Variational method for excited states

The Ritz theorem considered in the previous lecture can be generalized to excited states. In the generalized formulation it reads as follows:

The expectation value of the Hamiltonian is stationary in the neighborhood of its discrete eigenvalues.

Proof: Let us calculate an increment δE of the mean value $E\langle\psi|\psi\rangle = \langle\psi|H|\psi\rangle$ when ψ is changed to $\psi+\delta\psi$, using where $\delta\psi$ is assumed to be infinitely small. Neglecting terms of higher order than a linear term in $\delta\psi$, we obtain

$$\langle\psi|\psi\rangle\delta E = \delta\langle\psi|H|\psi\rangle - E\delta\langle\psi|\psi\rangle = \langle\delta\psi|H-E|\psi\rangle + \langle\psi|H-E|\delta\psi\rangle$$

The mean value E is stationary if $\delta E = 0$ for any infinitesimal $\delta\psi$, that is, if

$$\langle\delta\psi|H-E|\psi\rangle + \langle\psi|H-E|\delta\psi\rangle = 0$$

If $\delta\psi$ is chosen to be $\varepsilon(H-E)\psi$, where ε is an infinitesimal real number, the above equation implies that the norm of the function $(H-E)\psi$ is zero, and thus $(H-E)\psi$ must itself be a null function, or $H\psi = E\psi$. Therefore the mean value E is stationary if and only if the state ψ from which E is calculated is an eigenstate of H and the stationary value E is the corresponding eigenvalue of H. End of proof.

This generalized Ritz theorem allows an approximate determination of the eigenvalues of the Hamiltonian. If $E(\alpha)$, where α are some parameters of the trial function, has several extrema, they are the approximate values of some of the energies E_n.

In practical applications the most common way to approximate the wave function of a system is to form a linear combination of a finite number of independent functions $\psi(\alpha)$:

$$\psi = \sum_{i=1}^{N} c_i \psi_i(\alpha_i)$$

Functions $\psi_i(\alpha_i)$ do not have to be mutually orthogonal. In fact, it is quite common when they are not. They can always be made orthogonal, if necessary, by the Gram-Schmidt orthogonalization procedure. The variational method then reduces to the eigenvalue problem of the Hamiltonian inside the state space spanned by $\{\psi_i(\alpha_i)\}$. The restriction of the eigenvalue problem H to the subspace $\{\psi_i(\alpha_i)\}$ greatly simplifies the problem. Instead of finding the minimum ~~over~~ in an infinitely dimensional space $(\delta\psi)$ we need to minimize with respect to $\{c_i\}$ only (and also, possibly with respect to $\{\alpha_i\}$). Let us see how it works. For each i we want the above linear combination to satisfy the extremum condition:

$$\frac{\partial}{\partial c_i} \frac{\langle\psi|H|\psi\rangle}{\langle\psi|\psi\rangle} \qquad (1)$$

(For simplicity we will assume that all ψ_i's and c_i are real, although the method also works for complex functions ψ_i). We require the approximate function to remain normalized:

$$\langle\psi|\psi\rangle = 1 \qquad (2)$$

We may, in fact, satisfy both equations (1) and (2) by introducing a new quantity:

$$K = \langle\psi|H|\psi\rangle - \lambda\left[\langle\psi|\psi\rangle - 1\right]$$

and extend the minimization property to include the extra parameter λ:

$$\frac{\partial K}{\partial c_i} = \frac{\partial K}{\partial \lambda} = 0$$

which immediately yields

$$\langle \psi | \psi \rangle - 1 = 0$$

and proves that minimizing K with respect to all of the parameters involved ($\{c_i\}$ and λ) satisfies both of the conditions that we are aiming to satisfy. This sort of approach to solve the constrained minimization problem is known as the Lagrange multiplier method. Inserting our linear combination into $\frac{\partial}{\partial c_m}\langle \psi | H | \psi \rangle$ yields

$$\frac{\partial}{\partial c_m}\left[\left\langle \sum_i c_i \psi_i \middle| H \middle| \sum_j c_j \psi_j \right\rangle - \lambda\left(\left\langle \sum_i c_i \psi_i \middle| \sum_j c_j \psi_j \right\rangle - 1\right)\right]$$

$$= \frac{\partial}{\partial c_m} \sum_{ij} c_i c_j \left[\langle \psi_i | H | \psi_j \rangle - \lambda \langle \psi_i | \psi_j \rangle\right] =$$

$$= \sum_j c_j \left[\langle \psi_m | H | \psi_j \rangle - \lambda \langle \psi_m | \psi_j \rangle\right] = 0$$

or

$$\sum_j \langle \psi_k | H | \psi_j \rangle c_j = \lambda \sum_j \langle \psi_k | \psi_j \rangle c_k$$

We can recognize immediately the generalized eigenvalue problem:

$$H \cdot c = \lambda S c$$

where H and S are the matrix representations (in subspace $\{\psi_i\}$) and overlap of the Hamiltonian operators:

$$H_{nj} = \langle \psi_n | H | \psi_j \rangle \qquad S_{nj} = \langle \psi_n | \psi_j \rangle$$

The relationship between the eigenvalues of the truncated problem $Hc = \lambda Sc$ and the eigenvalues of the full Hamiltonian is elucidated by the Mini-Max theorem:

Let H be a hermitian operator with discrete eigenvalues $E_1 \leq E_2 \leq E_3 \leq \ldots$ Let $\lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \lambda_4 \leq \ldots \leq \lambda_N$ be the eigenvalues of H restricted to the subspace $\mathcal{V}_N$ of a linearly independent set of N functions $\{\psi_i\}$ of $\mathcal{V}_N$. Then

$$E_1 \leq \lambda_1, \quad E_2 \leq \lambda_2, \quad E_3 \leq \lambda_3, \ldots, \quad E_N \leq \lambda_N$$

Proof: Let ω_N be the subspace spanned by the orthonormal eigenstates $\phi_1, \ldots \phi_N$ of the operator H. We will first show that there is at least one normalized function in $\mathcal{V}_N$ with the property

$$\langle \psi | H | \psi \rangle \geq E_N$$

Consider the projected function $P\psi$ for any normalized function ψ in $\mathcal{V}_N$, where

$$P = \sum_{i=1}^{N} |\phi_i\rangle\langle\phi_i|$$

There are two possibilities: $P\psi_0 = 0$

1) there exists a function ψ_0 in $\mathcal{V}_N$ such that
2) $P\psi \neq 0$ for all functions ψ in $\mathcal{V}_N$

In case (1) ψ_0 is a linear combination of $\phi_{N+1}, \phi_{N+2}, \ldots$ and hence the normalized ψ_0 has the mean value

$$\langle \psi_0 | H | \psi_0 \rangle \geq E_{N+1} \geq E_N.$$

In case (2) any two different normalized functions ψ_1 and ψ_2 in $\mathcal{V}_N$ are projected to different functions $P\psi_1$ and $P\psi_2$ in ω_N because otherwise $P(\psi_1 - \psi_2) = 0$, which contradicts the assumption. Namely, any two different normalized functions in $\mathcal{V}_N$ are projected by P to different functions in ω_N. As both $\mathcal{V}_N$ and ω_N have the

same dimension, it follows that there must exist a normalized function ψ in V_N with the property $P\psi = a\phi_k$ $(a \neq 0)$. Then ψ can be expressed as $a\phi_k + b\phi$ where ϕ is a normalized function given as a linear combination of $\phi_{N+1}, \phi_{N+2}, \ldots$ and $|a|^2 + |b|^2 = 1$. This function has the mean value

$$\langle\psi|H|\psi\rangle = |a|^2 E_N + |b|^2\langle\phi|H|\phi\rangle = E_N + |b|^2(\langle\phi|H|\phi\rangle - E_N) \geq E_N$$

Because λ_N is the largest eigenvalue of H restricted to V_N we have

$$\lambda_N \geq \langle\psi|H|\psi\rangle \geq E_N$$

Next we can just define the $N-1$ dimensional state space V_{N-1} to be the set of all those functions in V_N which are orthogonal to the eigenvector belonging to λ_N and by induction show

$$\lambda_{N-1} \geq E_{N-1} \quad \text{and so on.}$$

The Mini-Max theorem provides a tool to obtain upper bounds to the energies of excited states and is widely used in variational calculations