Introduction to Gaussian

<u>Stationary points on the potential energy surfaces (PES):</u>

A *potential energy surface* (PES) is a $E = f(\mathbf{x})$ function, where E is the *total electronic energy* of a specific molecule with N atoms, $\mathbf{x} = (x_1, x_2, \ldots, x_M)^{\mathrm{T}}$ is the vector of *geometric parameters* (bond lengths, bond angles, and torsion angles) related to this molecule, and M is its *number of degree of freedoom*, defined as

$$M = \begin{cases} 3N - 5, & \text{if the molecule is linear,} \\ 3N - 6, & \text{otherwise.} \end{cases}$$

If the $f(\mathbf{x})$ function is twice partially differentiable with respect to all its variables at a given $\mathbf{x}_0 = (x_{1,0}, x_{2,0}, \ldots, x_{M,0})^{\mathrm{T}}$ point, then $f(\mathbf{x})$ can be approximated around $\mathbf{x}_0$ with a *second-degree Taylor polynomial*:

$$f(\mathbf{x}) \approx f(\mathbf{x}_0) + \sum_{i=1}^{M} \left(\frac{\partial f}{\partial x_i} \right)_{\mathbf{x}=\mathbf{x}_0} (x_i - x_{i,0}) + \frac{1}{2} \sum_{i=1}^{M} \sum_{j=1}^{M} \left(\frac{\partial^2 f}{\partial x_i \partial x_j} \right)_{\mathbf{x}=\mathbf{x}_0} (x_i - x_{i,0})(x_j - x_{j,0}).$$

Let us introduce the *gradient vector*, $\mathbf{g} = (g_1, g_2, \ldots, g_M)^{\mathrm{T}}$, and the *Hessian matrix*, $\mathbf{H} = \{h_{ij}\}$, as follows:

$$g_i = \left(\frac{\partial f}{\partial x_i} \right)_{\mathbf{x}=\mathbf{x}_0},$$

$$h_{ij} = \left(\frac{\partial^2 f}{\partial x_i \partial x_j} \right)_{\mathbf{x}=\mathbf{x}_0}.$$

Using $\mathbf{g}$ and $\mathbf{H}$, the former quadratic approximation can be written in the following way:

$$f(\mathbf{x}) \approx f(\mathbf{x}_0) + \mathbf{g}^{\mathrm{T}}(\mathbf{x} - \mathbf{x}_0) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^{\mathrm{T}} \mathbf{H}(\mathbf{x} - \mathbf{x}_0)$$

It can be derived that since $\mathbf{H}$ is a real symmetric matrix, all its eigenvalues are real.

If $\mathbf{g} = \mathbf{0}$, then $\mathbf{x}_0$ is called a *stationary point* of $f(\mathbf{x})$. The $\mathbf{x}_0$ stationary point is said to be *Kth order saddle point* if $\mathbf{H}$ has exaclty K negative eigenvalues. Note that the zeroth and the Mth order saddle points are referenced as *minima* and *maxima*, respectively.

In quantum chemistry, we are mainly interested in the physico-chemical characterization of the minima and the first-order saddle points. <u>Minima</u> of the $f(\mathbf{x})$ function represent the <u>stable structures</u> of the examined molecule, while the <u>first-order saddle points</u> correspond to the *<u>transition states</u>* (TS) among the stable geometries.
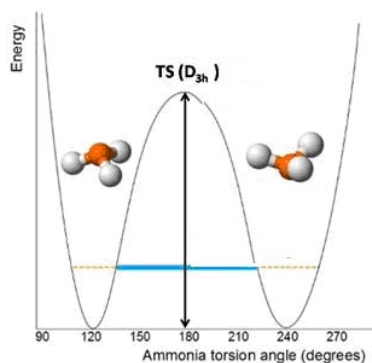
Figure 1: *Potential energy surface of the umbrella motion in ammonia*

As to the total electronic energy associated to a given geometry of a specific molecule, this quantity is obtained by solving the Schrödinger equation describing the microscipic behavior of the molecule:

$$\hat{\mathcal{H}}\Psi = E\Psi,$$

where $\hat{\mathcal{H}}$ is the Hamiltonian operator of the molecular system, and Ψ is the wavefunction corresponding to the E eigenvalue. Note that E can be decomposed in the following form:

$$E = E_{nn} + E_{ne} + E_{ee},$$

where E_{nn} is the *nuclear repulsion energy*, E_{ne} is the *nuclear-electron attraction energy*, and E_{ee} is *electron repulsion energy*.

During solving the Schrödinger equation, the following approximations are usually applied: (a) Ψ is expanded on a finite basis set, and (b) some terms of $\hat{\mathcal{H}}$ are neglected to simplify the complexity of the underlaying mathematical problem. The individual approximation schemes introduced for the reduction of the $\hat{\mathcal{H}}$ operator are the *level of theory* of the related quantum-chemical computation. During the course of this seminar, the so-called *Pople-type bases* (*e.g.*, 3-21G and 6-31G*) and the *Hartree-Fock* (HF) level of theory is utilized to estimate the total electronic energy of the investigated molecule.

If we would like to determine a particular $\bar{\mathbf{x}}$ stationary point on the PES, (a) we have to find an initial point (*initial geometry*), $\mathbf{x}_0$, that is adequately close to $\bar{\mathbf{x}}$, and (b) we must apply an optimization method, such as the well-known Newton–Raphson (NR) technique:

$$\mathbf{x} \approx \mathbf{x}_0 - \mathbf{H}^{-1}\mathbf{g},$$

where $\mathbf{x}$ is a better approximation of $\bar{\mathbf{x}}$ than $\mathbf{x}_0$. When $\mathbf{x}$ is available, we have to replace $\mathbf{x}_0$ with $\mathbf{x}$ and provide an even better approximation with the former equation until $\mathbf{g} \approx \mathbf{0}$ and $\mathbf{x} \approx \mathbf{x}_0$ are reached. In practice, when the computation of $\mathbf{H}$ is too expensive, the NR scheme is substituted with quasi-Newton methods (*e.g.*, BFGS and DFP) in the optimization process.

Basic Linux commands:

1. make a directory: `mkdir` `dirname`

2. go to a directory („change" directory): `cd` `dirname`

3. go back from a directory to the directory before that: `cd` `..`

4. list the files in a folder: `ls`

5. list the files with a specific extension: `ls *.ext`

6. copy a file to another: `cp` `oldfile newfile`

7. remove a file with confirmation: `rm` `filename`

8. remove a file without confirmation (in a forced way): `rm -f filename`

9. remove a directory with confirmation: `rm -r dirname`

10. remove a directory without confirmation: (in a forced way): `rm -rf dirname`

11. move (and rename) files and directories: `mv` `oldname newname`

12. complete a filename or directory name: `tabulator`

13. search for lines with a specific string: `grep "string" filename`

14. search for lines with a specific string (exact match): `grep -w "string" filename`

15. print *n* lines before a lines with a specific string: `grep -Bn "string" filename`

16. print *n* lines after a lines with a specific string: `grep -An "string" filename`

17. count the matches related to a specific string: `grep -c "string" filename`

18. redirect the input/output of a command: `</>` `filename`

19. append the output of a command to a file: `>>` `outfile`

20. further commands: `sed, awk, cat, cut, tr, head, tail, time, top, ctrl + c,` `ctrl + z, bg, fg, history,` *etc.*

19. inspect the manual of a specific command: `man` `comname`

Basic commands of the `vi` editor:

1. open a file: `vi` `filename`

2. edit („insert") mode: `i`

3. command mode: `Esc`

4. copy a line: `yy`

5. paste a line: `p`

6. cut (delete) a line: `dd`

7. forward search for a string: `/string`

8. backward search for a string: `?string`

9. go to the next found: `n`

10. go to the previous found: `shift + n`

11. copy/cut lines between the *n*th and *m*th lines: `:n,m y/d`

12. undo: `u`

13. redo: `ctrl + r`

14. go to the beginning/end of a line: `home/end`

15. insert a new line before a given line: `o`

16. insert a new line after a given line: `shift + o`

17. save the file content: `:w`

18. quit without editing the file: `:q`

19. quit without saving the file content: `:q!`

20. quit with saving the file content: `:wq`

Gaussian package:

A. Structure of the input (`*.com`) files:

1st line: `%nprocs=`*n* [*n*: `number of cores;` **optional**]

2nd line: `%mem=`*m* [*m*: `memory size;` **optional**]

3rd line (parameters for the computation):

$$\#method/basis \; \text{opt=(calcfc,ts) freq}$$

where

(a) `method` is the level of theory (*e.g.*, `method=hf` denotes the Hartree–Fock method),

(b) `basis` is the basis set applied (*e.g.*, `basis=3-21g`, `basis=6-31g*`),

(c) `opt` is used to perform geometry optimization

~ if we search for a transition state (TS) structure, we have to type `opt=(calcfc,ts)`,

~ in the case that a stable structure is pursued, `=(calcfc,ts)` should be omitted, and

(d) `freq` is the keyword for the frequency analysis (`NImag` is an important parameter in the Gaussian output: [i] if it is `NImag=0`, then a minimum structure is found, [ii] if `NImag=1`, then a TS is reached, and [iii] if `NImag>1`, then a higher-order saddle point perceived)

4th line: blank line

5th line: title (arbitrary text)

6th line: blank line

7th line: `charge multiplicity`

(a) `charge` is the charge of the molecule

(b) `multiplicity` is <u>the number of unpaired electrons plus one</u>

From 8th line to (8+N-1)th line: list of coordinates related to the N atoms of the molecule (using Z-matrix or Cartesian coordinates)

(a) in the case of Cartesian coordinates:

$$s(1)\ x(1)\ y(1)\ z(1)$$
$$s(2)\ x(2)\ y(2)\ z(2)$$
$$\vdots$$
$$s(N)\ x(N)\ y(N)\ z(N)$$

where

~ `s(1), s(2), ..., s(N)` are the symbols of the atoms (H, O, N, etc.),

~ `x(1), x(2), ..., x(N)` are the x coordinates of the atoms given in Å,

~ `y(1), y(2), ..., y(N)` are the y coordinates of the atoms given in Å, and

~ `z(1), z(2), ..., z(N)` are the z coordinates of the atoms given in Å

(b) in the case of Z-matrix:

```
s(1)
s(2) i₂ r[s(2),s(i₂)]
s(3) i₃ r[s(3),s(i₃)] j₃ a[s(3),s(i₃),s(j₃)]
s(4) i₄ r[s(4),s(i₄)] j₄ a[s(4),s(i₄),s(j₄)] k₃ d[s(4),s(i₄),s(j₄),s(k₄)]
s(5) i₅ r[s(5),s(i₅)] j₅ a[s(5),s(i₅),s(j₅)] k₅ d[s(5),s(i₅),s(j₅),s(k₅)]
```
$$\vdots$$
```
s(N) iₙ r[s(N),s(iₙ)] jₙ a[s(N),s(iₙ),s(jₙ)] kₙ d[s(N),s(iₙ),s(jₙ),s(kₙ)]
```

where

~ `s(1), s(2), ..., s(N)` are the *symbols* of the atoms (H, O, N, etc.),

~ the (i_m, j_m, k_m) triplets form the *connectivity list* of the molecule

~ `r[s(m),s(iₘ)]` is the *distance* of atoms `s(m)` and `s(iₘ)` in Å

~ `a[s(m),s(iₘ),s(jₘ)]` is the *angle* of atoms `s(m)`, `s(iₘ)`, and `s(jₘ)` in deg, and

~ `d[s(m),s(iₘ),s(jₘ),s(kₘ)]` is the *angle of two planes* P_1={`s(m)`, `s(iₘ)`, `s(jₘ)`} and P_2={`s(iₘ)`, `s(jₘ)`, `s(kₘ)`} in deg

(8+N)th (last) line: blank line

B. Examples for Gaussian input files:

1. Water molecule (*Z-matrix*):

```
%nproc=1
%mem=20MB
#opt freq hf/3-21G

H2O test

0 1
O
H 1 0.95
H 2 0.95 1 104.5

```

2. Ammonia molecule (Cartesian):

```
%nproc=1
%mem=20MB
#opt freq hf/3-21G

NH3(cart) test

0 1
N   0.000000   0.000000   0.116489
H   0.000000   0.939731  -0.271808
H   0.813831  -0.469865  -0.271808
H  -0.813831  -0.469865  -0.271808

```

3. Ammonia molecule (*Z-matrix*):

```
%nproc=1
%mem=20MB
#opt freq hf/3-21G

NH3 test

0 1
N
H 1 1.0
H 2 1.0 1 110.0
H 3 1.0 1 110.0 2 120.0

```

4. Ammonia-TS molecule (*Z*-matrix):

```
%nproc=1
%mem=20MB
#opt=(calcfc,ts) freq hf/3-21G

NH3-TS test

0 1
N
H 1 1.0
H 2 1.0 1 120.0
H 3 1.0 1 120.0 2 180.0

```

C. Performing Gaussian computations: `g03 filename.com`

D. Useful information in the Gaussian output (`*.log`*) files*:

1. Hartree–Fock (self-consistent field, SCF) energy: `grep "SCF D" filename.log | tail -1 | cut -d "=" -f2 | cut -d "A" -f1 | tr -d " "`

2. optimized parameters (distances and angles): `awk "/Stat/,/Grad/" filename.log | sed "/Grad/q"`

3. nuclear repulsion energy: `grep "nuclear rep" filename.log`

4. dipole moment: `grep –A1 "Dipole mom" filename.log`

5. frequencies of the normal modes (cm^{-1}) in one column: `grep "Frequencies" filename.log | awk "{$1=""; $2=""; print}" | xargs | tr -s '[:blank:]' '\n'`

6. number of imaginary frequencies (`NImag`): `cat filename.log | tr -d "\n" | grep –Po "(?<=NImag\=).*(?=\s*)" | sed "s@^[^0-9]*\([0-9]\+\).*@\1@"`

7. list of the optimization steps: `grep –A4 "Converged" filename.log`

8. number of optimization steps: `grep –c "Converged" filename.log`

9. point-group symmetry: `grep "Full point group" filename.log | tail -1 | awk "{print $4}"`

E. Molecular graphics using GaussView: `gview filename.log`

1. Calculate the activation energies of the $XCY \to XYC$ isomerization reactions, where $X = H, F, Cl, Br, I$ and $Y = N, P, As$. Make plots about the minima and the TS structures. Observe and explain the trends in the geometric parameters and the activation energies.

2. Calculate the activation energies for the umbrella motion of the YX_3 and H_3Z^+ molecules, where X and Y are the same as in the previous exercise, and $Z = O, S, Se, Te$. Make plots about the minima and the TS structures. Observe and explain the trends in the geometric parameters and the activation energies.