

Introduction to Theoretical organic Chemistry

Lecture 5.

1.2.7 Basis Functions

Throughout the years some magic has been attached to everything that is theoretical in chemistry and by now a certain amount of mythology has evolved around Quantum Chemistry as it may be applied to study organic problems. For this reason some demythologization is in order. This is hoping to be archived in five brief statements.

Statement no.1 The term „orbital” is a synonym for the term „one-electron” function (OEF).

Statement no.2 A single-centred OEF is synonymous with „Atomic Orbital (AO)”. A multi-centred OEF is synonymous with „Molecular Orbital (MO)”. The single and multi centred nature of these one-electron functions are illustrated in Figure 1.2.7—1

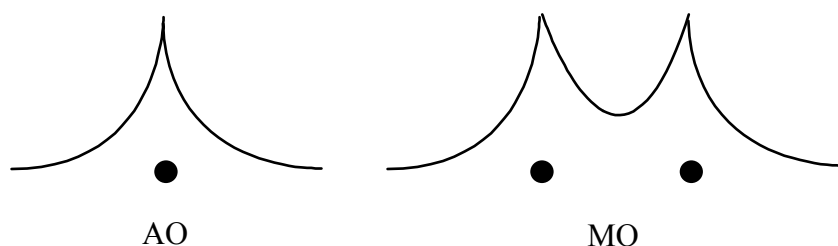


Figure 1.2.7—1 A schematic representation of single (AO) and multi (MO) centred one-electron functions.

An orbital (AO or MO) has as much to do with physical and chemical reality as the following functions do

$$\begin{aligned} &e^x \\ &\{x^n\} \\ &\{\sin k\alpha, \cos k\alpha\} \end{aligned}$$

1.2.7—1. eq.

Nevertheless, these functions (AO and MO) enable us to construct molecular wave functions that may be used to compute molecular electron density and molecular properties which are manifestations of physical and chemical realities. It is axiomatic, therefore, that these latter computed quantities must have a one-to-one correspondence to the corresponding physical properties of the molecule as determined experimentally

Statement no.3 There are three ways to express a mathematical function:

- Explicitly in analytic form

$$f(x) = e^x$$

1.2.7—2. eq.

the hydrogen-like AO are usually expressed in this form.

- As a table of numbers

x	f(x)
0.0	1.000
0.1	1.105
0.2	1.221
⋮	⋮

The Hartree-Fock type AO computed for numerous atoms are usually expressed in this numerical form.

- In the form of an expansion

$$f(x) \equiv e^x = C_0x^0 + C_1x^1 + C_2x^2 + C_3x^3 + \dots = \frac{f(0)}{0!}x^0 + \frac{f'(0)}{1!}x^1 + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 + \dots$$

1.2.7—3. eq.

which is analogous to the expression of MO in terms a set of AO:

$$\phi = C_0\eta_0 + C_1\eta_1 + C_2\eta_2 + C_3\eta_3 + \dots$$

1.2.7—4. eq.

Statement no. 4. The generation of MO(ϕ) from AO(η) is equivalent to the transformation of an N-dimensional vector space to another N-dimensional vector space where $\{\eta\}$ is the original set of non-orthogonal functions. After orthogonalization of the non-orthogonal AO basis set $\{\eta\}$ the orthogonal set $\{\chi\}$ is rotated to the another orthogonal set $\{\phi\}$. This overall process is illustrated in Figure 1.2.7—2

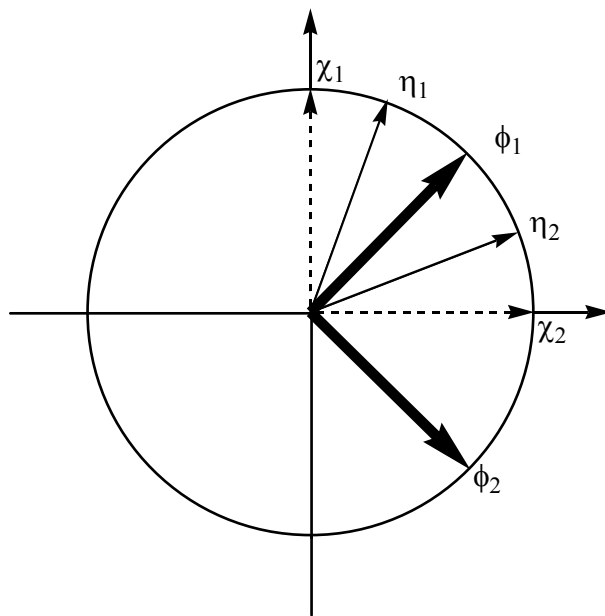
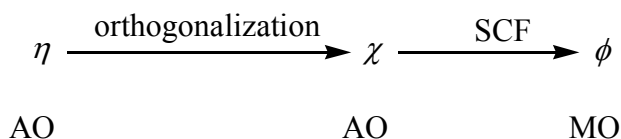


Figure 1.2.7—2 Two dimensional Vector Model of AO→MO transformation



Statement no.5. There are certain differences between the shape of numerical Hartree-Fock atomic orbitals (HF-AO), the analytic Slater type orbitals (STO) and the analytic Gaussian type functions(GTF),however , these differences are irrelevant to the final results as the MO can be expanded in terms of any of these complete sets of functions to any desired degree of accuracy.

Atomic orbital basis sets

The generation of MO from AO requires the generation and transformation of the Fock matrix into diagonal form. The elements of the Fock matrix are assembled from integrals in the following fashion:

$$f_{ij} = \langle i | \hat{h} | j \rangle + \sum_k \sum_l [2\{ij|kl\} - \{ik|jl\}]$$

1.2.7—5. eq.

Where the first term is a one-electron integral and the latter terms are two-electron integrals having the following forms.

One - electron Integrals :

$$\langle i | \hat{h} | j \rangle \equiv \langle \eta_i | \hat{h} | \eta_j \rangle \equiv \int_1 \eta_i(1) \hat{h}_1 \eta_j(1) d\tau_1$$

1.2.7—6. eq.

Two - electron integrals

$$\{ij|kl\} \equiv \left\langle \eta_i \eta_k \left| \frac{1}{r_{12}} \right| \eta_l \eta_j \right\rangle \equiv \int_1 \int_2 \eta_i(1) \eta_k(2) \frac{1}{r_{12}} \eta_l(2) \eta_j(1) d\tau_2 d\tau_1 = \int_1 \eta_i(1) \left\{ \int_2 \eta_k(2) \frac{1}{r_{12}} \eta_l(2) d\tau_2 \right\} \eta_j(1) d\tau_1$$

1.2.7—7. eq.

As the running indices i, j, k, l, range from 1 to N therefore the number of one – and two-electron integrals are calculable by the following formula

$$\text{Number of 1 - electron integrals : } p = \frac{N(N+1)}{2}$$

$$\text{Number of 2 - electron integrals : } q = \frac{p(p+1)}{2}$$

The table below illustrates how rapidly the number of one– and two–electron integrals grows with the basis set size N.

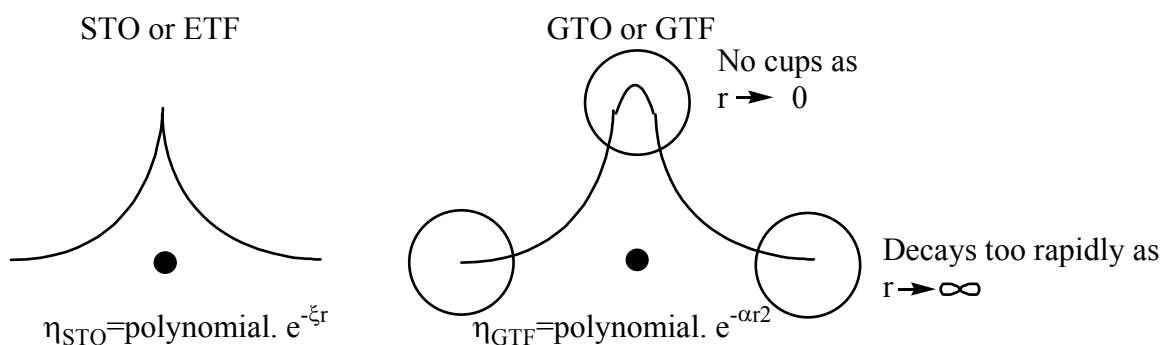
Table 1.2.7-1 The increase of the number of one-electron (p) and two electron (q) integrals with increasing basis set size (N)

N	P	q
1	1	1
10	55	1,540
50	1,275	814,725
100	5,050	12,751,250
150	11,325	64,133,475
200	20,100	202,015,050
300	45,150	1,019,261,250

For the atomic orbital $\{\eta\}$ two types of analytic functions are used in molecular computations.

- 1) Slater-type orbital (STO) or exponential type functions.(ETF).
- 2) Gaussian-type orbital (GTO) or Gaussian-type function (GTF).

Some characteristics of these two sets of function are illustrated in Figure 1.2.7—3 As far as STO are concerned the integral evaluation is very slow but a relatively small N gives fairly accurate results.

**Figure 1.2.7—3 Some characteristics of STO and GTF**

The GTF are more popular as it is possible to compute the integrals over Gaussian very quickly, but only relative large N gives accurate results. These AO basis sets need to be optimized for molecular calculation. This may be achieved by minimizing the electronic energy with respect to all orbital exponents. Figure 1.2.7—4 illustrates the simulation for a two-orbital case:

$$\eta_1 = \text{polynomial}_1 \cdot e^{-\alpha_1 r^2}$$

$$\eta_2 = \text{polynomial}_2 \cdot e^{-\alpha_2 r^2}$$

eq. 1.2.7—8

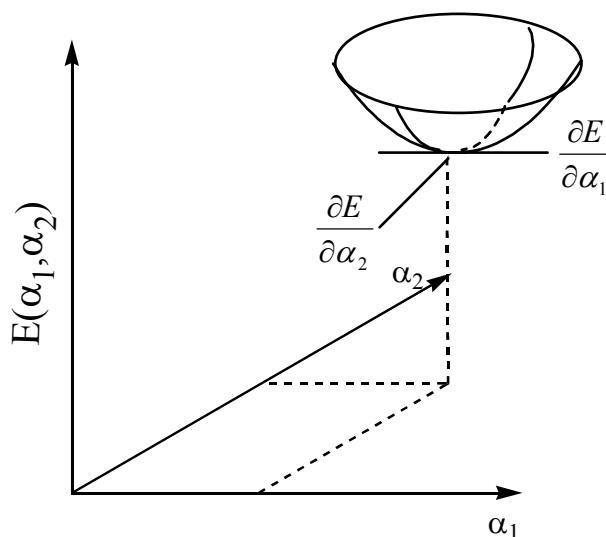


Figure 1.2.7—4 Electronic Energy surface as a function of orbital exponents $\{\alpha\}$

Sometimes STO are expanded in terms of GTF or, in other words, a GTF basis set is contracted to a set of STO. The contraction of a set of three Gaussian-type functions to single Slater function used to be very popular (STO-3G):

$$\eta_i^{STO} = d_{i1}\eta_1^{GTF} + d_{i2}\eta_2^{GTF} + d_{i3}\eta_3^{GTF}$$

1.2.7—9. eq.

This means that the total number of integrals to be stored can be reduced by contraction as illustrated below:

$$\begin{aligned}
& \langle \eta_i^{STO} | \hat{h} | \eta_j^{STO} \rangle = \\
& \langle d_{i1} \eta_1^{GTF} + d_{i2} \eta_2^{GTF} + d_{i3} \eta_3^{GTF} | \hat{h} | d_{j1} \eta_1^{GTF} + d_{j2} \eta_2^{GTF} + d_{j3} \eta_3^{GTF} \rangle = \\
& d_{i1} d_{j1} \langle \eta_1^{GTF} | \hat{h} | \eta_1^{GTF} \rangle \\
& + d_{i2} d_{j1} \langle \eta_2^{GTF} | \hat{h} | \eta_1^{GTF} \rangle \\
& + d_{i3} d_{j1} \langle \eta_3^{GTF} | \hat{h} | \eta_1^{GTF} \rangle \\
& \vdots \\
& d_{i3} d_{j3} \langle \eta_3^{GTF} | \hat{h} | \eta_3^{GTF} \rangle = \\
& \sum_k^3 \sum_l^3 d_{ik} d_{jl} \langle \eta_k^{GTF} | \hat{h} | \eta_l^{GTF} \rangle
\end{aligned}$$

1.2.7—10. eq.

Such the contraction of basis set reduces the size of the Fock matrix which will be more manageable to find the solution.

In the above case 9 integrals have been evaluated, but only their sum total, that is, a single integral, is stored. The STO-3g basis function in Figure 1.2.7—5 is a contracted Gaussian consisting of three primitive Gaussian each of which has a contraction coefficient (0.4446, 0.5353 and 0.1543). Typically, an ab initio basis function consist of a set of primitive Gaussians bundled together with a set of contraction coefficient.

$$\eta(\text{Slater}) = \left(\frac{\zeta^3}{\pi} \right)^{\frac{1}{2}} e^{-\zeta r} = 0.7790 e^{-1.24r}$$

1.2.7—11. eq.

$$\eta(\text{STO} - 3G) = 0.4446 \left(\frac{2\alpha}{\pi} \right)^{\frac{3}{4}} e^{-\alpha r^2} + 0.5353 \left(\frac{2\alpha}{\pi} \right)^{\frac{3}{4}} e^{-\alpha r^2} + 0.1543 \left(\frac{2\alpha}{\pi} \right)^{\frac{3}{4}} e^{-\alpha r^2} =$$

$$0.0835 e^{-0.1689r^2} + 0.2678 e^{-0.6239r^2} + 0.2769 e^{-3.4253r^2}$$

1.2.7—12. eq.

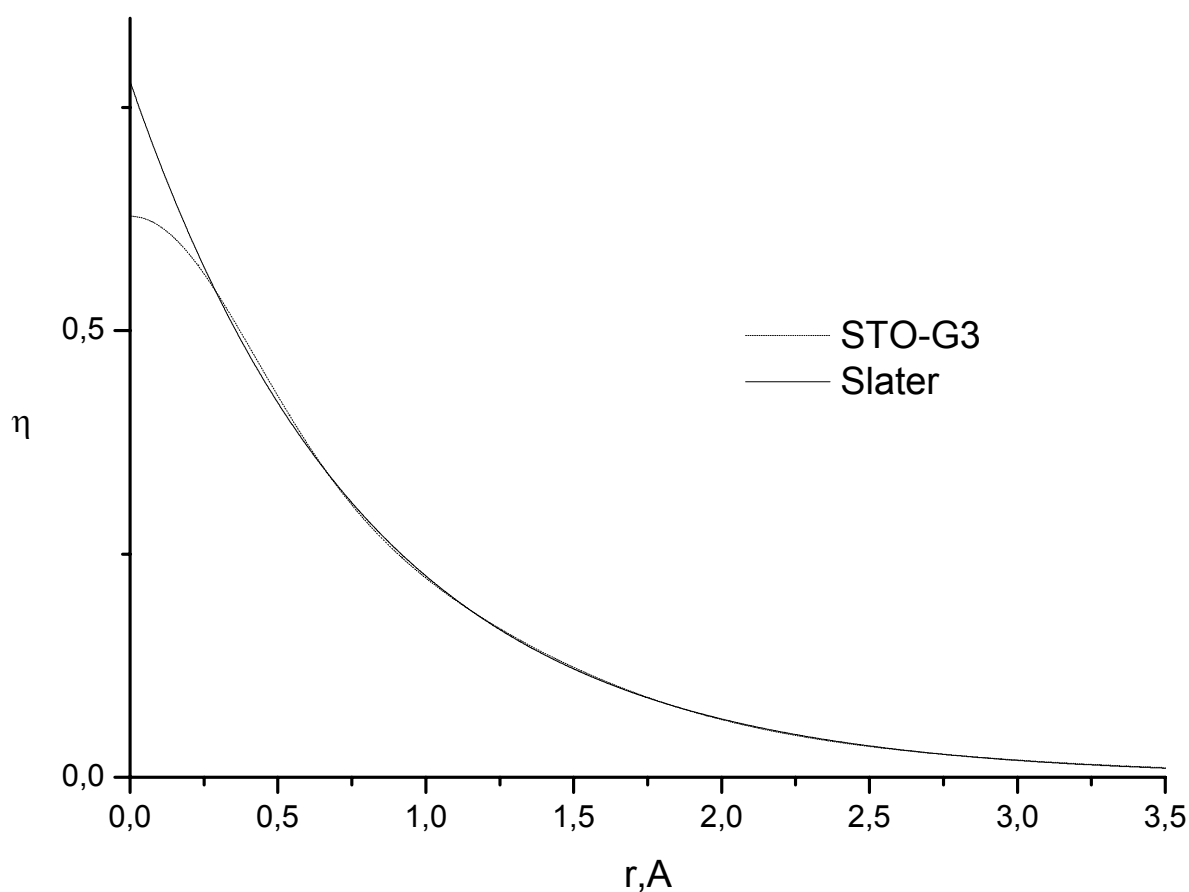


Figure 1.2.7—5 Comparison of Slater and STO-3G functions for hydrogen. The Slater function shown is the most appropriate one for hydrogen in a molecular environment.

The Basic advantage of Gaussian type functions is due to the fact that the product of any two Gaussian is also a Gaussian with its centre on a line between the centres of the two original Gaussian functions, as illustrated in Figure 1.2.7—6

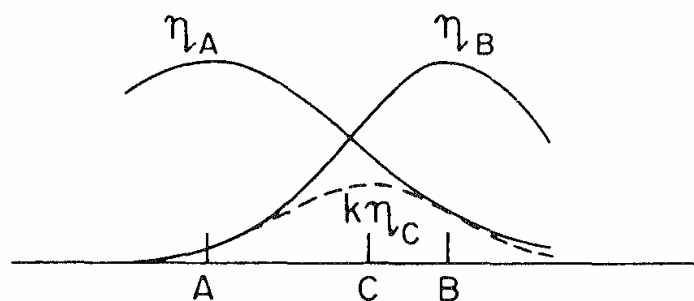


Figure 1.2.7—6 An illustration of the theorem that the product of two Gaussians is also a Gaussian.

Consequently, all integrals have explicit analytical expressions and may be evaluated rapidly. A similar theorem does not exist for STO. The principal disadvantages of GTF are their smooth behaviour (lack of cusp) at the nucleus and their too rapid (by e^{-r^2} rather than by e^{-r}) decrease at large distances (cf. Figure 1.2.7—7). This improper asymptotic behaviour requires the use of a larger number of GTF than STO for equivalent accuracy.

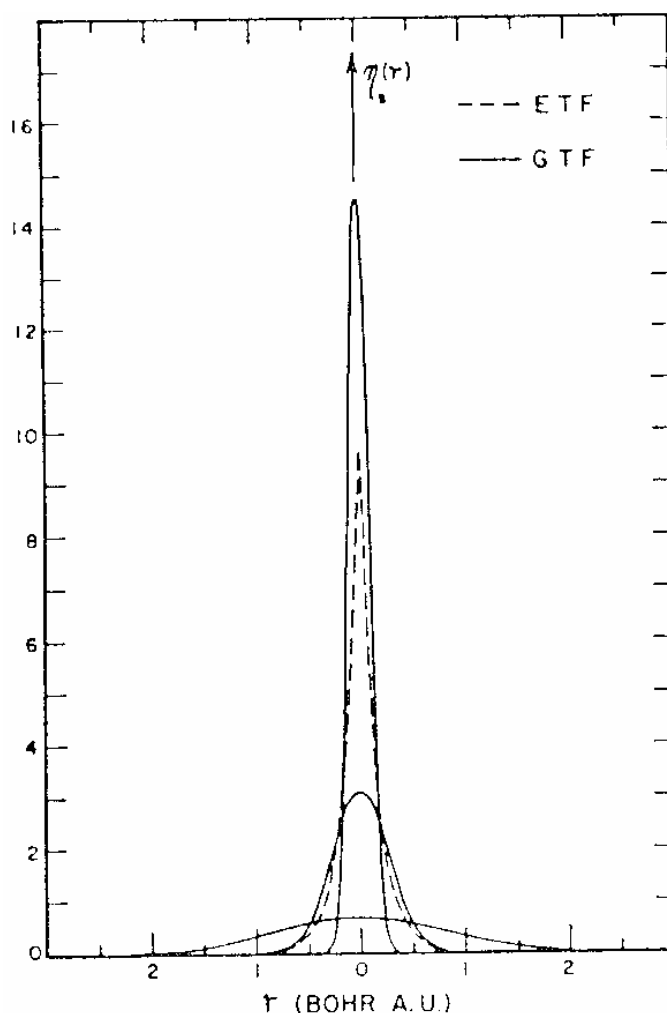


Figure 1.2.7—7 A comparison of a 1s-STO and 1s-GTF as a function of r.

However, the much greater speed per integral evaluation in terms of GTF as opposed to STO allows for this greater total number of integrals. Numerical values of the exponents and contraction coefficients are illustrated for selected basis set in the tables given below.

Table 1.2.7-2 XII-5 STO-3G and STO-4G minimal basis sets for H and C

	STO	STO-3G	STO-4G
		Exponents	
H	1s	3.42525D+00 6.23913D-01 1.68855D-01	8.02142D+00 1,46782D+00 4,07777D-01 1.35337D-01
C	1s	7.16168D+01 1.30451D+01 3.53051D+00 2,94125D+00 6,83483D-01 2.22290D-01	1.67716D+02 3,06899D+01 8,52600D+00 2.82970D+00 6.87386D+00 1,48804D+00 4,83819D-01 1.85818D-01
		Contraction Coefficients	
H,C	1s	1.54329D-01 5.35328D-01 4.44635D-01	5.67524D-02 2,60141D-01 5,32846D-01 2.91625D-01
C	2s	-9.99672D-02 3.99513D-01 7.00115D-01	-6.22071D-02 2.97680D-05 5.58855D-01 4.97767D-01
C	2p	1.55916D-01 6.07684D-01 3.91957D-01	4.36843D-02 2.86379D-01 5.83575D-01 2.46313D-01
* $n\alpha_s = n\alpha_p$ orbital exponent constraint			

Table 1.2.7-3 Split valence shell 4-31G Gaussian basis sets for H and C

	STO	Exponents 31G	Expansion Coefficients
H	1s	1.87311D+01	3.34946D-02
		2.82539D+00	2.34727D-01
		6.40121D-01	8.13757D-01
	1s'	1.61278D-01	1.00000D+00
	STO	Exponents 4-31G	Expansion Coefficients
C	1s	4.86967D+02	1.75983D-02
		7.33711D+01	1.22846D-01
		1.64135D+01	4.33782D-01
		4.34498D+00	5.61418D-01
			s
	2sp	8.67353D+00	-1.17489D-01
		2.09662D+00	-2.13994D-01
		6.04651D-01	1.17450D+00
			p
			6.40203D-02
		3.11203D-01	
		7.52748D-01	
		s or p	
2sp'	1.83558D-01	1.00000D+00	

1.2.8 Geometry optimization on PEF

There are three types of internal motions of molecules: stretch, bend and torsion as illustrated by Figure 1.2.8—1. The torsion is a periodic motion even if its periodicity low so that it repeats itself only after a 360° rotation. The bending motion usually governs a double-well potential. If the bonds that undergo bending motion are attached to N or O then the barrier to inversion is quite low (from a few to a few tens kcal/mole) but if it involves C then the inversion potential is very high as the inversion would pass through a planar carbon.

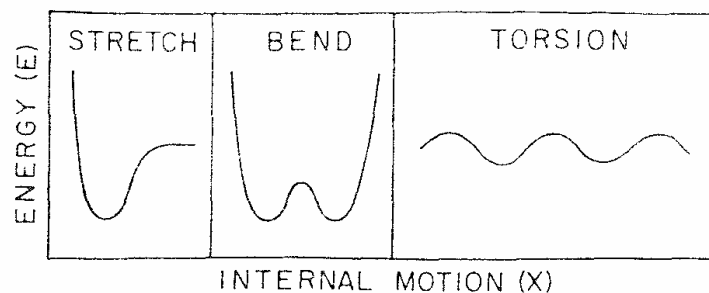


Figure 1.2.8—1 Three types of potentials associated with three internal modes of motion.

If we knew the potential curves that are characteristic to a given molecular system then we could determine the whereabouts of the minimum energy points that in fact correspond to the equilibrium geometry. Unfortunately, these potential functions are not only unknown but they change from molecule to molecule and from bond to bond. However, we do know that they may be approximated, near any of their minima by same quadratic potential, as illustrated by Figure 1.2.8—2, since quadratic and true functions osculate at the minimum.

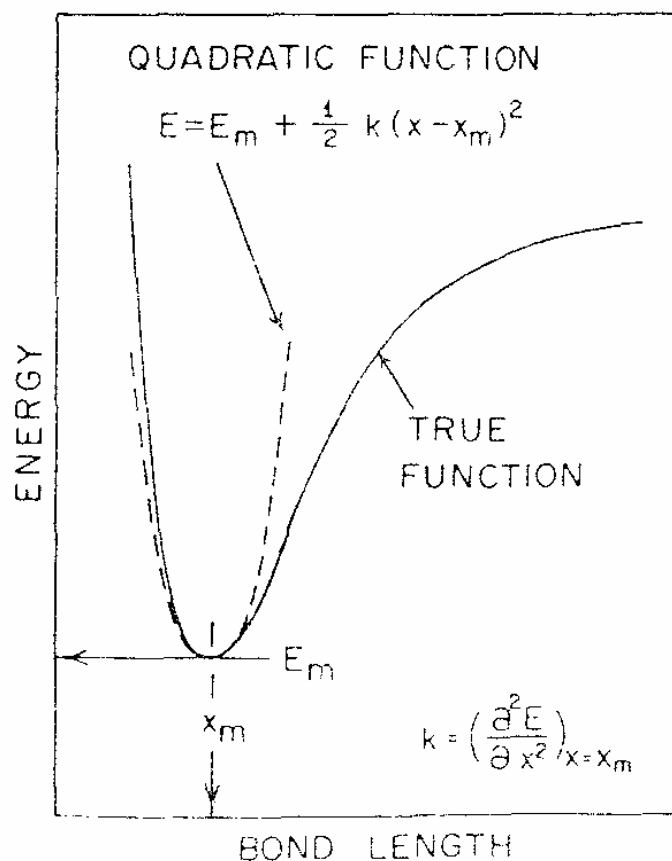


Figure 1.2.8—2 The osculation of the quadratic and the true functions at the minimum as illustrated for a bond stretch.

The quadratic function is the traditional Hooke's law:

$$E = E_m + \frac{1}{2}G(x - x_m)^2$$

1.2.8—1. eq.

where G , the second derivative of E with respect to x , is the force constant usually denoted by k

$$\frac{d^2 E}{dx^2} = G \equiv k$$

1.2.8—2. eq.

The minimum energy point is denoted by E_m and x_m in equation 1.2.8—3. eq. as well as in Figure 1.2.8—2. For a multi-dimensional problem the generalized Hooke's law may be written as follows:

$$E = E_m + \frac{1}{2}(\underline{x} - \underline{x}^m)^T G(\underline{x} - \underline{x}^m)$$

1.2.8—3. eq.

$$E = E_m + \frac{1}{2} \begin{bmatrix} [x_1 - x_1^m] & [x_2 - x_2^m] & \dots \end{bmatrix} \begin{bmatrix} G_{11} & G_{12} & \dots \\ \vdots & G_{22} & \dots \\ \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} [x_1 - x_1^m] \\ [x_2 - x_2^m] \\ \vdots \end{bmatrix}$$

1.2.8—4. eq.

where is the $(x-x^m)$ displacement vector and G , the Hessian matrix, collects all diagonal and off-diagonal or interaction force constants

$$\begin{bmatrix} \frac{\partial^2 E}{\partial x_1^2} & \frac{\partial^2 E}{\partial x_1 \partial x_2} & \dots \\ \vdots & \frac{\partial^2 E}{\partial x_2^2} & \dots \\ \vdots & \vdots & \ddots \end{bmatrix} = \begin{bmatrix} G_{11} & G_{12} & \dots \\ \vdots & G_{22} & \dots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

1.2.8—5. eq.

For a two dimensional case the situation is illustrated in terms of a schematic energy contour diagram in Figure 1.2.8—3. This figure also shows that the internal coordinates $\{x_i\}$ are not the same as the normal coordinates $\{y_i\}$.

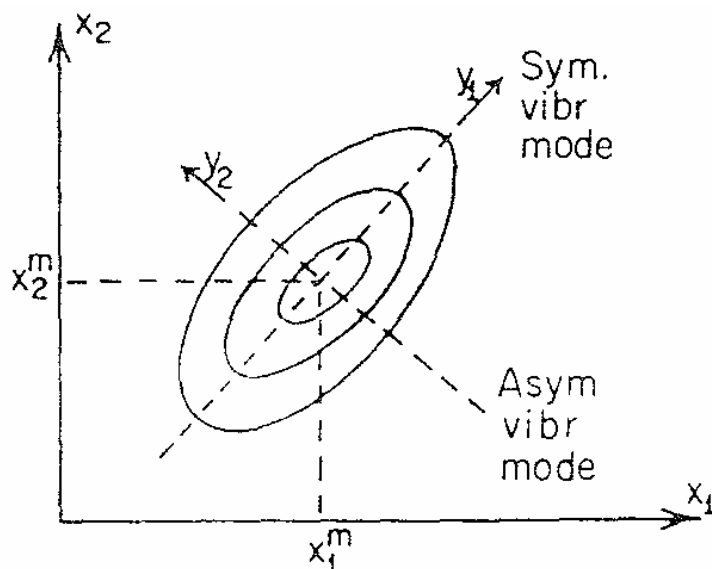


Figure 1.2.8—3 A schematic illustration of a two-dimensional potential energy surface: $E = E(x_1, x_2)$ in terms of energy levels contours. Note that the normal coordinates $\{y_i\}$ are different from the internal coordinates $\{x_i\}$

Molecular geometry optimization involves the finding of the minimum energy (E_m) point or in other words locating x_1^m, x_2^m, \dots . This can most effectively be done by evaluating the gradient vector:

$$\langle g | \equiv g = \left(\frac{\partial E}{\partial x_1}, \frac{\partial E}{\partial x_2}, \dots \right)$$

1.2.8—6. eq.

and searching for the point where the gradient vector is a zero vector

$$\langle g | = (0, 0, \dots)$$

1.2.8—7. eq.

since at the minimum the gradient vanishes. This is illustrated for the minima and higher order critical points for one- and two-dimensions in Figure 1.2.8—5 and Figure 1.2.8—6 respectively. Note that force (F) exerted on the atomic nuclei is the negative of the energy gradient vector:

$$F = -\text{grad}E = -\langle g \rangle$$

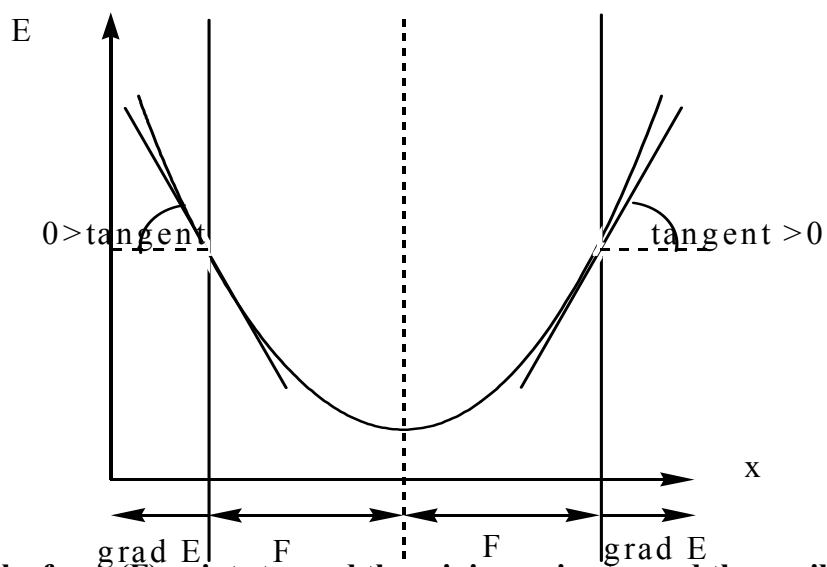


Figure 1.2.8—4 The force (F) points toward the minimum i.e. toward the equilibrium position while the gradient ($\text{grad} E$) points away from the minimum. Thus $F = -\text{grad} E$

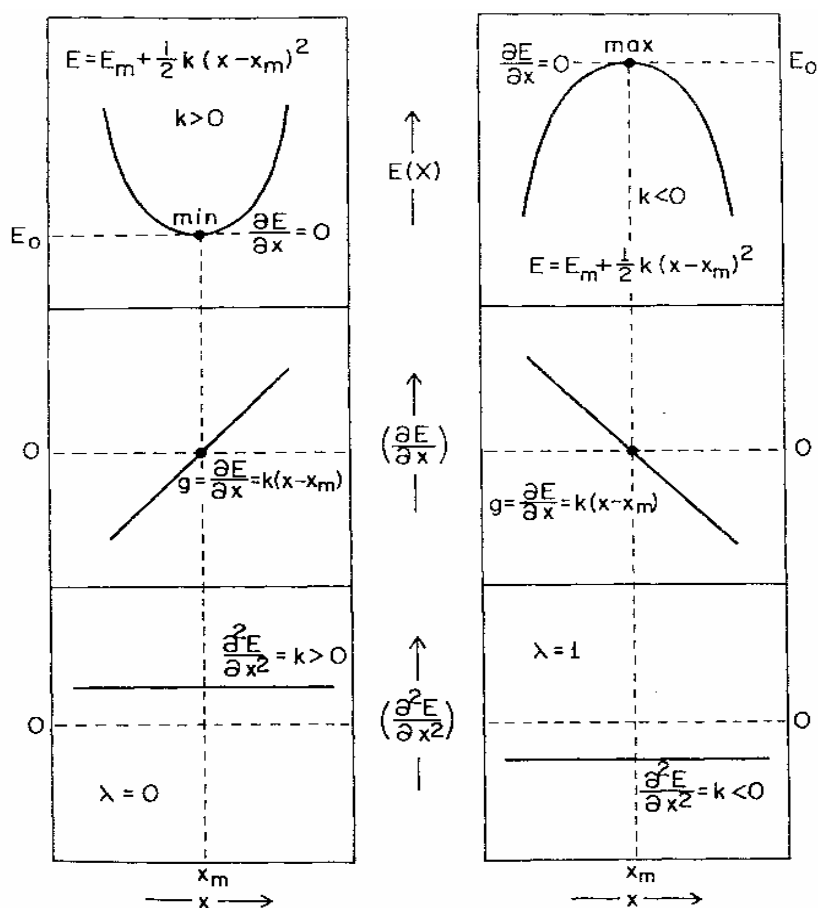


Figure 1.2.8—5 Characteristics of minima and maxima of a potential energy curve, (Note that g is the gradient, k is the force constant of the potential energy function and λ is the index of the critical point in question.)

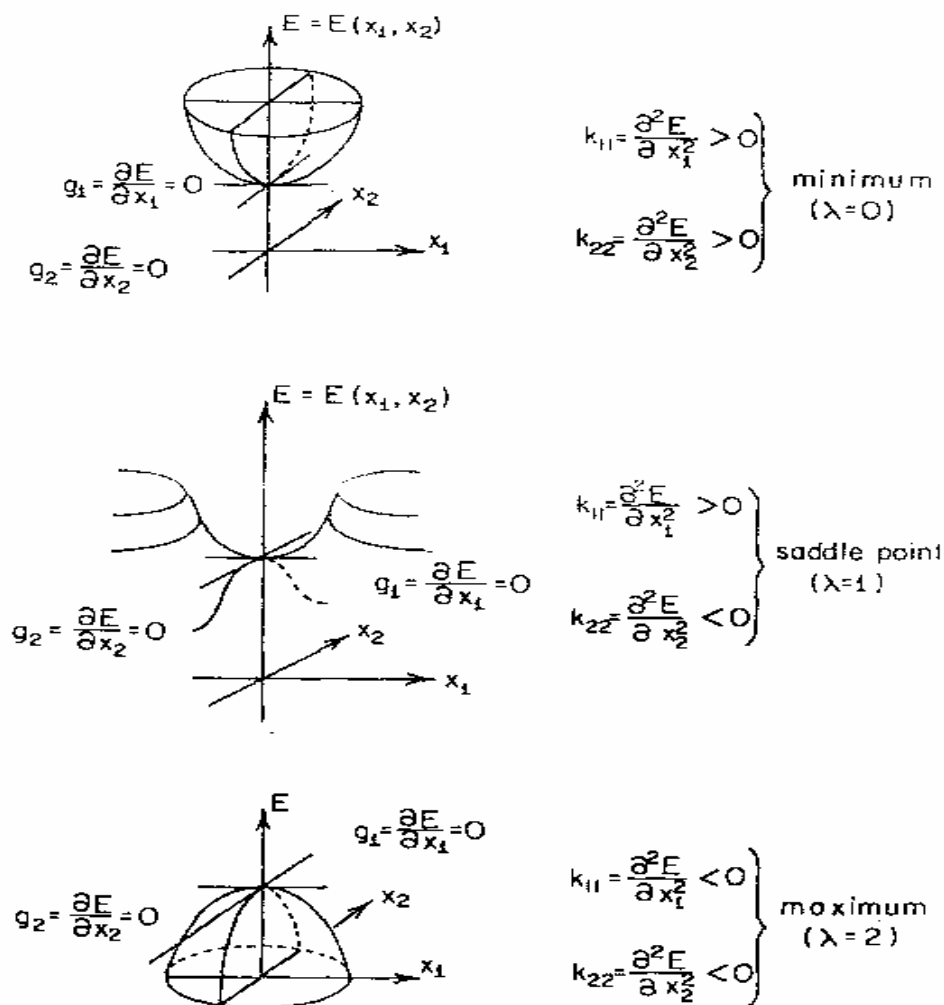


Figure 1.2.8—6 Three types of critical points of a potential energy surface and their characteristics in terms of second partial derivatives. (Note that the index (λ) of a critical point is the number of negative second derivatives.)

Even though the gradient vector provides a powerful tool in locating the minimum we always carry out the multi-dimensional optimization as a sequence of one-dimensional optimizations. The one-dimensional optimization, however, involves a quadratic or parabolic interpretation.

Parabolic interpolation

Basic procedure for one-dimension

First of all, we must choose an initial point: x_0 and a distance: d which lead to three equidistant points $(x_0 + d)$, x_0 , $(x_0 - d)$. Secondly we need to evaluate $E(x)$ at the three equidistance points leading to three energy values E_+ , E_0 and E_- . Thirdly, we locate the minimum energy (E_m) point which is located at x_m . If x_m falls within the range $x_0 - d$ to $x_0 + d$, and if d is sufficiently small then quit. Of course we want to avoid extrapolation. Alternatively we should get enough (3) equidistant points to bracket $x_m^{(1)}$,

and predetermine $x_m^{(2)}$. This may involve using some of the original points again if $x_m^{(1)}$ was close to x_0 .

Minimum of a parabola

A general equation of a parabola may be written as:

$$E(x) = ax^2 + bx + c$$

1.2.8—8. eq.

and the function may be evaluated at three equidistant points $(x_0 - d)$, x_0 , $(x_0 + d)$

$$E_+ = E(x_0 + d) = a(x_0 + d)^2 + b(x_0 + d) + c$$

1.2.8—9. eq.

$$E_0 = E(x_0) = ax_0^2 + bx_0 + c$$

1.2.8—10. eq.

$$E_- = E(x_0 - d) = a(x_0 - d)^2 + b(x_0 - d) + c$$

1.2.8—11. eq.

Expanding 1.2.8—9. eq., 1.2.8—10. eq., and 1.2.8—11. eq. gives:

$$E_+ = ax_0^2 + 2adx_0 + ad^2 + bx_0 + bd + c$$

1.2.8—12. eq.

$$E_0 = ax_0^2 + bx_0 + c$$

1.2.8—13. eq.

$$E_- = ax_0^2 - 2adx_0 + ad^2 + bx_0 - bd + c$$

1.2.8—14. eq.

Adding 1.2.8—12. eq. and 1.2.8—14. eq. together we obtain:

$$E_- + E_+ = 2ax_0^2 + 2ad^2 + 2bx_0 + 2c$$

1.2.8—15. eq.

Subtracting from 1.2.8—15. eq. twice of equation 1.2.8—13. eq. leads to 1.2.8—16. eq.

$$E_- + E_+ - 2E_0 = 2ad^2$$

1.2.8—16. eq.

This may be solved for the unknown parameters: a

$$a = \frac{E_- + E_+ - 2E_0}{2d^2}$$

1.2.8—17. eq.

We may obtain parameter b, as 1.2.8—19. eq., after subtracting 1.2.8—14. eq., from 1.2.8—12. eq.

$$E_+ - E_- = 4adx_0 + 2bd$$

1.2.8—18. eq.

$$b = \frac{E_+ - E_-}{2d} - 2ax_0$$

1.2.8—19. eq.

Finally parameter c may be obtained from equation 1.2.8—10. eq.

$$c = E_0 - ax_0^2 - bx_0$$

1.2.8—20. eq.

The location of minimum, where $x = x_m$ may be achieved by- setting the gradient equal to zero

$$g \equiv \frac{\partial E}{\partial x_1} = 2ax_m + b = 0$$

1.2.8—21. eq.

$$x_m = -\frac{b}{2a}$$

1.2.8—22. eq.

After substituting to 1.2.8—22. eq. the values of a and b from equations 1.2.8—17. eq. and 1.2.8—19. eq. respectively we obtain:

$$x_m = x_0 - d \left(\frac{E_+ - E_-}{2(E_+ + E_- - 2E_0)} \right)$$

1.2.8—23. eq.

It is convenient to introduce the following notations:

$$t = E_+ - E_-$$

1.2.8—24. eq.

$$s = E_+ - E_- - 2E_0$$

1.2.8—25. eq.

$$x_m = x_0 - d \left(\frac{t}{2s} \right)$$

1.2.8—26. eq.

This last equation resulted from 1.2.8—23. eq.. The three parameters a (from 1.2.8—17. eq.), b (from 1.2.8—19. eq.) and c (from 1.2.8—20. eq.) may be rewritten in terms of t and s.

$$a = \frac{s}{2d^2}$$

1.2.8—27. eq.

$$b = \frac{t}{2d} - \frac{s}{d^2} x_0$$

1.2.8—28. eq.

$$c = E_0 - \frac{s}{2d^2} x_0^2 - \frac{t}{2d} x_0 + \frac{s}{d^2} x_0^2 = E_0 + \frac{s}{2d^2} x_0^2 - \frac{t}{2d} x_0$$

1.2.8—29. eq.

The energy at x_m may be written according to 1.2.8—8. eq. as:

$$E(x_m) = ax_m^2 + bx_m + c$$

1.2.8—30. eq.

using 1.2.8—27. eq., 1.2.8—28. eq. and 1.2.8—29. eq. the above expression may be rewritten as follows:

$$\begin{aligned} E(x_m) &= \frac{s}{2d^2} x_m^2 + \frac{t}{2d} x_m - \frac{s}{d^2} x_0 x_m + E_0 + \frac{s}{2d^2} x_0^2 - \frac{t}{2d} x_0 = \\ &= \frac{s}{2d^2} (x_m^2 - 2x_0 x_m + x_0^2) + \frac{t}{2d} (x_m - x_0) + E_0 = \\ &= \frac{s}{2d^2} (x_m - x_0)^2 + \frac{t}{2d} (x_m - x_0) + E_0 \end{aligned}$$

1.2.8—31. eq.

Finally, the deviation $(x_m - x_0)$ may be expressed from equation 1.2.8—26. eq.

$$(x_m - x_0) = -\frac{dt}{2s}$$

1.2.8—32. eq.

Substituting 1.2.8—32. eq. Into 1.2.8—31. eq. we obtain:

$$E(x_m) = \frac{s}{2d^2} \left(\frac{d^2 t^2}{4s^2} \right) + \frac{t}{2d} \left(\frac{-dt}{2s} \right) + E_0$$

1.2.8—33. eq.

which simplifies to 1.2.8—34. eq.

$$E(x_m) = E_0 - \frac{t^2}{8s}$$

1.2.8—34. eq.

This predicted $E(x_m)$ is the minimum energy of the quadratic function. This value can be compared to $E(x_m)$ calculated from the function. Note also that an estimate of the force constant, the parameter a , is readily obtained.

$$\frac{\partial^2 E}{\partial x^2} = a \equiv k \equiv G$$

1.2.8—35. eq.

As it is clear from Figure 1.2.8—5

for a minimum $a > 0$

for a maximum $a < 0$

Procedure for N-dimensions

First of all, from the start position, one should find the minimum along the first direction (x_1). Subsequently, one must cycle through the remaining independent variables, (x_2, x_3, \dots, x_n), fitting parabolas to sets of 3 points in each direction in turn. To reach M_1 , the best point after the 1st cycle through all the parameters. An example with 2 independent variables, $E = E(x_1, x_2)$ is graphically illustrated by Figure 1.2.8—7. If m_i and m_{i-1} are sufficiently close together then quit. Alternatively the use M_i as the new start position, and continue. As the optimization progresses, it is desirable to cut down the step size d

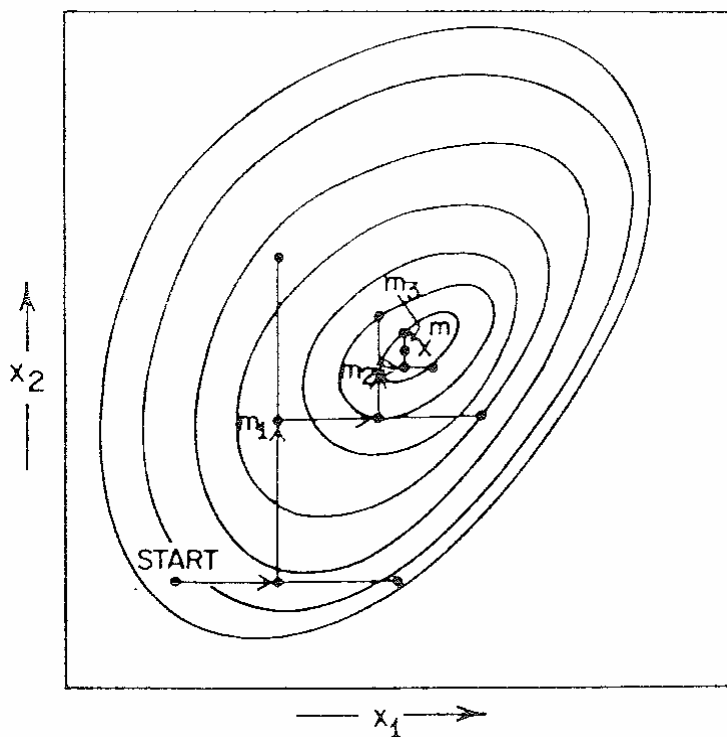


Figure 1.2.8—7 Typical parabolic interpolation optimization of $E(x_1, x_2)$ (Note that after the start position, m_i : end of cycle i , m : true minimum.)

While this type of sequential optimization along, one internal coordinate before the next one, is a working method but it is rather pedestrian. Also, if the variables are coupled, as they very often are, the method is very inefficient as illustrated by Figure 1.2.8—8

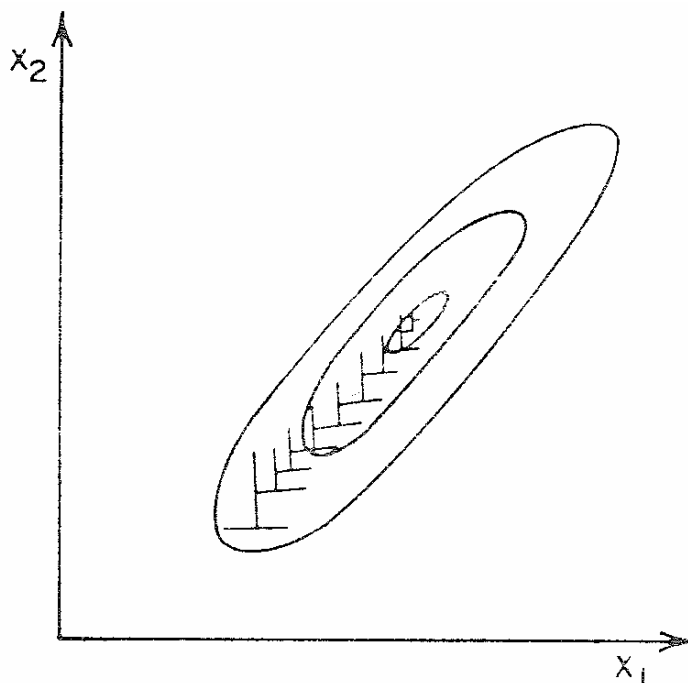


Figure 1.2.8—8 A schematic illustration of the inefficiency of sequential optimization method along one of the independent coordinates at a time is shown for a strongly coupled system.

$$v_m \sim \frac{1}{2\pi} \sqrt{\frac{\pm a^2 m}{\mu}}$$

1.2.8—38. eq.

Therefore, the first λ eigenvalues of 1.2.8—36. eq. correspond to imaginary frequency. For $m > \lambda$ we obtain 1.2.8—38. eq.

$$v_m \sim \frac{1}{2\pi} \sqrt{\frac{-a^2 m}{\mu}} = \frac{a_m}{2\pi\sqrt{\mu}} \sqrt{-1} = \frac{a_m}{2\pi\sqrt{\mu}} i$$

1.2.8—39. eq.

At the end of the optimization it is advisable to check the order (λ) of the critical point. If $\lambda = 0$ then we can rest assured that the critical point in our optimization we have converged to is indeed a minimum.

General introduction to gradient methods of optimization

Methods for unconstrained optimization of a variables (n dimensions) are designed to produce the answer in a finite number of steps when the function, $f(x)$, is quadratic. It is then hoped that the method will be efficient on more general functions, but especially those with slowly varying second derivatives.

For a general function the notation $f(x)$ is used rather than the $E(x)$ specified before. This is done to emphasize that the method is applicable to any routine differentiable function.

At the minimum or very close to it the functions has a quadratic form, in addition to the constant $f(x_m)$, without a linear term. This function may be written, as follows:

$$f(x) = f(x^m) + \frac{1}{2}(x - x^m)^{\dagger} G(x - x^m)$$

1.2.8—40. eq.

However, some distance away from the minimum at point x^k the inclusion of a linear term is advisable (c.f. Figure 1.2.8—9) as specified by the next equation

$$f_k(x) = \frac{1}{2}(x - x^k)^\dagger G(x - x^k)(x - x^k)^\dagger g_k + f(x^k)$$

1.2.8—41. eq.

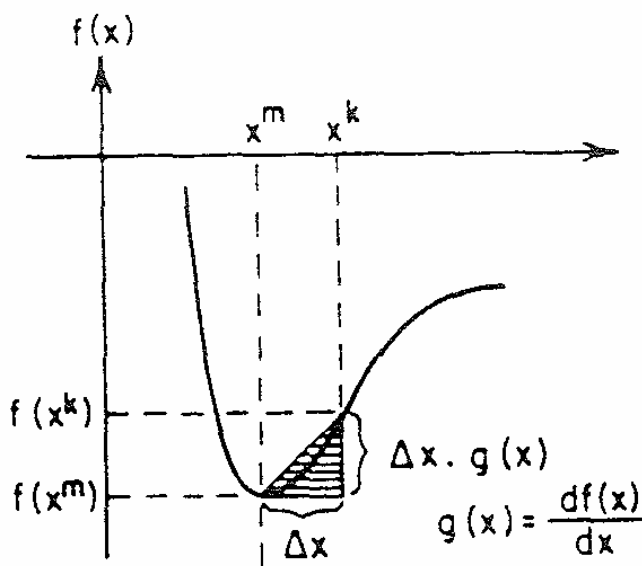


Figure 1.2.8—9 A linear segment that takes the minimum energy point from x^m to x^k .

The components of the gradient vector $g(x)$ are the first partial derivatives of $f(x)$ written in a fashion that is analogous to 1.2.8—6. eq.:

$$\text{grad } f(x) = \nabla f(x) = g = \langle g | = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots \right)$$

1.2.8—42. eq.

In general, the objective function $f(x)$ is

$$f_k(x) = f_m + \sum_{i=1}^n a_i x_i + \frac{1}{2} \sum_{i,j} G_{ij} x_i x_j$$

1.2.8—43. eq.

or in Dirac notation

$$f_k(x) = f_m + \langle x | a \rangle + \frac{1}{2} \langle x | G | x \rangle$$

1.2.8—44. eq.

Where

$$G_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

1.2.8—45. eq.

and the Hessian matrix" is constructed according to 1.2.8—5. eq. The gradient of $f(x)$, $g(x)$ is:

$$g_i(x) = a_i + \sum_j G_{ij} x_j$$

1.2.8—46. eq.

or

$$|g \rangle = |a \rangle + |G|x \rangle$$

1.2.8—47. eq.

where $g_i(x)$ is the i 'th component of $|g\rangle$. At the extremum of f ,

$$|g \rangle = 0$$

1.2.8—48. eq.

Letting $|x_m\rangle$ be the coordinates of the extremum. From 1.2.8—47. eq. we may express $|a\rangle$

$$|a \rangle + |G|x_m \rangle = 0$$

1.2.8—49. eq.

Therefore

$$|x_m \rangle = -|G^{-1}|a \rangle$$

1.2.8—50. eq.

From 1.2.8—47. eq., substitute for $|a\rangle = |g\rangle - |G|x\rangle$

$$\begin{aligned} |x_m \rangle &= -G^{-1} [|g \rangle + |G|x \rangle] \\ &= -G^{-1} |g \rangle + G^{-1} G |x \rangle \\ &= -G^{-1} |g \rangle + |x \rangle \end{aligned}$$

1.2.8—51. eq.

$$|x_m \rangle = |x \rangle - G^{-1} |g \rangle$$

1.2.8—52. eq.

Which is the Newton (or Newton-Raphson) equation. Given exact G and $|g\rangle$, the minimum can be found in one step for a quadratic function. Methods which do not use an exact G often use an approximate matrix H (a positive definite symmetric matrix) such that

$$HG = 1$$

1.2.8—53. eq.

In these cases, the step size equation 1.2.8—52. eq. is usually written:

$$|x^{(i+1)}\rangle = |x^{(i)}\rangle - \lambda_i H |g^{(i)}\rangle$$

1.2.8—54. eq.

where λ_i is to be determined. Note that $|x^{(i+1)}\rangle$ will not necessarily be the extremum desired for any value of λ_i , as H is not exact. Besides if f is not quadratic, 1.2.8—51. eq. is not valid anyways so 1.2.8—54. eq. is always used in practice. Such methods, based on 1.2.8—54. eq., are called quasi-Newton methods.

Steepest descents

A very simple optimization method called steepest descents arises from 1.2.8—54. eq. by assuming \mathbf{H} is the unit matrix, so that the search direction is always $-|g^{(i)}\rangle$. A line search is carried out along the direction $-|g^{(i)}\rangle$ to obtain λ_i . This method however, poses some problems. Subsequent search directions tend to be linearly dependent, so only a small subspace of the total space is explored.

Directions with large components of $|g^{(i)}\rangle$ are always favoured whereas progress can sometimes be made by searching in orthogonal directions to reach a part of the surface that would allow better progress. Also, the method converges very slowly near the minimum as $g(x)$ is getting smaller in that vicinity.

Summary

The connection between the Hartree-Fock-SCF computations of molecular orbitals and the gradient optimization of the geometry.

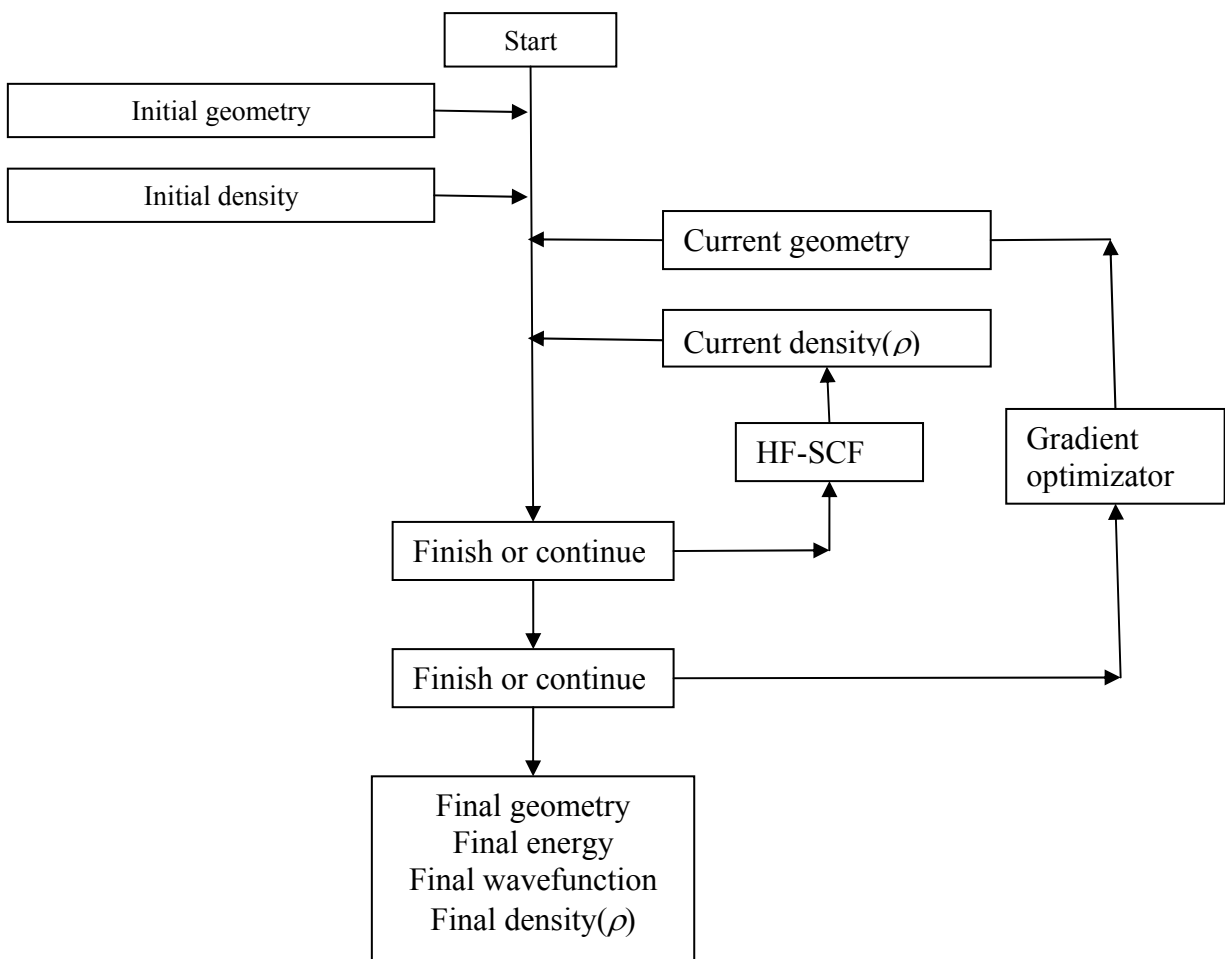


Figure 1.2.8—10 Flowchart for Non-empirical (ab-initio) and semi-empirical (AM1 or PM3) Molecular orbital (MO) computations with geometry optimization