THEORETICAL CHEMISTRY INSTITUTE

THE UNIVERSITY OF WISCONSIN

THE VARIATIONAL PRINCIPLE AND THE VARIATIONAL METHOD I

by

Saul T. Epstein

WIS-TCI-336 20 February 1969

MADISON, WISCONSIN

The Variational Principle and the Variational Method I[*]

by

Saul T. Epstein

Theoretical Chemistry Institute and
Physics Department
University of Wisconsin

I WOULD GREATLY APPRECIATE YOUR CRITICISMS, COMMENTS, AND SUGGESTIONS CONCERNING THE CONTENT AND ORGANIZATION OF THESE LECTURES.

S.T.E.

---

[*] Research supported by National Aeronautics and Space Administration Grant NGL 50-002-001.

I. Variational Principle

We will be concerned with the internal dynamics of a bound system, the internal Hamiltonian being denoted by H . Our starting point is the observation that whatever the state of the system, the average value of the energy cannot be less than the energy of the ground state. Thus given any wave function $\tilde{\Psi}$ for which the requisite integrals exist (henceforth we will refer to $\tilde{\Psi}$ as a "trial function") we then have that the average energy is an upper bound to the energy of the ground state, thus

$$\tilde{E} \equiv (\tilde{\Psi}, H\tilde{\Psi})/(\tilde{\Psi},\tilde{\Psi}) \geqslant E_G \qquad \text{(I-1)}$$

Problem: Give a formal proof of (I-1). Hint: expand $\tilde{\Psi}$ in terms of the eigenfunctions of H .

If H contains a kinetic energy operator and one is working in configuration space, then the existence of $(\Psi, H\tilde{\Psi})$ requires that $\tilde{\Psi}$ be twice differentiable. However this condition can be relaxed. If one uses $-(\nabla\tilde{\Psi}, \nabla\tilde{\Psi})$ instead of $(\tilde{\Psi}, \nabla^2\tilde{\Psi})$ then one can show that one still has an upper bound even if $\tilde{\Psi}$ is only once differentiable (see Courant Hiblert "Methods of Mathematical Physics" Vol. I, bottom p. 457). Also even if $\tilde{\Psi}$ is twice differentiable the $(\nabla\tilde{\Psi},\nabla\tilde{\Psi})$ form is often more convenient numerically. However, we will continue to use the form (I-1) since it is much easier to deal with formally.

Further we will now show that if we consider $\tilde{E}$ as a functional of $\tilde{\Psi}$, then the various bound state eigenvalues of H are stationary points for $\tilde{E}$. In particular then we will have shown that not only is $\tilde{E}$ an upper bound to E_G but that $\tilde{E} = E_G$ is actually a minimum point for $\tilde{E}$ is a functional of $\tilde{\Psi}$.

To prove all this we write

$$\tilde{\Psi} = \Psi + \tilde{\Delta} \qquad (1\text{-}2)$$

where Ψ is a bound state eigenfunction of H belonging to eigenvalue E :

$$(H - E)\Psi = 0 \qquad (1\text{-}3)$$

and where $\tilde{\Delta}$ is orthogonal to Ψ:

$$(\Psi, \tilde{\Delta}) = 0 \qquad (1\text{-}4)$$

(since we have not specified the normalization of either Ψ or $\tilde{\Psi}$, this latter requirement represents no loss of generality). Thus $\tilde{\Delta}$ tells us the amount by which $\tilde{\Psi}$ differs from the eigenfunction Ψ.

If we insert this expression for $\tilde{\Psi}$ into the formula for $\tilde{E}$, and use (3) and the hermiticity of H, then we find successively

$$\tilde{E} = (\Psi + \tilde{\Delta}, H(\Psi + \tilde{\Delta}))/(\tilde{\Psi}, \tilde{\Psi})$$

$$= \{E[(\Psi,\Psi) + (\tilde{\Delta},\Psi) + (\Psi,\tilde{\Delta})] + (\tilde{\Delta}, H\tilde{\Delta})\}/(\tilde{\Psi},\tilde{\Psi})$$

$$= E + (\tilde{\Delta}, (H - E)\tilde{\Delta})/(\tilde{\Psi},\tilde{\Psi}) \qquad (1\text{-}5)$$

which expresses in symbols what we said earlier in words: When $\tilde{\Delta}$ is small so that $\tilde{\Psi}$ is near Ψ, then $\tilde{E}$, as a functional of $\tilde{\Psi}$, differs from E by terms which are at least of order $\tilde{\Delta}^2$ and this is of course the mark of a stationary point, i.e., there are no terms linear in $\tilde{\Delta}$. Our initial observation that $\tilde{E} \geq E_G$ (and hence, in light of the present result, that $\tilde{E} = E_G$ is a minimum point) can also be read from (I-5). Namely the smallest eigenvalue of $H - E_G$ is evidently zero whence if $\tilde{\Delta}$ is orthogonal to ψ_G and if E_G is non degenerate then

$$(\tilde{\Delta}, (H - E_G)\tilde{\Delta}) > 0$$

which proves the point.

Problem: What can we say if E_{GROUND} is degenerate?

Problem: Show that the higher eigenvalues are neither maxima nor minima, i.e., they are simply stationary points.

Thus we have shown that the eigenvalues of H are stationary points of $\tilde{E}$ as a functional of $\tilde{\psi}$. In Sec. IV we will prove the converse: If E is a stationary point for $\tilde{E}$ then E is an eigenvalue and the corresponding $\tilde{\psi}$ is an eigenfunction. Together these results constitute a statement of the Variational Principle -- "variational" because in order to test for stationarity one must vary $\tilde{\psi}$ about the suspected value.

II. The Variational Method

In theory the variational principle provides an alternative but equivalent method to the direct solution of the Schroedinger equation for determining the ψ and E. However in practice one usually can't carry it to completion -- one can't examine all $\tilde{\psi}$ and look for stationary points. However, what one can do (and this constitutes a statement of the <u>variational method</u> or <u>variational approximation</u>) is to try to <u>approximate</u> the ψ and E by examining a restricted set of trial functions and determining those $\tilde{\psi}$ which yield $\tilde{E}$ which are stationary for variations of $\tilde{\psi}$ <u>within</u> the set, i.e., which are stationary with respect to the <u>restricted</u> class of variations. We will denote the $\tilde{\psi}$ which yield stationary $\tilde{E}$ in this restricted sense by $\hat{\psi}$ (possibly with a subscript) and call them optimal trial functions. The corresponding $\tilde{E}$ will be denoted by $\hat{E}$. Evidently it is then natural to put forward the $\hat{\psi}$ and $\hat{E}$ as the "best approximations" to the ψ and E to be found witin the set since in particular one would expect that as one enlarged the set of trial functions the $\hat{\psi}$ and $\hat{E}$ would steadily become better and better approximations to the ψ and E (of course it is ALSO not excluded that one may, by chance, already have an exact eigenfunction in a restricted set). However, note that from what we have learned so far, this expectation is really justified <u>only</u> for the <u>energy</u> of the <u>ground</u> state. Namely we know what whatever the set of trial functions, the lowest $\hat{E}$, call it $\hat{E}_G$, will satisfy

$$\hat{E}_G \geq E_G$$

and since enlarging the set of trial functions cannot raise $\hat{E}_G$, and will usually lower it, it follows that for the ground state, the approximation to the energy will improve steadily.

It is more difficult to make statements about $\hat{\psi}_G$, in part because how good an approximation $\hat{\psi}_G$ is to ψ_G depends on how one chooses to compare them. We will therefore not pursue this question further, except for a few related remarks in what follows.

For the higher states however, all that one can say in general is that as one enlarges the set, the $\hat{E}$ become "more stationary", but whether or not they become numerically more accurate one cannot in general predict. This is not to say that there are not techniques for ensuring that variational calculations give energy bounds for excited states so that we have more control of the situation; there are, and we will mention one IMMEDIATELY, and discuss another in Sec. VI. It is [illegible] to say, however, that they will require special precautions. For example suppose that one knows that each member of the set of trial functions in orthogonal to all the eigenfunctions with eigenvalues less than E'. Then clearly, instead of (1) one can write

$$\tilde{E} \geq E'$$

and thus have a variational bound on a higher state. However, in practice this is not very useful because one usually doesn't know the eigenfunctions of the lower states, and hence can't be sure that the $\tilde{\Psi}$ are orthogonal to them. An *important exception* occurs when the eigenfunctions of H can be classified in an *à priori* way according to some symmetry property. Then if we restrict the set of $\tilde{\Psi}$ to belong to a particular symmetry type we can say that *$\hat{E}$ will be an upper bound to the lowest eigenvalue associated with that symmetry*, which may of course not be the lowest eigenvalue of H. As an example, assuming a spin independent H, symmetry alone can ensure us of having a bound for the $1s2s\,{}^3S$ state of Helium since there is no triplet state below it. On the other hand to get a bound for the $1s2s\,{}^1S$ state will require a different technique (see Sec. VI) since the $(1s)^2\,{}^1S$ ground state lies below.

For the moment then we will confine attention to the ground state, or lowest state of a given symmetry (and drop the label G). However before continuing we want to make one further important general remark about the range of applicability of the variational method. In our discussion thus far, as in our discussions of perturbation theory, we have had in mind that H is an atomic or molecular Hamiltonian. However as far as the mathematics is concerned H could be a finite dimensional Hermitian matrix and correspondingly $\hat{\Psi}$ a finite dimensional column vector. Thus the variational method (and perturbation theory) can be and are used to approximate the solution of finite matrix problems. (Of course there are also many very efficient direct methods available for solving finite matrix problems of fairly large size, especially with the help of a computer.) Further, as we will see, starting in Sec. V, such problems in turn often arise in applying the variational method to the atomic or molecular Hamiltonian. Even more generally, one can envisage the use of the variational method to approximate the solution of a mathematical problem, this problem having arisen in the course of using the variational method to approximate the solution of another problem, etc. We will mention examples of this sort in Sec. VII.

The matrix problems often have the apparently more general form

$$\mathcal{H} C = \lambda \mathcal{S} C$$

where λ is the eigenvalue, C the eigenvector, $\mathcal{H}$ a Hermitian matrix and $\mathcal{S}$ a positive definite Hermitian matrix. However, they can be reduced to standard eigenvalue form by transforming with $\mathcal{S}^{1/2}$ (which exist since $\mathcal{S}$ is Hermitian and positive definite) according to $\mathcal{H} = \mathcal{S}^{1/2} \mathcal{H}' \mathcal{S}^{1/2}$ $C = \mathcal{S}^{-1/2} C'$ to yield the equation

$$\mathcal{H}'c' = \lambda c'$$

> which is of the form which we have been discussing. However, note that by reversing the prodecure, i.e., by writing $\mathcal{H}' = \mathcal{S}^{-1/2}\mathcal{H}\mathcal{S}^{-1/2}$, $c' = \mathcal{S}^{1/2}c$ one can express all quantities of interest, and in particular the variatonal principle, in terms of $\mathcal{H}$ and $\mathcal{S}$, i.e., one does not need to calculate $\mathcal{S}^{1/2}$ explicitly. <u>Problem</u>: Write the variational principle in terms of $\mathcal{H}$ and $\mathcal{S}$.

For the ground state then the variational method for finding approximations to ψ and E consists in examining a limited set of trial functions and determining that $\tilde{\psi}$, to be denoted by $\hat{\psi}$, which minimizes $\tilde{E}$, the minimum value of $\tilde{E}$ being denoted by $\hat{E}$. $\hat{\psi}$ and $\hat{E}$ are then taken as the "best" approximations to ψ and E to be found within the set.

> Note that for the ground state one really doesn't have to envisage a variational process, i.e., a continuous process. One could base the procedure wholly on (1) by examining a discrete set of $\tilde{\psi}$, and selecting the one which gives the smallest value of $\tilde{E}$. In practice, however, one usually uses continuous sets.

Now clearly from the minimal property (1), $\hat{E}$ is certainly the best approximation to E that we can find among the $\tilde{E}$ provided by the set. However, with respect to $\hat{\psi}$ the situation is not so clear. There will almost certainly be other members of the set which are superior to $\hat{\psi}$ in other ways, for example give more accurate expectation values for quantities other then the energy, or giving a smaller value for the "energy variance" $(\tilde{\psi}, (H-\tilde{E})^2\tilde{\psi})$ etc. (A recent reference: Keaveny and Christoffersen, J. Chem. Phys. <u>50</u>

80 (1969).) Indeed one often says that since $\hat{\mathcal{E}} - \mathcal{E}$ is of second order in $\hat{\Delta}$ while $\hat{\Psi} - \Psi$ is (by definition) of first order, that energies are more accurate than wave functions, and as a rule of thumb this is probably true. However, as noted earlier, to make such a statement precise one must give some norm or norms by which one judges the accuracy with which $\hat{\Psi}$ approximates Ψ since $\hat{\Psi} - \Psi$ is, after all, a function, usually of many variables. Then having settled on a norm there is still "the question of coefficients". This is one says that Ax^2 is of higher order than x (and hence, in this theoretical sense, "smaller") whatever the value of the constant coefficient A , because there is a range of X around zero such that $|A|x^2 \leq |x|$ However this range depends on A , being $|x| \leq |A|^{-1}$ and practical case A may be such that though the value of X of inte seems small, still $|A|x^2 > |x|$. To be more specific suppose ide to measure the accuracy of $\hat{\Psi}$ by the single parameter

$$\xi \equiv \left[\frac{(\hat{\Psi} - \Psi, \hat{\Psi} - \Psi)}{(\hat{\Psi}, \hat{\Psi})} \right]^{1/2} = \left[\frac{(\delta, \hat{\Delta})}{(\hat{\Psi}, \hat{\Psi})} \right]^{1/2}$$

and suppose that $\hat{\Delta}$ is an excited state of H with energy $\mathcal{E}'$. Then

$$\frac{(\hat{\mathcal{E}} - \mathcal{E})}{\mathcal{E}} = \frac{(\mathcal{E}' - \mathcal{E})}{\mathcal{E}} \xi^2$$

which <u>is</u> of second order in ξ but which, depending on the numerical values of $\mathcal{E}'$, and $\mathcal{E}$, and ξ for the case in point, may well be larger than ξ . In summary then the concept of "order" is a mathematical one and it is not necessarily so, that quantities which involve first order errors are less accurate than those which involve second order errors. That the variational method yields a second order error for the energy is important, but equally as important is that it yields an *improvable bound* to E_G so that one has some feeling for what one is about in trying to make more and more complicated calculations involving larger and larger sets of trial functions, namely one is sure that one is therby obtaining a better and better approxmation to E_G.

III. The Variational Method -- More Details

Let us now examine the variational method in more detail. We start with the definition of $\tilde{E}$ written as

$$(\tilde{\psi}, (H-\tilde{E})\tilde{\psi}) = 0 \qquad \text{(III-1)}$$

Then given a $\tilde{\psi}$ from the set of trial functions we consider a neighboring function which, to *first order* in the change of parameters and/or functions which label the various $\tilde{\psi}$, we write as

$$\tilde{\psi} + \delta\tilde{\psi}$$

Note that in general $\tilde{\psi}+\delta\tilde{\psi}$ is a member of the set *only through first order*, i.e., it is not itself actually a member of the set. Thus if the set consists of the functions $\tilde{A}\exp{-\tilde{\alpha}x}$ when $\tilde{A}$ and $\tilde{\alpha}$ are arbitrary parameters then $\tilde{\psi}+\delta\tilde{\psi}$ has the general form

$$\tilde{A}e^{-\tilde{\alpha}x} + \delta\tilde{A}e^{-\tilde{\alpha}x} - x\delta\tilde{\alpha}\tilde{A}e^{-\tilde{\alpha}x}$$

which, for $\delta\tilde{\alpha} \neq 0$, is *not* in the set. There are cases however -- linear spaces which we will discuss in detail later on -- in which $\tilde{\psi}+\delta\tilde{\psi}$ *is* in the set.

Then to first order, the change in $\tilde{E}$ is, from eq. (1), determined by

$$(\delta\tilde{\Psi}, (H-\tilde{E})\tilde{\Psi}) + (\tilde{\Psi}, (H-\tilde{E})\delta\tilde{\Psi}) - \delta\tilde{E}(\tilde{\Psi},\tilde{\Psi}) = 0 \qquad \text{(III-2)}$$

Now $\hat{\Psi}$ is that $\tilde{\Psi}$ which minimizes $\tilde{E}$ within the set. Hence when $\tilde{\Psi} = \hat{\Psi}$ and $\tilde{E} = \hat{E}$ all possible first order variations of $\tilde{E}$ **allowed** <u>within the set</u> must vanish. Thus we have

$$(\delta\hat{\Psi}, (H-\hat{E})\hat{\Psi}) + (\hat{\Psi}, (H-\hat{E})\delta\hat{\Psi}) = 0 \qquad \text{(III-3)}$$

which together with

$$(\hat{\Psi}, (H-\hat{E})\hat{\Psi}) = 0 \qquad \text{(III-4)}$$

are the equations to determine $\hat{\Psi}$ and $\hat{E}$. They constitute the <u>mathematical statement of the variational method</u>. It should be kept in mind, however, that in practice, having chosen a set of trial functions, one is often not able to solve equs. (3) and (4) exactly, i.e., analytically in closed form. Indeed one may well not be able to write them down exactly (!), usually because of the appearance of difficult integrals. In such cases one must resort to numerical methods, methods which are inevitably of finite accuracy, to write down and/or solve the equations. The potential interplay here between physical approximation and numerical approximation can clearly be quite important and interesting. However, we will ignore it in what follows. In particular when we discuss and derive properties the solutions $\hat{E}$ and $\hat{\Psi}$ of (3) and (4), we will be talking about the <u>exact</u> solutions and will not discuss the effects of possible numerical inaccuracies.

We will often write (3) as

$$\delta(\hat{\Psi}, (H-\hat{E})\hat{\Psi}) = 0 \qquad \text{(III-5)}$$

with the understanding, of course, that $\delta\hat{E} = 0$.

Note that the $\delta\hat{\Psi}$ which satisfy (3) form a real linear space. That is if $\delta_1\hat{\Psi}$ and $\delta_2\hat{\Psi}$ satisfy (3) then so do $A\delta_1\hat{\Psi} + B\delta_2\hat{\Psi}$ where A and B are any two real constants. On the other hand the set of trial functions is not usually a linear space though the case in which it is of great practical importance and is discussed in more detail in several of the sections which follow. Thus consider again the set $\tilde{A}\exp -\tilde{\alpha}x$. Then $3\exp -x$ and $5\exp -2x$ belong to the set but clearly $3\exp -x + 5\exp -2x$ does not.

An obvious procedure at this point would be to eliminate $\hat{E}$ from (3) by use of (4), solve the resultant equations for $\hat{\Psi}$ and then return to (4) to determine $\hat{E}$. Indeed in simple situations this is just what one does. Thus suppose that H is a one dimensional Hamiltonian and the $\tilde{\Psi}$ are as given in the previous note. Then what one would do is calculate $\tilde{E}$ as a function of $\tilde{\alpha}$ (evidently it doesn't depend on $\tilde{A}$), set $\frac{\partial\tilde{E}}{\partial\tilde{\alpha}} = 0$ $\left(\delta\tilde{E} = \frac{\partial\tilde{E}}{\partial\tilde{\alpha}}\delta\tilde{\alpha}\right)$, determine $\hat{\alpha}$ and then return to (4) to calculate $\hat{E}$.

<u>Problem</u>: With Gaussian trial functions $\tilde{A}\exp -\tilde{\alpha}r^2$ find the best approximation to the ground state of hydrogen.

However for theoretical purposes it is often convenient to retain the forms (3) and (4), and indeed it is often of practical use too (in particular for making contact with the Schroedinger equation and also for example in discussing linear trial functions). Thus we will leave

(3) and (4) as they are and not combine them directly in the way indicated. However, it is useful to note that they can often be profitable combined <u>indirectly</u>.

Namely suppose that the set of trial functions has no fixed overall scale. That is suppose that if $\tilde{\Psi}$ is a member then so is $A\tilde{\Psi}$ where A is an arbitrary real number. Then clearly among the variations which will have been found to satisfy (3) will be

$$\delta\hat{\Psi} = \delta A\,\hat{\Psi}$$

But if now we insert this into (3) we find (4). Thus if the set of trial functions has this property <u>then we need not refer to (4) explicitly</u>.

We have referred to (3) and (4) as the equations to be solved to determine $\hat{\Psi}$ and $\hat{E}$. A natural question is then, do they have solutions? Although clearly a question of great importance, it is also a question which is naturally of great mathematical complexity, and therefore we choose to ignore it. In general we will assume without further comment that solutions do exist, though in any particular case one must be prepared for the possibility that the contrary is true.

Now let us become even more specific. Let us consider the common situation in which the typical $\tilde{\Psi}$ is specified by a number of (real and independent) parameters $\tilde{\alpha}_1, \ldots, \tilde{\alpha}_M$. That is we can write

$$\tilde{\Psi} = \phi(\tilde{\alpha}_1 \ldots \tilde{\alpha}_M)$$

where ϕ specifies the functional form which we assume given (ϕ of course also depends on coordinates, spins, etc.). If now we vary

$\hat{a}_k$ then $\hat{\psi}$ is changed according to

$$\delta\hat{\psi} = \delta_k\hat{\psi} \equiv \frac{\partial\hat{\psi}}{\partial\hat{a}_k}\,\delta\hat{a}_k$$

(Note that a general $\delta\hat{\psi}$ is simply linear combination of the $\delta_k\hat{\psi}$.) Inserting this into (3) for each k value then yields the M equations

$$\left(\frac{\partial\hat{\psi}}{\partial\hat{a}_k}, (H-\hat{E})\hat{\psi}\right) + \left(\hat{\psi}, (H-\hat{E})\frac{\partial\hat{\psi}}{\partial\hat{a}_k}\right) = 0 \qquad \text{(III-6)}$$

which together with (4) are to be used to determine the $M+1$ unknowns $\hat{a}_k$ and $\hat{E}$. However what if the scale of the trial functions isn't fixed? Then as we have seen (4) is a redundant equation and so we seem to be left with M equations for $M+1$ unknown. However this is (Of course) not the case. Our assumption about the trial functions implies that they can be written in the form

$$\tilde{\psi} = \tilde{b}_1\,\phi'(\tilde{b}_2 \cdots \tilde{b}_M)$$

where the $\tilde{b}_k$ are real independent functions of the $\tilde{a}_k$. Namely with this form $\tilde{b}_1$ yields the arbitrary scale, and conversely the arbitrariness of the scale implies the existence of $\tilde{b}_1$.

Inserting (see below)

$$\delta\hat{\psi} = \bar{\delta}_1\hat{\psi} \equiv \frac{\partial\hat{\psi}}{\partial\hat{b}_1}\,\delta\hat{b}_1$$

into (3) of course then yields (4), and note that cancels out of (4). Then inserting

$$\delta\hat{\psi} = \bar{\delta}_k\hat{\psi} \equiv \frac{\partial\hat{\psi}}{\partial\hat{b}_k}\,\delta\hat{b}_k \qquad k=2,\cdots M$$

yields $M-1$ other equations which again don't involve $\hat{b}_1$. Thus we have consistency in that we have M equations for the M unknowns $\hat{E}$ and $\hat{b}_2 \cdots \hat{b}_M$. Since these M equations are equivalent to the original set (see below) it follows that the latter must also be consistent, first impressions to the contrary not withstanding. The value of $\hat{b}_1$, of course, remains arbitrary. That is with this sort of set of trial functions the overall scale of $\hat{\psi}$ is left arbitrary, just as the Schroedinger equation leaves the overall scale of ψ arbitrary.

The equivalence of the sets of equation follows from the observation that

$$\bar{\delta}_K \hat{\psi} = \frac{\partial \hat{\psi}}{\partial \hat{b}_K} \delta \hat{b}_K = \sum_{L=1}^{M} \frac{\partial \hat{\psi}}{\partial \hat{a}_L} \frac{\partial \hat{a}_L}{\partial \hat{b}_K} \delta \hat{b}_K$$

$$= \sum_{L=1}^{M} \delta_L \hat{\psi} \left(\frac{\partial \hat{a}_L}{\partial \hat{b}_K} \frac{\delta \hat{b}_K}{\delta \hat{a}_L} \right)$$

That is the $\bar{\delta}_K \hat{\psi}$ are <u>linear</u> combinations of the $\delta_L \hat{\psi}$ with real coefficients and hence the use of the $\hat{b}_K$ as variational parameters is equivalent to the use of the $\hat{a}_L$ and conversely.

We conclude this section with two further comments on equations (4) and (6)

(i) If one is convinced that one has found all solutions of these equations then the lowest $\hat{E}$ is of course the minimum, and is the one to take to approximate E_G (but see below). However, if one has found only <u>a</u> solution then one should test it to see if it is at least a local minimum and not a local maximum or stationary point. In numerical work this has presumably been done by the search procedure.

Analytically one must go on to second order to find

$$\tilde{E} - \hat{E} = \frac{1}{2} \sum_{k,l} \left(\frac{\partial^2 \tilde{E}}{\partial \tilde{a}_k \partial \tilde{a}_l} \right)_{\tilde{a} = \hat{a}} (\tilde{a}_k - \hat{a}_k)(\tilde{a}_l - \hat{a}_l) + \cdots$$

and then one must apply one of the standard tests to see whether or not this quadratic form is postitive. For example one can look to see if the matrix $\left(\frac{\partial^2 \tilde{E}}{\partial \tilde{a}_k \partial \tilde{a}_l} \right)_{\tilde{a} = \hat{a}}$ has any negative eigenvalues.

> Problem: Prove that the matrix must have at least one zero eigenvalue. Hint: What happens if $\tilde{\psi}$ is simply a multiple of $\hat{\psi}$?

(ii) On occasion one may have doubts that the standard calculus sort of approach for finding minima which we have been discussing is really yielding the smallest $\hat{E}$. Namely as we discussed at the outset, in order for various integrals to exist the $\tilde{\psi}$ must satisfy certain conditions and this may well mean that the range of the $\tilde{a}_l$ are not unlimited but that there are "boundaries". If this is the case it might then happen that the lowest $\tilde{E}$ occurred on the boundary. Thus

IV. The Variational Principle and the Schroedinger Equation

In Sec. I we stated but did not prove that if $\tilde{E} = E$ is a stationary point for $\tilde{E}$ as a functional of $\tilde{\Psi}$, then E is an eigenvalue and the corresponding $\tilde{\Psi}$ is an eigenfunction. We now want to justify this statement. The proof takes only a few lines. We have postponed it this long because it will immediately suggest some new points which we will want to explore, and so it seemed best to get the other basic material out of the way first.

First we use the Hermiticity of H to write eq. (III-3) as

$$(\delta\hat{\psi}, (H-\hat{E})\hat{\psi}) + ((H-\hat{E})\hat{\psi}, \delta\psi) = 0 \qquad \text{(IV-1)}$$

from which we can immediately read off our earlier result: If $(H-\hat{E})\hat{\psi} = 0$ i.e., if $\hat{E}$ and $\hat{\psi}$ satisfy the Schroedinger equation, then (1) is satisfied for *any* $\delta\hat{\psi}$, i.e. $\tilde{E} = \hat{E}$ *is* a stationary point for $\tilde{E}$. Moreover, and this the new point, we will now show that if (1) is satisfies for *any* $\delta\hat{\psi}$, i.e., *if* $\tilde{E} = \hat{E}$ is a stationary point, then $\hat{E}$ and $\hat{\psi}$ will satisfy the Schroedinger equation.

To prove this we need only note that if (1) is to be satisfied by any $\delta\hat{\psi}$ it must be satisfied by

$$\delta\hat{\psi} = \delta\eta\,(H-\hat{E})\hat{\psi}$$

where $\delta\eta$ is a real constant. But inserting this into (1) yields

$$((H-\hat{E})\hat{\psi}, (H-\hat{E})\hat{\psi}) = 0$$

and hence

$$(H-\hat{E})\hat{\psi} = 0$$

which proves the point.

V. The Variational Method and "Moments" of the Schroedinger Equation

The two terms which appear on the left hand side of (IV-1) are evidently on another's complex conjugate. Suppose then, as if often the case, that one is dealing with an H which is explicitly real (in the representation in which one is working) and one is using a set of trial functions which is real. Then these two terms will be equal and we may replace (IV-1) by the simpler form

$$(\delta\hat{\psi}, (H-\hat{E})\hat{\psi}) = 0 \tag{V-1}$$

Further this same equation will also hold with complex H (for example coordinate representation for a particle in a magnetic field) and complex trial functions if the set of trial functions is sufficiently flexible so that if $\tilde{\psi}+\delta_k\tilde{\psi}$ is included through first order then so is $\tilde{\psi}+i\,\delta_k\tilde{\psi}$ $(i=\sqrt{-1})$ for all possible $\delta_k\tilde{\psi}$. Namely if this is the case then clearly in determining $\hat{\psi}$ both $\hat{\psi}+\delta_k\hat{\psi}$ and $\hat{\psi}+i\delta_k\hat{\psi}$ will have been examined whence we can apply (IV-1) twice, first with $\delta\hat{\psi}=\delta_k\hat{\psi}$ to find

$$(\delta_k\hat{\psi}, (H-\hat{E})\hat{\psi}) + ((H-\hat{E})\hat{\psi}, \delta_k\hat{\psi}) = 0$$

and then with $\delta\hat{\psi} = i\,\delta_k\hat{\psi}$ to find

$$-(\delta_k\hat{\psi}, (H-\hat{E})\hat{\psi}) + ((H-\hat{E})\hat{\psi}, \delta_k\hat{\psi}) = 0$$

whence

$$(\delta_k\hat{\psi}, (H-\hat{E})\hat{\psi}) = 0 \qquad \text{All } k$$

which is the content of (1). Usually these conditions are met by not imposing *à priori* reality conditions on the variational parameters and/or functions.

Note that when eq. (1) applies the space of variations becomes a <u>complex</u> linear space. Thus if (1) is satisfied by $\delta_1\hat\psi$ and $\delta_2\hat\psi$ then it is also satisfied by $A\delta_1\hat\psi + B\delta_2\hat\psi$ where A and B are arbitrary complex numbers.

When eq. (1) applies, the variational method can be given an interesting and suggestive interpretation. If in a general way we understand by a "moment" of a function F, quantities of the sort

$$(G, F)$$

for various choices of G, then we may say that as an approximation to making $(H-\hat E)\hat\psi = 0$, i.e., to solving the Schroeding equation, the variational method requires that we make a restricted set of moments of $(H-\hat E)\hat\psi$ vanish. (If we make all moments vanish then of course we will satisfy the Schroedinger equation. In this connection we again remark, it may be by chance that a restricted set of trial functions contains an exact eigenfunction.) The definition of $\hat E$:

$$(\hat\psi, (H-\hat E)\hat\psi) = 0$$

can clearly also be interpreted in the same way. (One is satisfying the Schroedinger equation "on the average".)

Now this sort of approach, requiring various moments of $(H-\hat E)\hat\psi$ to vanish, is certainly one which one might come upon (and indeed one which poeple <u>have</u> came upon) without reference to the variational method.

In particular consider the method of <u>linear variational parameters</u> (the <u>Ritz varitional method</u>) which we will discuss in more detail in succeeding sections, in which the set of trial functions consists of functions of the form

$$\tilde\psi = \sum_{L=1}^{M} \tilde{a}_L \phi_L$$

where the ϕ_L (the "basis set") are a given set of linearly independent functions, and where the $\tilde{a}_L$ are arbitrary parameters (thus here is a case in which the space of trial functions *is* a linear space).

Since no reality conditons are imposed on the $\tilde{a}_L$, (1) applies whence we find by inserting

$$\hat{\psi} = \sum_{L=1}^{M} \hat{a}_L \phi_L$$

$$\delta\hat{\psi} = \delta_K\hat{\psi} = \delta\hat{a}_K \phi_K \qquad K=1,\cdots M,$$

the set of homogeneous linear equations

$$\sum_{L=1}^{M} (\phi_K, (H-\hat{E})\phi_L)\hat{a}_L = 0 \tag{V-2}$$

Now the point we want to make is that one can arrive at these same equations, and people often do, by first writing down the "Schroedinger Equation": (the reason for the " " will be discussed in a moment)

$$(H-\hat{E})\sum_{L=1}^{M} \hat{a}_L \phi_L = 0 \tag{V-3}$$

and following the standard procedure of "multiplying through by ϕ_K^* and integrating."

This sort of approach however raises further questions and possibilities. Since the use of ^ has special reference to the variational method let us replace (3) by the more neutral equations

$$(H-\bar{E})\sum_{L=1}^{M} \bar{a}_L \phi_L = 0 \tag{V-4}$$

The procedure of "multiplying through . . ." then provides one way for trying to evaluate $\bar{E}$ and the $\bar{a}_L$. But what about other methods? For example one might try to satisfy (4) indentically at M selected points, or more generally one might try multiplying through by quite another set of functions and integrate to find

$$\sum_{L=1}^{M} (\chi_K, (H-\bar{E})\phi_L)\,\bar{a}_L = 0 \qquad K=1 \cdots M \qquad \text{(V-5)}$$

⟨<u>Problem</u>: What if the χ_K are δ functions?⟩

What is the status of these various approaches? Are they equivalent? Is one superior to the other?

First as to the equivalence; in general (unless perchance an exact eigenfunction is contained in the $\tilde{\Psi}$.) the different procedures lead to different answers. The point is simply that (4) as it stands is not a consistent equation -- there are no $\bar{a}_L$ and $\bar{E}$ which satisfy it (hence our earlier use of "   ") since if there were, then we would have an eigenfunction and eigenvalue of H. If it were a consistent equation then different methods of solving it would lead to the same results. Since it is not consistent, different methods of "solution" will in general lead to different results.

Now as to the advantages of one method over another; as we have seen the variational method leads to (2) and therefore as we know endows it with the virtue that the lowest $\hat{E}$ is a guaranteed upper bound to E_G. Indeed, as we shall see in the next section, it is even more virtuous, the M solution of (2) are, in order, guaranteed upper bounds to the lowest M eigenvalues of H . Thus there is considerable reason to choose (2)

Equations (5) can also be given a variational basis. Suppose we define an energy $\underset{\sim}{E}$ by

$$(\underset{\sim}{\Theta}, (H-\underset{\sim}{E})\underset{\sim}{\Psi}) = 0 \qquad \text{(V-6)}$$

Then one can readily show that if

$$\underset{\sim}{\Psi} = \Psi + \underset{\sim}{\Delta}$$
$$\underset{\sim}{\Theta} = \Psi + \underset{\sim}{\Gamma}$$

then

$$\underset{\sim}{E} = E + \text{terms of order } \underset{\sim}{\Delta}\underset{\sim}{\Gamma},\ \underset{\sim}{\Delta}\underset{\sim}{\Delta} \text{ and } \underset{\sim}{\Gamma}\underset{\sim}{\Gamma}$$

i.e., there are no terms of order $\underset{\sim}{\Delta}$ or of order $\underset{\sim}{\Gamma}$. Thus, in this sense the eigenvalues of H are stationary points for $\underset{\sim}{E}$ as a functional of $\underset{\sim}{\Theta}$ and $\underset{\sim}{\Psi}$, and conversely. This then suggests determining "best" $\underset{\sim}{E}$, $\underset{\sim}{\Psi}$ and $\underset{\vee}{\Theta}$ from (6) and

$$(\delta\underset{\vee}{\Theta}, (H-\underset{\vee}{E})\underset{\vee}{\Psi}) + (\underset{\vee}{\Theta}, (H-\underset{\vee}{E})\delta\underset{\vee}{\Psi}) = 0 \qquad \text{(V-7)}$$

However $\underset{\vee}{E}$ has no bound properties. Also note that it is not necessarily real.

⟨Problem: Prove this.⟩

The connection with (5) comes if we choose

$$\underset{\vee}{\Psi} = \sum_{L=1}^{M} a_L \phi_L \qquad \underset{\vee}{\Theta} = \sum_{L=1}^{M} b_L \chi_L$$

Then inserting

$$\delta \underset{\vee}{\Theta} = \delta b_k \chi_k \qquad k=1 \cdots M$$

into (7) yields (5) in a slightly different notation.

Inserting $\delta \underset{\vee}{\psi} = \delta a_k \phi_k$ leads to the equations

$$\sum_{L=1}^{M} b_L^* \left(\chi_L, (H - \underset{\vee}{E}) \phi_k\right) = 0$$

Note that, as must be the case for consistency, these equations yield the same values of $\underset{\vee}{E}$ as do the equations (5).

All of this is to say that eqs. (5) do have some theoretical foundation over and above the "multiplying . . ." point of view. They have been extensively discussed in the mathematical and applied mathematical literature. However until recently they have not been used much in atomic and molecular calculations, primarily because they don't yield a bound. However recently Boys has argued that the more general (7) may well have definite computational advantages over the variational principle when one is trying to use trial functions involving complicated explicit electron correlation (something which has so far been computationally impossible in the context of the variational principle when more than 3 or 4 electrons are involved) and has backed this up by very successful calculations for several systems. Thus we may expect to hear much more of this approach in the future. However we will not discuss it further here.

VI. Linear Variational Parameters -- The Variational Method for Excited States

Let us return now to equations (V-2). This is a set of linear homogeneous equations to determine the $\hat{a}_k$ and $\hat{E}$. As is well known it has non-trivial solutions (i.e., all $\hat{a}_k$ not identically equal to zero) only for certain values of $\hat{E}$, namely those for which the determinant of coefficients, the so-called secular determinant, vanishes:

$$\left| (\phi_k, (H-\hat{E})\phi_L) \right| = 0 \qquad \text{(VI-1)}$$

> Note that we have here an example of the situation discussed in Sec. II. Namely in accord with the fact that the scale of the $\tilde{\psi}$ was not fixed a priori. We see that the scale of $\hat{\psi}$ is not fixed either, since only the ratios of the $\hat{a}_k$ are determined by (V-2), and also we see that we do not have to invoke $(\hat{\psi}, (H-\hat{E})\hat{\psi}) = 0$ explicitly in order to determine $\hat{E}$.

Equation (1) yields an Mth order algebraic equation for $\hat{E}$, the secular equation. We will denote the roots by $\hat{E}_K$, $K = 1 --- M$ where $\hat{E}_1 \le \hat{E}_2 \le \cdots$ and we will denote the corresponding $\hat{\psi}_K$ by and will choose them to be normalized to one:

$$(\hat{\psi}_K, \hat{\psi}_K) = 1 \qquad \text{(VI-2)}$$

In what follows we will need the following properties of the $\hat{\psi}_K$ which follow from the fact that the "Hamiltonian matrix" $(\phi_K, H\phi_L)$ and the "overlap matrix" (ϕ_K, ϕ_L) are Hermitian matrices: (If you are unsure of these results accept them for now. We will prove them in Sec. VII).

$$(\hat{\psi}_K, \hat{\psi}_L) = \delta_{KL} \tag{VI-3}$$

$$(\hat{\psi}_K, H\hat{\psi}_L) = \hat{E}_K \delta_{KL} \tag{VI-4}$$

Equation (2) is of course a special case of (3). Equations (3) And (4) follow automatically if $\hat{E}_K \neq \hat{E}_L$. If there is degeneracy one can arrange that they will be satisfied with no loss in generality. These properties are (For $K \neq L$ in the case of (4)) of course also shared by the actual eigenfunctions of H.

> Problem: Show that $(\phi_K, H\phi_L)$ And (ϕ_K, ϕ_L) are Hermitian matrices and that the latter is a positive definite matrix (recall the discussion in Sec. II).

All that we know about the $\hat{E}_K$ at the moment is that they are all upper bounds to E_G. We will now prove that they are successively upper bounds to the lowest M eigenvalues of H, with appropriately sharper statements if one can invoke symmetry. Thus through the use of linear variational parameters one can get variational upper bounds for excited states.

> In our discussion we have been taking the ϕ_K to be given, fixed functions. Often however one imbeds variational parameters in them (so-called "non-linear parameters"). Since the result we have stated holds for any value of these parameters one will still have a bound if one chooses the parameters in each $\hat{\psi}_K$ so as to further minimize each $\hat{E}_K$. In general this will require a different parameter choice in each $\hat{\psi}_K$ whence the "price" that one will pay will be that (3) and (4) will no longer be satisfied for $K \neq L$

In Sec. VIII we will give an elegant proof. Here we will proceed in a more pedestrian fashion. We have a basis set of M functions. Let us note this explicitly by writing $\hat{E}_K(M)$ instead of $\hat{E}_K$. Thus in particular (4) becomes

$$(\hat{\psi}_K, H\hat{\psi}_L) = \hat{E}_K(M)\,\delta_{KL} \tag{VI-5}$$

Now we ask for the effect of adding one more function ϕ to our basis set. Clearly we may assume without loss of generality that ϕ is normalized and orthogonal to all the ϕ_K, and hence orthogonal to all the $\hat{\psi}_K$:

$$(\phi,\phi)=1 \qquad , \qquad (\phi,\hat{\psi}_K)=0 \quad K=1\ldots M \tag{VI-6}$$

Let us then write our new optimal wave function as

$$\hat{\Psi} = \sum_{L=1}^{M} \hat{b}_L \hat{\psi}_L + \hat{b}\phi \tag{VI-7}$$

Where for convenience we have used the $\hat{\psi}_K$ instead of the ϕ_K , a step which is certainly allowed since the $\hat{\psi}_K$ span the same space as the ϕ , i.e., among other things, they <u>are</u> M linearly independent linear combinations of the ϕ_K.

If now we insert $\delta\hat{\Psi} = \delta\hat{b}_K \hat{\psi}_K, K=1\ldots M$; and $\delta\hat{\Psi} = \delta\hat{b}\phi$ into (V-1) then we readily derive, using (3), (5) and (7), the equations

$$(\hat{E}_K(M) - \hat{E})\hat{b}_K + (\hat{\psi}_K, H\phi)\hat{b} = 0 \tag{VI-8}$$

$$\sum_{L=1}^{M} (\phi, H\hat{\psi}_L)\hat{b}_L + [(\phi, H\phi) - \hat{E}]\hat{b} = 0 \tag{VI-9}$$

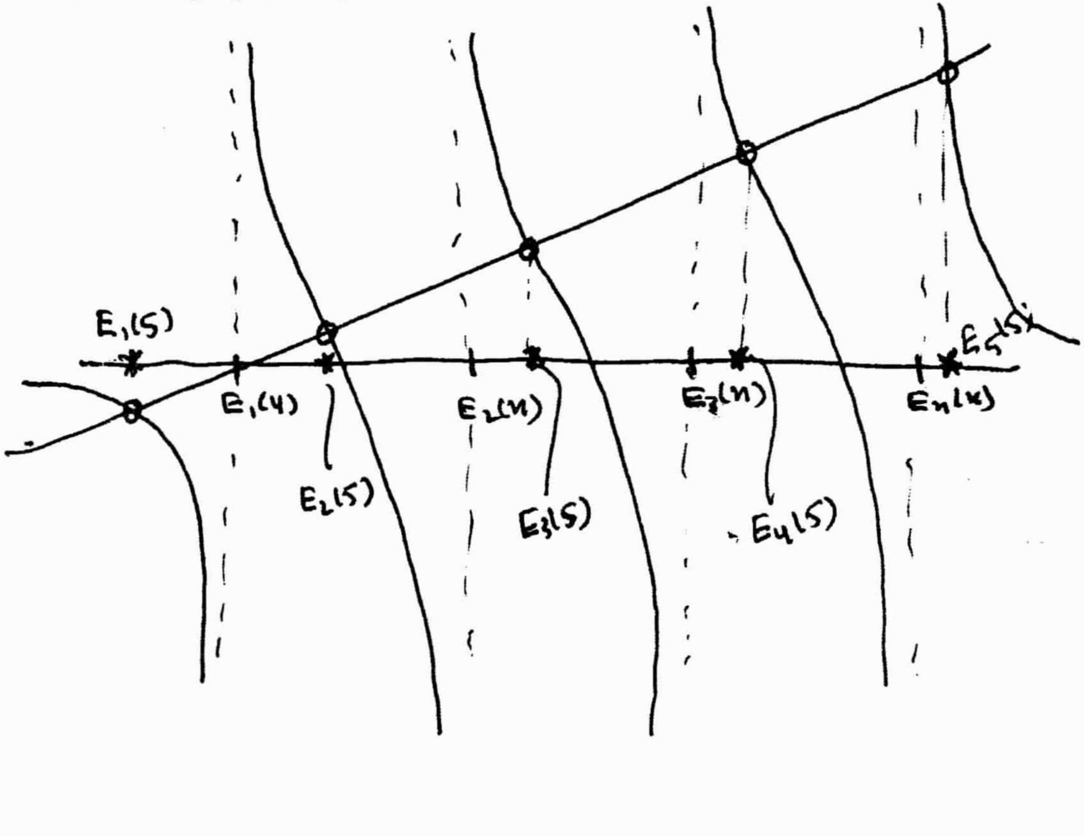
From (8) then we have

$$\hat{b}_L = (\hat{\psi}_L, H\phi)b / \hat{E} - \hat{E}_L(N)$$

Whence (9) yields an equation for $\hat{E}$:

$$\hat{E} - (\Phi, H\Phi) = \sum_{L=1}^{M} \frac{|(\Phi, H\hat{\Psi}_L)|^2}{\hat{E} - \hat{E}_L(M)} \qquad \text{(VI-10)}$$

Consider now the quantity on the right hand side as a function of $\hat{E}$ and assume that all the $\hat{E}_L(M)$ are distinct. We will discuss the case in which there is degeneracy later. Also we will assume that none of the $|(\Phi, H\hat{\Psi}_L)|^2$ vanishes. Then the function has poles when $\hat{E} = \hat{E}_L(M)$. It is negative immediately to the left of the poles and positive to the right of the poles. It goes to zero through positive (negative) values when $\hat{E}$ tends to positive (negative) infinity. The solutions of (10), let us denote them by $\hat{E}_K(M+1)$ are determined by the intersection of this function with the straight line $\hat{E} - (\Phi, H\Phi)$. The situation is shown graphically below for $M = 4$.



Evidently then we have (in general) the "separation theorem"

$$\hat{E}_{K-1}(M) \leq \hat{E}_K(M+1) \leq \hat{E}_K(M) \tag{VI-11}$$

In particular then $\hat{E}_K(M+1) \leq \hat{E}_K(M)$ tells us that as M increases and the basis ϕ_K becomes complete, the $\hat{E}_K$ approach their limiting values, which, assuming convergence, are the eigenvalues of H, from above which proves the point.

> If the eigenfunctions of H can be classified according to symmetry, and if all the ϕ_K have the same symmetry then evidently we can say that the $\hat{E}_K$ will by upper bounds to the M lowest eigenvalues of that symmetry.

> Problem: What if m_1, of the ϕ_K have one symmetry, m_2 another, etc.? Hint: If ϕ_K and ϕ_L have different symmetries then $(\phi_K,\phi_L)=0$ and $(\phi_K,H\phi_L)=0$

The theorem which we have just proven, that the $\hat{E}_K$ yield upper bounds to excited states energies also has a useful converse: If we have a set of functions χ_K which satisfy (VI-3) and (VI-4) that is which satisfy

$$(\chi_K,\chi_L) = \delta_{KL} \qquad K,L = 1 \cdots M$$

$$(\chi_K, H\chi_L) = \epsilon_K \delta_{KL}$$

Then the ϵ_K will be successively upper bounds to the first M eigenvalues of H (of appropriate symmetry, etc.). The proof is trivial. Let us use the χ_K in the linear variational method:

$$\hat{\psi} = \sum_{K=1}^{M} \hat{a}_K \chi_K$$

Then evidently we will find $\hat{\psi}_K = \chi_K$ and $\hat{E}_K = \epsilon_K$ which proves the point.

Even if there is degeneracy among the $\hat{E}_k(M)$ it is clear that the qualitative picture isn't changed since one can consider degeneracy as a limiting case of no degeneracy. Graphically what happens is of course that the appropriate $\backslash$ become steeper and steeper as successive $\hat{E}_k(M)$ come closer together, and in the limit become vertical lines. In particular note that if say $E_2(M) = E_3(M) = \epsilon$ then $^{r}E_3(5)$ would again equal ϵ, though there would be no more degeneracy.

Problem: Show that if there is an n fold degeneracy among the $\hat{E}_k(M)$ at the value ϵ, then the $\hat{E}_k(M+1)$ will have an n-1 fold degeneracy also at the value ϵ.

Problem: If one wants to treat the degenerate case on its own merits rather than as a limiting case, show that one cannot base the discussion on eq. (10). For a discussion based directly on the secular determinant see D.W. Davies, J. Chem. Phys. 33, 781 (1960).

Problem: What happens if one or more of the $(\phi_1, H\hat{\phi}_L)$ vanishes?

One choice of the ϕ_k which has been discussed quite a bit in the literature is

$$\phi_k = H^{k-1}\phi_1$$

(the so-called "method of moments"). Most of its practical applications however have been to problems in which H is a finite matrix. The reason is of course that if H is an atomic Hamiltonian then $(\phi_1, H^L\phi_1)$ may will not exist for $L > 2$ or so even with a "reasonable" choice for ϕ_1. For more details and a discussion of these points see J.B. Delos and S.M. Blinder, J. Chem. Phys. 47, 2784 (1967) and references given there. See also C-Y. Hu, Phys. Rev. 152, 1116 (1966) and 167, 112 (1968). In this latter paper there is also some discussion and use of lower bounds. A careful reading of this paper shows that the methods of moments is being applied not to the atomic H but to a finite Hamiltonian matrix (recall the discussion in Sec. II).

Although primarily of theoretical interest, it is useful to compare the excited state bound one gets using only linear variational parameters, to what one would get if one used linear variational parameters and in addition could also impose orthogonality to lower states as discussed in Sec. II. As might be expected, the latter procedure, if it could be carried out, would generally yield a better bound. Our discussion will be a "pedestrian version" of one given by Perkins, J. Chem. Phys. 45, 2156 (1966). We will reproduce his discussion in Sec. VIII. Consider the first excited state. Then suppose that instead of simply using trial functions of the form $\sum_1^M \tilde{a}_L \phi_L$ we would further require that they be orthogonal to ground state eigenfunction ψ, i.e., we would require that $\sum \tilde{a}_L(\phi_L,\psi)=0$. But then we can use this equation to determine one of the $\tilde{a}_L$ for which $(\phi_L,\psi)\neq 0$, in terms of the others. Let this one be $\tilde{a}_M$. Eliminating a_M then we see that this procedure is equivalent to using as trial functions

$$\sum_1^{M-1} \tilde{a}_L\left[\phi_L-\phi_N\frac{(\psi,\phi_L)}{(\psi,\phi_N)}\right]$$

If we denote the lowest approximate energy that we get from this procedure by $\overset{*}{E}_1$ then clearly it follows from (11), with K = 2, that

$$\overset{*}{E}_1 \le \hat{E}_2(N)$$

On the other hand we also know (Sec. II) that if E_2 is the eigenvalue for the first excited state of H then

$$E_2 \le \overset{*}{E}_1$$

So we have

$$E_2 \le \overset{*}{e}_1 \le \hat{E}_2(M) \qquad \text{(VI-12)}$$

which shows as expected that $\overset{*}{E}_1$, if we sould calculate it, would be a better approximation to E_2 than is $\hat{E}_2(M)$

Problem: Generalize this last result to higher excited states.

VII. Quantum Mechanics in a Subspace

Many of the properties of the method of linear variational parameters become obvious (if they were not obvious already) when one realizes that although the $\hat{E}_k$ and $\hat{\psi}_k$ are (probably) only approximations to the eigenvalues and eigenfunctions of H, they are exact eigenfunctions and eigenvalues of the "subspace Hamiltonian"

$$\bar{H} \equiv \Pi H \Pi \tag{VII-1}$$

where Π is the projection operator onto the subspace of Hilbert space spanned by the trial functions:

$$\Pi \hat{\psi}_k = \hat{\psi}_k \qquad k=1\cdots M \tag{VII-2}$$

$$\Pi \chi = 0 \qquad \text{if } (\hat{\psi}_k, \chi) = 0, \; k=1\cdots M \tag{VII-3}$$

⟨Problem: Prove that Π and $\bar{H}$ are Hermitian operators.⟩

A direct proof is quite trivial. Since from (VI-3) the $\hat{\psi}_k$ are orthonormal, we can write (using Dirac's bracket notation),

$$\Pi = \sum_{L=1}^{M} |\hat{\psi}_L\rangle\langle\hat{\psi}_L|$$

whence it follows that

$$\Pi H \Pi |\hat{\psi}_k\rangle = \sum_L |\hat{\psi}_L\rangle (\hat{\psi}_L, H \hat{\psi}_k)$$

which from (VI-4) becomes

$$\Pi H \Pi |\hat{\psi}_k\rangle = \hat{E}_k |\hat{\psi}_k\rangle$$

which proves the point.

However we also want to give a second somewhat less direct proof which, incidentally, involves no appeal to (VI-3) and (VI-4), since, as we will see, it can be readily generalized to other cases in interest. This proof is as follows: From Sec. I we know that the eigenfunctions and eigenvalues of $\bar{H}$ satisfy

$$(\delta\bar{\Psi}, (\bar{H}-\bar{E})\bar{\Psi}) + (\bar{\Psi}, (\bar{H}-\bar{E})\delta\bar{\Psi}) = 0 \quad \text{All } \delta\bar{\Psi} \qquad \text{(VII-4)}$$

We now want to show that $\hat{\Psi}_k$ and $\hat{E}_k$ are solutions of these equations, i.e., that

$$(\delta\bar{\Psi}, (\bar{H}-\hat{E}_k)\hat{\Psi}_k) + (\hat{\Psi}_k, (\bar{H}-\hat{E}_k)\delta\bar{\Psi}) = 0 \quad \text{All} \qquad \text{(VII-5)}$$

The proof is quite trivial. Namely any $\delta\bar{\Psi}$ can be decomposed into a part <u>in</u> the space of the $\hat{\Psi}_k$, and a part orthogonal to that space

$$\delta\bar{\Psi} = \delta_\parallel\bar{\Psi} + \delta_\perp\bar{\Psi}$$

Then evidently from (3) and the Hermitian property of Π, the contribution of $\delta_\perp\bar{\Psi}$ to the left hand side of (5) vanishes identically, whence we have only to show that

$$(\delta_\parallel\bar{\Psi}, (\bar{H}-\hat{E}_k)\hat{\Psi}_k) + (\hat{\Psi}_k, (\bar{H}-\hat{E}_k)\delta_\parallel\bar{\Psi}) = 0 \quad \text{All } \delta_\parallel\bar{\Psi}$$

But since $\delta_\parallel\bar{\Psi}$ and $\hat{\Psi}_k$ are both in the space, we may write this as

$$(\delta_\parallel\bar{\Psi}, (H-\hat{E}_k)\hat{\Psi}_k) + (\hat{\Psi}_k, (H-\hat{E}_k)\delta_\parallel\bar{\Psi}) = 0 \quad \text{All } \delta_\parallel\bar{\Psi}$$

which of course <u>is</u> true since these are just the basic equations of the method of linear variational parameters.

Thus we have indeed shown (without use of (VI-3) or (IV-4)) that the $\hat{\Psi}_k$ and the $\hat{E}_k$ are eigenfunctions and eigenvalues of $\bar{H}$.

H is an operator over all of Hilbert space. Evidently its other eigenfunctions in addition to the $\hat{\psi}_K$, are any functions χ such that $\Pi\chi = 0$, and all belong to eigenvalue zero.

From this fact it then follows in the usual way (either automatically or as a possible choice) that

$$(\hat{\psi}_K, \hat{\psi}_L) = \delta_{KL} \tag{VIII-12}$$

$$(\hat{\psi}_K, \bar{H}\hat{\psi}_L) = \hat{E}_K\delta_{KL} \tag{VIII-13}$$

and since the last equation can be written

$$(\hat{\psi}_K, H\hat{\psi}_L) = \hat{E}_K\delta_{KL} \tag{VIII-14}$$

we see that we have derived eqs. (VI-3) and (VI-4).

There are other interesting examples of sets of trial functions which form linear spaces. For example there has been considerable interest in the so-called "S-limit" for the Helium atom, namely in finding the optimal trial function of the form $\hat{\psi}(r_1, r_2)$ where r_1 and r_2 are the distances of the two electrons from the nucleus. Clearly such functions (even when restricted to be symmetric in r_1 and r_2) form a linear space. Now the point we want to make is that in <u>all</u> such cases results like the above apply. Namely we can define a Hermitian projection operator Π with the properties

$\Pi\chi = \chi$, χ wholly in the space

$\Pi\chi = 0$, χ orthogonal to the space and then we can simply repeat the second proof given above, symbol by symbol, to show that if $\hat{\Psi}_k$ and $\hat{E}_k$ are derived from the variational method, then they are an eigenfunction and eigenvalue of

$$\bar{H} = \Pi H \Pi$$

In particular then (12) and (14) will hold, whence, in accord with the discussions of Sec. VI it will follow that <u>whenever the trial functions form a linear space the $\hat{E}_k$ will yield upper bounds to excited state energies</u>.

For the particular example of the "S-limit", if the $\hat{\Psi}$ are restricted to be symmetric (antisymmetric) then since such a $\hat{\Psi}$ has angular momentum zero, one will get bounds for the energies of excited singlet (triplet) S-states of He .

A general variational calculation of course involves some restricted set of trial functions. One might then envisage defining a projection operator Π onto the set of trial functions. However, unless the set of trial functions form a linear space neither Π nor $\bar{H}$ will be <u>linear</u> operators and none of the above will apply.

<u>Problem</u>: Prove that if the set of trial functions form a linear space then Π and $\bar{H}$ <u>are</u> linear operators.

Of course unless the space of trial functions can be <u>finitely</u> parametrized, as in the linear variational method, one can rarely expect to solve the $\bar{H}$ problem exactly. In practice what one does then is to restrict oneself to an examination of a <u>subset</u> of functions of the appropriate type, a subset which often does <u>not</u> form a linear subspace. Thus one has the sort of situation mentioned in Sec. I -- as a vari approximation to the H problem one is led to the $\overline{H}$ problem, and in turn one uses the variational method to approximate its solution (usually without mentioning $\overline{H}$ explicitly).

One final point concerning quantum mechanics in a subspace. If for a general operator A we define

$$\overline{A} = \pi A \pi$$

then as long as $\tilde{\Psi}$ and $\tilde{\Psi}'$ are in the subspace we have

$$(\tilde{\Psi}, \overline{A}\tilde{\Psi}') = (\tilde{\Psi}, A\tilde{\Psi}')$$

i.e., we may "drop the bars". However one should keep in mind that in general

$$(\tilde{\Psi}, \overline{AB}\,\tilde{\Psi}') \neq (\tilde{\Psi}, \overline{A}\,\overline{B}\,\tilde{\Psi}')$$

though of course

$$(\tilde{\Psi}, \overline{AB}\,\tilde{\Psi}') = (\tilde{\Psi}, AB\tilde{\Psi}')$$

<u>Problem</u>: Under what conditions will $(\tilde{\Psi}, \overline{A}\,\overline{B}\,\tilde{\Psi}') = (\tilde{\Psi}, \overline{AB}\,\tilde{\Psi}')$ for <u>any</u> $\tilde{\Psi}$ and $\tilde{\Psi}'$ in the subspace?

Thus for example if x and p_x are coordinate and momentum operators respectively then in general

$$(\tilde{\Psi}, (\overline{x}\,\overline{p}_x - \overline{p}_x\overline{x})\tilde{\Psi}') \neq (\tilde{\Psi}, (xp_x - p_x x)\tilde{\Psi}') = i\hbar(\tilde{\Psi}, \tilde{\Psi}')$$

We will return to this point in a later chapter when we discuss frequency dependent polarizibilities and the "f-sum rule".

VIII. The Max-Min Theorem

From the discussion in Sec. I., it is clear that we can characterize E_k , the k'th eigenvalue of H $(E_k \le E_{k+1})$ by

$$E_k = \min_{\tilde{\psi}} (\tilde{\psi}, H\tilde{\psi})/(\tilde{\psi},\tilde{\psi}) \qquad (\psi_i,\tilde{\psi})=0 \quad i=1\ldots k-1 \qquad \text{(VIII-1)}$$

where the ψ_i are the eigenfunctions of H associated with the lower eigenvalues. That is one minimizes $\tilde{E}$ subject to the constraint that the $\tilde{\psi}$ be orthogonal to the lower eigenfunctions. As noted earlier, however, this approach is of little practical use unless one can invoke symmetry, since it requires a knowledge of the ψ_i , $i=1\ldots k-1$ However, there exists another variational approach which is free from this defect, the so-called "Max-Min Theorem". Namely one can show that

$$E_k = \max_{w_i} \min_{\tilde{\psi}} (\tilde{\psi}, H\tilde{\psi})/(\tilde{\psi},\tilde{\psi}) \qquad (w_i,\tilde{\psi})=0 \quad i=1\ldots k-1 \qquad \text{(VIII-2)}$$

where the w_i are $k-1$ arbitrary functions. One first fixes them and determines the minimum of $\tilde{E}$ subject to the contraint that $\tilde{\psi}$ be orthogonal to the w_i . This minimum $\tilde{E}$ is then a functional of the w_i . To find E_k one then maximizes with respect to the w_i .

We will not give a proof here that these two definitions of E_k are equivalent. A proof with references and historical comment can be found in S. H. Gould, "Variational Methods for Eigenvalue Problems" Second edition (Oxford, 1966) Sec. II.6.

However we will give a geometrical picture of the procedure by describing the analogous procedure for finding the principal axis of an ellipsoid. The analogue of $\tilde{\psi}$ is then a "trial axis" (through the center of the ellipsoid) and the analogue of $\tilde{E}$ is the length of this axis.

(1) and (2) yield the same presecription for finding the smallest axis -- minimize $\tilde{E}$ with no constraints on $\tilde{\psi}$.

To find the next smallest, (1) says minimize $\tilde{E}$ subject to the constraint that $\tilde{\psi}$ be orthogonal to the smallest axis. However (2) offers a prescription which is independent of any knowledge of the smallest axis. Namely given a vector w_1 choose $\tilde{\psi}$ to be in the plane perpendicular to w_1 . This plane intersects the ellipsoid in an ellipse. The $\tilde{\psi}$ which yields the minimum $\tilde{E}$ in this plane will then obviously be the minor axis of the ellipse. Now vary the choice w_1 , and find the largest of these minor axes. This will be the intermediate principal axis of the ellipsoid. To find the largest axis the simplest procedure of course is to maximize $\tilde{E}$ with no constraints on $\tilde{\psi}$. In fact, prescription (2) says the same. Namely one is to pass two planes through the ellipsoid, the normals to the planes being w_1 and w_2 . Since these planes will intersect in a line, $\tilde{\psi}$ is uniquely fixed as that line. Then one varies w_1 and w_2, the normals to the two planes, in such a way that $\tilde{E}$ is maximized. Obviously this is equivalent to simply maximizing $\tilde{E}$ directly with no constraints on $\tilde{\psi}$.

We now want to use (2) to give elegant derivation of some of the results in Sec. VI. First we will derive the separation theorem (VI-11). Let $\bar{H}$ be the projection of H onto the M+[illegible] dimensional subspace spanned by the ϕ_K and ϕ. Then evidently

$$\hat{E}_K(M+1) = \underset{w_i}{\mathrm{Max}}\ \underset{\tilde{\psi}}{\mathrm{Min}}\ (\tilde{\psi}, \bar{H}\tilde{\psi})/(\tilde{\psi},\tilde{\psi}) \qquad (w_i,\tilde{\psi})=0 \quad i=1\ldots K-1$$

$$\hat{E}_K(M) = \underset{w_i}{\mathrm{Max}}\ \underset{\tilde{\psi}}{\mathrm{Min}}\ (\tilde{\psi}, \bar{H}\tilde{\psi})/(\tilde{\psi},\tilde{\psi}) \qquad \begin{array}{l}(w_i,\tilde{\psi})=0 \quad i=1\ldots K-1\\ (\phi,\tilde{\psi})=0\end{array}$$

$$\hat{E}_{K-1}(M) = \underset{w_i}{\mathrm{Max}}\ \underset{\tilde{\psi}}{\mathrm{Min}}\ (\tilde{\psi}, \bar{H}\tilde{\psi})/(\tilde{\psi},\tilde{\psi}) \qquad \begin{array}{l}(w_i,\tilde{\psi})=0 \quad i=1\ldots K-2\\ (\phi,\tilde{\psi})=0\end{array}$$

And where the w_i and $\tilde{\psi}$ are to selected from the subspace.

Problem: Justify our definitions of $\hat{E}_K(M)$ and $\hat{E}_{K-1}(M)$. Why did we restrict $\tilde{\psi}$ and w_i to be in the subspace?
Hint: Recall that $\bar{H}$ has other eigenvalues

Comparing $\hat{E}_{K-1}(M)$ and $\hat{E}_K(M+1)$ we see that the prescriptions are similar except that for $\hat{E}_K(M+1)$, w_{K-1} is permitted to vary while in $\hat{E}_{K-1}(M)$ it is in effect fixed at ϕ. Thus the Max in the latter case can't be higher than in the former case and we have

$$\hat{E}_{K-1}(M) \leq \hat{E}_K(M+1)$$

Now let us compare $\hat{E}_K(M+1)$ and $\hat{E}_K(M)$. As far as the w_i are concerned the prescriptions are the same. However in the latter case $\tilde{\psi}$ is more restricted whence the Min can't be lower and we have

$$\hat{E}_K(M+1) \leq \hat{E}_K(M)$$

which completes the derivation of (VI-11).

Now we will derive (VI-12) generalized to the K'th state. Evidently the two procedures can be characterized by ($\bar{H}$ is now the projection of H onto the M dimensional subspace spanned by the ϕ_K)

$$\hat{E}_k = \underset{v_i}{\mathrm{Max}} \; \underset{\tilde{\Psi}}{\mathrm{Min}} \; (\tilde{\Psi}, \bar{H}\tilde{\Psi})/(\tilde{\Psi},\tilde{\Psi}) \qquad (v_i,\tilde{\Psi})=0 \quad i=1\cdots k-1$$

and

$$\overset{*}{E}_k = \underset{\tilde{\Psi}}{\mathrm{Min}} \; (\tilde{\Psi}, \bar{H}\tilde{\Psi})/(\tilde{\Psi},\tilde{\Psi}) \qquad (\psi_i,\tilde{\Psi})=0 \quad i=1\cdots k-1$$

whence follows that

$$\overset{*}{E}_k \leq \hat{E}_k$$

This result is of course independent of the choice of the ψ_i. However with the ψ_i the eigenfunctions of H associated with the lower eigenvalues we also know that

$$E_k \leq \overset{*}{E}_k$$

whence we have

$$E_k \leq \overset{*}{E}_k \leq \hat{E}_k$$

which generalizes (VI-14).

IX. The Unrestricted Hartree-Fock Approximation (URHF)

Although we have on several occasions mentioned the possibility of including arbitrary functions, as well as arbitrary parameters, in the set of trial functions, the only detailed example we gave was in Sec. V where we allowed the $\tilde{\Psi}$ to be quite arbitrary, and then showed that $\hat{\Psi}$ satisfies the Schroedinger equation. (Also in Sec. VIII we briefly mentioned the "S-limit" approximation for Helium.) In this section we will discuss another case which is of great practical importanct. From a formal point of view it is the simplest of the widely used Hartree-Fock (HF) or Self-Consistent-Field (SCF) approximation methods wherein one approximates an eigenfunction of an

N-electron problem by a finite sum of Slater determinants, the number of determinants being kept as small as possible consistent with whatever other restrictions (usually symmetry restrictions of one kind or another) that one wishes to impose. The optimal spin orbitals (subject to possible restrictions) are then determined by use of the variational method.

As the name suggests, in the unrestricted approximation one goes to the limit and uses only a single determinant, and imposes no further restrictions, at least restrictions of a physical nature. However in developing the theory one usually, for convenience, imposes a certain mathematical restriction which we will introduce in the next paragraph.

Thus we wish to use the variational method to determine the "best" single determinant appropriate to a Hamiltonian $\mathcal{H}$ which, typically is a sum of one- and two-electron terms:

$$\mathcal{H} = \sum_{s=1}^{N} h(s) + \frac{1}{2} \sum_{s} \sum_{\substack{t \\ s \neq t}} g(st) \qquad \text{(IX-1)}$$

We now wish to show that there is no loss of generality in assuming that the spin-orbitals which make up the determinant are orthonormal and that the determinant is normalized. Having done this we will then make use of these assumptions since they simplify the analysis. (For

---

Recently there has been great interest in so-called multiconfiguration SCF schemes in which one goes beyond the minimum number of determinants. For example see E. Clementi, Chem. Rev. 68, 341 (1968) and references given there. Also Sabelli & Hinze, J.C.P. 50 685 (1969) and references there a more general discussion see for example T. Gilbert in Molecular Orbitals in Chemistry, Physics and Biology, P. O. Löwdin ed.)

The point is simply the following: Any determinant

$$\chi = |u_1 \cdots u_N| \tag{IX-2}$$

can be written as a numerical multiple of a normalized determinant formed from orthonormal spin orbitals:

$$\tilde{\Psi} = 1/\sqrt{N!}\, |\tilde{\varphi} \cdots \tilde{\varphi}_N| \tag{IX-3}$$

$$(\tilde{\varphi}_s, \tilde{\varphi}_t) = \delta_{st} \tag{IX-4}$$

Granting this, then since χ and $\tilde{\Psi}$ yield the same trial energy, it follows that we can confine attention to determinants of the form $\tilde{\Psi}$.

One way to establish the equivalence is to apply the familiar Gram-Schmidt orthogenalization procedure to the u_t: First we define the orthogonal set of functions

$$\begin{aligned} \eta_1 &= u_1 \\ \eta_2 &= u_2 - u_1 (u_1, u_2) \\ \eta_3 &= u_3 - u_1 (u_1, u_3) - \eta_2 (\eta_2, u_3) \\ &\vdots \end{aligned}$$

Then, we note that from the rules for evaluating determinants it follows that

$$\chi = |\eta_1 \cdots \eta_N|$$

whence if we put

$$\eta_s = (\eta_s, \eta_s)^{1/2} \tilde{\varphi}_s$$

we will have

$$\chi = |\tilde{\varphi}_1 \cdots \tilde{\varphi}_N| \prod_s (\eta_s, \eta_s)^{1/2} = \text{multiple of } \tilde{\Psi}$$

which proves the point. Of course given one set of $\tilde{\varphi}_s$ which do the job, we can find an infinite number of other sets

$$\tilde{\varphi}_t' = \sum_s \tilde{\varphi}_s U_{st}$$

where the U_{st} are numerical coefficients. Namely if the U_{st} form a unimodular matrix (unitary and unit determinant) one finds that

$$\frac{1}{\sqrt{N!}} |\tilde{\varphi}_1' \cdots \tilde{\varphi}_N'|$$

again equals $\tilde{\Psi}$.

<u>Problem</u>: Prove this.

The set of Slater determinants does <u>not</u> form a linear space since in general the sum of two determinants is not a determinant. However there is sufficient linearity, thus

$$|\tilde{\varphi}_1 \, \tilde{\varphi}_2 \cdots \tilde{\varphi}_N| + |\tilde{\varphi}_1' \, \tilde{\varphi}_2 \cdots \tilde{\varphi}_N| = |\tilde{\varphi}_1 + \tilde{\varphi}_1' \, \tilde{\varphi}_2 \cdots \tilde{\varphi}_N|$$

to permit statements about upperbounds to excited states in some instances. See J. F. Perkins, J. Chem. Phys. <u>42</u>, 3827 (1965).

We will now proceed to derive the equations which determine the $\hat{\varphi}_s$, i.e., the best $\tilde{\varphi}_s$. However since only $\hat{\Psi}$ is directly involved we can expect, in light of the remarks at the end of the last paragraph, that the equations which we find will not fix the $\hat{\varphi}_s$ uniquely.

Because we have fixed the overall scale of our trial functions we will have need of both

$$(\delta\hat{\psi}, (H-\hat{E})\hat{\psi}) + (\hat{\psi}, (H-\hat{E})\delta\hat{\psi}) = 0 \qquad \text{(IX-5)}$$

and

$$(\hat{\psi}, (H-\hat{E})\hat{\psi}) = 0 \qquad \text{(IX-6)}$$

However the restriction to normalized trial functions implies that through first order

$$(\delta\hat{\psi}, \hat{\psi}) + (\hat{\psi}, \delta\hat{\psi}) = 0 \qquad \text{(IX-7)}$$

which we can use to simplify (5). Namely any $\delta\hat{\psi}$ can be decomposed into a part parallel to $\hat{\psi}$ and a part orthogonal to $\hat{\psi}$, thus

$$\delta\hat{\psi} = \delta A\,\hat{\psi} + \delta_{\perp}\hat{\psi} \qquad \text{(IX-8)}$$

where δA is a number. Inserting (8) into (7) we then find that δA must be pure imaginary

$$(\delta A)^{*} + \delta A = 0 \qquad \text{(IX-9)}$$

Then inserting (8) into (5) and using (9) we find that the contributions of the δA terms to the left hand side of (5) vanish identically whence we are left with

$$(\delta_{\perp}\hat{\psi}, H\hat{\psi}) + (\hat{\psi}, H\,\delta_{\perp}\hat{\psi}) = 0 \qquad \text{(IX-10)}$$

to determine $\hat{\Psi}$. Now what does a general $\delta_\perp\hat{\Psi}$ look like? A general $\delta\hat{\Psi}$ takes the form

$$\delta\hat{\Psi} = \sum_s \frac{1}{\sqrt{N!}} \left| \hat{\varphi}_1 \cdots \delta\hat{\varphi}_s \cdots \hat{\varphi}_N \right| \tag{IX-11}$$

Now one can certainly write

$$\delta\hat{\varphi}_s = \sum_{t=1}^{N} \delta a_t \hat{\varphi}_t + \delta_\perp \hat{\varphi}_s$$

where the δa_t are constants and where $\delta_\perp\hat{\varphi}_s$ is orthogonal to all the "occupied" spin orbitals:

$$(\hat{\varphi}_t, \delta_\perp\hat{\varphi}_s) = 0 \qquad t = 1 \cdots N \tag{IX-12}$$

But since

$$\frac{1}{\sqrt{N!}} \left| \hat{\varphi}_1 \cdots \sum_{t=1}^{N} \delta a_t \hat{\varphi}_t \cdots \hat{\varphi}_N \right| = \delta a_s \hat{\Psi}$$

it then follows that

$$\delta_\perp\hat{\Psi} = \sum \frac{1}{\sqrt{N!}} \left| \hat{\varphi}_1 \cdots \delta_\perp\hat{\varphi}_s \cdots \hat{\varphi}_N \right|$$

Further since $\delta_\perp\hat{\Psi}$ is a *sum* of terms and since the $\delta_\perp\hat{\varphi}_s$ need have no relation to one another (they are restricted only by (12)), it follows that we may confine attention to a typical term and use

$$\delta_\perp\hat{\Psi} = \frac{1}{\sqrt{N!}} \left| \hat{\varphi}_1 \cdots \delta_\perp\hat{\varphi}_s \cdots \hat{\varphi}_N \right| \tag{IX-13}$$

$$(\hat{\varphi}_t, \delta_\perp\hat{\varphi}_s) = 0 \qquad t = 1 \cdots N \tag{IX-14}$$

Such variations are, for obvious reasons, called one-electron excitations (of $\hat{\Psi}$). Finally, since we have placed no reality restrictions on our trial functions we may replace (10) by

$$(\delta\hat{\Psi}, H\hat{\Psi}) = 0 \qquad \text{(IX-15)}$$

Using the standard rules for calculating the matrix element of H between two Slater determinants which differ by one spin orbital (15) then yields

$$(\delta\hat{\varphi}_s, h\hat{\varphi}_s) + \sum_{t=1}^{N}\left[(\delta\hat{\varphi}_s\hat{\varphi}_t, g\hat{\varphi}_s\hat{\varphi}_t) - (\delta\hat{\varphi}_s\hat{\varphi}_t, g\hat{\varphi}_t\hat{\varphi}_s)\right] = 0 \qquad \text{(IX-16)}$$

Eq. (16) can be written succinctly in terms of the single particle Hartree-Fock Hamiltonian h_{HF} defined by

$$h_{HF}\chi = h\chi + \sum_{t=1}^{N} (\hat{\varphi}_t, g\hat{\varphi}_t)\chi - (\hat{\varphi}_t, g\chi)\hat{\varphi}_t \qquad \text{(IX-17)}$$

where χ is an arbitrary spin orbital, (<u>Problem</u>: Prove that h_{HF} is Hermitian), namely

$$(\delta\hat{\varphi}_s, h_{HF}\hat{\varphi}_s) = 0 \qquad \text{(IX-18)}$$

Comparing (18) with (14) we then conclude that the $\hat{\varphi}_s$ satisfy the set of coupled integro-differential equations

$$h_{HF}\hat{\varphi}_s = \sum_{t=1}^{N} \hat{\varphi}_t \epsilon_{ts} \qquad \text{(IX-19)}$$

where the ϵ_{ts} can, for the moment, be arbitrary numbers.

However we must now ensure, if everything is to be consistent that the $\hat{\varphi}_s$ <u>are</u> orthonormal. Evidently since h_{HF} is Hermitian, we can ensure orthogonality by putting $\epsilon_{st} = 0$ for $s \neq t$ since then the $\hat{\varphi}_s$ will be eigenfunctions of common Hermitian operator (for a comment on degeneracy see below) the ϵ_{ss} then being the eigenvalues.

The question of normalization might appear a bit trickier since if we write out $h_{HF}\hat{\varphi}_s$ using (16), the equations appear quite non-linear. The practical approach to the normalization question (and also to the degeneracy problem, i.e., orthogonality is <u>automatic</u> only if $\epsilon_{ss} \neq \epsilon_{tt}$) is as follows. To solve (18) one proceeds in an iterative fashion. One first guesses some orthonormal spin orbitals, call them $\overset{1}{\varphi}_s$. From these one constructs, in the obvious way, a first approximation to h_{HF} , call it ${}^{1}h_{HF}$. One then proceeds to solve

$$ {}^{1}h_{HF}\,\overset{2}{\varphi}_s = \overset{2}{\epsilon}_{ss}\,\overset{2}{\varphi}_s $$

which is an ordinary eigenvalue problem for the $\overset{2}{\varphi}_s$, [illegible] can be freely made orthonormal. One then proceeds on to ${}^{2}h_{HF}$, $\overset{3}{\varphi}_s$ etc. etc., stopping when a sufficient degree of self-consistency has been attained. Thus in practice there is no difficulty with normalization (or degeneracy).

> If one wishes one can also incorporate this procedure into the theory by simply replacing h_{HF} by an ${}^{*}h_{HF}$ which is independent of normalization and which equals h_{HF} when the spin orbitals are normalized. An obvious choice is
>
> $$ {}^{*}h_{HF}\chi = h\chi + \sum_{t=1}^{N} \frac{(\hat{\varphi}_t, g\hat{\varphi}_t)}{(\hat{\varphi}_t, \hat{\varphi}_t)}\chi - \frac{(\hat{\varphi}_t, g\chi)\hat{\varphi}_t}{(\hat{\varphi}_t, \hat{\varphi}_t)} $$

The equations

$$ h_{HF}\,\hat{\varphi}_s = \epsilon_s\hat{\varphi}_s \qquad \text{(IX-20)} $$

define the so-called canonical URHF spin orbitals (we have dropped the second subscript of ϵ_{ss}). However they are not necessarily the most useful. Other sets derived from the $\hat{\varphi}_s$ by a unimodular transformation, may have more desireable properties, for example they may be better localized (<u>Problem</u>: Show that, as must be the case, such sets satisfy equations of the form (19). Show that the ϵ_{ts} form a Hermitian matrix). Also it has on occasion been suggested that certain non-orthogonal sets could be useful too. (See article by T. Gilbert referred to earlier).

We have not yet calculated $\hat{E}$. To do so we use (6) and readily find

$$\hat{E} = \sum_{s=1}^{N} \hat{\epsilon}_s - \left(\hat{\Psi}, \tfrac{1}{2} \sum\sum g(st)\, \hat{\Psi}\right) \qquad \text{(IX-21)}$$

Thus $\hat{E}$ is not simply the sum of the "spin-orbital energies $\hat{\epsilon}_s$" because such a sum counts the two particle interactions twice. We will discuss the physical significance of the $\hat{\epsilon}_s$ as ionization energies in another chapter; our primary interest in this chapter being to describe the formalism.(<u>Problem</u>: Koopman's Theorem: Consider the $N-1$ particle determinant $\tilde{\Psi} = \frac{1}{\sqrt{(N-1)!}} |\hat{\varphi}_1 \cdots \hat{\varphi}_{N-1}|$. Show that the average $N-1$ particle energy calculated from $\tilde{\Psi}$ differs from $\hat{E}$ by $\hat{\epsilon}_N$, supporting the interpretation of the $\hat{\epsilon}_N$ as ionization energies.)

As we mentioned at the outset, the unrestricted approximation is the most extreme of the Hartree-Fock schemes, and usually the resultant $\hat{\Psi}$ will not have the symmetry properties of the corresponding eigenfunction of H-spin, angular momentum, etc. An important exception occurs in the case of closed shells. Here one can show that with the usual non-relativistic, fixed nucleus Hamiltonian, then, in the absence of external fields, a completely symmetrical $\hat{\Psi}$ <u>is</u> a self-consistent solution of eq. (20). Thus for a closed shell atom the would take the familiar form of radial function times shperical harmonic times spin function, the spatial functions being doubly occupied. In particular for He one can find a solution of the form

$$|\hat{s}\alpha\ \hat{s}\beta|$$

where $\hat{s}$ is a spherically symmetric spatial function and α and β are spin functions. Similarly for Be one can find a solution of the form $|\hat{s}\alpha\ \hat{s}\beta\ \hat{s}'\alpha\ \hat{s}'\beta|$ for Ne a solution of the form $|\hat{s}\alpha\ \hat{s}\beta\ \hat{s}'\alpha\ \hat{s}'\beta\ \hat{p}_+\alpha \cdots \hat{p}_-\beta|$ etc. Thus in these cases one may say that the restricted Hartree-Fock functions satisfy the unrestricted equations. Further one finds that, not unexpectedley, this is then also true in its presence of external fields.

However we should also point out that though for closed shells the completely symmetric solution is self-consistent it may well not yield the lowest energy of all single determinants. For recent discussions see Kaplan and Kleiner Phys. Rev. 156, (1967) and Cizek and Paldus, J. Chem. Phys. 47, 3976 (1967). See also Löwdin in "Quantum Theory of Atoms, Molecules and Solids", P. O. Löwdin, ed. (1965) P. 601.

When one goes to open shells however the situation changes. Thus consider Li. One can check (Problem: Do this.) that a determinant of the form $|\hat{s}\alpha\ \hat{s}\beta\ \hat{s}'\alpha|$ will not satisfy eq. (19) while a function of the form $|\hat{s}\alpha\ \hat{s}'\beta\ \hat{s}''\alpha|$ will. However the latter function is unsatisfactory in that it is not pure spin ½ but contains some spin 3/2. A similar situation will exist also with respect to orbital angular momentum when one goes on to open shells which nominally would involve orbitals with non-zero angular momentum.

Thus for open shells one must impose further restrictions if one wishes $\hat{\psi}$ to have appropriate symmetry. For Li one could simply require that the trial functions take the form $|\tilde{s}\alpha\,\tilde{s}\beta\,\tilde{s}'\alpha|$ and determine $\hat{s}$ and $\tilde{s}'$ from the variational method or, as a more flexible alternative, one might use the appropriate linear combination of $|\tilde{s}\alpha\,\tilde{s}'\beta\,\tilde{s}''\alpha|$ and $|\tilde{s}\alpha\,\tilde{s}''\beta\,\tilde{s}'\alpha|$ which will give spin $\frac{1}{2}$. Similarly for, non-zero angular momentum, one can specify the angular and spin dependence of the spin orbitals, vector coupling together various determinants to produce the desired angular momentum, and then determine the radial functions variationally. General schemes of this sort go under such names as restricted HF (or SCF) or multiconfigurational restricted HF, etc.

Returning to the case of Li, it is possible to imbed $|\hat{s}\alpha\,\hat{s}\beta\,\hat{s}'\alpha|$ and similar restricted single determinants appropriate to systems with spatially closed half filled shells outside of closed shells in what we might call a "slightly restricted scheme" (SRHF). Namely suppose one looks at determinants of the form

$$\tilde{\psi} = \frac{1}{\sqrt{N!}} |\tilde{v}_1\alpha\,\tilde{v}_1\beta\,\tilde{v}_2\alpha\,\tilde{v}_2\beta ---- \tilde{v}_{s-1}\alpha\,\tilde{v}_s\alpha|$$

where the $\tilde{v}_t$ are arbitrary orthonormal spatial orbitals. That is we have a number of doubly occupied orbitals and then a number of valence orbitals all with the same spin.

If one writes down the equations for the optimal $\tilde{v}_t$, i.e., for the $\hat{v}_t$, then one can show that for systems with spatially closed half filled shells outside closed shells, the appropriate restricted orbitals satisfy these equations. Further this also holds in the presence of external fields.

Problem: Derive the equations for

$$\hat{\Psi} = \frac{1}{\sqrt{3!}} | \hat{v}_1 \alpha \hat{v}_1 \beta \hat{v}_2 \alpha | \qquad (\hat{v}_i, \hat{v}_j) = \delta_{ij}$$

Show that one cannot put the analog of ϵ_{12} equal to zero. Show that these equations are satisfied by $\hat{v}_1 = \hat{s}$, $\hat{v}_2 = \hat{s}'$

We will conclude this chapter by some comments on how one actually goes about solving the eq. (19) and their analogue for restricted HF schemes. For closed shell atoms, and for open shells with angular and spin restrictions one can, assuming an symmetric solution, reduce them to a set of non-linear, coupled, 1-dimensional integro differential equations for the radial functions. It has then proved possible to solve these, apparantly with good accuracy, by direct numerical techniques using an interative procedure of the type mentioned earlier.

For molecules such a direct approach still seems pretty much out of the question and one resorts to so-called "analytic" methods. Namely what one really does is to go back to (13) and (14) and restrict the set of trial functions further by replacing the $\tilde{\varphi}_s$ by finite expansions in some basis sets (usually the same set for each s) so that $\delta_\perp \hat{\varphi}_s$ involves varying the finite number of expansion coefficients and any non-linear parameters which may be imbedded in the basis set. Qualitatively the situation here is the same as that discussed near the end of Sec. VII -- the (non-linear in this case) space of trial functions which yields (19) is multiply infinite. One is therefore forced to restrict attention to a finitely parameterized subset.

For fixed values of the non-linear parameters, varying the expansion coefficients leads to a set of coupled non-linear algebraic equations for the expansion coefficients, equations which one can then solve iteratively to self-consistency. One then changes the value of the non-linear parameters, etc.

X. Allowed Variations

The concept of an allowed variation is one which we will find useful in the discussion which follows. We will say that $\Delta\hat{\psi}$ is an <u>allowed</u> variation if it satisfies

$$(\Delta\hat{\psi}, (H-\hat{E})\hat{\psi}) + (\hat{\psi}, (H-\hat{E})\Delta\hat{\psi}) = 0 \tag{X-1}$$

Evidently if $\hat{\psi}$ is an eigenfunction then all $\Delta\hat{\psi}$ are allowed. Allowed variations can be loosely classified into one of three types

(i) Since $\hat{\psi}$ satisfies

$$(\delta\hat{\psi}, (H-\hat{E})\hat{\psi}) + (\hat{\psi}, (H-\hat{E})\delta\hat{\psi}) = 0 \tag{X-2}$$

it follows that any $\delta\hat{\psi}$ is an allowed variation. Such allowed variations one might call "built in". That is from the structure of the set of trial functions one can directly see that $\hat{\psi}+\Delta\hat{\psi}$ will belong to the set through first order and therefore this variation will have been explored in determining $\hat{\psi}$. Thus it follows that $\Delta\hat{\psi}$ will be a $\delta\hat{\psi}$ and hence allowed.

(ii) Certain variations will be allowed for reasons of symmetry. Thus for example if H is rotationally invariant and if $\hat{\psi}$ has a definite angular momentum, then any $\Delta\hat{\psi}$ with a different angular momentum is allowed. We have put such allowed variations in a separate category because they are usually not "built in". Namely, being aware of the symmetry properties of H , one will usually have restricted the class of trial functions to be of the desired symmetry type. Also note that only the symmetry type of $\hat{\Psi}$ is important, not the fact that $\hat{\Psi}$ is an optimal trial function. Thus with $\Delta\hat{\Psi}$ of a different symmetry type from $\hat{\Psi}$, $\hat{\Psi}$ in (1) can be any $\tilde{\Psi}$ of the same symmetry as $\hat{\Psi}$ and the equation will continue to be true.

(iii) The rest. Again let us note that the $\hat{\Psi}$ which arises from a limited set of trial functions could be an eigenfunction. Then whatever one's à priori expectations of the type (i) and (ii), one would find à postiori that all $\Delta\hat{\Psi}$ were allowed.

In some of the later sections we will introduce other variation methods, other optimal trial functions, etc. In all cases the concept of allowed variation will be used analogously to our use here.

XI. Special Theorems Satisfied by Optimal Trial Functions

As a consequence of the Schoedinger equation, eigenfunctions satisfy many special theorems which are often of interest in themselves -- Hellmann-Feynman Theorem, virial theorem, etc. In particular we will be interested in theorems whose satisfaction does not depend simply on symmetry properties. In this section we will derive several such theorems by assuming that various variations are allowed. Conversely then if, for a given set of trial $\tilde{\Psi}$, we know à priori that certain variations are allowed, then we can be sure that the optimal trial function will automatically satisfy the corresponding theorem.

What we will be giving then are sufficient conditions that the theorem be satisfied -- namely if the variation falls into class (i) of the previous section, then the corresponding theorem will be satisfied by $\hat{\psi}$. The existence of class (iii), however, makes it impossible to formulate a useful necessary condition. As a rule of thumb however, for the cases we will discuss, if it isn't clear *a priori* that the variation is allowed, then probably it isn't.

> To come back to a point raised in Sec. III: In saying, as we will say, that such and such approximation satisfies such and such a theorem, we will be assuming that all arithmetic has been done exactly. We will not discuss the effect of errors in numerical analysis; also, in a related vein, for approximations which don't obviously satisfy a theorem, we will *not* attempt to estimate how close they may come.

In this section our approach to the various theorems is to find conditions under which they will be satisfied "naturally" as a result of the choice of the $\tilde{\psi}$ which one uses in the variational method. In another chapter we will discuss a rather different approach in which the theorems are directly imposed as constraints on the variational method.

(A) Generalized Hellmann-Feynman Theorem

Suppose that H contains a real parameter σ. Then from

$$(\hat{\psi}, (H - \hat{E})\hat{\psi}) = 0 \tag{XI-1}$$

we have

$$\left(\frac{\partial\hat{\psi}}{\partial\sigma}, (H - \hat{E})\hat{\psi}\right) + \left(\hat{\psi}, (H - \hat{E})\frac{\partial\hat{\psi}}{\partial\sigma}\right) + \left(\hat{\psi}, \left(\frac{\partial H}{\partial\sigma} - \frac{\partial\hat{E}}{\partial\sigma}\right)\hat{\psi}\right) = 0 \tag{XI-2}$$

Suppose now that

$$\Delta\hat{\psi} \equiv \frac{\partial\hat{\psi}}{\partial\sigma}\,\delta\sigma \qquad \text{(XI-3)}$$

is an allowed variation. Then the sum of the first two terms on the left hand side of (1) will vanish, and we will be left with the generalized Hellmann-Feynman Theorem:

$$\frac{\partial\hat{E}}{\partial\sigma} = (\hat{\psi}, \frac{\partial H}{\partial\sigma}\hat{\psi})/(\hat{\psi},\hat{\psi}) \qquad \text{(XI-4)}$$

The one way we know for ensuring that (3) will be an allowed variation is for the set of trial functions to be chosen in a way which is independent of σ or, to put it more mathematically and more precisely, to be closed under the operation $\sigma \rightarrow \sigma + \delta\sigma$. To see that this works one need only note that then, whatever the σ value, one will search through the same set for $\tilde{\psi}$. Thus if $\hat{\psi}(\sigma_1)$ is the optimal trial function when σ has the value σ_1, then $\hat{\psi}(\sigma_1 + \delta\sigma)$ must also belong to the set whence clearly $\frac{\partial\hat{\psi}}{\partial\sigma}\delta\sigma$ will be a $\delta\sigma$, and hence allowed.

More specifically (A. C. Hurley, Proc. Roy. Soc. A226, 179, (1959)) suppose that the $\tilde{\psi}$ are specified by a number of real variational parameters $\tilde{\alpha}_1 \cdots \tilde{\alpha}_M$. Then the simplest way for the set to be independent of σ is for the $\tilde{\psi}$ not to depend on σ in any way. Thus $\hat{\psi}$ will depend on σ only because the $\hat{\alpha}_1 \cdots \hat{\alpha}_M$ do. Therefore

$$\frac{\partial\hat{\psi}}{\partial\sigma}\delta\sigma = \sum_i \frac{\partial\hat{\psi}}{\partial\hat{\alpha}_i}\frac{\partial\hat{\alpha}_i}{\partial\sigma}\delta\sigma = \sum \delta_i\hat{\psi}\left(\frac{\partial\hat{\alpha}_i}{\partial\sigma}\frac{\delta\sigma}{\delta\hat{\alpha}_i}\right)$$

which is clearly a possible $\delta\hat{\Psi}$ and hence allowed.

Obvious examples of situations in which (3) is allowed are (i) the linear variational method with the Φ_k independent of σ; (ii) most SCF schemes, since although they may involve restrictions an angular or spin dependence of the spin orbitals, usually no *a priori* requirements are imposed on how the functions should depend on possible σ's like nuclear charges, nuclear configurations, strength of external fields, etc; (iii) analytic SCF schemes in which the basic functions are independent of σ. *Problem*: In a linear space we have $\frac{\partial\hat{E}}{\partial\sigma} = (\hat{\Psi}, \frac{\partial H}{\partial\sigma}\hat{\Psi})/(\hat{\Psi}, \hat{\Psi})$. Show that this reduces to (4) when the space is independent of σ. Hint: What is $\partial\Pi/\partial\sigma$?

Although the situation with respect to SCF seems quite clear, still there has been confusion in the literature. See for example W. Huo, J. Chem. Phys. 45, 1554 (1966)--corrected in footnote 11, J. Chem. Phys. 40, 1482 (1968). See also Goddard, J. Chem. Phys. 48, 5337 (1968)--no correction as yet.

If $H = H^{(0)} + \nu H^{(1)}$ where $H^{(0)}$ and $H^{(1)}$ are independent of ν, then if (3) is satisfied with $\sigma = \nu$, i.e., if

$$\frac{\partial\hat{E}}{\partial\nu} = (\hat{\Psi}, H^{(1)}\hat{\Psi})/(\hat{\Psi}, \hat{\Psi})$$

then $\hat{\Psi}$ is said to be "stable" under the "perturbation" $H^{(1)}$ (G. G. Hall, Phil. Mag. 6, 249 (1961). Thus we see that the $\hat{\Psi}$ which arise from the linear variation method (with the independent of σ) and from SCF schemes are stable under arbitrary perturbations. In particular, in connection with SCF, we would stress that $H^{(1)}$ need not be a one-electron operator as has sometimes been implied in the literature.

Conversely if the $\tilde{\Psi}$ depend explicitly on σ, then, unless they are closed under $\sigma \to \sigma + \delta\sigma$, one would not expect the theorem to be satisfied. Thus for example in molecular calculations which involve *finite* basis sets (analytic SCF OR linear variational), usually the functions used involve the nuclear geometry explicity and are different for different geometries. Under these conditions then one does not expect that $\frac{\partial \hat{E}}{\partial R}$ will equal $(\hat{\psi}, \frac{\partial H}{\partial R}\hat{\psi})/(\hat{\psi},\hat{\psi})$ where R is a nuclear separation. (However, note that this statement is "coordinate dependent" -- see below.)

Problem: Would one expect the "method of moments" (Sec. VI) to satisfy (4)?

In our discussion of the generalized Hellmann-Feynman Theorem we have made no reference to the coordinate system (more generally, the representation) which we were using. This of course does not enter in computing $\partial \hat{E}/\partial\sigma$ but it can change the nature of $\partial H/\partial\sigma$ completely. If we denote by $d\xi$ the product of coordinate differentials which are involved in the calculation of $(\hat{\psi}, H\hat{\psi})/(\hat{\psi},\hat{\psi})$ and if we write

$$H = H(\xi, \tfrac{\partial}{\partial\xi}, \sigma)$$

then $\frac{\partial H}{\partial\sigma}$ evidently means $\left(\frac{\partial H}{\partial\sigma}\right)_{\xi, \frac{\partial}{\partial\xi}}$. Similarly by $\hat{\psi}(\sigma_1)$ and $\hat{\psi}(\sigma_1+\delta\sigma)$ we meant $\hat{\psi}(\xi,\sigma_1)$ and $\hat{\psi}(\xi,\sigma_1+\delta\sigma_1)$ respectively; and for example to say that the $\tilde{\psi}$ are independent of σ means that they depend only on ξ and variational parameters. In some other coordinate system they may well then depend on σ. In particular then one may satisfy (4) in one coordinate system but not in others. Thus we may have

$$\frac{\partial \hat{E}}{\partial\sigma} = \left(\hat{\psi}, \left(\frac{\partial H}{\partial\sigma}\right)_{\xi, \frac{\partial}{\partial\xi}} \hat{\psi}\right) \Big/ (\hat{\psi},\hat{\psi}) \neq \left(\hat{\psi}, \left(\frac{\partial H}{\partial\eta}\right)_{\eta, \frac{\partial}{\partial\eta}} \hat{\psi}\right) \Big/ (\hat{\psi},\hat{\psi})$$

For some examples of the effect of different choices of ξ on $\partial H/\partial\sigma$

see C. A. Coulson and A. C. Hurley, J. Chem. Phys. 37, 448 (1962). For further references and a more general discussion see S. T. Epstein, J. Chem. Phys. 42, 3813 (1965).

> We might however indicate one case in point. To derive the Hellmann-Feynman Theorem for a diatomic molecule one uses $\sigma = R$ and Cartesian coordinates with all the electrons referred to the same origin. Then H depends on R only through the electron-nucleus, and nucleus-nucleus terms. Use of (4) then yields the well-known Anschaulåche result. On the other hand if, as is appropriate to a discussion of long range forces, one refers different electrons to different origins, then the two electron terms in H will also involve R and the shape of $\partial H/\partial\sigma$ changes drastically. Use of non-Cartesian coordinate systems will further complicate the picture (see the references given above).

This discussion of different coordinate systems also raises another point. In general a volume element involves not only a product of coordinate differentials, but also a Jacobian J. What then if J depends on σ? (One might also envisage the limits of integration depending on σ. However such dependence can be lumped with J by use of appropriate step functions.) In our discussion up to now we have evidently implicitly ignored this by differentiating only H and $\hat{\varphi}$.

We will not give a general discussion here (see S. T. Epstein, J. Chem. Phys. 46, 571 (1967)) but will note that in many cases of interest, J involves σ only in the form

$$J = f(\sigma)\, g(\xi)$$

whence f cancels out of (XI-1) can we may ignore it.

Problem: Consider a linear harmonic oscillator of frequency ω . Using a Cartesian coordinate, what is $\partial E/\partial\omega$? Using a scaled cartesian coordinate $\xi = \sqrt{\omega}\, x$ what is $\partial E/\partial\omega$? Show that if the resulting two formulae for $\partial\hat{E}/\partial\omega$ are both true, then the virial theorem is satisfied.

(B) On the Prevalence of Interchange Theorems

In chapter   we have seen that if

$$H = H_0 + \lambda V + \mu W$$

then interchange theorems follow from the Hellmann-Feynman theorems

$$\frac{\partial E}{\partial \mu} = (\psi, W\psi)/(\psi,\psi)$$

$$\frac{\partial E}{\partial \lambda} = (\psi, V\psi)/(\psi,\psi)$$

if E is a non-degenerate eigenvalue and ψ the eigenfunction

In a similar way then we will have interchange theorems for a variational approximation if Hellmann-Feynman Theorems are satisfied for $\sigma = \mu$ and $\sigma = \lambda$. Thus in particular to calculate $\langle w\rangle \equiv (\hat\psi, w\hat\psi)(\hat\psi,\hat\psi)$ to any order in μ and first order in we need compute only $(\hat\psi, w\hat\psi)/(\hat\psi,\hat\psi)\big|_{\lambda=0}$ and $\frac{\partial}{\partial\mu}(\hat\psi, V\hat\psi)/(\hat\psi,\hat\psi)\big|_{\lambda=0}$ (the latter yielding $\frac{\partial\langle w\rangle}{\partial\lambda}\big|_{\lambda=0}$). That is, we need $\hat\psi$ only for $\lambda=0$. In particular then there will be interchange theorems within SCF theories, for linear variational approximations (with the ϕ_k independent of λ and μ) etc.

(C) Hypervirial Theorems

Let us suppose that

$$A\hat\psi = i\delta\hat\eta\, \mathcal{H}\hat\psi \qquad \text{(XI-5)}$$

is an allowed variation where $\mathcal{S}$ is Hermitian and $\delta\hat{\eta}$ is a real number. Then from (1) we have

$$(\hat{\Psi}, [\mathcal{S}H - H\mathcal{S}]\hat{\Psi}) = 0 \qquad \text{(XI-6)}$$

Thus if (5) is an allowed variation then $\hat{\Psi}$ satisfies the hypervirial theorem for $\mathcal{S}$. (The average value of the times derivative of $\mathcal{S}$ vanishes in a stationary state.)

On the other hand if

$$\Delta\hat{\Psi} = \delta\hat{\eta}\,\mathcal{S}\hat{\Psi} \qquad \text{(XI-7)}$$

is allowed then we find

$$(\hat{\Psi}, (\mathcal{S}H + H\mathcal{S})\hat{\Psi}) = 2\hat{E}(\hat{\Psi}, \mathcal{S}\hat{\Psi}) \qquad \text{(XI-8)}$$

and if both are allowed we have

$$(\hat{\Psi}, \mathcal{S}H\hat{\Psi}) = (\hat{\Psi}, H\mathcal{S}\hat{\Psi}) = \hat{E}(\hat{\Psi}, \mathcal{S}\hat{\Psi}) \qquad \text{(XI-9)}$$

The one way which has been found to ensure that such variations are allowed is to choose the set of trial functions in such a way that if $\tilde{\Psi}$ is a member of the set then so is $(1 + \delta\tilde{\eta}\,\mathcal{S})\tilde{\Psi}$ to first order, whence, depending on the reality conditions imposed on $\delta\tilde{\eta}$, (5) or (7) or both will be allowed.

Before illustrating some of the possibilities by means of examples it might be well to mention two $\mathcal{S}$'s of interest:

(A) If $\mathcal{S}$ is the "VIRIAL"

$$\mathcal{S} = \frac{1}{2}\sum_{s=1}^{N}(\vec{p}_s \cdot \vec{r}_s + \vec{r}_s \cdot \vec{p}_s)$$

where $\vec{p}_s$ and $\vec{r}_s$ are the Cartesian momenta and coordinates of the s'th particle, then one finds that, with H the non-relativistic fixes nucleus Hamiltonian for atoms or molecules in the absence of external fields, that the hypervirial theorem for $\mathcal{G}$ is just the familiar Virial Theorem, at least for atoms. For diatomic molecules what one finds is

$$\hat{T} + \hat{E} + R\left(\hat{\psi}, \frac{\partial H}{\partial R}\hat{\psi}\right)/(\hat{\psi},\hat{\psi}) = 0 \qquad \text{(XI-10)}$$

where $\hat{T}$ is the average kinetic energy and R is the nuclear separation and where we are using Cartesian coordinates. (Problem: Derive these results for atoms and diatomic molecules.) This then becomes what is usually called the virial theorem, namely

$$\hat{T} + \hat{E} + R\frac{d\hat{E}}{dR} = 0 \qquad \text{(XI-11)}$$

if the Hellmann-Feynman Theorem for $\sigma = R$ is also satisfied in Cartesian coordinates.

Depending on whether H is the Electronic Hamiltonian or the total Hamiltonian, including nuclear repulsion, $\hat{E}$ in (10) is the electronic or the total energy. In (11) either may be used since the energy of nucleus repulsion satisfies

$$E_{NR} + R\frac{dE_{NR}}{dR} = 0$$

Problem: Derive the analogous results for polyatomic molecules.

(B) If g is a component of the total electronic momentum

$$g = \sum_{j=1}^{N} p_i \equiv P_i$$

then the hypervirial theorem for g says that on the average there is no net force in the i'th direction acting on the electrons. As an interesting application consider an atom in a uniform electric field $\vec{\mathcal{E}}$. Then the net force on the electrons is equal to the force due to the nucleus plus the force due to the electric field, the electron-electron forces cancelling out by action equals reaction. Thus if the hypervirial theorems are satisfied for the three P_i we will have

$$0 = -N\vec{\mathcal{E}} + \vec{F}_{Ne} \qquad \text{(XI-12)}$$

On the other hand the force on the nucleus is

$$\vec{F}_N = Z\vec{\mathcal{E}} - \vec{F}_{Ne} \qquad \text{(XI-13)}$$

where Z is the nuclear charge. But from (12) we can then write $\vec{F}_N$ as

$$\vec{F}_N = Z\left(1 - \frac{N}{Z}\right)\vec{\mathcal{E}} \equiv Z(1-\beta)\vec{\mathcal{E}}$$

Thus when the hypervirial theorems for all the P_i are satisfied, the so-called "dipole shielding factor" β will take on the (correct) value N/Z.

Now let us turn to some examples of situations in which hypervirial theorems are satisfied.

(i) Let g be a one-particle operator and $\hat{\Psi}$ the optimal URHF Slater determinant. Thus we consider

$$\mathcal{G} = \sum_s G(s)$$

$$\Delta\hat{\Psi} = \delta\hat{\eta}\,\mathcal{G}\hat{\Psi} = \sum_s \frac{1}{\sqrt{N!}}\left|\hat{\varphi}_1 \cdots \delta\hat{\eta}\,G\hat{\varphi}_s \cdots \hat{\varphi}_N\right|$$

with $\delta\hat{\eta}$ either real or pure imaginary.

The latter is qualitatively the sort of variation which we considered in Sec. IX in connection with URHF, i.e., a sum of spin orbital variations. However unless $\delta\hat{\eta}$ is pure imaginary it does not satisfy

$$(\Delta\hat{\Psi}, \hat{\Psi}) + (\hat{\Psi}, \Delta\hat{\Psi}) = 0$$

which was a further condition which we imposed there. However since we argued at that time that such restrictions were only mathmatical and not physical, we can infer that $\Delta\hat{\Psi}$ must be allowed both for $\delta\hat{\eta}$ pure imaginary and for $\delta\hat{\eta}$ real, and hence that (6), (8), and (9) should be satisfied. We can also prove this explicitly. Namely clearly

$$\Delta\hat{\Psi} = \delta\hat{\eta}\,\frac{(\hat{\Psi}, \mathcal{G}\hat{\Psi})}{(\hat{\Psi}, \hat{\Psi})}\,\hat{\Psi} + \delta_\perp\hat{\Psi}$$

where $\delta_\perp\hat{\Psi}$ is the sort of variation which we considered there following (IX-12) and which we found to satisfy $(\delta_\perp\hat{\Psi}, H\hat{\Psi}) = 0$. Thus on the one Hand we have

$$(\Delta\hat{\Psi}, H\hat{\Psi}) = (\delta\hat{\eta})^*(\hat{\Psi}, \mathcal{G}H\hat{\Psi})$$

while on the other hand

$$(\Delta\hat{\Psi}, H\hat{\Psi}) = (\delta\hat{\eta})^*\,\frac{(\hat{\Psi}, \mathcal{G}\hat{\Psi})}{(\hat{\Psi}, \hat{\Psi})}(\hat{\Psi}, H\hat{\Psi}) = (\delta\hat{\eta})^*(\hat{\Psi}, \mathcal{G}\hat{\Psi})\hat{E}$$

Equating these then yields

$$(\hat{\Psi}, \mathcal{G}H\hat{\Psi}) = \hat{E}\,(\hat{\Psi}, \mathcal{G}\hat{\Psi})$$

which is part of (9), its complex conjugate yielding the rest. Then since (6) and (8) follow from (9) the point is proven. Thus URHF satisfies (9) and all hypervirial theorems (6) for one-electron operators and hence in particular the virial theorem and the theorem for β. (Note we did not add the qualifier "Hermitian". <u>Problem</u>: Show that if these theorems are satisfied for all Hermitian one-electron $\mathcal{D}$, they are satisfied for all one-electron operators.) In particular then restricted Hartree-Fock for closed shells atoms will also satisfy these theorems since, as mentioned in Sec. IX, in this case the restricted functions satisfy the unrestricted equations. (For some recent calculations and further references see Kaneko and Aron, Phys, Soc. Japan <u>26</u>, 110 (1969).)

Also it is easy to see that if $\mathcal{D}$ is any spin <u>independent</u> one-electron operator then the SRHF mentioned in Sec. IX will satisfy the hypervirial theorem for $\mathcal{D}$. (<u>Problem</u>: Prove this.) Since the virial theorem and the dipole shielding theorem involve $\mathcal{D}$'s of this type it follows in particular that they will also be satisfied by restricted Hartree-Fock approximations for atoms with spatially closed half filled shells outside of closed shells.

When there are external fields present one usually adds the adjective "coupled" to further delineate the various HF schemes which we have been discussing. This is to distinguish them from various "uncoupled" HF schemes which we will discuss in another chapter. Thus M. Cohen (Proc. Roy. Soc. <u>A293</u>, 365 (1966), Proc. Phys. Soc. <u>92</u>, 23 (1967)) has reported some calculations for the alkalis which seem to contradict the results found above. However what he calls HF in these papers is not coupled HF but is some version of uncoupled HF.

It is also useful to note that with the origin of coordinates taken at the nucleus $s = \frac{1}{2}\Sigma(\vec{p}_i\cdot\vec{r}_i + \vec{r}_i\cdot\vec{p}_i)$ really involves only the radial coordinate and radial momentum of each particle. Thus even if $\hat{\Psi}$ is a sum of determinants derived from a restricted HF approximation, since $s\hat{\Psi}$ involves only variations of radial functions, it will follow that $\hat{\Psi}$ will satisfy the virial theorem.

As implied in the previous note the virial depends on the origin of coordinates. Problem: Show that if the hypervirial theorems for $\vec{p}$ are satisfied then if the virial theorem is satisfied for one origin, it is satisfied for all origins.

(ii) One may [illegible] satisfy a hypervirial theorem by explicitly introducing a variational parameter to do the job. Thus if one uses trial functions of the form

$$\tilde{\Psi} = e^{\tilde{\eta} s}\,\tilde{\Theta} \qquad \text{(XI-14)}$$

where $\tilde{\eta}$ is a pure imaginary (real) variational parameter and is independent of $\tilde{\eta}$, but may involve other parameters and/or functions, then clearly the variations (5) (the variations (7)) will be allowed.

In the constrained variational approach mentioned at the beginning of this Sec. XI, one instead constrains the variational parameters already available in order to satisfy the theorem. Thus in the present approach, using $\tilde{\Psi}$ instead of $\tilde{\Theta}$ cannot raise the optimal $\tilde{E}$ and probably lowers it, while in the constrained approach $\hat{E}$ cannot be lowered and will probably be raised.

In particular for $\mathcal{S} = P_K$ and $\tilde{\eta}$ pure imaginary, one readily finds that

$$\tilde{\psi}(\vec{r}_1 \cdots \vec{r}_N, \tilde{\eta}) = \tilde{\Theta}(\vec{r}_1 + \vec{\tilde{\epsilon}}, \cdots, \vec{r}_N + \vec{\tilde{\epsilon}}) \qquad \text{(XI-15)}$$

where

$$\tilde{\epsilon}_K = |\tilde{\eta}| \quad , \quad \tilde{\epsilon}_L = 0 \quad L \neq K \qquad \text{(IX-16)}$$

(<u>Problem</u>: Derive this result.) Thus if the trial functions are not *a priori* linked to a particular origin along the K'th axis, the hypervirial theorem for P_K will be satisfied. <u>Problem</u>: What of the other theorems?

<u>Problem</u>: Prove that if $\tilde{\psi} = e^{i\vec{\tilde{\epsilon}} \cdot \vec{P}} \tilde{\Theta}$ with $\vec{\tilde{\epsilon}}$ now arbitrary but real, then $\tilde{\psi} = \tilde{\Theta}(\vec{r}_1 + \vec{\tilde{\epsilon}}, \cdots, \vec{r}_N + \vec{\tilde{\epsilon}})$ and the hypervirial theorems for all ~~three~~ components of $\vec{P}$ will be satisfied.

The preceding problem raises the general question of satisfying several hypervirial theorems simultaneously using an approach of the type (14). <u>Problem</u>: Show that if the $\mathcal{S}_i$ commute, that this can be done as above by use of

$$\tilde{\psi} = e^{\sum_k \tilde{\eta}_k \mathcal{S}_k} \tilde{\Theta}$$

where the $\tilde{\eta}_k$ are pure imaginary and $\tilde{\Theta}$ is independent of the $\tilde{\eta}_k$. For a discussion of the more general case (and for more details on finding $\tilde{\psi}$ explicitly given $\mathcal{S}$ and $\tilde{\Theta}$) see S. T. Epstein and J. O. Hirschfelder, Phys. Rev. <u>12[illegible]</u>, 1495 (1961). See also D. Pandres, Phys. Rev. <u>131</u>, 886 (1963).

Of course in the absence of external fields the theorems for P_K will often be satisfied simply because of symmetry. Thus consider an atom. Then if $\hat{\psi}$ has a definite parity we will have

$$(\hat{\psi}, \vec{P}\hat{\psi}) = (\hat{\psi}, H\vec{P}\hat{\psi}) = (\hat{\psi}, \vec{P}H\hat{\psi}) = 0$$

Note again that this has nothing to do with $\hat{\Psi}$ being an <u>optimal</u> trial function. All that is important is that it have the right symmetry. Also even in the present of an electric field the theorems for $\vec{P}_\perp$ (the components of $\vec{P}$ perpendicular to $\vec{\mathcal{E}}$) can be satisfied by symmetry. In this connection one can prove the following (<u>Problem</u>: Do this.) If the hypervirial theorem for $\mathcal{A}$ is satisfied simply for reasons of symmetry, then if the $\tilde{\Theta}$ have the appropriate symmetry, $\hat{\eta}$ will equal zero.

As another example of the sort of thing, this time related to time reversal invariance, suppose that H and $\hat{\Psi}$ are real, then if $\mathcal{A}$ is explicitly real

$$(\hat{\Psi}, (H\mathcal{A} - \mathcal{A}H)\hat{\Psi}) = 0$$

<u>Problem</u>: Prove this. Hint: Note that from our hypotheses $i(H\mathcal{A}-\mathcal{A}H)$ is a pure imaginary Hermitian operator.

<u>Problem</u>: Discuss H and $\hat{\Psi}$ real, $\mathcal{A}$ pure imaginary.

Let us now choose $\mathcal{A}$ to be the "virial" i.e., we take

$$\mathcal{A} = \frac{1}{2}\sum_{s=1}^{N}(\vec{p}_s\cdot\vec{r}_s + \vec{r}_s\cdot\vec{p}_s)$$

Then with $\tilde{\eta}$ pure imaginary one can show that

$$\Psi(\vec{r}_1 \cdots \vec{r}_N, \tilde{\eta}) = (\tilde{\zeta})^{\frac{3N}{2}}\,\tilde{\Theta}(\tilde{\zeta}\vec{r}_1 \cdots \tilde{\zeta}\vec{r}_N) \qquad \text{(XI-17)}$$

with

$$\tilde{\zeta} = e^{i\tilde{\eta}} \qquad \text{(XI-18)}$$

Thus we can say that if the scale of electronic coordinates can be freely varied then the virial theorem will be satisfied.

<u>Problem</u>: Discuss the possibilities for complex scaling.

Problem: Derive (17) and (18). Hint: Considering one dimension prove that

$$\frac{i}{2}(px+xp)\left\{e^{\frac{\tilde\eta}{2}}\tilde{\Theta}(e^{\tilde\eta}x)\right\} = \frac{\partial}{\partial\tilde\eta}\left(e^{\frac{\tilde\eta}{2}}\tilde{\Theta}(e^{\tilde\eta}x)\right)$$

In discussing the virial we have had in mind that H was an atomic or molecular Hamiltonian in the absence of external fields, with neglect of spin and velocity dependent forces. However we should emphacize that whatever the nature of H (atomic, molecular, nuclear, with or without external fields, spin dependent forces, etc.) scaling will assure that the average value of the commutator of H and the virial will vanish. In particular for a single particle in a potential V this then yields the general virial theorem

$$2\hat{T} - \frac{(\hat\Psi, \vec{r}\cdot\vec{\nabla}V\hat\Psi)}{(\hat\Psi,\hat\Psi)} = 0$$

For diatomic molecules then this means the theorem (10). In order that (11) also hold we could then also require that $\tilde{\Theta}$ be explicitly independent of R.

Such an approach would then satisfy both (10) and (11). Not surprisingly there is an alternative approach which in general satisfies only (11). Namely let $\tilde{\Theta}$ depend on R, but scale that dependence as well. Thus use

$$\tilde\Psi = (\tilde\zeta)^{\frac{3N}{2}}\tilde\Theta(\tilde\zeta\vec{r}_1,\dots,\tilde\zeta\vec{r}_N,\tilde\zeta R) \qquad \text{(XI-19)}$$

One sees that this works as follows: If one calculates (Problem: Carry out these calculations) $\partial\hat{E}/\partial R$ one will find the Hellmann-Feynman result plus an extra term, call it X. If one then makes use of the fact that $\Delta\hat\Psi = \frac{\partial\hat\Psi}{\partial\hat\eta}\delta\hat\eta$ is an allowed variation one will derive $(10)+X=0$ Combining these results then yields (11). (See P. O. Löwdin, J. Mol. Sepec. 3, 46 (1959). Also E. C. Herrmann, Chem. Phys. Lett. 1, 253 (1967)).

Problem: For what $\mathcal{A}$'s would hypervirial theorems be satisfied if one scaled the and coordinates of the particles independently?

(iii) Finally let us consider the situation when the set of trial functions forms a linear space as in Sec. VII. A feature of such cases which we have not commented on explicitly earlier, but which it is now useful to note, is that, in contrast to the general situation, the $\delta\hat{\Psi}$ belong to the space. This is obvious when one has linear variational parameters and in the general case it follows from the ramark that $\tilde{\Psi}-\hat{\Psi}$ belongs to the space for any $\tilde{\Psi}$ in the space.

Thus if $\Delta\hat{\Psi}$ is to be allowed because it is a $\delta\hat{\Psi}$, it follows that $\Delta\hat{\Psi}$ must belong to the space. Thus, with respect to the theorems of interest at the moment, this means that $\mathcal{A}\hat{\Psi}$ should belong to the space. The one way we know to ensure this is to have $\mathcal{A}\tilde{\Psi}$ belong to the space if $\tilde{\Psi}$ does. For the $\mathcal{A}$'s which we have been talking about, this is difficult unless the space is ∞ dimensional. Thus one does not expect the linear variational method to satisfy interesting hypervirial theorems. (We are assuming the ϕ_k to be fixed. If one allows non-linear parameters, scaling, etc., then that is a different story -- the space is no longer linear, and no longer finite dimensional.) On the other hand if one is dealing with infinite spaces then the prospects are better. Thus in view of our remarks at the end of the discussion (i) of this section, the "S-limit" for H_2 will satisfy the virial theorem.

Problem: Define $\bar{\mathcal{A}} = n\,\partial/\partial n$ Show that the theorems for all $\bar{\mathcal{A}}$ are satisfied in the linear variational method

(C) Orthogonality

Let $\hat{\psi}$ and $\hat{\psi}'$ be two solutions of the variational equations associated with energies $\hat{\epsilon}$ and $\hat{\epsilon}'$ respectively. Further let us suppose that

$$\Delta\hat{\psi} = \delta\hat{\eta}\,\hat{\psi}'$$

$$\Delta\hat{\psi}' = \delta\hat{\eta}'\,\hat{\psi} \qquad \text{(XI-20)}$$

are allowed variations of $\hat{\psi}$ and $\hat{\psi}'$ respectively where $\delta\hat{\eta}$ and $\delta\hat{\eta}'$ may be arbitrary complex numbers. Choosing $\delta\hat{\eta}$ and $\delta\hat{\eta}'$ first real and then pure imaginary one finds from the fact that $\Delta\hat{\psi}$ is allowed

$$(\hat{\psi}', (H - \hat{\epsilon})\hat{\psi}) = 0 \qquad \text{(XI-21)}$$

and from the fact that $\Delta\hat{\psi}'$ is allowed

$$(\hat{\psi}', (H - \hat{\epsilon}')\hat{\psi}) = 0$$

Subtracting then we have

$$(\hat{\epsilon}' - \hat{\epsilon})(\hat{\psi}', \hat{\psi}) = 0 \qquad \text{(XI-22)}$$

that is, if (20) are allowed and if $\hat{\epsilon} \neq \hat{\epsilon}'$, then $\hat{\psi}$ and $\hat{\psi}'$ will automatically be orthogonal. Further note that if we use (22) in (21) we have

$$(\hat{\psi}', H\hat{\psi}) = 0$$

and the considerations in Sec. VI concerning upper bounds to excited states can be applied.

Now when is it likely that the variations (20) <u>will</u> be allowed because $\Delta\hat{\psi}$ and $\Delta\hat{\psi}'$ are a $\delta\hat{\eta}$ and a $\delta\hat{\eta}'$? One case, which we have already discussed, is of course that in which the trial functions form a linear space, and in fact we know of no other (!). However one can use the result in reverse. Thus in URHF method the different solutions, ground state, first excited state of the same symmetry, etc., etc., in general differ from one another not just by one spin-orbital, as would be needed if (20) were to be allowed, but in all spin-orbitals. Thus one does not expect, and indeed one does not find, exact orthogonality. Similar remarks clearly also apply to other SCF schemes. However, of course, near orthogonality is not excluded. For some representative calculations and more references see for example Bagus, Phys. Rev. <u>139</u>, A619 (1965).

Though mainly only of formal interest note that if $\Delta\hat{\psi} = \delta\hat{\eta}\,\mathcal{J}\hat{\psi}'$ and $\Delta\hat{\psi}' = \delta\hat{\eta}\,\mathcal{J}\hat{\psi}$ are allowed with $\delta\hat{\eta}$ and $\delta\hat{\eta}'$ real or pure imaginary, then if the off idagonal hypervirial theorem for $\mathcal{J}$ will be satisfied:

$$(\hat{\psi}', [\mathcal{J}, H]\hat{\psi}) = (\hat{E} - \hat{E}')(\hat{\psi}', \mathcal{J}\hat{\psi})$$

For similar reasons as above one does not expect such relations to be satisfied in SCF schemes. See for example Vetchinkin, Opt. Specta (USSR) <u>14</u>, 169 (1963); La Paglia and Sinanoglu, J. Chem. Phys. <u>44</u>, 1088 (1966); and La Paglia J. Mol. Spec. <u>24</u>, 302 (1967), for discussion of oscillator strengths in SCF schemes.

(D) <u>Brillouin's Theorem</u>

Suppose that

$$\Delta\hat{\psi} = \delta\hat{\eta}\,\psi' \qquad \text{(XI-23)}$$

is an allowed variation for $\delta\hat{\eta}$ either real or pure imaginary, and that ψ' is orthogonal to ψ

$$(\hat{\psi}, \psi') = 0 \qquad \text{(XI-24)}$$

Then it follows that

$$(\psi', H\hat{\psi}) = 0 \qquad \text{(XI-25)}$$

i.e., the matrix element of H between ψ' and ψ vanishes. This result we call (the generalized) Brillouin's Theorem, since in the special case of URHF it becomes the Brillouin's Theorem.

This theorem is really not so much a "special theorem" like the virial theorem, but rather is another way of expressing the basic equations of the variational method. (For a derivation of various HF theories from this point of view see Nesbet, Rev. Mod. Phys. 33, 28 (1961) and Kaldos, J. Chem. Phys. 48, 835 (1968).) However the point is that when expressed in this way, one sees that the equations of the variational method directly contain information about what may happen if one tries to improve upon $\hat{\psi}$ by means of perturbation theory.

To do this one will introduce a zero-order Hamiltonian $H^{(0)}$ with the property

$$H^{(0)}\hat{\psi} = E^{(0)}\hat{\psi}$$

where $E^{(0)}$ may or may not equal $\hat{E}$. (See the end of this section -- note that in any case $E^{(0)} + E^{(1)} = \hat{E}$.) Then introducing the remaining orthonormal eigenfunctions $\psi_n^{(0)}$ and eigenvalues $e_n^{(0)}$ of H, the first order correction to $\hat{\psi}$ can be written in the well-known form (we assume that $E^{(0)}$ is non-degenerate)

$$\hat{\Psi}^{(1)} = -\sum_n \Psi_n^{(0)} \frac{(\Psi_n^{(0)}, H\hat{\Psi})}{(E_n^{(0)} - E^{(0)})} \tag{XI-26}$$

Now the import of Brillouin's Theorem is clear -- if $\Delta\hat{\Psi} = \delta\hat{\eta}\,\Psi_n^{(0)}$ is an allowed variation then $\Psi_n^{(0)}$ won't appear in the expression for $\hat{\Psi}^{(1)}$ and hence will have no effect on the energy through third order (Problem: Show that $\Psi_n^{(0)}$ may well appear in $\Psi^{(2)}$).

Another ~~way~~ method which is often used to improve a $\hat{\Psi}$ is to do a linear variational calculation, the basis functions being $\hat{\Psi}$ and a selected set Φ_k. This is the so-called configuration interaction (CI) method. Brillouin's Theorem then tells us that any Φ_k for which $\Delta\hat{\Psi} = \delta\hat{\eta}\,\Phi_k$ were allowed variations, will not be connected directly to $\hat{\Psi}$ in the resulting Hamiltonian matrix. For an application to multi-configurational Hartree-Fock theory see Levy and Berthier, Int. J. W. Chem. 2, 307 (1968) and references given there.

URHF is a particularly interesting case in point. The Ψ' for which (25) holds are, as we know, just the one-electron excitations of $\hat{\Psi}$. Thus if we choose for $H^{(0)}$ a one-electron operator (the choice $H^{(0)} = H_{HF} \equiv \sum_{s=1}^{N} h_{HF}(s)$ clearly being the most natural) so that the $\Psi_n^{(0)}$ are single determinants involving various excitations of $\hat{\Psi}$, then it will follow that $\Psi^{(1)}$ will involve no one-electron excitations.

This then implies the further interesting and useful result that for such an $H^{(0)}$, there are no first order corrections to the average value of any one-electron operator. Namely the first order corrections would be

$$\left(\Psi^{(1)}, \left(W - \frac{(\hat{\Psi}, W\hat{\Psi})}{(\hat{\Psi}, \hat{\Psi})}\right)\hat{\Psi}\right) + c.c.$$

But if W is a one-electron operator then

$$\left[W - \frac{(\hat{\varphi}, W\hat{\varphi})}{(\varphi, \hat{\varphi})} \right] \varphi$$

contains only one-electron excitations, whence, from the above discussion, the first order correction vanishes.

Thus there will be no first order corrections to average dipole moments, average kinetic energy, average electron-nucleus energy, etc. Also, if $H^{(0)}$ contains all the interaction with external fields, our result will hold to <u>all</u> orders in the external fields and hence there will be no first order corrections to static polaribilities, static susceptibilities, etc.

In a general way one may say that there will be no first order corrections to the one-electron density matrix $\hat{\gamma}$. One can see this by arguing that the average value of W is given by

$$\mathrm{Av}\, W = \mathrm{Tr}\, \hat{\gamma} W$$

whence if there are no first order corrections to $\mathrm{Av}\, W$ for any W, there can be no first order corrections to $\hat{\gamma}$. Alternatively one can note that $\hat{\gamma}$ itself is the expectation value of a one-electron operator. (See for example McWeeny and Mizund, Proc. Roy. Soc. <u>A259</u>, 534 (1961). See also Somorjai, J. Chem. Phys. <u>44</u>, 3041 (1966) -- these authors work in configuration space. The result is most obvious when one uses second quantization. See for example Yang ,
Rev. Mod. Phys. <u>34</u>, 6946 (1962).)

In any case a further consequence is then that the eigenfunctions of $\hat{\gamma}$, the natural spin orbitals, as given by URHF (namely the $\hat{\chi}_i$)

and the eigenvalues of $\hat{\gamma}$ (namely $\frac{1}{N}$) are also correct through first order. (Actually the result concerning the eigenvalues is true whether or not $\hat{\Psi}$ is an optimal determinant -- see Kutzelnigg and Smith, J. Chem. Phys. 42, 2791 (1965) and references given there.)

More specifically let us consider an atom and let $H^{(0)} = H_{HF}$. Then the perturbation $H - H_{HF}$ is just the difference between the electron-electron repulsion terms in H and those in H_{HF}, terms which, as we have seen in our discussions of the "1/Z Method" are of the order $1/Z$ where Z is the nuclear charge. Thus in this case we can conclude that URHF yields expectation values of one-electron operators which are accurate through order $\frac{1}{Z}$, the errors being of order $\left(\frac{1}{Z}\right)^2$. Similar conclusions then follow as we have discussed for restricted HF for closed shell atoms. Moreover let us again emphasize that these conclusions are true to any order in external fields which may be present (for calculations illustrating these points we refer to a series of papers by Dalgarno and collaborators, especially Cohen, which have been published in recent years, mainly in Proc. Roy. Soc. and Proc. Phys. Soc.).

Having said all this we must now point out that the preceding argument contains a flaw and is not completely valid. The reason is connected with the peculiar degeneracy of hydrogenic energy levels. Thus for example consider Be in the absence of external fields (for explicit calculations, see the references given above). In the $Z \to \infty$ limit (neglect electron-electron interaction) $\hat{\Psi}$, which is a single determinant, becomes the single determinant of hydrogenic functions $(1s)^2(2s)^2\ {}^1S$. On the other hand the correct result is a certain linear combination of the degenerate pair $(1s)^2(2s)^2\ {}^1S$ and

$(1s)^2(2p)^2\,{}^1S$. (For more details see A. Dalgarno, Advan, in Phys. 11, 281 (1962) esp. P. 307 et. seq. and references there. For an "extended HF scheme" designed to avoid this difficulty see for example Cohen and Dalgarno, J. Mol. Spec. 10, 378 (1963) and subsequent papers.) This already indicates that something must be wrong with our argument since evidently we don't get correct results even in zero order, i.e., the leading term in the limit $Z\to\infty$. The point is simply that our arguments about orders fails is this case because of the fact that some of the energy denominators in (24) and its higher order analog, will vanish as $Z\to\infty$ and hence $\psi^{(n)}$ etc. will contain contributions of lower order in $1/Z$ than our original argument suggested. On the other hand for Li the $(1s)^2 2s\,{}^2S$ and $(1s)^2 2p\,{}^2P$ degeneracy causes no complication if there are no external fields since these states, because they have different angular momentum, don't get mixed. However it is still generally true that if we introduce an artificial order parameter λ and write $H(\lambda)= H_{HF}+\lambda(H-H_{HF})$ then corrections of order λ to one-electron properties do vanish without exception. However, in cases like Be the terms of order λ^n are not also of order $(1/Z)^n$. Similar remarks apply also in the presence of external fields (and then also Li becomes a case in point. See Drake + Cohen, J. Chem. Phys. 48 1168 (1968) and references given there).

In any case the arguments given above fail for two-electron operators so that in general one expects and finds that one-electron expectation values will be given more accurately by URHF than two-electron expectation values. Notice the energy operator H occupies an intermediate position here. Namely we know from the variational principle that its expectation value, like that of one-electron operators, contains only a second order error. On the other hand H involves both one- and two-electron operators. However the point is that the two-electron operator occurs multiplied by λ whence a λ error for the two-electron operator contributes a λ^2 error as far as the energy is concerned.

As another example consider SRAF. In a general way the extension is clear -- no first order corrections to the expectation values of spin-independent one-electron operators. In particular then for Li -like ions one expects no corrections of order $1/Z$ for spin-independent operators, but one does expect correction for spin-dependent operators. These expectations are borne out by the calculations (and the general theorem) of Dalgarno and Cohen in Proc. Roy. Soc. A275, 492 (1963). However it should be pointed out that our line of argument to this result is not completely clear at this point. Namely because, as we noted in a problem, the analogue of ϵ_{ts} , $t \neq s$ don't all vanish, it is not immediately obvious that we can write down an $H^{(0)}$ which is a one-electron operator and which is such that $H - H^{(0)}$ is order $1/Z$. We will return to this point in chapter .

In introducing $H^{(0)}$ we noted that $E^{(0)}$ may or may not equal $\hat{E}$. In particular if $H^{(0)} = H_{HF}$ then we know from eq. (IX-21) that $E^{(0)} = \sum \epsilon_s \neq \hat{E}$. Of course one can always add a constant to $H^{(0)}$ and some authors do this (notably Dalgarno) for Hartree-Fock theory. One theory for which $E^{(0)} = \hat{E}$ "naturally" is the shielding approximation. (Note that when $E^{(0)} = \hat{E}$, then $E^{(1)} = 0$). There you will recall one uses

$$H^{(0)} = T - \hat{\zeta} \sum_{s=1}^{N} \frac{1}{r_s} \tag{XI-27}$$

where T is the kinetic energy operator. The corresponding $\hat{\Phi}$ is a Slater determinant (or sum of Slater determinants) made up of hydrogenic spin orbitals, the effective charge $\hat{\zeta}$ having been chosen by use of the variational method.

Now since $\hat{\psi}$ is an eigenfunction of $H^{(0)}$ it satisfies the virial theorem for $H^{(0)}$, i.e.,

$$E^{(0)} = -(\hat{\psi}, T\hat{\psi})/(\hat{\psi},\hat{\psi})$$

On the other hand hydrogenic functions of nuclear charge $\tilde{\zeta}$ depend on $\vec{r}_i$ only as $\tilde{\zeta}\vec{r}_i$, i.e., we have scaling, whence the optimal functions will also satisfy the virial theorem for H, which for an atom, is

$$\hat{E} = -(\hat{\psi}, T\hat{\psi})/(\hat{\psi},\hat{\psi})$$

whence we have $E^{(0)} = \hat{E}$ as announced.

Problem: Consider a diatomic molecule. Choose for $H^{(0)}$ (20) plus nuclear repulsion, with $\hat{\zeta}$ again determined variationally. With $\hat{E}$ the optimal total energy including nuclear repulsion, show that at equilibrium $E^{(0)} = \hat{E}$. Hint: Note that $\tilde{\psi}$ does not involve R explicitly. For an extension to polyatomic molecules and references to related work see Gliemann, Theoret. Chem. Acta. 11, 75 (1968).

(E) Gauge Invariance

Two Hamiltonians H and H' which are related by a unitary transformation

$$H' = U H U^{-1}, \quad UU^{\dagger} = U^{\dagger}U = 1 \qquad \text{(XI-28)}$$

have the same eigenvalues. A natural question to ask is then, under what circumstances will variational calculations based on H and H' yield the same answers? One trivial answer is: if one uses trial functions $\tilde{\psi}$ with H, then use trial functions $U\tilde{\psi}$ with H' since

$$(\tilde{\psi}, H\tilde{\psi})/(\tilde{\psi},\tilde{\psi}) = (U\tilde{\psi}, H'U\tilde{\psi})/(U\tilde{\psi},U\tilde{\psi})$$

More interesting is the answer: Use the same set of trial functions for both, but choose the set of trial functions to be invariant under the transformation U. (Problem: Show that this works.)

As a specific example consider an atom or molecule in a magnetic field. Then as is well known a gauge transformation of the vector potential

$$\vec{A}(\vec{r}) \longrightarrow \vec{A}'(\vec{r}) = \vec{A}(\vec{r}) + \vec{\nabla}\Lambda(\vec{r}) \qquad \text{(XI-29)}$$

can be produced by unitary transformation

$$U = e^{\frac{i}{c}\sum_{s=1}^{N}\Lambda(\vec{r}_s)} \qquad \text{(XI-30)}$$

Thus we conclude that if the set of trial functions in invariant under the transformation (30), the results of calculations will be gauge invariant for all Λ. In particular, since U times a determinant is again a determinant with changed spatial dependence of the orbitals it follows that URHF and SRHF are gauge invariant for arbitrary Λ. (For actual calculations see for examples a series of papers by W. N. Lipscomb and collaborations in J. Chem. Phys. starting around 1964. See also his review article in "Advances in Magnetic Resonance" 2 (1966).

Problem: What can one say about the Probable gauge invariance of the linear variational method?

If a set of trial fucntions $\tilde{\Psi}$ is not invariant under U one might ask, is there an optimal gauge in which to do the calculations? One criterion which suggests itself is to require that $\mathcal{E}'$ (the optimal energy calculated with Ψ') be stationary with respect to (general or restricted) variations of the gauge. Evidently this is the same as using $U\tilde{\Psi}$ as the set of trial functions for H and treating Λ as a further variational function.

(F) Integral Hellmann-Feynman Theorem

We now consider two Hamiltonians H_x and H_y and correspondingly two optimal trial functions $\hat{\Psi}_x$ and $\hat{\Psi}_y$. Now let us suppose that

$$\Delta\hat{\Psi}_x = \delta\hat{\eta}\,\hat{\Psi}_y$$

and (XI-31)

$$\Delta\hat{\Psi}_y = \delta\hat{\eta}\,\hat{\Psi}_x$$

are allowed variations of $\hat{\Psi}_x$ and $\hat{\Psi}_y$ respectively for $\delta\hat{\eta}$ both real and pure imaginary. Then it follows that

$$(\hat{\Psi}_y, (H_x - \hat{E}_x)\,\hat{\Psi}_x) = 0$$

$$(\hat{\Psi}_y, (H_y - \hat{E}_y)\,\hat{\Psi}_x) = 0$$

whence by subtraction one finds the so-called "integral Hellmann   man Theorem"

$$(\hat{E}_x - \hat{E}_y) = \frac{(\hat{\Psi}_y, (H_x - H_y)\hat{\Psi}_x)}{(\hat{\Psi}_y, \hat{\Psi}_x)} \qquad \text{(XI-32)}$$

which has aroused some interest in the literature (see for example various papers by Parr and collaborators in J. Chem. Phys. starting around (1964)). In practice the one way which has been found to insure that the variations (31) are allowed is to draw $\hat{\Psi}_x$ and $\hat{\Psi}_y$ from a common linear space.

If the variations (31) are allowed then it follows that the right hand side of (32) will differ from the true eigenvalue difference $E_x - E_y$ by terms of second order in the errors in $\hat{\Psi}_x$ and $\hat{\Psi}_y$, since we know that $\hat{E}_x$ and $\hat{E}_y$ involve only second order errors. On the other hand if the conditions are not met, then the quantity on the right hand side will usually not equal $\hat{E}_x - \hat{E}_y$ and will contain first order errors. (For some representative numerical results see the papers by Parr referred to earlier, and Rothstein and Blender, J. Chem. Phys. 49, 1283 (1968). See also Lowe and Mazziotti, J. Chem. Phys. 48, 877 (1967).)

We could go on and introduce η's into this two Hamiltonian situation, however the formal method is clear so we will not pursue the matter further.

We said that in practice the one way which has been found to ensure that the variations (31) are allowed is to draw $\hat{\psi}_x$ and $\hat{\psi}_y$ from a common linear space. The way in which this has usually been done, and which allows the introduction of non-linear parameters to some extent, is as follows: Let $\phi(z)$ be an approximate wave function derived in some way, usually from the variational method applied to H_z using non-linear parameters, etc. To ensure that the variations (31) are allowed we then put (as a minimum)

$$\hat{\psi}_\lambda = \hat{a}_{\lambda x}\,\phi(x) + \hat{a}_{\lambda y}\,\phi(y)$$

Then we determine the linear variational parameters $\hat{a}_{\lambda x}$ and $\hat{a}_{\lambda y}$ from the variational method applied to H_λ. More generally (Hurley, Int. J. Q. Chem. 1s, 677 (1967)) if one wants to satisfy (32) for a continuous range of x and y values lying between a and b one can use (as a minimum)

$$\hat{\psi}_\lambda = \int_a^b \hat{a}_{\lambda\lambda'}\,\phi(\lambda')\,d\lambda'$$

The $\hat{a}_{\lambda\lambda'}$ being determined from the variational method applied to H_λ. (Note that with integrals instead of sums, the set of linear homogeneous algebraic equations becomes a homogeneous linear integral equation.)