

Développement d'une méthode Full-CI sélective par Machine Learning

Bastien Casier[†], bastien.casier@univ-artois.fr

[†] Univ. d'Artois, CNRS, UCCS, UMR 8181, F-62300 Lens, France

Version française – Offre de Stage M2, 4 à 5 mois

L'approche Full-CI est une méthode de chimie quantique post-Hartree-Fock qui permet de décrire avec la plus grande précision la corrélation électronique. Toutefois, résoudre ses équations d'une manière directe reste impossible en pratique. Le problème vient du fait que le nombre de déterminants de Slater impliqués dans le développement de la fonction d'onde électronique croît exponentiellement suivant la taille de la base de fonctions utilisée. Par exemple, pour une base de fonctions minimale (*e.g.* STO-3G), l'espace de Hilbert associé à la molécule d'éthane (C_2H_6) contient plus d'un million de déterminants de Slater. Aussi une manière simple de contourner la complexité du problème est-elle de limiter l'expansion de la fonction d'onde FCI aux simples et doubles excitations (CISD) [1]. Toutefois, outre résulter d'un choix arbitraire, cette troncation est quelque peu brutale, principalement aux limites de dissociation où de nombreux déterminants deviennent dégénérés.

Une autre stratégie consiste donc à déterminer les déterminants excités (composant la fonction d'onde FCI) les plus significatifs. Par exemple, une des méthodes les plus populaires sélectionne les déterminants par des perturbations successives de l'hamiltonien. Celle-ci est connue sous le nom de CIPSI pour *Configuration Interaction using a Perturbative Selection done Iteratively* [2]. Toutefois, de part sa nature perturbative, la méthode ne fonctionne pas de manière précise pour de larges distances interatomiques.

Pour ce stage, je propose de dépasser l'approche CIPSI par une méthode itérative de classification basée sur un algorithme de *Machine Learning* [3] – *e.g.* un réseau de neurones. L'information apprise par la machine sera encodée sous la forme d'un vecteur « binaire » contenant la population de chaque spin-orbitale – *i.e.* 0 ou 1 – et l'apprentissage sera, quant à lui, basée sur la minimisation de l'entropie croisée binaire. La méthode présente actuellement des résultats très prometteurs pour de petites molécules comme N_2 , H_2O ou encore C_2H_6 , où la précision FCI a été atteinte pour chaque point de la courbe d'énergie potentielle (CEP) obtenue le long de leur axe principal. Cette étude préliminaire est une preuve de principe de l'utilisation de la classification binaire des déterminants de Slater pour les méthodes CI sélectives. Toutefois, je souhaiterais étendre cette approche à l'apprentissage simultané d'un ensemble de points pour *in fine* obtenir des potentiels d'interaction FCI.

Mots-clés : Calcul *ab-initio*, Full-CI, Machine Learning, Classification, Python, Entropie croisée.

Références

- [1] Attila SZABO et Neil S. OSTLUND. *Modern Quantum Chemistry : Introduction to Advanced Electronic Structure Theory*. First. Dover Publications, Inc., 1996.
- [2] B. HURON, J. P. MALRIEU et 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), p. 5745-5759.
- [3] B. HERZOG et al. *Solving the Schrödinger Equation in the Configuration Space with Generative Machine Learning*. 2022.

Présentation du laboratoire

L'Unité de Catalyse et Chimie du Solide (UCCS) résulte de la fusion de deux laboratoires régionaux et se partage entre deux sites universitaires que sont l'Université d'Artois et l'Université de Lille.

Les recherches menées à l'UCCS portent sur deux principaux champs scientifiques : l'*Énergie* et le *Développement Durable*, lesquels sont déclinés en trois axes, à savoir la catalyse homogène et hétérogène, la chimie moléculaire et la chimie du solide.

Le présent stage est proposé au sein du site de l'Université d'Artois (Faculté des Sciences Jean Perrin, 62300 Lens, France).

Profil du (de la) candidat(e)

Nous recherchons pour ce stage un(e) étudiant(e) de Master 2 (**Chimie** ou **Physique**) désireux(se) d'élargir ses compétences en **modélisation moléculaire** au **domaine du Machine Learning**. L'étudiant(e) devra, au demeurant, montrer une **motivation** toute particulière pour ce nouveau champ de recherches. Il est également attendu, de la part de l'étudiant(e), de solides connaissances en matière de **calculs *ab-initio*** comme la méthode Hartree-Fock, ainsi que certaines compétences en **programmation** (Python, C/C++, Fortran) – une maîtrise du système d'exploitation Linux serait un plus.

Le dossier de candidature devra **impérativement** contenir un CV et une lettre de motivation. *Tout dossier incomplet ne sera pas retenu.*