Chemistry Summer Project

pp583

July 2025

1 Procedure for finding parameters to feed into a machine learning model

1. Get Localized Molecular Orbitals

- Generate a molecule.
- Using `mf.mo_occ` and `mf.mo_coeff`, split the molecular orbital (MO) coefficients into those corresponding to occupied and virtual (unoccupied) orbitals.
- The molecular orbitals (ψ_i) are expressed as:

$$\psi_i = \sum_{\mu=1}^{K} \phi_\mu C_{\mu i}$$

  where K is the number of atomic orbitals

- After splitting, we define:

$$\psi_i^{\text{occ}} = \sum_{\mu=1}^{K} \phi_\mu C_{\mu i} \qquad \text{and} \qquad \psi_a^{\text{vir}} = \sum_{\mu=1}^{K} \phi_\mu C_{\mu a}$$

  where $i = 1, 2...N_{occ}$ and $a = 1, 2...N_{vir}$

- Then we localize the orbitals by finding a unitary rotation:

$$\widetilde{\psi}_{\text{occ}} = \psi_{\text{occ}} \text{U}_{\text{occ}} \qquad \text{and} \qquad \widetilde{\psi}_{\text{vir}} = \psi_{\text{vir}} \text{U}_{\text{vir}}$$

- In our coordinate system, set by the basis functions, we represent this as:

$$\widetilde{C}_{\mu I}^{\text{occ}} = \sum_i C_{\mu i}^{\text{occ}} U_{iI} \qquad \text{and} \qquad \widetilde{C}_{\mu A}^{\text{vir}} = \sum_a C_{\mu a}^{\text{vir}} U_{aA}$$

- Then we stack the two matrices to obtain the new matrix with the coefficients of the localized molecular orbitals: $\widetilde{C} = (\widetilde{C}_{\mu I}^{\text{occ}}, \widetilde{C}_{\mu I}^{\text{vir}})$

2. Perform Population Analysis on the Molecular Orbitals

***!* usually we should orient orient all the bonds along the z-axis when performing these calculations but I am not sure how this will affect our orbitals, so for now I am working with diatomic molecules that are oriented along the z-axis**

- The number of electrons in each orbital is:

$$n_i = \left\langle \widetilde{\psi}_i \middle| \widetilde{\psi}_i \right\rangle = 1$$

- In our chosen basis, this becomes:

$$n_i = \sum_{\mu,\nu} \widetilde{C}^*_{\mu i} \widetilde{C}_{\nu i} \int \phi^*_\mu(r) \phi_\nu(r)\, dr = \sum_{\mu,\nu} \widetilde{C}^*_{\mu i} \widetilde{C}_{\nu i} S_{\mu\nu}$$

- We define the **atomic population** of orbital ψ_i on atom A as:

$$q_A = \sum_{\mu \in A} \sum_{\nu} \widetilde{C}^*_{\mu i} \widetilde{C}_{\nu i} S_{\mu\nu} = \sum_{\mu \in A} \sum_{\nu} \widetilde{C}^\dagger_{i\mu} S_{\mu\nu} \widetilde{C}_{\nu i}$$

- By looking at q_A for the different atoms, we can determine which atoms the orbitals are centered on (along with knowing interatomic distances).

3. Calculate the Expectation Values of $\hat{L}_z$ for the localized Molecular Orbitals

- The expectation value is:

$$\left\langle \widetilde{\psi}_i \right| \hat{L}_z \left| \widetilde{\psi}_i \right\rangle = \sum_{\mu,\nu} \widetilde{C}^*_{\mu i} \left\langle \phi_\mu \right| \hat{L}_z \left| \phi_\nu \right\rangle \widetilde{C}_{\nu i} = \sum_{\mu,\nu} \widetilde{C}^\dagger_{i\mu} L_{\mu\nu} \widetilde{C}_{\nu i} = (\widetilde{C}^\dagger L \widetilde{C})_{ii}$$

  where L is the matrix representation of the angular momentum operator $\hat{L}_z$ in the basis $\{\phi_\mu\}$

  However, since the matrix representation of $\hat{L}_z$ in a real-valued atomic orbital basis has vanishing diagonal elements, it is more appropriate to use the squared operator $\hat{L}_z^2$ to characterize the angular momentum of molecular orbitals.

$$\left\langle \widetilde{\psi}_i \right| \hat{L}_z^2 \left| \widetilde{\psi}_i \right\rangle = \sum_{\mu,\nu} \widetilde{C}^*_{\mu i} \left\langle \phi_\mu \right| \hat{L}_z^2 \left| \phi_\nu \right\rangle \widetilde{C}_{\nu i} = \sum_{\mu,\nu} \widetilde{C}^*_{\mu i} \sum_{\lambda} (\left\langle \phi_\mu \right| \hat{L}_z \left| \phi_\lambda \right\rangle \left\langle \phi_\lambda \right| \hat{L}_z \left| \phi_\nu \right\rangle) \widetilde{C}_{\nu i} = \sum_{\mu,\nu} \widetilde{C}^\dagger_{i\mu} L^2_{\mu\nu} \widetilde{C}_{\nu i} = (\widetilde{C}^\dagger L^2 \widetilde{C})_{ii}$$

  The `pyscf` code for obtaining the $\hat{L}_z$ matrix is:

```
from pyscf.gto import moleintor
# Build molecule
mol = gto.Mole()
mol.atom = 'H 0 0 0; F 0 1.1 0'
mol.basis = 'sto-3g'
mol.build()

# Compute angular momentum operator integrals: r × p
lz_3comp = moleintor.getints('int1e_cg_irxp_sph',
                             mol._atm, mol._bas, mol._env,
                             comp=3)
lz_matrix = lz_3comp[2]

```

4. Calculate the expectation values of $\hat{F}$) for the localized Molecular orbitals(i.e estimate their energies)

- The expectation value is:

$$\left\langle \widetilde{\psi}_i \right| \hat{F} \left| \widetilde{\psi}_i \right\rangle = \epsilon_i$$

$$\epsilon_i = \int (\sum_p \psi^*_p U^*_{pi}) \hat{F} (\sum_q \psi_q U_{qi}) = \int (\sum_p \psi^*_p U^*_{pi}) (\sum_q \epsilon_q \psi_q U_{qi}) = \sum_{pq} U^*_{pi} U_{qi} \epsilon_q \delta_{pq} = \sum_{pq} U^\dagger_{ip} U_{qi} \epsilon_q \delta_{pq} = \sum_p |U_{pi}|^2 \epsilon_p$$

  this is a **weighted average** of the energies of the canonical MOs ϵ_p U has the form:

$$U = \begin{pmatrix} U^{occ} & 0 \\ 0 & U^{vir} \end{pmatrix}$$