Introduction to Q-Chem and QCMagic

A brief tutorial

© Thom Group Cambridge, 2019

Maintainer: Bang C. Huynh (cbh31[at]cam.ac.uk)

Contents

Before we begin, make sure that you have a copy of Q-Chem 5.2 compiled with parallelisation on your workstation and that you are sufficiently familiar with Unix shell commands to navigate around and perform various day-to-day tasks. This tutorial assumes that you already know how to carry out such tasks and focuses entirely on Q-Chem and QCMagic.

1 Single-Point Calculations

1.1 Restricted Hartree–Fock (RHF)

We begin with a simple Hartree–Fock (HF) calculation for the simplest closed-shell system, a He atom, to familiarise ourselves with the syntax of Q-Chem. Every Q-Chem run requires an input file in which the molecular configuration and all relevant calculation parameters are specified. Listing 1 gives an example of such an input file.

Listing 1: codesnippets/singlepoint/He.RHF.STO-3G.inp

```
$molecule
    0 1
    He  0   0   0
$end
$rem
    BASIS STO-3G
    EXCHANGE hf
    CORRELATION none
    UNRESTRICTED false
    SCF_GUESS CORE
    SCF_ALGORITHM DIIS
    SCF_CONVERGENCE 13
    SCF_MAX_CYCLES 1000
    MOM_START 1
    SCF_PRINT 1
    PRINT_ORBITALS true
$end
```

We observe the following in Listing 1:

1. There are two sections in this file: `$molecule` specifies the molecular configuration of the system and `$rem` specifies the parameters to be used in the calculation.

2. The system of interest consists of a single He nucleus located at the origin (line 3).

3. The overall charge of the system is zero (first number on line 2), so there are two electrons in total.

4. The "spin multiplicity" of the system is supposed to be 1 (second number on line 2). However, one cannot in general constrain the value of S in a conventional HF calculation. Therefore, the "multiplicity" specified here is actually equal to $2M_S + 1$ instead of $2S + 1$. This system thus has $M_S = 0$, but whether or

not S is a "good" quantum number for the system is dependent upon the HF method and the solutions obtained.

5. A self-consistent-field (SCF) restricted HF (RHF) calculation will be run on this system in the STO-3G basis set using the DIIS converging algorithm starting from the guess molecular orbital (MO) coefficients obtained by diagonalising the one-electron Hamiltonian matrix (lines 6–11).

6. The convergence criterion is when the maximum DIIS error becomes smaller than 1×10^{-13} (line 12). This is a very tight criterion.

7. The maximum number of SCF cycles is 1000 (line 13). This is more than enough here because convergence can be reached very quickly in this particular example.

8. During each SCF cycle, the code maximises overlap with the orbitals from the previous cycle. This behaviour starts from cycle 1 (line 14) where the code maximises overlap with the guess orbitals.

9. After each cycle, the code prints out only the minimal and useful output together with the component breakdown of SCF electronic energy (line 15).

10. After convergence, the code prints out the occupied orbitals plus five virtual orbitals (line 16).

To carry out the above Q-CHEM calculation, copy Listing 1 from the `git` repository to a location of your choice and then execute the following command in a shell:

```
$   qchem -nt 6 He.RHF.STO-3G.inp He.RHF.STO-3G.out
```

which instructs Q-CHEM to run a calculation with the input parameters specified in `He.RHF.STO-3G.inp` using six OpenMP threads and write the outputs to `He.RHF.STO-3G.out`.

Task 1.1

Examine the output file and determine the following:

    (a) the DIIS error at convergence;
    (b) the SCF energy of the converged solution;
    (c) the contributing components of the SCF energy;
    (d) the MO energies; and
    (e) the MO coefficients in terms of the atomic orbital (AO) basis functions.

What is the value of $\langle \hat{S}^2 \rangle$ of the converged solution and why? Assign a term symbol to this solution.

1.2 Unrestricted Hartree–Fock (UHF)

Duplicate `He.RHF.STO-3G.inp` and give it a different name such as `He.UHF.STO-3G.inp`. Then, change the value of `UNRESTRICTED` on line 9 to `true` as shown in Listing 2 and execute

```
$ qchem -nt 6 He.UHF.STO-3G.inp He.UHF.STO-3G.out
```

Listing 2: `codesnippets/singlepoint/He.UHF.STO-3G.inp`

```
9    UNRESTRICTED true
```

Task 1.2

Examine `He.UHF.STO-3G.out` and compare the following with those in the RHF case:

   (a) the SCF energy of the converged solution;

   (b) the contributing components of the SCF energy;

   (c) the value of $\langle \hat{S}^2 \rangle$;

   (d) the structure of the MO coefficients;

   (e) the MO energies; and

   (f) the MO coefficients in terms of the atomic orbital (AO) basis functions.

Is this converged solution identical to the RHF one? If so, explain the origin of the identicality.

1.3 Effects of Basis Sets

1.3.1 SCF Energy

Choosing an appropriate basis set is an important aspect of electronic structure calculation: a suitable basis set not only gives sensible and sufficiently accurate results but also saves computation time and storage space. To investigate basis set effects on the solutions obtained so far, duplicate `He.RHF.STO-3G.inp` and `He.UHF.STO-3G.inp` to `He.RHF.6-31GSTAR.inp` and `He.UHF.6-31GSTAR.inp`, respectively, then change the value of `BASIS` on line 6 to `6-31G*` as shown in Listing 3. When ready, execute

```
$ qchem -nt 6 He.RHF.6-31GSTAR.inp He.RHF.6-31GSTAR.out
$ qchem -nt 6 He.UHF.6-31GSTAR.inp He.UHF.6-31GSTAR.out
```

```
6      BASIS 6-31G*
```

Task 1.3

Examine `He.RHF.6-31GSTAR.out` and `He.UHF.6-31GSTAR.out` and answer the following:

(a) Are the RHF and UHF solutions located in 6-31G* identical to each other?

(b) Can these solutions be assigned to the same term as those in STO-3G? If so, are they a better description of this term and why?

But how exactly does 6-31G* differ from STO-3G for a He atom? One possible way to determine this is to ask Q-CHEM to print out the basis set definition. This is achieved by setting `PRINT_GENERAL_BASIS` to `true` in the `$rem` section. Another way is to look it up on `https://www.basissetexchange.org/`.

Shown in Listing 4 is the 6-31G* basis set definition for a He atom as printed by Q-CHEM. This tells us that, in 6-31G*, each electron in a He atom is described by two *s* shells. The radial part of the first shell is constructed by contracting three primitive Gaussian functions whereas the radial part of the second shell is constructed by contracting only a single primitive Gaussian function. The Gaussian primitive exponents and the contraction coefficients are given by the first and second columns respectively within each shell. Since both of these shells are of angular momentum *s*, their angular parts are simply 1.

Listing 4: `codesnippets/singlepoint/He.RHF.6-31GSTAR.withbasis.out`

```
122  Basis set in general basis input format:
123  ----------------------------------------------------------------
124  $basis
125  He    0
126  S    3    1.000000
127     3.84216340E+01    2.37660000E-02
128     5.77803000E+00    1.54679000E-01
129     1.24177400E+00    4.69630000E-01
130  S    1    1.000000
131     2.97964000E-01    1.00000000E+00
132  ****
133  $end
134  ----------------------------------------------------------------
```

Task 1.4

   (a) In 6-31G*, how many basis functions are used to describe each orbital of He?

   (b) How does STO-3G differ from 6-31G* for He?

    6-31G** is another small basis set for He in which each electron is described by two *s* shells and one *p* shell. Run an RHF calculation in this basis and answer the following:

   (c) How many basis functions are used to describe each orbital of He in this basis set? Why is this number different from the number of shells in this basis set?

   (d) Examine the orbital coefficient matrix carefully. Why does it have a very particular block-diagonal form?

   (e) Does 6-31G** offer any improvement over 6-31G* for the ground RHF solution of He? Why or why not?

1.3.2 Numbers of Solutions

It should not be surprising that a larger basis set inevitably introduces more solutions to the non-linear HF equations. Before proceeding, think about the three basis sets STO-3G, 6-31G*, and 6-31G** for He and answer the questions in the following Task.

Task 1.5

   (a) How many RHF solutions exist for He in STO-3G and 6-31G*?

   (b) How many *linearly independent* RHF solutions exist for He in 6-31G**? How does your answer change if linearly dependent RHF solutions are also counted?

Having thought through the number of possible RHF solutions in 6-31G* for He, we will now attempt to locate another RHF solution in this basis. One approach to doing this is to start from the RHF solution located in Task 1.3 which has one occupied orbital and one virtual orbital (identical for both spin spaces), then ask Q-CHEM to "swap" these orbitals so that the virtual orbital is now considered occupied and *vice versa*. This is then use as the initial guess for another SCF run, hoping to converge to another RHF solution whose occupied orbital is similar (but not necessarily identical) to the virtual one in the previous solution. To this end, duplicate `He.RHF.6-31GSTAR.inp` to `He.RHF.6-31GSTAR.excited.inp` and **append** to it the lines shown in Listing 5. Then execute

```
$ qchem -nt 6 He.RHF.6-31GSTAR.excited.inp He.RHF.6-31GSTAR.excited.out
```

and observe that there are now two calls to Q-Chem. Examine He.RHF.6-31GSTAR.excited.out and notice how it now contains the outputs for two jobs. The solution converged to in the first job should be identical to that obtained in Task 1.3. The solution found in the second job is a new RHF solution with properties as expected from the discussion earlier.

Listing 5: codesnippets/singlepoint/excited/He.RHF.6-31GSTAR.excited.inp

```
18    @@@
19    $molecule
20        read
21    $end
22    $rem
23        BASIS 6-31G*
24        EXCHANGE hf
25        CORRELATION none
26        UNRESTRICTED false
27        SCF_GUESS read
28        SCF_ALGORITHM DIIS
29        SCF_CONVERGENCE 13
30        SCF_MAX_CYCLES 1000
31        MOM_START 1
32        SCF_PRINT 1
33        PRINT_ORBITALS true
34    $end
35    $occupied
36        2
37        2
38    $end
```

So how does this work exactly? We note that He.RHF.6-31GSTAR.excited.inp actually contains inputs for two jobs. These two inputs are separated by @@@ (line 18 in Listing 5). In the first job, Q-Chem converges to the first RHF solution. But after this job is done, Q-Chem does not remove the scratch files just yet but will keep them for the next job. These scratch files are binary files with very specific formats that Q-Chem uses to efficiently store and retrieve data such as MO coefficients during its run. By specifying the inputs of the two jobs in the same file, the second job also has access to the scratch files produced by the first job. Therefore, the option read in the $molecule section (line 20) instructs Q-Chem to read from the scratch files the molecular configuration of the first job and use it for the second job. Then, on line 27, the option

read of `SCF_GUESS` tells Q-Chem to use the set of MOs from the scratch files, which should correspond to the MOs of the RHF solution located in the first job, as the initial guess. But more crucially, there is now a new section called `$occupied` (lines 35–38) through which we can control which of the read-in MOs should be occupied. In our example, we know that the RHF solution from the first job has in each spin space two MOs, one occupied and one virtual, with index 1 and 2 respectively. Since we wish to use the virtual orbital as guess for this job, we "occupy" it by specifying its index for both α and β spin spaces in lines 36 and 37 respectively.

Task 1.6

Study `He.RHF.6-31GSTAR.excited.out` and answer the following:

  (a) How do the occupied MOs of the two solutions differ? Are the occupied MOs of the second solution identical to the virtual MOs of the first solution? Why or why not?
  (b) How do the SCF energies of the two solutions differ? Is the energy difference between the two solutions the same as twice the gap between the occupied and virtual MOs in each solution?
  (c) What term symbol can be assigned to the second solution?

Use a similar method to locate all linearly independent RHF solutions of He in 6-31G** and answer the following:

  (d) Why are some of these solutions degenerate?
  (e) Is it possible to assign term symbols to all of these solutions? Why or why not?
  (f) Can this method be used to obtain other solutions that are linear combinations of the degenerate solutions found above?

Even though this method allows us to find multiple SCF solutions, it has several drawbacks. It is clearly inefficient because we have to specify the occupation pattern for each solution we seek to find and will quickly become unmanageable if there are more electrons and if a larger basis set is used. This of course assumes that we already know what solutions to expect, so the solutions we find are only as good as our guesses. In addition, as you should have realised in Task 1.6, this method can only swap occupied and virtual orbitals exactly but does not allow for an arbitrary mixing between them according to

$$\chi = \chi^{\text{occ}} \cos \vartheta + \chi^{\text{vir}} \sin \vartheta$$

for $0 \leq \vartheta \leq 2\pi$, and might therefore miss out several solutions. Later on in the tutorial, we will meet *metadynamics*, a method developed to overcome some of these problems.[1]

1.4 Restarting from Scratch

Another major challenge with the method in Section 1.3.2 that will become apparent for larger systems involves the existence and consistency of the read-in MOs. Note that so far, whenever we want to find an excited solution, we always have to run two linked jobs with the first one to converge to the reference solution from which we construct the guess MOs for the second one. In He, with a small basis set, the reference solution is almost always the same solution that gets converged to very quickly. However, for larger systems in larger basis sets, there is no guarantee that the same solution is always converged to and that the amount of time taken to reach convergence is always sensible. After all, even if we ignore all the above uncertainties, it is still a waste of time and computational resources to keep having to search for the same solution over and over again only to use it as an initial guess. It would be much better if we could store the MO coefficients of the reference solution once and for all and read them in as the initial guess whenever we need them. It turns out that Q-CHEM does allow us to do this, albeit in a rather cumbersome way. In what follows, we will recalculate the excited RHF solution for He in 6-31G*, but instead of running both jobs at once using the input file in Listing 5, we will run the first job and save the scratch files, then run the second job *separately* reading in the MO coefficients from these scratch files.

In a shell, execute

```
$   qchem -nt 6 -save He.RHF.6-31GSTAR.inp He.RHF.6-31GSTAR.out savedscratch
```

which saves all the binary scratch files in $QCSCRATCH/savedscratch, where QCSCRATCH is an environment variable that tells Q-CHEM the location in which it should read and write scratch files. Inspect the new He.RHF.6-31GSTAR.out file and verify that the MO coefficients are those expected for the ground RHF solution of He. Once happy, create a new file called He.RHF.6-31GSTAR.excited.secondjobonly.inp, paste into it lines 19–38 of Listing 5 (*i.e.*, without @@@), and then execute

```
$   qchem -nt 6 -save He.RHF.6-31GSTAR.excited.secondjobonly.inp He.RHF.6-31\
    >   GSTAR.excited.secondjobonly.out savedscratch
```

which will now read in the molecular configuration and MO coefficients from $QCSCRATCH/savedscratch and use the requested MOs in $occupied as the initial guess. The solution in the output file should be the same as that found by the second job in Listing 5.

References

[1] Alex J. W. Thom and Martin Head-Gordon. "Locating Multiple Self-Consistent Field Solutions: An Approach Inspired by Metadynamics". In: *Physical Review Letters* 101.19 (November 2008), p. 193001. DOI: 10.1103/PhysRevLett.101.193001.