High Dimensional Learning rather than Computing in Quantum Chemistry

Matthew Hirn*, Stéphane Mallat*, Nicolas Poilvert**

*École normale supérieure
**Pennsylvania State University
Motivation

• Chemists want to build “Google of molecules”

• Pharmaceutical industry
  Materials science

• Need to compute energy of each molecule

• Billions of molecules

• Complex, time consuming computation
Energy Computation

• **Exact:**
  Schrödinger’s Equation
  Extremely high dimensional eigenvalue problem
  Example: Alcohol $\text{C}_2\text{H}_6\text{O}$ is $\sim 2^{300}$ dimensional

• **Approximate:**
  Coupled cluster methods
  Density functional theory
  Scales as $O(N^a)$ where $4 \leq a \leq 7$
  Number of electrons
Regression

- High dimensional $x \in \mathbb{R}^d$

- Approximate a functional $f(x)$
  given $n$ sample values $\{x_i, f(x_i)\}_{i=1}^n$

- Many body problems:
  Energy $f(x)$ of a state $x = \{(p_k, q_k)\}_k$
Curse of Dimensionality

• $f(x)$ can be approximated from samples $\{x_i, f(x_i)\}_{i=1}^{n}$ by local interpolation if $f$ is regular and there are close examples

• Need $n = \epsilon^{-d}$ points to cover $[0, 1]^d$ at a Euclidean distance $\epsilon \rightarrow \|x - x_i\|$ is always large
Sparse Linear Regression

- Representation of $x$: $\Phi(x) = \{\phi_p(x)\}_p$

- Regression $\tilde{f}(x)$ of $f(x)$ linear in $\Phi(x)$:
  $$\tilde{f}(x) = \langle \alpha, \Phi(x) \rangle = \sum_p \alpha_p \phi_p(x)$$

- Interpolates: $\tilde{f}(x_i) = f(x_i)$

- Few samples $\{x_i, f(x_i)\}_{i=1}^n$
  $\implies$ can only learn small number of coefficients $\{\alpha_p\}_p$
  $\implies$ must have a sparse expansion of $f$ in $\{\phi_p\}_p$ to obtain good regression

- Sparsity $\implies$ $\tilde{f}$ inherits the properties of $\{\phi_p\}_p$
  $\implies$ $\{\phi_p\}_p$ must possess the properties of $f$
Energy Properties

- State: \( x = \{(p_k, q_k)\}_k \)

- Energy: \( f(x) \)

1. Invariant to actions of the isometry group:
   \[
   E(d) = \mathbb{R}^d \rtimes O(d)
   \]

2. Multiscale potential

3. Lipschitz continuous to the action of diffeomorphisms

- Want a representation \( \Phi \) with these three properties
Classical Physics

- Energy of $N$ interacting bodies (Coulomb, gravitation)
- Invariant to isometries
- Multiscale potential

- Point charges/masses: $x \mapsto \rho(u) = \sum_{k=1}^{N} q_k \delta(u - p_k)$

- Potential: $V(u) = |u|^{-\beta} \implies f(x) = f(\rho) = \sum_{k \neq \ell} \frac{q_k q_\ell}{|p_k - p_\ell|^\beta}$

- Diagonalized by Fourier modulus:

$$O(N) \quad f(\rho) = \int \hat{V}(\omega) \left| \hat{\rho}(\omega) \right|^2 d\omega$$

$\alpha_\omega \quad \phi_\omega(\rho)$

coefficients
Wavelets

• Complex wavelet:

\[ \psi(u) = g(u) e^{i \xi \cdot u}, \quad u \in \mathbb{R}^d, \quad d = 2, 3 \]

• Dilated and rotated:

\[ \psi_{j,r}(u) = 2^{-d j} \psi(2^{-j} ru), \quad (j, r) \in \mathbb{Z} \times O(d) \]
Potential Diagonalized by Wavelet Energy Coefficients

- Properly defined $L^2$ wavelet energy coefficients define a representation that is invariant over isometries and gives a multiscale decomposition of the potential.

**Theorem** (H., Mallat, Poilvert; 2014): For any $\epsilon > 0$ there exists wavelets with

\[
 f(\rho) = (1 + \epsilon) \sum_j \alpha_j \int_{O(3)} \|\rho \ast \psi_{j,r}\|^2 \, dr
\]

$O(\log N)$ coefficients
Quantum Chemistry
Density Functional Theory

- State: \( x = \{ (p_k, q_k) \}_k \) = Positions and total protons charges of the atoms
  \( f(x) \) = Energy of the molecule

- Invariant to isometries
  - Multiscale potential
  - Stability to diffeomorphisms

- Electronic density: \( x \mapsto \rho(u) \)

- Hohenberg-Kohn, 1964:
  \[
  f(x) = f(\rho) = \min_{\tilde{\rho}} E(\tilde{\rho}), \quad \rho = \arg \min_{\tilde{\rho}} E(\tilde{\rho})
  \]

- Will have to learn representation as \( x \mapsto \Phi(\tilde{\rho}) \) where \( \tilde{\rho} \) is an approximate electronic density that can be efficiently derived from \( x \)
Electronic Density

Solution to variational problem:

$$\rho = \arg \min_{\tilde{\rho}} E(\tilde{\rho})$$

$$E(\rho) =$$

$$T(\rho)$$  
 Kinetic energy

$$+ \int \rho(u)V(u)$$  
 Electron-nuclei attraction

$$+ \frac{1}{2} \int \int \frac{\rho(u)\rho(v)}{|u - v|}$$  
 Electron-electron Coulomb repulsion

$$+ E_{xc}(\rho)$$  
 Exchange correlation energy
Electronic Density

Locally Kato Cusp Condition:

\[ \rho(u) \sim e^{-2q_k |u - p_k|} \]
Approximate Density

\[ \tilde{\rho}(u) = \sum_k q_k^4 e^{-2q_k |u-p_k|} \]
Stability to Diffeomorphisms

- Diffeomorphism $1 - \tau$:
  $$D_\tau \rho(u) = \rho (u - \tau(u))$$

- Amplitude of diffeomorphism: $\|\nabla \tau\|_\infty$

- Want **Lipschitz stability** to diffeomorphisms:
  $$\tilde{\rho} = D_\tau \rho \iff \| \Phi(\rho) - \Phi(D_\tau \rho) \| \leq C \|\nabla \tau\|_\infty \| \rho \|$$
Fourier Unstable to Diffeomorphisms

- Fourier modulus representation:
  \[ \Phi(\rho) = \{ \phi_\omega(\rho) \}_\omega = \{ |\hat{\rho}(\omega)| \}_\omega \]

- Fourier:
  Unstable to small diffeomorphisms \( \rho_\tau(u) = \rho(u - \tau(u)) \)
  \[ ||\hat{\rho}(\omega)| - |\hat{\rho}_\tau(\omega)|| \] is large at high frequencies

\[ \implies ||\Phi(\rho) - \Phi(\rho_\tau)|| \gg ||\nabla\tau||_\infty ||\rho|| \]
Scattering Representation

Layer 0
\[ p = \emptyset \]

\[ \tilde{\rho} \]

\[ \phi_0(\tilde{\rho}) = \int_{\mathbb{R}^d} \tilde{\rho}(u) \, du \]
Scattering Representation

Layer 0
\( p = \emptyset \)

Layer 1
\( p = j \)

\[ \tilde{\rho}_j(u, r) = |\tilde{\rho} * \psi_{j,r}(u)| \]

\[ \phi_0(\tilde{\rho}) = \int_{\mathbb{R}^d} \tilde{\rho}(u) \, du \]

\[ \phi_j(\tilde{\rho}) = \int_{E(d)} \tilde{\rho}_j(u, r) \, du \, dr \]
Scattering Representation

Layer 0
\[ p = \emptyset \]

Layer 1
\[ p = j \]

Layer 2
\[ p = (j, \lambda_2) \]

\[ \tilde{\rho} \]

\[ \tilde{\rho}_j''(u, r) \]

\[ \tilde{\rho}_j'(u, r) \]

\[ \tilde{\rho}_j(u, r) = |\tilde{\rho} \ast \psi_{j,r}(u)| \]

\[ \phi_0(\tilde{\rho}) = \int_{\mathbb{R}^d} \tilde{\rho}(u) \, du \]

\[ \phi_j(\tilde{\rho}) = \int_{E(d)} \tilde{\rho}_j(u, r) \, du \, dr \]

\[ \phi_{j,\lambda_2}(\tilde{\rho}) = \int_{E(d)} |\tilde{\rho}_j \ast \Psi_{\lambda_2}(u, r)| \, du \, dr \]
Scattering Representation

Layer 0
\[ p = \emptyset \]

Layer 1
\[ p = j \]

Layer 2
\[ p = (j, \lambda_2) \]

\[ \phi_0(\tilde{\rho}) = \int_{\mathbb{R}^d} \tilde{\rho}(u) \, du \]
\[ \phi_j(\tilde{\rho}) = \int_{E(d)} \tilde{\rho}_j(u, r) \, du \, dr \]
\[ \phi_{j, \lambda_2}(\tilde{\rho}) = \int_{E(d)} |\tilde{\rho}_j \odot \Psi_{\lambda_2}(u, r)| \, du \, dr \]

\[ \Phi(\tilde{\rho}) = \{ \phi_p(\tilde{\rho}) \}_p \]
Scattering Properties

\[
\Phi(\tilde{\rho}) = \begin{pmatrix}
\|\tilde{\rho}\|_{L^1(\mathbb{R}^d)} \\
\|\tilde{\rho} \ast \psi_j, \cdot\|_{L^1(E(d))} \\
\|\|\tilde{\rho} \ast \psi_j, \cdot \otimes \Psi \lambda_2\|_{L^1(E(d))}
\end{pmatrix}_{j, \lambda_2}
\]

- Invariant to isometries… yes
- Multiscale family of functions… yes
- Lipschitz stable to diffeomorphisms… yes

Mallat, 2012
Augment Scattering

- From classical physics, we know $L^2$ energy coefficients are needed to expand the Coulomb potential, which is also present in the quantum chemistry molecular energy.

\[
\Phi(\tilde{\rho}) = \begin{pmatrix}
\|\tilde{\rho}\|_{L^p(\mathbb{R}^d)}^p \\
\|\tilde{\rho} \ast \psi_j, \cdot \|_{L^p(E(d))}^p \\
\|\|\tilde{\rho} \ast \psi_j, \cdot \| \ast \Psi \lambda_2 \|_{L^p(E(d))}^p
\end{pmatrix}
\]

$j, \lambda_2; \ p=1,2$
Quantum Chemistry Regression

- Two data bases \( \{ x_i, f(x_i) \}_i \) of planar, organic molecules with up to 20 atoms

- Regression on Fourier and scattering coefficients:
  \[
  \{ \phi_p \}_p = \begin{cases} 
  L^1/L^2 \text{ Fourier modulus coefficients} \\
  \text{or} \\
  L^1/L^2 \text{ Scattering coefficients}
  \end{cases}
  \]

- M-term sparse regression with greedy Orthogonal Least Squares computed on a training set:
  \[
  f_M(x) = \sum_{k=1}^{M} \alpha_k \phi_{p_k}(x)
  \]
M-term Error

$$\log_2 \mathbb{E}(|f(X) - f_M(X)|)$$

Scattering 1 Layer (Wavelets)

Scattering 2 Layers

Fourier

Coulomb
Numerical Results

- Mean absolute error $\mathbb{E}(|f(X) - f_M(X)|)$ in kcal/mol:

<table>
<thead>
<tr>
<th></th>
<th>Fourier</th>
<th>Coulomb</th>
<th>Scattering</th>
</tr>
</thead>
<tbody>
<tr>
<td>400 molecules</td>
<td>21.40</td>
<td>13.09</td>
<td>6.61</td>
</tr>
<tr>
<td>4000 molecules</td>
<td>18.61</td>
<td>4.16</td>
<td>2.05</td>
</tr>
</tbody>
</table>

- Scattering expansion terms:
  - First term: $\phi_{n_1}(\tilde{\rho}) = \int \tilde{\rho}(u) = \sum_k q_k = \text{total charge}$
  - Selected scales: Important geometric scales
Conclusion

• The scattering transform defines a representation that captures the fundamental properties of molecular energy.

• One can learn the energy through data and compute it fast.

• Can we learn other physical functionals?

http://www.di.ens.fr/~hirn/