University of Regensburg
Institute of Theoretical Physics
Summer term 2024
Discussion on: 6 June 2024
Dr. Jan Wilhelm, Dr. Štěpán Marek, Maximilian Graml
https://www.ur.de/physics/wilhelm/teaching/sose24-computational-nanoscience

Computational Nanoscience: Exercise Sheet No. 4

Exercise 4.1: Computational scaling of Hartree-Fock

In this exercise, we investigate how the computational cost of closed-shell Hartree-Fock calculations grows with the number of atoms in a system. For simplicity, suppose we only have the same atom type in the system and we keep the same basis set and the same numerical parameters when increasing the system size. As practical example of such a test system, you can think of a chain of H_2 molecules with increasing number of H_2 molecules. As basis for this exercise, have a look at Fig. 5.2 in the lecture notes.

(a) How does the number of basis functions N_b and the number of occupied molecular orbitals increase with the number of atoms N_atom in the calculation? **[1]**

(b) The computational cost of every single step in the Hartree-Fock algorithm in Fig. 5.2 increases (= scales) as N_atom^α, $\alpha \in \mathbb{N}$. Identify the steps where the computational cost increases with the highest exponent on N_atom (without reformulating the algorithm). **[2]**

(c) We introduce the density matrix

$$D_{\lambda\sigma} = \sum_{m=1}^{N/2} C_{m\lambda}\, C_{m\sigma}\,.$$

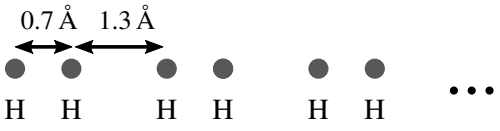
How can we use the density matrix to reduce the scaling of Hartree-Fock to N_atom^4? **[2]**

(d) Suppose, we have localized basis functions $\phi_\nu(\mathbf{r})$, for example atom-centred Gaussians as in CP2K, and suppose we have a large molecule (e.g. system size in one direction 20 Å). What is the condition on the basis functions $\phi_\nu(\mathbf{r})$, $\phi_\mu(\mathbf{r})$, $\phi_\sigma(\mathbf{r})$, $\phi_\lambda(\mathbf{r})$ in four-centre Coulomb integrals $(\mu\nu|\lambda\sigma)$ such that the integrals have a small numerical value (for example smaller than 10^{-12})? **[1]**

(e) Suppose, we neglect all four-centre integrals $(\mu\nu|\lambda\sigma)$ with small numerical value (for example: threshold 10^{-12} in the Hartree-Fock calculation). Argue that the number of kept integrals $(\mu\nu|\lambda\sigma)$ scales with N_atom^2 for reasonably large molecular systems (e.g. system size in largest direction ≈ 20 Å). **[2]**

(f) Based on rigorous screening in large molecules, we only keep $O(N_\mathrm{at}^2)$ four-centre integrals $(\mu\nu|\lambda\sigma)$. In such a scenario, how is the computational cost of Hartree-Fock increasing with the system size? **[2]**

(g) Measure the CPU time of Hartree-Fock for H_2 molecules and fill the table on the next page. You will find the relevant CPU time in the CP2K output after "integrate_four_center". Use a def2-QZVPP basis and a geometry of the H_2 molecules as follows. **[5]**



Number of H_2 molecules, N_{H_2}	CPU time (s)
1	
2	
4	
6	
8	

Hint: Set the environment variable `OMP_NUM_THREADS` to the number of CPU threads on your machine (can be checked by running `lscpu`) to get faster parallel performance. You can use the following input file

```
&GLOBAL
 RUN_TYPE ENERGY
 PROJECT_NAME HF
&END GLOBAL
&FORCE_EVAL
 METHOD QS
 &DFT
  &XC
   &HF
    FRACTION 1.0
   &END HF
   &XC_FUNCTIONAL NONE
   &END XC_FUNCTIONAL
  &END XC
  &QS
   METHOD GAPW
   ! All overlaps are considered
   MIN_PAIR_LIST_RADIUS -1.0
  &END QS
  &POISSON
   PERIODIC NONE
   POISSON_SOLVER ANALYTIC
  &END POISSON
  BASIS_SET_FILE_NAME def2-qzvpp.1.cp2k
  POTENTIAL_FILE_NAME POTENTIAL
 &END DFT
 &SUBSYS
  &CELL
   ! Leave sufficient room for the whole chain
   ABC 30.0 10.0 10.0
   PERIODIC NONE
  &END CELL
  &TOPOLOGY
   &CENTER_COORDINATES
   &END CENTER_COORDINATES
   COORD_FILE_FORMAT XYZ
```

```
    ! Here you can change the molecule that is being investigated
    COORD_FILE_NAME H2_1.xyz
   &END TOPOLOGY
   &KIND H
    BASIS_SET def2-QZVPP
    POTENTIAL ALL
   &END KIND
  &END SUBSYS
&END FORCE_EVAL
```

(h) Assume the CPU time (dominated by integrate_four_center routine) scales as

$$\text{CPU time} = C \cdot N_{\text{H}_2}^{\alpha} .$$

Determine the exponent α when using your data points (1) $2\,\text{H}_2$ and $4\,\text{H}_2$ and (2) $4\,\text{H}_2$ and $8\,\text{H}_2$. **[3]**

(i) What is the reason in the Hartree-Fock algorithm for the difference between the exponents you have determined in (h) (1) and (2)? **[2]**