

Apprentissage sous contraintes physiques

Molecule Energy prediction

Hanna Bekkare - Maxime Moshfeghi

June 27, 2025

Objectif du projet : prédire l'**énergie d'atomisation** d'une molécule à partir de sa structure géométrique 3D.

Contraintes physiques à respecter (invariances) :

- **Invariance par translation** :

$$E(\{\mathbf{r}_i + \mathbf{t}\}) = E(\{\mathbf{r}_i\}) \quad \text{pour tout } \mathbf{t} \in \mathbb{R}^3$$

- **Invariance par rotation** :

$$E(\{R\mathbf{r}_i\}) = E(\{\mathbf{r}_i\}) \quad \text{pour toute rotation } R \in SO(3)$$

- **Invariance par permutation** des atomes de même nature :

$$E(\{\mathbf{r}_{\pi(i)}, Z_{\pi(i)}\}) = E(\{\mathbf{r}_i, Z_i\}) \quad \text{pour toute permutation } \pi$$

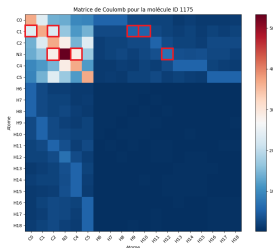
Enjeu : construire des *descripteurs moléculaires* qui respectent naturellement ces invariances pour entraîner un modèle de régression fiable. Dynamic of the system is controlled by the equation of motion:

Descripteur : Matrice de Coulomb

Définition de la matrice :

$$M_{ij} = \begin{cases} 0.5Z_i^{2.4} & \text{si } i = j \\ \frac{Z_i Z_j}{R_{ij}} & \text{si } i \neq j \end{cases}$$

Visualisation de la matrice de Coulomb :



Matrice de Coulomb d'une molécule

Descripteur : Scattering

Scattering Harmonique 3D : méthode mathématique inspirée de la théorie des ondelettes pour exploiter les caractéristiques des densités électroniques de la molécule.

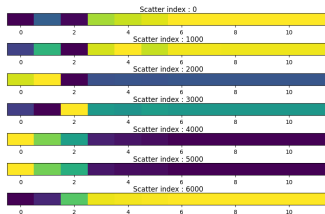
Apport des invariance :

$$\underbrace{\rho_x(u) = \sum_k \gamma_k g(u - r_k)}_{\text{Permutation}}$$

$$\underbrace{U[j, l] \rho(u) = \left(\sum_{-l}^l |\rho * \psi_{j, l}^m(u)|^2 \right)^{1/2}}_{\text{Rotation}}$$

$$\underbrace{S\rho[j, l, q] = \int_{\mathbb{R}^3} |U[j, l] \rho(u)|^q du}_{\text{Translation}}$$

Visualisation du scattering :



Analyse du scattering au premier ordre centré de 7 molécules

Regressor	Méthode d'encodage	RMSE Train	RMSE Test
XGBRegressor + SOAP	Confusion matrix with sorted_l2	0.110	0.466
Ridge, alpha = 0.001	Scattering, J = L = 3, (M, N, O)=(160, 112, 80)	0.108	0.108

Table: Comparaison des performances en RMSE des différents modèles testés

Les meilleurs résultats sont obtenus avec les méthodes contraintes par la physique.

Formule des ondelettes solides

$$\psi_{\ell}^m(u) = \frac{1}{\left(\sqrt{2\pi}\right)^3} e^{-|u|^2/2} |u|^{\ell} Y_{\ell}^m \left(\frac{u}{|u|} \right)$$