Hartree-Fock electronic structure solver with Gaussian type oribtals

The engine of all modern computational chemistry programs

Problem description:

- The foundation of quantum chemistry is the time-independent Schrödinger wave equation:
  $H\,\psi = E\,\psi$
- Here H is the Hamiltonian, the total, i.e., kinetic T plus potential V, energy operator
- ψ is the “oracle”, i.e., the many-electron wave function of the molecule, where $|\psi(\mathbf{x})|^2$ corresponds to the electronic charge density $\rho(\mathbf{x})$ at the point **x**
- E is the “answer” (i.e., the total energy of the many-electron systemfor the given molecular geometry)
- An additional difficulty is that the Hamiltonian is also a functional of the charge density (since the electrons interact with both nuclei and themselves, and thus the charge density contributes to the potential energy), i.e., $H = H[\rho]$
- A typical approximate solution to this nonlinear problem is iterative: starting with a guess for ψ and keep improving until a self-consistent consistent form for ψ is found

Hartree-Fock molecular orbital (MO) theory:

- $H = T_e + T_n + V_{ne} + V_{ee} + V_{nn}$, where n and e stand for nuclei and electron contributions, respectively.
- In the operator form (assuming *e=ħ=m=1* for the electron charge, Plancks constant, and all masses):

$$\hat{H}|\Psi\rangle = E|\Psi\rangle$$

$$\begin{aligned}\hat{H} = \; & -\sum_{i}^{N}\frac{1}{2}\nabla_i^2 - \sum_{A}^{N}\frac{1}{2}\nabla_A^2 \\ & -\sum_{i}^{N}\sum_{A}^{M}\frac{Z_A}{r_{iA}} \\ & +\sum_{i}^{N}\sum_{j}^{N}\frac{1}{r_{ij}} + \sum_{A}^{M}\sum_{B}^{M}\frac{Z_A Z_B}{r_{AB}}\end{aligned}$$

- static geometry: $T_n = 0$, V_{nn} *trivial, leaving only the electrons* ($T_e + V_{ne} + V_{ee}$) *to worry about*
- Independent electron approximation: each electron is moving in the mean field of the other electrons
- The wave function ansatz: a product (antisymmetrized, fermion) of single electron molecular spin-oribtals ψ_i (i.e., with up to two electrons with opposite spins per MO), determinant:

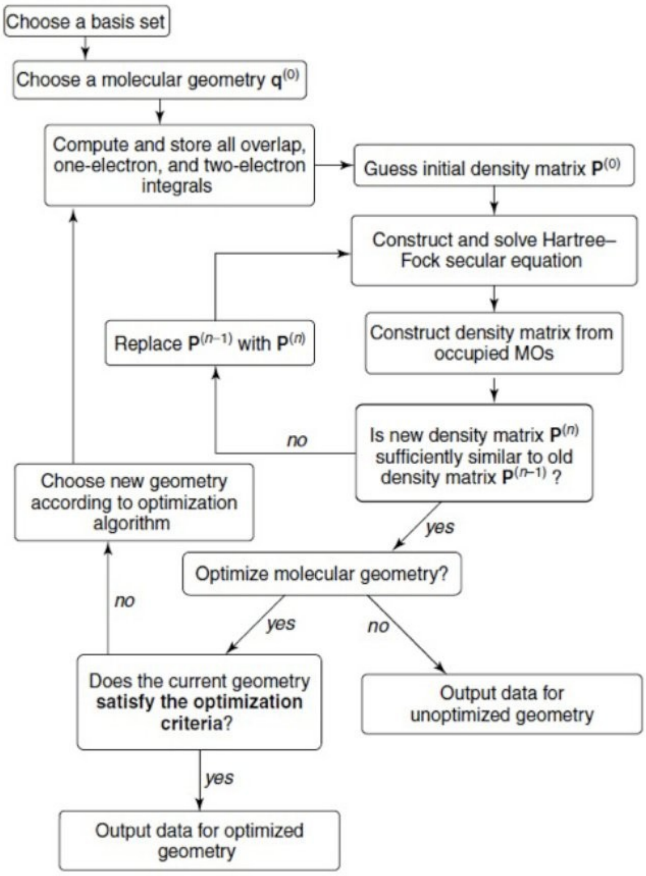
$$\Psi_{\text{molc}} = \frac{1}{\sqrt{N!}}\begin{vmatrix} \psi_1(1) & \psi_1(2) & \cdots & \psi_1(N) \\ \psi_2(1) & \psi_2(2) & \cdots & \psi_2(N) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_N(1) & \psi_N(2) & \cdots & \psi_N(N) \end{vmatrix}$$

- Another assumption: MO as a linear combination of atomic orbitals (LCAO) ϕ_μ: $\psi_i = \Sigma_\mu\, C_{\mu i}\, \phi_\mu$
- A natural choice for the ϕ_μ orbitals would be the standard *s*, *p*, *d*, *f* etc type atomic orbitals, with exponential decay radially away from theeir nuclei, the so-called Slater-type orbitals (STO). The one and two electron multicenter integrals typically need numerical integration, which is inefficient.
- STOs used only in semi-empirical quantum chemistry models, where many of the integrals are either discarded or approximated
- A more efficient way: approximate each atomic orbital with a linear combination of cartesian Gaussians:

$$\phi(x,y,z;\alpha,i,j,k) = \left(\frac{2\alpha}{\pi}\right)^{3/4} \left[\frac{(8\alpha)^{i+j+k}\, i!\,j!\,k!}{(2i)!(2j)!(2k)!}\right]^{1/2} x^i y^j z^k e^{-\alpha(x^2+y^2+z^2)}$$

- Here *x*, *y*, *z* are electron's *i* coordinates relative to its atom *A* nuclear coordinates, $x = x_i - x_A$ etc
- The GTO basis is not orthogonal, with non-zero overlaps $S_{\mu\nu} = \langle \mu | \nu \rangle = \int \phi_\mu \phi_\mu \, d\mathbf{x}$
- A seminal work by L.E. McMurchie and E.R. Davidson, *J. Comput. Phys.* 26 (1978) 218 https://doi.org/10.1016/0021-9991(78)90092-X
- All one and two-electron integrals evaluated analytically via Hermite polynomials
- The eigenvalue problem, Roothaan-Hall equations: $F\,C = S\,C\,E$, where C is the matrix of MO coefficients $C_{\mu i}$ we wish to determine, *S* is the overlap matrix between the atomic GTOs, *E* is the vector of MO energies, and *F* is the Fock operator/matrix whose elements contain the nuclear-electron attraction integrals, electron-electron Coulomb integrals and exchange integrals
- *F* is a functional of the unknown $C_{\mu i}$ coefficients, i.e., $F = F\,[C]$, which requires a self-consistent solution, via iterations

Flow chart:


Choose a basis set
Choose a molecular geometry q(0)
Compute and store all overlap, one-electron, and two-electron integrals
Guess initial density matrix P(0)
Construct and solve Hartree–Fock secular equation
Replace P(n−1) with P(n)
Construct density matrix from occupied MOs
no
Is new density matrix P(n) sufficiently similar to old density matrix P(n−1) ?
Choose new geometry according to optimization algorithm
yes
Optimize molecular geometry?
no
yes
no
Does the current geometry satisfy the optimization criteria?
Output data for unoptimized geometry
yes
Output data for optimized geometry

- N.B. the bottom part, geometry optimization (bond lengths and angles) has not been implemented, the program only produces output for the initial geometry.

Fortran implementation:

- Ten `*.f90` files in the https://github.com/snazzybloke/Hartree_Fock_Gaussian implementing the actual workflow and providing useful definitions/utilities.
- The main program `abmol` starts the workflow execution by calling the `readmol` subroutine, which reads the input file, provided by the user as the only command line argument. The file needs to be in a format closely resembling the XMol xyz format. For example, for the CO molecule the input file (`CO_631.au`) should look something like this:

```
2
6-31G**   0
O     0.0     12.105    0.0     8
C     0.0     10.0      0.0     6
```

- Here the first line (`2`) shows the total number of atoms. The second line shows the choice of the GTO basis set (the popular `6-31G**`) and the total charge (`0`, neutral molecule). The atom xyz coordinates, preceded by the element names (`O`, `C`), are then given in atomic units (i.e., not Ångstroms), with their atomic numbers provided in the final column (`8`, `6`)
- Subsequently, the Coulomb repulsion between the positively charged nuclei (the V_{nn} term) is evaluated by calling the `nuclrep` subroutine.
- Next the size of the matrices `msize` is determined by calling the `matsize` subroutine, based on the choice of the GTO basis set (`STO-3G`, `3-21G`, or `6-31G**`)
- To set the initial guess for the electronic charge density the `atomics` subroutine is then invoked. It evalues the initial density matrix `mP` as the sum of the individual atomic contributions, i.e., it solves the Hartree-Fock secular equation (Roothan-Hall eqs) for the isolated atoms first, by calling the `integral`, `collect` and `scf` subroutines (see below)for the details about these). Without a reasonable initial guess for the molecular charge density a congverging self-consistent solution quite often may not be found.
- Now the self-consistent loop on the right hand side of the flow chart can begin. Firstly, the `integral` subroutine is invoked to calculate: (1) the orbital overlap matrix (*S* as the 2D matrix `si`), (2) the electron kinetic energy contributions (T_e as the 2D matrix `ti`), (3) the one-electron (the V_{ne} term as the `v1ei` 3D matrix), and (4) two-electron (the V_{ee} term as the `v2ei` 4D matrix) contributions to the electrostatic energy.
- The subroutine `collect` is invoked to evaluate: (1) the “core” Hamiltonian matrix (`mH`, without the two-electron V_{ee} term) and (2) the canonical orthogonalization which diagonalizes the overlap matrix `si` (also named `mS`) with the new, transformed basis of orthogonal orbitals, thus providing the transformation matrix `mX`.
- Finally the `scf` subroutine is called to carry out the iterative (self-consistent) solution of the Roothaan-Hall eqs. Firstly, the Fock operator (the matrix `mF`) is evalauted as the sum of the “core” Hamiltonian (`mH`) and the two-electron Coulomb energy contributions (`mG`), transformed into the new basis of orthogonal basis orbitals and diagonalized. The eigenvectors (`mCp`), when back-transformed into the original orbital basis, produce the matrix of the MO coefficients (`mC`). The eignevalues are the MO energies (`mEig`). The solution is found when density matrix (`mP`), calculated from the new matrix of MO coefficients, has fully converged.

Limitations:

- GTO basis sets have been provided only the eighteen light elements (from H to Ar)
- GTO basis set choice currently limited to only four (`STO-3G`, `3-21G`, `4-31G`, or `6-31G**`)

Repository files :

- ten `*.f90` Fortran source files
- lib/libnag.a provides the NAG (Mk 14, 1989) Fortran77 implementation of gamma and incomplete gamma functions (S14AAF and S14BAF, respectively), compiled using the Intel `ifort` compiler
- `Makefile` to compile the source (type “`make abmol`”) using the Intel `ifort` compiler
- Sample input files for several small molecules in the directory `inputs/*.au`
- this file
- `README.md` with the brief description

Requirements, installation and running

- Any old Linux/UNIX/MacOS powered machine
- `make` command line utility
- Intel `ifort` compiler and MKL library
- NAG library (or at least its minimal subset with S14BAF and S14BAF subroutines, https://www.nag.com/numeric/fl/nagdoc_fl25/html/s/s14baf.html, as provided in this repo)
- To compile, enter “`make abmol`”
- To execute, enter e.g., “./`abmol inputs/CH4_631.au > CH4_631.out`”

Literature:

1. A. Szabo & N.S. Ostlund, *"Modern Quantum Chemistry"*, Dover, Mineola N.Y., 1996
2. L.E. McMurchie and E.R. Davidson, *J. Comput. Phys.* 26 (1978) 218