Solid Harmonic Wavelet Scattering for Molecular Energy Regression

Michael Eickenberg [1] **Georgios Exarchakis** [1] **Matthew Hirn** [2] **Stéphane Mallat** [1]

Abstract

We introduce a solid harmonic wavelet scattering representation, which is invariant to rigid movements and stable to deformations, for regression and classification of 2D and 3D images. Solid harmonic wavelets are computed by multiplying solid harmonic functions with Gaussian windows dilated to different scales. Invariant scattering coefficients are obtained by cascading such wavelet transforms with the complex modulus nonlinearity. We study an application of solid harmonic scattering invariants to the estimation of quantum molecular energies, which are also invariant to rigid movements and stable with respect to deformations. Linear or bilinear regressions over scattering invariants provide close to state of the art results over a database of organic molecules. [1]

1. Introduction

Prior information on invariants of a classification or regression problem can be used to reduce its dimensionality and derive precise estimations from a limited number of training examples. Many image or 3D data regression problems are invariant to translations and rotations of the input signal, and vary continuously with small deformations. For instance, this is the case for visual texture classification on satellite images, or for the regression physical quantities such as energies of isolated systems. In this paper we concentrate on applications to quantum molecular energy regressions, given the position and charges of the constituent atoms.

The main difficulty in building translation and rotation invariant image representations is to create a sufficiently rich set of invariants which does not lose important information and which is stable to small deformations. A small data deformation should produce a small modification of its invariant vector. Fourier transform representations may be invariant to translations and rotations but are not stable to deformations at high frequencies. Wavelet scattering representations are translation and potentially rotation invariant representations, which are stable to deformations (Mallat, 2011; Sifre & Mallat, 2013). They are computed by cascading wavelet transforms and modulus non-linearities, with a computational architecture similar to a convolutional neural network. Rotation invariant representations based on oriented Gabor wavelets have been studied for texture image analysis (Sifre & Mallat, 2013) and used to regress quantum energies of planar molecules (Hirn et al., 2016).

For 3D signals, computing scattering coefficients with oriented wavelets requires calculating a large number of 3D convolutions along orientation angles which uniformly sample the unit sphere. Such algorithms are computationally expensive. This paper introduces a faster algorithm based on solid harmonic functions, which are solutions of Laplace's equation. Section 2 defines solid harmonic wavelets by localizing these functions with Gaussian windows. The corresponding wavelet transform is covariant to translations and rotations, which means that wavelet coefficients are translated and rotated when the input signal is translated and rotated. Section 2.3 explains how to compute invariant descriptors by summing the modulus of these wavelet coefficients. Scattering coefficients are higher order invariants obtained by cascading two spherical harmonic wavelet transform and modulus operators which suppress their complex phase. This solid harmonic scattering provides a rich set of translation and rotation invariant descriptors which are computed with fast algorithms for 2D and 3D signals.

As a concrete application, this paper concentrates on employing solid harmonic scattering invariants for the regression of quantum energies of molecules. Estimating the ground state energy of atoms and molecules is one of the most fundamental and most studied topic in computational quantum mechanics. In principle, molecular energies can be computed from the atomic configuration by solving the Schrödinger equation. This is intractable in practise due to the curse of dimensionality, but it can be shown that all information pertaining to ground state energy of the molecule is contained in its 3D electronic density. Approximate solvers with density functional theory (DFT) are still computationally very intensive, with a polynomial complexity of order 4 in the number of electrons, which limits the molecule size and the number of simulations. This has motivated the study of energy regression with machine learn-

[1]This work was supported by the anonymous grant

ing methods, which compute fast regressions from previous computational results stored in databases. There are considerable industrial applications for this type of procedure, for the development of fast screening procedures for chemical, pharmaceutical and materials industries. Section 3 reviews several approaches to attack this problem from a machine learning perspective (Rupp et al., 2012; Montavon et al., 2012; Hansen et al., 2015; De et al., 2016). Section 4 explains how to compute regression of quantum energies of 3D molecules with linear and bilinear regressions from solid harmonic scattering coefficients. Numerical experiments on the GDB7-12 chemical dataset (see (Rupp et al., 2012) for a description) are presented in Section 5, with close to state of the art results.

2. Solid harmonic wavelet scattering

Given an input 2D or 3D signal ρ, we want to compute a vector of coefficients which are invariant to translations and rotations of ρ, and which are stable to small deformations. Our goal is to regress and approximately linearize functions $f(\rho)$ which are also invariant to translations and rotations. We use the rotation properties of spherical harmonics and define appropriate solid harmonic wavelets to construct such invariants with a scattering transform.

2.1. Solid harmonics in 2D and 3D

Solid harmonics are solutions of Laplace's equation $\Delta f = 0$, usually expressed in spherical coordinates. In 2D, interpreting $\mathbb{R}^2$ as the complex plane, we find that $z \mapsto z^\ell$ is a solution for all $\ell \in \mathbb{N}$ due to its holomorphicity[2]. Expressing this solution in polar coordinates gives

$$(r, \varphi) \mapsto r^\ell e^{i\ell\varphi},$$

revealing an ℓth- order polynomial in radius and a so-called circular harmonic with ℓ angular oscillations per circle.

Solving Laplace's equation in 3D spherical coordinates (r, ϑ, φ) gives rise to spherical harmonics, the eigenvectors of the Laplacian on the sphere. Imposing separability of azimuthal and elevation contributions yields

$$Y_\ell^m(\vartheta, \varphi) = C(\ell, m) P_\ell^m(\cos\vartheta) e^{im\varphi},$$

where P_ℓ^m is an associated Legendre polynomial and $C(\ell, m) = \sqrt{\frac{(2\ell+1)(\ell-m)!}{4\pi(\ell+m)}}$, for $\ell \geq 0$ and $-\ell \leq m \leq \ell$. They form an orthogonal basis of $\mathcal{L}^2$ functions on the sphere.

Analogous to the 2D case, 3D solid harmonics are then defined as

$$(r, \vartheta, \varphi) \mapsto \sqrt{\frac{4\pi}{2\ell+1}} r^\ell Y_\ell^m(\vartheta, \varphi).$$

2.2. Solid harmonic wavelets

We now define solid harmonic wavelets in 2D and 3D. A wavelet $\psi(u)$ is a function with zero sum, which is localized in the sense that it has a fast decay along u. Let $\psi_j(u) = 2^{-dj}\psi(2^{-j}u)$ be a normalized dilation of ψ by 2^j in dimension d. A multiscale wavelet transform of ρ computes convolutions with these dilated wavelets at all scales 2^j:

$$\{\rho \star \psi_j(u)\}_{j\in\mathbb{Z}} .$$

If ρ is translated by τ then $\rho \star \psi_j$ is also translated by τ. This is a translation covariance property. Let us denote by $\hat\rho(\omega)$ the Fourier transform of $\rho(u)$. The Fourier transform of these convolutions are $\hat\rho(\omega)\hat\psi(2^j\omega)$.

A wavelet is defined from a solid harmonic by multiplying it by a Gaussian, which localizes its support. In the 2D case

$$\psi_\ell(r, \varphi) = \frac{1}{\sqrt{(2\pi)^2}} e^{-\frac{1}{2}r^2} r^\ell e^{i\ell\varphi}.$$

For $\ell > 0$, these functions have zero integrals and are localized around the origin. In 2D frequency polar coordinates $\omega = \lambda(\cos\alpha, \sin\alpha)^T$, one can verify that the Fourier transform of this solid harmonic wavelet is

$$\hat\psi_\ell(\rho, \alpha) = (-i)^\ell e^{-\frac{1}{2}\lambda^2} \lambda^\ell e^{i\ell\alpha}.$$

The solid harmonic wavelet transform inherits the rotation properties of the solid harmonics. In 2D, the rotation of a solid harmonic incurs a complex phase shift. Let $R_\gamma \in SO(2)$ be a rotation of angle γ. We first observe that

$$R_\gamma \psi_{j,\ell}(r, \varphi) = \psi_{j,\ell}(r, \varphi - \gamma) = e^{-il\gamma}\psi(r, \varphi).$$

One can derive that for any signal ρ rotating the input signal leads to the same rotation applied to the output times a phase factor encoding the rotational angle:

$$R_\gamma \rho \star \psi_{j,\ell}(u) = e^{il\gamma} R_\gamma(\rho \star \psi_{j,\ell})(u).$$

This phase term is removed by a modulus. It results that the modulus of wavelet coefficients $U[j, \ell]\rho(u) = |\rho \star \psi_{j,\ell}(u)|$ is covariant to rotations:

$$U[j, \ell]\, R_\gamma \rho(u) = R_\gamma U[j, \ell]\rho(u).$$

In 3D, wavelet solid harmonics are defined by

$$\psi_{\ell,m}(r, \vartheta, \varphi) = \frac{1}{\sqrt{(2\pi)^3}} e^{-\frac{1}{2}r^2} r^\ell\, Y_\ell^m(\vartheta, \varphi).$$

[2] Real and imaginary parts of holomorphic functions are harmonic - their laplacian is 0

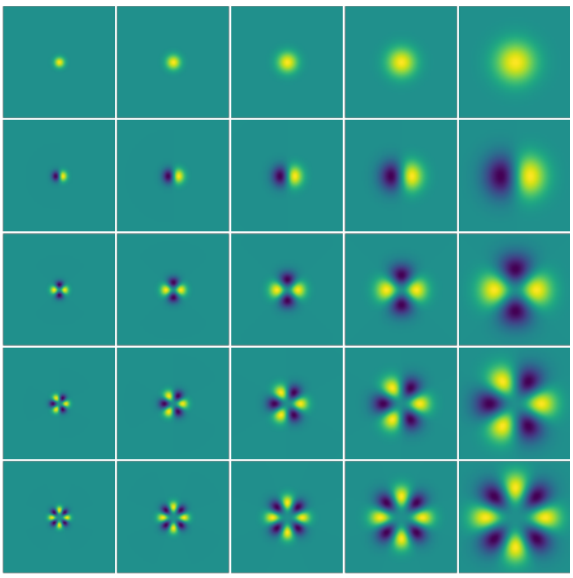

Figure 1. Real parts of 2D solid harmonic wavelets $\psi_{\ell,j}(u)$. The ℓ parameters increases from 0 to 4 vertically where as the scale 2^j increases from left to right. Cartesian slices of 3D spherical harmonic wavelets yield similar patterns.

We write $\psi_{\ell,m,j}$ its dilation by 2^j. In 3D frequency polar coordinates $\omega = \lambda(\cos\alpha\cos\beta, \cos\alpha\sin\beta, \sin\alpha)^T$, its Fourier transform is

$$\hat{\psi}_{\ell,m}(\lambda,\alpha,\beta) = \frac{4\pi(-i)^\ell}{\sqrt{(2\pi)^3}} e^{-\frac{1}{2}\lambda^2}\,\lambda^\ell\, Y_\ell^m(\alpha,\beta).$$

The 3D covariance to rotation is more involved due to the grouping of rotational subspaces in the spherical harmonics. The asymmetry of the azimuthal and elevation components of the spherical harmonics requires them to be treated differently. In order to obtain a rotation covariance property, it is necessary to sum the energy over all m indices. We shall define the modulus of wavelet spherical harmonics in 3D by

$$U[\ell,j]\rho(u) = \left(\sum_{m=-\ell}^{\ell} |\rho \star \psi_{\ell,m,j}(u)|^2\right)^{1/2}.$$

Similarly to the 2D case, one can prove that this summation over m defines a wavelet transform modulus which is covariant to 3D rotations. For a general rotation $R \in SO(3)$

$$U[j,\ell]\,R\rho = R\,U[j,\ell]\rho.$$

2.3. Solid harmonic scattering invariants

We showed that a wavelet transform modulus computed with solid harmonic wavelets is covariant to translations and rotations in 2D and 3D. One can thus compute a translation and rotation invariant representation by summing these coefficients over the spatial variable u at some power q. Since $U[j,\ell]\rho(u)$ is obtained by a wavelet scaled by 2^j, it is a regular function and an approximate sum is obtained by subsampling u at intervals $2^{j-\alpha}$ where α is an oversampling factor typically equal to 1. In 2D and 3D we thus define for any (j_1,ℓ_1) and exponent q:

$$S[j_1,\ell_1,q]\rho = \sum_u \left|U[j_1,\ell_1]\rho(2^{j_1-\alpha}u)\right|^q$$

Translating or rotating ρ does not modify $S[j_1,\ell_1,q]\rho$. Let $J > 0$ denote the number of scales j_1, and $L > 0$ the number of angular oscillations parameter ℓ_1. We choose $q \in \{1,2\}$ which yields $2JL$ invariant coefficients.

The summation eliminates the variability of the $U[j_1,\ell_1]\rho(u)$ along u. A scattering transform computes complementary invariants by calculating the variations of $U[j_1,\ell_1]\rho(u)$ by retransforming this function of u with a second wavelet transform modulus operator $U[j_2,\ell_2]$ for different (j_2,ℓ_2). Since $U[j_1,\ell_1]$ and $U[j_2,\ell_2]$ are covariant to translations and rotations, the product $U[j_2,\ell_2]\,U[j_1,\ell_1]\rho(u)$ is also covariant to translations and rotations of ρ. The variable u can be subsampled at intervals $2^{j_2-\alpha}$ because of the regularization produced by the second wavelet transform at the scale 2^{j_2}. Summing over the subsampled spatial variable u yields another set of invariants called second order scattering invariants.

$$S[j_1,\ell_2,j_2,\ell_2,q]\rho = \sum_u \left|U[j_2,\ell_2]\,U[j_1,\ell_1]\rho(2^{j_2-\alpha}u)\right|^q.$$

The coefficients are computed only for $j_2 > j_1$ because otherwise one can verify (Mallat, 2011) that the amplitude of these invariant coefficients is negligible. The total number of computed second order invariants is thus $2L^2J(J-1)/2$.

Because wavelets are localized and regular functions, the small deformation of a wavelet defines a similar function. Besides translation and rotation invariance, one can prove that a scattering transform is Lipschitz continuous to deformations. This means that if $\rho(u)$ is deformed by a "small" diffeomorphism applied to u, then the scattering vector is modified by the addition of an error whose Euclidean norm is bounded by a constant multiplied by this small amplitude (Mallat, 2011). This property is particularly important to linearly regress functions which are also stable to deformations.

3. Quantum molecular energy regression

Instead of performing costly ab initio calculations for every new molecule, machine learning algorithms compute

regressions from previously calculated energies which are stored in databases. The next section describes the invariance and stability properties of a quantum energy functional, and we then review state of the art machine learning approaches.

3.1. Molecular regression invariances

For quantum energy regression, many invariance and stability properties of the energy function are known and described below.

A molecule containing K atoms is entirely defined by its nuclear charges z_k and its nuclear position vectors r_k indexed by k. Denoting by x the state vector of a molecule, we have

$$x = \{(r_k, z_k) \in \mathbb{R}^3 \times \mathbb{R} : k = 1, \dots, K\}.$$

Since the target value that we are trying to regress is a scalar representing a physical energy, we know that:

Permutation invariance The energy is invariant to the permutation of the indexation of atoms in the molecule.

Isometry invariance The energy is invariant to translations, rotations, and symmetry of the molecule and hence to any orthogonal operator.

Deformation stability The energy is differentiable with respect to the distances between atoms.

Multiscale interactions The energy has a multiscale structure, with highly energetic covalent bonds between neighboring atoms, and weaker energetic exchanges at larger distances, such as Van-der-Waals interactions.

The deformation stability stems from the fact that a small deformation of the molecule induces a small modification of its energy. The primary difficulty is to construct a representation which satisfies these four properties, while simultaneously containing a rich enough set of descriptors to accurately regress the atomization energy of a diverse collection of molecules.

Density functional theory (DFT) provides efficient numerical algorithms to compute an approximation of the quantum energies of molecules, with a precision of 1kal to 2kcal per mol. Hohenberg and Kohn proved in (Hohenberg & Kohn, 1964) that the molecular energy E can be written as a functional of the electron density $\bar{\rho}(u) \geq 0$ which specifies the density of electronic charge at every point $u \in \mathbb{R}^3$. The minimization of $E(\bar{\rho})$ over a set of electron densities $\bar{\rho}$ leads to the calculation of the ground state energy

$$f(x) = E(\bar{\rho}_x) = \inf_{\bar{\rho}} E(\bar{\rho}) \, . \tag{1}$$

The functional E is separated into contributions from electron kinetic energy, electron-nucleus Coulomb attraction, electron-electron Coulomb repulsion and an exchange-correlation energy which carries all quantum effects. This last term is approximted which introduces errors. Solving the variational problem (1) with this approximate energy functional is computationally intensive and scales polynomially with an order 4 in the number of electrons.

3.2. Machine Learning State of the Art

In recent years, machine learning methods have gained considerable traction to estimate quantum molecular energies. The first approaches used Coulomb matrices, which encode pairwise nucleus-nucleus repulsion forces for each molecule (Rupp et al., 2012; Montavon et al., 2012; Hansen et al., 2013; Montavon et al., 2013). They are then used to interpolate the target chemical property using e.g. kernel Ridge regression. A drawback of this approach is that Coulomb matrices are not invariant to permutations of indices of atoms in the molecules, which leads to regression instabilities, despite the use of stabilization procedures by perturbing the column norm sorting with random values to break symmetries (Montavon et al., 2012). On the GDB7-13 dataset of small molecules (see (Montavon et al., 2013) for a description) , these methods have achieved a documented mean absolute error (MAE) atomization energy of 3.07 kcal/mol (Hansen et al., 2013).

Continued development has led to improvements with bag-of-bonds descriptors (Hansen et al., 2015), which groups matrix entries according to bond type (GDB7-12 MAE 1.5kcal/mol), or with fixed-length smooth bond-distance histograms (Collins et al., 2017) (MAE 1.19kcal/mol on GDB7-12).

Another successful type of descriptors are the smooth overlap of atomic position (SOAP) descriptors(De et al., 2016), which directly compute a similarity kernel measure between molecules built out of descriptors of the neighborhoods of every atom in the molecule. These achieve an error rate of 0.92 kcal/mol on GDB7-13. Their inconvenience is that they cannot take into account long range interactions.

The abovementioned machine learning algorithms compute a kernel ridge regression from these descriptors, with a Gaussian or Laplacian kernel. These kernels induce local interpolation between training descriptors to predict the target. As long as the descriptor is powerful enough, increasing amounts of data will improve the predictions, but will also increase the size of the dual representation and require each testing point to be compared to every training point for a prediction. Herein lies a potential drawback of such non-parametric methods. They do not distill a fixed-size decision procedure from the available data, which would

guarantee constant execution time at testing time and be more amenable to analysis.

Recent deep tensor networks (Schütt et al., 2017) combine pairwise distance matrix representations with a deep learning type approach. However, deep networks require large data bases for training, which are potentially not available to the desired chemical accuracy.

Gabor wavelet scattering transforms have been applied to quantum energy regression of planar molecules (Hirn et al., 2016) because they can define representations which satisfy the invariance and stability properties of molecular energies given in Section 3.1. At order 1 this is achieved by averaging over the coefficients extracted at different orientations, which creates an invariant to rotations out of a set of descriptors that otherwise vary with rotation. At order 2 invariant coefficients are extracted by averaging all couples of orientations of first layer and second layer wavelet modulus integrals that are at the same angle to each other. However, the 3D analog of this approach requires using many 3D Gabor wavelets whose orientations define a regular and dense sampling of the unit sphere. It thus has an important computational complexity which is considerably reduced by using solid harmonic wavelets.

4. Solid Harmonic Scattering for energy regression

We study applications of solid harmonic scattering coefficients to regress quantum molecular energies, and show that this representation approximatively linearizes the energy. Improved results are obtained with a bilinear fit which achieves close to state-of-the-art performance on the GDB7-12 dataset. From a chemical point of view, wavelet solid harmonics also have similarities with the Gaussian basis sets typically used in DFT calculations. In DFT, an electronic density is represented as the square sum of a linear combination of cartesian versions of Gaussian solid harmonics centered around the atoms.

4.1. Permutation invariant molecule embeddings

In the spirit of the electron density we create a 3D spatial embedding of the positional and charge data describing the molecule x as a sum of non-interacting electronic densities, creating a zeroth-order approximation of true electron densities in the case it is sufficiently localized. For K atoms placed at $\{r_k\}_{k=1}^K$ having charges $\{z_k\}_{k=1}^K$, input images take the general form

$$\rho_x(r) = \sum_{k=1}^{K} c(z_k) g(r - r_k),$$

where g is a Gaussian "blob" roughly representing an electron density localized around the nucleus, and $c(z_k)$ is a vector-valued "electronic channel" function, which can encode different aspects of the atomic structure, making them immediately accessible to the invariant operators. Here we choose three channels:

1. nuclear charge z_k;
2. valence shells v_k;
3. core electronic shells $z_k - v_k$,

and can thus write $c(z_k) = (z_k, v_k, z_k - v_k)^T$ The molecule embedding verifies $\int \rho_x(u)\mathrm{d}u = \sum_k (z_k, v_k, z_k - v_k)^T$, thus counting the total number of nucleus charges and valence and core electrons. By construction, this approximate density is invariant to permutations of atom indices k.

4.2. Rotation and translation invariant scattering descriptors

The molecule representation ρ_x is invariant to permutations of atom indices but it is not invariant to isometries (rigid body transformations lead to density in different positions), nor does it separate scales. These missing invariances and the separation of scales into different channels are obtained by computing its scattering representation with solid harmonic wavelets.

Letting $J \in \mathbb{N}$ be the number of scales and $L \in \mathbb{N}$ the maximal spherical harmonic order, we extract

$$S[j_1, \ell_1, q]\rho_x \quad \text{and} \quad S[j_1, \ell_1, j_2, \ell_2, q]\rho_x$$

For $j_1, j_2 \in \{0, \dots, J-1\}$ with $j_1 < j_2$ and $\ell_1, \ell_2 \in \{0, \dots, L-1\}$ and $q \in \{1, 2\}$. The relation of the exponent with chemical properties has been shown in (Hirn et al., 2016): The electron-electron Coulomb repulsion term can be recovered from wavelet modulus integrals of the true electron density using $q = 2$. Adding $q = 1$ adds a term that is approximately linear in number of charges modulo destructive interferences of wavelets that occur when the support of one wavelet scale covers several atoms.

4.3. Regressions on invariant descriptors

Let $p = (j_1, \ell_1, q)$ or $p = (j_1, \ell_1, j_2, \ell_2, q)$ be a multiindex that enumerates all valid scattering paths. The last machine learning step is a supervised regression of quantum energy from $S\rho = \{S[p]\rho_x\}_p$, given a training data base. The goal is to establish a link between the extracted invariants and the predictive target, the atomization energy. We show here that solid harmonic scattering invariants of ρ_x linearize the atomization energy by expressing it as a linear combination thereof.

$$\tilde{f}(x) = \tilde{E}(\rho_x) = b + \langle S\rho_x, w \rangle \tag{2}$$

We use Ridge regression to fit the linear model on the scattering coefficients. Ridge regression weights are the solution of

$$\text{argmin}_w \frac{1}{2N} \sum_{n=1}^{N} (E_n - b - \langle S\rho_{x_n}, w\rangle)^2 + \frac{\lambda}{2}\|w\|^2$$

Extending this affine function to a quadratic permits pairwise multiplicative interactions between invariants and can be seen as the creation of new linearizing features:

$$\tilde{E}(\rho_x) = b + \langle S\rho_x, w\rangle + \langle S\rho_x, WS\rho_x\rangle$$

We fit a quadratic functional to the regress the energies from the scattering coefficients, which we frame as a bilinear neural network with multiplicative gating units (Memisevic & Hinton, 2010).

For a generic vector $s \in \mathbb{R}^d$ and $i \in \{1, 2\}$ (indexing either side of a bilinear product), let

$$h^i(s) = b^i + W^i s.$$

Further, let

$$G(s) = b + \langle h_k^1(s), Dh_k^2(s)\rangle,$$

where $D = \text{diag}(d)$ is a diagonal matrix, with the entries $d \in \mathbb{R}^\kappa$. This structure makes the estimation of any inhomogeneous polynomial of order 2 theoretically possible. We fit this network to the target energy by minimizing a mean squared error loss using stochastic gradient descent:

$$\text{argmin}_{W^1, W^2, d, b^1, b^2, b} \sum_{n=1}^{N} (E_n - G(S\rho_{x_n}))^2$$

5. Numerical Experiments on Chemical Databases

We performed atomization energy regression on the GDB7-12 molecules dataset, consisting of 7165 molecules of up to 23 atoms among H, C, O, N and S. They are maximally around 10Å long. We performed rigid affine coordinate transforms to align each molecule with its principle axis, making it possible to fit every molecule in a box of dimensions $12.8\text{Å} \times 9.6\text{Å} \times 6.4\text{Å}$. We discretize this space by creating a grid of stepsize 0.1333.The Gaussians $g(r)$ have small widths on the order of the grid step ($\sigma = 0.1\text{Å}$), in order to interpolate between grid points in case the position does not fall exactly onto a grid point. Each Gaussian is renormalized to sum to exactly $c(z_i)$ before summing all Dirac contributions to obtain a multi-channel density image. c is chosen to represent *full* nucleus as well as *valence* and *core* electronic charges. Density images are padded with zeros in order to be able to perform convolutions in Fourier space. After this operation, each density image is a multidimensional numerical array of size $3 \times 144 \times 144 \times 96$, where the first dimension indicates the input channel. Solid harmonic wavelets are chosen for $l \in \{0, 1, 2, 3\}$ and $\sigma = 2^{-J+1}$.

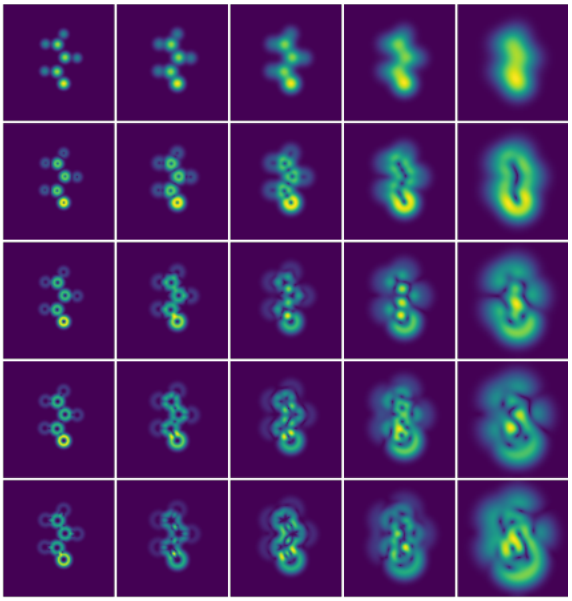

Figure 2. Solid harmonic wavelet moduli of a molecule. The interference patterns at the different scales are reminiscent of molecular orbitals obtained in e.g. density functional theory.

We extract the solid harmonic wavelet invariants using GPUs. When appropriate (permitted by the filter bandwidth), the wavelet modulus images are subsampled to a lower resolution in order to speed up calculations.

Discrete grids introduce sampling errors which are countered by data augmentation.

5.1. Results

The resulting representation is used to fit a linear model using ridge regression ($\lambda = 10^{-10}$ fixed) and a bilinear (gating) neural network as described in section 4.3. For the linear model (L-Scat), we standardize the data to be mean free and unit variance. The average of the mean absolute error (MAE) over 5 folds of the for both first and second order coefficients is 2.2 kcal/mol. Similarly, for the bilinear model (B-Scat), we standardize the coefficients to be mean free and unit variance. Additionally, we perform dimensionality reduction and whitening of the data set using principle component analysis while retaining 50% of the dimensions of the invariant representation. The average validation error, in terms of mean absolute error, over

	RSCM	BoB	SOAP	DTN	CBoB	L-Scat	B-Scat
MAE	3.1	1.5	0.9	1	1.2	2.2	1.2

Table 1. Comparison of solid harmonic invariant regression with related work. (RSCM: Random Sorted Coulomb Matrix, BoB: Bag of Bonds, SOAP: smooth overlap of atomic positions, DTN: deep tensor networks, CBoB: Continuous bag of bonds, L-Scat: solid harmonic invariants followed by ridge regression, B-Scat: solid harmonic invariants followed by quadratic regression

5 data folds is 1.2kcal/mol.

Comparing to the succession of state-of-the-art methods described in table 1, both the linear and the bilinear regressions on solid harmonic scattering invariants lie in the same performance regime, with mean absolute error of 1.2 and 2.2 kcal/mol respectively. It should be noted that these are preliminary results obtained with little to no parameter tuning.

5.2. Discussion

While drawbacks of computational intensity are largely mitigated by the employment of GPUs, the method we propose has a number of advantages with respect to existing approaches. Compared to Coulomb matrix approaches it is guaranteed to be perfectly invariant to atom indexing and thus stable with respect to small deformations, which can cause a switch in column ordering in Coulomb matrices. Furthermore, similarly to (Collins et al., 2017), our approach provides constant size descriptors that do not scale quadratically in dimensionality with the size of the largest molecule. The SOAP similarity measure (De et al., 2016) is at least linear in atom number due to the comparison of all atomic neighborhoods. Our approach scales logarithmically with the size of the molecule, due to potentially larger box size to fit them. Finally, compared to deep tensor networks, which propose the learning of many parameters, here we create a powerful neural-network-type representation but with predefined and fixed filters. Since they linearize the predictive target of atomization energy, coefficients of learned models become interpretable and can be studied as to their role in describing molecular energies.

Introducing multiplicative interactions between solid harmonic wavelet invariants further improves the performance on the energy regression task, achieving near state of the art performance. In fact, the bilinear neural network on scattering coefficients likely proposes the only domain agnostic method with this level of accuracy.

6. Conclusion

We have presented a general method to extract rotation and translation invariant descriptors from 2D and 3D images. We use wavelet transform modulus integrals and second order scattering invariants with solid harmonic wavelets, which we introduce. Translation invariance is guaranteed by the spatial convolution operator. Rotation invariance is due to the specific design of the wavelets.

We show that solid harmonic wavelet scattering can be applied successfully to 3D quantum chemical energy regression problems. Molecular properties are independent of position and orientation of the molecule. Solid harmonic scattering invariants are shown to linearize this energy. Further, pairwise multiplicative interactions between solid harmonic scattering invariants permit close to state-of-the-art error rates.

Up to now, no optimization of the filters has been attempted. Recently, a 2D harmonic network, in which the radial part of the filters was learned, has been introduced for image classification (Worrall et al.). This approach can be introduced into the representation presented here, yielding flexibility of wavelet shape while retaining the invariance property.

References

Collins, Christopher R., Gordon, Geoffrey J., von Lilienfeld, O. Anatole, and Yaron, David J. Constant size molecular descriptors for use with machine learning. *arXiv*, 2017.

De, Sandip, Bartk, Albert P., Csnyi, Gbor, and Ceriotti, Michele. Comparing molecules and solids across structural and alchemical space. *Phys. Chem. Chem. Phys.*, 18(20):13754–13769, 2016. ISSN 1463-9084. doi: 10.1039/C6CP00415F.

Hansen, Katja, Montavon, Grégoire, Biegler, Franziska, Fazli, Siamac, Rupp, Matthias, Scheffler, Matthias, von Lilienfeld, O. Anatole, Tkatchenko, Alexandre, and Müller, Klaus-Robert. Assessment and validation of machine learning methods for predicting molecular atomization energies. *Journal of Chemical Theory and Computation*, 9(8):3404–3419, 2013. doi: 10.1021/ct400195d.

Hansen, Katja, Biegler, Franziska, Ramakrishnan, Raghunathan, Pronobis, Wiktor, von Lilienfeld, O. Anatole, Mller, Klaus-Robert, and Tkatchenko, Alexandre. Machine learning predictions of molecular properties: Accurate many-body potentials and nonlocality in chemical space. *The Journal of Physical Chemistry Letters*, 6(12):2326–2331, 2015. doi: 10.1021/acs.jpclett.5b00831. URL http://dx.doi.org/10.1021/acs.jpclett.5b00831. PMID: 26113956.

Hirn, Matthew, Mallat, Stéphane, and Poilvert, Nicolas. Wavelet scattering regression of quantum chemical energies. *arXiv*, 2016.

Hohenberg, P. and Kohn, W. Inhomogeneous electron gas. *Phys. Rev.*, 136:B864–B871, Nov 1964. doi: 10.1103/PhysRev.136.B864. URL http://link.aps.org/doi/10.1103/PhysRev.136.B864.

Mallat, Stéphane. Group Invariant Scattering. November 2011. URL http://arxiv.org/abs/1101.2286.

Memisevic, Roland and Hinton, Geoffrey E. Learning to represent spatial transformations with factored higher-order boltzmann machines. *Neural Computation*, 22(6): 1473–1492, 2010.

Montavon, Grégoire, Hansen, Katja, Fazli, Siamac, Rupp, Matthias, Biegler, Franziska, Ziehe, Andreas, Tkatchenko, Alexandre, von Lilienfeld, O. Anatole, and Müller, Klaus-Robert. Learning invariant representations of molecules for atomization energy prediction. In Bartlett, P., Pereira, F.C.N., Burges, C.J.C., Bottou, L., and Weinberger, K.Q. (eds.), *Advances in Neural Information Processing Systems 25*, pp. 449–457. 2012.

Montavon, Grégoire, Rupp, Matthias, Gobre, Vivekanand, Vazquez-Mayagoitia, Alvaro, Hansen, Katja, Tkatchenko, Alexandre, Müller, Klaus-Robert, and von Lilienfeld, O Anatole. Machine learning of molecular electronic properties in chemical compound space. *New Journal of Physics*, 15(9):095003, 2013. URL http://stacks.iop.org/1367-2630/15/i=9/a=095003.

Rupp, M., Tkatchenko, A., Müller, K.-R., and von Lilienfeld, O. A. Fast and accurate modeling of molecular atomization energies with machine learning. *Physical Review Letters*, 108:058301, 2012.

Schütt, Kristof T., Arbabzadah, Farhad, Chmiela, Stefan, Müller, Klaus R., and Tkatchenko, Alexandre. Quantum-chemical insights from deep tensor neural networks. *Nature Communications*, 8:13890 EP –, Jan 2017. URL http://dx.doi.org/10.1038/ncomms13890. Article.

Sifre, Laurent and Mallat, Stéphane. Rotation, scaling and deformation invariant scattering for texture discrimination. In *CVPR*, pp. 1233–1240. IEEE Computer Society, 2013. ISBN 978-0-7695-4989-7. URL http://dblp.uni-trier.de/db/conf/cvpr/cvpr2013.html#SifreM13.

Worrall, Daniel E., Turmukhambetov, Daniyar, Garbin, Stephan J., and Brostow, Gabriel J. Harmonic networks: Deep translation and rotation equivariance.