Hartree-Fock calculations in PySCF

This set of exercises comprises all the information you need to run a Hartree-Fock calculation in PySCF. After giving a detailed account of how to install the software, you will learn how to create the inputs and apply the Hartree-Fock knowledge to your own electronic structure calculation.

Installation of PySCF

Below you can find instructions on how to install PySCF for non-developers, adapted from the PySCF website (`https://pyscf.org/index.html`).

Please connect to meluxina and move to the `/project/home/p201028` directory:

```
cd /project/home/p201028
```

Then, request an interactive job on a CPU node for 1 hour:

```
salloc -A p201028 --reservation=cpudev -q dev -N 1 -t 1:0:0
```

Install PySCF using `pip`:

```
pip install --prefer-binary pyscf
```

The `pip` package provides a precompiled PySCF code. For more options on how to install PySCF, please refer to the website: `https://pyscf.org/user/install.html`

Inputs

To perform a Hartree–Fock calculation, two main inputs are required: the **molecular geometry (or crystal structure)** and the choice of **basis sets**.

Initializing a molecule

There are three ways to define a molecule.

The first one is to use the `Mole.build()` method to initialize a molecule, in this case H_2O:

```
from pyscf import gto
mol = gto.Mole()
mol.build(
    atom = '''O 0 0 0; H  0 1 0; H 0 0 1''',
    basis = '''sto-3g''')
```

To confirm that the molecular geometry and basis sets have been set up properly, you can run:

```
print(mol.atom)
print(mol.basis)
```

The second one is to assign the geometry, basis, to the `Mole` object, followed by calling the `build()` method:

```
from pyscf import gto
mol = gto.Mole()
mol.atom = 'O 0 0 0; H  0 1 0; H 0 0 1'
mol.basis = 'sto-3g'
mol.build()
```

In this case as well, you can confirm that the molecular geometry and basis sets have been set up correctly by running:

```
print(mol.atom)
print(mol.basis)
```

The third way is to use the shortcut functions `pyscf.M()` or `Mole.M()`. These functions pass all the arguments to the `build()` method:

```
import pyscf
mol = pyscf.M(
    atom = '''O 0 0 0; H  0 1 0; H 0 0 1''',
    basis = '''sto-3g''')

from pyscf import gto
mol = gto.M(
    atom = '''O 0 0 0; H  0 1 0; H 0 0 1''',
    basis = '''sto-3g''')
```

The molecular geometry is specified in Cartesian coordinates, with Angstrom as the default unit. The unit can be set using the `unit` attribute to either "Angstrom" or "Bohr".

```
mol.unit = 'B'
```

The attribute `mol.unit` is case-insensitive. Any string starting with "B" or "AU" is interpreted as Bohr; all other values are treated as Angstrom.

The input parser supports both Cartesian and Z-matrix formats, as well as explicit nuclear charges. Atoms can be labeled with numbers or special characters to distinguish them, allowing different basis sets, masses, or nuclear models to be assigned. Geometries can also be defined directly in the internal format of `Mole.atom` using Python lists, tuples, or NumPy arrays, or read from an .xyz file. For details, see: `https://pyscf.org/user/gto.html`.
After building the `Mole` object, the geometry can be accessed with `Mole.atom_coords()`, which returns an (N, 3) array of atomic positions. Note that the default unit is Bohr.

```
print(mol.atom_coords(unit='Bohr'))
[[0.  0.  0.]
 [0.  1.  0.]
 [0.  0.  1.]]
```

Basis sets

Assigning a basis set name (string) to `Mole.basis` applies it to all atoms. If different basis sets are needed for different elements, `Mole.basis` should be specified as a Python dictionary.

```
mol.basis = {'O': 'sto-3g', 'H': '6-31g'}
```

Custom basis sets can be defined with `gto.basis.parse()`, which parses a string in NWChem format (`https://www.basissetexchange.org/`).

```
mol.basis = {'O': gto.basis.parse('''
C    S
     71.6168370              0.15432897
     13.0450960              0.53532814
      3.5305122              0.44463454
C    SP
      2.9412494             -0.09996723             0.15591627
      0.6834831              0.39951283             0.60768372
      0.2222899              0.70011547             0.39195739
''')}
```

As with geometry input, the basis sets dictionary can use either atomic symbols (case-insensitive) or atomic nuclear charges. Numbers or special characters may be prefixed or suffixed to atom labels. If a decorated symbol is not found in the dictionary, the parser removes the decoration and looks up the plain atomic symbol.

In the following example, the `6-31G` basis set will be assigned to the atom `H1`, but the `STO-3G` basis will be used for the atom `H2`:

```
mol.atom = '8 0 0 0; H1 0 1 0; H2 0 0 1'
mol.basis = {'O': 'sto-3g', 'H': 'sto-3g', 'H1': '6-31G'}
```

Basis sets format

Basis sets data should be supplied in NWChem format, either as a text file or a Python script. PySCF converts them in the internal format, which looks like:

```
[[angular, kappa, [[exp, c_1, c_2, ..],
                   [exp, c_1, c_2, ..],
                   ... ]],
[angular, kappa, [[exp, c_1, c_2, ..],
                   [exp, c_1, c_2, ..]
                   ... ]]]
```

where `angular` is the angular momentum, `kappa` are the Gaussian exponents and `c` the contraction coefficients. `kappa` can have values; $-l-1$ (corresponding to spinors with $j = l + 1/2$), l (corresponding to spinors with $j = l - 1/2$) or 0. When `kappa` is 0, both types of spinors are assumed in the basis.

GTO basis functions are stored in the following order:

- atoms
- angular momentum
- shells
- spherical harmonics

Basis functions are first grouped by the atoms to which they belong. Within each atom, they are organized by angular momentum. For a given angular momentum, shells are ordered from inner to outer, i.e., from large to small exponents. A shell may represent a true atomic shell (a linear combination of Gaussians) or a single primitive Gaussian with multiple angular components.

Hartree-Fock calculations

Hartree-Fock is a self-consistent field method, therefore a self-consistent iterative procedure is applied, which begins from some initial guess. PySCF provides several schemes for the initial guess, controlled by the `init_guess` attribute of the SCF solver. The possible values are:

- `minao`: Default option.
- `1e`: Not accurate for molecular systems.
- `atom`
- `huckel`
- `chk`

Even with a very good initial guess, making the SCF procedure converge is sometimes challenging. PySCF implements two kinds of approaches for SCF, namely, **direct inversion in the iterative subspace** (**DIIS**) and **second-order SCF** (**SOSCF**).

Let's run our first Hartree-Fock calculation!

From now on, Hartree–Fock will be referred to as HF.

Perform a HF calculation of the hydrogen fluoride (HF) molecule to calculate its energy, orbital energies, and Mulliken population.

```
import pyscf

mol = pyscf.M(
    atom = 'H 0 0 0; F 0 0 1.1',
    basis = 'ccpvdz',
    symmetry = True,
)

myhf = mol.HF()
myhf.kernel()

myhf.analyze()
```

`symmetry`: Point group symmetry information
The point group symmetry information is held in the `Mole` object. The symmetry module of PySCF can detect arbitrary point groups.

Check the detected point group of the system, using:

```
print(mol.topgroup)
```

If `Mole.symmetry` is set, `Mole.build()` checks that the geometry satisfies the specified symmetry. If it does not, initialization stops and an error is raised.

`.kernel()`: Function to call HF driver.

`.analyze()`: Function that calls the Mulliken population analysis etc.

Re-run the calculation with different basis sets. For details, consult `https://www.basissetexchange.org/`.

Evaluate the effect of different initial guesses.

```
import tempfile
from pyscf import gto
from pyscf import scf


mol0 = gto.M(
    atom = '''
        C      0.    0.    0.625
        C      0.    0.   -0.625 ''',
    basis = 'cc-pVDZ',
)

mf = scf.RHF(mol0).run()

mol0.verbose = 9


# Save the density matrix as the initial guess for the next calculation

dm_init_guess = mf.make_rdm1()


mol = gto.M(
    atom = '''
        C      0.    0.    0.7
        C      0.    0.   -0.7 ''',
    basis = 'cc-pVDZ',
)

tmp_chkfile = tempfile.NamedTemporaryFile()
chkfile_name = tmp_chkfile.name
mf = scf.RHF(mol)
mf.chkfile = chkfile_name
mf.kernel(dm_init_guess)

# This is the default initial guess.  It is the superpostion from atomic
# density.  The atomic density is projected from MINAO basis.

mf = scf.RHF(mol)
mf.init_guess = 'minao'
mf.kernel()
```

`.make_rdm1()`: Function that saves the density matrix.

`scf.RHF()`: Command to perform a restricted HF calculation.

`.verbose`: Control of the print level globally. The print level can be 0 (quiet, no output) to 9 (very noisy).

Evaluate the effect of different initial guesses by first performing an HF calculation on Cr^{6+} to obtain its density matrix, and then using that matrix as the initial guess for an HF calculation on the neutral Cr atom.
Bonus: Use restricted HF method. Charge and spin can be assigned to `Mole` *object, such as:*

```
mol.charge = 1
mol.spin = 1
```

Perform a restricted and unrestricted calculation of H_2O molecule.

Restricted HF

```
from pyscf import gto, scf

mol = gto.Mole()
mol.verbose = 5
mol.atom = '''
O         0.000000     0.000000     0.117790
H         0.000000     0.755453    -0.471161
H         0.000000    -0.755453    -0.471161'''
mol.basis = 'ccpvdz'
mol.symmetry = 1
mol.build()

mf = scf.RHF(mol)
mf.kernel()
```

Unrestricted HF

```
from pyscf import gto, scf

mol = gto.Mole()
mol.verbose = 5
mol.atom = '''
O         0.000000     0.000000     0.117790
H         0.000000     0.755453    -0.471161
H         0.000000    -0.755453    -0.471161'''
mol.basis = 'ccpvdz'
mol.symmetry = 1
mol.build()

mf = scf.UHF(mol)
mf.kernel()
```

Perform both restricted (RHF) and unrestricted (UHF) calculations for the H_2 molecule at different bond lengths, and compare the resulting energies. The coordinates for two representative geometries are given below:

H 0.0 0.0 0.0
H 0.0 0.0 0.7414

H 0.0 0.0 0.0
H 0.0 0.0 5.6

Geometry optimization

Perform geometry optimization of N_2 molecule.

Install `geomeTRIC` and `PyBerny`.

```
pip install geometric
pip install pyberny
```

```
from pyscf import gto, scf
from pyscf.geomopt.geometric_solver import optimize
from pyscf.geomopt.berny_solver import optimize

mol = gto.M(atom='N 0 0 0; N 0 0 1.2', basis='ccpvdz')
mf = scf.RHF(mol)

# geometric
from pyscf.geomopt.geometric_solver import optimize
mol_eq = optimize(mf, maxsteps=100)
print(mol_eq.tostring())

# pyberny
from pyscf.geomopt.berny_solver import optimize
mol_eq = optimize(mf, maxsteps=100)
print(mol_eq.tostring())
```

Perform a geometry otpimization calculation for the H_2O you used above.

Submit jobs with slurm

Rather than running interactively, jobs can be submitted via slurm using a submission script. Example for MeluXina:

```
#!/bin/bash -l
#SBATCH --qos=default                       # SLURM qos
#SBATCH --nodes=1                           # number of nodes
#SBATCH --ntasks=1                          # number of tasks
#SBATCH --ntasks-per-node=1                 # number of tasks per node
#SBATCH --time=00:05:00                     # time (HH:MM:SS)
#SBATCH --partition=cpu                     # partition
#SBATCH -A p201028                          # project account
#SBATCH --cpus-per-task=1                   # CORES per task

python3 geom.py > out
```

Save the above commands in a plain text file. If you are not familiar with other text editors, you can use `vi`. You can open a new file, by typing `vi` and the name of the file:

```
vi job_script.sh
```

After hitting `enter` you will find yourself in `vi`. `vi` has three different working modes:

- Normal mode: If you start the program or if you hit `Esc` buttom. In this mode, you can give commands to `vi`, as specified later.
- Insert mode: By typing `i`, you will be in insert mode. If you type something now, it will be inserted in your file.

- Replace mode: By hitting `i` again in the insert mode, you will pass to replace mode. If you start typing, `vi` will replace the letters below your cursor with a new one, rather than adding them. By hitting `i` again, you switch back to regular insert mode.

In your freshly opened file switch to insert mode by pressing `i`. Then copy-paste the above commands. To save the file, you have to exit the insert mode by pressing `Esc`. In escape mode, type `:w` to save the file, or type `:w job_script.sh` to save it under a filename called `job_script.sh`. You can exit `vi` by typing `:q`. If there were changes since your last save, type `:q!`, where the `!` forces `vi` to quit, without applying the changes.

To perform the geometry optimization via slurm rather than interactively, the PySCF commands should be saved in a ***`python`*** *script file, as well. Try running the H_2O example from above using slurm, too.*