3 Large Basis Sets, Dissociation Energy and Geometry Optimisation

Albeit deceptively small and simple, the H_2 molecule with its two protons and two electrons is already too big to derive an analytic solution of the Schrödinger equation. However, one may construct a special basis set that is arbitrarily close to the complete basis set limit, and then run a post-Hartree Fock calculation. Using such techniques, it was found that, for the H_2 molecule, $E_{exact} = -1.174474$ a.u.

In this set of exercises, you will compare the influence of basis sets on the energy of H_2 at equilibrium distance, and you will then go on to compute the dissociation energy of H_2 in Hartree Fock. Subsequently, you will determine the equilibrium geometry of a species at the example of water and visualise the progress of the optimisation in the molecular viewing software *Molden*.

3.1 Effects of Basis Set Size: The Molecular Case

In exercise 2.4, you have already computed the equilibrium energy of H_2 at the Hartree-Fock level using the 6-31G basis set. You will now be employing the slightly larger 6-311G basis, as well as Dunning's correlation-consistent aug-cc-pV6Z basis. Correlation-consistent basis sets were defined such that systematic improvement over total energies and molecular properties is possible. Create the following input (blank line at the end!) that differs slightly from the ones you have written so far:

```
%NProcShared=2
%Mem=1GB
#P UHF/6-311G SP Guess=Mix

H2 equilibrium, UHF, 6-311G

0 1
H 0.0000 0.0000 0.0000
H 0.0000 0.0000 0.7414

```

Save this file under an appropriate name using the extension `.com` to specify that it is an input. Create a second input using the aug-cc-pV6Z basis (and, of course, use a different filename).

The new line in the input corresponding to `%Mem=` specifies the amount of memory that Gaussian will use instead of the default value, which is (only) 256 MB. Compared to the default settings you used so far, this will enable for a considerable speed-up especially for the large correlation-consistent basis set. Submit both jobs (as a reminder, the general syntax is

```
g09 < input_filename.com > output_filename.log &
```

where you denote the output by the extension `.log`.)

a) Complete the following table (use the 6-31G results from series 2):

Basis set	Number of basis functions	Total Energy [a.u.]
6-31G		
6-311G		
aug-cc-pVTZ		
Exact value	(analytical)	-1.174474

b) Plot the energy after each SCF cycle against the number of cycles. You may extract this information from the output file using `grep`. The information on the energy within an SCF cycle is given three lines after the line `Cycle`. By adding the option `-A 3` to `grep`, `grep` will not only output the line containing `Cycle`, but also the three following lines. The pipe `|` then redirects the output to the following command, which is again a `grep` instruction, extracting just the line containing the energies from the three initial lines.

```
grep -A 3 'Cycle' filename | grep 'E='
```

which will display all the lines corresponding to `SCF Done` on the screen. You may alternatively redirect the output to a file by appending `»` after the command:

```
grep -A 3 'Cycle' filename | grep 'E=' >> newfilename
```

which will create (or append to) a file called `newfilename`. You may use any plotting program you feel at ease with; if you would like to learn about elegant command line plotting with `gnuplot`, please let one of the assistants know (it will not take longer than using Excel or LibreOfficeCalc).

c) How does the basis set size influence the SCF convergence behaviour?

d) What is the meaning of the line `Delta-E` that is printed during the SCF cycle, and why is the last value not equal to zero? (Bonus: What could the significance of `Rises=F` be?)

e) Use your knowledge on the Hartree-Fock SCF procedure and the Gaussian output to derive a flow chart of how a wavefunction optimisation is carried out. Use Ψ_{guess} for the first trial wavefunction (`Guess` in the Gaussian output), Ψ_{opt} for the optimised wavefunction (`SCF Done` in the Gaussian output) and think of appropriate naming schemes for the intermediate steps. Clarify whether, during an iteration, the Fock operator acts on the previous Ψ or on the new one. Include checks for convergence at the appropriate points.

f) Try to find out why we are using `Guess=Mix` for the UHF runs. You may use the Gaussian manual on gaussian.com (or Google...)

3.2 Recording a Dissociation Curve for H$_2$: RHF vs. UHF

Calculate the energy of H$_2$ at the following internuclear distances at both the RHF/6-31G and UHF/6-31G level of theory:

Distance ($\mathbf{R}$) [Å]	0.40	0.60	0.80	1.00	1.50	2.50	4.00
$E_{H_2}^{6-31G}(\mathbf{R})$ UHF							
$E_{H_2}^{6-31G}(\mathbf{R})$ RHF							

Do not forget about `Guess=Mix` in the UHF inputs.

a) Plot the interaction energy:

$$E_{inter}^{6-31G}(\mathbf{R}) = E_{H_2}^{6-31G}(\mathbf{R}) - 2 \cdot E_H^{6-31G} \tag{80}$$

against the interatomic distance $\mathbf{R}$. For E_H^{6-31G}, you may use the energy of the isolated hydrogen atom that you obtained in the last set of exercises. Convert all energies to kcal mol^{-1}, and plot UHF and RHF energies on the same graph. You may quickly extract the converged SCF energy using

```
grep 'SCF Done' outputfile
```

b) Explain the physical origin of the difference between the two dissociation curves. Discuss how important this difference is by comparing to the energy that is normally required to break a bond.

3.3 Geometry Optimisation of a Water Molecule

Until now, you have been calculating the wavefunction for a given arrangement of the nuclei. Often, however, you may lack the neccessary information about the equilibrium geometry of a species, and you may be interested in the unknown equilibrium bond lengths and angles. You will thus have to find the optimal parameter in a geometry optimisation. The optimised geometry may be defined as one where the spatial arrangement of the nuclei is such that *the corresponding wavefunction is lowest in energy.*
Consider the simple case of the hydrogen molecule, where the only parameter that can be optimised is the distance between the nuclei. The energy of every possible configuration is given by a two-dimensional *potential energy surface* (PES): Starting from an initial guess for the geometry, one may follow the curvature of the PES down to the minimal energy, which will correspond to the equilibrium geometry. In a practical geometry optimisation, this will be done in discrete steps of a certain size, as illustrated in figure 1. At each iteration, the energy of the corresponding geometry is calculated in a SCF calculation (*cf.* figure 2). Then, the gradient with respect to the nuclear coordinates is determined, and the nuclei are moved to a new position on the PES that lies in the direction of the steepest gradient. The geometry optimisation converges once the minimum is attained.
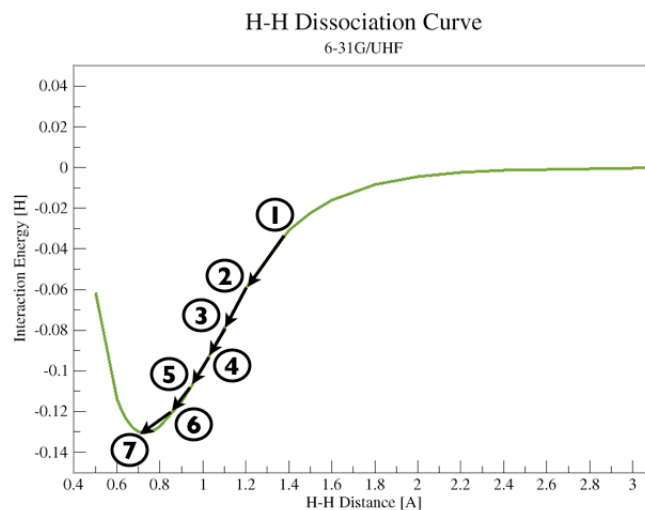
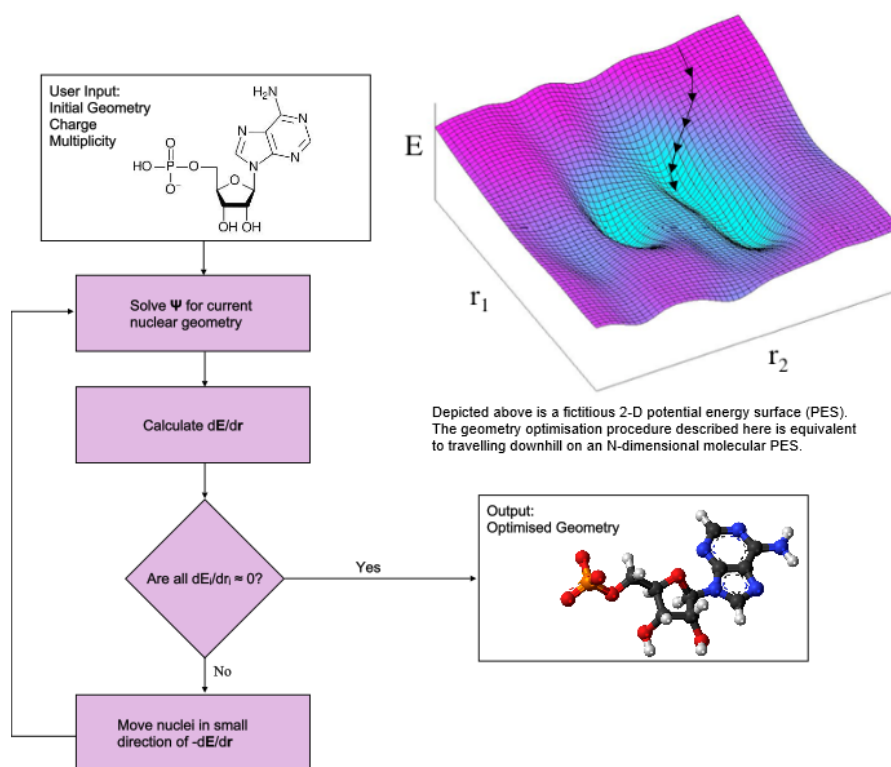Figure 1: 1D PES of the hydrogen molecule showing the discrete steps taken in a geometry optimisation.

Figure 2: Schematic representation of a geometry optimisation.

In practice, this implies that the forces on the nuclei fall below a certain treshold.

Larger systems will have more nuclear degrees of freedom than simple bimolecular species, and in these cases the PES becomes many-dimensional; each degree of freedom will correspond to one dimension. Geometry optimisations may not converge on very complex potential energy surfaces. If the program is fed with an input that resides in a region of the PES that has a difficult topology, the algorithm may not find a minimum, and it may be better to restart from a different initial geometry. Furthermore, a PES may exhibit multiple minima, and the geometry optimisation algorithm may only reach the next local minimum, but not neccessarily the global minimum (*cf.* figure 2).

Geometry Optimisation in Gaussian

You will carry out a geometry optimisation on a bad guess for H_2O with an unrealistic H-O-H angle of 90°. During the optimisation, Gaussian will move downhill on the potential energy surface until it finds a minimum. Create the following input file:

```
%NProcShared=2
%Mem=1GB
#P UHF/6-31G Opt Symmetry=None Guess=Mix

H2O geometry optimisation

0 1
O 0.0000 0.0000 0.0000
H 1.4000 0.0000 0.0000
H 0.0000 1.4000 0.0000

```

where the command `SP` for a single-point calculation was replaced by `Opt` for a geometry optimisation, and `Symmetry=None` is used as to not artificially impose a symmetry operation that may be found in the input. Create a second input file where you use the 6-31+G* basis and submit both of them to Gaussian.

Changes in the Output

The output for a geometry optimisation will be considerably longer than the previous ones: For each trial geometry, Gaussian will report the full SCF cycle data and energies. The last geometry iteration will be followed by - `Stationary point found`, after which Gaussian will print the new geometry both in the coordinate frame that you specified in the input (`input orientation`) and in the coordinate system that it used to internally optimise the geometry (`standard orientation`). Then follow all the properties of the optimised system, just like for the single point calculations.

   a) Open one of the output files in vi and scroll through one geometry optimisation step. The SCF cycle information that you already know (`SCF Done`) is followed by a

part called `GradGradGradGrad`. What could this additional information be related to? Match the statements (`SCF Done` and `GradGradGrad`) with the corresponding boxes in the flow chart of figure 2.

b) Go to the end of the geometry optimisation by searching for the string `Stationary point found`. What is the significance of the block `Derivative Info`?

Visualisation Software: Molden

For anyone not gifted with an exceptional xyz-to-3D converter in their brain (the existence of which the author doubts), keeping track of a geometry optimisation by simply looking at the changes in the molecular coordinates may be quantitatively near to impossible, especially for large molecules. Luckily, there exist plenty of molecular visualisation software. One rather straightforward example is *Molden*, which has the ability to read in Gaussian output. Open it using

```
molden outputfile.log
```

and you will see a graphical interface appear on the screen. You will later on learn how to use a more powerful and aesthetic visualisation tool, but Molden is the most pragmatic choice for the moment. In the *Molden Control* panel, you have different options to click on. You may first experiment a bit with *Draw Mode* (for instance, choose `Solid`, then `Ball & Stick`). The geometry that is displayed on your screen corresponds to the Gaussian input geometry. If you want to see how it changed during the geometry optimisation, click on `Next` in the menu *Select Point*, until `Last point` appears in the *Status line*. This last geometry corresponds to the optimised molecular structure. Clicking on the button *Geom. conv.* in *Convergence* will open a new window containing information on the energy, nuclear forces and number of SCF steps for each point along the geometry optimisation. You can view the corresponding geometry by clicking on a point in the plot.

If you wish to retrieve additional information, such as changes in bond lengths and angles, the `ZMAT Editor` is a reasonable choice for medium-sized species. The z-matrix is an alternative representation of the topology of the system, where the position of an atom is described in terms of a bond length with respect to its (nearest) neighbour, and the angle is given with respect to other atoms further away. If you open the z-matrix of water, you will see that it simply consists of the two O-H bond lengths and the H-O-H bond angle. By clicking on a point in the `Geometry Convergence` window, the z-matrix will be updated to the current coordinates accordingly.

c) For each step along the geometry optimisation, give the O-H bond length and the H-O-H bond angle along with the corresponding SCF energy. You may use Molden for everything, or you could choose to extract the SCF energies from the output using `grep` as in previous exercises. Do this for both basis sets, 6-31G and 6-31+G*.

d) Are there significant differences in bond lengths and angles between the two basis sets? Comment on what you observe. How do these quantities evolve as a function of geometry optimisation steps?

e) Is there a trend in the energies as the geometry optimisation progresses? If yes, explain.

f) Explain any qualitative difference between the behaviour of the forces and the total energy during the optimisation.