Feature Extraction in Deep Learning and Image Processing

Yiran Li

Applied Mathematics, Statistics, and Scientific Computation
Norbert Wiener Center
Department of Mathematics
University of Maryland, College Park

Norbert Wiener Center for Harmonic Analysis and Applications

Outline

Outline

Overview

My dissertation consists of the following topics:

- **Approximation Properties of Neural Networks**
- Maximal Function Pooling in Convolutional Sparse Coding
- **Quantum Energy Regression Using Gabor Transform**
- Detection of Epithelial versus Mesenchymal Regions in 2D Images of Tumor Biopsies Using Shearlets

Due to the time constraint, I will discuss the topics in bold in this presentation.

Outline

Approximation Properties of Neural Networks

Deep Neural Networks (DNNs) and deep learning algorithms have achieved successful results in many areas of machine learning, and there has been growing interest in the theoretical study of DNNs.
Some important topics in the theoretical analysis of neural networks include:

1. Specification of the network topology to obtain certain approximation properties of functions;
2. The stability analysis of the network;
3. Study of the training algorithms to obtain desired convergence rate.

Previous Works

1. Chui, C. K., Li, X., & Mhaskar, H. N. (1994). Neural networks for localized approximation. *Mathematics of Computation*, 63(208), 607-623.
2. Shaham, U., Cloninger, A., & Coifman, R. R. (2016). Provable approximation properties for deep neural networks. *Applied and Computational Harmonic Analysis*.
3. Bölcskei, H., Grohs, P., Kutyniok, G., & Petersen, P. (2017). Optimal approximation with sparsely connected deep neural networks. *preprint arXiv:1705.01714*.
4. Balan, R., Singh, M., & Zou, D. (2017). Lipschitz properties for deep convolutional networks. *arXiv preprint arXiv:1701.05217*.
5. Jacobs, R. A. (1988). Increased rates of convergence through learning rate adaptation. *Neural networks*, 1(4), 295-307.

Universal Approximation Theorem

The most well-known early result is by Cybenko in 1989 states that:

> *Any continuous function can be uniformly approximated by a continuous neural network having only one internal hidden layer and with arbitrary continuous sigmoidal nonlinearity.*

Theorem (Cybenko, 1989)

Let σ be any continuous discriminatory sigmoidal function. Then the finite sums

$$G(x) = \sum_{k=1}^{n} c_k \sigma(w_k \cdot x + b_k), \tag{1}$$

are dense in $C(I_d)$, where I_d is the unit cube in $\mathbb{R}^d$.

Here σ is the sigmoidal activation function, defined as $\sigma(u)$ with $\lim_{u\to-\infty} \sigma(u) = 0$ and $\lim_{u\to\infty} \sigma(u) = 1$.

Norbert Wiener Center
for Harmonic Analysis and Applications

Fourier Approximation

The number of neurons and number of layers required to yield an approximation rate of a given quantity is not addressed.
The first work to address this problem is by Barron in 1991:

Theorem (Fourier Approximation, Barron, 1991)

Given a function $f : \mathbb{R}^m \to \mathbb{R}$ with

$$C_f = \int_{\mathbb{R}^m} |\omega||\hat{f}(\omega)|d\omega < \infty, \tag{2}$$

there exsists a single layer artificial neural network (ANN) of N sigmoid units, s.t. the output of the network f_N satisfies

$$\|f - f_N\|_2 \leq \frac{c_f}{\sqrt{N}}, \tag{3}$$

with c_f proportional to C_f.

Approximation Properties with Wavelets

Smooth functions defined on low-dimensional subspace can be high dimensional in ambient space. The goal is to find approximation rate related to the dimension of the manifold, not the ambient space.

Theorem (Cloninger, Coifman, Shaham, 2015)

Let $\Gamma \subset \mathbb{R}^m$ be a smooth d-dimensional manifold, $f \in L^2(\Gamma)$ and let $\epsilon > 0$ be an approximation level. Then if a network has at least 4 layers, there exists a sparsely-connected neural network with N total units where $N = C_\Gamma m + C'_\Gamma d N_{f,\epsilon}$, computing function f_N such that

$$\|f - f_N\|_2^2 < \epsilon, \tag{4}$$

where $N_{f,\epsilon}$ depends on the complexity of f in terms of its local wavelet representation, and C_Γ on the curvature and dimension of the manifold Γ.

If $f \in C^2(\Gamma)$ and has bounded Hessian, then

$$\|f - f_N\|_\infty = \mathcal{O}(N^{-\frac{2}{d}}). \tag{5}$$

Construction of Wavelet Frame

- A wavelet like frame of $\mathbb{R}^d$ is constructed in which the frame elements are built using rectified linear units. A rectified linear unit (ReLU) is defined as
$$rect(x) = max\{0, x\}. \tag{6}$$
- Define a trapezoid-shaped function t: $\mathbb{R} \to \mathbb{R}$ by
$$t(x) = rect(x+3) - rect(x+1) - rect(x-1) + rect(x-3). \tag{7}$$
- Define the scaling function $\phi : \mathbb{R}^d \to \mathbb{R}$ by
$$\phi(x) = rect\left(\sum_{j=1}^{d} t(x_j) - 2(d-1)\right), \tag{8}$$
and normalize it so that the integral of ϕ is 1.
- Let $S_k(x, b) = 2^k \phi(2^{\frac{k}{d}}(x - b))$. Define the mother wavelet as $D_k(x, b) = S_k(x, b) - S_{k-1}(x, b)$. The wavelets are defined as $\psi_{k,b}(x) = 2^{-\frac{k}{d}} D_k(x, b)$.

Norbert Wiener Center for Harmonic Analysis and Applications

Construction of Wavelet Frame

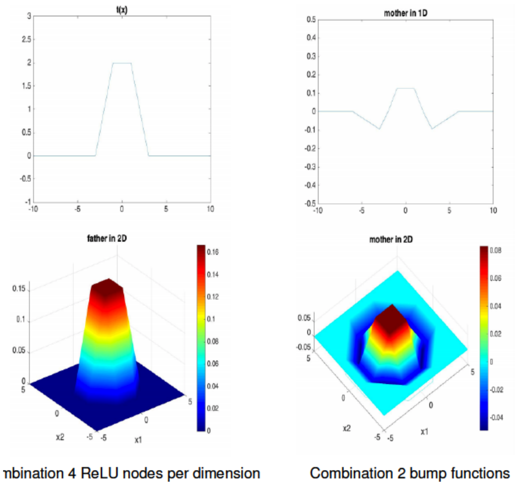
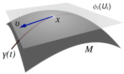


Figure: Wavelets constructed from ReLUs (Source: U. Shaham, A. Cloninger, R. R. Coifman, 2015.)

Approximation of Functions on Manifold

- Given a d-dimensional manifold $\Gamma \subset \mathbb{R}^m$, cover Γ by set of pairs $\{(U_i, \phi_i)\}_{i=1}^{C_\Gamma}$. Here ϕ_i is the orthogonal projection from U_i onto H_i, where H_i is the hyperplane tangent to Γ at x_i.



- Use the corresponding partition of unity $\{\eta_i\}$ to define

$$f_i(x) = f(x)\eta_i(x). \tag{9}$$

Define $\hat{f} \in \mathbb{R}^d$ as

$$\hat{f}_i(x) = \begin{cases} f_i(\phi_i^{-1}(x)), & x \in \phi_i(U_i), \\ 0, & otherwise. \end{cases} \tag{10}$$

- For all $x \in \Gamma$, we have

$$\sum_{i:x\in U_i} \hat{f}_i(\phi_i(x)) = f(x). \tag{11}$$

Assuming $\hat{f}_i \in L^2(\mathbb{R}^d)$, it can be expanded using wavelet frame.

Gabor System in Machine Learning

- Current study of the approximation properties of neural networks is mainly in terms of affine transformations (e.g., wavelet transform).
- We extend the study of approximation properties of neural networks to functions in the modulation space.
- Gabor transform, or the short-time Fourier transform, arises naturally in analysis of 1D data such as speech and music.
- There have been algorithms developed for voiced-unvoiced speech discrimination in noise, where short segments of speech are modeled as a sum of basis functions from a Gabor dictionary.
- Gabor filters and Gabor wavelets are widely used as convolutional kernels for neural networks for 2D image processing.

Motivation for Gabor System in Neural Networks

- There are cortical receptive fields that best respond to signals with orientation. They also capture spatial frequency information.
- Two-dimensional spatial linear filters are constrained by general uncertainty relations. The theoretical lower limit for the uncertainty is achieved by Gabor functions.
- Gabor filters have been used as units of a neural network to model the profile of cortical receptive fields (Daugman, 1988).

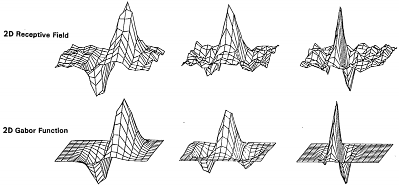


Figure: Illustration of experimentally measured 2D receptive-field profiles of three simple cells in cat striate cortex (top row). Each plot shows the excitatory or inhibitory effect of a small flashing light or dark spot on the firing rate of the cell, as a function of the (x, y) location of the stimulus. Best fit using Gabor functions (second row).

Approximation Properties of Neural Networks

We design a novel type of neural network and prove its theoretical approximation rate to functions f based on the network topology. Informally, we show that

Theorem (Informal)

Let $f \in L^2(\mathbb{R})$, and let $\delta > 0$ be an approximation level. There exists a 4-layer sparsely-connected neural network with N units where $N = N(f, \delta)$, computing f_N with

$$\|f - f_N\|_\infty \leq \delta.$$

- We demonstrate a method to build a Gabor frame of $L^2(\mathbb{R})$ based on a type of activation function in neural networks: rectified linear units.
- We construct a 4-layer neural network based on the Gabor frame and demonstrate its approximation properties.

Gabor System

We first introduce the notion of Gabor system. Let the time shift T of a function $g \in L^2(\mathbb{R}^d)$ by $x \in \mathbb{R}^d$ be defined by

$$T_x g(t) = g(t - x),$$

and let the modulation of g by $\omega \in \mathbb{R}^d$ be defined by

$$M_\omega g(t) = e^{2\pi i \omega \cdot t} g(t).$$

Definition

A Gabor system $G(g, \alpha, \beta)$ is the set of time-frequency shifts of a non-zero window function $g \in L^2(\mathbb{R}^d)$ with lattice parameters $\alpha, \beta > 0$:

$$\{T_{\alpha k} M_{\beta n} g : k, n \in \mathbb{Z}^d\}.$$

Frame

We introduce the notion of a frame. A frame can be thought of as a generalization of a basis that may be linearly dependent.

Definition

A sequence $\{e_j, j \in J\}$ in a separable Hilbert space $\mathcal{H}$ is called a *frame* if there exist positive constants $A, B > 0$ such that for all $f \in \mathcal{H}$

$$A\|f\|^2 \leq \sum_{j \in J} |\langle f, e_j \rangle|^2 \leq B\|f\|^2.$$

Any two constants A, B where $0 < A \leq B < \infty$ satisfying the above statement are called *frame bounds*. If $A = B$, then $\{e_j : j \in J\}$ is called a *tight frame*.

A frame provides a redundant way of representing a signal.

Construction of a Gabor Frame Using ReLUs

We build the window function g using rectified linear units.
Rectified linear unit, or ReLU, is commonly used as activation function of the neuron of deep neural networks. A rectified linear unit is defined as:

$$rect(x) = max\{0, x\}.$$

We define the window function g as a triangular-shaped window function:

$$g(x) = rect(\frac{1}{2}x + 1) - rect(x) + rect(\frac{1}{2}x - 1). \tag{12}$$

We take g as the window function of a Gabor system $G(g, \alpha, \beta)$.

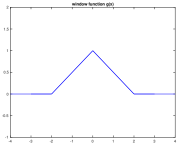


Figure: Window function g definied in (12).

Construction of a Gabor Frame Using ReLUs

It can be shown that $G(g, \alpha, \beta)$ is a Gabor frame with specific choices of α and β.

Lemma

Given window function $g(x) = rect(\frac{1}{2}x + 1) - rect(x) + rect(\frac{1}{2}x - 1)$, the Gabor system $G(g, \alpha, \beta)$ is a Gabor frame for $L^2(\mathbb{R})$ with values of α, β satisfying $\alpha = 1$ and $\beta \leq \frac{1}{6}$.

Correlation Functions and Wiener Space

We introduce the following definitions for the proof.

Definition

Given $g, \gamma \in L^2(\mathbb{R}^d)$ and $\alpha, \beta > 0$, the *correlation functions* of the pair (g, γ) are defined to be

$$G_n(x) = G_n^{(\alpha,\beta)}(x) = \sum_{k \in \mathbb{Z}^d} \bar{g}(x - \frac{n}{\beta} - \alpha k)\gamma(x - \alpha k) \tag{13}$$

for $n \in \mathbb{Z}^d$.

Denote the cube $[0, \alpha]^d$ by Q_α and write $Q = Q_1 = [0, 1]^d$ for the unit cube.

Definition

A function $g \in L^\infty(\mathbb{R}^d)$ belongs to the *Wiener space* $W = W(\mathbb{R}^d)$ if

$$\|g\|_W = \sum_{n \in \mathbb{Z}^d} \operatorname*{ess\,sup}_{x \in Q} |g(x + n)| < \infty. \tag{14}$$

Existence of Gabor Frames

We introduce the following Theorem on conditions of existence of Gabor frame.

Theorem (Walnut, 1992)

Suppose that $g \in W(\mathbb{R}^d)$ *and that* $\alpha > 0$ *is chosen such that for constants* $a, b > 0$

$$a \leq \sum_{k \in \mathbb{Z}^d} |g(x - \alpha k)|^2 \leq b < \infty \quad x - a.e. \tag{15}$$

Then there exists value $\beta_0 = \beta_0(\alpha) > 0$*, such that* $G(g, \alpha, \beta)$ *is a Gabor frame for all* $\beta \leq \beta_0$*. Specifically, if* $\beta_0 > 0$ *is chosen such that*

$$\sum_{n \in \mathbb{Z}^d, n \neq 0} \|G_n^{(\alpha, \beta_0)}\|_\infty < \operatorname*{ess\,inf}_{x \in \mathbb{R}^d} |G_0(x)|, \tag{16}$$

then $G(g, \alpha, \beta)$ *is a frame for all* $\beta \leq \beta_0$*.*

Proof of Lemma

By Theorem (Walnut, 1992), we need to show that $g \in W(\mathbb{R})$ and that g satisfies

$$a \leq \sum_{k \in \mathbb{Z}^d} |g(x - \alpha k)|^2 \leq b < \infty \quad x - a.e. \tag{17}$$

for some $a, b > 0$.
We know that $\text{supp } g = [-2, 2]$, and that $\sup|g| = 1$ by construction of g.
Since $x \in Q_1 = [0, 1]$, and $n \in \mathbb{Z}$, we have

$$\|g\|_W = \sum_{n \in \mathbb{R}} ess \sup_{x \in Q} |g(x + n)| \leq 4 \sup|g| = 4 < \infty. \tag{18}$$

We can choose $\alpha = 1$ so that the infinite sum in (18) has only four non-zero terms for all $x \in \mathbb{R}$. Given any $x \in \mathbb{R}$, we have

$$\sum_{k \in \mathbb{Z}} |g(x - k)|^2 \leq 4 \sup|g|^2 = 4.$$

Thus the upper bound b is $b = 4$.

Norbert Wiener Center
for Harmonic Analysis and Applications

Proof of Lemma

Note that the window function g can be expressed as a piecewise linear function:

$$g(x) = \begin{cases} \frac{1}{2}x + 1, & -2 \leq x \leq 0; \\ -\frac{1}{2} + 1, & 0 < x \leq 2. \end{cases} \tag{19}$$

Hence in order to find the lower bound a, we simplify the sum in (17) for some $x \in [-2, -1]$, and rewrite the equation as

$$\begin{aligned} \sum_{k \in \mathbb{Z}} |g(x-k)|^2 &= |g(x)|^2 + |g(x+1)|^2 + |g(x+2)|^2 + |g(x+3)|^2 \\ &= (x+1)^2 + \frac{5}{2}. \end{aligned} \tag{20}$$

Therefore, the minimum is reached when $x = -1$ and $a = \frac{5}{2}$.
Given $\alpha = 1$, we can choose $\beta \leq \beta_0 = \frac{1}{6}$ so that the the condition for β_0 listed in Theorem (Walnut, 1992) is satisfied.

Norbert Wiener Center
for Harmonic Analysis and Applications

Approximation Property

Now that we have introduced the Gabor frame, we will build a 4-layer neural network that can be used to approximate functions, and we show that

Lemma (Approximation Property, Czaja, Li, 2017)

Let $f \in L^2(\mathbb{R})$ be s times continuously differentiable, and let $\|f^{(s)}\|_1 < \infty$. Then for every $x \in \mathbb{R}$, there exists a construction f_N using Gabor coefficients of modulations up to scale N such that:

$$|f - f_N| = \mathcal{O}\left(\frac{1}{N^{s-1}}\right), \tag{21}$$

where $|\cdot|$ denotes the point-wise absolute value.

Construction of the Neural Network

We construct the neural network with specified number of nodes and layers as the following.

- The input layer: $x \in \mathbb{R}$.
- The first layer: all the shifts $\{x - \alpha k\}$ of x for $k \in [-K, K]$.
- The second layer: shifted x's are activated by modulated ReLUs of three types: $rect(\frac{1}{2}x + 1)$, $-rect(x)$, $rect(\frac{1}{2}x - 1)$, with each of them modulated by $M_{\beta n}$ for $n \in [-N, N]$.
- The third layer: outputs from different ReLUs of the same modulation term are added together to obtain $T_{\alpha k} M_{\beta n} g$ for all $k \in [-K, K]$ and $n \in [-N, N]$.
- The output layer: outputs from the third layer are added to produce the final output function:

$$f_{K,N} = \sum_{|k| \leq K} \sum_{|n| \leq N} w_{k,n} T_{\alpha k} M_{\beta n} g. \tag{22}$$

Construction of the Neural Network

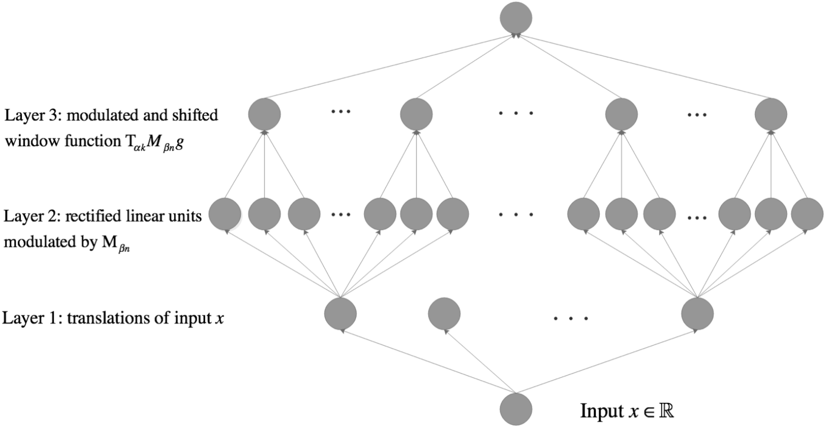


Figure: Illustration of the neural network

Approximation Rate of Neural Networks

Theorem (Czaja, Li, 2017)

Let $f \in L^2(\mathbb{R})$. If f is at s times continuously differentiable for $s \geq 2$, then f can be approximated on the order of $\mathcal{O}(\frac{1}{N^{s-1}})$ using a 4-layer network with $(2K+1)(4(2N+1)+1)$ units. There are $2K+1$ linear units in the first layer; $(2K+1) \times 3 \times (2N+1)$ units in the second layer; $(2K+1)(2N+1)$ linear units in the third layer and a single linear unit in the fourth layer.
Here K is the number of translations and N is the number of modulations in the Gabor system used to construct the neural network.

Norbert Wiener Center for Harmonic Analysis and Applications

Proof of Theorem

The output of the neural network can be written as

$$f_{K,N} = \sum_{|k|\leq K} \sum_{|n|\leq N} w_{k,n} T_{\alpha k} M_{\beta n} g \tag{23}$$

with weights $w_{k,n}$. It remains to prove the Lemma (approximation property). We have shown that $G(g, \alpha, \beta)$ is a Gabor frame for $L^2(\mathbb{R})$ with $\alpha = 1$ and $\beta \leq \frac{1}{6}$.

Proposition (Grochenig)

If $G(g, \alpha, \beta)$ is a frame for $L^2(\mathbb{R}^d)$, then there exists a dual window $\gamma \in L^2(\mathbb{R}^d)$, such that the dual frame of $G(g, \alpha, \beta)$ is $G(\gamma, \alpha, \beta)$. Consequently, every $f \in L^2(\mathbb{R}^d)$ possesses the expansions

$$\begin{aligned} f &= \sum_{k,n\in\mathbb{Z}^d} \sum \langle f, T_{\alpha k} M_{\beta n} g \rangle T_{\alpha k} M_{\beta n} \gamma \\ &= \sum_{k,n\in\mathbb{Z}^d} \sum \langle f, T_{\alpha k} M_{\beta n} \gamma \rangle T_{\alpha k} M_{\beta n} g \end{aligned} \tag{24}$$

with unconditional convergence in $L^2(\mathbb{R}^d)$.

Proof of Theorem

Let $f_{K,N}$ be the approximation obtained by the first $(2K+1)(2N+1)$ terms in the expansion:

$$f_{K,N} = \sum_{|k|\leq K} \sum_{|n|\leq N} \langle f, T_{\alpha k} M_{\beta n} \gamma \rangle T_{\alpha k} M_{\beta n} g. \tag{25}$$

Then for any $x \in \mathbb{R}$,

$$\begin{aligned} |f(x) - f_{K,N}(x)| &= \left| \sum_{|k|>K} \sum_{|n|>N} \langle f, T_{\alpha k} M_{\beta n} \gamma \rangle T_{\alpha k} M_{\beta n} g(x) \right| \\ &\leq \sum_{|k|>K} \sum_{|n|>N} |\langle f, T_{\alpha k} M_{\beta n} \gamma \rangle| \cdot |e^{2\pi i \beta n \cdot (x - \alpha k)}| \cdot |g(x - \alpha k)|. \\ &\leq \sum_{|k|>K} \sum_{|n|>N} |\langle f, T_{\alpha k} M_{\beta n} \gamma \rangle| |g(x - \alpha k)|. \end{aligned} \tag{26}$$

Proof of Theorem

Note that we can consider the Gabor coefficients as

$$\langle f, T_{\alpha k} M_{\beta n} \gamma \rangle = \hat{H}_{\alpha,\beta,k}(n), \quad \textit{where} \quad H_{\alpha,\beta,k}(t_0) = f(\frac{1}{\beta} t_0 + \alpha k)\overline{\gamma(\frac{1}{\beta} t_0)}, \tag{27}$$

for $t_0 \in \mathbb{R}$. We need to discuss the properties of the dual window function γ.

In fact, from the work by Christensen, Kim, Kim on regularity of dual Gabor windows, we obtain the following Lemma:

Lemma (Construction of Smooth Dual Window)

Given window function $g = rect(\frac{1}{2}x + 1) - rect(x) + rect(\frac{1}{2}x - 1)$, there exists dual window function γ such that γ is smooth with finite support.

Proof of Theorem

Since $f \in C^s(\mathbb{R})$, $\gamma \in C^\infty(\mathbb{R})$ and γ has finite support, we have

$$|\hat{H}_{\alpha,\beta,k}(n)| < \frac{C_s}{n^s}, \tag{28}$$

and

$$\sum_{|n|>N} |\hat{H}_{\alpha,\beta,k}(n)| < \sum_{|n|>N} \frac{C_s}{n^s} < s\int_N^\infty \frac{2C_s}{n^s}dn = \frac{2sC_s}{N^{s-1}}, \tag{29}$$

where C_s is a constant proportional to the L^1 norm of the sth derivative of $H_{\alpha,\beta,k}$. we plug in the bound in (29) back into (26) and obtain

$$|f - f_{K,N}| < \sum_{|k|>K} \frac{2sC_s}{N^{s-1}} |g(x-\alpha k)|. \tag{30}$$

Note that g is compactly supported on $[-2,2]$. Then, for any x, there are only finitely many k's ($\lceil \frac{4}{\alpha} \rceil$) with $|k| > K$ such that $g(x-\alpha k) \neq 0$. Therefore,

$$|f - f_{K,N}| < \sum_{|k|>K} \frac{2sC_s}{N^{s-1}} |g(x-\alpha k)| < \frac{\lceil \frac{4}{\alpha} \rceil 2sC_s}{N^{s-1}}. \tag{31}$$

Representation of Network Dictionary

We illustrate with figures the neurons of our neural network in the third layer and the output of our neural network with random weights.

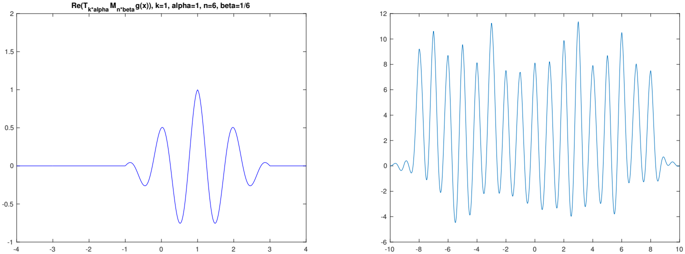


Figure: Translated and modulated ReLUs (left); Output of the neural network with random weights when $K = 8$, and $N = 8$ (right).

Outline

Background on Quantum Energy Regression

- Computation of energy of a single chemical molecule has become an essential topic in computational chemistry.
- A chemical molecule is represented by its state $x = \{r_k, z_k\}_k$, where $r_k \in \mathbb{R}^3$ is the position of the kth nuclei and $z_k > 0$ is the charge of kth nuclei.
- The molecular energy E can be written as a functional of the electron density $\rho(u) \geq 0$ at every position $u \in \mathbb{R}^3$.
- The ground state energy $f(x)$ which is unique for every molecule x, can be obtained by minimizing energy E over a set of electronic densities ρ:

$$f(x) = E(\rho_x) = \inf_{\rho} E(\rho).$$

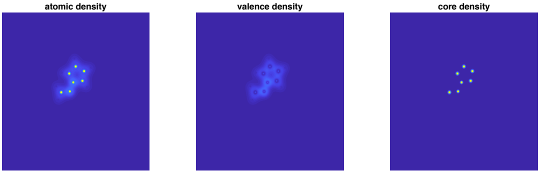


Figure: Approximated Electron Density $\tilde{\rho}_x$

Invariant Properties of Chemical Molecules

The quantum energy $f(x)$ of molecule x must satisfy the following invariant properties:

- **Permutation invariance** The energy functional $f(x)$ is invariant under permutation of indices $\{k = 1, ..., K\}$ in $x = \{r_k, z_k\}_k$.
- **Isometric invariance** The energy functional $f(x)$ is invariant under global translations, rotations, and symmetries of atomic positions r_k.

In machine learning, one way to avoid direct computation of $f(x)$ is to build set of dictionaries of functions $\Phi(x) = \{\phi_i(x)\}_i$ such that the energy $f(x)$ can be approximated by $\tilde{f}(x)$, where

$$\tilde{f}(x) = \langle w, \Phi(x) \rangle = \sum_i w_i \phi_i(x).$$

The weights $\{w_i\}$ are computed such that the error $\sum_{j=1}^n \left|\tilde{f}(x_j) - f(x_j)\right|^2$ on the training data set is minimized.
We intend to build a set of dictionaries from the electronic density of molecules with desired invariant properties.

Gabor Transform on Electronic Density of Molecules

We take the Gabor transform $G_\rho(t, \gamma)$ for the electronic density ρ by window function g:

$$G_\rho(t, \gamma) = \int \rho(x)\overline{g(x-t)}e^{-2\pi i x \gamma} dx,$$

where t is the center location of the window.

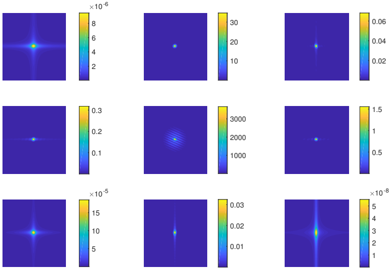


Figure: Gabor transform of electron density at different translation locations

Norbert Wiener Center for Harmonic Analysis and Applications

Translation Invariance

For translation invariance, we take the modulus of $G_\rho(t,\gamma)$ and integrate over all t: $G_\rho(\gamma) = \int_{\mathbb{R}^3} |G_\rho(t,\gamma)|dt$.

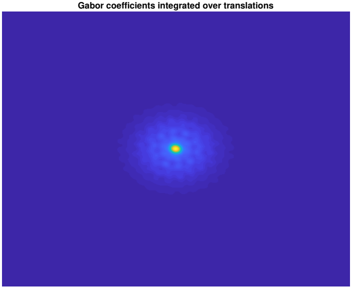


Figure: Gabor coefficients integrated over translations

Rotation Invariance

In order to obtain rotation invariance in the representation, we take average of the coefficients across locations of the same distance to the center.

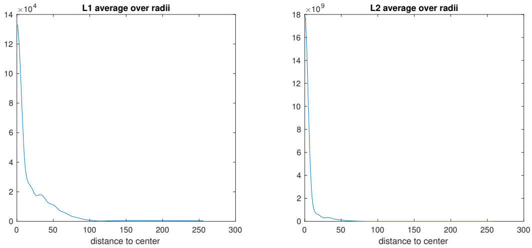


Figure: Rotation and translation invariant representation

To capture information of f of different widths at t, we adopt two different Gaussian functions g_1 and g_2. The Gabor invariant dictionary is defined as:

$$\Phi_\rho = \{\|\rho\|_1, \|G^1_{\rho,k\epsilon}\|_1, \|G^1_{\rho,k\epsilon}\|_2^2, \|G^2_{\rho,k\epsilon}\|_1, \|G^2_{\rho,k\epsilon}\|_2^2\}_{0\le k\le \epsilon^{-2}}. \quad (32)$$

Norbert Wiener Center for Harmonic Analysis and Applications

Quantum Energy Regression

We use sparse orthogonal least squares regression in dictionaries $\Phi(x) = \{\phi_k(x)\}_k$. We compare our results with state-of-the-art methods:

	$\bar{M}$	RMSE	MAE
Coulomb Matrix		6.7 ± 2.8	14.8 ± 12.2
Fourier	73 ± 27	6.7 ± 0.7	8.5 ± 0.9
Wavelet	38 ± 13	6.9 ± 0.6	9.1 ± 0.8
Scattering 16	74	6.9	9.0
Gabor	**71 ± 31**	**5.3 ± 0.3**	**7.0 ± 0.6**
Scattering 17	107 ± 41	3.2 ± 0.1	4.5 ± 0.2

Table: Average Error $\pm$ Standard Deviation over the five folds in kcal/mol

The Gabor invariant representation is promising for its extend-ability to 3D.

Thank You!

References I

Helmut Bölcskei, Philipp Grohs, Gitta Kutyniok, and Philipp Petersen, *Optimal approximation with sparsely connected deep neural networks*, arXiv preprint arXiv:1705.01714 (2017).

Radu Balan, Maneesh Singh, and Dongmian Zou, *Lipschitz properties for deep convolutional networks*, arXiv preprint arXiv:1701.05217 (2017).

Ole Christensen, Hong Oh Kim, and Rae Young Kim, *Regularity of dual gabor windows*, Abstract and Applied Analysis, vol. 2013, Hindawi, 2013.

CK Chui, Xin Li, and HN Mhaskar, *Neural networks for localized approximation*, Mathematics of Computation **63** (1994), no. 208, 607–623.

George Cybenko, *Approximation by superpositions of a sigmoidal function*, Mathematics of control, signals and systems **2** (1989), no. 4, 303–314.

John G Daugman, *Complete discrete 2-d gabor transforms by neural networks for image analysis and compression*, IEEE Transactions on acoustics, speech, and signal processing **36** (1988), no. 7, 1169–1179.

References II

- Karlheinz Gröchenig, *Foundations of time-frequency analysis*, Springer Science & Business Media, 2013.
- Matthew Hirn, Stéphane Mallat, and Nicolas Poilvert, *Wavelet scattering regression of quantum chemical energies*, Multiscale Modeling & Simulation **15** (2017), no. 2, 827–863.
- Matthew Hirn, Nicolas Poilvert, and Stéphane Mallat, *Quantum energy regression using scattering transforms*, arXiv preprint arXiv:1502.02077 (2015).
- Arthur P Lobo and Philipos C Loizou, *Voiced/unvoiced speech discrimination in noise using gabor atomic decomposition*, Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on, vol. 1, IEEE, 2003, pp. I–I.
- Uri Shaham, Alexander Cloninger, and Ronald R Coifman, *Provable approximation properties for deep neural networks*, Applied and Computational Harmonic Analysis (2016).