

Note scattering transform et wavelet leaders

Leo Davy

April 2, 2022

1 Scattering transform

Soit G un groupe de rotations r d'angles $\frac{2k\pi}{K}$, on note les ondelettes obtenues à partir d'un filtre passe bande ψ ¹:

$$\psi_\lambda(u) = 2^{-2j}\psi(2^{-j}r^{-1}u), \quad \text{avec } \lambda = 2^{-j}r. \quad (1)$$

Ainsi, si $\hat{\psi}$ est centré à la fréquence η , alors $\hat{\psi}_{2^{-j}r}$ est centré en $2^{-j}r\eta$, avec un support de taille proportionnelle à 2^{-j} (donc en augmentant j on analyse à une fréquence plus basse et de façon plus locale (dans le domaine de Fourier)).

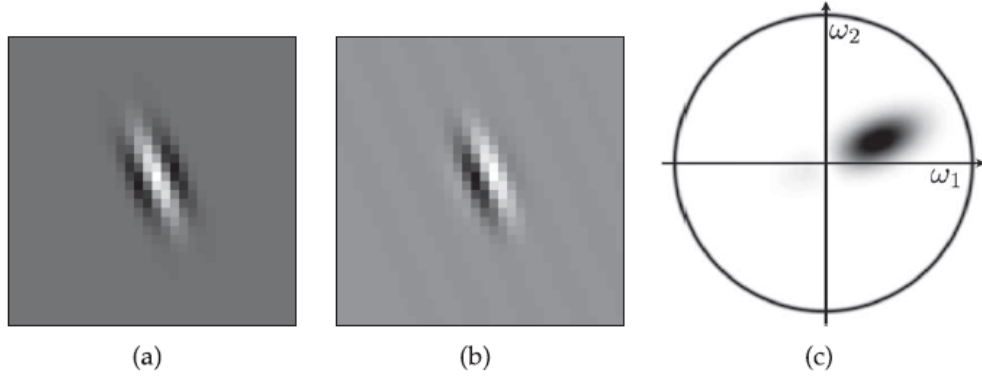


Fig. 1. Complex Morlet wavelet. (a) Real part of $\psi(u)$. (b) Imaginary part of $\psi(u)$. (c) Fourier modulus $|\hat{\psi}(\omega)|$.

A l'ordre 1, les coefficients de scattering d'un signal x sont définis par

$$Sx(u, \lambda, \alpha) = \rho(x \star \psi_{\lambda, \alpha}) \star \phi_J(2^J u) \quad (2)$$

où $\psi_{\lambda, \alpha} = \text{Re}(e^{-i\alpha}\psi_\lambda)$, et ρ est un ReLU, $x \mapsto x_+$. L'action du ReLU permet de récupérer seulement les coefficients d'ondelettes positivement corrélés² avec le signal. La

¹Dans les références, l'ondelette (complexe) de Morlet est toujours utilisée ("afin d'avoir un ensemble parcimonieux de coefficients non négligeables")

²Plus précisément, ayant une partie réelle positive après déphasage d'angle α , mais par la suite on oublie α et on utilise directement le module complexe.

convolution par ϕ_J , une gaussienne permet d'éliminer les variations locales (à une échelle plus fine que 2^J).

L'information perdue en moyennant avec ϕ_J est récupérée par les coefficients d'ordre 2, en appliquant à $\rho(x \star \psi_{\lambda, \alpha})$ une transformée en ondelette, puis on réeffectue un moyennage avec ϕ_J :

$$Sx(u, \lambda_1, \lambda_2, \alpha_1, \alpha_2) = \rho(\rho(x \star \psi_{\lambda_1, \alpha_1}) \star \psi_{\lambda_2, \alpha_2}) \star \phi_J(2^J u). \quad (3)$$

Afin de réduire le nombre de coefficients, le ReLU qui dépend de la phase est remplacé par un module complexe, ce qui donne les coefficients de second ordre:

$$Sx(u, \lambda_1, \lambda_2) = ||x \star \psi_{\lambda_1}| \star \psi_{\lambda_2}| \star \phi_J(2^J u) \quad (4)$$

où $|\lambda_2| < |\lambda_1|$ ($j_2 > j_1$) (on part des plus hautes fréquences et on récupère les moins hautes fréquences au fur et à mesure).

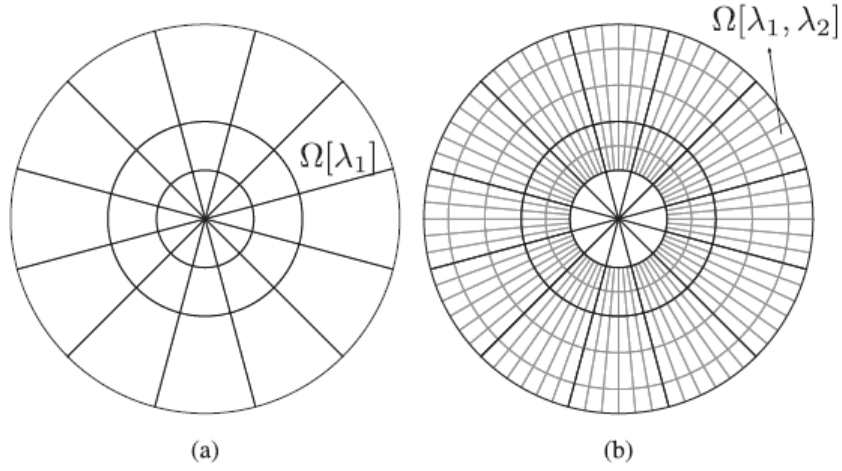


Fig. 3. To display scattering coefficients, the disk covering the image frequency support is partitioned into sectors $\Omega[p]$, which depend upon the path p . (a) For $m = 1$, each $\Omega[\lambda_1]$ is a sector rotated by r_1 that approximates the frequency support of $\hat{\psi}_{\lambda_1}$. (b) For $m = 2$, all $\Omega[\lambda_1, \lambda_2]$ are obtained by subdividing each $\Omega[\lambda_1]$.

Pour généraliser les notations à des ordres plus grands, on peut noter $p = (\lambda_1, \dots, \lambda_m)$ un chemin, et définir

$$U[p](x) = |\dots| |x \star \psi_{\lambda_1}| \star \psi_{\lambda_2} | \dots | \star \psi_{\lambda_m} | \quad (5)$$

et $U[\emptyset]x = x$, et on a les coefficients de scattering en u au bout du chemin p

$$S[p]x(u) = U[p]x \star \phi_{2^J}(u). \quad (6)$$

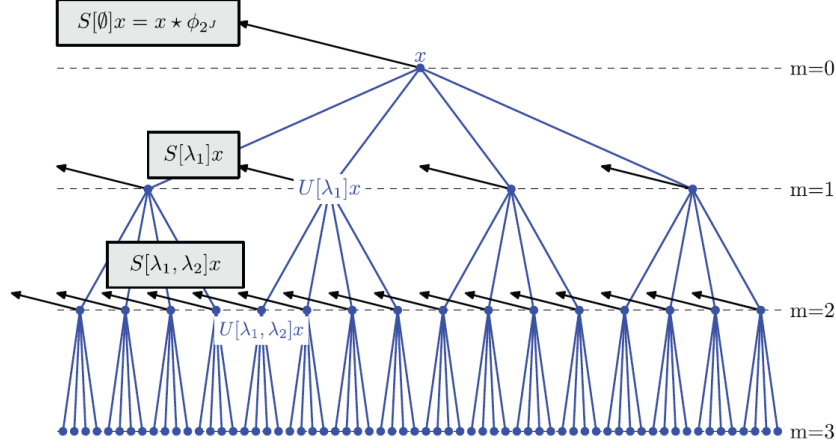


Fig. 2. A scattering propagator \tilde{W} applied to x computes the first layer of wavelet coefficients modulus $U[\lambda_1]x = |x \star \psi_{\lambda_1}|$ and outputs its local average $S[0]x = x \star \phi_{2^J}$ (black arrow). Applying \tilde{W} to the first layer signals $U[\lambda_1]x$ outputs first-order scattering coefficients $S[\lambda_1] = U[\lambda_1] \star \phi_{2^J}$ (black arrows) and computes the propagated signal $U[\lambda_1, \lambda_2]x$ of the second layer. Applying \tilde{W} to each propagated signal $U[p]x$ outputs $S[p]x = U[p]x \star \phi_{2^J}$ (black arrows) and computes the next layer of propagated signals.

Quelques propriétés :

1. Invariance par translation (grâce à $\star \phi_J$ (?))
2. Possible d'adapter à d'autres symétries (rotation, changement d'échelle, ...)
3. En notant $Sx = \{S[p]x\}_{p \in \mathcal{P}_\infty}$:

$$\|Sx\|_2^2 = \sum_{p \in \mathcal{P}_\infty} \|S[p]x\|^2 \quad (7)$$

4. Lipschitz continue par rapport aux déformations, soit $x_\tau(u) = x(u - \tau(u))$, alors

$$\|Sx_\tau - Sx\| \leq C|\bar{m}|||x|| (2^{-J}\|\tau\|_\infty + \|\nabla\tau\|_\infty) \quad (8)$$

5. S est non-expansive (car la transformée en ondelette W et le module sont non-expansifs)
6. Conservation de l'énergie si W est unitaire³

$$\|x\|^2 = \sum_{m=0}^l \sum_{p \in \mathcal{P}^m} \|S[p]x\|^2 + \sum_{p \in \mathcal{P}^{l+1}} \|U[p]x\|^2 \quad (9)$$

7. Réseau peu profond (car le dernier terme de l'équation précédente tend vers 0, donc l'énergie se concentre sur les premières couches)

³ \mathcal{P}^m est l'ensemble des chemins de longueur p .

8. Nombre total de coefficients, avec profondeur \bar{m} , K rotations,

$$P = N 2^{-2J} \sum_{m=0}^{\bar{m}} K^m \binom{J}{m} \quad (10)$$

9. Compléxité de calcul de \bar{m} couches

$$\mathcal{O} \left(\left(\frac{K}{3} \right)^{\bar{m}} N \log N \right) \quad (11)$$

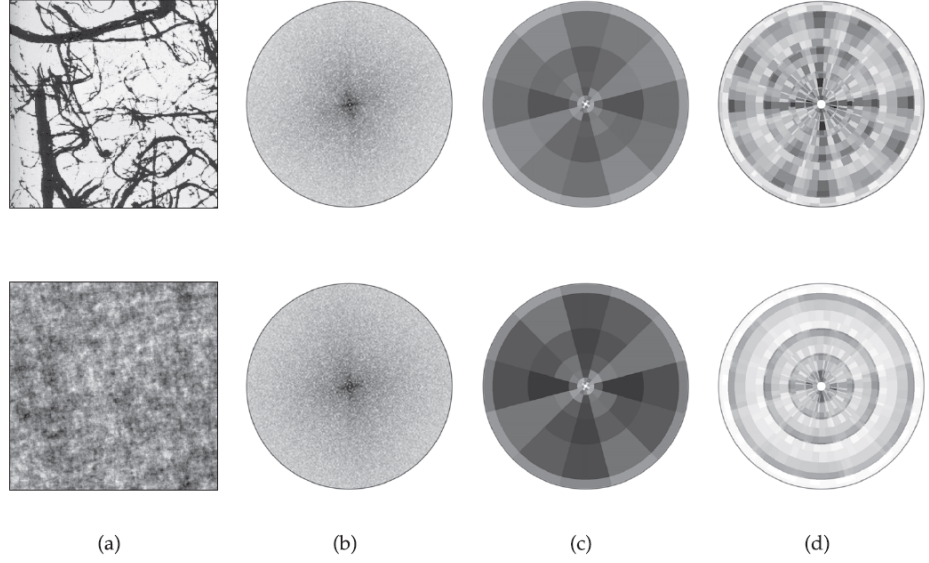


Fig. 5. (a) Realizations of two stationary processes $X(u)$. Top: Brodatz texture. Bottom: Gaussian process. (b) The power spectrum estimated from each realization is nearly the same. (c) First-order scattering coefficients $S[p]X$ are nearly the same for 2^J equal to the image width. (d) Second-order scattering coefficients $S[p]X$ are clearly different.

Quelques applications :

1. Processus stationnaires, les coefficients de scattering sont aussi stationnaires, la variance est conservée

$$\sum_{p \in \mathcal{P}_\infty} \mathbb{E} |S[p]X|^2 = \mathbb{E} |X|^2 \quad (12)$$

et la variance des coefficients d'une ondelette ψ_λ est donnée par

$$\sum_{p \in \mathcal{P}_\infty} \mathbb{E} |S[\lambda + p]X|^2 = \mathbb{E} |X \star \psi_\lambda|^2. \quad (13)$$

Peut distinguer des textures qui ont des seconds moments identiques, mais des moments de plus grand ordre distincts.

2. PCA/classification: Soit k une classe, et X_k des réalisations de la classe k , alors les coefficients moyens sont donnés par $\mathbb{E}SX_k$. La différence $SX_k - \mathbb{E}SX_k$ peut être approximée dans un espace de dimension d ($d \ll P$ où P est le nombre total de coefficients), et la meilleure approximation est donnée par les d principaux vecteurs de la matrice de covariance de SX_k , on note V_k l'espace linéaire engendré par les d vecteurs principaux. Cela nous donne un espace affine pour chaque classe k

$$\mathbb{A}_k = \mathbb{E}SX_k + V_k \quad (14)$$

et on associe à un nouveau signal x sa classe :

$$\hat{k}(x) = \operatorname{argmin}_{k \leq C} \|Sx - P_{\mathbb{A}_k}(Sx)\| \quad (15)$$

3. Classification : On apprend un classifieur (PCA ou SVM) à partir des coefficients de scattering (qui par invariance, ne dépendent pas de la position, rotation, éclairage, ...), état de l'art sur MNIST, USPS et CURET (classification de texture)

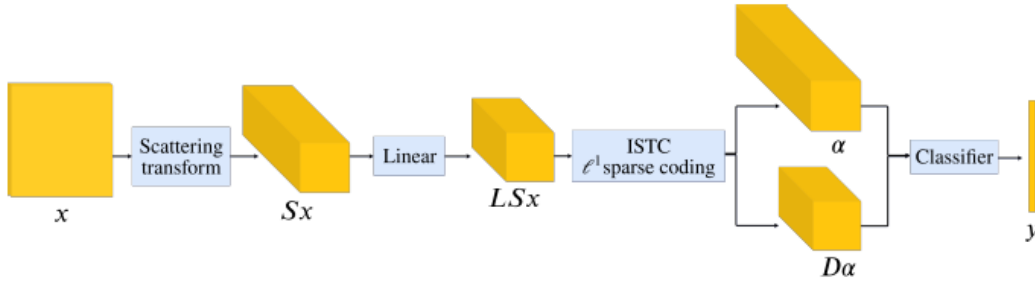


Figure 3: Two variants of the image classification architecture: one where the input for the classifier is the sparse code α , and the other where the reconstruction Da is the input for the classifier.

Quelques références :

1. Référence principale *Invariant Scattering Convolution Networks*, BRUNA, MALLAT http://people.ee.duke.edu/~lcarin/Bruna_Mallat.pdf
2. Détails pour de la classification et liens avec opérateurs proximaux et (A)(L)ISTA *Deep Network classification by scattering and homotopy dictionary learning*, ZARKA, THIRY, ANGLES, MALLAT, <https://arxiv.org/abs/1910.03561>
3. Détails sur comment construire des coefficients de scattering ayant des groupes d'invariance prédéfinis, *Group invariant scattering*, MALLAT, <https://www.di.ens.fr/~mallat/College/CPAM-Mallat-Scat.pdf>

2 Wavelet leaders

Quelques définitions, pour x un signal, a l'échelle, k position et $T_X(a, k)$ une quantité multi-échelle :

1. Fonction de structure :

$$S(a, q) = \frac{1}{n_a} \sum_{k=1}^{n_a} |T_x(a, k)|^q = \mathbb{E}_{k \in \lambda_a} |T_x(a, k)|^q \sim a^{\zeta(q)} \quad (16)$$

correspond au q -ème moment de $T_x(a, \cdot)$ (donc un moment généralisé de x à l'échelle a) en donnant le même poids à tous les n_a coefficients accessibles à l'échelle a (donc invariant par translation).

2. Fonction d'échelle $\zeta(q)$ donne le taux de croissance de $S(a, q)$ (en fonction de a).
3. Régularité en u :

$$h(u) = \liminf_{a \rightarrow 0} \frac{\log(T_x(a, k(u)))}{\log(a)} \quad (17)$$

donne la régularité (minimale) du signal en u aux échelles les plus fines.⁴

Et en adaptant les définitions précédentes aux coefficients d'ondelettes

$$c_{j,k}^{(i)} = \int_{\mathbb{R}^d} X(u) 2^{-dj} \psi^{(i)}(2^{-j}u - k) du \quad (18)$$

$$= \langle X, \tilde{\psi}_{j,k}^{(i)} \rangle = X \star \tilde{\psi}_j^{(i)}(k) \quad (19)$$

et pour alléger les notations, on peut poser

$$c_\lambda^p := \sum_i |c_\lambda^{(i)}|^p, \quad \text{où } \lambda = (j, k), \quad (20)$$

et on a alors:

1. Fonction de structure :

$$S(j, p) = 2^{dj} \sum_k \sum_i |c_{j,k}^{(i)}|^p = 2^{dj} \sum_k c_{j,k}^p \quad (21)$$

2. Fonction d'échelle :

$$\eta(p) = \liminf_{j \rightarrow -\infty} \frac{\log(S(j, p))}{\log(2^j)} \quad (22)$$

3. Leaders :

$$l^p(j, k) = \left(\sum_{j' \leq j, \lambda' \subset 3\lambda} 2^{-d(j-j')} c_{\lambda'}^p \right)^{\frac{1}{p}} \quad (23)$$

4. p -exposant :

$$h_p(u) = \liminf_{j \rightarrow -\infty} \frac{\log(l_{\lambda_j(u)}^{(p)})}{\log(2^j)} \quad (24)$$

⁴Si h ne dépend pas de u , alors $h = \zeta(1)$ (?)

2.1 Leaders et coefficients de scattering

Le cas $p = \infty$ des leaders correspond à la définition classique des wavelets leaders :

$$l_\lambda^{p=\infty} = \sup_{\lambda' \subset 3\lambda} |c_{\lambda'}^{(i)}| = \sup_{\lambda'} |c_{\lambda'}|. \quad (25)$$

Le cas $p = 1$ semble correspondre à la norme 1 des coefficients de scattering à une couche avec des fréquences arbitrairement grandes :

$$l_\lambda^{p=1} := \sum_{j' \leq j, \lambda' \subset 3\lambda} \sum_i |c_{\lambda'}^{(i)}| 2^{-d(j-j')} \quad (26)$$

$$= 2^{-dj} \sum_i \sum_{j' \leq j} \sum_{\lambda_{j'} \subset 3\lambda} \sum_{k \in \lambda_{j'}} |c_{j',k}^{(i)}| 2^{dj'} \quad (27)$$

$$= 2^{-dj} \sum_i \sum_{j' \leq j} \sum_{\lambda_{j'} \subset 3\lambda} \sum_{k \in \lambda_{j'}} |X \star \tilde{\psi}_{j'}^{(i)}|(k) \quad (28)$$

$$= 2^{-dj} \sum_i \sum_{j' \leq j} \sum_{\lambda_{j'} \subset 3\lambda} \sum_{k \in \mathbb{Z}^d} |X \star \tilde{\psi}_{j'}^{(i)}|(k) \mathbb{1}_{3\lambda}(k) \quad (29)$$

$$= 2^{-dj} \sum_i \sum_{j' \leq j} \sum_{k \in \mathbb{Z}^d} |X \star \tilde{\psi}_{j'}^{(i)}|(k) \mathbb{1}_{3\lambda}(k) \quad (30)$$

$$= 2^{-dj} \sum_i \sum_{j' \leq j} \sum_{k \in \mathbb{Z}^d} |X \star \tilde{\psi}_{j'}^{(i)}| \star \mathbb{1}_{[-2^{j'}, 2^{j'}]^d}(2^j k) \quad (31)$$

$$= 2^{-dj} \sum_i \sum_{j' \leq j} \sum_{k \in \mathbb{Z}^d} S^{(i)}[j'] x(k) \quad (32)$$

$$(33)$$

où $S^{(i)}[j']x(k)$ correspondent à des coefficients de scattering pour lesquels on a remplacé la convolution par une gaussienne ϕ_J par la convolution avec une indicatrice sur un cube dyadique.