Variational method for excited states

The Ritz theorem considered in the previous lecture can also be generalized for excited states The generalized form reads as follows: The expectation value of the Hamiltonian is stationary in the neighborhood of its discrete eigenvalues

Proof:

Let us calculate the increment δE of the mean value using the relation $E\langle\psi|\psi\rangle = \langle\psi|H|\psi\rangle$ when ψ is changed from ψ to $\psi+\delta\psi$. Here $\delta\psi$ is assumed to be an infinitely small variation. Neglecting terms of orders higher than linear in $\delta\psi$ we obtain:

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

or

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

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

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 be a null function, or it in other words it must satisfy the equation $H\psi = E\psi$ Therefore the mean value E is stationary if (and only if) ψ is an eigenstate of H and the stationary value E is the corresponding eigenvalue of H.

End of proof.

The generalized Ritz theorem allows to determine (approximately) the ground and excited energies of a quantum system. If $E(\{\alpha\})$, where $\{\alpha\}$ are some parameters of the trial function, has several extrema, they are approximate values of certain energy levels E_n.

In practical applications the wave function is commonly represented as a linear combination of a finite number of independent functions $\phi_i(\alpha_i)$ [Rayleigh-Ritz method]

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

Function $\phi_i(\alpha_i)$ do not have to be mutually orthogonal. In fact, in practical applications they are often not orthonormal. However, this set of N functions can always be made orthonormal, if necessary, by means of the Gram-Schmidt orthogonalization procedure. The use of the variational method then amounts to solving an eigenvalue problem in a subspace spanned by $\{\phi_i(\alpha_i)\}$. The restriction to a finite-dimensional subspace greatly simplifies the problem. Instead of finding the solution (ψ_{exact}) in an infinitely-dimensional Hilbert space, we can find a minimum (with respect to $\{c_i\}$) in N dimensions. (it may be useful to find a minimum with respect to $\{\alpha_i\}$, if feasible).

For simplicity let us assume that all ϕ_i's and c_i's are real (although the method also works for complex functions ϕ_i and coefficients c_i).

We require that the approximate wave function to remain normalized regardless of what happens to c_i's :

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

This introduces a constraint into our minimization of $\langle \psi | H | \psi \rangle$ with respect to c_i's. The constrained minimization can be handled by means of the Lagrange multiplier method (see appendix in order to refresh your memory about this method). We form a new quantity:

$$F(c_1, \ldots c_N; \lambda) = \langle \psi | H | \psi \rangle - \lambda \left(\langle \psi | \psi \rangle - 1 \right)$$

and find an unconstrained minimum with respect to both $\{c_i\}$ and λ:

$$\frac{\partial F}{\partial c_k} = \frac{\partial F}{\partial \lambda} = 0 \qquad k = 1, \ldots, n$$

which immediately yields $\langle \psi | \psi \rangle - 1 = 0$

Inserting the liner combination $\sum_{i=1}^{N} c_i \phi_i$ in place of ψ gives

$$\frac{\partial}{\partial c_k} \left[\langle \sum_i{}' c_i \phi_i | H | \sum_j c_j \phi_j \rangle - \lambda \left(\langle \sum_i c_i \phi_i | \sum_j c_j \phi_j \rangle - 1 \right) \right] =$$

$$= \frac{\partial}{\partial c_k} \sum_{ij}{}' c_i c_j \left[\langle \phi_i | H | \phi_j \rangle - \lambda \langle \phi_i | \phi_j \rangle \right]$$

Let us denote $\langle \phi_i | H | \phi_j \rangle = H_{ij}$ and $\langle \phi_i | \phi_j \rangle = S_{ij}$

Then taking into account that $\frac{\partial c_i c_j}{\partial c_k} = \delta_{ik} c_j + \delta_{jk} c_i$,

$$\frac{\partial F}{\partial c_k} = \sum_j{}' c_j \left[\langle \phi_k | H | \phi_j \rangle - \lambda \langle \phi_k | \phi_j \rangle \right] + \sum_i c_i \left[\langle \phi_k | H | \phi_i \rangle - \lambda \langle \phi_k | \phi_i \rangle \right]$$

or

$$\sum_j H_{kj} c_j = \lambda \sum_j S_{kj} c_j$$

We can now recognize the generalized eigenvalue problem:

$$H\vec{c} = \lambda S \vec{c}$$

where H and S are $N \times N$ matrices (matrix representations of the Hamiltonian and overlap in subspace $\{\phi_i\}$) and $\vec{c}$ is a column vector

$$\vec{c} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_N \end{pmatrix}$$

There are N solutions of the generalized eigenvalue problem, that is there are N pairs $\lambda, \vec{c}$. We can label them with an index

$$\lambda_i \quad \vec{c}^{(i)} \qquad i = 1, \ldots N$$

The relationship between the eigenvalues of the truncated problem $H\vec{c} = \lambda S \vec{c}$ and the eigenvalues E_i of the original Hamiltonian is elucidated by the Mini-Max theorem (proved in appendix), which states that if $\{\phi_i\}$ is a set of N linearly independent functions then

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

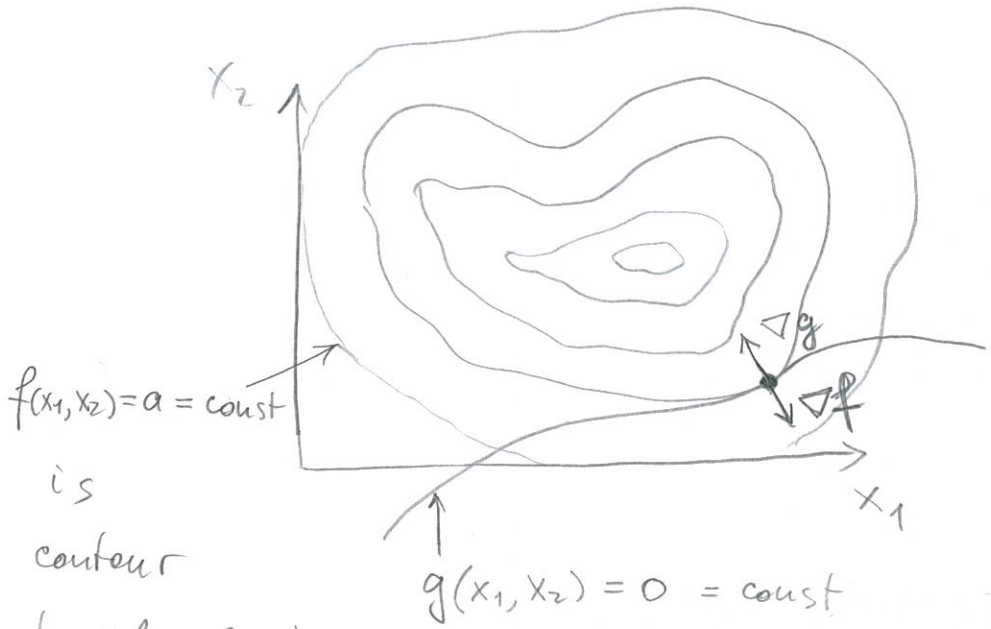
The Mini-Max theorem provides a foundation for computing upper bounds to the exact energies of excited states.

Appendix: Method of Lagrange multipliers

The method of Lagrange multipliers gives a recipe of finding a (local) maximum or minimum of a function of several variables $f(x_1, \ldots, x_n)$ subject to equality constraints $g_1(x_1, \ldots x_n) = 0, \ldots, g_m(x_1, \ldots x_n) = 0$

For simplicity let us assume $n = 2$ and $m = 1$, so that the task is

$$\begin{cases} \min\limits_{\{x_1, x_2\}} f(x_1, x_2) \\ g(x_1, x_2) = 0 \end{cases}$$



The gradient of f is perpendicular to its contour lines. The gradient of g is also perpendicular to its contour lines. $g(x_1, x_2) = 0$ is a contour line. Hence for the contour lines of f and g to be parallel (the conditional minima are at such points where the contour lines are parallel) the following must be true

$$\nabla f = \lambda \nabla g \quad \text{where } \lambda \text{ is a constant (could be positive or negative)}$$

It can be seen that if we rather than $f(x)$ minimize the following function

$$F = f - \lambda g$$

with respect to x_1, x_2 as well as with respect to λ the above condition ($\nabla f = \lambda \nabla g$) is automatically met

$$\left.\begin{aligned} \frac{\partial F}{\partial x_1} &= 0 \\ \frac{\partial F}{\partial x_2} &= 0 \\ \frac{\partial F}{\partial \lambda} &= 0 \end{aligned}\right\} \Rightarrow \begin{cases} \frac{\partial f}{\partial x_1} = \lambda \frac{\partial g}{\partial x_1} \\ \frac{\partial f}{\partial x_2} = \lambda \frac{\partial g}{\partial x_2} \\ g = 0 \end{cases}$$

at the same time at the point we are looking for $g(x_1, x_2) = 0$.

Hence to solve our problem – finding a minimum of a function subject to constraint we can find an unconstrained minimum of $F(x_1, x_2, \lambda)$ with respect to x_1, x_2, and λ. Quantity λ is called the Lagrange multiplier.

In the case of multiple (m) constraints we will have m Lagrange multipliers $\lambda_1, \dots, \lambda_m$ and we will need to find an unconstrained minimum of

$$F(x_1, \dots x_n; \lambda_1, \dots \lambda_m) = f(x_1, \dots x_n) - \sum_{i=1}^{m} \lambda_i g_i(x_1, \dots x_n)$$

Appendix: The Mini-Max Theorem

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

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

Proof: Let ω_N be the subspace spanned by the orthonormal eigenstates $\psi_1, \ldots, \psi_N$ of operator H. We will first show that there is at least one normalized function in ν_N with the property

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

Consider the projected function $P\phi$ for any normalized function ϕ in ν_N, where

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

There are two possibilities:

1) there exists a function ϕ_0 in ν_N such that $P\phi_0 = 0$

2) $P\phi \neq 0$ for all functions ϕ in ν_N

In case 1) ϕ_0 is a linear combination of $\phi_{N+1}, \phi_{N+2}, \ldots$ and hence the normalized ϕ_0 has the mean value

$$\langle \phi_0 | H | \phi_0 \rangle \geq E_{N+1} \geq E_N$$

In case 2) any two different normalized functions ϕ_i and ϕ_j in ν_N are projected to different functions $P\phi_i$ and $P\phi_j$ in ω_N because otherwise $P(\phi_i - \phi_j) = 0$,

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 ϕ in $\mathcal{V}_N$ with the property $P\phi = a\psi_K$ $(a \neq 0)$. Then ϕ can be expressed as $a\psi_K + b\psi$ where ψ is a normalized function given as a linear combination of $\psi_{N+1}, \psi_{N+2}, \ldots$ and $|a|^2 + |b|^2 = 1$. This function has the mean value

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

Because λ_N is the largest eigenvalue of H restricted to $\mathcal{V}_N$ we have

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

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

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