Variational Method
(See CTDL 1148-1155, [Variational Method]
252-263, 295-307[Density Matrices])

Last time:
Quasi-Degeneracy → Diagonalize a part of infinite $H$
* sub-matrix: $H^{(0)} + H^{(1)}$
* corrections for effects of out-of-block elements: $H^{(2)}$
  (the Van Vleck transformation)
* diagonalize $H_{\text{eff}} = H^{(0)} + H^{(1)} + H^{(2)}$

coupled H-O’s example: the 2 : 1 ($\omega_1 \approx 2 \omega_2$) Fermi resonance polyads

1. Perturbation Theory vs. Variational Method: non-orthogonal → $S$ (overlap matrix)
2. Variational Theorem
3. Stupid nonlinear variation
4. Linear Variation → new kind of secular Equation
5. Linear combined with nonlinear variation
6. Strategies for criteria of goodness — various kinds of variational calculations

1. Perturbation Theory vs. Variational Method

**Perturbation Theory** effectively uses $\infty$ basis set
non-degenerate: diagonalize $H_{\text{eff}}$
quasi-degenerate: non-diagonal $H_{\text{eff}}$ (model with quantum number scaling)
goals: parametrically parsimonious fit model, $H_{\text{eff}}$
fit parameters (molecular constants) $\leftrightarrow$ parameters that define $V(x)$

\[
\text{order-sorting } \frac{H^{(1)}_{nk}}{E^{(0)}_n - E^{(0)}_k} < 1
\]
— errors smaller than this “mixing angle” times the previous order non-zero correction term

(n is in-block, k is out-of block) because diagonalization is to $\infty$ order (within block).

**Variational Method**

★ best possible estimate for lowest few $E_n, \psi_n$ (and properties derivable from these) using a finite (often huge) basis set and the exact form of $H$. 

modified 8/13/20 8:25 AM
Vast majority of computer time in Chemistry is spent in variational calculations. 
Goal is numbers. Insight is secondary.

“Ab Initio” vs. “semi-empirical” or “fitting”

[intentionally bad basis set: Hückel, tight binding – qualitative behavior obtained by a fit to a few microscopic–like control parameters]

2. Variational Theorem

If \( \phi \) is approximation to eigenfunction of \( \hat{A} \) that belongs to the lowest eigenvalue, \( a_0 \), then

\[
\alpha \equiv \frac{\langle \phi | \hat{A} | \phi \rangle}{\langle \phi | \phi \rangle} \geq a_0
\]

the Variational Theorem

PROOF: eigenbasis (which we do not know – but know that it must exist)

\[
\hat{A} | n \rangle = a_n | n \rangle
\]

expand \( | \phi \rangle \) in eigenbasis of \( \hat{A} \), exploiting completeness

\[
| \phi \rangle = \sum_n | n \rangle \langle n | \phi \rangle
\]

\[
\langle \phi | \hat{A} | \phi \rangle = \sum_{n,n'} \langle \phi | n \rangle \langle n | \hat{A} | n' \rangle \langle n' | \phi \rangle = \sum_n \langle \phi | n \rangle^2 a_n
\]

\[
\langle \phi | \phi \rangle = \sum_n \langle \phi | n \rangle \langle n | \phi \rangle = \sum_n \langle \phi | n \rangle^2
\]

\[
\alpha \equiv \frac{\langle \phi | \hat{A} | \phi \rangle}{\langle \phi | \phi \rangle} = \frac{\sum_n a_n \langle n | \phi \rangle^2}{\sum_{n'} \langle n' | \phi \rangle^2}
\]

all terms in both sums are \( \geq 0 \)

subtract \( a_0 \) from both sides

\[
\alpha - a_0 = \sum_n (a_n - a_0) \langle n | \phi \rangle^2 \geq 0
\]

again, all terms in both sums are \( \geq 0 \)

because \( \sum_n \langle n | \phi \rangle^2 = \sum_{n'} \langle n' | \phi \rangle^2 \)
We are done because, by definition of \( a_0 \), \( a_n \geq a_0 \) for all \( n \) and all terms in sum are \( \therefore \geq 0 \).

\[
\therefore \alpha \geq a_0.
\]

QED \( \left\{ \text{but useless because we do not know } a_n \text{ or } \langle n | \phi \rangle \text{ in advance} \right\} \)

It is possible to perform a variational calculation for any \( A \), not limited to \( H \).

3. Stupid Nonlinear Variation

Use the wrong functional form or the wrong variational criterion to get poor results — illustrates that the variational function must have sufficient flexibility and the variational criterion must be as it is specified in the variational theorem, as opposed to a clever shortcut.

The \( H \) atom Schrödinger Equation \((\ell = 0)\)

\[
\mathbf{H} = \begin{bmatrix}
\frac{1}{2} \frac{\partial}{\partial r} r^2 \frac{\partial}{\partial r} & -\frac{1}{r} \\
\end{bmatrix}
\]

and we know \( \psi_{1s} (r) = \langle r | 1s \rangle = \pi^{-1/2} e^{-r} \)

\[
E_{1s} = -1/2 \text{ au}
\]

\[
\left[ 1 \text{ au} = 219475 \text{ cm}^{-1} \right]
\]

but try \( \langle r | \phi \rangle = \left[ \frac{\xi^3}{2\pi} \right]^{1/2} (\xi r) e^{-\xi r} \) normalized for all \( \xi \)

\( \xi \) is a scale factor that controls overall size of \( \phi (r) \)

[actually this is the form of \( \psi_{2p} (r) \)] which at \( \xi = 1 \) is necessarily orthogonal to \( \psi_{1s} \)!

STUPID!

\[
\left( \phi(0) = 0 \quad \text{but} \quad \psi_{1s} (0) = \pi^{-1/2} \right)
\]

\[
\varepsilon = \frac{\langle \phi | \mathbf{H} | \phi \rangle}{\langle \phi | \phi \rangle} = \frac{4}{3} \left( \frac{\xi^2 - 3\xi}{8} \right) \text{ skipped a lot of algebra}
\]

minimize \( \varepsilon \): \( \frac{d\varepsilon}{d\xi} = 0 \quad \xi_{\text{min}} = 3/2 \rightarrow \varepsilon_{\text{min}} = -3/8 \text{ au} \)

FAILURE! \( \left[ \text{c.f. the true values: } E_{1s} = -1/2 \text{ au}, E_{2s} = -\frac{1}{8} \text{ au} \right] \)

[insufficiently flexible variational function.]
5.73 Lecture #18

Try something clever (but lazy):
What is the value of $\xi$ that maximizes $\langle \phi | 1s \rangle$?

for the best variational $\xi = 3/2, C_{1s} \equiv \langle \phi (\xi = 3/2) | 1s \rangle = 0.9775$

but if instead we maximize $C_{1s}$ wrt $\xi$: $\xi = 5/3 \rightarrow C_{1s} = 0.9826$  

Is this better?

No. The value of $\varepsilon = -0.370$ results from maximizing $C_{1s}$, which is a poorer bound to $\varepsilon$ than obtained by minimizing $\varepsilon$.

$\xi = 3/2 \rightarrow \varepsilon = -0.375$

* need flexibility in $\phi$

* can't improve on $\frac{d\varepsilon}{d\xi} = 0$ by employing an alternative variational strategy

This was stupid anyway because we would never use the variational method when we already know the answer!

4. Linear Variation $\rightarrow$ Secular Equation

$$\phi = \sum_{n=1}^{N} c_n \chi_n$$

$$\langle \chi_n | H | \chi_{n'} \rangle = H_{nn'}$$

$$\langle \chi_n | \chi_{n'} \rangle = S_{nn'}$$

overlap integrals

(non-orthogonal basis sets are often convenient)

also not necessarily normalized

$$\varepsilon = \frac{\langle \phi | H | \phi \rangle}{\langle \phi | \phi \rangle} = \frac{\sum_{n,n'} c_n c_{n'} H_{nn'}}{\sum_{m,m'} c_m c_{m'} S_{mm'}}$$

Rearrange this equation.

To find the minimum value of $\varepsilon$,

take $\frac{\partial}{\partial c_j}$ for each $j$ and require that

$$\frac{\partial \varepsilon}{\partial c_j} = 0$$

linear variation!

We are seeking to minimize $\varepsilon$ with respect to each $c_j$. Find the global minimum of the $\varepsilon(c_1,c_2,\ldots,c_N)$ hypersurface.

The only terms in the sums that survive $\frac{\partial}{\partial c_j}$ are those that include $c_j$. 
These are all of the surviving terms (i.e. those that include j). Each j term appears twice in both sums, once as a bra and once as a ket.

We get one such equation for each j (same set of unknown \{c_n\}).

N linear homogeneous equations in N unknown \( c_n \)’s

Non trivial \( \{c_n\} \) only if \( |\textbf{H} - \varepsilon \textbf{S}| = 0 \)

(Not the same simple form as \( |\textbf{H} - \textbf{1E}| = 0 \), but we can deal with this extra computational complexity.)

The result is N special values of \( \varepsilon \) that satisfy this equation.

CTDL show: all N \( \varepsilon \)-values are upper bounds to the lowest-N \( E_n \)'s and all obtained \( \{ \phi_n \}'s are orthogonal! (orthogonal provided they belong to different values of \( E_n \))

**How to solve \( |\textbf{H} - \varepsilon \textbf{S}| = 0 \):**

series of transformations on both \( \textbf{S} \) and \( \textbf{H} \)

1. Diagonalize \( \textbf{S} \)

   \( \textbf{U}^{\dagger} \textbf{S} \textbf{U} = \tilde{\textbf{S}} \quad \tilde{S}_{ij} = s_i \delta_{ij} \)

   (orthogonalize \( \{ \chi \} \) basis)

2. Normalize \( \tilde{\textbf{S}} \)

   \[
   \left( \tilde{\textbf{S}} \right)^{-1/2} \tilde{\textbf{S}} \left( \tilde{\textbf{S}} \right)^{-1/2} = 1 = \tilde{\textbf{S}} = \textbf{T}^{\dagger} \textbf{S} \textbf{T}
   \]

   where \( \textbf{T} = \textbf{U} \tilde{\textbf{S}}^{-1/2} \quad \left( \tilde{\textbf{S}}^{-1/2} \right)^{\dagger} = \tilde{\textbf{S}}^{-1/2} = \left( \begin{array}{ccc} s_1^{-1/2} & 0 & 0 \\ 0 & s_2^{-1/2} & 0 \\ 0 & 0 & \ddots \end{array} \right) \)

   This is not an orthogonal transformation of \( \tilde{\textbf{S}} \), but it does not destroy orthogonality because each eigenfunction of \( \textbf{S} \) is only being multiplied by a constant.
3. Transform $H$ to orthonormalized basis set

$$H = \tilde{S}^{-1/2} \left( U^T H U \right) \tilde{S}^{-1/2}$$

U diagonalizes $S$ not $H$

Obtain a new secular equation:

$$| \tilde{H} - \epsilon \tilde{S} | = 0 \quad \text{but} \quad \tilde{S} = 1$$

thus $| \tilde{H} - \epsilon I | = 0$ by which $\tilde{H}$ is diagonalized by the usual procedure.

5. Combine Linear and Nonlinear Variation

typically done in ab initio electronic structure calculations

Basis set: $\chi_n(\xi_n r)$

$\psi = \sum_n c_n \chi_n(\xi_n r)$

linear variation in $\{ \chi_n \}$, but where $\xi_n$ is a radial scale factor, one for each $\chi_n$

get $S_{nn'}(\xi_n, \xi_n'), H_{nn'}(\xi_n, \xi_n')$

nonlinear variation

0. pick arbitrary set of $\{ \xi_i \}$

1. calculate all $H_{ij}(\xi_i, \xi_j)$ and $S_{ij}(\xi_i, \xi_j)$

2. Solve $| H - \epsilon S | = 0$

   a. $S \rightarrow \tilde{S}$ diagonalize $S$ (orthogonalize)
   b. $(\tilde{S})^{-1/2}$ (normalize)
   c. $H \rightarrow \tilde{H}$
   d. diagonalize $\tilde{H}$

and now the nonlinear variation begins — find global minimum of $\epsilon_{\text{lowest}}$ with respect to $\xi_i$. 
3. change $\xi_i$ from $\xi_i^{(0)} \rightarrow \xi_i^{(1)} = \xi_i^{(0)} + \delta$
4. recalculate all integrals in $H$ and $S$ involving $\chi_i$
5. Solve $|H-\varepsilon S| = 0$ to obtain a new set of $\{\varepsilon_i\}$.
   Pick lowest $\varepsilon_i$.
6. calculate $\frac{\partial \varepsilon_{\text{lowest}}}{\partial \xi_i} = \frac{\varepsilon_{\text{old}} - \varepsilon_{\text{new}}}{\xi_i^{(0)} - \xi_i^{(1)}}$
7. repeat #3 – 6 for each $\xi_i$ (always looking only at lowest $\varepsilon_i$)
   This defines a gradient on a multidimensional lowest-$\varepsilon$($\xi_1, \ldots, \xi_N$) surface.
   We seek the minimum of this hypersurface. Take a step in direction of
   steepest descent by an amount determined by $| \partial \varepsilon / \partial \xi_{\text{steepest}} |$ (small
   slope, small step; large slope, large step).

   This completes 1st iteration. All values of $\{\xi_i\}$ are improved.

8. Return to #3, iterate #3-7 until convergence is obtained.

Nonlinear variations are much slower than linear variations.
Typically use ENORMOUS LINEAR $\{\chi\}$ basis set.
Contract this basis set by optimizing the nonlinear parameters (exponential scale
factors) in a SMALL BASIS SET to match the lowest $\{\phi\}'s$ that had
initially been expressed in large basis set. Least squares fitting of
wavefunctions.

* Large linear variation to get the linear combination of $\phi$'s that have the lowest
  energies;
* define a set of functions $\{\psi_i(\xi_i x)\}$ that contain non-linear scale parameters;
* perform a least squares fit of the $\{\xi\}$ to match the lowest few energy states from
  the linear variation;
* ortho-normalize the small set of $\psi_i(\xi_i x)$ functions and use them in a linear
  variation, thereby replacing the many-component functions from the massive
  linear variation by few-component functions (contracted basis set) from the
  hybrid linear variation.
6. Alternative Strategies

* rigorous variational minimization of $E_{\text{lowest}}$: “ab initio”
* constrain variational functions to be orthogonal to specific subset of variationally optimized functions
  e.g. orthogonal to ground state – to get variational convergence.
  Applied to higher energy members of specific symmetry class
  or orthogonal to core: frozen-core approximation.
  “Pseudopotentials” (use some observed energy levels to determine $Z^{\text{eff}}(r)$ of frozen core)

* least squares fitting to truncated $H$ i.e. $H^{\text{eff}}$

   minimize differences between a set of measured energy levels (or other properties) and a set of computed variational eigen-energies (or other properties computed from variational wavefunctions).

$$\{\text{observed } E_n\} \leftrightarrow \{\text{parameters in } H^{\text{eff}}\}$$
  molecular constants
  ↓
  experimental $\psi$’s in finite
  variational basis set

* semi-empirical model

   replace exact $\hat{H}$ by a grossly simplified form and restrict basis set to a simple form too. Then adjust parameters in $H$ to match some observed pattern of energy splittings. Confirm by using parameters to predict unobserved properties. Use values of fit parameters to build insight.

Never-ending battle between accuracy and insight!