)}\rangle = \det(\mathbf{U}_\mathrm{oo})$$

   following notation A.1. Therefore, $\widetilde{\Phi} = \Phi'/\det(\mathbf{U}_\mathrm{oo}) = \Phi'\det(\mathbf{U}^{-1}_\mathrm{oo})$ and the rows of $\widetilde{\Phi}$ are given by

   $$|\widetilde{\boldsymbol{\psi}}_\mathrm{o}\rangle = |\boldsymbol{\psi}'_\mathrm{o}\rangle\,\mathbf{U}^{-1}_\mathrm{oo} = |\boldsymbol{\psi}_\mathrm{o}\rangle + |\boldsymbol{\psi}_\mathrm{v}\rangle\,\mathbf{U}_\mathrm{vo}\mathbf{U}^{-1}_\mathrm{oo}$$

   where we have expanded $|\boldsymbol{\psi}'_\mathrm{o}\rangle$ according to eq A.1. The columns of this equation are $\widetilde{\psi}_i = \psi_i + \sum_a \psi_a\,(\mathbf{U}_\mathrm{vo}\mathbf{U}^{-1}_\mathrm{oo})_{ai}$. Referring back to the first proposition, this shows that $\widetilde{\Phi} = e^{T_1}\Phi$ with $t^i_a = (\mathbf{U}_\mathrm{vo}\mathbf{U}^{-1}_\mathrm{oo})_{ai}$.

B Orbital rotations

Definition B.1. ***Normal matrix.*** A square matrix satisfying $\mathbf{N}^\dagger\mathbf{N} = \mathbf{N}\mathbf{N}^\dagger$ is termed *normal*. Several important kinds of matrices meet this criterion: *Hermitian matrices*, $\mathbf{H}^\dagger = \mathbf{H}$; *anti-Hermitian matrices*, $\mathbf{A}^\dagger = -\mathbf{A}$; and *unitary matrices*, $\mathbf{U}^\dagger = \mathbf{U}^{-1}$. Note that Hermitian and anti-Hermitian matrices can always be written as $\mathbf{X} + \mathbf{X}^\dagger$ and $\mathbf{X} - \mathbf{X}^\dagger$.

Remark B.1. The Spectral Theorem[4] says that every normal matrix can be diagonalized by a unitary transformation. A direct corollary[5] is that the eigenvalues of Hermitian, anti-Hermitian, and unitary matrices can be written as follows.

$$h^* = h \implies h = \phi \qquad a^* = -a \implies a = i\phi \qquad u^* = u^{-1} \implies u = e^{i\phi} \qquad \phi \in \mathbb{R} \tag{B.1}$$

In words, Hermitian eigenvalues are real, anti-Hermitian eigenvalues are pure imaginary, and unitary eigenvalues lie on the unit circle. Note that unitary eigenvalues have the form $u = \exp(a)$ where a is an anti-Hermitian eigenvalue. This implies that any unitary matrix $\mathbf{U}$ can be written as $\exp(\mathbf{A})$, where $\mathbf{A}$ is anti-Hermitian.

Remark B.2. According to def B.1 and rmk B.1, unitary transformations of the spin-orbitals can be parametrized as

$$\psi'_p = \sum_q \psi_q (\exp(\mathbf{X} - \mathbf{X}^\dagger))_{qp} \tag{B.2}$$

in terms a square matrix $\mathbf{X}$. The anti-Hermitian form of this parametrization leads to redundancies. In particular, notice that $\mathbf{X} = [z\,\delta_{pq}]$ generates the same transformation as $\mathbf{X}^\dagger = [-z^*\,\delta_{qp}]$. These redundancies are eliminated by setting the upper or lower triangle of $\mathbf{X}$ to zero. For most single-reference methods, transformations within the occupied and virtual blocks are also redundant because they don't change the energy.[6] Setting all but one of the off-diagonal blocks to zero yields a non-redundant parametrization. Here, we will choose $\mathbf{X}_{\text{vo}} = [x^i_a]$.

Proposition B.1. *The identity* $\exp(G)\, a^\dagger_p \exp(-G) = \sum_q a^\dagger_q \,(\exp(\mathbf{G}))_{qp}$ *holds for any* $G = \sum_{pq} (\mathbf{G})_{pq}\, a^\dagger_p a_q$.

Proof: This follows from $[G, \cdot]^m(a^\dagger_p) = \sum_q a^\dagger_q (\mathbf{G}^m)_{qp}$, which we will prove by induction. For $m = 0$ the statement is trivially true. If we assume it holds for m, then the following shows that it also holds for $m+1$,[7]

$$[G, \cdot]^{m+1}(a^\dagger_p) = \textstyle\sum_q [G, a^\dagger_q]\,(\mathbf{G}^m)_{qp} = \sum_{qr} a^\dagger_r (\mathbf{G})_{rq} (\mathbf{G}^m)_{qp} = \sum_r a^\dagger_r (\mathbf{G}^{m+1})_{rp}$$

which completes the induction. Substituting this result into the Hausdorff expansion of $\exp(G)\, a^\dagger_p \exp(-G)$ recognizing the Taylor expansion of $\exp(\mathbf{G})$ completes the proof.

Remark B.3. The identity $a^\dagger_p|\text{vac}\rangle = |\psi_p\rangle$ implies that creation operators transform like orbitals. Therefore, the creation operator corresponding to ψ'_p in equation B.2 is given by $a'^\dagger_p = \sum_q a^\dagger_q (\exp(\mathbf{X} - \mathbf{X}^\dagger))_{qp}$. By prop B.1, this is equivalent to

$$a'^\dagger_p = \exp(X - X^\dagger) a^\dagger_p \exp(X^\dagger - X) \qquad X \equiv \sum_{ai} x^i_a\, a^a_i \tag{B.3}$$

where we have eliminated the redundant parameters. Substituting this into $|\Phi'_{(p_1 \cdots p_n)}\rangle = a'^\dagger_{p_1} \cdots a'^\dagger_{p_n}|\text{vac}\rangle$ gives a convenient expression for determinants of the transformed orbitals. [8]

$$|\Phi'_{(p_1 \cdots p_n)}\rangle = \exp(X - X^\dagger)|\Phi_{(p_1 \cdots p_n)}\rangle \qquad \langle\Phi'_{(p_1 \cdots p_n)}| = \langle\Phi_{(p_1 \cdots p_n)}|\exp(X^\dagger - X) \tag{B.4}$$

[4] See `https://en.wikipedia.org/wiki/Spectral_theorem`

[5] Since there exists a basis in which $\mathbf{N}$ is diagonal, statements about $\mathbf{N}$ translate into statements about its eigenvalues.

[6] This is what allows us to diagonalize the Fock matrix in canonical Hartree-Fock theory.

[7] The second equality here follows from expanding G and using $[a^\dagger_r a_s, a^\dagger_q] = :a^\dagger_r \overbrace{a_s a^\dagger_q}: = a^\dagger_r\, \delta_{sq}$.

[8] This follows from $\exp(X^\dagger - X)\exp(X - X^\dagger) = 1$ and $\exp(X^\dagger - X)|\text{vac}\rangle = 0$.