Machine learning under physical constraints
Wavelet scattering representations

Sixin Zhang
(sixin.zhang@toulouse-inp.fr)

Outline

Wavelet scattering representations

Rotational invariance

Stability properties

Outline

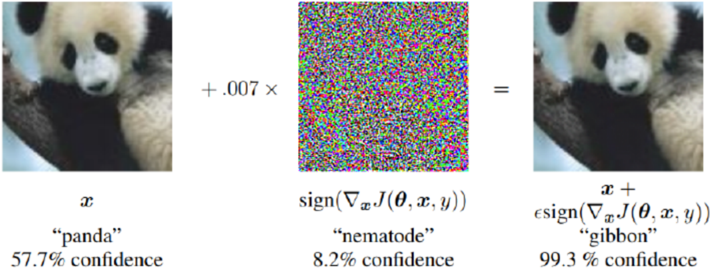
Wavelet scattering representations

Rotational invariance

Stability properties

Adversarial attacks in deep learning

- Small perturbations of input lead to big changes on output.



[Goodfellow et al., 2014]

Instability of Fourier representation

- Fourier representation $\Phi(x) = |\hat{x}|$ is invariant to translations, but unstable to deformations of $x \in L^2(\mathbb{R})$.
- Example: deform a high-frequency signal $x(u) = e^{i\xi u}\theta(u)$.
  - Scale x by deformation: $\tau(u) = su$, $0 < s < 1$
  - $x_\tau(u) = x(u - \tau(u)) = x((1-s)u) = e^{i\xi(1-s)u}\theta((1-s)u)$
  - $\hat{x}_\tau(\omega) = \hat{\theta}(\omega/(1-s))/(1-s)$ has little support overlap with $\hat{x}(\omega)$ if $|\xi|$ is big.
- This lecture: construct stable and informative representations by wavelet scattering transform in 1d,2d,3d.

Wavelet scattering in 1d

- Wavelet transform in 1d: dilate a wavelet ψ with a scale sequence $(2^j)_{j\in\mathbb{Z}}$.
- High-frequency information captured by $\psi_j(u)$ for $j < J$.
- Low-frequency information captured by $\phi_J(u)$.

$$Wx = \{\underbrace{x\star\psi_j}_{W_j x}, \underbrace{x\star\phi_J}_{A_J x}\}_{j<J}$$

- **Zero-th order** scattering invariant: $\int A_J x(u)du$.
- Unfortunaly, $\int W_j x(u)du = 0$, which has no information of x.

First-order and second-order scattering

- Idea: apply a non-linear operator ρ to $W_j x$ to capture information beyond zero-th order. Let

$$U_j x(u) = \rho(W_j x(u)) = \rho(x \star \psi_j(u))$$

- **First-order** scattering invariant:

$$\int U_j x(u) du = \int \rho W_j x(u) du$$

- This captures the average of $U_j x(u)$ (at Fourier frequency 0): more than the zero-th order.
- But it loses high-frequency information in $U_j x$.

First-order and second-order scattering

- Question: How to capture high-frequency information in $U_j x$?
- Compute the wavelet transform of $U_{j_1} x$ at scale j_2,

$$W_{j_2} U_{j_1} x$$

- Second-order scattering transform: apply ρ

$$U_{j_1,j_2} x = \rho W_{j_2} U_{j_1} x = \rho W_{j_2} \rho W_{j_1} x$$

- Second-order scattering invariant:

$$\int U_{j_1,j_2} x(u) du$$

Choice of non-linear operator ρ

- Choose ρ so that $U_j x$ captures informative information.
- Example: Modulus
$$\rho(z) = |z|^p$$
e.g. $p = 1$ and $p = 2$: $\int U_j x(u) du$ captures ℓ_1 and ℓ_2 norms of the wavelet coefficients $W_j x$.
- Example: Generalized rectifier
$$\rho_\alpha(z) = \mathrm{Relu}(\mathrm{Real}(e^{i\alpha} z)), \quad \alpha \in [0, 2\pi]$$
Similar to Relu in neural networks, this captures the phase information in $W_j x$.

m-th order scattering

- In general ,we compute for all $(j_1, \cdots, j_m) \in \{0, 1, \cdots, J-1\}^m$,

$$U_{j_1,j_2,\cdots,j_m}x = \rho W_{j_m} \cdots \rho W_{j_2} \rho W_{j_1} x.$$

- Problem with ρ: If $\rho(z) = |z|^2$, then it is hard to control the stability of $\rho W_{j_m} \cdots \rho W_{j_2} \rho W_{j_1} x$ as its amplitude may grow quickly with m.
- To control the amplitude, we use the modulus non-linearity (or generalized rectifier): $\rho(z) = |z|$.

Invariant scattering coefficients

- Rewrite scattering coefficients using a path variable $p \in \{\emptyset, (j_1), (j_1, j_2), (j_1, j_2, j_3), \cdots\}$.

$$\bar{S}x(p) = \int U_p x(u) du$$

- Order 0: $p = \emptyset$
- Order 1: $p = (j_1)$
- Order 2: $p = (j_1, j_2)$
- Order m: $p = (j_1, j_2, \cdots, j_m)$

Locally invariant scattering coefficients

- To analyze data which are only locally invariant (e.g. MNIST classification), we use the low-pass filter Φ_J to compute

$$S_J x(p, u) = U_p x \star \phi_J(u)$$

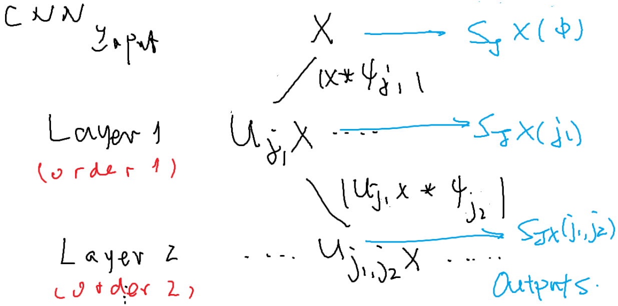
- Since $A_J x = x \star \phi_J$, we write $S_J x(p, u) = A_J U_p x(u)$.
- Relation with invariant $\bar{S}x(p)$ as $J \to \infty$

$$\forall u \in \mathbb{R}, \quad 2^J S_J x(p, u) \to \phi(0) \bar{S} x(p)$$

i.e. local invariance becomes global invariance as J grows.

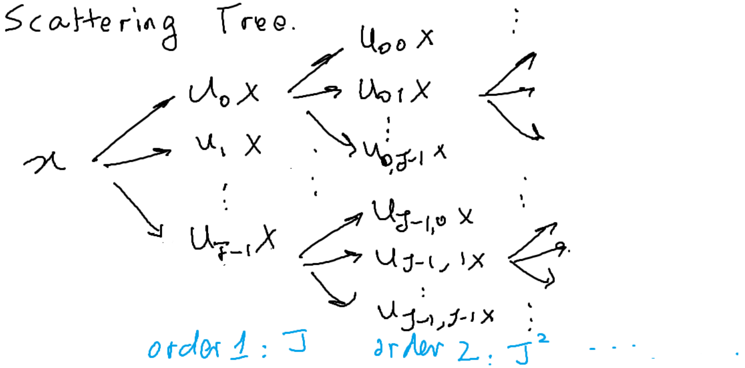
Relation with CNN in deep learning

- Scattering coefficients can be computed using a convolutional neural network (CNN).



- Convolutional kernels: $\{\psi_j\}$.
- Non-linearity: $\rho(z) = |z|$.
- Pooling layer: $S_J x(p) = A_J U_p x$.

Issue of Scattering



- **Issue**: the size of the tree grows in the order J^m as m grows.
- In practice, how to reduce the size?

Scattering in practice: order limitation

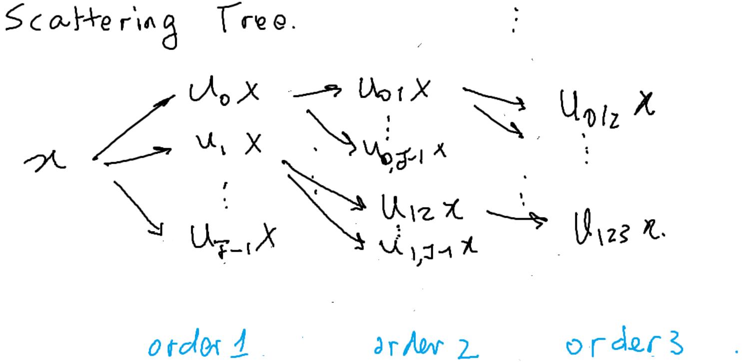
- High-order scattering coefficients tend to be very small, i.e. for $m > 2$,

$$\int |U_p x(u)|^2 du \approx 0, \quad p = (j_1, j_2, j_3, \cdots j_m)$$

Thus only order $m = 1$ and $m = 2$ are used in practice.

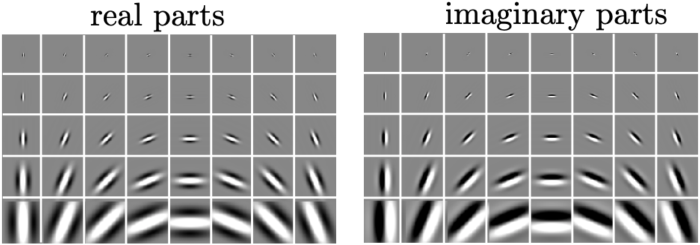
Scattering in practice: scale limitation

- For $m = 2$, one further select the scale $j_2 \geq j_1 + 1$ based on the frequency support overlap of $U_{j_1} x = |x \star \psi_{j_1}|$ and ψ_{j_2}, e.g. x is Dirac.



Scattering in 2d

- Morlet wavelet transform: $Wx = \{x \star \psi_{j,\ell}, x \star \phi_J\}$



  Top to bottom: increasing $j < J$. Left to right: increasing $\ell < L$.

- Define a similar path trajectory:
  $p \in \{\emptyset, (j_1, \ell_1), (j_1, \ell_1, j_2, \ell_2), (j_1, \ell_1, j_2, \ell_2, j_3, \ell_3), \cdots\}$

Scattering in 2d

- Define a similar scattering propagator

$$U_p x(u) = \begin{cases} x(u) & \text{if } p = \emptyset; \\ |x \star \psi_{j_1,\ell_1}(u)| & \text{if } p = (j_1, \ell_1); \\ ||x \star \psi_{j_1,\ell_1}| \star \psi_{j_2,\ell_2}(u)| & \text{if } p = (j_1, \ell_1, j_2, \ell_2); \\ \dots \end{cases}$$

- Invariant scattering coefficients $\bar{S}x(p) = \int U_p x(u) du$.
- Local scattering coefficients $S_J x(p) = A_J U_p x = U_p x \star \phi_J$.

Scattering 2d in practice

- Usually we take $m = 2$, as in scattering 1d.
- Choice of scales j_1, j_2: $j_2 \geq j_1 + 1$.
- Choice of angles ℓ_1, ℓ_2:
  - To compute rotational invariant coefficients, we choose all $0 \leq \ell_1 < 2L$ and $0 \leq \ell_2 < 2L$.
  - We may also consider (ℓ_1, ℓ_2) such that they are nearby angles (recall $\theta_{\ell_1} = \frac{\pi \ell_1}{L}$ and $\theta_{\ell_2} = \frac{\pi \ell_2}{L}$).

Scattering 2d in practice: angle limitations

- There are redundancies in the scattering coefficients using Morlet wavelets, because

$$\psi_{j,\ell+L}(u) = \psi_{j,\ell}(u)^*.$$

  Proof: use $r_{\theta+\pi} u = -r_\theta u, \forall u \in \mathbb{R}^2$.

- Thus we limit the angles to $0 \leq \ell_1 < L$ and $0 \leq \ell_2 < L$,

$$\begin{aligned}
&\big| |x \star \psi_{j_1,\ell_1}| \star \psi_{j_2,\ell_2} \big| \\
&= \big| |x \star \psi_{j_1,\ell_1+L}| \star \psi_{j_2,\ell_2} \big| \\
&= \big| |x \star \psi_{j_1,\ell_1}| \star \psi_{j_2,\ell_2+L} \big| \\
&= \big| |x \star \psi_{j_1,\ell_1+L}| \star \psi_{j_2,\ell_2+L} \big|.
\end{aligned}$$

Scattering 2d in practice: spatial limitations

- In practice, wavelet transform is discretized on a finite grid, i.e. $u \in [0, N-1]^2$.
- The scattering propogator $U_j x$ becomes

$$U_j x(u) = \rho(x \star \psi_j(u)) = \rho\Bigg(\sum_{v \in [0,N-1]^2} x(u-v)\psi_j(v) \Bigg)$$

- To remove spatial redundancies in $U_j x(u)$, one can further sub-sample the “image” $U_j x$ by 2^j, by keeping only $u = 2^j n$ for $n \in [0, N/2^j - 1]^2$.
  Similarly for $S_J x(p, u)$ at $u = 2^J n$ for $n \in [0, N/2^J - 1]^2$.

Outline

Rotational symmetry

- Physical processes which are rotational invariant are called isotropic.
- **Materials science**: In the study of mechanical properties of materials, "isotropic" means having identical values of a property in all directions.

  This sand grain made of volcanic glass is isotropic, and thus, stays extinct when rotated between polarization filters on a petrographic microscope.
- **Fluid dynamics**: Fluid flow is isotropic if there is no directional preference (e.g. in fully developed turbulence). See: `https://en.wikipedia.org/wiki/Isotropy`

Rotational invariant scattering in 2d

- Can we compute scattering coefficients which are invariant to rotations of x in 2d?
- Focus on Morlet wavelets: $u \in \mathbb{R}^2$,

$$\psi_{j,\ell}(u) = 2^{-2j} \psi(2^{-j} r_{\theta_\ell} u)$$

  where $\theta_\ell = \frac{\ell\pi}{L}$ with $0 \leq \ell < 2L$.
- **Question**: Is $\int |x \star \psi_{j,\ell}(u)| du$ rotational invariant?

Rotational invariant scattering in 2d

- Let $\Theta = \{\theta_\ell = \frac{\ell\pi}{L} | 0 \leq \ell < 2L\}$.
- Take $\theta_k \in \Theta$, and apply r_{θ_k} to $x \star \psi_{j,\ell}(u)$. Show

$$(r_{\theta_k} x) \star \psi_{j,\ell}(u) = x \star \psi_{j,\ell-k}(r_{\theta_k} u)$$

  Therefore $\int |x \star \psi_{j,\ell}(u)| du$ is not rotational invariant.
- However, the following first-order scattering coefficients are rotational invariant,

$$\frac{1}{2L} \sum_{\ell=0}^{2L-1} \int |x \star \psi_{j,\ell}(u)| du$$

- Only need to consider $\ell < L$ due to redundancies. The total number of **first-order coefficients** is J.

Rotational invariant scattering in 2d

- Similar to the first-order coefficients, we have

$$|(r_{\theta_k} x) \star \psi_{j_1,\ell_1}| \star \psi_{j_2,\ell_2}(u) = |x \star \psi_{j,\ell_1 - k}| \star \psi_{j_2,\ell_2 - k}(r_{\theta_k} u)$$

- Thus the following second-order scattering coefficients are rotational invariant,

$$\frac{1}{2L} \sum_{k=0}^{2L} \int ||x \star \psi_{j_1,\ell_1 - k}| \star \psi_{j_2,\ell_2 - k}(u)| du$$

- Only need to consider L pairs of (ℓ_1, ℓ_2) due to redundancies. The total number of second-order coefficients is $J(J-1)L/2$.

Outline

Wavelet scattering representations

Rotational invariance

Stability properties

Lipschitz stability

- Consider the robustness property of $\Phi(x)$ to additive perturbations of x to "avoid" adversarial attacks, i.e. we want

$$\text{For small } \epsilon, \quad \Phi(x+\epsilon) \approx \Phi(x).$$

- **Lipschitz stability**: Φ is Lipschitz stable if there is $C > 0$ such that for all $x, x' \in L^2(\mathbb{R}^d)$,

$$\|\Phi(x) - \Phi(x')\| \leq C\|x - x'\|$$

- The modulus non-linearity $\rho(z) = |z|$ is also Lipschitz stable with $C = 1$: for all $z, z' \in \mathbb{C}$,

$$|\rho(z) - \rho(z')| \leq |z - z'|$$

Lipschitz stability of wavelet coefficients

- Focus on 1d case: assume wavelets satisfy the Littlewood-Paley condition with $0 < \epsilon < 1$, i.e. $\forall \omega \in \mathbb{R}$,

$$1 - \epsilon \leq |\hat{\phi}_J(\omega)|^2 + \frac{1}{2} \sum_{j<J} |\hat{\psi}_j(\omega)|^2 + |\hat{\psi}_j(-\omega)|^2 \leq 1$$

- By Plancherel formula, for any $x \in L^2(\mathbb{R}^1)$, the wavelet transform $Wx = \{x \star \phi_J, x \star \psi_j\}_{j<J}$ satisfies

$$(1-\epsilon)\|x\|^2 \leq \|Wx\|^2 \leq \|x\|^2.$$

- As a consequence, the wavelet transform is Lipschitz stable,

$$\|Wx - Wx'\| \leq \|x - x'\|$$

Lipschitz stability of local scattering coefficients

- Is $S_J x = \{S_J x(p)\}_p = \{A_J U_p x\}_p$ Lipschitz stable?
- First-order coefficients (order less than or equal to 1):

$$S_J x = \{A_J x, A_J \rho W_j x\}_{j<J}$$

- Second-order coefficients (order less than or equal to 2):

$$S_J x = \{A_J x, A_J \rho W_{j_1} x, A_J \rho W_{j_2} \rho W_{j_1} x\}_{j_1, j_2 < J}$$

Lipschitz stability of first-order coefficients

- Show $S_J x = \{A_J x, A_J \rho W_j x\}_{j<J}$ is **Lipschitz stable**

$$\|S_J x - S_J x'\|^2 \leq \|x - x'\|^2$$

- By definition,

$$\|S_J x - S_J x'\|^2 = \|A_J x - A_J x'\|^2 + \sum_{j<J} \|A_J \rho W_j x - A_J \rho W_j x'\|^2$$

  - Step 1: As the wavelet transform is Lipschitz stable, so is A_J.
  - Step 2: As ρ is also Lipschitz stable, check that $\|A_J \rho W_j x - A_J \rho W_j x'\|^2 \leq \|W_j x - W_j x'\|^2$.
  - Step 3: Apply the Lipschitz stability of the wavelet transform to conclude.

Lipschitz stability of second-order coefficients

- By the Lipschitz stability of the wavelet transform and ρ,

$$
\begin{aligned}
\|S_J x - S_J x'\|^2 &= \|A_J x - A_J x'\|^2 + \sum_{j_1 < J} \|A_J \rho W_{j_1} x - A_J \rho W_{j_1} x'\|^2 \\
&\quad + \sum_{j_1 < J, j_2 < J} \|A_J \rho W_{j_2} \rho W_{j_1} x - A_J \rho W_{j_2} \rho W_{j_1} x'\|^2 \\
&\leq \|A_J x - A_J x'\|^2 + \sum_{j_1 < J} \|W_{j_1} x - W_{j_1} x'\|^2 \\
&\leq \|x - x'\|^2
\end{aligned}
$$

Deformation stability

- The modulus of Fourier coefficients are not stable to deformations, we study the deformation stability of local scattering coefficients $S_J x$ for $x \in L^2(\mathbb{R})$.
- Let τ be a deformation on $\mathbb{R}$, and $x_\tau(u) = x(u - \tau(u))$.
- **Main idea**: Let $S_J x = A_J U x$, we are going to control the difference between $S_J x_\tau$ and $S_J x$ by the size of τ and Ux.
  - Order 1: $Ux = \{x, \rho W_j x\}_{j<J}$
  - Order 2: $Ux = \{x, \rho W_{j_1} x, \rho W_{j_2} \rho W_{j_1} x\}_{j_1<J, j_2<J}$

Deformation stability of local scattering transform

- Assumption 1: for $\tau \in C^2(\mathbb{R})$ with

$$\|\nabla\tau\|_\infty = \sup_u |\nabla\tau(u)| \leq 1/2$$

- Assumption 2: for $x \in L^2(\mathbb{R})$,

$$\|Ux\|_1 = \sum_p \|U_p x\| < \infty$$

- **Deformation stability**: There exists a constant $C > 0$ such that for x and τ satisfying Assumption 1 and 2,

$$\|S_J x_\tau - S_J x\| \leq C\|Ux\|_1 K(\tau)$$

where $K(\tau)$ is a function depending on J and norms of τ.

Deformation stability: proof sketch

- Let L_τ be the deformation such that $L_\tau x(u) = x_\tau(u)$.
- Step 1: verify

$$\|S_J L_\tau x - S_J x\| \leq \|L_\tau S_J x - S_J x\| + \|L_\tau S_J x - S_J L_\tau x\|$$

- Step 2: verify

$$\|L_\tau S_J x - S_J x\| \leq \|L_\tau A_J - A_J\| \|Ux\|, \quad \|Ux\| \leq \|Ux\|_1$$

- Then use Lemma 1:

$$\|L_\tau A_J - A_J\| \leq C 2^{-J} \|\tau\|_\infty$$

From Lemma 2.12 in Group Invariant Scattering, S. Mallat, 2012

Deformation stability: proof sketch

- Let $U_J x = \{A_J x, \rho W_j x\}_{j<J}$ and $\| \triangle \tau \|_\infty = \sup_{(u,u') \in \mathbb{R}\times\mathbb{R}} |\tau(u) - \tau(u')|$.
- Step 3: verify

$$\|L_\tau S_J x - S_J L_\tau x\| \leq \|Ux\|_1 \|U_J L_\tau - L_\tau U_J\|$$

$$\|U_J L_\tau - L_\tau U_J\| \leq \|W L_\tau - L_\tau W\|$$

- Then use Lemma 2:

$$\|W L_\tau - L_\tau W\| \leq C(\|\nabla \tau\|_\infty (\max(\log \frac{\| \triangle \tau\|_\infty}{\|\nabla \tau\|_\infty}, 1)) + \|\nabla^2 \tau\|_\infty)$$

From Lemma 2.14 in Group Invariant Scattering, S. Mallat, 2012. Therefore

$$K(\tau) = 2^{-J} \|\tau\|_\infty + \|\nabla \tau\|_\infty (\max(\log \frac{\| \triangle \tau\|_\infty}{\|\nabla \tau\|_\infty}, 1)) + \|\nabla^2 z\|_\infty$$