THEORETICAL CHEMISTRY INSTITUTE

THE UNIVERSITY OF WISCONSIN

THE VARIATION METHOD - I

Saul T. Epstein

FACILITY FORM 602

N71-36518	
(ACCESSION NUMBER)	(THRU)
141	63
(PAGES)	(CODE)
CR-122953	#06
(NASA CR OR TMX OR AD NUMBER)	(CATEGORY)

WIS-TCI-431

(A revision of WIS-TCI-336)

19 February 1971

MADISON, WISCONSIN

1 STOCK LOCATION

2 DATE RECEIVED YR 71 MO 12 DAY 28

12 SCREEN
☐ REJECT
☐ OBTAIN BETTER COPY ☐ OUT OF PRINT SOD
☐ OBTAIN AUTHORITY ☐ ERRATA

17 ACCESSION NUMBER N71-36518

3 RECEIPT TYPE & FORMAT ☐ MAGNETIC TAPE
☐ LOAN ☒ PC ☐ 35 MM ☐ CARDS
☒ RETAIN ☐ MF ☐ 16 MM ☐ OTHER

13A ANNOUNCEMENT VOL 72 ISSUE 02 ☐ USG ☐ TAB ☒ UNANN

13B FAS ☐ YES ☒ NO

18 PAGES 141

19 SHEETS —

20 LOW LIMIT PC MF

21 ☐ SUBSCRIPTION

4 STOCK RECEIVED FOR SALE PC 0 MF

14 REPRODUCTION INSTRUCTIONS
BLOWBACK →

PRINT ↓	NO	1UP	2UP
NO	1	4	7
1UP	2	[✓ 5]	8
2UP	3	6	9
MIX	SAME SIZE	ORDER STOCK FROM	0

MAKE MICROFICHE ☐ YES ☒ NO

15 PRESTOCK

	NO1	# COPIES
PC	X	
MF		M
PC DUE IN		

SOURCE TO ORDER COPIES

22 PRICES
☒ U UNIT ☐ P PC →
☐ E PC + MN BOX 16 ☐ M MN →

23A CATEGORY

23B CATEGORY

24 DISTR CODE —

25 INITIALS ACC VH A B

5 LOAN DOCUMENT RETURNED
DUE
OUT

6 TRANSACTION
NEW ☐ ITEM
DUPE ☒
SUPERSEDES ☐
PRIOR ☐ NUMBER

26 FILL FROM
PAPER COPY ETC BW
☐ ↓
MICRO-NEGATIVE XN

27 PUBLIC RELEASABILITY C

7 PB-198 667

8 REPORT NUMBERS (XREF) ☐ →
WIS-TCI-431

16 REMARKS

9 RELATED DOCUMENT ☐ →

10 CONTRACTING OFFICE — BILLING CODE
NASA –

11 NOT FULLY
LEGIBLE ☐
COLOR ☐

28 FORM & PRICE

1 ARCHIVES

12/28

FORM NTIS-77
(10/70)
DOCUMENT TRAVELER

NATIONAL TECHNICAL INFORMATION SERVICE
U. S. DEPARTMENT OF COMMERCE

THE VARIATION METHOD - I.*

by

Saul T. Epstein
Theoretical Chemistry Institute
and
Physics Department
University of Wisconsin
Madison, Wisconsin 53706

(I would be most grateful for any comments, or any suggestions for improvement).

---

* Research supported by Grant GP-12832 from the National Science Foundation and National Aeronautics and Space Administration Grant NGL 50-002-001.

INTRODUCTION

In his first paper on wave mechanics Schroedinger[1] presented his famous equation in the form of a variation principle, indeed just the variation principle which we will be discussing in the next section. Thus our subject had deep roots in quantum mechanics, and of course, the general use of variation principles goes back much further.[2] Similarly the variation method, the general approximation method based on the variation principle, which we will be discussing in detail in subsequent sections, is one of the pillars of applied quantum mechanics since most approximation procedures are either direct applications of the variation method, or can be related to it in one way or another (and of course the use of variation methods to approximate the solution of physical problems has an even longer history).

Finally, if our choice of subject is in need of further justification, let us note that in recent years Ruedenberg[3] has shown that by taking the variation principle rather literally and imagining that as a molecule forms it actually does, so to speak, try one wave function and then another, relaxing a bit here, tightening a bit there, before finding the most suitable wave function, one can get real insight into the nature of chemical binding.

I. POSITIVE HERMITIAN OPERATORS

A positive Hermitian operator is a Hermitian operator whose eigenvalues are all non-negative. For our purposes the most important consequence of this is that the expectation value of such an operator is always non-negative. Proof: Let 0 be a positive operator, φ_i a complete orthonormal set of its eigenfunctions, and θ_i its eigenvalues. Then, using a discrete notation, $(\psi, O\psi) = \sum_i \theta_i \,|(\varphi_i, \psi)|^2$ which is ≥ 0, since by assumption the θ_i are ≥ 0.

We will also have use for the following trivial extension of these ideas: Suppose that although O is not positive, we deal only with ψ's such that $(\varphi_i, \psi) = 0$ unless $\theta_i \geq 0$. Then clearly we still have $(\psi, O\psi) \geq 0$ for all such ψ. We will say that such an O is positive with respect to the functions ψ.

II. THE VARIATION PRINCIPLE

Given any function $\tilde{\psi}$ for which the requisite integrals exist[A] (we will refer to such functions as "trial functions") we can calculate the real number

$$\tilde{E} \equiv (\tilde{\psi}, H\tilde{\psi})/(\tilde{\psi}, \tilde{\psi}) \qquad \text{(II-1)}$$

Evidently $\tilde{E}$ would be the average energy of the system if the system were in the state described by the function $\tilde{\psi}$.[B] Similarly if $\overset{\times}{\psi}$ is another trial function we can calculate the corresponding average energy

$$\overset{\times}{E} \equiv (\overset{\times}{\Psi}, H\overset{\times}{\Psi}) / (\overset{\times}{\Psi}, \overset{\times}{\Psi}) \tag{II-2}$$

One property of $\tilde{E}$ and $\overset{\times}{E}$ is clear immediately: since each is an average energy, neither can be less than the smallest possible energy, that is neither can be smaller than the smallest eigenvalue of H. We will return to this point in a moment. To derive other properties of $\tilde{E}$ and $\overset{\times}{E}$ we now write $\tilde{\Psi}$ as

$$\tilde{\Psi} = \overset{\times}{\Psi} + \tilde{\Delta} \tag{II-3}$$

thereby defining $\tilde{\Delta}$. Then using (1) and (2) we find

$$\tilde{E} = \frac{\overset{\times}{E}(\overset{\times}{\Psi}, \overset{\times}{\Psi}) + (\overset{\times}{\Psi}, H\tilde{\Delta}) + (\tilde{\Delta}, H\overset{\times}{\Psi}) + (\tilde{\Delta}, H\tilde{\Delta})}{(\tilde{\Psi}, \tilde{\Psi})} \tag{II-4}$$

However from (3), $(\overset{\times}{\Psi}, \overset{\times}{\Psi}) = (\tilde{\Psi}, \tilde{\Psi}) - (\tilde{\Psi}, \tilde{\Delta}) - (\tilde{\Delta}, \tilde{\Psi}) - (\tilde{\Delta}, \tilde{\Delta})$ so we can write (4) as

$$\tilde{E} = \overset{\times}{E} + \frac{(\overset{\times}{\Psi}, (H - \overset{\times}{E})\tilde{\Delta}) + (\tilde{\Delta}, (H - \overset{\times}{E})\overset{\times}{\Psi}) + (\tilde{\Delta}, (H - \overset{\times}{E})\tilde{\Delta})}{(\tilde{\Psi}, \tilde{\Psi})} .$$

Finally we use the Hermiticity of $(H - \overset{\times}{E})$ to find

$$\tilde{E} = \overset{\times}{E} + \frac{((H - \overset{\times}{E})\overset{\times}{\Psi}, \tilde{\Delta}) + (\tilde{\Delta}, (H - \overset{\times}{E})\overset{\times}{\Psi}) + (\tilde{\Delta}, (H - \overset{\times}{E})\tilde{\Delta})}{(\tilde{\Psi}, \tilde{\Psi})} \tag{II-5}$$

We will now draw several important conclusions from this result.

[1] Suppose that

$$(H - \overset{\times}{E})\overset{\times}{\Psi} = 0 \tag{II-6}$$

that is, suppose that $\overset{\times}{\Psi}$ and $\overset{\times}{E}$ are an eigenfunction and the corresponding eigenvalue of H. Then (5) becomes

$$\tilde{E} = \overset{\times}{E} + \frac{(\tilde{\Delta}, (H-\overset{\times}{E})\tilde{\Delta})}{(\tilde{\Psi}, \tilde{\Psi})} \tag{II-7}$$

which tells us, among other things, that when $\tilde{\Delta}$ is small[c] so that $\tilde{\Psi}$ is nearly the eigenfunction $\overset{\times}{\Psi}$, then $\tilde{E}$ differs from the corresponding eigenvalue $\overset{\times}{E}$ by terms which are at least of second order in $\tilde{\Delta}$. Therefore the eigenvalues of H are stationary points of $\tilde{E}$ as a functional of $\tilde{\Psi}$.

[2] We will now show that $\tilde{E}$ has no other stationary points. Thus suppose that $\overset{\times}{E}$ is a stationary point of $\tilde{E}$ as a functional of $\tilde{\Psi}$. This then requires that the first order term in (5)

$$((H-\overset{\times}{E})\overset{\times}{\Psi}, \tilde{\Delta}) + (\tilde{\Delta}, (H-\overset{\times}{E})\overset{\times}{\Psi})$$

must vanish for all sufficiently small $\tilde{\Delta}$ and hence in particular must vanish for

$$\tilde{\Delta} = \epsilon\, (H-\overset{\times}{E})\overset{\times}{\Psi}$$

where ϵ is an arbitrarily small real number. Thus we must have

$$((H-\overset{\times}{E})\overset{\times}{\Psi}, (H-\overset{\times}{E})\overset{\times}{\Psi}) = 0$$

which can be satisfied only if

$$(H-\overset{\times}{E})\overset{\times}{\Psi} = 0$$

Therefore we have the result that if $\overset{\times}{E}$ is a stationary point then $\overset{\times}{E}$ is an eigenvalue and the corresponding $\overset{\times}{\Psi}$ is an eigenfunction. The characterization of the eigenvalues and eigenfunctions of H provided by [1] and [2] constitutes a statement of the variation principle.

[3] Suppose now that $\overset{\times}{E}$ is the smallest eigenvalue of H. Then $(H-\overset{\times}{E})$ has only non-negative eigenvalues and thus is a

positive operator. In particular then this means that however large $\tilde{\Delta}$ may be

$$(\tilde{\Delta}, (H-\overset{\times}{E})\tilde{\Delta}) \geq 0$$

and hence, from (7) we see, as we noted at the outset, that <u>the lowest</u> eigenvalue of H is the absolute minimum of $\tilde{E}$ as a functional of $\tilde{\Psi}$.[D] On the other hand if $\overset{\times}{E}$ is not the smallest eigenvalue then by choosing $\tilde{\Delta}$ to be an arbitrarily small linear combination of the lower (higher) eigenfunctions we can make $(\tilde{\Delta}, (H-\overset{\times}{E})\tilde{\Delta})$ less than (greater than) zero. Thus the higher eigenvalues of H are only stationary points of $\tilde{E}$ as a functional of $\tilde{\Psi}$, and are neither maxima or minima.

[4] That the lowest eigenvalue is an absolute minimum of $\tilde{E}$ is a very striking result. However it does not in general serve to characterize the energies of the ground states of atoms or molecules since, because of the requirements of the Pauli Principle, these ground states are usually not the lowest states of the Hamiltonian; for example the ground state of the lithium atom is $(1s)^2 2s$ and not $(1s)^3$.

Happily however there is a similar theorem which is applicable to physical ground states and to various excited states as well. Namely suppose that H commutes with certain commuting operators $\mathcal{T}$ so that we can form a complete orthonormal set of eigenfunctions of H which are also eigenfunctions of $\mathcal{T}$, i.e., can be labelled by the eigenvalues of $\mathcal{T}$. We will say that functions having the same $\mathcal{T}$ quantum numbers have the same symmetry, and we will say that the two functions which have different $\mathcal{T}$ quantum numbers, and which are therefore orthogonal to one another, have different symmetry. (We will have no need to compare functions associated with different sets $\mathcal{T}$.) If then $\overset{\times}{E}$ is the smallest eigenvalue of H for states of a given symmetry (for example for states satisfying the Pauli Principle) it follows that $(H-\overset{\times}{E})$ will be a positive operator with respect to functions of that symmetry because clearly such functions will be orthogonal to all the lower eigenfunctions of $(H-\overset{\times}{E})$.

If now we confine attention to $\tilde{\Psi}$ with the given symmetry, then $\tilde{\Delta}=\tilde{\Psi}-\overset{\times}{\Psi}$ will also have that symmetry and therefore however large $\tilde{\Delta}$ may be, still

$$(\tilde{\Delta},(H-\tilde{E})\tilde{\Delta})\geq 0$$

Thus we have the result that the lowest eigenvalue of H associated with a given symmetry is the absolute minimum of $\tilde{E}$ as a functional of trial functions of that symmetry.

[5] We now note that if H commutes with $\mathcal{T}$ and if $\overset{\times}{\Psi}$ has a definite symmetry then the variation $\epsilon(H-\overset{\times}{E})\overset{\times}{\Psi}$ which played the decisive role in [2] will have that same symmetry. Thus we may generalize the result found there as follows: If H commutes with $\mathcal{T}$ and if $\overset{\times}{\Psi}$ has a certain symmetry, and if $\overset{\times}{E}$ is stationary with respect to all variations of that symmetry, then $\overset{\times}{\Psi}$ is an eigenfunction and $\overset{\times}{E}$ is the associated eigenvalue. In short, combining this last result with [1], the variation principle applies separately to each symmetry type.

[6] As a generalization of [4] we have the following: Let $\overset{\times}{E}$ be an arbitrary eigenvalue of H and confine attention to $\tilde{\Psi}$ which are orthogonal to all eigenfunctions of H whose associated

eigenvalues are less than $\overset{\times}{E}$. Then clearly we will have

$$(\tilde{\Delta}, (H-\overset{\times}{E})\tilde{\Delta}) \geq 0$$

and thus *an arbitrary eigenvalue of H is an absolute minimum of $\tilde{E}$ as a functional of trial functions orthogonal to eigenfunctions associated with smaller eigenvalues*.

III. THE VARIATION METHOD

The results of the previous section are of great practical importance because they suggest a soundly based method for approximating the eigenvalues and eigenfunctions of H . According to the variation principle we can find the eigenvalues and eigenfunctions of H by calculating $\tilde{E}$ for all $\tilde{\Psi}$, and then looking for stationary points. In practice this is usually impossible - one cannot examine *all* $\tilde{\Psi}$. However what one *can* do is to examine a restricted class of trial functions, a class no larger than one can handle, and then take the stationary points of $\tilde{E}$ *within this restricted class* as *approximations* to the eigenfunctions and eigenvalues of H .

This procedure is known as the *variation method*. We will call the $\tilde{\Psi}$ which yield stationary values, optimal trial functions and denote them by $\hat{\Psi}$, possibly with a subscript. The corresponding $\tilde{E}$ we will denote by $\hat{E}$, again possibly with a subscript.

We said that this is a soundly based method. To support this assertion consider first the lowest state of a given symmetry. Then [4] of Section I tells us that we have a good approximation scheme in that it is capable of systematic improvement. Namely if we enlarge the class of trial functions (assumed to be of appropriate symmetry) then the minimum $\hat{E}$ will almost certainly decrease (in any case it cannot increase) whence, from [4] we will have a better approximation to the energy. Note also that we have a quadratic convergence to the eigenvalue in the sense that, as follows from [1], the error in the eigenvalue is of second order in the error of the eigenfunction.

For the higher states of a given symmetry the situation at this point is not so clear. Result [6] of Section I is of little practical use since one usually cannot guarantee the required orthogonality. We can of course say that if we enlarge the class of trial functions we will make the higher $\hat{E}$ "more stationary", but this may or may not represent a numerical improvement. However in a later section we will discuss a practical way of choosing trial functions (the linear variation method) which does permit a systematic improvement in the approximation to higher eigenvalues.

Even from these brief remarks it should be clear that the variational approximation to eigenvalues is a soundly based one. It is harder to make a definite statement about the quality of the eigenfunction approximation, mainly because there are so many figures of merit which one might use -- the overlap between the approximation and the eigenfunction, the accuracy of particular expectation values, the energy variance $(\hat{\Phi}, (H-\hat{E})^2\hat{\Phi})/(\hat{\Phi}, \hat{\Phi})$ etc. etc. We will not attempt a quantitative discussion of these many possibilities but we will return to these questions from time to time in the succeeding sections.

In a general way however one usually says that the approximation to the eigenvalue furnished by $\hat{E}$ is better than the approximation to the eigenfunction furnished by $\hat{\Psi}$ because, as we have already noted, the error in the former is of second order in the error in the latter. In this connection though it should be kept in mind that to some extent "order" is a theoretical concept, and that second order quantities are guaranteed to be smaller than first order quantities only if the order parameter is "sufficiently" small. Thus $A x^2$ is less than x only for $x < 1/A$.

Also it should be admitted that the preceding discussion of energy is directly relevant only for very light systems since usually it is only in such cases that total energies are of immediate interest. Rather one is usually interested in comparatively small energy differences, excitation energies, ionization energies, changes in molecular energy with nuclear configuration, etc. Therefore since the difference of two upper bounds is not in general a bound, and since improving the individual upper bounds will not necessarily improve the difference, improvement of the accuracy of the total energy is not of immediate concern.[A] (Of course the difference between an upper bound and a lower bound is an upper bound but that is another story). These considerations might then lead one to a pessimistic view of the applicability of the variation method to atoms and molecules. However in practice the opposite situation prevails - differencing of results of only moderate individual accuracy often giving results of even very high accuracy. In some cases this can be understood as a cancellation of obvious common errors[1] but in other cases, for example in recent calculation on He_2,[2] the process is by no means well understood.

IV. THE VARIATION METHOD: MORE DETAILS

In general the set of trial functions will be labelled by arbitrary numerical parameters and/or arbitrary functions. To implement the variation method then what one does in principle, and often in practice, is to calculate $\tilde{E}$ as a function of the variational parameters and/or functions and then determine their optimal values by setting equal to zero the derivatives of $\tilde{E}$ with respect to each parameter and function. This approach, when carried out formally, yields a set of equations which must be solved to determine the optimal values of the variational parameters and/or functions. We should however point out that in many practical calculations which must be largely numerical rather than analytical, such equations are often partially or completely bypassed in favor of some sort of direct numerical search procedure to locate the stationary points of $\tilde{E}$.

For theoretical purposes, and sometimes also for practical purposes, it is, however, convenient to proceed a little more indirectly in the formal discussions. Starting from a given $\tilde{\Psi}$, suppose that we change the parameters and/or functions infinitesimally in some way so that we go from $\tilde{\Psi}$ to a "neighboring" function. If we denote the first order change in $\tilde{\Psi}$, the <u>variation</u> in $\tilde{\Psi}$, by $\delta\tilde{\Psi}$, and if we write (II-1) as

$$(\tilde{\Psi}, (H-\tilde{E})\tilde{\Psi}) = 0 \tag{IV-1}$$

then we see that $\delta\tilde{E}$, the first order change in $\tilde{E}$, is determined by

$$(\delta\tilde{\Psi}, (H-\tilde{E})\tilde{\Psi}) + (\tilde{\Psi}, (H-\tilde{E})\delta\tilde{\Psi}) - \delta\tilde{E}(\tilde{\Psi}, \tilde{\Psi}) = 0 \tag{IV-2}$$

Now the $\hat{\Psi}$ are those $\tilde{\Psi}$ which make $\tilde{E}$ stationary with respect to all variations possible within the set. Thus we must have

$$(\delta\hat{\Psi}, (H-\hat{E})\hat{\Psi}) + (\hat{\Psi}, (H-\hat{E})\delta\hat{\Psi}) = 0 \qquad \text{"all" } \delta\hat{\Psi} \tag{IV-3}$$

where the quotation marks are to remind us that we are requiring that equation (3) hold only for variations within the set. We will sometimes write (3) as

$$\delta(\hat{\Psi}, (H-\hat{E})\hat{\Psi}) = 0 \tag{IV-4}$$

without explicitly stating the qualifications $\delta\hat{E} = 0$ and "all" $\delta\hat{\Psi}$.

Equation (3) together with

$$(\hat{\Psi}, (H-\hat{E})\hat{\Psi}) = 0 \tag{IV-5}$$

are then the equations to be used to determine the $\hat{\Psi}$ and $\hat{E}$ and these are the equations which we will use to characterize the variation method. An obvious procedure at this point would be to eliminate $\hat{E}$ from (3) by means of (5), solve the resultant equations for $\hat{\Psi}$ and then return to (5) to determine $\hat{E}$; and indeed this just leads to the straightforward procedure which we outlined at the beginning of this section. To see this, and also to make our notation a bit clearer, suppose for example that the set of trial functions is labelled by a set of N (independent) real parameters.

$\tilde{a}_1, \tilde{a}_2, \text{----}\ \tilde{a}_M$. That is

$$\tilde{\psi} = \phi(\tilde{a}_1 \cdots \tilde{a}_M) \equiv \phi(\tilde{a})$$

For example for a single particle we might use $\tilde{\psi} = \exp -(\tilde{a}_1 x^2 + \tilde{a}_2 y^2 + \tilde{a}_3 z^2)$

Then evidently the most general $\delta\hat{\psi}$ is

$$\delta\hat{\psi} = \sum_{i=1}^{M} \frac{\partial\phi(\hat{a})}{\partial\hat{a}_i}\,\delta\hat{a}_i$$

with the $\delta\hat{a}_i$ real but otherwise arbitrary. Inserting this into (3) we then have

$$\sum_{i=1}^{M}\left[\left(\frac{\partial\phi(\hat{a})}{\partial\hat{a}_i}, (H-\hat{E})\,\phi(\hat{a})\right) + \left(\phi(\hat{a}), (H-\hat{E})\frac{\partial\phi(\hat{a})}{\partial\hat{a}_i}\right)\right]\delta\hat{a}_i = 0 \qquad \text{(IV-6)}$$

which, we will now show implies that

$$\left(\frac{\partial\phi(\hat{a})}{\partial\hat{a}_j}, (H-\hat{E})\,\phi(\hat{a})\right) + \left(\phi(\hat{a}), (H-\hat{E})\frac{\partial\phi(\hat{a})}{\partial\hat{a}_j}\right) = 0\ ;\ j=1 \cdots M \qquad \text{(IV-7)}$$

Proof: That conditions (7) are sufficient to satisfy (6) is obvious. That they are necessary follows from the observation that since the $\delta\hat{a}_i$ are arbitrary we can choose $\delta\hat{a}_i = 0$ for $i \neq j$ and $\delta\hat{a}_j \neq 0$. Then (6) yields (7) directly.

Let us now calculate $\partial\hat{E}/\partial\hat{a}_j$. From (5) we have

$$\frac{\partial\hat{E}}{\partial\hat{a}_j} = \frac{\partial}{\partial\hat{a}_j}\frac{(\phi(\hat{a}), H\phi(\hat{a}))}{(\phi(\hat{a}),\phi(\hat{a}))} = \frac{\left(\frac{\partial\phi(\hat{a})}{\partial\hat{a}_j}, H\phi(\hat{a})\right) + \left(\phi(\hat{a}), H\frac{\partial\phi(\hat{a})}{\partial\hat{a}_j}\right)}{(\phi(\hat{a}),\phi(\hat{a}))}$$

$$- \frac{(\phi(\hat{a}), H\phi(\hat{a}))}{(\phi(\hat{a}),\phi(\hat{a}))^2}\left[\left(\frac{\partial\phi(\hat{a})}{\partial\hat{a}_j}, \phi(\hat{a})\right) + \left(\phi(\hat{a}), \frac{\partial\phi(\hat{a})}{\partial\hat{a}_j}\right)\right]$$

$$\frac{\partial \hat{E}}{\partial \hat{\alpha}_j} = \text{FIRST LINE} - \hat{E}\left[\left(\frac{\partial \phi(\hat{\alpha})}{\partial \hat{\alpha}_j}, \phi(\hat{\alpha})\right) + \left(\phi(\hat{\alpha}), \frac{\partial \phi(\hat{\alpha})}{\partial \hat{\alpha}_j}\right)\right] \Big/ (\phi(\hat{\alpha}), \phi(\hat{\alpha}))$$

and hence finally

$$\frac{\partial \hat{E}}{\partial \hat{\alpha}_j} = \left[\left(\frac{\partial \phi(\hat{\alpha})}{\partial \hat{\alpha}_j}, (H-\hat{E})\, \phi(\hat{\alpha})\right) + \left(\phi(\hat{\alpha}), (H-\hat{E})\, \frac{\partial \phi(\hat{\alpha})}{\partial \hat{\alpha}_j}\right)\right] \Big/ (\phi(\hat{\alpha}), \phi(\hat{\alpha}))$$

Comparing this result with (7) we see, that as we stated earlier, equation (7) with (5) are equivalent to

$$\frac{\partial \hat{E}}{\partial \hat{\alpha}_j} = 0 \ ; \quad j = 1 \ldots M. \qquad \text{(IV-8)}$$

We have implied that the use of equation (3) offers certain advantages over the use of equation (8). As a first illustration of this point we now remark that often (5) is a special case of (3), so that (3) alone then suffices to completely characterize the variation method. Namely frequently (the linear variation method which we will discuss in detail in subsequent sections is an important case in point) a set of trial functions will have no fixed overall scale, that is if $\tilde{\Psi}$ is a member of the set then so is $A\tilde{\Psi}$ where A is an arbitrary constant. In such cases then, among the neighbors of $\hat{\Psi}$ in the set will be $(1+\delta A)\hat{\Psi}$ where δA is a small real constant, which in turn implies that (3) must be satisfied by $\delta\hat{\Psi} = \delta A\,\hat{\Psi}$. Inserting this $\delta\hat{\Psi}$ into (3) and cancelling a factor of δA then yields (5) which proves the point.

V. THE VARIATION METHOD AND MOMENTS OF THE SCHRODINGER EQUATION

Using the Hermiticity of $(H-\hat{E})$, Eq. (IV-3) can be written as

$$(\delta\hat{\Psi}, (H-\hat{E})\hat{\Psi}) + ((H-\hat{E})\hat{\Psi}, \delta\hat{\Psi}) = 0, \quad \text{"all"}\ \delta\hat{\Psi} \tag{V-1}$$

or

$$\mathrm{Re}\,(\delta\hat{\Psi}, (H-\hat{E})\hat{\Psi}) = 0, \quad \text{"all"}\ \delta\hat{\Psi} \tag{V-2}$$

We now note that if, as is often the case in practice, there are no *à priori* reality conditions on the variational parameters and/or functions then the equations (1) are equivalent to the seemingly stronger equations

$$(\delta\hat{\Psi}, (H-\hat{E})\hat{\Psi}) = 0, \quad \text{"all"}\ \delta\hat{\Psi} \tag{V-3}$$

To see this let us suppose that the $\tilde{\Psi}$ are labelled by a single arbitrary function $\tilde{\chi}$, the generalization to several functions and/or parameters then being obvious. Then in general

$$\delta\hat{\Psi} = \frac{\partial\hat{\Psi}}{\partial\hat{\chi}}\delta\hat{\chi}$$

so that (1) becomes

$$\left(\frac{\partial\hat{\Psi}}{\partial\hat{\chi}}\delta\hat{\Psi}, (H-\hat{E})\hat{\Psi}\right) + \left((H-\hat{E})\hat{\Psi}, \frac{\partial\hat{\Psi}}{\partial\hat{\chi}}\delta\hat{\chi}\right) = 0, \quad \text{all}\ \delta\hat{\chi} \tag{V-4}$$

where, of course, there is no need to qualify "all". That there are no *à priori* reality restrictions on $\hat{\chi}$ means that we can vary its real and imaginary parts separately. Therefore we may replace (4) by

$$\left(\frac{\partial\hat{\Psi}}{\partial\hat{\chi}}(\delta\hat{\chi})_R,\,(H-\hat{E})\hat{\Psi}\right)+\left((H-\hat{E})\hat{\Psi},\,\frac{\partial\hat{\Psi}}{\partial\hat{\chi}}(\delta\hat{\chi})_R\right)=0\;;\;\text{all }(\delta\hat{\chi})_R \qquad \text{(V-5)}$$

and

$$-\left(\frac{\partial\hat{\Psi}}{\partial\hat{\chi}}(\delta\hat{\chi})_I,\,(H-\hat{E})\hat{\Psi}\right)+\left((H-\hat{E})\hat{\Psi},\,\frac{\partial\hat{\Psi}}{\partial\hat{\chi}}(\delta\hat{\chi})_I\right)=0\;,\;\text{all }(\delta\hat{\chi})_I \qquad \text{(V-6)}$$

where $(\delta\hat{\chi})_R$ and $(\delta\hat{\chi})_I$ are the real and imaginary parts of $\delta\hat{\chi}$ and where in (6) we have cancelled out a factor of i. We now derive (3) as follows: (6) is to be true for all $(\delta\hat{\chi})_I$ and therefore in particular, for a given $(\delta\hat{\chi})_R$, (6) must be true with $(\delta\hat{\chi})_I = (\delta\hat{\chi})_R$. Making this substitution in (6) and comparing with (5) then immediately yields

$$\left(\frac{\partial\hat{\Psi}}{\partial\hat{\chi}}(\delta\hat{\chi})_R,\,(H-\hat{E})\hat{\Psi}\right)=0\;,\;\text{all }(\delta\hat{\chi})_R \qquad \text{(V-7)}$$

Similarly for a given $(\delta\hat{\chi})_I$, (5) must be satisfied with $(\delta\hat{\chi})_R = (\delta\hat{\chi})_I$ which then leads to

$$\left(\frac{\partial\hat{\Psi}}{\partial\hat{\chi}}(\delta\hat{\chi})_I,\,(H-\hat{E})\hat{\Psi}\right)=0\quad\text{all }(\delta\hat{\chi})_I \qquad \text{(V-8)}$$

If now we multiply (8) by (−i) and add to (7) then the result is (3).

Having gone through this in detail it is now useful to note the following quick derivation: If there are no _a priori_ reality restrictions then if $\delta_1\hat{\Psi}$ is a possible variation of $\hat{\Psi}$ within the set then so is $\delta_2\hat{\Psi} = i\delta_1\hat{\Psi}$ ..

(Proof: Choose $(\delta_2\hat{\psi})_R = -(\delta_1\hat{\psi})_R$, $(\delta_2\hat{\psi})_I = (\delta_1\hat{\psi})_R$). Therefore (1) must also be satisfied if we replace $\delta\hat{\psi}$ by $i\delta\hat{\psi}$. Doing this and cancelling a factor of i then yields

$$-(\delta\hat{\psi}, (H-\hat{E})\hat{\psi}) + ((H-\hat{E})\hat{\psi}, \delta\hat{\psi}) = 0 \qquad \text{(V-9)}$$

Comparison with (1) then yields (3). Finally we note that (1) and (3) are trivially equivalent if H is explicitly real and if one restricts oneself to real trial functions.

When (3) applies it provides an interesting and suggestive interpretation of the variation method. In a general way, given a function F , quantities of the form

$$(G, F)$$

for various choices of G, are referred to as "Moments" of F . Thus we can say that when (3) applies, the variation method approximates making $(H-\hat{E})\hat{\psi} = 0$ by requiring the vanishing of a restricted set of moments of $(H-\hat{E})\hat{\psi}$ (Note that the other basic equation, $(\hat{\psi}, (H-\hat{E})\hat{\psi}) = 0$, is also in moment form).

The approximation of requiring only that certain moments of $(H-\hat{E})\hat{\psi}$ vanish, is certainly one which one might come upon, and indeed one which people have come upon, without reference to the variation method. In particular consider the linear variation method (which we will discuss in more detail in succeeding sections) in which the set of trial functions consists of functions of the form

$$\tilde{\Psi} = \sum_{L=1}^{M} \tilde{a}_L \phi_L \tag{V-10}$$

where the ϕ_L (the "basis set") are a given set of linearly independent functions, and where the $\tilde{a}_L$ are arbitrary parameters.

If no reality conditions are imposed on the $\tilde{a}_L$, then (3) applies so that with

$$\hat{\Psi} = \sum_{L=1}^{M} \hat{a}_L \phi_L \quad , \quad \delta\hat{\Psi} = \sum_{L=1}^{M} \delta\hat{a}_L \phi_L \quad , \tag{V-11}$$

we have

$$\sum_{K=1}^{M} \sum_{L=1}^{M} \delta\hat{a}_K^{*} \left(\phi_K, (H-\hat{E})\phi_L\right) = 0 \quad , \text{all } \delta\hat{a}_K^{*}$$

and therefore (recall the proof following (IV-7)

$$\sum_{L=1}^{M} \left(\phi_K, (H-\hat{E})\phi_L\right) \hat{a}_L = 0 \qquad K=1 \ldots M. \tag{V-12}$$

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 "Schrodinger Equation" (the reason for the " " will be discussed in a moment):

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

and then simply taking the scalar product with each ϕ_K in turn.

This sort of approach to the derivation of equations (11) suggests other possibilities. Since the use of $\hat{}$ has special reference to the variation method let us consider the more neutral "equation"

$$(H-\check{E})\sum_{L=1}^{M}\check{a}_L\phi_L=0 \qquad \text{(V-13)}$$

Then we note that although the procedure of "taking the scalar ------ " provides one way of trying to determine $\check{E}$ and the $\check{a}_L$, there are other possibilities. For example one might try to satisfy (13) identically at M selected points, or more generally one might try multiplying through by quite another set of M functions and integrate to find

$$\sum_{L=1}^{M}(\chi_K,(H-\check{E})\phi_L)\check{a}_L=0 \qquad k=1\ldots M. \qquad \text{(V-14)}$$

[Evidently this reduces to the second suggestion if the χ_K are Dirac delta functions]. These observations of course raise questions as to the relative status of these various approaches. Are they equivalent? Is one superior to the other? First as to the equivalence: In general the different procedures (different choices for the set of χ_K) will lead to different answers. The point is simply that (13) as it stands is almost certainly an inconsistent equation - there are no $\check{a}_L$ and $\check{E}$ which satisfy it (hence our use of " "), or more precisely, it is a consistent equation only if there happens to be an eigenfunction of H which can be written as a linear combination of the ϕ_L . Since in practice in a complicated problem this is unlikely, we may take it that "equation" (13) is not consistent and, hence it follows that 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 variation method leads to (2) and therefore, as we know, this endows it with the virtue that the lowest $\hat{E}$ is a guaranteed upper bound to the lowest eigenvalue of H of appropriate symmetry. Indeed, as we shall see in the next section, it is even more virtuous: the $\hat{E}$ which are solutions of (2) are, in order, guaranteed upper bounds to the M lowest eigenvalues of H of appropriate symmetry. Thus there is considerable reason to choose (2). However recently there has been a revival of interest in the use of equations of the form

$$(\chi_k, (H-\check{E})\check{\Psi}) = 0 \qquad k=1,\ldots \qquad \text{(V-15)}$$

where $\check{\Psi}$ may be of the form (10), but may also be of a much more complicated structure, and where the χ_k may be given functions, or given functions multiplying operators, or may involve some of the arbitrary parameters and/or functions in $\check{\Psi}$ which are to be determined from equations (15). In any case the reason for the interest is quite simply that with the forms of $\hat{\Psi}$ which are in use (or which one would like to use) in the applications of (3) to atoms and molecules, the integrals in (13) are often quite difficult (or impossible in practice) whereas with a $\check{\Psi}$ of similar form and with a suitable choice of the χ_k, the integrals in (15) are quite tractable. We will not discuss such methods further here but instead will refer the interested reader to the original literature.[1-] We would emphasize however, that such methods do not in general yield bounds.

VI. THE LINEAR VARIATION METHOD

Let us now return to Eq. (V-12). This is a set of linear homogeneous equations to determine the $\hat{a}_K$. It has non-trivial solutions (that is not all $\hat{a}_K \equiv 0$) only for certain values of $\hat{E}$, those for which the determinant of coefficients (the "secular determinant") vanishes

$$\left| (\phi_K, (H-\hat{E})\phi_L) \right| = 0 \tag{VI-1}$$

Equation (1), the "secular equation", is an M'th order algebraic equation to determine $\hat{E}$ A; and incidently note that in accordance with the discussion at the end of Sec. IV, we have not had to invoke (IV-5) explicitly since the set of $\tilde{\Psi}$ which we are using clearly has no fixed overall scale. We will denote the roots of (1) by $\hat{E}_K$ with $K=1,2,\ldots M$ and $\hat{E}_1 \leq \hat{E}_2 \leq$ -------- . Similarly we will denote the corresponding $\hat{\Psi}$ by $\hat{\Psi}_K$.

The set of trial functions (V-10) has the special property of forming a linear space (a subspace of Hilbert space) since any linear combination of such trial functions is again a member of the set. In contrast the set of functions $e^{-\tilde{\alpha} x^2}$ with $\tilde{\alpha}$ a variational parameter do not form a linear space since for example $e^{-x^2} + e^{-7x^2}$ is not of the form $e^{-\tilde{\alpha} x^2}$. There are other interesting sets of trial functions which form linear spaces. Thus the set of all functions of a given symmetry form a linear space. Also there has been considerable interest in the so called "S - limit" for Helium-like ions in which one deals with all functions of the form $\tilde{\Psi}(r_1, r_2)$ where r_1 and r_2 are distances of the two electrons from the nucleus. Clearly the set of all such functions, or indeed the subset of all such functions which are symmetric (antisymmetric) form a linear space.[1]

We will now show that whenever the set of trial functions forms a linear space, then although the $\hat{\Psi}_k$ and $\hat{E}_k$ (we will use the same notation for the general case as for the special case (V-10)) are generally only approximations to the eigenfunctions and eigenvalues of H , they are exact eigenfunctions and eigenvalues of the "projected Hamiltonian"

$$\bar{H} \equiv \Pi H \Pi \qquad \text{(VI-2)}$$

where Π is the Hermitian projection operator onto the linear space spanned by the trial functions:

$\Pi \phi = \phi$ if ϕ is in the space

$\Pi \phi = 0$ if ϕ is orthogonal to the space

Proof: From [1] and [2] of Sec. II the conditions

$$(\Delta, (\bar{H}-\hat{E}_k)\hat{\Psi}_k) + (\Psi_k, (\bar{H}-\hat{E}_k)\Delta) = 0 \quad \text{all } \Delta \qquad \text{(VI-5)}$$

are both necessary and sufficient for $\hat{\Psi}_k$ and $\hat{E}_k$ to be an eigenfunction and corresponding eigenvalue respectively of $\bar{H}$

Now any Δ can be written

$$\Delta = \Delta_{\parallel} + \Delta_{\perp} \qquad \text{(VI-6)}$$

where $\Delta_\parallel$ is wholly in the linear space and where $\Delta_\perp$ is orthogonal to the space. In particular then, this means that $\Delta_\perp$ is orthogonal to $(\bar{H} - \hat{E}_k)\hat{\psi}_k$, and therefore that the contribution of $\Delta_\perp$ to the left hand side of (5) vanishes identically. Thus we are now left with showing that

$$(\Delta_\parallel, (\bar{H}-\hat{E}_k)\hat{\psi}_k) + (\hat{\psi}_k, (\bar{H}-\hat{E}_k)\Delta_\parallel) = 0 \quad , \text{all } \Delta_\parallel \qquad \text{(VI-7)}$$

We now note that since

$$\Pi\,\Delta_\parallel = \Delta_\parallel \qquad \text{and} \qquad \Pi\,\hat{\psi}_k = \hat{\psi}_k$$

we may replace $\bar{H}$ in (7) by H which yields

$$(\Delta_\parallel, (H-\hat{E}_k)\hat{\psi}_k) + (\hat{\psi}_k, (H-\hat{E}_k)\Delta_\parallel) = 0 \quad , \text{all } \Delta_\parallel \qquad \text{(VI-8)}$$

But now we are finished because we know that (8) is true, since it is just the basic equation of the variation method written in somewhat different notation. Namely since we are dealing with a linear space it should be clear that the functions which are close to $\hat{\psi}_k$ can all be written in the form $\hat{\psi}_k + \epsilon\Delta_\parallel$ where $\Delta_\parallel$ is an arbitrary member of the space and where ϵ is a small real number. Thus a general $\delta\hat{\psi}_k$ is of the form $\epsilon\Delta_\parallel$ and this, when inserted in (IV-3) yields (8) which proves the point. Of course any function orthogonal to the space is also an eigenfunction of $\bar{H}$ with eigenvalue zero.

The observation that for a linear space the $\hat{\Psi}_k$ and $\hat{E}_k$ are eigenfunctions and corresponding eigenvalues of the Hermitian operator $\bar{H}$ immediately leads to the following important results which among other things show that the $\hat{\Psi}_k$ and $\hat{E}_k$ have some of the formal properties of the eigenfunctions and eigenvalues of H (these results can also be derived directly from the variational equations, for example from the equations (V-12) in the case of the linear variation method):

(0) Considering $\tilde{E}$ as a function of the variation parameters and/or functions, $\tilde{E} = \hat{E}_1$ is an absolute minimum, $\tilde{E} = \hat{E}_M$ is an absolute maximum and $\tilde{E} = \hat{E}_k$, $k \neq 1, M$ are only stationary points, neither maxima or minima.

(1) The $\hat{E}_k$ are real - this of course has been clear since Sec. II.

(ii) If $\hat{E}_k \neq \hat{E}_L$, $\hat{\Psi}_k$ and $\hat{\Psi}_L$ are automatically orthogonal, while if $\hat{E}_k = \hat{E}_L$ but $\hat{\Psi}_k \neq \hat{\Psi}_L$, i.e. if there is degeneracy, $\hat{\Psi}_k$ and $\hat{\Psi}_L$ can be chosen to be orthogonal. A general degenerate $\hat{\Psi}$ is then some linear combination of the orthogonal ones. If further we assume, as we shall that the $\hat{\Psi}_k$ are normalized then we have

$$(\hat{\Psi}_k, \hat{\Psi}_L) = \delta_{kL} \tag{VI-9}$$

(iii) Since $\bar{H}\,\hat{\Psi}_L = \hat{E}_L\,\hat{\Psi}_L$ we have, from (i) and (ii), that

$$(\hat{\Psi}_k, \bar{H}\,\hat{\Psi}_L) = \hat{E}_L\,(\hat{\Psi}_k, \hat{\Psi}_L) = \hat{E}_L\,\delta_{kL}$$

or since $\Pi\hat{\varphi}_K = \hat{\varphi}_K$ and $\Pi\hat{\varphi}_L = \hat{\varphi}_L$

$$(\hat{\varphi}_K, H\hat{\varphi}_L) = \hat{E}_L\delta_{KL} \tag{VI-10}$$

Thus in words: the $\hat{\varphi}_K$ are a finite orthonormal set of functions which diagonalize H within the finite space which they span.

(iv)
$$H\hat{\varphi}_L = \hat{E}_L\hat{\varphi}_L + \psi_\perp \tag{VI-11}$$

where $\psi_\perp$ is orthogonal to the finite space. Proof: Since the $\hat{\varphi}_K$ span the space we can certainly write $H\hat{\varphi}_L = \sum_K A_{LK}\hat{\varphi}_K + \psi_\perp$. But then from (9) and (10) and the fact that $(\hat{\varphi}_K, \psi_\perp) = 0$ we have $A_{LK} = (\hat{\varphi}_K, H\hat{\varphi}_L) = \hat{E}_K\delta_{KL}$ and therefore (11).

(v) As a converse theorem, if we have a set of functions with the properties (9), (10), and hence (11) then if we use them as basis functions in a linear variation calculation, the optimal trial functions will just be the $\hat{\varphi}_K$ again and the $\hat{E}$ will be the $\hat{E}_K$. Proof: Let Π be the projection onto the space of the $\hat{\varphi}_K$. Then we have $\bar{H}\hat{\varphi}_L = \Pi H\hat{\varphi}_L$ which from (11) equals $\hat{E}_L\Pi\hat{\varphi}_L + \Pi\psi_\perp = \hat{E}_L\hat{\varphi}_L$. Thus the $\hat{\varphi}_K$ are eigenfunctions of $\bar{H}$ with $\hat{E}_K$ the eigenvalues, and this proves the point.

Although the set of trial functions (V-10) and the "s-limit" functions each form a linear subspace, there is one obvious difference between them; the space formed by the former is of finite dimensionality while that formed by the latter is infinite. This has important consequences in practice. One can fairly readily solve finite problems, particularly algebraic problems, to arbitrary accuracy (and the same is true for ordinary differential equations). However really infinite problems typically involving the solution of non separable partial differential or integral equations, are usually intractable and make it necessary to introduce further approximations, although recently partial differential equations in two variables, such as occur with the S-limit problem have begun to come under direct attack.[2] Usually these further approximations consist simply in again using the variation method but with the $\tilde{\Psi}$ finitely, though not necessarily linearly, parametrized subset of the infinite linear space.[3] Of course if the finite subspace is linear then the $\hat{\Psi}$ and $\hat{E}$ will be eigenfunctions and eigenvalues of $\pi H \pi$ where π is the projection onto the finite subspace.

Since the linear variation method leads to a finite problem it has been widely used and goes under various names: The Ritz method, the Rayleigh-Ritz method, the method of linear variational parameters, etc. As the name Rayleigh suggests its use predates quantum mechanics; it has been applied to all kinds of vibration problems, and quite generally wherever eigenvalue problems occur.

In atomic and molecular problems one common application of the linear variation method is in the configuration interaction method (CI).[4] Here, with H a fixed nucleus Hamiltonian, the Φ_K are Slater determinants made out of given spin orbitals, (the spin orbitals often also involving non-linear parameters - see end of Sec. VII). If one uses all the determinants of appropriate symmetry which one can make from the given spin orbitals then one speaks of complete CI; otherwise one speaks of incomplete CI. In this connection it is important to keep in mind that even with a modest number of spin orbitals the complete CI problem, though finite, may become impractically large. For example if one has 10 electrons and 20 spin orbitals one can form

$$\frac{20!}{10!\ 10!} = 184,756$$

Slater determinants! Of course probably for reasons of symmetry not all of these need be used but still the numbers can become enormous. Thus partial CI, involving a selection of (hopefully) the most important "configurations" Φ_k becomes the practical alternative when one deals with even moderately complicated systems.[B]

VII. LINEAR SPACES AND EXCITED STATES[1]

We have by now mentioned several times that all the $\hat{E}_k$ furnished by the linear variation method have bounding properties. We now want to prove this. More generally we will show that whenever the set of trial functions form a linear space (having a definite symmetry if symmetry considerations are applicable) then the successive $\hat{E}_K$ are upper bounds to the corresponding successive bound state eigenvalues of $\mathbb{H}$ (of that symmetry).[A]

To prove this we first note that from (VI-9) and (VI-10) the average energy in a state described by $\hat{\Phi} = \sum_{k=1}^{M} b_k \hat{\Phi}_k$ is

$$\tilde{E} = \frac{\sum_{k=1}^{M}\sum_{l=1}^{M} b_k^* b_l (\hat{\Phi}_k, H\hat{\Phi}_l)}{\sum_{k=1}^{M}\sum_{l=1}^{M} b_k^* b_l (\hat{\Phi}_k, \hat{\Phi}_l)} = \frac{\sum_{k=1}^{M} |b_k|^2 \hat{E}_k}{\sum_{k=1}^{M} |b_k|^2} \qquad \text{(VII-1)}$$

and therefore is not greater than $\hat{E}_{k'}$ where k' is the largest k for which $b_k \neq 0$. Further we note that there is at least one linear combination of the first N $\hat{\phi}_k$ which is orthogonal to the lowest (N-1) eigenfunctions of H (having the same symmetry). From what we have just proven the average energy for this function will be less than or equal to $\hat{E}_N$ while from [6] of Sec. II it is certainly not less than E_N the N'th smallest eigenvalue of H (of that symmetry). Thus we have, as announced, that

$$E_N \leq \hat{E}_N \tag{VII-2}$$

We will now show further that the bounds (2) are improvable bounds in that if we are dealing with a finite space, then enlarging the space will improve or at any rate not worsen them. Thus as already mentioned in Sec. III, the linear variation method provides a soundly based method for approximating the higher eigenvalues of H.

We start with 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 (VI-10) becomes

$$(\hat{\phi}_k, H\hat{\phi}_l) = \hat{E}_k(M)\,\delta_{kl} \tag{VII-3}$$

Suppose now that we add one more function ϕ to our basis set. We may assume without loss of generality that ϕ is normalized and orthogonal to all the ϕ_k, and hence orthogonal to all the $\hat{\phi}_k$:

$$(\phi, \phi) = 1 \qquad\qquad (\phi, \hat{\phi}_k) = 0 \tag{VII-4}$$

and of course we continue to have

$$(\hat{\varphi}_K, \hat{\varphi}_L) = \delta_{KL} \tag{VII-5}$$

Let us write our new optimal trial function as

$$\hat{\varphi} = \sum_{L=1}^{M} \hat{b}_L \hat{\varphi}_L + \hat{b}\,\phi \tag{VII-6}$$

where, for convenience, we will use the $\hat{\varphi}_K$ instead of the ϕ_K .

If we now insert $\delta\varphi = \sum_{L=1}^{M} \delta\hat{b}_L \hat{\varphi}_L + \delta\hat{b}\,\phi$ with $\delta\hat{b}_L$ and $\delta\hat{b}$ arbitrary, into (V-3) and use (3), (4) and (5), the following equations result:

$$(\hat{E}_K(M) - \hat{E})\hat{b}_K + (\hat{\varphi}_K, H\phi)\hat{b} = 0 \tag{VII-7}$$

and

$$\sum_{L=1}^{M} (\phi, H\hat{\varphi}_L)\hat{b}_L + [(\phi, H\phi) - \hat{E}]\hat{b} = 0 \tag{VII-8}$$

From (7) we then have

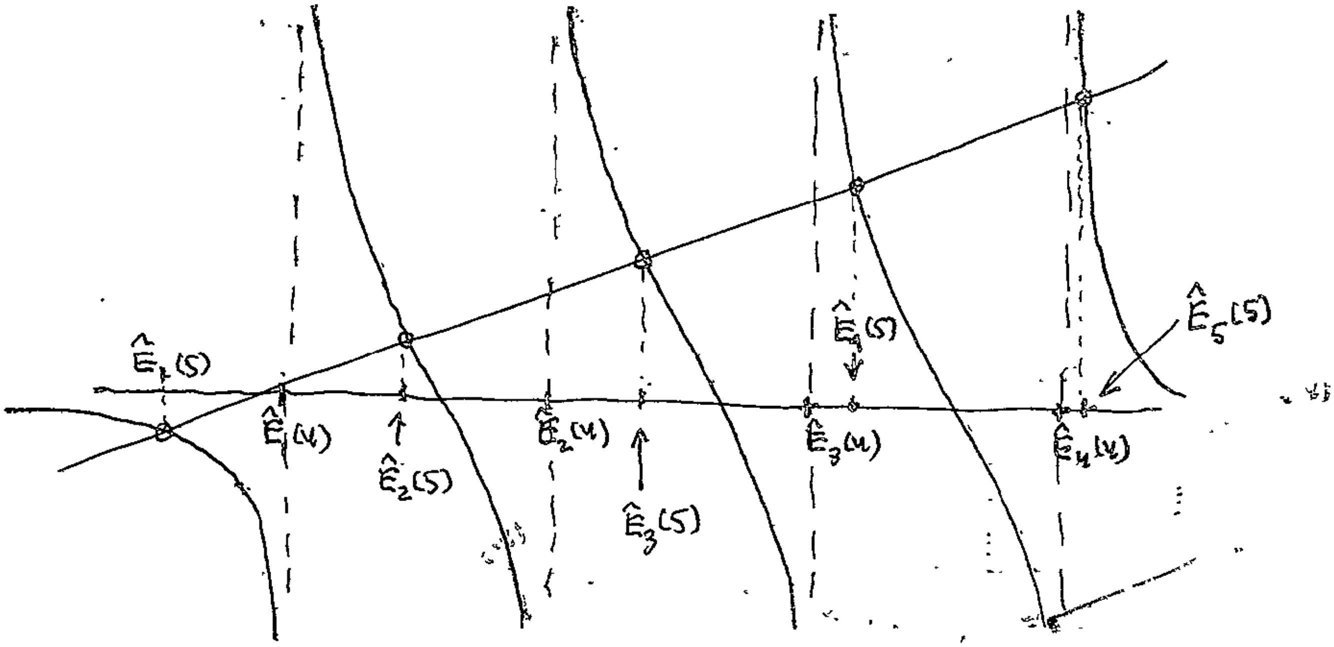
$$\hat{b}_L = (\hat{\varphi}_L, H\phi)\hat{b} \,/\, (\hat{E} - \hat{E}_L(M))$$

which when inserted into (8), yields an equation for $\hat{E}$:

$$\hat{E} - (\phi, H\phi) = \sum_{L=1}^{M} \frac{|(\phi, H\hat{\varphi}_L)|^2}{\hat{E} - \hat{E}_L(M)} \equiv \Omega \tag{VII-9}$$

If the $\hat{E}_K(M)$ are all distinct and if none of the $(\phi, H\hat{\varphi}_L)$ vanishes (we will shortly remove these restrictions) then Ω as a function of $\hat{E}$ obviously has the following properties: It has simple poles at $\hat{E} = \hat{E}_L(M)$, $L=1,\dots,M$. 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 (9), let us denote them by $\hat{E}_k(M+1)$, are then the intersections of Ω with the straight line $\hat{E} - (\Phi, H\Phi)$. The situation is shown graphically below for M = 4.



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

$$\hat{E}_{k-1}(M) \leq \hat{E}_k(M+1) \leq \hat{E}_k(M) \qquad \text{(VII-10)}$$

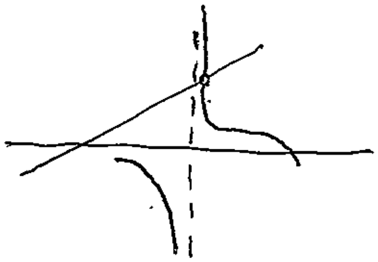
and hence in particular

$$\hat{E}_k(M+1) \leq \hat{E}_k(M) \qquad \text{(VII-11)}$$

which is what we wanted to prove (for a more elegant proof see Appendix A).[C]

Turning now to the case of degeneracy among the $\hat{E}_k(M)$, the qualitative picture isn't changed since one can consider degeneracy as a limiting case of non-degeneracy.[2] Graphically what happens is that the appropriate  segments becomes steeper and steeper as successive $\hat{E}_k(M)$ come closer together, and in the limit become vertical lines. In particular then if (say) $E_2(4) = E_3(4) = \mathcal{E}$ then $E_3(5)$ will again equal $\mathcal{E}$, though in general there would be no more degeneracy; or more generally that if there is an n - fold degeneracy among the $\hat{E}_k(M)$ at the value $\mathcal{E}$, then the $\hat{E}_k(M+1)$ will have at least an n-1 fold degeneracy, also at the value $\mathcal{E}$. However in any case (10) still holds.

Finally let us consider the possibility that one, or more, of the $(\Phi, H\hat{\Psi}_L)$ vanish. Suppose in particular that in our example $(\Phi, H\hat{\Psi}_2)$ becomes very small. This will mean that the sections  on either side of the vertical assymptote at $\hat{E} = \hat{E}_2(4)$ will hug the vertical assymptote more and more closely since the strength of the pole is being diminished. Thus we will have



In the limit then as $(\Phi, H\hat{\Psi}_2)$ becomes very small we will evidently find $E_3(5) = E_2(4)$. and in general the vanishing of $(\Phi, H\hat{\Psi}_L)$ means that $\hat{E}_L(M)$ remains an $\hat{E}_k(M+1)$ as one goes on to the next stage; however Eq. (10) still holds.

Although it is primarily of theoretical interest, we will now compare the excited state bound one gets using only linear variation parameters, to what one would get if one used linear variation parameters, and in addition could also impose orthogonality to lower states as discussed in [6], Sec. II. As might be expected, the latter procedure, if it can be carried out, will generally yield a better bound. Consider the first excited state. Then suppose that instead of simply using an (M+1) dimensional basis set and trial functions of the form $\sum_{L=1}^{M+1} \tilde{\alpha}_L \phi_L$ we further require that $\sum_{L=1}^{M+1} \tilde{\alpha}_L (\psi, \phi_L) = 0$, where ψ is the lowest eigenfunction of H (with appropriate symmetry). Thus we can use this last equation to determine one of the $\tilde{\alpha}_L$ for which $(\psi, \phi_L) \neq 0$, in terms of the others. Let this one be $\tilde{\alpha}_{M+1}$. Eliminating $\tilde{\alpha}_{M+1}$ in this way then we see that this procedure is equivalent to using as trial functions the set

$$\Psi' = \sum_{L=1}^{M} \tilde{\alpha}_L \left[\phi_L - \phi_{M+1} \frac{(\psi, \phi_L)}{(\psi, \phi_{M+1})} \right] \equiv \sum_{L=1}^{M} \tilde{\alpha}_L \chi_L$$

Thus this procedure corresponds to using the linear variation method with the M functions χ_L as the basis set. We now note that if we adjoin the function ϕ_{M+1} to the functions χ_L we will effectively recover our original M+1 dimensional basis set and therefore it follows from (1) with k = 2 that, in obvious notation,

$$\hat{E}_1(M) \leq \hat{E}_2(M+1)$$

On the other hand we also know from [6] of Sec. II that if E_2 is the 1st excited eigenvalue of H , then

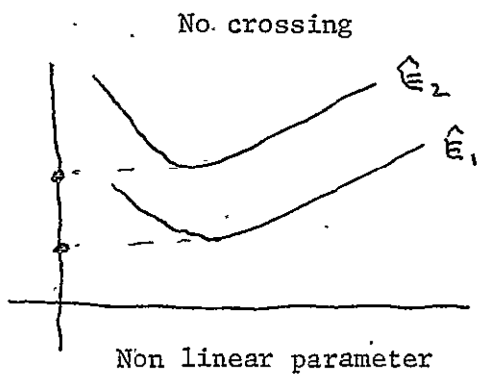
$$E_2 \leq \hat{E}_1(M)$$

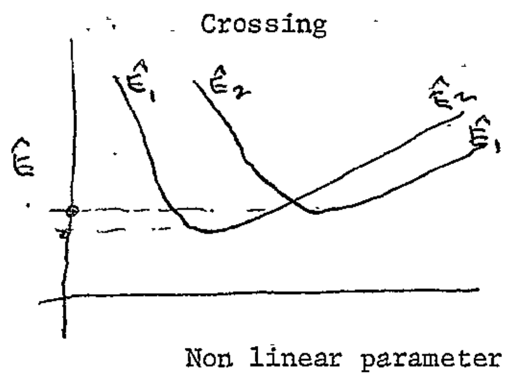
so we have

$$E_2 \leq \hat{E}_1(M) \leq \hat{E}_2(M+1) \tag{VII-12}$$

which shows as expected that $\hat{E}_1(M)$, if we could calculate it, would be a better approximation to E_2 than is $\hat{E}_2(M+1)$. Another proof of this theorem and of its generalization to higher states is given in Appendix A.

The results which we have found in this section hold for any given choice of the ϕ_k . In practice one often imbeds parameters ("non-linear parameters") in the ϕ_k and varies them as well. The reason for introducing non-linear parameters is that they are usually very effective in that one non-linear parameter can often do the work of many linear parameters. Thus a single parameter $\tilde{\alpha}$ can produce an optimal exponent in $e^{-\tilde{\alpha} r}$ whereas it will in general take several terms to do as well by linearly superposing for example e^{-r}, re^{-r}, r^2e^{-r} etc. However the difficulties of dealing with non-linear parameters (see for example the paper by Handler and Joy cited in reference 7) coupled with the increasing power of modern computing machinery often swings the balance in favor of more linear parameters, i.e. more basis functions. When non-linear parameters are used one usually chooses the parameters in each $\hat{\Psi}_k$ so as to minimize each $\hat{E}_k$ separately. This in general will mean different parameter values in each $\hat{\Psi}_k$, however from what we have just said, the bounds are still valid though one has to be aware of the possibility of "curve crossing" as shown in the graphs below.





Thus in the curve crossing case illustrated if one used the two minimum values, one actually gets two guaranteed upper bounds to E_1, rather than guaranteed upper bounds to E_1 and E_2. Such possibilities aside, one price one pays for having different parameters in different $\hat{\Psi}_k$ is of course that (VI-9) to (VI-10) no longer hold for $k \neq l$ and so, to this extent the $\hat{\Psi}_k$ are less like eigenfunctions of H than before.

VIII. SELF CONSISTENT FIELD METHODS - INTRODUCTION

In dealing with systems in which many particles move about at not too high densities and interact by means of long range forces, a natural, and one would expect quite accurate, approximation is to replace the detailed interactions acting on any one particle by a smooth field in which the particle is then assumed to move independently of the others, the smooth field representing the averaged effect of all the other particles calculated in some self consistent way (particle motions → fields → particle motions). Such self consistent field (SCF) models; and various approximations thereto, have been widely used to approximate the behavior of nuclei,[A] atoms, molecules, solids, liquids, plasmas, galaxies, etc., so much so that they together with the linear variation method comprise the bulk of the approximation methods used for atoms and molecules.

In quantum mechanics the self consistent field idea is made precise by using the variation method to determine the smooth fields. Most simply let $\mathcal{H}$ be a fixed nucleus Hamiltonian for the atom, molecule, or solid under consideration so that we need consider only the electrons. Then in accord with the above ideas we associate a single spin orbital with each electron and use trial functions of the form

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

the use of a Slater determinant rather than a simple independent particle product being required by the Pauli principle. The optimal spin orbitals $\hat{\varphi}_i$ are then determined by the variation method. This procedure is known as the unrestricted Hartree-Fock procedure (UHF), and we will discuss it in detail in the next two sections.

From such simple ideas however SCF methods have been developed in many forms and varieties. Without attempting a complete review let us note some of the reasons for this. For a more detailed summary and references to the literature we refer the reader to recent papers by Kaldor and Harris; by Kaldor; and by Larsson.[1]

(i) As we will discuss in Section X, for open shell states which on simple one particle models would be described by single determinants made up of spin orbitals of appropriate symmetry, UHF often fails to yield $\hat{\psi}$ of proper symmetry, (total spin, angular momentum, etc.). To ensure proper symmetry one may then further restrict the $\tilde{\varphi}_i$ in some way, for example in atoms one can require that the self consistent fields be effectively a central field which is the same for all the orbitals. Also, if the symmetry requires it, one may have to superpose several determinants (vector coupling) formed from such restricted spin orbitals. To put the matter more physically - already the Pauli principle which requires the use of determinantal wave functions, rather than simple products, is to some extent in conflict with the original independent particle picture. It is therefore not surprising that requiring further "cooperation" among the particles in order to ensure proper overall symmetry may require further concessions. We should however point out that in making these remarks we have in fact inverted history. The restricted schemes were developed first. Later (for a brief review see Larsson's paper) there were reasons, both formal and physical, for relaxing the restrictions on the $\tilde{\varphi}_i$, total relaxation yielding UHF, but various intermediate stages have also been discussed and used.

(ii) The physical reasons for relaxing the restrictions, and therefore the symmetry requirements, often had to do with the fact that, say for open shell atoms in a simple central field model, the closed shells are quite inert, particularly as regards magnetic properties, thereby yielding poor agreement with experiment. However since the symmetry properties (spin, angular momentum, etc.) are equally well experimental facts, it is natural that in addition to the restricted schemes, other SCF type methods have been developed which <u>do</u> meet the symmetry requirements, however not by further restrictions on $\tilde{\Psi}$ but rather by making $\tilde{\Psi}$ more flexible. In addition many of these schemes (for a review of the spin symmetry problem see the papers by Kaldor and Harris, and by Kaldor[1]) stick fairly close to the original physical picture in that they still are based on spin orbitals, and most importantly for the physical picture, the number of orbitals (though not necessarily the number of spin orbitals) involved is no greater than N so that one can still make some correspondence between electrons and "states of motion".

For example in some schemes of this type ("extended Hartree-Fock"), the trial functions are of the form (1) with possibly some restrictions on the $\tilde{\varphi}_i$ but also multiplied by appropriate projection operators to enforce the desired symmetry.[2] In such schemes then there are still only N spin orbitals. An example in which there are more than N spin orbitals but still only N orbitals is to be found in the use of "open shell" wave functions of the form

$$|\tilde{\varphi}_1\alpha\,\tilde{\varphi}_2\beta| + |\tilde{\varphi}_2\alpha\,\tilde{\varphi}_1\beta|$$

where $\tilde{R}_1$ and $\tilde{R}_2$ are radial functions and α and β are spin functions, to describe the ground state of helium. These are ^2S functions but since they involve two radial functions they are more flexible than the restricted form

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

(which in this case is equivalent to UHF - see Sec. X) and can therefore yield a lower $\hat{E}$ (one *does* find that $\hat{R}_2 \neq \hat{R}_1$). However one can still speak of one electron being in the orbital $\hat{R}_1$ and the other in $\hat{R}_2$.

(iii) The departures from the simple form (1) described in (i) and (ii) all stayed rather close to the original physical picture. The third large class of departures (multi configuration SCF theorems or MC SCF)[3], though there is really no sharp distinction between these and those of (ii) except possibly in spirit, tend to start with the more formal view that in UHF one seeks the best single determinant approximation, and generalizes this by seeking the best sum of two, three, ... determinants (perhaps subject to a priori restrictions of one kind or another on the spin orbitals). Indeed MC SCF theories can probably best be viewed as an economical CI in that one attempts to fully optimize a few configurations, hoping thereby to do the work of many more fixed configurations chosen more or less arbitrarily.[3]

In the next two sections we will discuss UHF in some detail. It is formally the simplest of the SCF schemes, but it serves to illustrate most of the general features of these methods; and further it has been widely used. In Sec. X we will briefly develop another more restricted, but still rather general, SCF scheme, which is also of practical importance, and which will illustrate some further formal points.

One final note: In general the sum of two determinants is not a determinant. Therefore in UHF and the like we are not dealing with a linear space. Nevertheless there is still sufficient linearity in that

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

does equal

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

so that in some circumstances one can assert that certain higher $\hat{E}$ provided by the method furnish upper bounds to certain excited states of H . For details we refer to the original paper of Perkins.[5]

IX. THE UNRESTRICTED HARTREE-FOCK APPROXIMATION

As we discussed in the previous section UHF is formally the simplest of the SCF schemes in that in this approximation one uses only a single determinant with no further restrictions. Thus one calculates $\tilde{E}$ for $\tilde{\Psi}$ a single determinant

$$\tilde{\Psi} = |\tilde{\varphi}_1, \tilde{\varphi}_2, \ldots \tilde{\varphi}_N| \qquad \text{(IX-1)}$$

with H typically of the form

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

where g(st) = g(ts), and then determines the $\hat{\varphi}_i$ in such a way that $\delta \hat{E} \simeq 0$. (We are here taking the straightforward approach to the variational method as discussed at the beginning of Sec. IV.)

Now in calculating $\tilde{E}$ one gets simple formulae ("Slater's rules") if the $\tilde{\varphi}_i$ are orthonormal.[1] We will now show that there is no loss in generality in assuming this. First, however, we will prove a more general theorem: Consider any linear transformation of the spin orbitals

$$\tilde{\varphi}_i' = \sum_j A_{ij} \varphi_j \qquad \text{(IX-3)}$$

where the A_{ij} are any set of numbers. It then follows that

$$|\tilde{\varphi}_1' \cdots \tilde{\varphi}_N'| = |A|\,|\tilde{\varphi}_1 \cdots \tilde{\varphi}_N| \qquad \text{(IX-4)}$$

where $|A|$ is the determinant of the matrix A_{ij}. Proof: Let us write out the left hand side of (4) in more detail, using (3) and explicitly introducing the particle labels. Thus

$$|\tilde{\varphi}_1' \cdots \tilde{\varphi}_N'| = \begin{vmatrix} \sum_j A_{1j}\varphi_j(1) & \cdots & \sum_j A_{1j}\varphi_j(N) \\ \vdots & & \\ \sum_j A_{Nj}\varphi_j(1) & \cdots & \end{vmatrix}$$

which will be recognized as the determinant of the product of the two "matrices" A_{ij} and $\varphi_j(s)$. Equation (4) then follows from the standard theorem that the determinant of a product of matrices is equal to the product of the determinants of the separate matrices.

Turning now to the question at hand, we first note that there are many linear transformations of the type (3), for example the well known Schmidt procedure, which, starting from a given linearly independent set of $\tilde{\varphi}_i$ (and the $\tilde{\varphi}_i$ must be linearly independent to start with or else $\tilde{\Psi} \equiv 0$) will produce an orthogonal set of $\tilde{\varphi}_i'$.[2] From (4) then we have that $\tilde{\Psi}$ is proportional to the Slater determinant formed from the $\tilde{\varphi}_i'$, and since the proportionality constant $|A|$ will simply cancel out in calculating $\tilde{E}$, we have the desired result. Notice however that the $\tilde{\varphi}_i'$ are certainly not unique since given one orthonormal set any unitary transformation will produce another set, and from the theorem which we have just proven, this new set will yield essentially the same $\tilde{\Psi}$.

Assuming then that the $\tilde{\varphi}_i$ are orthonormal we will now derive the equations which determine the $\hat{\varphi}_i$, the optimal spin orbitals. Applying Slater's rules one finds[A]

$$\tilde{E} = \sum_j (\tilde{\varphi}_j, h\tilde{\varphi}_j) + \frac{1}{2}\sum_j \sum_k \left[(\tilde{\varphi}_j \tilde{\varphi}_k, g\, \tilde{\varphi}_j \tilde{\varphi}_k) - (\tilde{\varphi}_j \tilde{\varphi}_k, g\, \tilde{\varphi}_k \tilde{\varphi}_j) \right] \qquad \text{(IX-5)}$$

We now vary the spin orbitals and set $\delta\hat{E} = 0$. Then after a bit of rearranging we find

$$0 = \sum_j \left[(\delta\hat{\varphi}_j, \hat{f}\hat{\varphi}_j) + (\hat{f}\hat{\varphi}_j, \delta\hat{\varphi}_j) \right] \qquad \text{(IX-6)}$$

where the operator $\hat{f}$ is the so called "Hartree-Fock single particle Hamiltonian" and is defined by its action on an arbitrary spin orbital according to

$$\hat{f}(s)\,\chi(s) = h(s)\,\chi(s) + \sum_j (\hat{\varphi}_j, g\hat{\varphi}_j)\,\chi(s) - (\hat{\varphi}_j, g\chi)\,\hat{\varphi}_j(s) \tag{IX-7}$$

It is then easy to show that $(\chi', \hat{f}\chi) = (\hat{f}\chi', \chi)$, i.e. that $\hat{f}$ is Hermitian.

In deriving (5) and (6) we have assumed that the $\hat{\varphi}_j$ are orthonormal. Thus we need not require that (6) be true for all $\delta\hat{\varphi}_j$ (in fact it would be in general impossible to find non-zero $\hat{\varphi}_j$ which satisfy this) but rather we need only require it to be true for all $\delta\hat{\varphi}_j$, which also satisfy the further conditions $\delta(\hat{\varphi}_i, \hat{\varphi}_j) = 0$, i.e.

$$(\delta\hat{\varphi}_i, \hat{\varphi}_j) + (\hat{\varphi}_i, \delta\hat{\varphi}_j) = 0 \tag{IX-8}$$

To take account of these constraints we introduce Lagrange multipliers ϵ_{ij} (see Appendix B) and replace (6) and (8) by

$$0 = \sum_j (\delta\hat{\varphi}_j, \hat{f}\hat{\varphi}_j) + (\hat{f}\hat{\varphi}_j, \delta\hat{\varphi}_j)$$

$$- \sum_i \sum_j \left[(\delta\hat{\varphi}_j, \hat{\varphi}_i) + (\hat{\varphi}_j, \delta\hat{\varphi}_i) \right] \epsilon_{ij}$$

or, rearranging a bit

$$0 = \sum_j \left[\left(\delta\hat{\varphi}_j, \hat{f}\hat{\varphi}_j - \sum_i \hat{\varphi}_i \epsilon_{ij}\right) + \left(\hat{f}\hat{\varphi}_j - \sum_i \hat{\varphi}_i \epsilon^*_{ji}, \delta\hat{\varphi}_j\right) \right] \tag{IX-9}$$

which <u>is</u> (see Appendix B) to be true for <u>all</u> $\delta\hat{\varphi}_j$ . . .

We now note that the right hand side of (9) may be complex and hence its real and imaginary parts must vanish separately. In particular then since $\sum_j [(\delta\hat{\varphi}_j, f\hat{\varphi}_j) + (f\hat{\varphi}_j, \delta\hat{\varphi}_j)$ is obviously real this means that

$$0 = \mathrm{Im} \sum_i \sum_j \left[(\delta\hat{\varphi}_j, \hat{\varphi}_i)\,\epsilon_{ij} + (\hat{\varphi}_i, \delta\hat{\varphi}_j)\,\epsilon_{ji} \right]$$

which, after some rearranging, becomes

$$0 = \sum_i \sum_j \left[(\delta\hat{\varphi}_j, (\epsilon_{ij} - \epsilon_{ji}^*)\hat{\varphi}_i) - ((\epsilon_{ij} - \epsilon_{ji}^*)\hat{\varphi}_i, \delta\hat{\varphi}_j) \right] \tag{IX-10}$$

Since this is to be satisfied for all $\delta\hat{\varphi}_j$, it must in particular be satisfied by

$$\delta\hat{\varphi}_j = \eta \sum_K (\epsilon_{KL} - \epsilon_{LK}^*)\,\hat{\varphi}_K \qquad j = L$$

$$= 0 \qquad j \neq L$$

where η is a small pure imaginary number. Inserting this in (10) then yields

$$0 = \left(\sum_K (\epsilon_{KL} - \epsilon_{LK}^*)\,\hat{\varphi}_K \,,\, \sum_K (\epsilon_{KL} - \epsilon_{LK}^*)\hat{\varphi}_K \right)$$

which implies that

$$\sum_K (\epsilon_{KL} - \epsilon_{LK}^*)\,\hat{\varphi}_K = 0$$

or, since the $\hat{\varphi}_K$ are supposed to be linearly independent

$$\epsilon_{KL} - \epsilon_{LK}^* = 0 \tag{IX-11}$$

i.e. *the Lagrange multipliers must form a Hermitian matrix*, and clearly this is not only a necessary condition to satisfy (10), it is also sufficient.

Assuming (11) then, (9) becomes

$$0 = \sum_j \left[\left(\delta\hat{\varphi}_j , \hat{f}\hat{\varphi}_j - \sum_i \hat{\varphi}_i \epsilon_{ij}\right) + \left(\hat{f}\hat{\varphi}_j - \sum_i \hat{\varphi}_i \epsilon_{ij} , \delta\hat{\varphi}_j\right)\right]$$

from which, following what should by now be a familiar pattern, we are led to

$$\hat{f}\,\hat{\varphi}_j = \sum_i \hat{\varphi}_i \epsilon_{ij} \qquad \text{(IX-12)}$$

where the ϵ_{ij} are to be determined in such a way as to ensure the orthonormality of the $\hat{\varphi}_i$.

We now note that, conversely, if the $\hat{\varphi}_i$ are orthonormal, then (11) will automatically be satisfied, since then from (12)

$$\epsilon_{ij} = (\hat{\varphi}_i , \hat{f}\hat{\varphi}_j) \qquad \text{(IX-13)}$$

which since $\hat{f}$ is Hermitian, equals $(\hat{\varphi}_j , \hat{f}\hat{\varphi}_i)^*$ which in turn *is* ϵ_{ji}^* . Further we note that *one* choice of the Lagrange multipliers which will guarantee orthogonality is to put $\epsilon_{ij} = 0$ for $i \neq j$, since then the $\hat{\varphi}_i$ will all be eigenfunctions of the same Hermitian operator namely $\hat{f}$. The diagonal elements of ϵ_{ij} then remain to be used to enforce normalization. Using just a single subscript for the diagonal elements then, the equations

$$\hat{f}\hat{\varphi}_i = \epsilon_i \hat{\varphi}_i \qquad \text{(IX-14)}$$

define the "canonical" UHF spin orbitals; and clearly from the form of $\hat{f}$ as given in Eq. (7), these equations do conform to the physical picture of Sec. VIII though admittedly the exchange terms, the terms with the minus sign and $j \neq i$ in $\hat{f}$, do not admit of a simple physical interpretation.

It should be pointed out however that the canonical spin orbitals are not necessarily the most usefully ones physically. Other sets of spin orbitals derived from these by a unitary transformation may have more desirable properties, for example they may be better localized.[B] Also it has been suggested that certain non-orthogonal sets may also be useful. We will not pursue these matters further here,[3] except to note that considerations of this sort are not limited to UHF, they can be applied to the spin orbitals of any Slater determinant.

Equations (14), when (As a practical matter) they can be solved at all (see below), are usually solved by an iterative process. One first "guesses" a set of orthonormal spin orbitals, call them $\overset{1}{\varphi}_j$. From them one constructs, in an obvious way, a first approximation to $\hat{f}$, call it $\overset{1}{f}$. One then proceeds to solve

$$\overset{1}{f}\,\overset{2}{\varphi}_i = \overset{2}{\epsilon}_i\,\overset{2}{\varphi}_i$$

which is an ordinary eigenvalue problem for $\overset{2}{\varphi}_i$. Thus the solutions are automatically orthogonal if there is no degeneracy, and if there is degeneracy they can be chosen orthogonal. Also they can be normalized. One then proceeds to calculate $\overset{2}{f}$ etc, etc, stopping when (hopefully) a sufficient degree of self consistency has been reached, that is when the $\overset{N+1}{\varphi_i}$ are sufficiently similar to the $\overset{N}{\varphi_i}$.

Since $\hat{f}$ involves all the $\hat{\varphi}_j$, (14) is really a set of coupled non-linear integro-partial-differential equations. For real atoms and molecules it does not seem possible to exhibit closed form solutions, so other approaches must be used. If one can reduce the equations to coupled equations in one variable then a direct iterative numerical attack is possible.[4] In particular if H is the non-relativistic Hamiltonian for an isolated atom in the fixed nucleus approximation then, as we shall discuss in the next section, one can often find solutions of (14) in which the $\hat{\varphi}_j$ have the spin and angular dependence that one expects on the basis of the central field model of the atom, with only the radial dependence remaining to be determined from (14).[C] However even for the simplest of molecules, that is diatomic molecules, the most that one can hope to get "for free" is the dependence of the $\hat{\varphi}_j$, on the azimuthal angle around the internuclear axis, and therefore one is still left with two independent variables to deal with.

To get a finite problem the standard procedure[5] is to further restrict the $\hat{\varphi}_j$, by requiring that they may be expandable in <u>finite</u> basis sets (which may contain non linear parameters, however we will not consider them explicitly.)[6] The optimal values of the expansion coefficients are then determined from (9). We will refer to this procedure as analytic unrestricted Hartree-Fock (AUHF). Similar analytic approximations can be, and regularly are, made to other SCF and MCSCF type approximations. Thus we write

$$\hat{\varphi}_j = \sum_{\alpha=1}^{M} \hat{C}_{\alpha j}\, \chi_\alpha \qquad \text{(IX-15)}$$
$$\delta\hat{\varphi}_j = \sum_{\alpha=1}^{M} \delta\hat{C}_{\alpha j}\, \chi_\alpha$$

where to avoid notational complexity we have required that each $\hat{\varphi}_j$ be expandable in the same finite basis set.[8] (Note however that this still permits different $\hat{\varphi}_j$ to have different symmetry: one includes among the χ_α functions of various symmetries so that if the appropriate $\hat{C}_{\alpha j}$ turn out to vanish, $\hat{\varphi}_j$ will have a definite symmetry). Then (9) becomes

$$0 = \sum_j \sum_\alpha \sum_\beta \delta\hat{C}^*_{\alpha j} \left[(\chi_\alpha, \hat{f}\chi_\beta)\hat{C}_{\beta j} - \sum_i (\chi_\alpha, \chi_\beta)\hat{C}_{\beta i}\, \epsilon_{ij} \right]$$
$$+ \sum_j \sum_\alpha \sum_\beta \left[\hat{C}^*_{\beta j} (\chi_\beta, \hat{f}\chi_\alpha) - \sum_i (\chi_\beta, \chi_\alpha)\hat{C}^*_{\beta i}\, \epsilon_{ji} \right] \delta\hat{C}_{\alpha j} \qquad \text{(IX-16)}$$

where now

$$(\chi_\alpha, \hat{f}\chi_\beta) = (\chi_\alpha, h\chi_\beta) + \sum_\ell \sum_\gamma \sum_\delta \hat{C}^*_{\gamma\ell}\hat{C}_{\delta\ell} [\qquad \text{(IX-17)}$$
$$[(\chi_\alpha\chi_\gamma, g\chi_\beta\chi_\delta) - (\chi_\alpha\chi_\gamma, g\chi_\delta\chi_\beta)]$$

At this point it is helpful to introduce some matrix notation. Thus we introduce the M x M Hermitian matrix $\hat{f}$ whose elements are $(\chi_\alpha, \hat{f}\chi_\beta)$ and the M x M positive definite Hermitian "overlap matrix" S whose elements are $(\chi_\alpha, \chi_\beta)$, and finally we introduce N , M element column vectors $\hat{c}_j$ whose elements are $\hat{C}_{\alpha j}$. In terms of these quantities then (15) can be written

$$0 = \sum_j \left\{ \delta\hat{c}_j , \hat{f}\hat{c}_j - \sum_i S\hat{c}_i\, \epsilon_{ij} \right\} + \sum_j \left\{ \hat{f}\hat{c}_j - \sum_i S\hat{c}_i\, \epsilon^*_{ji} , \delta\hat{c}_j \right\} \qquad \text{(IX-18)}$$

where $\{ v_j , Q v_i \}$ means the usual scalar product of two complex vectors

$$\{V_j, O V_i\} \equiv \sum_{\alpha\beta}\sum V_{\alpha j}^{*} O_{\alpha\beta} V_{\beta i} \qquad \text{(IX-19)}$$

Also in this same notation the orthonormality requirement

$$(\hat{\varphi}_i, \hat{\varphi}_j) = \delta_{ij} \qquad \text{(IX-20)}$$

becomes

$$\{\hat{C}_i, S\hat{C}_j\} = \delta_{ij} \qquad \text{(IX-21)}$$

From this point on the argument proceeds much as before. Indeed the steps are identical if we introduce the vectors $\hat{D}_j$ and the Hermitian matrix $\hat{g}$ defined by

$$\begin{aligned}\hat{D}_j &= \sqrt{S}\,\hat{C}_j \\ \hat{f} &= \sqrt{S}\,\hat{g}\,\sqrt{S}\end{aligned} \qquad \text{(IX-22)}$$

so that (15) and (21) become

$$0 = \sum_j \{\delta\hat{D}_j, \hat{g}\hat{D}_j - \sum_i \hat{D}_i \epsilon_{ij}\} + \sum_j \{\hat{g}\hat{D}_j - \sum_i \hat{D}_i \epsilon_{ji}^{*}, \delta\hat{D}_j\}$$

and

$$\{\hat{D}_i, \hat{D}_j\} = \delta_{ij}$$

which have precisely the same structure as (9) and (20). Thus we are again led to (11) and instead of (14) we have

$$\hat{g}\,\hat{D}_i = \epsilon_i \hat{D}_i$$

or from (22)

$$\hat{f}\,\hat{C}_i = \epsilon_i S \hat{C}_i \qquad \text{(IX-23)}$$

Incidently, recalling the discussion in Sec. V, note that these same equations can also be derived by taking "moments" of (14). That is if we substitute for the $\hat{\varphi}_\kappa$ from (15) into (14) (including $\hat{f}$) and take the scalar product of the resulting "equation" with χ_β one arrives at (23).

The equations (23), with (17) constitute a set of non-linear algebraic equations for the $\hat{C}_j$. Again the usual solution procedures are iterative. One chooses some $\overset{1}{C}_j$, computes $\overset{1}{f}$, solves

$$\overset{1}{f}\,\overset{2}{C}_i = \overset{2}{\epsilon}_i\, S\, \overset{2}{C}_i \qquad \text{(IX-24)}$$

as an ordinary algebraic eigenvalue problem, calculates $\overset{2}{f}$ etc. etc. For the details of practical procedures for doing this for atoms as well as a discussion of a treatment of non-linear parameters we refer the reader to the article by Roothaan and Bagus.[9] For molecules, especially for large molecules, this procedure may still not be practical at all, or it may only be practical only using the largest of computers, because of the sheer number of integrals which must be calculated to evaluate $\hat{f}$ and often, depending on the nature of the χ_α because of difficulty in calculating them. These difficulties, in addition to spawning large literatures on the choice of the χ_α , and on integral evaluation,[10] have also led to the development of many methods which further approximate $\hat{f}$ in some way or other, however we will not attempt to review these methods here.[11]

Thus far we have not looked at $\hat{E}$, and that it has not emerged automatically is of course a consequence of the fact that our $\tilde{\Psi}$ all had a fixed scale: $(\tilde{\Psi}, \tilde{\Psi}) = N!$. Clearly it would be nice, and would yield a simple physical interpretation of the ϵ_i if $\hat{E}$ were to equal $\sum_i \epsilon_i$. However this is not the case since from (5) we have

$$\hat{E} = \sum_i (\hat{\varphi}_i, h\hat{\varphi}_i) + \frac{1}{2}\sum_i \sum_j \left[(\hat{\varphi}_i \hat{\varphi}_j, g\, \hat{\varphi}_i \hat{\varphi}_j) - (\hat{\varphi}_i \hat{\varphi}_j, g\hat{\varphi}_j \hat{\varphi}_i)\right] \qquad \text{(IX-25)}$$

or

$$\hat{E} = \sum_i (\hat{\varphi}_i, h\hat{\varphi}_i) + \sum_i \sum_j \left[\qquad \text{''} \qquad \right]$$
$$- \frac{1}{2}\sum_i \sum_j \left[\qquad \text{''} \qquad \right]$$

In this form the first line will be recognized as $\sum_i (\hat{\varphi}_i, \hat{f}\hat{\varphi}_i)$, which therefore from (14) and the fact that the $\hat{\varphi}_i$ are normalized, does equal $\sum_i \epsilon_i$. However there is still the second line which is clearly just the negative of the average of the two body interaction $\frac{1}{2}\sum\sum_{s \neq t} g(st)$. Therefore we have

$$\hat{E} = \sum_{i=1}^{N} \epsilon_i - (\varphi, \tfrac{1}{2}\sum\sum_{s\neq t} g(st)\varphi)/(\varphi, \varphi) \qquad \text{(IX-27)}$$

so that, as we said, $\hat{E}$ is not just the sum of the ϵ_i (an analogous result holds in AUHF). The physical basis for this result is of course the fact that $\sum_i \epsilon_i$ counts each electron-electron interaction twice. Thus one cannot immediately ascribe physical significance to the ϵ_i as "one electron energies". However there is a good empirical correlation between the ϵ_i and ionization energies, a correlation that is supported by the following theoretical result known as Koopman's theorem:[13] Suppose that as an approximation to the optimal UHF function for the N-1 electron ionized system one uses the determinant gotten from $\hat{\Psi}$ by deleting the k'th spin orbital. Then the $\tilde{E}$ for the N-1 particle system will differ from $\hat{E}$ in (25) by the removal of all terms involving $\hat{\varphi}_k$. However these terms are precisely

$$\Delta E = (\hat{\varphi}_k, h\hat{\varphi}_k) + \frac{1}{2}\sum_j \left[(\hat{\varphi}_k\hat{\varphi}_j, g\,\hat{\varphi}_k\hat{\varphi}_j) - (\hat{\varphi}_k\hat{\varphi}_j, g\,\hat{\varphi}_j\hat{\varphi}_k)\right]$$
$$+\frac{1}{2}\sum_j \left[(\hat{\varphi}_j\hat{\varphi}_k, g\,\hat{\varphi}_j\hat{\varphi}_k) - (\hat{\varphi}_j\hat{\varphi}_k, g\,\hat{\varphi}_k\hat{\varphi}_j)\right]$$

or

$$\Delta E = (\hat{\varphi}_k, h\hat{\varphi}_k) + \sum_j \left[(\hat{\varphi}_k\hat{\varphi}_j, g\,\hat{\varphi}_k\hat{\varphi}_j) - (\hat{\varphi}_k\hat{\varphi}_j, g\,\hat{\varphi}_j\hat{\varphi}_k)\right]$$

or finally

$$\Delta E = (\hat{\varphi}_k, \hat{f}\,\hat{\varphi}_k) = \epsilon_k$$

Thus in this approximation ϵ_k is precisely the ionization energy.[D] Of course this kind of trial function for the N-1 particle system with so to speak "frozen orbitals" would not seem to be a very good approximation, especially if inner shell electrons are being ionized; presumably it would be better to do a complete UHF calculation for the N-1 particle system.[E] Nevertheless the result does give some feeling for the empirical situation.[14]

X. RESTRICTED HARTREE FOCK AND OHF

If for an isolated atom or molecule, simple one particle models predict that single determinant closed shell states are possible then one can show that the UHF equations do have solutions of the expected symmetry.[1] [A] Thus for two electron atoms (with H the nonrelativistic, fixed nucleus, spin free Hamiltonian) one can find solutions of the form

$$|\hat{R}\alpha\,\hat{R}\,\beta| \tag{X-1}$$

and similarly for ten electron atoms one can find solutions of the form

$$|\hat{R}_1\alpha\,\hat{R}_1\beta\,\hat{R}_2\alpha\,\hat{R}_2\beta \text{ ------ } Y_1^{-1}\hat{R}_3\alpha\,Y_1^{-1}\hat{R}_3\beta|$$

However when one goes to open shells the situation changes in that, considering atoms to be definite, even though a simple central field model would allow the state to be described by a single determinant (i.e. no vector coupling needed), still as we mentioned in Sec. VIII, UHF will not in general have solutions of the expected symmetry. Thus, confining ourselves for the moment to s-orbitals, for two electron systems we can find solutions of the form

$$|\hat{R}_1\alpha\,\hat{R}_2\alpha| \tag{X-2}$$

which is a member of a spin triplet. However for three electron systems it is easy to see by trying that though there are pure quartet solutions $|\hat{R}_1\alpha\,\hat{R}_2\alpha\,\hat{R}_3\alpha|$ there are no pure doublets

$$|\hat{R}_1\alpha\,\hat{R}_1\beta\,\hat{R}_2\alpha| \tag{X-3}$$

but rather there are solutions of the form

$$|\hat{R}_1\alpha\,\hat{R}_2\beta\,\hat{R}_3\alpha| \tag{X-4}$$

which is a linear combination of doublet and quartet.

A similar situation exists with respect to orbital angular momentum when one investigates what would normally be single determinant open shell states involving orbitals of non zero angular momentum, for example $1s^2 2p\ {}^2P$ for three electrons. One finds instead solutions which are eigenfunctions of L_z but not of L^2. Further, although the orbitals are eigenfunctions of ℓ_z, the component of one particle angular momentum, they are not eigenfunctions of ℓ^2.[2] In Sec. VIII we briefly sketched and gave references to the responses which have been made to this situation. Here we want to pursue the restricted Hartree-Fock approach in more detail.

Applied to closed shells, and to be specific let us consider the two electron atomic example, it would consist in restricting the $\tilde{\Psi}$ to be of the form

$$|\tilde{R}\alpha\ \tilde{R}\beta|$$

from the outset, and then determining $\hat{R}$ from the variation method. Since as we have said, UHF does have solutions of this type it follows that for closed shells the restricted Hartree-Fock functions satisfy the UHF equation. We mention this because this then implies that the RHF functions for closed shells will also satisfy the many interesting special theorems which are satisfied by UHF functions, theorems which we will be discussing in Sec. XII.

Turning now to open shells, consider again the three electron doublet example. There one would use trial functions of the form

$$|\tilde{R}_1\alpha\ \tilde{R}_1\beta\ \tilde{R}_2\alpha|$$

More generally let us consider trial functions of the form

$$|\tilde{v}_1\alpha\ \tilde{v}_1\beta\ \tilde{v}_2\alpha| \qquad \text{(X-5)}$$

where we do not restrict the form of the orbitals v_i , i.e. they need not be simply radial functions. Since adding a multiple of $\tilde{\hat{v}}_1$ to $\tilde{\hat{v}}_2$ will not change the value of the determinant we may assume that $\tilde{\hat{v}}_1$ and $\tilde{\hat{v}}_2$ are orthogonal and of course we may assume that each is normalized. Then if h and g are spin independent one readily finds from (IX-5) that

$$\tilde{E} = 2(\tilde{\hat{v}}_1, h\tilde{\hat{v}}_1) + (\tilde{\hat{v}}_2, h\tilde{\hat{v}}_2) + (\tilde{\hat{v}}_1\tilde{\hat{v}}_1, g\tilde{\hat{v}}_1\tilde{\hat{v}}_1) + 2(\tilde{\hat{v}}_1\tilde{\hat{v}}_2, g\tilde{\hat{v}}_1\tilde{\hat{v}}_2)$$
$$- (\tilde{\hat{v}}_1\tilde{\hat{v}}_2, g\tilde{\hat{v}}_2\tilde{\hat{v}}_1)$$

$\delta\hat{E}$ = 0 subject to the conditions $(\hat{v}_i, \hat{v}_j) = \delta_{ij}$ then leads to

$$2h\hat{v}_1 + 2(\hat{v}_1, g\hat{v}_1)\hat{v}_1 + 2(\hat{v}_2, g\hat{v}_2)\hat{v}_1 - (\hat{v}_2, g\hat{v}_1)\hat{v}_2 = \epsilon_{11}\hat{v}_1 + \epsilon_{12}\hat{v}_2 \qquad \text{(X-7)}$$

and

$$h\hat{v}_2 + 2(\hat{v}_1, g\hat{v}_1)\hat{v}_2 - (\hat{v}_1, g\hat{v}_2)\hat{v}_1 = \epsilon_{12}^*\hat{v}_1 + \epsilon_{22}\hat{v}_2$$

There are now two main points which we want to make:

(i) Without trying to write these equations in very elegant form it should be clear that we cannot ensure the orthogonality of $\hat{v}_1$ and $\hat{v}_2$ by simply putting $\epsilon_{12} = 0$ since $\hat{v}_1$ and $\hat{v}_2$ will not then be eigenfunctions of a common Hamiltonian. Rather one must deal with the off diagonal Lagrange multipliers explicitly, and this is a rather general feature of restricted open shell SCF calculations. Various procedures have been devised to do this and we will simply refer the interested reader to the original literature on the subject[3]. Also direct search procedures, such as we mentioned at the outset of Sec. IV, which effectively bypass the equations, have been used.[4]

(Such procedures are anyway almost invariably used to find the optimal values of non linear parameters).

(ii) It is easy to see that the equations do admit solutions of the form (3). Thus the restricted Hartree-Fock function for this case will share the general properties of solutions of (7). More generally we will denote (for no particular reason) by OHF the procedure, in which one deals with trial functions of the form

$$|\tilde{v}_1\alpha\,\tilde{v}_1\beta\,\tilde{v}_2\alpha\,\tilde{v}_2\beta - - - \tilde{v}_c\alpha\,\tilde{v}_c\beta\,\tilde{v}_{c+1}\alpha\,\tilde{v}_{c+2}\alpha - - - \tilde{v}_{N-c}\alpha|$$

which are therefore eigenfunctions of S_z with eigenvalue $\frac{N-2c}{2}$ and of S^2 with eigenvalue $\left(\frac{N-2c}{2}\right)\left(\frac{N-2c}{2}+1\right)$; our discussion above then being for the special case $N=3$, $c=1$. One can then show that restricted Hartree-Fock functions for atoms and molecules describing closed shells plus a spatially closed shell with all spins aligned will satisfy the OHF equations, and therefore will also satisfy the theorems for OHF which will be discussed in Sec. XII. However since no essentially new points of principle arise we will not write out the detailed equations for general OHF and AOHF.

XI. THE GENERALIZED BRILLOUIN THEOREM

In general whatever sort of trial functions one uses, any $\hat{\psi}$ will almost certainly be only an approximation to an eigenfunction of H, and so the question naturally arises, how can we improve on the approximation? One approach of course is simply to enlarge the set of trial functions in some way. Another would be to use

Rayleigh-Schroedinger perturbation theory, and it is this approach which we want to discuss in this section.

In order to use RS perturbation theory we must introduce a zero order Hamiltonian H_0 which has $\hat{\Psi}$ as an eigenfunction:

$$H_0 \hat{\Psi} = \mathcal{E} \hat{\Psi} \tag{XI-1}$$

where $\mathcal{E}$, the eigenvalue, may or may not equal $\hat{E}$. Having chosen H_0 one now treats $H-H_0$ as a perturbation. The first order correction to $\hat{\Psi}$ is therefore (we will assume that $\mathcal{E}$ is non-degenerate and as usual will use a discrete notation; also we will assume that $\hat{\Psi}$ is normalized)

$$\hat{\Psi}^{(1)} = -\sum_{\alpha} \Theta_\alpha \frac{(\Theta_\alpha, (H-H_0)\hat{\Psi})}{\mathcal{E}_\alpha - \mathcal{E}} \tag{XI-2}$$

where the Θ_α are the other orthonormal eigenfunctions of H_0, and where the $\mathcal{E}_\alpha$ are the corresponding eigenvalues. Finally using (1) and the fact that the Θ_α are orthogonal to $\hat{\Psi}$ we can write (2) as

$$\hat{\Psi}^{(1)} = -\sum_{\alpha} \Theta_\alpha \frac{(\Theta_\alpha, H\hat{\Psi})}{\mathcal{E}_\alpha - \mathcal{E}} \tag{XI-3}$$

Note however that this perturbation scheme involves a great deal of arbitrariness since if we write

$$H_0 = \mathcal{E} |\hat{\Psi}\rangle\langle\hat{\Psi}| + \sum_{\alpha} \mathcal{E}_\alpha |\Theta_\alpha\rangle\langle\Theta_\alpha| \tag{XI-4}$$

then (1) will be satisfied whatever we choose for the $\mathcal{E}_\alpha$ and the Θ_α so long as the latter are orthogonal to $\hat{\Psi}$. However different choices of the $\mathcal{E}_\alpha$ and the Θ_α can make profound changes in $\hat{\Psi}^{(1)}$.

We now note that as a consequence of the variation method, certain $(\Theta_\alpha, H\hat{\Psi})$ may vanish so that the corresponding Θ_α will not occur in $\hat{\Psi}^{(1)}$, though of course they may appear in $\hat{\Psi}^{(2)}, \hat{\Psi}^{(3)}---$, etc. Namely suppose that with $\delta\eta$ an infinitesimal, but otherwise arbitrary complex number,

$$\delta\eta \quad \Theta_\alpha \tag{XI-5}$$

is among the variations of $\hat{\Psi}$ possible within the set of trial functions. Then it follows from (V-3) and the orthogonality of Θ_α and $\hat{\Psi}$ that

$$(\Theta_\alpha, H\hat{\Psi}) = 0 \tag{XI-6}$$

which from (3) tells us that Θ_α will not appear in $\hat{\Psi}^{(1)}$. In the context of UHF where as we will see the Θ_α satisfying (6) are one-electron excitations of $\hat{\Psi}$, this is known as Brillouin's theorem.[1] (In our earlier discussion of UHF we did not specifically invoke Eq. (V-3). However it clearly must be satisfied since in UHF one imposes no à priori reality restrictions). Therefore we will call it the generalized Brillouin theorem. More precisely, and quite apart from its application to perturbation theory, the generalized Brillouin theorem is the following: Let $\delta\eta\,\Psi'$ with $\delta\eta$ an infinitesimal but otherwise arbitrary complex number, and with Ψ' orthogonal to $\hat{\Psi}$, be a possible variation of $\hat{\Psi}$ within the set of trial functions, then it follows from (V-3) that

$$(\Psi', H\hat{\Psi}) = 0 \tag{VI-7}$$

In many ways then the generalized Brillouin theorem is really not a new result but merely a restatement of the variation method and is so used by many authors.[2]

Returning to perturbation theory, let us in particular consider UHF. Then a natural choice for H_0 is[3, A]

$$H_0 = F \equiv \sum_{s=1}^{N} f(s) \tag{IX-8}$$

Since F is a one electron operator this means that the Θ_α are single determinants involving 1, 2, ..N electron excitations of $\hat{\Psi}$ (For ground states of neutral systems these Θ_α are usually in the continuum).[4] We will now show that the one electron excitations do not appear in $\hat{\Psi}^{(1)}$ (and therefore do not affect the energy till fourth order). Proof:

For $\hat{\Psi}$ we have

$$\hat{\Psi} = \frac{1}{\sqrt{N!}} \{\hat{\varphi}_1, \ldots, \hat{\varphi}_N\} \tag{XI-9}$$

where the $\hat{\varphi}_i$ are orthonormal spin orbitals, and where we have included the factor $1/\sqrt{N!}$ so that $\hat{\Psi}$ will be normalized. Therefore the most general $\delta\hat{\Psi}$ is of the form

$$\delta\hat{\Psi} = \frac{1}{\sqrt{N!}} \sum_j |\hat{\varphi}_1, \ldots \delta\hat{\varphi}_j, \ldots \hat{\varphi}_N| \tag{XI-10}$$

where the $\delta\hat{\varphi}_j$ are arbitrary. If in particular we choose the $\delta\hat{\varphi}_j$ to be orthogonal to the $\hat{\varphi}_i$ but otherwise arbitrary, then (10) is an arbitrary sum of one-electron excitations of $\hat{\Psi}$ which proves the point.

Similarly for OHF the Ψ' which satisfy the generalized Brillouin theorem are a general superposition of two types: One electron excitation of the singly occupied orbitals without change in spin, and paired excitations of the doubly occupied orbitals, again without change in spin. Thus for the example (X-5) these types are

$$|\hat{v}_1\alpha\,\hat{v}_1\beta\,\delta\hat{v}_2\alpha|$$

and

$$|\delta\hat{v}_1\alpha\,\hat{v}_1\beta\,\hat{v}_2\alpha| + |\hat{v}_1\alpha\,\delta\hat{v}_1\beta\,\hat{v}_2\alpha|$$

where for orthogonality to $\hat{\Psi}$ we need

$$(\delta\hat{v}_2, \hat{v}_2) = (\delta\hat{v}_1, \hat{v}_1) = 0$$

The second type of excitation may then be further classified according to whether or not $\delta\hat{v}_1$ is orthogonal to $\hat{v}_2$. We will discuss the linear variation method later.

Returning now to UHF, the fact that with the choice $H_0 = F$, $\Psi^{(1)}$ contains no one electron excitations of $\hat{\Psi}$ has an interesting consequence:[9] there are no first order corrections to the average value of any one electron operator; or equivalently there are no first order corrections to the one-electron density matrix (Indeed the one-electron density matrix is itself the expectation value of a one electron operator), and hence to its eigenvalues, and to its eigenfunctions the natural spin orbitals. Proof: Let W be a one-electron operator. Then the first order correction to its expectation value is

$$(\Psi^{(1)}, W\hat{\Psi}) + (\hat{\Psi}, W\Psi^{(1)})$$

or

$$(\Psi^{(1)}, W\hat{\Psi}) + (W^{\dagger}\hat{\Psi}, \Psi^{(1)})$$

which, since from (3) $\Psi^{(1)}$ is orthogonal to $\hat{\Psi}$, can be written

$$(\hat{\Phi}^{(1)}, [W - (\hat{\Phi}, W\hat{\Phi})]\hat{\Phi}) + ([W^{\dagger} - (\hat{\Phi}, W^{\dagger}\hat{\Phi})]\hat{\Phi}, \hat{\Phi}^{(1)}) \tag{X-12}$$

However if W is a one-electron operator then, either from Slater's rules or from

$$W\hat{\Phi} = \sum_{s} |\hat{\varphi}_1, \dots, w\hat{\varphi}_s \dots \hat{\varphi}_N|$$

one readily sees that

$$[W - (\hat{\Phi}, W\hat{\Phi})]\hat{\Phi} \quad \text{and} \quad [W^{\dagger} - (\hat{\Phi}, W^{\dagger}\hat{\Phi})]\hat{\Phi}$$

involve only one-electron excitations of $\hat{\Phi}$ so that (12) vanishes as claimed.

Because of the freedom in the choice of H_0 this result though interesting, is rather more formal than physical. Thus for UHF we could certainly choose the Θ_α and hence H_0 so that none of the Θ_α was a pure one-electron excitation. In such a case then Brillouin's theorem would have no especially interesting consequences for UHF. For UHF applied to atoms a more physical result can be derived as follows:

We want to consider an isoelectronic sequence in the limit that nuclear charge $Z \to \infty$. Therefore to keep things under control we will, as usual, use scaled coordinates $Z\vec{r}_s$ and measure all energies and Hamiltonians in "units" of Z^2. For H we take the usual fixed nucleus Hamiltonian (divided by Z^2) possibly including external fields, and for $Z^2 H_0$ we will again use F. Then since H and H_0 differ only in their treatment of the Coulomb interaction between the electrons, and since in scaled coordinates, the Coulomb interaction is of order $1/Z$ it follows that

$$\lim_{Z\to\infty} (H - H_0) = O\left(\frac{1}{Z}\right) \qquad \text{(XI-13)}$$

and that therefore, if we denote by ψ the normalized eigenfunction of H to which $\hat{\psi}$ is an approximation and write

$$\psi = \hat{\psi} + \Delta \qquad \text{(XI-14)}$$

then Δ is at least of order $1/Z$. Further since $(\psi, \psi) = (\hat{\psi}, \hat{\psi}) = 1$ we have

$$(\hat{\psi}, \Delta) + (\Delta, \hat{\psi}) + (\Delta, \Delta) = 0 \qquad \text{(XI-15)}$$

i.e. $(\hat{\psi}, \Delta) + (\Delta, \hat{\psi})$ is at least of order $(1/Z)^2$.

Now consider any operator O such that in scaled coordinates

$$\lim_{Z\to\infty} O = O_1 + O' \qquad \text{(XI-16)}$$

where O_1 is a one-electron operator and where O' is at least of order $1/Z$ with respect to O. (Thus H itself is an example of such an operator). Then correct through terms of relative order $1/Z$ the exact average of O is

$$(\psi, O\psi) = (\hat{\psi}, O\hat{\psi}) + (\hat{\psi}, O\Delta) + (\Delta, O\hat{\psi})$$

which to the same order can be written

$$(\psi, O\psi) = (\hat{\psi}, O\hat{\psi}) + \left([O_1^\dagger - (\hat{\psi}, O_1^\dagger\hat{\psi})]\hat{\psi}, \Delta\right) + \left(\Delta, [O_1 - (\hat{\psi}, O_1\hat{\psi})]\hat{\psi}\right) \qquad \text{(XI-17)}$$

since as noted above

$$\left((\hat{\psi}, O_1^\dagger\hat{\psi})\hat{\psi}, \Delta\right) + \left(\Delta, (\hat{\psi}, O_1\hat{\psi})\hat{\psi}\right) = (\hat{\psi}, O\hat{\psi})[(\hat{\psi}, \Delta) + (\Delta, \hat{\psi})]$$

is of relative order $1/Z^2$. But now O_1 and $O_1^\dagger$ are one-electron operators and since from our previous formal result we know that to first order in ($H - H_0$), and therefore at least through first order in $1/Z$, Δ contains no one-electron excitations of $\hat{\Psi}$ the last two terms in (9) vanish to this order. Thus we have

$$(\Psi, O\Psi) = (\hat{\Psi}, O\hat{\Psi})\left[1 + O\left(\frac{1}{Z^2}\right)\right] \tag{XI-18}$$

That is averages calculated using a UHF $\hat{\Psi}$ for any operator satisfying (16), and hence in particular for any one electron operator, are accurate through terms of relative order $1/Z$. Since, as we have noted, H satisfies (8) it then follows that, if we reinstate the factor Z^2 and write

$$\hat{E} = Z^2 \hat{E}_0 \left(1 + \frac{\hat{a}}{Z} + \frac{\hat{b}}{Z^2} + \cdots\right)$$

and if we write the corresponding eigenvalue in a similar way

$$E = Z^2 E_0 \left(1 + \frac{a}{Z} + \frac{b}{Z^2} + \cdots\right)$$

then $\hat{E}_0 = E_0$ and $\hat{a} = a$ but $\hat{b} \neq b$. Thus the correlation energy $E - \hat{E}$ is in first approximation independent of Z. Note also that these results for UHF are true in arbitrary external fields and therefore hold for all manner of polarizabilities, susceptibilities, etc. (For calculations which, among other things, illustrate these points we refer the interested reader to a series of papers by A. Dalgarno and collaborators, notably M. Cohen, which have been published mainly in Proc. Roy. Soc. and Proc. Phys. Soc. starting around 1960).

Having said all this in great detail we now want to make two further points, one major and one minor. First the minor point. Having understood the derivation in detail it is clearly possible to simplify it drastically: In the $\mathcal{Z} \to \infty$ limit $\hat{\Psi}$ differs from Ψ by terms of relative order $1/\mathcal{Z}$. Therefore for any operator O which in the $\mathcal{Z} \to \infty$ limit becomes a one-electron operator, it follows from our earlier formal result that the $1/\mathcal{Z}$ correction to its UHF expectation value vanishes identically, which of course yields (18). For $O = H$, (18) can also be inferred directly from the variation principle, i.e. $\hat{E}$ involves only a second order error. In Sec. XVI we will give another derivation of these results, a derivation which will show that (18) also holds for OHF provided that H_1 and O_1 are both spin free.

Now for the major point. This is that the preceding arguments contain a flaw, and are not completely valid! The reason is connected with the peculiar degeneracy of hydrogenic energy levels. Thus for example consider the ground state of the Be atom in the absence of external fields. In the $\mathcal{Z} \to \infty$ limit $\hat{\Psi}$ which is a single determinant, becomes the function $(1s)^2(2s)^2\ {}^1S$. On the other hand the correct result is a certain linear combination of the degenerate pair $(1s)^2(2s)^2\ {}^1S$ and $(1s)^2(2p)^2\ {}^1S$.[6] Thus something is wrong with our argument since $\hat{\Psi}$ and Ψ don't agree even in zero order! Incidently note that the MCSCF schemes can avoid this difficulty, and indeed to some extent the MCSCF method was first introduced to deal with this problem.[7] The difficulty is of course that our estimate of the order of $\hat{\Psi}^{(1)}$ is incorrect since as $\mathcal{Z} \to \infty$, there is an energy denominator in $\hat{\Psi}^{(1)}$ which, instead of being of order 1 is of order $1/\mathcal{Z}$ like the numerator, thus making $\hat{\Psi}^{(1)}$ actually of order 1 rather than of order $1/\mathcal{Z}$, and similarly for $\hat{\Psi}^{(2)}$ etc.

On the other hand for an isolated Lithium atom the $(1s)^2(2s)^2\,{}^2S$ and $(1s)^2\,(2p)^2\,{}^2P$ degeneracy causes no problem since they have quite different symmetry, but in external fields Li becomes a case in point as well.[8] However it should be noted that in any case our earlier formal argument is still generally valid. More precisely if we introduce an order parameter λ and write

$$H(\lambda) = F + \lambda(H-F)$$

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 $(1/Z)^n$.

Well what can one learn from all this as regards the accuracy of UHF? Since none of the arguments applies to two, three, etc. electron operators one expects, and one finds, that expectation values of one electron operators or more generally operators like H which satisfy (18) are given more accurately by UHF than are expectation values of two electron operators. However the arguments about the order of accuracy are a bit shaky. We have already noted the formal character of our first argument and the 1/Z statements for atoms strictly apply only in the $Z \to \infty$ limit and therefore not to neutral or near neutral atoms. Nevertheless as a general rule UHF does quite well as regards one-electron properties,[9] however there is definite evidence that for some one-electron properties of molecules, second and higher order corrections are not negligible and, in particular that one electron excitations of $\hat{\Psi}$ which, as we have seen, don't appear in $\hat{\Psi}^{(1)}$ with the choice of $H_0 = F$, but which can appear in $\hat{\Psi}^{(2)}$ etc., can have a non-negligible effect.[10]

Turning briefly to another case, let $\hat{\Psi}$ be a $\hat{\Psi}_k$ of the linear variation method. Then a natural choice for H_0 is $\overline{H}$ or some linear combination of $\overline{H}$ and $(1-\Pi)H(1-\Pi)$. With this choice, all the $\hat{\Psi}_L$ are eigenfunctions of H_0 and therefore from (VI-9) and (VI-10) (i.e. the generalized Brillouin theorem) it follows that the $\hat{\Psi}_L$ with $L \neq k$ will not appear in $\hat{\Psi}^{(1)}$. Finally let us note that a common means for improving a $\hat{\Psi}$ is to do a further linear variation calculation with the basis set Φ_k consisting of $\hat{\Psi}$ and some other functions χ_α. The generalized Brillouin theorem then tells us that if a χ_α satisfies the conditions of the theorem then it will not be directly coupled to $\hat{\Psi}$ in the secular equation. Also if, as is becoming increasingly popular,[11] one solves the linear variation problem by a perturbation technique then, with appropriate choice of $\overline{H}_0$, χ_α will not appear in $\hat{\Psi}^{(1)}$ and will not affect the energy till $\hat{E}^{(4)}$.

XII. SPECIAL THEOREMS SATISFIED BY OPTIMAL TRIAL FUNCTIONS - INTRODUCTION

Eigenfunctions of H satisfy various physically interesting and useful special theorems -- Hypervirial theorems, generalized Hellmann-Feynman theorems, etc. and may have certain symmetries. In this section we will show how one can choose the set of trial functions in such a way as to be sure a priori that the optimal trial functions $\hat{\Psi}$ will have analogous properties. These theorems, when applicable, can then provide physical insight into the nature of the $\hat{\Psi}$ and

$\hat{E}$, and the degree to which they approximare the behavior of actual eigenfunctions and eigenvalues. Also if one cannot determine $\hat{\psi}$ and $\hat{\epsilon}$ exactly, the extent to which the applicable theorems are satisfied by the approximate $\hat{\psi}$ and $\hat{\epsilon}$ can give one an indication of how accurate the approximation is; for example how accurately an AUHF calculation approximates UHF.

Of course one approach to symmetries is the one which we have mentioned repeatedly - constrain the $\tilde{\psi}$ so that each has the desired symmetry. Also in the last few years there has developed a considerable literature in which, usually through the use of Lagrange multiplier techniques, the $\hat{\psi}$ are constrained to have various properties and satisfy various theorems.[1] However in what follows we will be interested in more general possibilities in which the symmetries and/or theorems are satisfied "naturally".

In all cases we will give only sufficient conditions, and it seems that one can hardly do better than this in any useful way because any set of $\tilde{\psi}$ might contain, as one unique member, an eigenfunction which had all the desired properties. However it is empirically an excellent rule of thumb that if the sufficient conditions we give are not met, then it is very likely that the $\hat{\psi}$ won't have the desired properties. Presumably this is the case because the sufficient conditions which we will give are rather natural ones.

XIII. REALITY

If, in the representation in which one is working, H is explicitly real then, as is well known, if an eigenvalues is non degenerate, the corresponding eigenfunction will automatically be essentially real (i.e. $\psi = \alpha \psi^*$ where α is some constant) while if the eigenvalue is degenerate the eigenfunctions may be chosen to be real, an arbitrary degenerate eigenfunction then being some linear combination of the real ones.

We will now show that if H is real, and if the set of trial functions is invariant to complex conjugation,

then if $\hat{E}$ is non degenerate, $\hat{\phi}$ will be essentially real.

We first note that since $\hat{E}$ is anyway real, the assumption that H is real implies that

$$\hat{E} = \frac{(\hat{\phi}, H\hat{\phi})}{(\hat{\phi}, \hat{\phi})} = \frac{(\hat{\phi}, H\hat{\phi})^*}{(\hat{\phi}, \hat{\phi})^*} = \frac{(\hat{\phi}^*, H\hat{\phi}^*)}{(\hat{\phi}^*, \hat{\phi}^*)} \qquad \text{(XIII-1)}$$

Thus $\hat{\phi}$ and $\hat{\phi}^*$ yield the same energy. However under our assumption both $\hat{\phi}$ and $\hat{\phi}^*$ are in the set of trial functions, and therefore if $\hat{E}$ is non degenerate it follows that $\hat{\phi}$ and $\hat{\phi}^*$ must be proportional to one another as we wanted to prove. UHF and OHF, and the linear variation method if the ϕ_k can be chosen real, are examples involving sets of trial functions which are invariant to complex conjugation.

If $\hat{E}$ is degenerate there does not seem to be a simple general theorem unless one has a situation in which the trial functions can be labelled in such a way that they are complex only because certain <u>linear</u> variational parameters and/or functions can be complex. Then we can show[1] that the $\hat{\Psi}$ are either automatically essentially real or if there is degeneracy can be chosen real, an arbitrary degenerate $\hat{\Psi}$ then being some linear combination of the real $\hat{\Psi}$.

Proof: In such a case since, for fixed values of the non linear quantities, the real functions $(\tilde{\Psi} + \tilde{\Psi}^*)$ and $i(\tilde{\Psi} - \tilde{\Psi}^*)$ also belong to the set, it follows that as far as varying the linear quantities is concerned we can deal with a real basis set. For fixed values of the non linear quantities, the optimal trial functions are thus eigenfunctions of a real $\bar{H}$. Therefore if there is no degeneracy at this stage, the optimal function is real and hence, under our assumptions, will stay real as one determines the optimal values of the non linear quantities. On the other hand if there is degeneracy at the linear stage then we can anyway write our trial functions at this stage as a linear combination of say $\mathcal{D}$ real functions $\check{\Psi}_\alpha$ which, from (VI-9) and (VI-10) satisfy

$$(\check{\Psi}_\alpha, \check{\Psi}_\beta) = \delta_{\alpha\beta} \qquad (\check{\Psi}_\alpha, H\check{\Psi}_\beta) = \check{E}\,\delta_{\alpha\beta};\ \alpha,\beta = 1 \ldots \mathcal{D} \quad \text{(XIII-2)}$$

and where $\check{E}$ and the $\check{\Psi}_\alpha$ depend on the non linear quantities. But now going on to vary the non linear quantities one sees that a general $\tilde{\check{\Psi}}$, which will now have the form $\sum_{\alpha=1}^{\mathcal{D}} \tilde{a}_\alpha \check{\Psi}_\alpha$, yields the energy $\check{\tilde{E}}$ whatever the values of the $\tilde{a}_\alpha$ since from (2) we have

$$\frac{(\tilde{\psi}, H\tilde{\psi})}{(\tilde{\psi}, \tilde{\psi})} = \frac{\sum_\alpha |\tilde{a}_\alpha|^2 \check{E}}{\sum_\alpha |\tilde{a}_\alpha|^2} = \check{E}$$

Therefore the degeneracy persists and the $\tilde{a}_\alpha$ are left undetermined by variation of the non linear quantities. Thus an arbitrary degenerate $\hat{\psi}$ will be some linear combination of the real $\hat{\psi}_\alpha$, which completes the proof.

One final note: Whether or not $\hat{\psi}^*$ was in the set of trial functions, one can in any case produce optimal real functions by doing a <u>further</u> linear variation calculation with $\hat{\psi}$ and $\hat{\psi}^*$ as the basis set. Proof: Instead of using $\hat{\psi}$ and $\hat{\psi}^*$ as the basis for this further calculation we may equally well use the real basis set $\chi_1 = \psi + \hat{\psi}^*$ and $\chi_2 = i(\hat{\psi} - \hat{\psi}^*)$. Therefore it follows from the above discussion that the result will be two real (orthogonal) linear combinations of χ_1 and χ_2, and of course, as a bonus, at least one of these combinations will have an energy less than (or at least not greater than) the original $\hat{E}$. Indeed even without a further detailed calculation one can see that one or the other of the real functions χ_1 or χ_2 will themselves yield a lower energy, or at least an energy which is no higher than $\hat{E}$! Proof: From (1) and the fact that $(\hat{\psi}, \hat{\psi}) = (\hat{\psi}^*, \hat{\psi}^*)$ one can easily verify that

$$\hat{E} = \left[(\chi_1, H\chi_1) + (\chi_2, H\chi_2)\right] \Big/ \left[(\chi_1, \chi_1) + (\chi_2, \chi_2)\right]$$

which is certainly not less than the smaller of $(\chi_1, H\chi_1)/(\chi_1, \chi_2)$ and $(\chi_2, H\chi_2)/(\chi_2, \chi_2)$. However in general neither χ_1 nor χ_2 will be optimal functions in the sense of this paragraph. We will discuss this point further in Section XV below.

XIV. UNITARY INVARIANCE

The eigenvalues of H are invariant to a unitary transformation of H. Let U be a unitary operator, and suppose that one uses the same set of trial functions in the variation calculation for U^+HU as one uses in the calculation for H. Then we can show that if the set of trial functions is invariant to the transformation U then the $\hat{E}$ are invariant.

. The proof is as follows: To find the $\hat{E}$ we look for the stationary values of

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

as $\tilde{\Psi}$ ranges through the set. To find variational approximations to the eigenvalues of U^+HU we look for the stationary values of

$$\tilde{E}^U = (\tilde{\Psi}, U^+HU\tilde{\Psi})/(\tilde{\Psi},\tilde{\Psi})$$

which, since U is unitary, we can write as

$$\tilde{E}^U = (U\tilde{\Psi}, HU\tilde{\Psi})/(U\tilde{\Psi},U\tilde{\Psi})$$

If now the set of $U\tilde{\Psi}$ is the same as the set of $\tilde{\Psi}$ it is clear that the stationary values of $\tilde{E}^U$ and $\tilde{E}$ will be the same, which proves the point. Note that if the set is invariant to U then it is also invariant to all powers of U and to $U^{-1} = U^+$.

In particular a spatial translation is a unitary transformation, the operator U being $\exp i\vec{a}\cdot\vec{P}$ where $\vec{a}$ is the amount of the translation and $\vec{P}$ is the momentum operator for the particle or particles being translated.[1] Thus if the set of trial functions is invariant to a rigid translation of the electronic coordinates, then one will get the same $\hat{E}$ whether one uses H as given or whether in H one replaces the $\vec{r}_s$ by $\vec{r}_s + \vec{\alpha}$ where $\vec{\alpha}$ is a constant vector, leaving everything else unchanged. Similar remarks apply to rotations where $U = \exp i \vec{\theta} \cdot \vec{J}$ with $\vec{\theta}$ the angle (and axis) of rotation and $\vec{J}$ the angular momentum of the rotated particle or particles;[1] and to inversion where U is the product of inversions applied to each particle separately.

Since the operator U for a rigid translation, or rotation, or inversion of all the electrons takes the form

$$U = \prod_{s=1}^{N} u(s) \tag{XIV-1}$$

they transform a Slater determinant into a Slater determinant:

$$\prod_{s=1}^{N} u(s) \, | \tilde{\varphi}_1 \cdots \tilde{\varphi}_N | = | u\tilde{\varphi}_1 \; u\tilde{\varphi}_2 \cdots u\tilde{\varphi}_N |$$

Therefore UHF is invariant to rigid rotations, translations and inversions, and to any other U of this type. Further if we consider only spatial rotations, i.e. we don't rotate the spins, then since the U's for such rotations, and for translations, and inversion are spin independent it follows that OHF is also invariant to such rotations and to translations and inversions. The situation with respect to analytic approximations depends of course on the nature of the basis sets.

Another interesting U having the structure (3) is that of a gauge transformation.[1] Thus we have the result that UHF is gauge invariant and further, since the appropriate $u(s)$ is spin independent we can assert that OHF is also gauge invariant, with again the properties of AUHF etc. depending on the nature of the basis sets.

As a final example of a unitary transformation consider the transformation from the coordinate representation to the momentum representation. Tn the abstract operator approach we have been using if $H = H(\vec{r}_1 \cdots \vec{r}_i \cdots)$ then $U^{+}HU = H(-\vec{r}_1, \cdots \vec{r}_i \cdots)$ and $U^{+}\tilde{\Psi}(\vec{r}_1 \cdots) = \tilde{\Phi}(\vec{r}_i \cdots)$ where $\tilde{\Phi}$ is the Fourier transform of $\tilde{\Psi}$. (Here and in what follows we will suppress spin labels unless needed). That is if in momentum space we continue to use the symbol $\vec{r}$ to denote the independent variable, then momentum is represented by $\vec{r}$ and coordinate by $-\frac{\hbar}{i}\vec{\nabla} \equiv \vec{p}$. Thus for invariance the set of $\tilde{\Phi}(\vec{r}_i \cdots)$ should be the same as the set of $\tilde{\Psi}(\vec{r}_i \cdots)$. Since here too U is a product of spin independent transformations of each particle separately, it follows that the UHF and OHF are invariant, i.e. the Fourier transformation of a Slater determinant is a Slater determinant. Also since orbital angular momentum is symmetric in coordinate and momentum it should be no surprise that the Fourier transform of a spherical harmonic is a spherical harmonic and therefore that most RHF approximations for atoms are invariant to transformation to momentum space.[2]

In many interesting cases a transformation U of the dynamical variables is equivalent to a change in some parameter(s) σ in the Hamiltonian:

$$U^{+} H(\sigma) U = H(\sigma') \qquad ; \sigma' = f(\sigma) \qquad (XIV-2)$$

The eigenvalues of H are therefore invariant to the transformation $\sigma \rightarrow f(\sigma)$, and evidently H is invariant to the combined transformations U and $f(\sigma) \rightarrow \sigma$. We will now show that the $\hat{E}$ will be invariant to the transformation $\sigma \rightarrow f(\sigma)$ if the set of trial functions is invariant to the combined transformations U^{+} and $f(\sigma) \rightarrow \sigma$.

Proof:

$$\tilde{E}(\sigma') = (\tilde{\Phi}(\sigma'), H(\sigma')\tilde{\Phi}(\sigma'))/(\tilde{\Phi}(\sigma'), \tilde{\Phi}(\sigma'))$$

which from (2) can be written as

$$\tilde{E}(\sigma') = (U\tilde{\Phi}(\sigma'), H(\sigma) U\tilde{\Phi}(\sigma'))/(U\tilde{\Phi}(\sigma'), U\tilde{\Phi}(\sigma')).$$

Therefore if the set $U\tilde{\Phi}(\sigma')$ is the same as the set $\tilde{\Phi}(\sigma)$, the $\hat{E}(\sigma')$ will be the same as the $\hat{E}(\sigma)$, which proves the point. In particular if the set is independent of σ then the $\hat{E}$ will be invariant to $\sigma \rightarrow f(\sigma)$ if the set of trial functions is invariant to the transformation U^{+}.

The special case U = 1 is also of interest. Thus suppose that H is invariant to some transformation $\sigma \rightarrow f(\sigma)$. Then we have the result that if the set of trial functions is invariant to the SAME transformation, then the $\hat{E}$ will be invariant.

XV. SYMMETRY

If H commutes with a unitary operator U then there are certain consequences which we now want to discuss. However first we must settle a point of notation: If H commutes with a Hermitian operator T then it commutes with the whole set of unitary operators $\exp i\alpha T$ where α is a real parameter, and conversely if it commutes with the set it commutes with T . In such a case we will use the symbol U to represent the whole set, so that for example the statement that a wave function is an eigenfunction of U will therefore also mean that it is an eigenfunction of T .

If H commutes with U then if an eigenvalue of H is nondegenerate, the corresponding eigenfunction is automatically an eigenfunction of U while if the eigenvalue is degenerate one can find a set of functions which are simultaneous eigenfunctions of H and U and such that an arbitrary degenerate eigenfunction is some linear combination of the members of this set.

We will now show that if H commutes with U, and if the set of trial functions is invariant to U , then if $\hat{E}$ is nondegenerate, the corresponding $\hat{\Psi}$ is an eigenfunction of U. The proof follows a pattern similar to that in Sec. XIII. We observe that if H commutes with U then

$$\hat{E} = \frac{(\hat{\Psi}, H\hat{\Psi})}{(\hat{\Psi}, \hat{\Psi})} = \frac{(\hat{\Psi}, U^{\dagger} H U \hat{\Psi})}{(\hat{\Psi}, U^{\dagger} U \hat{\Psi})} = \frac{(U\hat{\Psi}, HU\hat{\Psi})}{(U\hat{\Psi}, U\hat{\Psi})} .$$

Thus both $\hat{\Psi}$ and $U\hat{\Psi}$ yield the same energy. Therefore since both are in the set it follows that if $\hat{E}$ is nondegenerate, then $\hat{\Psi}$ and $U\hat{\Psi}$ must be proportional which is what we want to prove.

We now turn to the degenerate case. Eigenvalues of H usually are degenerate because there exist U's which commute with H but not with each other. Such degeneracies however can always be removed by applying suitable extra external fields so that in the fields all the U's which commute with H do commute with one another. <u>If</u> in addition such external fields serve to remove the degeneracy in the variational calculation, then we know from the discussion of the previous paragraph that the $\hat{\Psi}$ in the fields will automatically be eigenfunctions of those U's which commute with H in the fields, and which leave the set invariant. If now we let the fields tend to zero, but continue to use the same set of trial functions, it follows that when the external fields have been reduced to zero the $\hat{\Psi}$ will still be eigenfunctions of these U's (we are of course assuming that the U's are independent of the external fields). Since different external fields single out different U's we therefore have the result that , if there is degeneracy then, among the degenerate $\hat{\Psi}$, will be eigenfunctions of any set of U's which commute with H and with each other, and which leave the set of trial functions invariant. However whether or not an arbitrary degenerate $\hat{\Psi}$ can be written as a linear combination of the degenerate $\hat{\Psi}$ which are also eigenfunctions of a particular set of U's will depend on how much linearity there is in the set of trial functions.

The above discussion, though quite general, contains the qualification that all degeneracies in the variational calculation should be removable if one would only apply suitable external fields. Therefore its applicability to actual calculations is not immediately evident. Nevertheless it is clearly consistent with the results discussed in Section X for atomic UHF. Thus $e^{i\alpha L_z}$ and $e^{i\alpha S_z}$ and the U for parity are all products of single particle operations and therefore leave the set of $\tilde{\Psi}$ invariant while $e^{i\alpha S^2}$ and $e^{i\alpha L^2}$ are not. Therefore the present discussion would correctly suggest that for a given $\hat{E}$, UHF $\hat{\Psi}$ can be found which are eigenfunctions of parity, of a component of $\vec{L}$ and of a component of $\vec{S}$, but that in general one will not find eigenfunctions of either L^2 or S^2. Also consistent with this point of view is the fact that in those cases in which one does find eigenfunctions of L^2 and S^2 this is usually forced by the behavior with respect to L_z and S_z. Thus for example a closed shell state is a simultaneous eigenfunction of all components of $\vec{L}$ and $\vec{S}$ with eigenvalues zero and therefore must be an eigenfunction of L^2 and S^2 with eigenvalues zero. Also the function mentioned before (X-3) is a quartet function "because" it has the maximum S_z possible for the given N .

Less problematical but more specialized is the following theorem which is similar to one in Sec. XIII (a special case of this theorem was discussed in footnote A, Sec. VII): one can find $\hat{\Psi}$ which are eigenfunctions of U, and such that an arbitrary degenerate $\hat{\Psi}$ is some linear combination of these eigenfunctions of U, if the set of trial functions is invariant to U in such a way that U induces changes only in linear parameters and/or functions; that is if $U\tilde{\Psi}$ involves the same non linear quantities as $\tilde{\Psi}$, but possibly different linear ones. The proof follows from the fact that under these conditions the $\overline{\Pi}$ appropriate to the linear part of the calculation will commute with U and therefore so will $\overline{H}$.

The pattern of the proof is then identical to the proof of the analogous theorem in Sec. XIII.

Also, again analogously to the discussion in Sec. XIII, whether or not the functions $\hat{\psi}, U\hat{\psi}, U^2\hat{\psi}, \ldots U^{+}\hat{\psi} \ldots$ all of which yield the same energy, are all in the set, one can in any case produce an optimal set of functions which are eigenfunctions of U, by doing a further linear variation calculation in the space spanned by these functions (U here could represent a complete commuting set of operators). That this will produce eigenfunctions of U is guaranteed by the fact that this linear space is obviously invariant to the action of U, and therefore the $\overline{H}$ appropriate to it will commute with U. Indeed it is easy to see what the result of this calculation will be. Namely we can expand $\hat{\psi}$ in normalized eigenfunctions of U belonging to different eigenvalues, thus

$$\hat{\psi} = \sum_{n'} A_{n'} \zeta_{n'} \tag{XV-1}$$

where the sum may be infinite, though hopefully it is only finite. Then since $U^{m}\hat{\psi}$ and $U^{+m}\hat{\psi}$ are simply linear combinations of these same functions with different coefficients, it is clear that the functions $\zeta_{n'}$ involved in the sum span the linear space formed from $\hat{\psi}, U\hat{\psi} \ldots$. Further since

$$(\zeta_{n'}, \zeta_{n''}) = \delta_{n'n''} \qquad (\zeta_{n'}, H\zeta_{n''}) = 0 \quad n' \neq n'' \tag{XV-2}$$

it follow from the "converse theorem" (V) of Sec. VI that the $\zeta_{n'}$ will be the $\hat{\psi}_{K}$ which would result from the linear variation calculation. Thus instead of doing the linear variation calculation we can simply project out of $\hat{\Psi}$ the various symmetry components which it contains. This procedure, and approximations to it have been extensively applied to UHF functions, particularly to produce functions of a definite total spin S^2 (this is a case in which $U\hat{\Psi}, U^2\hat{\Psi}, \ldots$ don't belong to the original set since, as noted earlier S^2 is not a one particle operator). Of course in general one can do even better if, as mentioned in Sec. VIII, one projects before carrying out the original variation calculation.

There are certain similarities between the procedure we have just discussed and that in Sec. XIII. However there are also certain differences and we would now like to draw attention to these in more detail. If we write, in the notation of Sec. XIII

$$\hat{\Psi} = \tfrac{1}{2}\chi_1 - \tfrac{1}{2}(i\chi_2)$$

then we have written $\hat{\Psi}$ as a linear combination of (unnormalized) eigenfunctions of the complex conjugation operator K belonging to different eigenvalues:

$$K\chi_1 = \chi_1 \qquad K(i\chi_2) = -(i\chi_2)$$

So here we have a certain similarity with (1). However as we remarked in Sec. XIII, χ_1 and χ_2 are in general not the functions which result from the further linear variation calculation and this is to be contrasted with the result of the present section, that the $\zeta_{m'}$ are the functions which result from a further linear variation calculation. The difference arises because while U is a linear operator of the familiar sort, K is not. Rather it is what is called anti-linear: $K\alpha\tilde{\Psi} = \alpha^* K\tilde{\Psi} = \alpha^*\tilde{\Psi}^*$ if α is a number, and thus is neither Hermitian nor unitary. Rather one has

$$(\chi, K\chi') = (\chi, \chi'^*) = (\chi^*, \chi')^* = (K\chi, \chi')^*$$

In particular then, from $K\chi_1 = \chi_1$ and $K(i\chi_2) = -(i\chi_2)$ one concludes, not that χ_1 and $i\chi_2$ are orthogonal but only that $(\chi_1, i\chi_2)$ is pure imaginary, which is anyway obvious: Thus following the canonical pattern

$$(\chi_1, Ki\chi_2) = -(\chi_1, i\chi_2)$$

but also

$$(\chi_1, Ki\chi_2) = (K\chi_1, i\chi_2)^* = (\chi_1, i\chi_2)^*$$

which completes the proof. Note also that although χ_1 and $i\chi_2$ belong to different eigenvalues of K, χ_1 and χ_2 belong to the same eigenvalue. Similarly from the fact that $H = H^*$ one can not conclude that $(\chi_2, H i\chi_1) = 0$. Finally let us note that in a certain sense χ_1 and χ_2 are not very well defined, since, if α is a number, $\hat\varphi$ and $\hat\varphi' = e^{i\alpha}\hat\varphi$ are physically equivalent yet χ_1' and χ_2' are quite different from χ_1 and χ_2.

XVI. GENERALIZED HELLMANN - FEYNMAN THEOREM

Suppose that H contains a real parameter σ. Then by differentiating

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

with respect to σ we find

$$\left(\frac{\partial\hat\varphi}{\partial\sigma}, (H-\hat E)\hat\varphi\right) + \left(\hat\varphi, (H-\hat E)\frac{\partial\hat\varphi}{\partial\sigma}\right) + \left(\hat\varphi, \left(\frac{\partial H}{\partial\sigma} - \frac{\partial\hat E}{\partial\sigma}\right)\hat\varphi\right) = 0 \tag{XVI-1}$$

where in carrying out the differentiations we have of course kept the integration variables fixed. In this connection it should be especially noted that if one changes variable then in general $\partial H/\partial\sigma$ will change if the change of variable is σ dependent. Also we have assumed that the volume element in the integration does not depend on σ. For many cases of interest in which the volume element does depend on σ the dependence is only multiplicative and hence cancels out.[1]

We will now show that if the set of trial functions is invariant to changes in the value of σ then the first line of Eq. (1) will be separately equal to zero so that we will have

$$\left(\hat{\psi}, \left(\frac{\partial H}{\partial\sigma} - \frac{\partial\hat{E}}{\partial\sigma}\right)\hat{\psi}\right) = 0 \qquad \text{(XVI-2)}$$

which is the variational version of the generalized Hellmann-Feynman theorem for σ.[A] The proof is as follows:

Though as a whole the set of trial functions is, by assumption, independent of σ, still which particular members of the set are selected by the variational method as optimal trial functions will in general depend on the value of σ. Therefore let $\hat{\psi} = \hat{\psi}(\sigma_1)$ be an optimal trial function when $\sigma = \sigma_1$, and let $\hat{\psi}(\sigma_1 + \delta\sigma_1)$ be the corresponding optimal trial function for a slightly different value of σ (we are assuming that the dependence on σ is continuous). Now by our assumption $\hat{\psi}(\sigma_1)$ and $\hat{\psi}(\sigma_1 + \delta\sigma_1)$ both belong to the set from which $\hat{\psi}(\sigma_1)$ was selected. Therefore $\hat{\psi}(\sigma_1 + \delta\sigma_1)$ was one of the neighboring functions which was examined in testing for the stationarity of $\hat{E}(\sigma_1)$. This in turn implies that

$$\frac{\partial \hat{\psi}}{\partial \sigma_1} \delta\sigma_1$$

must be a $\delta\hat{\psi}$ satisfying (IV-3); that is, cancelling the factor $\delta\sigma_1$ we must have

$$\left(\frac{\partial \hat{\psi}}{\partial \sigma_1} , (H-\hat{E}) \hat{\psi} \right) + \left(\hat{\psi}, (H-\hat{E}) \frac{\partial \hat{\psi}}{\partial \sigma_1} \right) = 0$$

which, since σ_1 could be any value of σ, proves the point.

The theorem which we have just proven, is in its essentials due to Hurley,[2] and we would like to emphasize its simplicity and its generality since this does not seem to be widely enough appreciated. In particular there are in the literature (subsequent to Hurley's work) many very detailed derivations of special Hellmann-Feynman theorems (i.e. special choices of σ) for particular variational approximations; derivations which are quite unnecessary since the results are immediate consequences of Hurley's theorem.

Obvious examples of sets of trial functions which are invariant to changes in σ are (i) the trial functions of most SCF (UHF, OHF, restricted AF etc. etc) type approximations since there are usually no à priori requirements as to how the spin orbitals should depend on possible σ like nuclear charge, nuclear configuration, strength of external fields, etc. (ii) the trial functions of analytic SCF approximations if the basis set is independent of σ, (iii) linear spaces in which the basis set is independent of σ.

In the latter two cases, if the basis set is fixed, invariance requires either that the individual basis functions don't involve σ at all or more generally that $\phi_k(\sigma+\delta\sigma)$ is a linear combination of the

$\phi_k(\sigma)$. However if the basis functions also involve non linear parameters (thus not really a linear space) then more flexibility is possible.

As an application of the generalized Hellmann-Feynman theorem we will use it to derive, and extend, some of the results found in Sec. XI. In the spirit of double perturbation theory let us include in the "Hamiltonian" an additive term of the form $\mu\Theta$ where μ is a real parameter and where we will be interested in the limit $\mu \to 0$. If the set of trial functions is invariant to changes in μ then we will have

$$(\hat{\Psi}, \Theta\hat{\Psi})/(\hat{\Psi}, \hat{\Psi}) = \frac{\partial \hat{E}}{\partial \mu} \tag{XVI-3}$$

Now suppose that in the limit $\gamma \to 0$, where γ is some other parameter, $\hat{\Psi}$ becomes an eigenfunction of the "Hamiltonian". Then barring problems with degeneracy the error in $\hat{\Psi}$ must be of order $\gamma(1+\beta\mu)$ where β is some constant, and therefore the error in $\hat{E}$ will be of order $\gamma^2(1+\beta\mu)^2$. But now if Ψ is the eigenfunction to which $\hat{\Psi}$ approximates, and if E is the corresponding eigenvalue then

$$(\Psi, \Theta\Psi)/(\Psi, \Psi) = \frac{\partial E}{\partial \mu} \tag{XVI-4}$$

Comparing (3) and (4) we see that the error in the expectation value of Θ is the same as the error in $\partial\hat{E}/\partial\mu$ which from the above is of order γ^2 in the limit $\mu \to 0$.

This result then evidently includes that of Sec. XI for UHF as a special case if we identify γ with $1/Z$. Also for OHF we see that, with the further qualification that $\mathcal{O}_1$ and H_1 be spin independent, we have a similar result since in the $\mathcal{Z}_0 \to \infty$ limit the "Hamiltonian" $H + \mu\mathcal{O}$ will be a spin free one-electron operator and hence will have the $\hat{\Psi}$ of the OHF as eigenfunctions.[3]

XVII. HYPERVIRIAL THEOREMS - GENERAL

Let $\mathcal{W}$ be a Hermitian operator and suppose that among the variations of $\hat{\Psi}$ which are possible within the set of trial functions is

$$i\,\delta\alpha\,\mathcal{W}\,\hat{\Psi} \tag{XVII-1}$$

where $\delta\alpha$ is a small real parameter. Then (IV-3) must be satisfied with $\delta\hat{\Psi}$ given by (1). Thus we have

$$(i\,\delta\alpha\,\mathcal{W}\hat{\Psi}, (H-\hat{E})\Psi) + (\hat{\Psi}, (H-\hat{E})\,i\,\delta\alpha\,\mathcal{W}\hat{\Psi}) = 0 \tag{XVII-2}$$

which immediately simplifies to

$$(\Psi, (H\mathcal{W} - \mathcal{W}H)\Psi) = 0 \tag{XVII-3}$$

and we have the result that under the given conditions, $\hat{\Psi}$ satisfies the hypervirial theorem for $\mathcal{W}$.[1]

We now note that (1) is the term of first order in $\delta\alpha$ in the expression

$$e^{i\delta\alpha\mathcal{W}}\,\hat{\Psi} - \hat{\Psi} \tag{XVII-4}$$

From this observation it then follows that a sufficient condition for $\hat{\Psi}$ to satisfy the hypervirial theorem for $\mathcal{W}$ is that the set of trial functions be invariant with respect to all unitary transformations $V = e^{i\alpha\mathcal{G}}$ where α is an arbitrary real number. Proof: If the set is invariant to such transformations then the first term on the left hand side of (4) will be a function near to $\hat{\Psi}$ in the set and therefore (1) will be a possible variation of $\hat{\Psi}$ within the set, whence the result follows.

It should be clear that this condition is only sufficient and not necessary. Thus we really need only that (1) be a possible variation for <u>one</u> real value of δa. Also under our conditions $i\delta a\mathcal{G}\tilde{\Psi}$ is a possible variation of <u>any</u> $\tilde{\Psi}$ within the set, whereas we need this to be true only for $\hat{\Psi}$.

As an application of these results let us consider UHF. If $\mathcal{G}$ is a one particle operator

$$\mathcal{G} = \sum_{s=1}^{N} g(s)$$

then $e^{i\alpha\mathcal{G}}$ takes the form $\prod_{s=1}^{N} u(s)$. Thus it follows from the discussion in Sec. XII-B that UHF satisfies the hypervirial theorem for any one-particle $\mathcal{G}$ (Hermitian or not, since the hypervirial theorem is linear in $\mathcal{G}$ and any one-particle operator $\mathcal{O}$ can be written as a linear combination of two one-particle Hermitian operators, for example $(\mathcal{O} + \mathcal{O}^{\dagger})$ and $i(\mathcal{O} - \mathcal{O}^{\dagger}$)). Correspondingly weaker statements hold for restricted HF schemes. Thus consider OHF. Then clearly we have the result that the optimal trial functions will satisfy (3) for any spin independent one electron operator $\mathcal{G}$.

In our discussion so far we have insisted that δa be real. However UHF is formally invariant to transformation by $e^{i\alpha\mathcal{G}}$ whether or not α is real so long as $\mathcal{G}$ is a one-particle operator.

Returning then to (2) and assuming δa to be pure imaginary and $\mathcal{M}$ Hermitian one then finds that the $\hat{\Phi}$ of UHF will also satisfy

$$(\hat{\Phi}, (H\mathcal{M} + \mathcal{M}H)\hat{\Phi}) = 2\hat{E}\,(\hat{\Phi}, \mathcal{M}\hat{\Phi}) \qquad \text{(XVII-5)}$$

or, combining (3) and (5)

$$(\hat{\Phi}, H\mathcal{M}\hat{\Phi}) = \hat{E}\,(\hat{\Phi}, \mathcal{M}\hat{\Phi}) = (\hat{\Phi}, \mathcal{M}H\hat{\Phi}) \qquad \text{(XVII-6)}$$

Similar results of course also apply to OHF with $\mathcal{M}$ any spin independent Hermitian one-particle operator. Since however, non unitary transformations are often not pleasant to deal with (they may transform a normalizable function into an unnormalizable one), the following derivation of (6) for UHF and OHF based directly on the variation method may be more convincing.

With $\mathcal{M}$ a Hermitian one-particle operator it follows from the consideration of Sec. XI that for UHF (and for OHF if $\mathcal{M}$ is also spin independent) that if we write

$$\mathcal{M}\hat{\Phi} = \frac{(\hat{\Phi}, \mathcal{M}\hat{\Phi})}{(\hat{\Phi}, \hat{\Phi})}\hat{\Phi} + \textcircled{H}$$

then $\textcircled{H}$ satisfies the conditions of Brillouin's theorem. Thus we have

$$\begin{aligned} 0 &= (\textcircled{H}, H\hat{\Phi}) = (\mathcal{M}\hat{\Phi}, H\hat{\Phi}) - (\hat{\Phi}, \mathcal{M}\hat{\Phi})(\hat{\Phi}, H\hat{\Phi})/(\hat{\Phi}, \hat{\Phi}) \\ &= (\hat{\Phi}, \mathcal{M}H\hat{\Phi}) - \hat{E}\,(\hat{\Phi}, \mathcal{M}\hat{\Phi}) \end{aligned}$$

which is the right hand side equality in (6). The left hand side equality then follows by complex conjugation.

Turning now to situations in which the set of trial functions forms a linear space let us return to (1) and write it as

$$(1+i\delta a\,\mathcal{S})\Psi - \Psi \qquad \text{(XVII-7)}$$

Then we see that if $\mathcal{S}$ applied to Ψ yields a function in the space, then (1), with no restriction on δa, will certainly be a possible variation of Ψ within the space since, by linearity, the first term in (7) is a neighboring function in the space. Therefore, as a sufficient condition we can say that if the set of trial functions forms a linear space then (6) and hence (3), will be satisfied if $\mathcal{S}$ applied to any function in the space yields a function in the space. Note however that this sufficient condition, though stated a bit differently, is formally equivalent to our general sufficient condition applied to this case: first of all it implies that $e^{i\alpha\mathcal{S}}$ with α an arbitrary complex number applied to any function in the space yields a function in the space, since if $\mathcal{S}\tilde{\Psi}$ is in the space so is $\mathcal{S}^2\tilde{\Psi}, \mathcal{S}^3\tilde{\Psi}$ etc. and therefore by linearity so is $e^{i\alpha\mathcal{S}}\tilde{\Psi}$. Conversely if $e^{i\alpha\mathcal{S}}\tilde{\Psi}$ is in the space then by linearity so is $\lim_{\alpha\to 0}(e^{i\alpha\mathcal{S}}\tilde{\Psi}-\tilde{\Psi})/i\alpha = \mathcal{S}\tilde{\Psi}$.

The following direct derivation of (11) for the linear case is also of interest: Since $\hat{\Psi}_k$ is an eigenfunction of $\overline{H}$ we have

$$(\hat{\Psi}_k, \overline{H}\mathcal{S}\hat{\Psi}_k) = \hat{E}_k(\hat{\Psi}_k, \mathcal{S}\hat{\Psi}_k) = (\hat{\Psi}_k, \mathcal{S}\overline{H}\hat{\Psi}_k)$$

But $\Pi\hat{\Psi}_k = \hat{\Psi}_k$ and if $\mathcal{S}\hat{\Psi}_k$ is in the space $\Pi\mathcal{S}\hat{\Psi}_k = \mathcal{S}\hat{\Psi}_k$ so that we can replace $\overline{H}$ by H which yields (11).

We have now seen two distinct consequences of the invariance of a set of trial functions to unitary transformations (i) invariance of the optimal energies, and (ii) hypervirial theorems. We now want to bring these two results together with the help of the generalized Hellmann-Feynman theorem. Introducing a real parameter ζ we define

$$H(\zeta) \equiv e^{i\zeta\mathscr{d}}\, H\, e^{-i\zeta\mathscr{d}}$$

with $\hat{\phi}(\zeta)$ and $\hat{E}(\zeta)$ the corresponding optimal trial functions and energies for $H(\zeta)$. Now suppose that the set of trial functions is invariant to transformation by $e^{i\zeta\mathscr{d}}$ for all ζ. Then from the results of Sec. XII-B it follows that the $\hat{E}$ are in fact independent of ζ. Further since changing ζ to $\zeta+\delta\zeta$ simply replaces one transformation by another we see that the set of trial functions is invariant to changes in ζ. Therefore from Sec. XII-D and the result just found, that $\hat{E}$ doesn't depend on ζ, we have that

$$0 = \left(\hat{\phi}(\zeta), \frac{\partial H(\zeta)}{\partial\zeta}\hat{\phi}(\zeta)\right)$$

However

$$\frac{\partial H}{\partial \zeta} = i\left(H(\zeta)\mathscr{d} - \mathscr{d}H(\zeta)\right)$$

Thus, putting $\zeta=0$ equal to zero, we have rederived the result that $\hat{\phi}$ satisfies the hypervirial theorem for $\mathscr{d}$.

We will now consider some specific hypervirial theorems of physical interest. For H we will always use the non-relativistic fixed nucleus electrostatic (i.e. no magnetic fields) Hamiltonian. Also we will work almost entirely in the coordinate representation.

XVIII. MOMENTUM THEOREMS

Let $\mathscr{d}$ be a component of

$$\vec{D} = \sum_{s=1}^{N} \vec{r}_s \qquad \text{(XVIII-1)}$$

Then one readily finds that $i(HD_k - D_kH) = \sum_s \frac{\partial H}{\partial p_{sk}}$ is the corresponding component of $\vec{P}$ the operator for total electronic momentum

$$\vec{P} = \sum_{s=1}^{N} \vec{p}_s \qquad \text{(XVIII-2)}$$

Thus if the hypervirial theorem for this $\mathcal{H}$ is satisfied then the average of the corresponding component of $\vec{P}$ calculated using $\hat{\Psi}$ will vanish. We will call this a momentum theorem.

Since the components of $\vec{D}$ are spinless one electron operators we then know that the $\hat{\Psi}$ of UHF and of OHF will satisfy all momentum theorems. One way to produce a set which is invariant to the transformations generated by $\vec{D}$ is to explicitly introduce real variational parameters $\vec{\tilde{K}}$ according to^{A}

$$\tilde{\Psi} = e^{i\vec{\tilde{K}}\cdot\vec{D}}\,\tilde{\Theta} \qquad \text{(XVIII-3)}$$

where the $\tilde{\Theta}$ are independent of K . That this works then follows from the observation that such a set of trial functions will be invariant to the unitary transformation $e^{i\vec{\xi}\cdot\vec{D}}$ since

$$e^{i\vec{\xi}\cdot\vec{D}}\,\tilde{\Psi} = e^{i(\vec{\tilde{K}}+\vec{\xi})\cdot\vec{D}}\,\tilde{\Theta} = e^{i\vec{\tilde{K}}'\cdot\vec{D}}$$

However having made these remarks it is very important to point out that often the momentum theorems will be satisfied simply for reasons of symmetry of one kind or another. For example since $\vec{P}$ is a pure imaginary Hermitian operator its average will automatically vanish if $\hat{\Psi}$ is real. Proof:

$$(\hat{\Psi}, \vec{P}\hat{\Psi}) = (\hat{\Psi}, \vec{P}\hat{\Psi})^{*} = (\hat{\Psi}^{*}, \vec{P}^{*}\hat{\Psi}^{*}) = -(\hat{\Psi}, \vec{P}\hat{\Psi})$$

And therefore $(\hat{\psi}, \vec{P}\,\hat{\psi}) = 0$. More generally if H is real the hypervirial theorem for any real Hermitian $\mathcal{H}$ will be satisfied if $\hat{\psi}$ is real since $i(H\mathcal{H} - \mathcal{H}H)$ is then a pure imaginary Hermitian operator.[1]

Another example is provided by an isolated atom. Then, reality aside, if, as is generally the case, $\hat{\psi}$ has a definite parity under inversion through the nucleus then $\hat{\psi}^{*}\hat{\psi}$ is invariant to inversion while $\vec{P}$ obviously changes sign. Therefore $(\hat{\psi}, \vec{P}\,\hat{\psi})$ will vanish. Or consider a diatomic molecule (or an atom in field which is invariant to rotation about an axis through the nucleus). If $\hat{\psi}$ has a definite component of angular momentum along the internuclear axis, then $\hat{\psi}^{*}\hat{\psi}$ will be invariant to a rotation of 180° about the internuclear axis while the components of $\vec{P}$ perpendicular to the axis will change sign. Therefore averages of these components will automatically vanish. Also they will vanish if the molecule, instead of having a definite component of angular momentum, has a definite parity for reflection in any plane containing the internuclear axis. The vanishing of the component along the internuclear axis is then guaranteed (by an adaptation of the earlier reality argument) if, as is often the case, $\hat{\psi}$ is complex only because it contains an angular factor on which the component of $\vec{P}$ along the axis does not act. Also for a homonuclear diatomic molecule one will usually arrange for the $\hat{\psi}$ to have a definite parity with respect to reflection in a plane perpendicular to the axis and through the mid-point and this will also ensure the variation of the average momentum along the internuclear axis.

XIX. FORCE THEOREMS

Let $\mathcal{P}$ be a component of the operator for total electronic momentum $\vec{P}$. Then one readily finds that $i(HP_k - P_kH) = -\sum_s \frac{\partial H}{\partial r_{sk}}$ is the corresponding component of $\vec{F}$, the operator for the total force on the electrons. $\vec{F}$ is of course the sum of the forces due to the nucleus and the forces due to whatever external fields may be present, the electron-electron forces cancelling. Thus for a molecule in the absence of external fields

$$\vec{F} = \sum_s \sum_A \mathcal{Z}_A \frac{(\vec{R}_A - \vec{r}_s)}{|\vec{R}_A - \vec{r}_s|^3} \qquad \text{(XIX-1)}$$

where $\mathcal{Z}_A$ is the charge on the A'th nucleus and $\vec{R}_A$ is its position vector. (Incidently the reader should keep in mind that this result and various others which we have been and are stating in the language of atoms and molecules and solids, are of course either quite general and do not depend on the detailed nature of the force laws or can be easily generalized). Thus if the hypervirial theorem for this $\mathcal{P}$ is satisfied then the average of the corresponding component of $\vec{F}$ calculated using $\hat{\Psi}$ will vanish. This we will call a force theorem.

Since the components of $\vec{P}$ are spinless one electron operators we then know that the $\hat{\Psi}$ of UHF and of OHF satisfy all force theorems. Further since a component of $\vec{P}$ generates a rigid translation of the electrons in the corresponding direction we know that in general if the set of trial functions is invariant to rigid translation in a particular direction, then the corresponding force theorem will be satisfied by the $\hat{\Psi}$

One way to produce a set of trial functions which is invariant to all translations is to explicitly introduce real variational parameters $\vec{\eta}$ attached in an additive way to each electron coordinate:

$$\tilde{\Psi}(\vec{r}_1 \cdots \vec{r}_N) = \tilde{\Theta}(\vec{r}_1 + \vec{\eta}, \vec{r}_2 + \vec{\eta} + \cdots \vec{r}_N + \vec{\eta}) \qquad \text{(XIX-2)}$$

$\tilde{\Theta}$ may of course involve other variation parameters and/or functions and may also depend on non variational parameters like nuclear coordinates, charges, etc. However here, and in analogous situations later, we will not indicate such dependencies explicitly unless they are relevant to the discussion (we have followed the same policy all along with respect to spin coordinates). That this works then follows from the observation that a rigid translation $\vec{r}_i \to \vec{r}_i + \vec{\alpha}$ is equivalent to $\vec{\eta} \to \vec{\eta} + \vec{\alpha}$ i.e. simply produces another member of the set. Indeed (2) formally has the same structure as (XVIII-3) since evidently we can write (2) as

$$\tilde{\Psi}(\vec{r}_1 \cdots \vec{r}_N) = e^{i\vec{\eta}\cdot\vec{D}}\,\tilde{\Theta}(\vec{r}_1 \cdots \vec{r}_N) \qquad \text{(XIX-3)}$$

Also (XVIII-3) transformed to momentum space takes the form

$$\tilde{\Phi} = \tilde{\chi}(\vec{r}_1 - \vec{\kappa}, \cdots \vec{r}_N - \vec{\kappa})$$

where $\tilde{\chi}$ is the Fourier transform of $\tilde{\Theta}$, i.e. $\vec{D}$ generates translations in momentum space.

However, just as with $\vec{P}$, this elaborate machinery may be unnecessary in that force theorems can often be satisfied simply by reason of symmetry. Thus consider again an isolated atom. Since $\vec{F}$ is odd under inversion through the nucleus, $(\hat{\Psi}, \vec{F}\hat{\Psi})$ will vanish if $\hat{\Psi}$ has a definite parity under inversion through the nucleus.

Also for a diatomic molecule (or an atom in an external field that is axially symmetric) force theorems perpendicular to the internuclear axis can be satisfied under the same conditions as the corresponding momentum theorems of the preceding section. Also this same symmetry will insure that the average net force on each nucleus separately will have only an axial component. However along the axis, symmetry is usually of no help except in the case of a homonuclear diatomic molecule. Hence to satisfy the force theorem along the axis usually requires the use of a set of trial functions which is explicitly invariant to translations along the axis.

As an interesting application of the force theorems, consider a molecule, first in an external electric field. The net force^A on all the nuclei is then the sum of the forces on the nuclei due to the electrons, and the forces on the nuclei due to the external field, the nucleus-nucleus forces cancelling. Thus

$$\begin{aligned}&\text{Net force on nuclei} = \text{Force on nuclei due to electrons} + \\ &\qquad \text{Force on nuclei due to external field}\end{aligned} \tag{XIX-4}$$

On the other hand the net force on the electrons is the sum of the forces on the electrons due to the nuclei plus the forces on the electrons due to the external field, and if the force theorems are satisfied these two contributions cancel. Thus

$$\begin{aligned}&0 \doteq \text{Force on electron due to nuclei} + \\ &\qquad \text{Force on electron due to external field}\end{aligned} \tag{XIX-5}$$

However

$$\begin{aligned}&\text{Force on electrons due to nuclei} = - \text{Force on nuclei} \\ &\qquad \text{due to electrons}\end{aligned} \tag{XIX-6}$$

Therefore combining (4), (5), and (6) we have the result

Net force on nuclei = Force on nuclei due to external field +
Force on electrons due to external field (XIX-7)

from which we can draw several interesting conclusions.

(i) Suppose that the molecule contains only one nucleus, i.e. is an atom. Then if we are dealing with a uniform external electric field $\vec{\mathcal{E}}$, the force on the nucleus due to the external field is simply $\mathcal{Z}\vec{\mathcal{E}}$ while the force on the electrons due to the field is $\sum_{s=1}^{N} -(\hat{\Psi}, \mathcal{E}\hat{\Psi})/(\hat{\Psi},\hat{\Psi}) = -N\vec{\mathcal{E}}$. Therefore if the force theorems are satisfied so that (18) applies we have the $\hat{\Psi}$ independent result

$$\text{Net force on nucleus} = (\mathcal{Z} - N)\vec{\mathcal{E}} = \mathcal{Z}\left(1 - \frac{N}{\mathcal{Z}}\right)\vec{\mathcal{E}} \qquad \text{(XIX-8)}$$

N/Z is called the dipole shielding factor and evidently N/Z is its exact value.[1]

(ii) If for an isolated atom the force theorems are satisfied then the force on the nucleus calculated from $\hat{\Psi}$, will vanish.

(iii) Returning to molecules, if there is no external field, then if the force theorems are satisfied so that (7) applies, we have that

Net force on nuclei = 0 (XIX-9)

In particular then, in a diatomic molecule the force on one nucleus calculated from $\hat{\Psi}$ will be equal and opposite to the force on the other.

XX. TORQUE THEOREMS

Let $\mathcal{M}$ be a component of $\vec{L}$ the operator for the total electronic orbital angular momentum

$$\vec{L} = \sum_{s=1}^{N} \vec{r}_s \times \vec{p}_s \qquad \text{(XX-1)}$$

Then one readily finds that $i(HL_k - L_k H)$ is the corresponding component of the operator for the net torque on the electrons, this net torque being provided by the nuclei and whatever external fields may be present, the electron-electron contributions cancelling. Thus if the hypervirial theorem for this $\mathscr{J}$ is satisfied then the average of the corresponding component of the net torque calculated using $\hat{\Psi}$ will vanish. This we will call a torque theorem.

Since the components of $\vec{L}$ are spinless one-electron operators we then know that the $\hat{\Psi}$ of UHF and of OHF satisfy all torque theorems. Further since a component of $\vec{L}$ generates a rigid rotation of the electrons about the corresponding axis we know that in general if the set of trial functions is invariant to rigid rotations about a particular axis then the corresponding torque theorem will be satisfied by the $\hat{\Psi}$.

The angular momentum and torque which we have been talking about are calculated about the origin of coordinate system. If more generally we calculate the angular momentum and torque about another point we will usually get a different answer. That is if we replace the $\vec{r}_s$ by $\vec{r}_s + \vec{\alpha}$ then the average angular momentum changes by

$$\vec{\alpha} \times (\hat{\Psi}, \vec{P}\, \Psi)/(\Psi, \Psi)$$

while the average torque changes by

$$\vec{\alpha} \times (\Psi, \vec{F}\, \Psi)/(\Psi, \Psi)$$

Therefore only if the momentum theorems are satisfied is the average angular momentum independent of origin, and only if the force theorems are satisfied are the average torques independent of origin.

We now note that for an isolated atom the operator for the torque about the nucleus vanishes identically ($\vec{L}$ about the nucleus is conserved) so that if the force theorems are satisfied all torque theorems about an arbitrary origin will be satisfied. Also for a diatomic molecule the torque theorem for torques about one of the nuclei will usually be satisfied simply by symmetry, and therefore if the force theorems are satisfied, the torque theorems about an arbitrary axis will be satisfied. Proof: About nucleus 1 the net torque on the electrons is due only to the other nucleus, the torque operator being proportional to

$$\vec{\mathcal{T}} = \sum_{s=1}^{N} (\vec{r}_s - \vec{R}_1) \times (\vec{r}_s - \vec{R}_2) / |\vec{r}_s - \vec{R}_2|^3$$

$$= \sum_{s=1}^{N} (\vec{r}_s - \vec{R}_1) \times (\vec{R}_1 - \vec{R}_2) / |\vec{r}_s - \vec{R}_2|^3 \equiv \vec{v} \times (\vec{R}_1 - \vec{R}_2)$$

Now the components of $\vec{v}$ which are perpendicular to $(\vec{R}_1 - \vec{R}_2)$ change sign if we rotate by 180° about the internuclear axis $\vec{R}_1 - \vec{R}_2$. Therefore if $\hat{\Psi}$ has a definite component of angular momentum along the internuclear axis (or a definite parity for reflection in any plane containing the internuclear axis) it follows that the average of these components of $\vec{v}$ will vanish. Therefore we may effectively replace $\vec{v}$ by its component along $\vec{R}_1 - \vec{R}_2$ whence $\vec{\mathcal{T}}$ is effectively zero, which completes the proof.

XXI. VIRIAL THEOREMS

Let $\mathcal{V}$ be the operator

$$\mathcal{V} \equiv \sum_{s=1}^{N} \vec{r}_s \cdot \vec{p}_s - \frac{3i}{2} N \qquad \text{(XXI-1)}$$

the addition of $-3i\frac{N}{2}$ ensuring that $\mathcal{V}$ is Hermitian. However it clearly plays no role in the hypervirial theorem for $\mathcal{V}$. Then one finds that $e^{ia\mathcal{V}}\hat{\Psi}(r_1 \cdots r_N)$ equals

$$\tau^{\frac{3N}{2}}\,\hat{\Psi}\,(\tau\vec{r}_1 \cdots \tau\vec{r}_N) \qquad \text{(XXI-2)}$$

where $\tau = e^{a}$ Proof: $\mathbb{X} \equiv e^{ia\mathcal{V}}\hat{\Psi}$ can also be defined by

$$\frac{1}{i}\frac{\partial}{\partial a}\mathbb{X} = \mathcal{V}\,\mathbb{X} \quad ; \; \mathbb{X} = \hat{\Psi} \text{ when } a = 0 \qquad \text{(XXI-3)}$$

It is then easy to verify that (2) satisfies (3).

Thus $\mathcal{V}$ produces a positive scaling of the electronic coordinates, the factor of $\tau^{3N/2}$ which arises from the $-i3N/2$ ensuring that the normalization of the scaled function is the same as that of $\hat{\Psi}$, as befits a unitary transformation. Since however this factor will cancel out in calculating energies we therefore have the result that if the set of trial functions is effectively invariant to such scaling then the $\hat{\Psi}$ will satisfy the hypervirial theorem for $\mathcal{V}$ A (we will discuss the content of this theorem in a moment). In particular since $\mathcal{V}$ is a spinless one-electron operator the $\hat{\Psi}$ of UHF and of OHF will satisfy the theorem. One common way to ensure that a set of trial functions will be invariant to positive scaling is to explicitly include a coordinate scaling parameter as a variational parameter. That is one uses trial functions of the form

$$\tilde{\Psi} = \tilde{\zeta}^{3N/2}\,\tilde{\Theta}\,(\tilde{\zeta}\vec{r}_1 \cdots \tilde{\zeta}\vec{r}_N) \qquad \text{(XXI-4)}$$

with $\tilde{\zeta}$ a real variational parameter, the factor of $\tilde{\zeta}^{3N/2}$ being optional. That such a set of invariant to positive scaling then follows from the observation that replacing the $\vec{r}_s$ by $\tau\vec{r}_s$ in $\tilde{\Psi}$ is equivalent to leaving the $\vec{r}_s$ alone and replacing $\tilde{\zeta}$ by $\tau\tilde{\zeta}$ and therefore effectively produces another member of the set. Moreover such a set is obviously invariant not only to positive scaling but also negative scaling $(\tau<0)$ and hence in particular to inversion $\vec{r}_s \to -\vec{r}_s$. Evidently UHF and OHF also have this property. In general when a set is invariant to both positive and negative scaling we will say simply that it is invariant to scaling.

The hypervirial theorem for $\mathcal{V}$ is essentially the virial theorem which is so often used in discussions of chemical binding, force constants, etc. To show this we first note that we can calculate $i(H\mathcal{V}-\mathcal{V}H)$ quite generally as

$$i(H\mathcal{V}-\mathcal{V}H) = \sum_{s=1}^{N}\left(\frac{\partial H}{\partial \vec{p}_s}\cdot\vec{p}_s - \vec{r}_s\cdot\frac{\partial H}{\partial \vec{r}_s}\right) \qquad \text{(XXI-5)}$$

Now let us specialize to an isolated molecule. Then

$$H = T + V \qquad \text{(XXI-6)}$$

where T , the kinetic energy operator, is a homogeneous function of degree 2 in the $\vec{p}_s$, while V , the potential energy operator, which we will take to include the nuclear repulsions so the $\hat{E}$ is the total molecular energy, is a homogeneous function of degree (-1) in the $\vec{r}_s$ and the $\vec{R}_A$. Thus from Euler's theorem on homogeneous functions we find

$$\sum_{s=1}^{N}\frac{\partial H}{\partial \vec{p}_s}\cdot\vec{p}_s = 2T \qquad \text{(XXI-7)}$$

and

$$\sum_{s=1}^{N}\vec{r}_s\cdot\frac{\partial H}{\partial \vec{r}_s} = -V - \sum_{A}\vec{R}_A\cdot\frac{\partial H}{\partial \vec{R}_A} \qquad \text{(XXI-8)}$$

Therefore the hypervirial theorem for $\mathcal{V}$ becomes

$$2\hat{T} + \hat{V} - \sum_A \vec{R}_A \cdot \hat{\vec{F}}_A = 0 \qquad \text{(XXI-9)}$$

where $\hat{T}$ is the average kinetic energy of the electrons, $\hat{V}$ the average potential energy of the electrons and nuclei, and $\hat{F}_A$ is the average force on nucleus A :

$$\hat{T} = (\Psi, T\Psi)/(\Psi,\Psi) \quad ; \hat{V} = (\Psi, V\Psi)/(\Psi,\Psi) \; ; \hat{\vec{F}}_A = -(\Psi, \frac{\partial H}{\partial \vec{R}_A}\Psi)/(\Psi,\Psi) \qquad \text{(XXI-10)}$$

And where

$$-\frac{\partial H}{\partial \vec{R}_A} \equiv \vec{F}_A = \sum_{s=1}^{N} {}^{e}\mathcal{Z}_A \frac{(\vec{r}_1 - \vec{R}_A)}{|\vec{r}_s - \vec{R}_A|^3} + \sum_{B \neq A} {}^{e}\mathcal{Z}_A {}^{e}\mathcal{Z}_B \frac{(\vec{R}_A - \vec{R}_B)}{|\vec{R}_A - \vec{R}_B|^3} \qquad \text{(XXI-11)}$$

We will call (9) the generalized virial theorem. To reduce it to more familiar form suppose that $\hat{\Psi}$, in addition to satisfying the hypervirial theorem for $\mathcal{V}$, also satisfies the generalized Hellmann-Feynman theorems in $\vec{r}_s$ coordinates for σ equal to the components of the $\vec{R}_A$. That is suppose that

$$\frac{\partial \mathcal{E}}{\partial \vec{R}_A} = (\Psi, \frac{\partial H}{\partial \vec{R}_A}\Psi)/(\Psi,\Psi) \qquad \text{(XXI-12)}$$

Then (9) yields

$$2\hat{T} + \hat{V} + \sum_A \vec{R}_A \cdot \frac{\partial \mathcal{E}}{\partial \vec{R}_A} = 0 \qquad \text{(XXI-13)}$$

Finally let us suppose that $\hat{\Psi}$ was derived from a set of trial functions which is invariant to translations and rotations of the electrons. Then since a translation or rotation of the electrons is equivalent, as far as H is concerned, to leaving the electrons alone and translating and rotating the nuclei the other way, and since our set is already assumed to be independent of the $\vec{R}_A$ in order to satisfy the generalized Hellmann-Feynman theorems for $\sigma = \vec{R}_A$

it follows from Sec. XIV that $\hat{E}$ can depend only on translationally and rotationally invariant quantities like bond lengths $R_{(ab)} \equiv |\vec{R}_A - \vec{R}_B|$ and bond angles $\cos^{-1} \frac{(\vec{R}_A - \vec{R}_B) \cdot (\vec{R}_C - \vec{R}_D)}{|\vec{R}_A - \vec{R}_B| \, |\vec{R}_C - \vec{R}_D|}$. Since the latter are homogeneous functions of degree zero in the $\vec{R}_A$ and since $\frac{\partial}{\partial \vec{R}_A} R_{(ab)} = -\frac{\partial}{\partial \vec{R}_B} R_{(ab)}$ one then readily finds that under these conditions (13) can be written as[11, B]

$$2\hat{T} + \hat{V} + \sum_i R_{(i)} \frac{\partial \hat{E}}{\partial R_{(i)}} \qquad \text{(XXI-14)}$$

For a diatomic molecule this becomes the familiar

$$2\hat{T} + \hat{V} + R \frac{d\hat{E}}{dR} = 0 \qquad \text{(XXI-15)}$$

where R is the internuclear separation, and for an isolated atom (R = 0) is the equally familar

$$2\hat{T} + \hat{V} = 0 \qquad \text{(XXI-16)}$$

Eq. (14), or its specializations (15) and (16), is what is usually called the virial theorem.[C]

Thus if one uses a set of trial functions which is invariant to positive scaling of the electronic coordinates, is invariant to changes in the nuclear coordinates, and is invariant to translations and rotations of the $\vec{r}_s$, then the $\hat{\Psi}$ will satisfy (9), (13), and (14). In addition we know from previous sections that these conditions will also guarantee that the net force on the electrons and the net torque on the electrons calculated from $\hat{\Psi}$ will both vanish, and that therefore the sum of the forces on all the nuclei as calculated from $\hat{\Psi}$ will vanish. Also the forces on the nuclei in a diatomic molecule will be only along the axis since if $\hat{E}$ is a function only of $R = |\vec{R}_2 - \vec{R}_1| \equiv |\vec{R}|$ then

$$-\frac{\partial \hat{E}}{\partial \vec{R}_1} = +\frac{\partial \hat{E}}{\partial \vec{R}_2} = \frac{\vec{R}}{|\vec{R}|}\frac{d\hat{E}}{dR} \qquad \text{(XXI-17)}$$

This then will also mean that the average net torque on the nuclei will vanish.[D]

As we have seen UHF and OHF satisfy all these conditions. For other sorts of trial functions one often explicitly introduces variation parameters to do the job. Since there are various ways of doing this it will help to avoid notational confusion to consider a simple example -- a one-electron diatomic molecule in a simple LCAO type approximation involving two 1s atomic orbitals. Generalizations should be obvious. A first choice (Heitler - London) for the set of $\tilde{\Psi}$ might be the functions

$$e^{-Z_1|\vec{r}-\vec{R}_1|} + \tilde{c}\, e^{-Z_2|\vec{r}-\vec{R}_2|} \qquad \text{(XXI-18)}$$

where $\tilde{c}$ is a variational parameter. However this set has none of the properties we want. It is not invariant to scaling, it depends explicitly on $\vec{R}_1$ and $\vec{R}_2$ and is not invariant to either rotation or translation of the electron's coordinates. To take care of all these deficiencies, but still keep the same sort of $\tilde{\Psi}$, it is then quite natural to introduce two real positive variational parameters $\tilde{\zeta}$ and $\tilde{\rho}$ and two real vector variational parameters $\tilde{\vec{\gamma}}$ and $\tilde{\vec{\mu}}$ and use the set of trial functions

$$e^{-|\tilde{\zeta}\vec{r}-\tilde{\vec{\gamma}}|} + \tilde{c}\, e^{-\tilde{\rho}|\tilde{\zeta}\vec{r}-\tilde{\vec{\mu}}|} \qquad \text{(XXI-19)}$$

This set is then obviously invariant to changes in $\vec{R}_1$ and $\vec{R}_2$ since each member of the set is separately invariant. Also it is obviously invariant to scaling. Further the set is invariant to translations since replacing $\vec{r}$ by $\vec{r}+\vec{\alpha}$ is equivalent to replacing $\tilde{\vec{\gamma}}$ and $\tilde{\vec{\mu}}$ by $\tilde{\vec{\gamma}}-\tilde{\xi}\vec{\alpha}$ and $\tilde{\vec{\mu}}-\tilde{\xi}\vec{\alpha}$. Finally the set is invariant to rotation since replacing $\vec{r}$ by $\mathcal{Q}\vec{r}$ where $\mathcal{Q}$ is a rotation dyadic is equivalent to replacing $\tilde{\vec{\gamma}}$ and $\tilde{\vec{\mu}}$ by $\mathcal{Q}^{-1}\tilde{\vec{\gamma}}$ and $\mathcal{Q}^{-1}\tilde{\vec{\mu}}$. Proof: $|\mathcal{Q}\tilde{\xi}\vec{r}-\tilde{\vec{\gamma}}|=|\mathcal{Q}(\tilde{\xi}\vec{r}-\mathcal{Q}^{-1}\tilde{\vec{\gamma}}|=$ since the length of a vector, in this case, is invariant to rotation.[E] (As we said, we wrote down (19) by analogy with (18). However it should be noted that either term in the sum (19) would yield a set with the same invariance properties.)

Thus far our discussion has been quite general, with a view toward application to a general polyatomic molecule. However for atoms and diatomic molecules for example, one has special symmetries which are usually taken advantage of in any variational calculation. First let us consider an atom. Then (14) becomes

$$2\hat{T}+\hat{V}-\vec{R}_1\cdot\hat{\vec{F}}_1=0 \tag{XXI-20}$$

which is equivalent to the virial theorem if the force theorems are satisfied, and we have seen that this will usually be the case for reasons of symmetry. However we can also derive the virial theorem from (20) simply by using the nucleus as the origin of coordinates so that $\vec{R}_1=0$; and this one would almost certainly do in any practical calculation. Thus if one chooses the nucleus as the coordinate origin then the virial theorem will be guaranteed simply by having the set of trial functions be invariant to positive scaling of electronic coordinates. Further, with this origin of coordinates we see that most restricted Hartree-Fock methods for atoms will also satisfy the virial theorem since positive scaling does not affect angles, i.e. only radial coordinates need be scaled.

Also for atoms, and with the origin of coordinates at the nucleus, the following alternative derivation of the virial theorem as a consequence of invariance to scaling is of interest.[2] Suppose that we calculate $\tilde{E}(\tilde{\zeta})$ using the trial function (4). Then by changing variables in the integrals from the $\vec{r}_s$ to the $\hat{\zeta}\vec{r}_s$ and using the homogeneity properties of T and V it is easy to show that, in obvious notation

$$\tilde{E}(\hat{\zeta}) = \tilde{\zeta}^2 \tilde{T}(1) + \tilde{\zeta}\,\tilde{V}(1) \qquad \text{(XXI-21)}$$

where to be definite we have assumed that $\tilde{\zeta}$ is positive. Requiring that $\partial\hat{E}(\hat{\zeta})/\partial\hat{\zeta} = 0$ then yields

$$2\,\hat{\zeta}\,\tilde{T}(1) + \tilde{V}(1) = 0$$

or, multiplying through by $\hat{\zeta}$

$$2\hat{T} + \hat{V} = 0$$

which is the virial theorem again.

Turning now to diatomic molecules, here one would almost certainly put the origin of coordinates on the internuclear axis and use the internuclear axis as one of the coordinate axes, say the x' axis. Under these circumstances V will be a homogeneous function of degree -1 in the $\vec{r}_s\,'$ and the X_A' so that if we use a set of $\tilde{\Psi}$ which is invariant to positive scaling of the $\vec{r}_s\,'$ we will find instead of (9), that $\hat{\Psi}$ satisfies

$$2\hat{T} + \hat{V} + \sum_{A=1}^{2} X_A' \left(\hat{\Psi}, \frac{\partial H}{\partial X_A'}\hat{\Psi}\right) \Big/ (\hat{\Psi},\hat{\Psi}) \qquad \text{(XXI-22)}$$

If further the $\hat{\Psi}$ satisfy the generalized Hellmann-Feynman theorem for $\sigma = X_A'$ in the $\vec{r}_s{}'$ coordinates:

$$\frac{\partial \hat{E}}{\partial X_A'} = (\Psi, \frac{\partial H}{\partial X_A'} \Psi)/(\Psi,\Psi) \tag{XXI-23}$$

then (22) yields

$$2\hat{T} + \hat{V} + \sum_{A=1}^{2} X_A' \frac{\partial \hat{E}}{\partial X_A'} = 0 \tag{XXI-24}$$

Eq. (24) will then be equivalent to the virial theorem if $\hat{E}$ depends only on the internuclear separation $R = |X_1' - X_2'|$. Since translation of the electrons along the x' axis and inversion of the electrons (strictly we need consider only $x_s' \to -x_s'$) is, as far as H is concerned, equivalent to leaving the electrons alone and translating and inverting the nuclei it follows from Sec. XIV that if the set of trial functions, in addition to being invariant to changes in the X_A', is also invariant to such translations and inversions of the electrons, then $\hat{E}$ will depend only on the translation and inversion invariant quantity R .

Thus if the set of $\tilde{\Psi}$ is invariant to scaling of the $\vec{r}_s{}'$, is invariant to changes in the X_A', and is invariant to translation of the $\vec{r}_s{}'$ along the internuclear axis then the $\hat{\Psi}$ will satisfy (22), (23), and (24) and (15). Also the force theorem will be satisfied along the internuclear axis. UHF and OHF of course, have all these properties. Returning to our one electron diatomic molecule, a set of trial functions with these properties would be

$$e^{-\sqrt{(\tilde{\zeta}x' - \tilde{\gamma})^2 + (\tilde{\zeta}y')^2 + (\tilde{\zeta}z')^2}} + \tilde{c}\, e^{-\tilde{\delta}\sqrt{(\tilde{\zeta}x' - \tilde{\mu})^2 + (\tilde{\zeta}y')^2 + (\tilde{\zeta}z')^2}} \tag{XXI-25}$$

and again, as with (19), either term in the sum yields a set with the same invariance properties.

If (22) is satisfied one may ask whether or not the more general (9) is also satisfied. The Cartesian coordinates in (9) are arbitrary, differing therefore from the x'y'z' coordinates by a rotation about the origin plus some translation. Now $\partial H/\partial x'_A$ is the x' component of

$$\vec{F}'_A = \sum_s Z_A \frac{(\vec{r}'_s - \vec{R}'_A)}{|\vec{r}'_s - \vec{R}'_A|^3} + \sum_{B \neq A} Z_A Z_B \frac{(\vec{R}'_A - \vec{R}'_B)}{|\vec{R}'_A - \vec{R}'_B|^3}$$

which, if we were to delete the ' would be precisely $\vec{F}_A$. Therefore since $y'_A = z'_A = 0$ it readily follows that if we write $\vec{r}'_s = \mathcal{Q}(\vec{r}_s + \vec{d})$ and $\vec{R}'_A = \mathcal{Q}(\vec{R}_A + \vec{d})$ where $\mathcal{Q}$ is a rotation dyadic and $\vec{d}$ a translation then

$$-x'_A \frac{\partial H}{\partial x'_A} = \vec{R}'_A \cdot \vec{F}'_A = \vec{R}_A \cdot \vec{F}_A + \vec{d} \cdot \vec{F}_A$$

Therefore we can write (22) referred to general coordinates (but with the <u>same</u> $\hat{\psi}$) as

$$2\hat{T} + \hat{V} - \sum_{A=1}^{2} \vec{R}_A \cdot \hat{\vec{F}}_A - \vec{d} \cdot \sum_A \hat{\vec{F}}_A = 0 \qquad \text{(XXI-26)}$$

where $\vec{d}$ is the vector connecting the origin of the x'y'z' coordinates to the origin of the x y z coordinates. Thus if all force theorems are satisfied so that $\sum_A \hat{\vec{F}}_A = 0$, (22) will imply (9). As we have discussed before, the force theorems perpendicular to the internuclear axis will usually be satisfied by symmetry, and for example such is the case with the functions (25) since they are individually invariant to rotation about the internuclear axis, and therefore have zero angular momentum about the internuclear axis. Therefore if $\vec{d}$ is perpendicular to the internuclear axis (22) and (9) can be made equivalent simply by symmetry. If $\vec{d}$ has a component along the internuclear axis then one also needs to explicitly satisfy the force theorem along the axis in order to have equivalence.

Returning to the x' y' z' coordinates, often they are chosen in such a way that whatever the nuclear separation, the coordinate origin is a fixed fractional distance along the internuclear axis, thus at a point

$$\vec{\mathcal{D}} = \frac{\alpha \vec{R}_1' + \beta \vec{R}_2'}{\alpha + \beta} \tag{XXI-26}$$

Introducing coordinates referred to this origin

$$\vec{r}_s'' = \vec{r}_s' - \vec{\mathcal{D}}$$

we find

$$\vec{r}_s' - \vec{r}_t' = \vec{r}_s'' - \vec{r}_t'' \quad , \quad \vec{r}_s' - \vec{R}_1' = \vec{r}_s'' + \frac{\beta}{\alpha+\beta}\vec{R} \quad , \quad \vec{r}_s' - \vec{R}_2' = \vec{r}_s'' - \frac{\alpha}{\alpha+\beta}\vec{R} \tag{XXI-27}$$

where $\vec{R}$ is the vector separation of the nuclei

$$\vec{R} = \vec{R}_2' - \vec{R}_1' \tag{XXI-28}$$

Thus since $\vec{R}$ has only one component, call it $\mathcal{R}$, in these coordinates we see that V is a homogeneous function of degree -1 in the $\vec{r}_s''$ and $\mathcal{R}$ and therefore if the set of $\tilde{\Psi}$ is invariant to positive scaling of the $\vec{r}_s''$ we will have

$$2\hat{T} + \hat{V} + \mathcal{R}\left(\Psi, \frac{\partial H}{\partial \mathcal{R}} \Psi\right) / (\Psi, \Psi) \tag{XXI-29}$$

where

$$-(\Psi, \frac{\partial H}{\partial \mathcal{R}} \Psi)/(\Psi,\Psi) = -\frac{\beta}{\alpha+\beta} \text{ (average axial component of force on nucleus 1)}$$
$$= +\frac{\alpha}{\alpha+\beta} \text{ (average axial component of force on nucleus 2)} \qquad \text{(XXI-30)}$$

If further the generalized Hellmann-Feynman theorem for $\sigma = \mathcal{R}$ in $\vec{r}_s{}''$ coordinates is satisfied

$$\frac{\partial \hat{E}}{\partial \mathcal{R}} = (\Psi, \frac{\partial H}{\partial \mathcal{R}} \Psi)/(\Psi,\Psi) \qquad \text{(XXI-31)}$$

then we will have

$$2\hat{T} + \hat{V} + \mathcal{R}\frac{d\hat{E}}{d\mathcal{R}} = 0 \qquad \text{(XXI-32)}$$

This equation will then be equivalent to the virial theorem if $\hat{E}$ depends only on $R = |\mathcal{R}|$. Since as far as H is concerned $\vec{r}_s{}'' \to -\vec{r}_s{}''$ is equivalent to $\mathcal{R} \to -\mathcal{R}$ (strictly all we need is $x_s{}'' \to -x_s{}''$) it follows from the discussion in Sec. XIV that if the set of trial functions (which we have assumed to be independent of $\mathcal{R}$) is also invariant to negative scaling $\hat{E}$ will depend only on R and so we will have the virial theorem. However these devices alone; invariance to scaling of $\vec{r}_s{}''$ and invariance in $\vec{r}_s{}''$ coordinates to changes in $\mathcal{R}$, will not in general ensure that either (22), (23) or the axial force theorem are satisfied. Of course if the axial force theorem is satisfied then, from (42) and (43) and assuming that $\mathcal{R} > 0$, we will have

$-dE/dR$ = average axial component of force on nucleus 2
= − average axial component of force on nucleus 1.

If $\mathcal{R}$ is < 0 the signs on the right are reversed.

Along these same lines it is of interest to note that Löwdin[F] has given a simple prescription which leads directly to the virial theorem. The prescription is to use trial functions of the form

$$\tilde{\Psi}(\tilde{\zeta}\vec{r}_1'', \tilde{\zeta}\vec{r}_2'', \cdots \tilde{\zeta}\vec{r}_N'', \tilde{\zeta}\mathcal{R}) \qquad \text{(XXI-33)}$$

and the proof is most easily given following the pattern centering around Eq. (21). Thus introducing the $\tilde{\zeta}\vec{r}_\ell''$ as integration variables, one readily finds that in obvious notation

$$\tilde{E} = \tilde{\zeta}^2\,\tilde{T}(1, \tilde{\zeta}\mathcal{R}) + \tilde{\zeta}\,\tilde{V}(1, \tilde{\zeta}\mathcal{R}) .$$

Then

$$\partial\hat{E}/\partial\hat{\zeta} = 0$$

yields

$$2\hat{\zeta}\,\hat{T}(1, \hat{\zeta}\mathcal{R}) + \hat{V}(1, \hat{\zeta}\mathcal{R}) + \hat{\zeta}^2\frac{\partial\hat{T}(1, \hat{\zeta}\mathcal{R})}{\partial\hat{\zeta}} + \hat{\zeta}\frac{\partial\hat{V}(1, \hat{\zeta}\mathcal{R})}{\partial\hat{\zeta}} = 0$$

or

$$2\hat{\zeta}\,\hat{T}(1, \hat{\zeta}\mathcal{R}) + \hat{V}(1, \hat{\zeta}\mathcal{R}) + \mathcal{R}\hat{\zeta}\frac{\partial\hat{T}(1, \hat{\zeta}\mathcal{R})}{\partial\mathcal{R}} + \mathcal{R}\frac{\partial\hat{V}(1, \hat{\zeta}\mathcal{R})}{\partial\mathcal{R}} = 0$$

Multiplying through by $\hat{\zeta}$ then yields (32) :

$$2\hat{T}(\hat{\zeta}, \hat{\zeta}\mathcal{R}) + \hat{V}(\hat{\zeta}, \hat{\zeta}\mathcal{R}) + \mathcal{R}\frac{d\hat{E}}{\partial\mathcal{R}} = 0$$

However since H is invariant to a simultaneous change in sign of the $\vec{r}_\ell''$ and of $\mathcal{R}$ and since our set of trial functions has the same property (the transformation is equivalent to $\tilde{\zeta} \to -\tilde{\zeta}$) it follows from Sec. XIV that the $\hat{E}$ do not depend on the sign of $\mathcal{R}$, and so we have the virial theorem. However unless the individual $\tilde{\Psi}$ don't depend on $\mathcal{R}$ at all, or involve $\mathcal{R}$ only in conjunction with another variational parameter (see below) such a set is in general not invariant to positive scaling and is not invariant to changes in $\mathcal{R}$ since neither separately is equivalent to a change in $\tilde{\zeta}$

Therefore although $\hat{\Psi}$ will satisfy the virial theorem it will not, in general, satisfy either (29) or (31). Also it will not in general satisfy the axial force theorem without further variational flexibility. On the other hand any set of the form $\tilde{\Psi}(\tilde{\zeta}\vec{r}_1^{\,n}, \cdots \tilde{\nu}\tilde{\zeta}R)$, since it is separately invariant to scaling of the $\vec{r}_s^{\,n}$ (equivalent to $\tilde{\zeta} \to \tau\tilde{\zeta}$, $\tilde{\nu} \to \tau^{-1}\tilde{\nu}$) and changes in R (equivalent to $\tilde{\nu} \to \tilde{\nu}(1+\frac{\delta R}{R})$), will yield $\hat{\Psi}$ which will satisfy (41) and (43) and (29). If further the variables occur in the combination $\sqrt{(\tilde{\zeta}(x_s^n-\tilde{\nu}R))^2+(\tilde{\zeta}y_s^n)^2+(\tilde{\zeta}z_s^n)^2}$ then one will also have invariance to translation along the internuclear axis ($x_s^n \to x_s^n + a$ equivalent to $\tilde{\nu} \to \tilde{\nu} - \frac{R}{a}$) and therefore the $\hat{\Psi}$ will also satisfy the axial force theorem. Indeed the two terms in (25) are each of this type, though in a slightly different notation, i.e. with $\tilde{\gamma}$ and $\tilde{\mu}$ instead of $\tilde{\zeta}\,\tilde{\nu}\,R$

An interesting specialization of (33) is provided by trial functions of the form

$$\tilde{\Psi}\left(\frac{\vec{r}_1^{\,n}}{R}, \frac{\vec{r}_2^{\,n}}{R} \cdots, \tilde{\zeta}R\right) \qquad \text{(XXI-34)}$$

Such a set then has the additional property that for fixed values of the

$$\vec{\rho}_s \equiv \vec{r}_s^{\,n}/R$$

it is independent of R . Therefore such a set will yield $\hat{\Psi}$'s which satisfy the generalized Hellmann-Feynman theorem for $\sigma = R$ in $\vec{\rho}_s$ coordinates, and hence in any coordinates derived from the by an R -independent transformation, for example the often used orthogonal confocal elliptic coordinates (with the nuclei at the foci). We will now show[3] that the generalized Hellmann-Feynman theorem for

$\sigma = \mathcal{R}$ in such coordinates is precisely the virial theorem. Proof: In $\vec{\rho}_3$ coordinates H takes the form

$$H = \frac{t}{\mathcal{R}^2} + \frac{v}{\mathcal{R}}$$

where we have assumed that $\mathcal{R}$ is positive and where t and v are independent of $\mathcal{R}$. Therefore the generalized Hellmann-Feynman theorem in these coordinates yields (the factor of $\mathcal{R}^3$ in the volume element cancels out)

$$\frac{\partial \hat{E}}{\partial \mathcal{R}} = -\frac{2}{\mathcal{R}^3}\frac{(\hat{\psi}, t\hat{\psi})}{(\hat{\psi}, \hat{\psi})} - \frac{1}{\mathcal{R}^2}\frac{(\hat{\psi}, v\hat{\psi})}{(\hat{\psi}, \hat{\psi})}$$

Multiplying through by $\mathcal{R}$ we therefore have (32) again.

$$2\hat{T} + \hat{V} + \mathcal{R}\frac{\partial \hat{E}}{\partial \mathcal{R}} = 0$$

However in general $H = \frac{t}{R^2} + \frac{v}{R}$ and is therefore independent of the sign of $\mathcal{R}$. Therefore since for fixed $\vec{\rho}_3$ the set of $\tilde{\psi}$ is independent of the sign of $\mathcal{R}$ ($\mathcal{R} \to -\mathcal{R}$ equivalent to $\tilde{\zeta} \to -\tilde{\zeta}$) it follows Sec. XIV that $\hat{E}$ will depend only on R and hence we have the virial theorem. Thus to satisfy the virial theorem it is sufficient to use a set of trial functions which in $\vec{\rho}_3$ type coordinates is independent of $\mathcal{R}$.

As another exercise in scaling we will state the following without detailed proof: Consider a general molecule in general coordinates, and following Hurley,^{4} let us refer the nuclear configuration to a similar configuration according to

$\vec{R}_A = s\vec{R}_A^{\,0}$, same S for all A, S > 0 (XXI-35)

Using the pattern of Eq. (33) et seq. it is then easy to show that if we use trial functions of the form

$$\hat{\Psi} = \Psi(s\vec{r}_1, s\vec{r}_2, \ldots, s\vec{R}_1, s\vec{R}_2, \ldots \qquad \text{(XXI-36)}$$

the $\hat{\Psi}$ will satisfy Hurley's[4] form of the virial theorem

$$2\hat{T} + \hat{V} + s\frac{\partial \hat{E}}{\partial s} = 0 \qquad \text{(XXI-37)}$$

XXII. ORTHOGONALITY AND RELATED THEOREMS

Eigenfunctions belonging to different eigenvalues of H are automatically orthogonal. As we will see in a moment, it is easy enough to give sufficient conditions such that $\hat{\Psi}_a$ and $\hat{\Psi}_b$ will be automatically orthogonal if $\hat{E}_a \neq \hat{E}_b$. However it must be stated that so far, the only known way of realizing these conditions à priori is to draw $\hat{\Psi}_a$ and $\hat{\Psi}_b$ from a common linear space, and for this case we have already discussed the orthogonality properties of the $\hat{\Psi}_K$ in Sec. XI. Also there are various other theorems of a similar type for which the same situation prevails. To put the matter another way round, if the $\tilde{\Psi}$ involve non linear parameters and/or functions as in UHF and OHF and CI with non linear parameters, then we do not expect the theorems to be satisfied, and this is in agreement with experience.[A]

To cover all the theorems at once let $\hat{\Psi}_a$ be an optimal trial function for a Hamiltonian H_a, and let $\hat{\Psi}_b$ be an optimal trial function for a Hamiltonian H_b. Then suppose that among the variations of $\hat{\Psi}_a$ which were possible in the set from which it was drawn, was $\delta a \mathcal{J} \hat{\Psi}_b$ where δa is a small but otherwise arbitrary complex number and $\mathcal{J}$ is a Hermitian operator. Then from (V-3) applied to $H = H_a$ we find, with obvious changes in notation

$$(\hat{\Phi}_b, \mathcal{G} H_a \hat{\Phi}_a) = \hat{E}_a (\hat{\Phi}_b, \mathcal{G} \hat{\Phi}_a) \tag{XXII-1}$$

Similarly if $\delta_a \mathcal{G} \hat{\Phi}_a$ was a possible variation of $\hat{\Phi}_b$ then from the complex conjugate of (V-3) applied to $H = H_b$ we find

$$(\hat{\Phi}_b, H_b \mathcal{G} \hat{\Phi}_a) = \hat{E}_b (\hat{\Phi}_b, \mathcal{G} \hat{\Phi}_a) \tag{XXII-2}$$

Subtracting these then we have

$$(\hat{\Phi}_b, (\mathcal{G} H_a - H_b \mathcal{G}) \hat{\Phi}_a) = (\hat{E}_a - \hat{E}_b)(\hat{\Phi}_b, \mathcal{G} \hat{\Phi}_a) \tag{XXII-3}$$

Various special cases are now of interest:

(i) $\mathcal{G} = 1$; $H_a = H_b = H$. Then Eq. (3) tells us that if $\hat{E}_a \neq \hat{E}_b$, then $\hat{\Phi}_a$ will be orthogonal to $\hat{\Phi}_b$. However as we noted at the outset of the section, the one way we know to implement the sufficient condition in an à priori manner is to draw $\hat{\Phi}_a$ and $\hat{\Phi}_b$ from a common linear space.

(ii) $\mathcal{G} \neq 1$; $H_a = H_b = H$. For $\hat{E}_a \neq \hat{E}_b$ Eq. (3) is the variational version of the so-called off diagonal hypervirial theorem for $\mathcal{G}$. The one way we know to implement the sufficient condition in an à priori manner is to draw $\hat{\Phi}_a$ and $\hat{\Phi}_b$ from a common linear space which is invariant to the action of $\mathcal{G}$.[1] However in such a case we can derive the theorem more directly as follows:

$$(\hat{\Phi}_b, (\mathcal{G} \bar{H} - \bar{H} \mathcal{G}) \hat{\Phi}_a) = (\hat{E}_a - \hat{E}_b)(\hat{\Phi}_b, \mathcal{G} \hat{\Phi}_a)$$

But $\Pi \hat{\Phi}_a = \hat{\Phi}_a$, $\Pi \hat{\Phi}_b = \hat{\Phi}_b$, $\Pi \mathcal{G} \hat{\Phi}_a = \mathcal{G} \hat{\Phi}_a$ and $\Pi \mathcal{G} \hat{\Phi}_b = \mathcal{G} \hat{\Phi}_b$. Therefore we can replace $\bar{H}$ by H and we have the theorem.

(iii) $\mathcal{A} = 1$, $H_a \neq H_b$. Eq. (3) is now the variational version of the integral Hellmann-Feynman theorem.[2] The one way known to implement the sufficient condition in an a priori manner is to draw $\hat{\psi}_a$ and $\hat{\psi}_b$ from a common linear space.[3] However in such a case we can derive the theorem directly as follows:

$$(\hat{\psi}_b, (\overline{H}_a - \overline{H}_b)\hat{\psi}_a) = (\hat{E}_a - \hat{E}_b)(\hat{\psi}_b, \psi_a)$$

But $\Pi \hat{\psi}_a = \psi_a$ and $\Pi \hat{\psi}_b = \psi_b$ (note that under our assumption the same Π is involved in $\overline{H}_a$ and $\overline{H}_b$) so that we may replace $\overline{H}_a$ and $\hat{H}_b$ by H_a and H_b thereby deriving the desired theorem.

(iv) $\mathcal{A} \neq 1$; $H_a \neq H_b$. We leave it to the reader to name and discuss this case.

APPENDIX A: THE MAX-MIN THEOREM

From [6] of Sec. II, we can characterize E_k , the k'th smallest eigenvalue of H by

$$E_k = \min_{\tilde{\Psi}} (\tilde{\Psi}, H\tilde{\Psi})/(\tilde{\Psi}, \tilde{\Psi}) \quad ; \quad (\Psi_i, \tilde{\Psi}) = 0 \,, i = 1, \ldots k-1 \qquad \text{(A1)}$$

where the Ψ_k 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. We now want to point out that there exists another variation approach, the so called "Max-Min Theorem", which does not require explicit information about lower states. Namely one can show that

$$E_k = \max_{w_i} \min_{\tilde{\Psi}} (\tilde{\Psi}, H\tilde{\Psi})/(\tilde{\Psi}, \tilde{\Psi}) \qquad (w_i, \tilde{\Psi}) = 0 \,, i = 1 \ldots k-1 \qquad \text{(A2)}$$

where the w_i are k - 1 arbitrary functions. In words one first fixes the w_i and determines the minimum of $\tilde{E}$ subject to the constraint that $\tilde{\Psi}$ be orthogonal to the w_i . This minimum is then a functional of the w_i . To find E_k one then maximizes with respect to the w_i . We will now give a brief proof that these two definitions of E_k are equivalent. A more detailed proof with references and historical comment can be found in S. H. Gould, <u>Variational Methods for Eigenvalue Problems</u> , Second Edition (Oxford, 1966) Sec. II.6.

We first note that whatever functions one chooses for the w_i , they span a space which is at most k - 1 dimensional, and that therefore there is at least one linear combination of the $\Psi_1 \ldots \Psi_k$ which is orthogonal to all the w_i and therefore is a suitable $\tilde{\Psi}$ for

(2). However for this $\tilde{\varphi}$ we have, writing it as $\tilde{\varphi} = \sum_{L=1}^{K} A_L \psi_L$ that

$$\tilde{E} = \frac{(\tilde{\varphi}, H\tilde{\varphi})}{(\tilde{\varphi}, \tilde{\varphi})} = \frac{\sum_{L=1}^{K} |A_L|^2 E_L}{\sum_{L=1}^{K} |A_L|^2} \leq E_K$$

Therefore we have the result that whatever functions are chosen for the w_i, the maximum of the minima in (2) cannot exceed E_K. On the other hand if we choose the w_i to be equal to the ψ_i, $i=1\cdots K-1$ then from (1) it follows that the minimum for this choice of the w_i is precisely E_K, hence (2) follows.

We will now use (2) to give an elegant derivation of some of the results in Sec. VII. First we will derive the separation theorem (VII-10). Let $\bar{\bar{H}}$ be the projection of H onto the M + 1 dimensional space spanned by the ϕ_K and ϕ. Then

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

$$\hat{E}_K(M) = \underset{w_i}{\text{Max}}\ \underset{\tilde{\varphi}}{\text{Min}}\ (\tilde{\varphi}, \bar{\bar{H}}\tilde{\varphi})/(\tilde{\varphi}, \tilde{\varphi}) \quad ; (w_i, \tilde{\varphi}) = 0,\ i=1\ldots K-1 \quad (\phi, \tilde{\varphi}) = 0$$

$$\hat{E}_{K-1}(M) = \underset{w_i}{\text{Max}}\ \underset{\tilde{\varphi}}{\text{Min}}\ (\tilde{\varphi}, \bar{\bar{H}}\tilde{\varphi})/(\tilde{\varphi}, \tilde{\varphi}) \quad (w_i, \tilde{\varphi}) = 0\ \ i=1\ldots K-2 \quad (\phi, \tilde{\varphi}) = 0$$

where the w_i are selected from the M + 1 dimensional space.

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 Ψ is more restricted so that the Min can't be lower and we have

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

which completes the derivation of (VII-10).

Now following Perkins, J. Chem. Phys. <u>45</u>, 2156 (1966), we will derive the analog of (VII-12) for an arbitrary excited state. (This paper also contains some numerical examples) If $\bar{H}$ is the projection of H onto the M + 1 dimensional space spanned by the Φ_k then the two procedures can be characterized by

$$\hat{E}_k(M+1) = \max_{w_i} \min_{\Psi} (\Psi, \bar{H}\Psi)/(\Psi,\Psi) ; \ (w_i, \Psi) = 0 \quad i = 1 \ldots k-1$$

and by

$$\hat{E}_1 = \min_{\Psi} (\Psi, \bar{H}\Psi)/(\Psi,\Psi) ; \ (\psi_i, \Psi) = 0 \quad i = 1 \ldots k-1$$

respectively. Thus whether or not the ψ_i are eigenfunctions of H , we have

$$\hat{E}_1 \leq \hat{E}_k(M+1)$$

However the ψ_i <u>are</u> the eigenfunctions of H associated with the lower eigenvalues so we also know that

$$E_k \leq \hat{E}_1$$

so that we have

$$E_k \leq \hat{E}_1 \leq \hat{E}_k(M+1)$$

which is the desired generalization of (VII-12).

APPENDIX B: LAGRANGE MULTIPLIERS

We wish to find the consequences of

$$\delta \hat{E} = 0 .$$

in a situation in which the parameters and/or functions which label $\hat{\psi}$ satisfy certain equations of constraint

$$\mathcal{C}_{\alpha} = 0 \qquad \alpha = 1 \cdots c \tag{B1}$$

To be specific, and since it is the most easily visualized case, suppose that $\hat{E}$ depends on M real parameters $\hat{\alpha}_1$, —— $\hat{\alpha}_M$. The direct approach is first to use the equations (1) to extract an independent set of parameters, say $\hat{\alpha}_1$ —— $\hat{\alpha}_R$ in terms of which all the others may be expressed. Next one writes $\hat{E}$ in terms of these independent parameters, and, denoting the result by $\hat{\mathcal{E}}$, calculates $\delta \hat{E}$ from

$$\delta \hat{E} = \sum_{i=1}^{R} \frac{\partial \hat{\mathcal{E}}}{\partial \hat{\alpha}_i} \delta \hat{\alpha}_i \tag{B2}$$

Then since the $\delta \hat{\alpha}_i$, $i = 1 \cdots R$ are aribtrary, $\delta \hat{E} = 0$ yields

$$\frac{\partial \hat{\mathcal{E}}}{\partial \hat{\alpha}_i} = 0 \tag{B3}$$

as the equations to be solved.

Another approach is the method of Lagrange multipliers. Here the prescription is to first require

$$\sum_{j=1}^{M} \left(\frac{\partial \hat{E}}{\partial \hat{\alpha}_j} + \sum_{\alpha=1}^{c} \lambda_\alpha \frac{\partial \mathcal{C}_\alpha}{\partial \hat{\alpha}_j} \right) \delta \hat{\alpha}_j = 0 \tag{B4}$$

without regard to the lack of independence of the $\delta \hat{\alpha}_j$; that is (*) one solves

$$\frac{\partial \hat{E}}{\partial \hat{a}_j} + \sum_{\alpha=1}^{c} \lambda_\alpha \frac{\partial \mathcal{C}_\alpha}{\partial \hat{a}_j} = 0 \qquad \text{(B5)}$$

where the λ_α, the Lagrange multipliers, are, for the moment, unknown parameters. The solutions of (5) will depend on the λ_α and the latter are then to be chosen so that the equations (1) are satisfied.

We now want to show that these two procedures are equivalent. The point is simply that if the equations (1) are satisfied so that we can use them to determine $\hat{a}_{R+1}$ -- -- $\hat{a}_M$ in terms of $\hat{a}_1$ -- -- $\hat{a}_R$ then the equations (1) will also imply that

$$\sum_{j=1}^{M} \frac{\partial \mathcal{C}_\alpha}{\partial \hat{a}_j} \cdot \frac{\partial \hat{a}_j}{\partial \hat{a}_i} = 0 \qquad i=1 \cdots R \qquad \text{(B6)}$$

Therefore, multiplying (5) by $\frac{\partial \hat{a}_j}{\partial \hat{a}_i}$ and summing over j, we find

$$\sum_{j=1}^{M} \frac{\partial \hat{E}}{\partial \hat{a}_j} \frac{\partial \hat{a}_j}{\partial \hat{a}_i} = 0 \qquad i=1 \cdots R \qquad \text{(B7)}$$

which clearly is the same as (3) since the left hand side is just $\partial \hat{E}/\partial \hat{a}_i$ with the understanding that we have used the constraints to express $\hat{a}_{R+1}$ -- -- $\hat{a}_M$ in terms of $\hat{a}_1 \ldots \hat{a}_R$, i.e. it is just $\partial \hat{E}/\partial \hat{a}_i$

---

*Equations (1) and (5) are equivalent to $\delta \hat{\mathcal{L}} = 0$ without constraint where $\hat{\mathcal{L}} = \hat{E} + \sum_{\alpha=1}^{c} \lambda_\alpha \mathcal{C}_\alpha$: variation of $\hat{a}_j$ yields (5) while variation of λ_α yields (1).

REFERENCES

Sec. 0

1. E. Schroedinger, Ann. der Phys. 79, 361 (1926).
2. See for example, W. Yourgrau and S. Mandelstam, Variational Principles in Dynamics and Quantum Theory 3rd ed. (W. B. Saunders, Philadelphia, 1968).
3. K. Ruedenberg, Rev. Mod. Phys. 34, 326 (1962). For a recent application and references see M. J. Feinberg, K. Ruedenberg, and E. L. Mehler in Advances in Quantum Chemistry 5, (1970) P. Lowdin (ed.) Academic Press, N. Y.

Sec. III

1. See for example, A. C. Wahl and G. Das in Advances in Quantum Chemistry 5, (1970) P. O. Löwdin (ed.) (Academic Press, New York).
2. a) H. D. Schaefer III, D. R. Mc Laughlin, F. E. Harris and B. J. Alder, Phys. Rev. Lett. 25, 988 (1970); b) B. Bertoncini, and A. C. Wahl, Phys. Rev. Lett. 25, 991 (1970).

Sec. V

1. M. G. Hegyi, M. Mezei, and T. Szondy, Theoret. Chem. Acta 15, 273 (1969) and references given there, discuss particularly the case in which the χ_κ are given functions. An application to the Li_2 molecule is given by P. Gombas and T. Szondy, Int. J. Q. Chem. 4, 603 (1970). The very promising "Transcorrelated Method" of S. F. Boys fits into the pattern of (15) with the χ_κ involving the parameters in Ψ. For a recent application and references to earlier papers see S. F. Boys and N. C. Handy,

Proc. Roy. Soc. (London) 311, 309 (1969). For some connections with a variation principle see N. C. Handy and S. T. Epstein, J. Chem. Phys. 53, 1392 (1970), also C. Schwartz, J. Comp. Phys. 2, 90 (1967). Some other references having to do with moments will be found in footnote B of Sec. VI.

Sec. VI

1. For some other examples and for references see J. Katriel and G. Adam, Chem. Phys. Lett. 6, 431 (1970).
2. See N. W. Winter, A. Laferrière, and V. McKoy, Phys. Rev. A2, 49 (1970). This paper also contains references to the earlier work of this group. See also C. G. Barraclough, and J. R. Mooney, J. Chem. Phys. 54, 35 (1971). This paper also contains some results of a direct numerical attack on the full ground state Helium problem (3 variables).
3. For the S-limit see H. L. Davis, J. Chem. Phys. 39, 1827 (1963). For a more recent calculation and further references see G. S. Handler and H. W. Joy, J. Chem. Phys. 47, 5074 (1967).
4. For a detailed discussion, in a special case, of the convergence as one increases the size of the basis set see C. F. Bunge, Theoret. Chem. Acta 19, 74 (1970).

Sec. VII

1. In the physics and chemistry literature the results of this section are associated with the names of E. A. Hylleraas and B. Undheim, Z. f. Phys. 65, 759 (1930); and J. K. L. MacDonald, Phys. Rev. 43, 830 (1933).

2. For a discussion of the degenerate case on its own merits rather than as a limiting case see D. W. Davies, J. Chem. Phys. 33, 781 (1960).

Sec. VIII

1. V. Kaldor and F. E. Harris, Phys. Rev. 183, 1 (1968). V. Kaldor, Phys. Rev. A1, 1586 (1970). S. Larsson, Phys. Rev. A2, 1248 (1970).
2. P. O. Lowdin, Phys. Rev. 97, 1509 (1955) and in Quantum Theory of Atoms, Molecules, and the Solid State, P. O. Löwdin ed. (Academic Press, New York, 1966) page 601. For a summary see the references 1. See also R. Pauncz, The Alternant Molecular Orbital Method (W. B. Saunders, Philadelphia, 1967), and reference 1 of Sec. XV below.
3. Concise bibliographies can be found in N. G. Mukherjee and R. McWeeny, Int. J. Q. Chem. 4, 97 (1970) and in A. C. Wahl and G. Das Advances in Quantum Chemistry 5 (1970), P. O. Löwdin (ed.) (Academic Press, New York).
4. Compare the calculations of reference 2, Sec. III.
5. J. F. Perkins, J. Chem. Phys. 42, 3827 (1925). See also W. A. Goddard III, Phys. Rev. 176, 106 (1968) Appendix C.

Sec. IX

1. See for example E. U. Condon and G. H. Shortley The Theory of Atomic Spectra (Cambridge 1935) Chap. 6. For the rules if the spin orbitals are not orthonormal see P. O. Löwdin, Phys. Rev. 97, 1474 (1955).
2. For a general discussion of such procedures see P. O. Löwdin in Advances in Quantum Chemistry 5 (1970), P. O. Lowdin (ed.) Academic Press, New York, page 185.

3. A general review of these matters has been given by T. Gilbert in *Molecular Orbitals in Chemistry, Physics, and Biology*, P. O. Löwdin and B. Pullman (ed.) (Academic Press, New York 1964). Also see C. Edmiston and K. Ruedenberg, Rev. Mod. Phys. 35, 457 (1963), and K. Ruedenberg in *Modern Quantum Chemistry* 1, O. Sinanoglu (ed.) (Academic Press, New York (1965)). A recent paper on localized orbitals is S. Rothenberg, J. Chem. Phys. 51, 3389 (1964). See also A. T. Amos and G. C. Hall, Proc. Roy. Soc. A263, 483 (1962) who have introduced the so-called "corresponding orbitals".
4. See for example D. R. Hartee *The Calculation of Atomic Structures* (John Wiley and Sons, New York (1957)). Also C. Froese-Fischer, Can. J. Phys. 41, 1895 (1963); J. Chem. Phys. 45, 1417 (1963).
5. C. C. J. Roothaan, Rev. Mod. Phys. 23, 69 (1951); and G. G. Hall, Proc. Roy. Soc. A205, 511 (1951).
6. For specific equations determining non-linear parameters see M. Simonetti and E. Gianinetti in *Molecular Orbitals in Chemistry, Physics and Biology* op cit. Also J. P. Olive, J. Chem. Phys. 51, 4340 (1969).
7. For a numerical comparison of some analytic and numerical functions see P. S. Bagus, B. Liu, and H. Schaeffer III, Phys. Rev. A2, 555 (1970).
8. For the details of a procedure in which different $\hat{\varphi}_j$ are expanded in essentially different basis sets see M. T. Marron, N. C. Handy, R. G. Parr, and H. J. Silverstone, Int. J. Q. Chem. 4, 245 (1970).

9. C. C. J. Roothaan, and P. S. Bagus in Methods in Computational Physics 2 (1963), B. Alder, S. Fernbach and M. Rotenberg (ed.) (Academic Press, New York).

10. The paper by H. J. Silverstone, J. Chem. Phys. 48, 4098 (1968) which is the first in a series contains an extensive list of references to integration techniques using Slater type orbitals, as well as many new results.

11. Recent reviews may be found in K. Jug, Theoret. Chim. Acta (Berl.) 14, 91 (1969); R. D. Brown and K. R. Roby, Theoret. Chim. Acta (Berl.) 16, 175 and 194 (1970); J. A. Pople and D. L. Beveridge, Approximate Molecular Orbital Theory (Mc Graw Hill, New York, 1970). A general discussion of the status of "non approximate" (within the SCF framework) methods has been given recently by E. Clementi, Chem. Rev. 68, 341 (1968).

12. Experimental Data: J. C. Slater Quantum Theory of Atomic Structure 1 (Mc Graw Hill, New York, (1963). $\mathcal{E}_i$: E. Clementi, IBM J. Res. and Devel. Suppl. 9, 2 (1962).

13. T. Koopmans, Physica 1, 105 (1933).

14. For a critical discussion see M. D. Newton, J. Chem. Phys. 48, 2825 (1968).

Sec. X

1. See for example C. C. J. Roothaan, Rev. Mod. Phys. 23, 69 (1951).

2. See for example S. Larsson, Phys. Rev. 1248 (1970).

3. C. C. J. Roothaan, Rev. Mod. Phys. 32, 174 (1960). A review is given by G. Berthier in Molecular Orbitals in Chemistry, Physics, and Biology, P. O. Löwdin and B. Pullman (ed.) (Academic Press,

New York, 1964). For some recent developments see W. J. Hunt, W. A. Goddard III, and J. H. Dunning Jr., Chem. Phys. Lett. 6, 147 (1970), and S. Huzingas, J. Chem. Phys. 51, 3071 (1969). For some numerical comparisons of various methods see D. H. Sleaman, Theoret. Chem. Acta 11, 135 (1968).

4. See for example R. Kari and B. T. Sutcliffe, Chem. Phys. Lett. 7, 149 (1970) for a recent discussion.

Sec. XI

1. L. Brillouin, Actualités Sci. Ind. No. 159 (1934). For MCSCF see B. Levy and G. Berthier, Int. J. Q. Chem. 2, 307 (1968).
2. See for example R. K. Nesbet, Rev. Mod. Phys. 33, 28 (1961); and R. Lefebvre in Modern Quantum Chemistry 1, O. Sinanoglu (ed.) (Academic Press, New York, 1965).
3. For other choices see S. T. Epstein, J. Chem. Phys. 41, 1045 (1964); H. P. Kelly, Phys. Rev. 136, B896 (1964); P. Claverie, S. Diner, and J. P. Malrieu, Int. J. Q. Chem. 1, 751 (1967); and subsequent papers by various of these authors in Theoret. Chim. Acta; S. T. Manson, Phys. Rev. 145, 35 (1966); E. Steiner, J. Chem. Phys. 46, 1717 and 1727 (1967); H. J. Silverstone and M. L. Yin, J. Chem. Phys. 49, 2026 (1968); S. Y. Chang, E. R. Davidson, and G. Vinlon, J. Chem. Phys. 49, 529 (1968) especially footnote 11.
4. H. P. Kelly, Phys. Rev. 131, 684 (1963).
5. C. Møller, and M. S. Plesset, Phys. Rev. 46, 618 (1934); M. Cohen and A. Dalgarno, Proc. Phys. Soc. (London) 77, 748 (1961); G. G. Hall, Phil. Mag. 6, 249 (1961).

6. For more details see A. Dalgarno, Advan. in Phys. 11, 281 (1962) esp. p. 307 et seq.
7. D. R. Hartree, W. Hartree and B. Swirles, Phil. Trans. Roy. Soc. 238, 229 (1939).
8. See for example G. W. F. Drake and M. Cohen, J. Chem. Phys. 48, 1168 (1968) and references given there.
9. For an extensive compilation for atoms see Many Electron Systems: Properties and Interactions, S. Fraga and G. Mali (Saunders, Phila. 1968). For some molecular properties see for example J. Geratt and I. M. Mills, J. Chem. Phys. 49, 1719 (1968).
10. See for example C. F. Bender and E. R. Davidson, J. Chem. Phys. 49, 4222 (1968); Phys. Rev. 183, 23 (1969); S. Green, J. Chem. Phys. 54, 827 (1971), and references therein.
11. P. Claverie, et. al. ref. 3; S. Y. Chang, et. al, ref. 3; S. Diner, J. P. Malrieu and P. Claverie, Theoret. Chim. Acta 13, 1 (1969); J. P. Malrieu, P. Claverie, and S. Diner, Theoret. Chim. Acta 13, 18 (1969); S. Diner et. al. Theoret. Chim. Acta 18, 86 (1970); Z. Gershgorn and I. Shavitt, Int. J. Q. Chem. 2, 751 (1968).

Sec. XII

1. For a recent applications and references to the earlier literature see R. S. Loeb and Y. Raisel, J. Chem. Phys. 52, 4995 (1970); D. R. Whitman and R. N. Carpenter, J. Chem. Phys. 53, 3830 (1970).

Sec. XIII

1. E. Brändas, J. Mol. Spec. 27, 236 (1968); E. Brändas and O. Martensson, Chem. Phys. Lett. 3, 315 (1969); C. A. Coulson and R. J. White, Mol. Phys. 18, 577 (1970).

Sec. XIV

1. See almost any (advanced) quantum mechanics text.
2. For some examples of $\hat{\Phi}$ and references to earlier work see I. Epstein, J. Chem. Phys. 53, 4418 (1970).

Sec. XV

1. See for example T. Amos and G. G. Hall, Proc. Roy. Soc. A263, 483 (1962); T. Amos and L. C. Snyder, J. Chem. Phys. 41, 1773 (1964). A recent reference is T. A. Claxton and D. McWilliams, Theoret. Chim. Acta 16, 346 (1970). For a review in the context of nuclear physics see N. MacDonald, Adv. in Phys. 19, 79 (1970).

Sec. XVI

1. For the more general case of a σ dependent volume element see S. T. Epstein, J. Chem. Phys. 46, 571 (1967), and for a discussion of the connection between theorems in different coordinate systems see J. Chem. Phys. 42, 3813 (1965).
2. A. C. Hurley, Proc. Roy. Soc. A226, 179 (1954). See also his article in *Molecular Oribtals in Chemistry, Physics and Biology* op. cit.
3. For a different proof and for some numerical confirmation see M. Cohen and A. Dalgarno, Proc. Roy. Soc. A275, 492 (1963).

Sec. XVII

1. J. O. Hirschfelder, J. Chem. Phys. 33, 1762 (1960).

Sec. XVIII

1. Ref. 1 of Sec. XVII.

Sec. XIX

1. R. Sternheimer, Phys. Rev. 96, 951 (1954).

Sec. XXI

1. B. Nelander, J. Chem. Phys. 51, 469 (1969).
2. E. A. Hylleraas, Z. f. Physik 54, 347 (1929). V. Fock, Z. f. Physik, 63, 855 (1930). P. O. Löwdin, J. Mol. Spec. 3, 46 (1959) contains an extensive discussion and bibliography.
3. C. A. Coulson and A. C. Hurley, J. Chem. Phys. 37, 448 (1962); P. Phillipson, J. Chem. Phys. 39, 3010 (1963).
4. A. C. Hurley, Proc. Roy. Soc. A226, 170 (1954).

Sec. XXII

1. D. P. Chong and M. L. Bengston, J. Chem. Phys. 49, 1302 (1968) discuss how to satisfy such theorems by contraining $\hat{\Psi}$.
2. R. G. Parr, J. Chem. Phys. 40, 3726 (1964); J. W. Richardson, J. Chem. Phys. 41, 897 (1964).

FOOTNOTES

Sec. II

A. If one works in configuration space the existence of $(\tilde{\Psi}, H\tilde{\Psi})$ requires among other things that $\tilde{\Psi}$ be twice differntiable. 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 the results which we will derive in this section will still hold even if $\tilde{\Psi}$ is only once differentiable. (See E. Courant and D. Hilbert Methods in Mathematical Physics 1 bottom of page 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 expression (1) because it is much easier to deal with formally.

B. In this and succeeding sections (and in the preceding footnote) we will use the language of molecular bound state quantum mechanics. In particular we refer to H as the Hamiltonian having in mind that it is the internal Hamiltonian (or some approximation thereto) of a typical atom or molecule. However it should be kept in mind as we proceed that many of our formal results hinge only on H being a Hermitian operator - be it differential, integral or finite matrix - with a (at least partially) discrete spectrum.

C. Since we have imposed no normalization requirements on $\tilde{\Psi}$ and $\hat{\Psi}$, the size of $\tilde{\Delta}$ is not a true measure of the difference between $\tilde{\Psi}$ and $\hat{\Psi}$ in that even if $\tilde{\Delta}$ is large they may still be describing the same state - thus let $\tilde{\Psi} = 7\hat{\Psi}$. An accurate measure provided by $\Delta_1/(\tilde{\Psi}, \tilde{\Psi}) \equiv \delta_1$ where

$$\Delta_\perp = \tilde{\Delta} - \hat{\Psi}(\hat{\Psi}, \tilde{\Delta})/(\hat{\Psi}, \hat{\Psi})$$

is that part of $\tilde{\Delta}$ which is orthogonal to $\hat{\Psi}$. It is then of some interest that with (6), Eq. (7) can be written as

$$\tilde{E} = \hat{E} + (\delta_\perp, (H - \hat{E})\delta_\perp)$$

D. If H is a finite matrix (recall footnote B) or more generally if its spectrum is bounded from above then clearly the largest eigenvalue is an absolute maximum of $\tilde{E}$.

Sec. III

A. Another approach to getting a bound on the difference is to fix one member by experiment. However in this connection it should be kept in mind that H is almost certainly an approximate Hamiltonian and therefore further corrections must be applied to the $\hat{E}$ (or to the experimental data) in order that the two numbers refer to the same physical (or mathematical) problem.

Sec. VI

A. In introducing the linear variation method in Sec. V we stated that the ϕ_k should be linearly independent. This requirement has played no real role until now. However if the ϕ_k are not linearly independent there will be less than M independent equations in (V-12) and consequently the secular determinant will vanish identically. Thus only if the ϕ_k are linearly independent is (1) really an equation for $\hat{E}$. In practice, particularly when using large non orthogonal basis sets, near linear dependence can often become a real numerical problem.

B. The authors of the first of references 1 refer to their method as the "method of moments". Unhappily this same name has also long been applied to a particular version of the linear variation method in which the Φ_k are taken to have the form

$$\Phi_k = H^{k-1}\Phi_1$$

where Φ_1 in some given function. This particular choice of the Φ_k however is of little use in molecular problems since the integrals $(\Phi_1, H^L \Phi_1)$ may well not exist for L larger than 2 or so even with a very reasonable choice for Φ_1. [In the paper by C-Y Hu, Phys. Rev. 167, 112 (1968) it might appear that the method of moments has been applied to the Helium Hamiltonian. However a careful reading of the paper shows that it is actually being applied to a finite matrix approximation to that Hamiltonian]. A detailed discussion, with bibliography, of this method, can be found in the paper of J. B. Delos, S. M. Blinder, J. Chem. Phys. 47, 2784 (1927). They also propose a method which is a computationally practical combination of the two methods of moments in which the χ_k of Sec. V take the form of a δ-function multiplying (not multiplied by) various powers of H . They refer to it as the "γ - method". It was also proposed independently together with a related method by H. Silverstone, M-L Yin, and R. L. Somarjai, J. Chem. Phys. 47, 4824 (1967), and some illustrative calculations have been made by J. M. Rothstein, J. E. Welch, and H. J. Silverstone, J. Chem. Phys. 51, 2932 (1969).

Sec. VII

A. If the linear space contains functions of various symmetries but in the form of a direct sum (that is the linear space can be decomposed into linear subspaces each having a definite symmetry) then since Π will commute with $\mathcal{T}$ so will $\overline{H}$ if H does. Therefore the $\hat{\psi}_k$, as eigenfunctions of $\overline{H}$, will automatically have, or can be chosen to have, definite symmetry. In such a situation then the $\hat{\psi}_k$ and $\hat{E}_k$ we are talking about in this section are the successive $\hat{\psi}_k$ and $\hat{E}_k$ of a given symmetry. Finally if H commutes with $\mathcal{T}$ but $\overline{H}$ does not, so that in general the $\hat{\psi}_k$ won't have definite symmetry, then all we will be able to say from the results of this section is that the successive $\hat{E}_k$ are upper bounds to the successive eigenvalues of H ordered without regard to symmetry.

B. If H has only $M'<M$ bound state eigenvalues (of appropriate symmetry) then for the $\hat{E}_{M'+1} \dots \hat{E}_M$ one will be able to conclude only that they are all upper bounds to the highest bound state of H . However in what follows we will not consider this possibility explicitly. Also we will not worry about such interesting things as bound states and quasi bound states imbedded in continua of the same symmetry. For a recent review of the application of the linear variation method to such situations see H. S. Taylor, Advances in Chemical Physics 18 (1970), I. Priggogine and S. A. Rice (ed.) (Interscience New York). Also A. U. Hazi and H. S. Taylor, Phys. Rev. A1, 1109 (1970).

C. If H is a finite matrix, or more generally an operator whose highest eigenvalues are all discrete and bounded from above, then one can clearly prove a theorem analogous to that of (6), Sec. II but with minimum replaced by maximum and smaller replaced by larger (recall also footnote D, Sec. II). Correspondingly, since $\tilde{\mathcal{E}}$ in (1) is also larger than $\hat{\mathcal{E}}_{k''}$ where k'' is the smallest k value for which $b_k \neq 0$, one can show that if H has dimension $D \geq M$ then

$$\hat{\mathcal{E}}_{M-m} \leq E_{D-m} \qquad m=0,1,\ldots M-1$$

Sec. VIII

A. For nuclei the Pauli principle plays a large role in validating the independent particle picture. See for example V. F. Weisskopf, Physics Today, July 1961, page 18.

Sec. IX

A. A point of notation. In our general discussion the symbol (,) has denoted a scalar product in the N-particle space. In the first sum in (5) however we use the same symbol for a scalar product in a one particle space and in the second sum it is used for a scalar product in a two particle space. This should cause no confusion if one keeps in mind always the nature of the operators and functions involved. For example Eq. (7) below contains $(\hat{\varphi}_i , g\hat{\varphi}_i)$ which is a scalar product over the variables of φ_i but which is still a function of the remaining variables in g .

B. These other sets are readily shown to satisfy (12) with the ϵ_{ij} the appropriate unitary transformation of the ϵ_i . The essential point is that since $\hat{f}$ involves the spin orbitals only in the form of a "scalar product" $\sum_i \hat{\varphi}_i^*(1)\varphi_i(2)$, it is invariant to such a transformation of spin orbitals. For a detailed discussion see C. C. J. Roothaan, Rev. Mod. Phys. <u>23</u>, 69 (1951).

C. However even with these one-dimensional equations the fact that the exchange terms involve an integral operator rather than being a local potential often makes calculation difficult. Therefore there has been considerable investigation and use of local approximations to the exchange terms. The original suggestion was due to J. C. Slater, Phys. Rev. <u>81</u>, 385 (1951). For recent discussions see T. M. Wilson, J. H. Wood, and J. C. Slater, Phys. Rev. <u>A2</u>, 620 (1970), and J. C. Slater and J. H. Wood, Int. J. Q. Chem. <u>S4</u>, 3 (1971).

D. The result which we have just proven is actually only a corrollary of what is really Koopman's theorem: Let the $\hat{\varphi}_j$ be some orthonormal set of UHF spin orbitals, not necessarily the canonical set. Then we delete one of them to yield a trial function for the N - 1 particle system. We now fix the $\hat{\varphi}_j$ i.e. fix the unitary transformation which relates them to the canonical spin orbitals, by requiring that they be such as to make the energy of the N - 1 particle system stationary. Since the energy of the N particle system is fixed this then means Δ E stationary and clearly we still have

$$\Delta E = (\hat{\varphi}_k, \hat{f}\hat{\varphi}_k)$$

since in the discussion in the text we used the canonical nature of the spin orbitals only in the final step. We now require that $\delta(\Delta E)=0$. Since $\delta\hat{f}=0$ (see footnote B above) this means

$$(\delta\hat{\varphi}_k, \hat{f}\hat{\varphi}_k) + (\hat{\varphi}_k, \hat{f}\,\delta\hat{\varphi}_k)=0$$

In addition we have constrained $\hat{\varphi}_k$ to be normalized so we also have

$$(\delta\hat{\varphi}_k, \hat{\varphi}_k) + (\hat{\varphi}_k, \delta\hat{\varphi}_k)=0$$

and thus we are led to

$$\hat{f}\hat{\varphi}_k = \epsilon_k\hat{\varphi}_k$$

That is (and this is really Koopman's theorem) the optimal $\hat{\varphi}_k$ to remove when one uses trial functions of this sort for the N-1 particle system is precisely the canonical one. An analogous theorem for excitation of the N particle system has been discussed by W. J. Hunt and W. A. Goddard III, Chem. Phys. Lett. 3, 414 (1969).

E. The frozen spin orbital wave function would seem more appropriate to describe the N-1 particle system immediately after the sudden ionization of the N particle system (assuming that UHF gives an adequate description of the N particle system). For a discussion of the implications of $\Delta E = \epsilon_k$ from this point of view see R. Manne and T. Åberg, Chem. Phys. Lett. 7, 282 (1970).

Sec. X

A. In the case of ground states there is a large literature discussing the question of whether or not such solutions, though they exist, represent an absolute minimum or indeed even a local minimum within UHF. For a recent discussion and an extensive list of references see for example J. Pauldus and J. Cizek, Phys. Rev. 2, 2268 (1970). See also J. I. Musher, Chem. Phys. Lett. 7, 397 (1970).

Sec. XI

A. With this choice of H_0, $\mathcal{E} = \sum_{i=1}^{N} \epsilon_i \neq \hat{E}$. Note however that quite generally, whatever $\hat{\Phi}$, H_0 and $\mathcal{E}$ that $\mathcal{E} + \mathcal{E}^{(1)} = (\hat{\Phi}, H_0 \hat{\Phi}) + (\hat{\Phi}, (H - H_0)\hat{\Phi}) = \hat{E}$. For some cases in which $\mathcal{E}$ rather naturally equals $\hat{E}$ see K. H. Hansen, Theoret. Chim. Acta 6, 87 (1966); G. Gliemann, Theoret. Chim. Acta 11, 75 (1968), and references therein.

B. For some special properties of the choice $\bar{H} + (1-\pi)H(1-\pi)$ see W. H. Adams, J. Chem. Phys. 45, 3422 (1966). Note however that except for the $\hat{\Phi}_L$, the eigenfunctions of this operator are by no means obvious. With the choice $H_0 = \bar{H}$ all the Θ_α which contribute to $\Phi^{(1)}$ have $E_\alpha \approx 0$. Therefore completeness immediately yields the simple result

$$\hat{\Phi}_k^{(1)} = \frac{(H - \hat{E}_k)}{\hat{E}_k} \hat{\Phi}_k$$

Sec. XII

A. An interesting question which we won't pursue: If the set "almost" satisfies the sufficient condition, then how nearly can one expect the $\hat{\Phi}$ to have the desired property?

Sec. XIII

A. It would be sufficient for our purposes if the Set was only "effectively" invariant to complex conjugation. That is if there were some constant A , depending perhaps on $\tilde{\Psi}$, such that $A\tilde{\Psi}^*$ is in the set if $\tilde{\Psi}$ is. Since $A\tilde{\Psi}^*$ is physically equivalent to $\tilde{\Psi}^*$ the use of the word "effectively" is clearly justified.

Also similar remarks could be made in later sections where we will assume various other invariances in order to derive other theorems. However the distinction between invariant and effectively invariant is anyway overly pedantic since clearly, without changing the results of the variational calculation in any way we can change an effectively invariant set into an invariant one simply by making the overall scale of the trial functions arbitrary. Thus here and in the sections which follow we will, to simplify the presentation, require invariance to various operations, though effective invariance would suffice.

Sec. XVI

A. Some discussion and references to the early history of this theorem can be found in S. T. Epstein, Am. J. Phys. 22, 613 (1954), and in J. I. Musher, Am. J. Phys. 34, 267 (1966). We will discuss the Hellmann-Feynman theorem in footnote D, Sec. XXI.

B. In spite of our comments in the text about the simplicity and sufficiency of Hurley's theorem, the following derivation is of interest for the linear case. Since $\hat{\varphi}_k$ is an eigenfunction of $\bar{H}$ we have

$$\frac{\partial \hat{E}_k}{\partial \sigma} = (\hat{\varphi}_k, \frac{\partial \bar{H}}{\partial \sigma} \varphi_k) / (\hat{\varphi}_k, \varphi_k)$$

But if the set is invariant to changes in σ then $\partial \bar{U} / \partial \sigma = 0$. Therefore

$$\frac{\partial \hat{E}_k}{\partial \sigma} = (\hat{\varphi}_k, \Pi \frac{\partial H}{\partial \sigma} U \varphi_k) / (\hat{\varphi}_k, \varphi_k)$$

$$= (\varphi_k, \frac{\partial H}{\partial \sigma} \varphi_k) / (\hat{\varphi}_k, \varphi_k)$$

Sec. XVIII

A. S. T. Epstein and J. O. Hirschfelder, Phys. Rev. 123, 1495 (1967) discuss the use of trial functions of the form $e^{i\tilde{\lambda}\cdot\mathbf{r}}\tilde{\Theta}$ for more general $\mathcal{H}$.

Sec. XIX

A. By average force on the nuclei we mean the average, using $\hat{\psi}$ of the familiar classical expression. See footnote D, Sec. XXI for some comments on this definition in the case of molecules.

Sec. XXI

A. $\mathcal{V}$ is the trace of the "tensor virial operator"

$$\mathcal{V}_{ij} = \sum_{s=1}^{N} r_{si}\, p_{sj} - \frac{iN}{2}\delta_{ij}$$

We have seen that by "isotropic scaling", i.e. by scaling all components of the $\vec{r}_s$ equally we can satisfy the hypervirial theorem for $\mathcal{V}$. Similarly by "anisotropic scaling" (scaling each component of $\vec{r}_s$ separately one can guarantee the hypervirial theorem for the diagonal elements of $\mathcal{V}_{ij}$ (for an application see W. J. Meath and J. O. Hirschfelder, J. Chem. Phys. 39, 1135 (1963)). Finally by tensor scaling (D. Pandres, Phys. Rev. 131, 886 (1963)) one can simultaneously guarantee the theorem for all components of $\mathcal{V}_{ij}$. However, as always, symmetry alone may be enough to guarantee some of these theorems (or to guarantee some, given others).

B. For a polyatomic molecule the $R_{\ell c)}$ and the bond angles are not all independent. Thus $\hat{E}$ may be written in various ways as functions of these quantities. In (28) the sum runs over those particular bond lengths, independent or not, which one has chosen to use in expressing $\hat{E}$.

C. If one does not use a fixed nucleus approximation then it is easy to show that if the internal Hamiltonian is written in terms of internal Cartesian coordinates $\vec{\xi}_\alpha$ and their canonically conjugate momenta $\vec{\pi}_\alpha$, then the hypervirial theorem for $\sum_\alpha \vec{\pi}_\alpha \cdot \vec{\xi}_\alpha$ is the statement that $2\hat{T} + \hat{V} = 0$ where $\hat{T}$ now is the average total (electronic and nuclear) internal kinetic energy and where $\hat{V}$ is the corresponding average potential energy. This theorem can be ensured by using a set of $\hat{\varphi}$ which is invariant to scaling of the $\vec{\xi}_\alpha$.

D. Equation (26) is the variational version of <u>the</u> Hellmann-Feynman theorem; H. Hellmann, <u>Einfuhring in die Quanten Chemie</u> (Franz Denticke; Leipzig, 1937) p. 285, and R. P. Feynman, Phys. Rev. <u>56</u>, 340 (1937). In words it says that, as classically, one can calculate the force on a nucleus by calculating the negative gradient of the energy with respect to nuclear coordinates. Actually the theorem is often read the other way around. That is, because of the use one makes of $\hat{E}$ in the Born-Oppenheimer approximation, the average force on nucleus 1 say is, a priori, taken to be $-\partial\hat{E}/\partial\vec{R}_1$. The theorem then says that it can also be calculated from the electronic charge density as the average value of the classical force operator $\vec{F}_1$. (See for example P. Pulay, Mol. Phys. <u>17</u>, 197 (1969) especially Sec. 3). Moreover the theorem often appears in other forms, forms which are all equivalent if various theorems are satisfied (and if $\hat{\Psi}$ is an eigenfunction, they all *are* satisfied). Thus consider a diatomic molecule. Then if the force theorems are satisfied we can write the force on nucleus one more generally as

$$-\left(\hat{\Psi},\left(a\frac{\partial H}{\partial \vec{R}_1}+b\frac{\partial H}{\partial \vec{R}_2}\right)\hat{\Psi}\right)/(\hat{\Psi},\hat{\Psi})$$

where a and b are arbitrary numbers except that $a-b=1$, and various choices have been used in the literature. Also if, as is usually the case, the $\hat{\vec{F}}_A$ are purely axial one needs only the axial component of this quantity.

E. The atomic orbitals in (18) are centered on the nuclei and are invariant to rotation about the internuclear axis. For reasons of symmetry one might expect the latter also to be true of the optimal orbitals derived from (19), i.e. that the points $\hat{\vec{\gamma}}$ and $\hat{\vec{\mu}}$ will be on the internuclear axis. However there is no reason to expect that $\hat{\vec{\gamma}}/\hat{\xi}$ will equal $\vec{R}_1$, and that $\hat{\vec{\mu}}/\hat{\xi}$ will equal $\vec{R}_2$. That is, as a price one pays for translational invariance (and hence the force theorem) with such a simple set of $\tilde{\Psi}$, the atomic orbitals will have their cusps off the nuclei. Following Hurley (ref. 7) $\tilde{\Psi}$ like (19) are often called "floating wave functions". Eigenfunctions for this problem, of course, have cusps at the nuclei.

F. P. O. Löwdin, J. Mol. Spec. 3, 46 (1959). Löwdin does not mention any particular coordinate system. However he assumes, without comment, that $\hat{V}$ depends only on R and not on $\vec{R}_1$ or $\vec{R}_2$ separately.

Sec. XXII

A. We have already discussed orthogonality to some extent at the end of Sec. VII. In some SCF calculations, orthogonality, though not exact, can however be very nearly realized. See for example P. Bagus, Phys. Rev. 139, A619 (1965). See also M. Cohen and A. Dalgarno, Rev. Mod. Phys. 35, 506 (1963). For a discussion of the situation for off diagonal hypervirial theorems see the references and discussions given in ref. 1. For the integrated Hellmann-Feynman theorem see S. T. Epstein, A. C. Hurley, R. E. Wyatt, and R. G. Parr, J. Chem. Phys. 47, 1275 (1967) and references cited there.

B. The way this is often done in practice is as follows. Let $\Phi_a [\Phi_b]$ be a function derived in some way, usually by a variational calculation for $H = H_a\ [H_b]$. Then one does new linear variational calculations for H_a and H_b, in each case using Φ_a and Φ_b as the basis set. If one wishes to satisfy a sequence of integral Hellmann-Feynman theorems involving a sequence of Hamiltonians $H_i\ ; i = 1, 2, \ldots$, then one does a new linear variational calculation for each H_i using the same set of Φ_i as the basis. In particular (A. C. Hurley, Int. J. Q. Chem. 13, 677 (1967)) if the sequence H_i is continuous the linear variational method leads to homogeneous integral equations instead of algebraic equations.