Variational Functionals for Excited States

Naoum C. Bacalis
Theoretical and Physical Chemistry Institute, National Hellenic Research Foundation,
Vasileos Constantinou 48, GR-11635 Athens, Greece

Functionals Ω_n that have local minima at the excited states of a non degenerate Hamiltonian are presented. Then, improved mutually orthogonal approximants of the ground and the first excited state are reported.

PACS: 31.15.xt, 31.15.-p, 31.10.+z, 02.70.?c

In the following the Hamiltonian expectation value of a trial wave function, ϕ, is denoted by $E\phi$ and is called energy of ϕ. The Hamiltonian eigenfunctions (assumed non-degenerate) are denoted by using the symbol ψ. All functions are assumed real and normalized.

According to the Hylleraas, Undheim, and McDonald [HUM] theorem[1] the higher roots of the secular equation tend to the excited state energies from *above*. But it should be observed that among all functions ϕ_1, which are orthogonal to an available ground state approximant ϕ_0, the Gram – Schmidt orthonormal to ϕ_0

$$\phi_1^+ \equiv \frac{\psi_1 - \phi_0 \langle \psi_1 | \phi_0 \rangle}{\sqrt{1 - \langle \psi_1 | \phi_0 \rangle^2}}$$

which is the *closest*[2] to the exact ψ_1 (i.e. with the largest projection $\langle \psi_1 | \phi_1 \rangle^2$ - not decreased by the presence of any other components) lies energetically *below the exact* $E\psi_1$, only if $E\phi_0 < E\psi_1$:

$$E\phi_1^+ = E\psi_1 - \frac{(E\psi_1 - E\phi_0)\langle \psi_1 | \phi_0 \rangle^2}{1 - \langle \psi_1 | \phi_0 \rangle^2} < E\psi_1 ,$$

Therefore, the 2nd HUM root, ϕ_1^{HUM}, lying higher than ψ_1, $E\phi_1^{HUM} > E\psi_1$, is necessarily *not* the closest to ψ_1 (while orthogonal to ϕ_0).

On the other hand, minimizing the energy orthogonally to the available ϕ_0, does not lead to the *closest* either: Passing through $E\phi_1^+$, it leads to an *even lower* energy: Because for any $\phi_1^{\perp +}$, chosen simultaneously orthogonal to both ϕ_0 and ϕ_1^+, the Hamiltonian opens the energy gap between $E\phi_1^{\perp +}$ and $E\phi_1^+$, so that, the lowest of the Hamiltonian eigenfunctions Ψ^-, Ψ^+, (both orthogonal to ϕ_0) on the subspace of { $\phi_1^{\perp +}$, ϕ_1^+ }, lies lower than $E\phi_1^+$, i.e. $E\Psi^- < E\phi_1^+ < E\psi_1$, so that the lowest, ϕ_1^{MIN}, of all such Ψ^- s, obtained by minimizing the energy orthogonally to ϕ_0, lies even lower than $E\phi_1^+$. Therefore, ϕ_1^{MIN} is *not* the closest to ψ_1 either (while orthogonal to ϕ_0). (In fact, an appropriate sum $\Psi = \Psi^- \sqrt{\frac{E\Psi^+ - E\psi_1}{E\Psi^+ - E\Psi^-}} \pm \Psi^+ \sqrt{\frac{E\psi_1 - E\Psi^-}{E\Psi^+ - E\Psi^-}}$, orthogonal to ϕ_0, has energy $E\Psi = E\psi_1$, with $\langle \psi_1 | \Psi \rangle^2$ not necessarily large.)

Thus, seeking ϕ_1, approximant to ψ_1, orthogonal to an approximant ϕ_0, either by the HUM theorem or by orthogonal optimization, does neither lead to ϕ_1^+, the *closest* to ψ_1, nor does it raise the energy going from ϕ_1^+ to ψ_1 (which is orthogonal to ψ_0, not to ϕ_0). As Shull and Löwdin[3] have shown, the excited states can be calculated without knowledge of ψ_0. Therefore, a variational functional for ϕ_1 would be desirable, that leads to ψ_1 not necessarily orthogonally to the available ϕ_0, allowing subsequent improvement of ϕ_0 orthogonally to ϕ_1:

Construction: For a non-degenerate Hamiltonian of (unknown) bound eigenstates of a specific type of symmetry, ψ_0, ψ_1, and eigenenergies $E\psi_0 < E\psi_1 < ...$, a normalized approximant of ψ_n can be expanded as

$$\phi_n = \sum_{i<n} \psi_i \langle \psi_i | \phi_n \rangle + \psi_n \sqrt{1 - \sum_{i<n} \langle \psi_i | \phi_n \rangle^2 - \sum_{i>n} \langle \psi_i | \phi_n \rangle^2} + \sum_{i>n} \psi_i \langle \psi_i | \phi_n \rangle \quad (1.1)$$

where the overlap coefficients are small. The energy is

$$E\phi_n = E\psi_n - \sum_{i<n}\left(E\psi_n - E\psi_i\right)\left\langle\psi_i\middle|\phi_n\right\rangle^2 + \sum_{i>n}\left(E\psi_i - E\psi_n\right)\left\langle\psi_i\middle|\phi_n\right\rangle^2 \equiv E\psi_n - P_L + P_H\,, \tag{1.2}$$

an n-order saddle point, where the lower and higher than-n parts, P_L and P_H, are positive (so that $E\psi_n - P_L \le E\phi_n \le E\psi_n + P_H$).

The minimum of the following paraboloid, defined by

$$E\psi_n + P_L + P_H = E\phi_n + 2P_L \tag{1.3}$$

determines $\phi_n \to \psi_n$, in terms of the lower than-n information (P_L). An expression for the behaviour of P_L can be found by first considering, to leading order in coefficients, the overlap and the Hamiltonian matrix elements in terms of the (similarly predetermined as described here) approximants $\phi_i, i<n$:

$$\begin{aligned}\left\langle\phi_i\middle|\phi_n\right\rangle &= \left\langle\psi_i\middle|\phi_n\right\rangle + \left\langle\psi_n\middle|\phi_i\right\rangle + \cdots\\ \left\langle\phi_i\middle|H\middle|\phi_n\right\rangle &= E\psi_i\left\langle\psi_i\middle|\phi_n\right\rangle + E\psi_n\left\langle\psi_n\middle|\phi_i\right\rangle + \cdots\,.\end{aligned} \tag{1.4}$$

Substituting $\left\langle\psi_i\middle|\phi_n\right\rangle$ from Eqs. (1.4) to each term of P_L in Eq. (1.2) gives, to leading order, $\left(E\psi_n\left\langle\phi_i\middle|\phi_n\right\rangle - \left\langle\phi_i\middle|H\middle|\phi_n\right\rangle\right)^2 \big/ \left(E\psi_n - E\psi_i\right)$, which suggests an examination, in terms of known quantities, of the expression $\sum_{i<n}\left[\left(E\phi_n\left\langle\phi_i\middle|\phi_n\right\rangle - \left\langle\phi_i\middle|H\middle|\phi_n\right\rangle\right)^2 \big/ \left(E\phi_n - E\phi_i\right)\right]$. This, as directly verified, when both $\phi_i = \psi_i$ and [in Eq.(1.2)] $P_H \to 0$, reduces to $P_L\left(1 - \sum_{i<n}\left\langle\phi_i\middle|\phi_n\right\rangle^2\right)$. Therefore, for $P_H \ne 0$ the behaviour of the paraboloid of Eq. (1.3) close to ψ_n is reasonably described by the functional Ω_n :

$$E\psi_n + P_L + P_H = E\phi_n + 2P_L \to \Omega_n \equiv E\phi_n + 2\,\frac{\displaystyle\sum_{i<n}\frac{\left(E\phi_n\left\langle\phi_i\middle|\phi_n\right\rangle - \left\langle\phi_i\middle|H\middle|\phi_n\right\rangle\right)^2}{E\phi_n - E\phi_i}}{1 - \displaystyle\sum_{i<n}\left\langle\phi_i\middle|\phi_n\right\rangle^2} \tag{1.5}$$

with a local minimum at $\phi_n = \psi_n$, which is paraboloidal, by construction, when $\phi_i = \psi_i$.

Proof: Ω_n has a true local minimum at $\phi_n = \psi_n$ when ϕ_i are *approximants* of ψ_i ($\phi_i \approx \psi_i$), while $E\phi_n$ has a saddle point there: By collecting the contribution of the higher than-n subspace for each ϕ_i wave function , $i \le n$, to the contribution of a normalized function $\phi_i^{\perp\{n\}}, i \le n$, orthogonal to all lower than-n ψ_i eigenfunctions, i.e.

$$\phi_i^{\perp\{n\}} = \sum_{j>n}\psi_j\left\langle\psi_j\middle|\phi_i\right\rangle \Bigg/ \sqrt{\sum_{j>n}\left\langle\psi_j\middle|\phi_i\right\rangle^2}\,,\; i \le n\,, \tag{1.6}$$

where the overlap and Hamiltonian matrix elements are generally non-zero, $\left\langle\phi_i^{\perp\{n\}}\middle|\phi_j^{\perp\{n\}}\right\rangle \ne 0$, $\left\langle\phi_i^{\perp\{n\}}\middle|H\middle|\phi_j^{\perp\{n\}}\right\rangle \ne 0$, $i, j \le n$, and whose energies, obviously, are $E\phi_i^{\perp\{n\}} > E\psi_n, i \le n$, it is directly verified that all the principal minors A_n^i , $i \le n$, of the Hessian determinant A_n^n of Ω_n , along the main diagonal, i.e. those which are required by the second derivatives theorems of calculus (Sylvester's theorem), are, at the desired place $\phi_n = \psi_n$, $\phi_i \ne \psi_i$, $i<n$, positive, if ϕ_i are close to ψ_i: Each principal minor determinant (denoted by the main diagonal)

$$A_n^{k<n} \equiv Det\left[\frac{\partial^2\Omega_n}{\partial\left\langle\psi_0\middle|\phi_n\right\rangle\partial\left\langle\psi_0\middle|\phi_n\right\rangle}\cdots\frac{\partial^2\Omega_n}{\partial\left\langle\psi_i\middle|\phi_n\right\rangle\partial\left\langle\psi_i\middle|\phi_n\right\rangle}\cdots\frac{\partial^2\Omega_n}{\partial\left\langle\psi_k\middle|\phi_n\right\rangle\partial\left\langle\psi_k\middle|\phi_n\right\rangle}\right]\Bigg|_{\phi_n=\psi_n\,,\,\phi_i\ne\psi_i\,,\,i<k}$$

equals

$$A_n^{k<n} = 2^k\prod_{i=0}^{k}\left(E\psi_n - E\psi_i\right) > 0\;\left(+O\left[\left\langle\psi_q\middle|\phi_r\right\rangle\left\langle\psi_s\middle|\phi_t\right\rangle\right]\right)\,. \tag{1.7}$$

where there are no coefficients (which depend on the quality of ϕ_i) of 1st power, while the Hessian itself

$$A_n^n \equiv Det\left[\frac{\partial^2\Omega_n}{\partial\langle\psi_0|\phi_n\rangle\partial\langle\psi_0|\phi_n\rangle}\cdots\frac{\partial^2\Omega_n}{\partial\langle\psi_i|\phi_n\rangle\partial\langle\psi_i|\phi_n\rangle}\cdots\frac{\partial^2\Omega_n}{\partial\langle\phi_n^{\perp\{n\}}|\phi_n\rangle\partial\langle\phi_n^{\perp\{n\}}|\phi_n\rangle}\right]_{\phi_n=\psi_n,\,\phi_i\neq\psi_i,\,i<n}$$

equals

$$A_n^n = 2^n\left(E\phi_n^{\perp\{n\}} - E\psi_n\right)\prod_{i=0}^{n-1}\left(E\psi_n - E\psi_i\right) > 0\ \left(+O\left[\langle\psi_q|\phi_r\rangle\langle\psi_s|\phi_t\rangle\right]\right)\ . \tag{1.8}$$

If ϕ_i are close to ψ_i, all these determinants of Eqs. (1.7 - 1.8) are positive, hence the Hessian matrix is positive definite, therefore, the functional Ω_n has a local minimum at $\phi_n = \psi_n$, which determines ψ_n if all ϕ_i approximants of ψ_i, $i \le n$, are known. Obviously, Ω_0 reduces to the Eckart[4] theorem for ψ_0.

The functional Ω_n passes from all ψ_i. A way to identify the desired ψ_n for atoms and for diatomic molecules, is to expand (for atoms) in a basis of Slater type exponentials whose prefactors are *not monomials*, but rather they are variationally optimized *polynomials*: initially starting from the *identifiable* associated Laguerre polynomials, because these are *not severely modified* during optimization; Also identifiable (for diatomic molecules) are the (separable into radial and angular parts) variationally optimized *one*-electron-diatomic-*molecule*-type orbitals. Both significantly reduce the size of a configuration interaction expansion. [5]

Improving ϕ_0 *orthogonally to* ψ_1: If ψ_1 were known it would be possible to improve ϕ_0 orthogonally to ψ_1:[2] On the subspace of { ϕ_0, ψ_1 } the highest Hamiltonian eigenvector, Ψ^+, is

$$\Psi^+ = \psi_1\ .$$

The lowest, Ψ^-, is orthogonal to ψ_1,

$$\Psi^- = \phi_0^+ \equiv \frac{\phi_0 - \psi_1\langle\psi_1|\phi_0\rangle}{\sqrt{1-\langle\psi_1|\phi_0\rangle^2}}$$

with energy

$$E\phi_0^+ = E\phi_0 - \frac{\left(E\psi_1 - E\phi_0\right)\langle\psi_1|\phi_0\rangle^2}{1-\langle\psi_1|\phi_0\rangle^2} \le E\phi_0 \tag{1.9}$$

(same or better than ϕ_0). Further, rotating ϕ_0^+ around ψ_1 improves ϕ_0^+ as follows: After introducing (e.g. by one more configuration) a function $\phi_0^{(2+)}$ orthogonal to both { ϕ_0^+, ψ_1 }, then, in the subspace of { $\phi_0^+, \phi_0^{(2+)}$ } (both orthogonal to ψ_1), the lowest Hamiltonian eigenvector $\Psi^- \equiv \phi_0^-$ has energy $E\phi_0^- \le E\phi_0^+$, closer to $E\psi_0$, because the Hamiltonian opens the energy gap between $\{E\phi_0^+, E\phi_0^{(2+)}\}$ (in a 3-dimensional function space { ψ_0, ψ_1, ψ_k } this would be exactly $E\psi_0$ as it can be directly verified). $E\phi_0^-$ can be further improved by further rotating around ψ_1 similarly, i.e. after introducing another function $\phi_0^{(3+)}$ orthogonal to both { ϕ_0^-, ψ_1 } by calculating in the subspace of { $\phi_0^-, \phi_0^{(3+)}$ } (both orthogonal to ψ_1) the lowest eigenvector $\Psi^- \equiv \phi_0^{(2-)}$ which has energy $E\phi_0^{(2-)} \le E\phi_0^-$ (even closer to $E\psi_0$); and so on.

Improving ϕ_0 *orthogonally to* ϕ_1: Since ψ_1 is never exactly known, then, it may still be possible to improve ϕ_0 orthogonally to ϕ_1, the best available approximant of ψ_1, by first computing ϕ_0^+ orthogonal to ϕ_1,

$$\phi_0^+ \equiv \frac{\phi_0 - \phi_1\langle\phi_1|\phi_0\rangle}{\sqrt{1-\langle\phi_1|\phi_0\rangle^2}} \tag{1.10}$$

if the condition

$$\mathrm{E}\phi_0^+ = \frac{\mathrm{E}\phi_0 + \mathrm{E}\phi_1\langle\phi_1|\phi_0\rangle^2 - 2\langle\phi_0|H|\phi_1\rangle\langle\phi_1|\phi_0\rangle}{1-\langle\phi_1|\phi_0\rangle^2} \le \mathrm{E}\phi_0 \tag{1.11}$$

is attainable. Indeed, by expanding about ψ_1, as directly verified, this condition, to leading order, reads $\left(\mathrm{E}\psi_1 - \mathrm{E}\psi_0\right)\left(1-\langle\psi_1|\phi_0\rangle^2\right) \ge \left(\mathrm{E}\phi_0^{\perp\{1\}} - \mathrm{E}\psi_0\right)\langle\phi_0^{\perp\{1\}}|\phi_0\rangle^2$, which is not impossible. Here [c.f. Eq. (1.6)] $\phi_0^{\perp\{1\}}$ is the normalized function, orthogonal to both $\{\psi_0,\psi_1\}$, collecting all higher than-1 terms of ϕ_0. For ϕ_0,ϕ_1 very close to ψ_0,ψ_1, as directly verified by expanding about ψ_0,ψ_1 the condition is satisfied when $\langle\psi_0|\phi_1\rangle^2 \le \langle\psi_1|\phi_0\rangle^2$ (indicative of the relative quality of the approximants). Incidentally, all other (small) components (out of the plane of ψ_0,ψ_1) are less relevant, so that the opposite procedure of optimizing ϕ_1 orthogonally to ϕ_0 can lead to ϕ_1^{MIN} unpredictably *far* from ψ_1 with *still* $E\phi_1^{MIN} \lesssim E\psi_1$, as shown in the following example.

Example: Even in the subspace { ψ_0, ψ_1, ψ_2 }, the orthonormal trial functions $\phi_0 = a\psi_0 + b\psi_2$, $\phi_1 = b\psi_0 - a\psi_2$ with $a = \sqrt{[(E\psi_1-\varepsilon)-E\psi_0]/(E\psi_2-E\psi_0)}$, $b = \sqrt{[E\psi_2-(E\psi_1-\varepsilon)]/(E\psi_2-E\psi_0)}$, (small ε), have energies $E\phi_0 = E\psi_0 + E\psi_2 - (E\psi_1-\varepsilon) \cong E\psi_0+\varepsilon$ (if $E\psi_2 - E\psi_1$ is small), $E\phi_1 = E\psi_1 - \varepsilon$, while ϕ_0 reasonably, but not particularly accurately, approximates ψ_0 (for instance, for He ^{1}S, in a.u., $E\psi_0 = -2.903$, $E\psi_1 = -2.146$, $E\psi_2 = -2.06$, $\phi_0 = 0.9476\,\psi_0 + 0.3194\,\psi_2$ has $E\phi_0 = -2.817$ and $\phi_1 = 0.3194\,\psi_0 - 0.9476\,\psi_2$ has $E\phi_1 = -2.146 = E\psi_1$, while ϕ_1 is orthogonal to both ϕ_0 *and* ψ_1), so that, *any function orthogonal to the same* ϕ_0 could be a minimization "result", ϕ_1^{MIN}, with *arbitrary* $\langle\psi_1|\phi_1^{MIN}\rangle$ and with $E\psi_1 - \varepsilon \le E\phi_1^{MIN} \le E\psi_1$.

Demonstration of Ω_1: Minimization of Ω_1, for the same ϕ_0 of He, as above, by varying $\phi_1 = c\,\psi_0 + d\,\psi_2 + \psi_1\sqrt{1-c^2-d^2}$, yields $c < tol = 10^{-8}$, $d < tol$, with $E\phi_1 = -2.146$ [so that $\phi_1 = \psi_1$ and , from Eq. (1.10), $\phi_0^+ = \phi_0$].

Further improvement of ϕ_0: If $\mathrm{E}\phi_0^+ \le \mathrm{E}\phi_0$ [Eq. (1.11)], then, by rotating around ϕ_1, as described above [after Eq. (1.9)], since the Hamiltonian always opens the energy gap between mutually orthogonal functions (all orthogonal to ϕ_1), ϕ_0^+ can be further improved (until $\langle\psi_0|\phi_1\rangle^2 > \langle\psi_1|\phi_0\rangle^2$), by always taking the lowest current eigenfunction $\phi_0^{(m-)} = \Psi^-$. At any step, $\phi_0^{(m-)}$ can be used as a new ϕ_0 to improve ϕ_1 via Ω_1 of Eq. (1.5). In the above example of He, rotating ϕ_0 around ϕ_1, gives $\phi_0^{(1-)} = \Psi^- = \psi_0$ (and $\Psi^+ = \psi_2$).

Technicalities: If the higher eigenvalues approach each other, then the second derivatives diminish and the paraboloid Ω_n flattens within the tolerance criterion ε_n, used in the Ω_n minimization. Then it might be desirable to steepen it near the minimum. The simplest way would be to multiply Ω_n by a large number N, so as to distinguish the differences within the *same* ε_n. Also, it might be possible, by introducing one more variable, E_F, to minimize the functional $F[\Omega_n, E_F] \equiv \Omega_n + \left|\dfrac{\Omega_n - E_F}{E_F T}\right|$: if T is chosen in the order of Ω_n 's curvature radius at ψ_n, ~(inverse of second derivatives, estimated by the Hessian minors, or by trial) , then, as directly verified by expanding about ψ_n, F is a paraboloid with minimum at $\phi_n = \psi_n$ with $F[E\psi_n, E\psi_n] = E\psi_n$.

Summary: Ω_n [Eq. (1.5)] has a local minimum at the excited state ψ_n, where $\Omega_n = E\psi_n$ and $\phi_n = \psi_n$. If ϕ_1 is a better approximant to ψ_1 than ϕ_0 is to ψ_0 [i.e. if, from Eq. (1.11), $\mathrm{E}\phi_0^+ \le \mathrm{E}\phi_0$], then ϕ_0 can be improved orthogonally to ϕ_1.

---

[1] E. Hylleraas and B. Undheim, Z. Phys. **65**, 759 (1930); J. K. L. McDonald, Phys. Rev. **43**, 830 (1933).
[2] N. C. Bacalis, International Conference of Computational Methods in Sciences and Engineering 2007 (ICCMSE 2007), Agios Ioannis Peristeron, Corfu, Greece, 25-30 September 2007, Lecture Series, edited by T. Simos and G. Maroulis (in print).
[3] H. Shull and P. -O. Löwdin, Phys. Rev. **110**, 1466 (1958).
[4] C. Eckart, Phys. Rev. **36**, 878 (1930).
[5] Z. Xiong and N. C. Bacalis, Commun. Math. Comput. Chem. **53**, 283 (2005); Z. Xiong, M. Velgakis, and N. C. Bacalis, Int. J. Q. Chem. **104**, 418 (2005); N. C. Bacalis, J. Phys. B **29**, 1587 (1996).