Solid harmonic scattering for quantum energy regressions

• Louis THIRY, Data Team, ENS Paris (Advisor Stéphane Mallat)
Molecular property regression

Methane molecule

Numerical representation

\[ \Phi(CH_4) \rightarrow \begin{cases} 
- \text{positions + charges: } \{r_k, z_k\}_k \\
- \text{graph: } H-C-H \\
- \text{...} 
\end{cases} \]

Learn a regressor for the target property

\[ \Phi(CH_4) \xrightarrow{\tilde{f}} E(CH_4) \]
Database

Source:
- Experimental measures
- Numerical simulations

QM9 database
- 130,000 organic molecules
- up to 9 atoms of type C / O / N / Fl
- 13 properties: atomization energies, polarizability, dipole moment, thermal capacity ...
- computed using density functional theory (B3LYP)
- Error in energies $\sim 5$ kcal.mol$^{-1}$
Digits classification

Handwritten digits classification

Invariance to translation, stability to deformation

3 6 8 1 7 9 6 6 9 1
6 7 5 7 8 6 3 4 8 5
2 1 7 9 7 1 2 8 4 5
4 8 1 9 0 1 8 8 9 4

3 3 5 5
Instability to translation
for the usual metric
Instability to translation for the usual metric
Averaging

→ Creates stability to translations and deformations
Create stability to translations and deformations

Convolution with Gaussian kernel $\phi_J$:
- stable to geometric deformations
- Allows dimensionality reduction via subsampling
- lots of details are lost
Recover lost information

- Convolution + modulus with wavelets

\[ |x * \psi_{j,\theta}| \]

\[ \rightarrow \]

Gabor wavelets $\psi_{j,\theta}$

→ reveals details at different orientations and scales

- low-pass filtering (create stability)
- subsampling (to reduce dimension)
State of the art of non Deep-Network methods for image classification:
- 99.5% accuracy in MNIST digits classification
- 83% accuracy in CIFAR 10 image classification
Molecular energy regressions

Methane molecule energy

Invariance to translation and rotation:

Stability to deformation:
Quantum Energy Regression using Scattering Transforms
M. Hirn, N. Poilvert, S. Mallat (2016)

- Only planar molecule to have 2D images
- Create a fictitious image of the molecule using Isolated atomic densities

$$\tilde{\rho}_x(u) = \sum_{k=1}^{K} \rho^{a(k)}_{at}(u - p_k)$$

- Apply the technique used for images
Quantum Energy Regression using Scattering Transforms
M. Hirn, N. Poilvert, S. Mallat (2016)

- Compute scattering coefficients of the image

- QM2D DB: energies of 4357 planar molecules
- Perform linear regression on scattering coefficients
- MAE : 1.8 kcal/mol, RMSE : 2.7 kcal/mol
Solid harmonic wavelet scattering for predictions of molecule properties
G. Exarchakis, M. Eickenberg, M. Hirn, S. Mallat, L. Thiry (JCP, 2018)

From 2D to 3D

- implementation issues

- Gabor filters detect edges → do not seem to be relevant

- use knowledge from chemistry
Solid harmonic wavelet scattering for predictions of molecule properties

Input: atom positions and charges \( \{ r_k, z_k \} \_k \)

3D density channels

Scattering transform

First Order

Second Order

convolution

modulus

Multilinear Regression \( \hat{f}(x) \)
3D densities

- Isolated atomic densities replaced by Gaussians to avoid the CUSP in $0$

$$\rho(u) = \sum_{n} w_{n} g(u - r_{n})$$
Solid harmonic wavelet scattering for predictions of molecule properties
3D Solid harmonic wavelets

$$\psi_{l,m}(r, \theta, \phi) = K_l \left( e^{-r^2/2} r^l \right) Y_{l}^{m}(\theta, \phi)$$

polynomial gaussian

spherical harmonic

Lq pooling

$$c_{l,q} = \int_{\mathbb{R}^3} |\rho \ast \psi_l|^q$$
Regression with scattering transform

Input: atom positions and charges \( \{r_k, z_k\}_k \)

Scattering transform

First Order
convolution + modulus

Second Order

\( \| \cdot \|_q \)

Multilinear Regression

\( \tilde{f}(x) \)

Slice plot of the 3D density
Multi-linear regression

Linear regression

\[ E = \sum_{k=1}^{K} \alpha_k c_k + b : K \text{ coefficients to learn} \]

Bilinear regression

\[ E = \sum_{k,l=1}^{K} \alpha_{k,l} c_k c_l + b : K^2 \text{ coefficients to learn} \]

Factorized bilinear regression

\[ E = \left( \sum_k \alpha_k c_k \right) \cdot \left( \sum_k \beta_k c_k \right) + b : 2K \text{ coefficients to learn} \]
## Linear / Kernel / Neural network

<table>
<thead>
<tr>
<th></th>
<th>Linear regression</th>
<th>Kernel regression</th>
<th>Neural network</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Choice</strong></td>
<td>$\phi(x)$</td>
<td>$K(x, x')$</td>
<td>${\phi_\theta(x)}$</td>
</tr>
<tr>
<td><strong>$F(x)$</strong></td>
<td>$\langle a, \phi(x)\rangle_K$</td>
<td>$\langle\alpha, K(x, x_n)\rangle_N$</td>
<td>$f_{\theta'} (\phi_\theta(x))$</td>
</tr>
<tr>
<td><strong>Optimization</strong></td>
<td>$(a_k)_{1...K}$</td>
<td>$(\alpha_n)_{1...N}$</td>
<td>$(\theta, \theta')$</td>
</tr>
</tbody>
</table>
Solid harmonic scattering on QM9

<table>
<thead>
<tr>
<th>property</th>
<th>Linear regression</th>
<th>Tri-linear regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atomization Energy at 0 K (kcal.mol(^{-1}))</td>
<td>1.89</td>
<td>0.5</td>
</tr>
<tr>
<td>Thermal capacity (cal.mol(^{-1}).K(^{-1}))</td>
<td>0.10</td>
<td>0.049</td>
</tr>
</tbody>
</table>
Regression with scattering transform

Hyperparameters:
- width of the Gaussian representing the atoms
- wavelets
  - maximal scale $J$
  - maximal harmonic $L$
- $K$ densities
- $Q$ values for the $l$-$q$ pooling
- choice of the regression

Size of the descriptor

$$Q \times K \times J(J+1)/2 \times L \quad (= 384 \text{ for QM9})$$

Computational time
- grows linearly with the volume of the molecule
- QM9 (130K molecules) : 1 week on a good GPU
Interpretation

\[ g \ast \psi_{l,m}(u) = k \psi_{l,m}(\frac{u}{s}) \]

\[ \rho(u) = \sum_n g(u - r_n) \]

\[ |\rho \ast \psi_l|(u) = \left( \sum_{m=-l}^{l} \left| \sum_n \psi_{l,m}(u - r_n) \right|^2 \right)^{1/2} \]

Interference patterns of solid harmonic wavelets
Interference patterns
Interference patterns
Interference patterns
Scattering coefficients

\[ c_{l,q} = \int_{\mathbb{R}^3} |\rho \ast \psi_l|^q \]

- account for different patterns at small to large scales
- computed like convolutional network features
- not a many body features like distances, angles, dihedral angles
- not based on local atomic neighborhood
- well suited for periodic systems since convolutions are periodic
Conclusion

- Global descriptors
- Not a many-body feature, not based on neighborhoods
- Inspired from image classification technique
- Can be seen as a (shallow) convolutional network
- PyTorch code: [www.github.com/louity/pyscatharm](http://www.github.com/louity/pyscatharm)
- Forces: differentiable w.r.t. atomic positions (WIP)
- Test the ability to describe long range interactions → physical system with long range interaction
Questions ?