



Mathematical
Institute

The Scattering Transform: a deterministic DNN

THEORIES OF DEEP LEARNING: C6.5, VIDEO 13

Prof. Jared Tanner

Mathematical Institute

University of Oxford

Oxford
Mathematics

Scattering Transform (Mallat 12')

A deterministic deep nonlinear transform

The Scattering Transform repeatedly applied a deterministic wavelet transform followed by $\sigma(x) = |x|$ as nonlinear activation

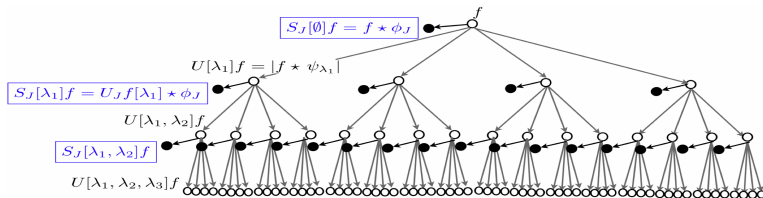


Figure 1: A scattering propagator U_J applied to f computes each $U[\lambda_1]f = |f \star \psi_{\lambda_1}|$ and outputs $S_J[\emptyset]f = f \star \phi_{2^J}$. Applying U_J to each $U[\lambda_1]f$ computes all $U[\lambda_1, \lambda_2]f$ and outputs $S_J[\lambda_1] = U[\lambda_1] \star \phi_{2^J}$. Applying iteratively U_J to each $U[p]f$ outputs $S_J[p]f = U[p]f \star \phi_{2^J}$ and computes the next path layer.

Depth allows the transform to become increasingly invariant to translation and small diffeomorphisms.

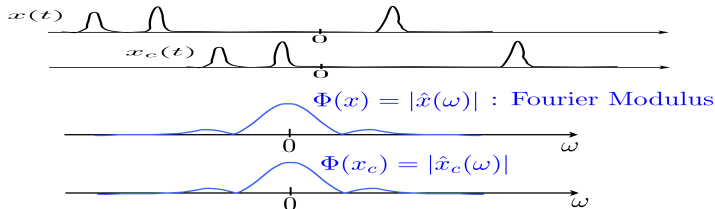
<https://arxiv.org/pdf/1101.2286.pdf>

Classification as learning invariance (Mallat '13)

Learning the nullspace for “nuisance” variables

Invariance to translations $x_c(t) = x(t - c)$

$$\forall c \in \mathbf{R} \quad , \quad \Phi(x_c) = \Phi(x) \quad .$$



Lipschitz stable to deformations $x_\tau(t) = x(t - \tau(t))$

small deformations of $x \implies$ small modifications of $\Phi(x)$

$$\forall \tau \quad , \quad \|\Phi(x_\tau) - \Phi(x)\| \leq C \sup_t |\nabla \tau(t)| \|x\| \quad .$$

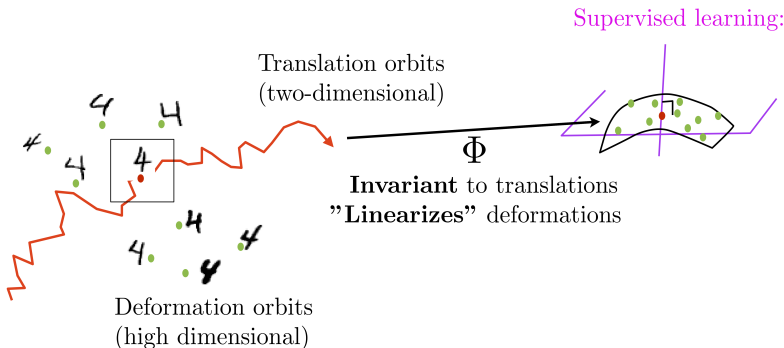
$\downarrow$
deformation size

<http://lcs1.mit.edu/ldr-workshop/Home.html>

Linearising deformations (Mallat '13)

Invariant to translation

- Specific deformation invariance must be learned.

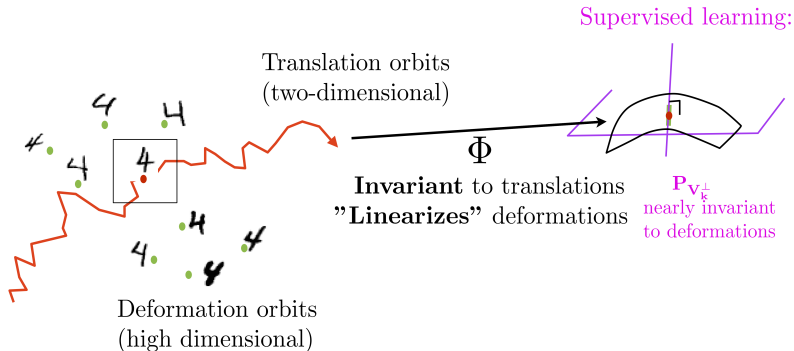


<http://lcs1.mit.edu/ldr-workshop/Home.html>

Linearising deformations (Mallat '13)

Project out the action of translation

- Specific deformation invariance must be learned.

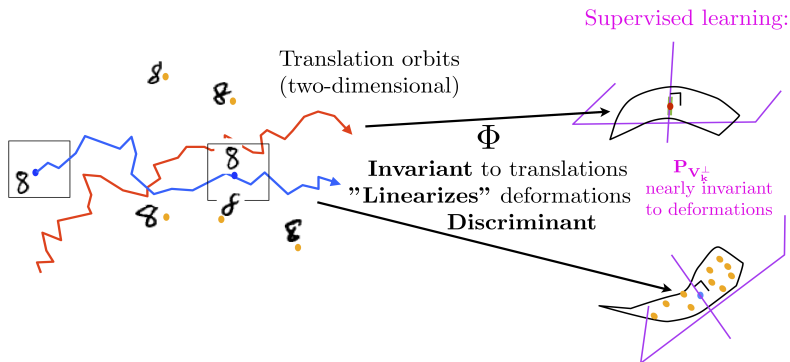


<http://lcs1.mit.edu/ldr-workshop/Home.html>

Linearising deformations (Mallat '13)

Classification after projecting out the invariant

- Specific deformation invariance must be learned.

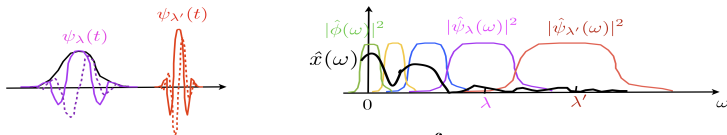


<http://lcs1.mit.edu/ldr-workshop/Home.html>

Wavelet transform as time-frequency tiling (Mallat '13)

Introduction to the wavelet transform

- Complex wavelet: $\psi(t) = \psi^a(t) + i \psi^b(t)$
- Dilated: $\psi_\lambda(t) = 2^{-j} \psi(2^{-j}t)$ with $\lambda = 2^{-j}$.



- Wavelet transform: $x \star \psi_\lambda(t) = \int x(u) \psi_\lambda(t - u) du$

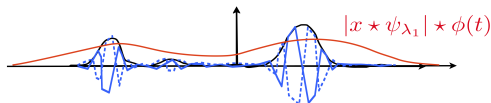
$$Wx = \begin{pmatrix} x \star \phi(t) \\ x \star \psi_\lambda(t) \end{pmatrix}_{t,\lambda}$$

Unitary: $\|Wx\|^2 = \|x\|^2$.

<http://lcs1.mit.edu/ldr-workshop/Home.html>

Modulus and averaging in wavelet domain (Mallat '13)

Depth through repeated application: analogous to DNN



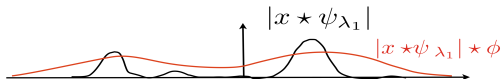
- The modulus $|x \star \psi_{\lambda_1}|$ is a regular envelop
- The average $|x \star \psi_{\lambda_1}| \star \phi(t)$ is invariant to small translations relatively to the support of ϕ .
- Full translation invariance at the limit:

$$\lim_{\phi \rightarrow 1} |x \star \psi_{\lambda_1}| \star \phi(t) = \int |x \star \psi_{\lambda_1}(u)| du = \|x \star \psi_{\lambda_1}\|_1$$

<http://lcs1.mit.edu/ldr-workshop/Home.html>

Second layer of the scattering transform (Mallat '13)

Developing translation invariance through depth



- The high frequencies of $|x \star \psi_{\lambda_1}|$ are in wavelet coefficients:

$$W|x \star \psi_{\lambda_1}| = \left(\begin{array}{c} |x \star \psi_{\lambda_1}| \star \phi(t) \\ |x \star \psi_{\lambda_1}| \star \psi_{\lambda_2}(t) \end{array} \right)_{t, \lambda_2}$$

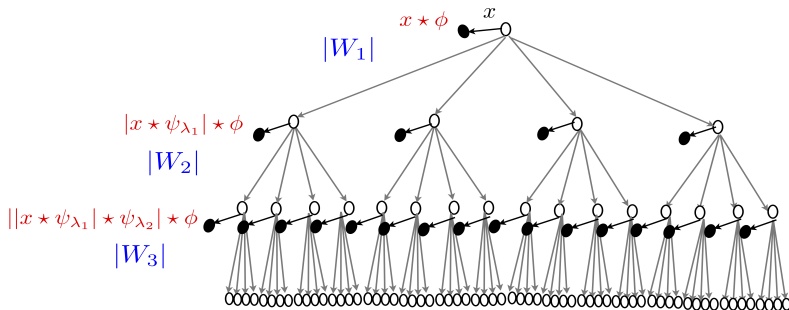
- Translation invariance by time averaging the amplitude:

$$\forall \lambda_1, \lambda_2, \quad ||x \star \psi_{\lambda_1}| \star \psi_{\lambda_2}| \star \phi(t)$$

<http://lcs1.mit.edu/ldr-workshop/Home.html>

Scattering Transform (Mallat '13)

Depth with a deterministic constructive operator



- Cascade of contractive operators

$$|||W_k|x - |W_k|x'| \leq \|x - x'\| \quad \text{with} \quad |||W_k|x| = \|x\| .$$

<http://lcs1.mit.edu/ldr-workshop/Home.html>

Scattering Transform properties (Mallat '13)

Stable to deformations

$$Sx = \begin{pmatrix} x \star \phi(u) \\ |x \star \psi_{\lambda_1}| \star \phi(u) \\ ||x \star \psi_{\lambda_1}| \star \psi_{\lambda_2}| \star \phi(u) \\ |||x \star \psi_{\lambda_2}| \star \psi_{\lambda_2}| \star \psi_{\lambda_3}| \star \phi(u) \\ \dots \end{pmatrix}_{u, \lambda_1, \lambda_2, \lambda_3, \dots}$$

Theorem: *For appropriate wavelets, a scattering is*

contractive $\|Sx - Sy\| \leq \|x - y\|$

preserves norms $\|Sx\| = \|x\|$

stable to deformations $x_\tau(t) = x(t - \tau(t))$

$$\|Sx - Sx_\tau\| \leq C \sup_t |\nabla \tau(t)| \|x\|$$

$\Rightarrow$ linear discriminative classification from $\Phi x = Sx$

<http://lcs1.mit.edu/ldr-workshop/Home.html>

Scattering Transform: energy decay (Mallat 12')

Decay of energy with depth allows truncation

Lemma

For suitably chosen wavelet transforms (see Theorem 2.6 in below article)
then for all $f \in L^2(\mathbb{R}^d)$

$$\lim_{m \rightarrow \infty} \|U[\Lambda_J^m]f\|^2 = \lim_{m \rightarrow \infty} \sum_{n=m}^{\infty} \|S_J[\Lambda_J^n]f\|^2 = 0$$

where $U[\lambda]f = |f \star \psi_\lambda|$ and $S_J[\lambda]f = \phi_j \star U[\lambda]f$ and $\|S_J[P_J]f\| = \|f\|$.
Moreover, for all $c \in \mathbb{R}^d$

$$\lim_{J \rightarrow \infty} \|S_J[P_J]f - S_J[P_J]L_c f\| = 0$$

where $L_c f = f(x - c)$ is the translation operator.

<https://arxiv.org/pdf/1101.2286.pdf>

Scattering Transform: energy decay (Mallat 13')

Example of energy decay

TABLE 1
Percentage of Energy $\sum_{p \in \mathcal{P}^m} \|S[p]x\|^2 / \|x\|^2$ of
Scattering Coefficients on Frequency-Decreasing Paths
of Length m , Depending upon J

J	$m = 0$	$m = 1$	$m = 2$	$m = 3$	$m = 4$	$m \leq 3$
1	95.1	4.86	-	-	-	99.96
2	87.56	11.97	0.35	-	-	99.89
3	76.29	21.92	1.54	0.02	-	99.78
4	61.52	33.87	4.05	0.16	0	99.61
5	44.6	45.26	8.9	0.61	0.01	99.37
6	26.15	57.02	14.4	1.54	0.07	99.1
7	0	73.37	21.98	3.56	0.25	98.91

These average values are computed on the Caltech-101 database, with zero mean and unit variance images.

Depth (m) and wavelet scale (J) fully represent $f \in \mathbb{L}^2(d)$.

<https://www.di.ens.fr/data/publications/papers/pami-final.pdf>

Scattering Transform: MNIST classification (Mallat 13')

Classification error as function of amount of training data

TABLE 4

Percentage of Errors of MNIST Classifiers, Depending on the Training Size

Training size	x		Wind. Four.		Scat. $\bar{m} = 1$		Scat. $\bar{m} = 2$		Conv. Net.
	PCA	SVM	PCA	SVM	PCA	SVM	PCA	SVM	
300	14.5	15.4	7.35	7.4	5.7	8	4.7	5.6	7.18
1000	7.2	8.2	3.74	3.74	2.35	4	2.3	2.6	3.21
2000	5.8	6.5	2.99	2.9	1.7	2.6	1.3	1.8	2.53
5000	4.9	4	2.34	2.2	1.6	1.6	1.03	1.4	1.52
10000	4.55	3.11	2.24	1.65	1.5	1.23	0.88	1	0.85
20000	4.25	2.2	1.92	1.15	1.4	0.96	0.79	0.58	0.76
40000	4.1	1.7	1.85	0.9	1.36	0.75	0.74	0.53	0.65
60000	4.3	1.4	1.80	0.8	1.34	0.62	0.7	0.43	0.53

<https://www.di.ens.fr/data/publications/papers/pami-final.pdf>