Machine Learning Development for a Smart Selective Full-CI Method

Bastien Casier†, bastien.casier@univ-artois.fr
† *Univ. d'Artois, CNRS, UCCS, UMR 8181, F-62300 Lens, France*

English version – M2 Internship, 4 to 5 months

Resolving in a straight manner the Full Configuration Interaction (FCI) equation stays an impossible way. The impracticable nature of this problem comes from the exponential growth of the number of Slater determinants according to the number of orbitals. For instance, the size of the Hilbert's space relative to the ethane molecule (C_2H_6) reaches more than one million of determinants in a minimal basis set (e.g. STO-3G). Consequently, a simple solution consists in limiting the full expansion of the FCI wave function to simply and doubly excited determinants (CISD) [1]. Nevertheless, this truncation, resulting from an arbitrary choice, may be too severe, especially at dissociation limits where a large amont of determinants becomes degenerated.

Another strategy consists in selecting the most important Slater determinants. To do so, the most popular method is the *Configuration Interaction using a Perturbative Selection done Iteratively* (CIPSI) [2]. This approach presents some outstanding results at the equilibrium geometries and around. However, the perturbative nature of the method yields a lack of accuracy for large interatomic distances.

In this internship we will use an iterative machine learning classification algorithm to smartly select the most important determinants [3]. The learnable information will be encoded in a binary representation of the spin-orbital populations of each Slaters determinant, while the learning will be based on the minimization of the binary cross-entropy. Nowadays, this method presents very promising results for small molecules (e.g. CO, N2, H2O, NH3, C2H6, etc.), as FCI accurate energies have been obtained along all the dissociation curve by only using a small fraction of the Hilbert's space. This primary study is a proof of principle of the binary classification of the Slater determinants for the selective CI methods. Our end goal would be to realize a global learning on the integral curves and to define the first interatomic potentials at the FCI level.

Keywords : *Ab-initio* calculations, Full-CI, Machine Learning, Classification, Cross entropy.

References

[1] Attila Szabo and Neil S. Ostlund. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. First. Dover Publications, Inc., 1996.

[2] B. Huron, J. P. Malrieu, and P. Rancurel. "Iterative Perturbation Calculations of Ground and Excited State Energies from Multiconfigurational Zeroth-Order Wavefunctions". In: *The Journal of Chemical Physics* 58.12 (1973), pp. 5745–5759.

[3] B. Herzog et al. *Solving the Schrödinger Equation in the Configuration Space with Generative Machine Learning*. 2022.

Laboratory

The Unité de Catalyse et Chimie du Solide (UCCS) results from the merger of two regional laboratories and is shared between two universities, the Université d'Artois and Université de Lille.

The research conducted at UCCS focuses on two main scientific fields: *Energy* and *Sustainable Development*, which are divided into three axes, namely homogeneous and heterogeneous catalysis, molecular chemistry and solid state chemistry.

This internship is offered within the site of Université d'Artois (Faculté des Sciences Jean Perrin, 62300 Lens, France).

Application

For this internship, we are looking for a Master 2 student (in **Chemistry** or **Physics**) wishing to broaden their skills in **molecular modeling** to the domain of **Machine Learning**. The student must show a very particular **motivation** for this new field of research. It will be also expected to have a solid knowledge in ***ab-initio* calculations**, such as the Hartree-Fock method, as well as some **programming skills** (Python, C/C++, Fortran) – knowledge of the Linux operating system would be a plus.

The application form must **imperatively** contain a CV and a cover letter. *Any incomplete form will not be retained.*