Can Signal Classification Speak Mathematics?

Stéphane Mallat,

IHES
Ecole Polytechnique
Deluge of Signals

- Video phones
- Audio
- Seismic data
- Satellite images
- HD Television
- Cameras
- Medical data

- Requires automatic retrieval and analysis.
• Considerable variability in each class.
• A Euclidean norm does not measure signal «similarities». 
A Wild World

Lots of brut force: try everything
Lots of algorithms and little mathematics
Great ideas (neural nets) and know how
Complex and different for different signals.

- Can we now make it simpler and more efficient?

Find simple mathematical concepts explaining the efficiency of algorithms.
Signal Classification Architecture

Signals
dimension \( \sim 10^6 \)

\[ x \quad y \quad ||x - y|| \]

Features
dimension \( \sim 10^5 \)

Audio: MFCC
Images: SIFT

\[ ||\Phi_1(x) - \Phi_1(y)|| \]
similarity measure

- Invariants to physical groups: translations, rotations, scaling...
- Stable to non-rigid deformations
- Preserve information for discriminability
Features

dimension $\sim 10^5$
no low dimens. manifold

Learning Parts: from grouping

dimension $\sim 10^4$

Audio: phonèmes, diphones

Images: ”visual words”

Algorithms: GMM, K-mean clustering
Deep Neural Networks
Dictionary learning
Sparse coding in parts: adaptive variability reduction

dimension $\sim 10^4$

- Audio: Hidden Markov Model
- Images: LLC, Pursuits...
  sparse projection and averaging
Deep Neural Networks

It works! (80% success on CalTech 101)
Signal Classification Architecture

Signals $\Phi_1 \rightarrow$ Features $\Phi_2 \rightarrow$ Code $\rightarrow$ Classes

Sparse coding in parts

dimension $\sim 10^4$

Statistical Classification

Classes

SVM, PCA, RBF

It works! (80% success on CalTech 101)
Representation Code:

progressively more invariant, but stable and discriminative

What are these stable invariants?
Stable Informative Invariants
For Audio and Images
Without Learning

Joakim Anden, Joan Bruna, Laurent Sifre, Irène Waldspurger

• Fourier instabilities ⇒ Mel Frequency Spectrum and Wavelets

• Loss of information ⇒ Deep Convolution Networks.

• Classification
Stable Translation Invariants

- **Invariance** to translations \( x_c(t) = x(t - c) \)

\[
\forall c \in \mathbb{R}, \quad \Phi(x_c) = \Phi(x).
\]

- **Stability** to deformations \( x_{\tau}(t) = x(t - \tau(t)) \)

\[
\begin{align*}
7 & 9 & 6 & 6 & 0 \\
8 & 6 & 3 & 4 & 8
\end{align*}
\]

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

\[
\forall \tau, \quad \| \Phi(x_{\tau}) - \Phi(x) \| \leq C \sup_{t} |\tau'(t)| \| x \|.
\]

deformation size
Fourier Translation Invariance

- Fourier transform \( \hat{x}(\omega) = \int x(t) e^{-i\omega t} dt \) invariance:

  \[
  \text{if } x_c(t) = x(t - c) \text{ then } |\hat{x}_c(\omega)| = |\hat{x}(\omega)|
  \]

- Spectrogram: windowed Fourier transform modulus

  Locally invariant on the window support
High Frequency Instabilites

Instabilites to small deformations $x_{\tau}(t) = x(t - \tau(t))$

$||\hat{x}_\tau(\omega)| - |\hat{x}(\omega)||$ is big at high frequencies

$\tau(t) = \epsilon t$

$\hat{x}(\omega)$ $\hat{x}_\tau(\omega)$
• High frequency instabilities are removed by averaging

with $Q$-constant frequency bandwidth filters $\hat{\psi}_\lambda$
• Spectrogram averaging over Mel frequency bands:

with \( Q \)-constant frequency bandwidth filters \( \hat{\psi}_\lambda \)
Loss of Information

Important loss of information for large invariant windows

$T = 23 \text{ ms}$  $T = 23 \text{ ms}$  $T = 370 \text{ ms}$
• Dilated wavelets: \( \psi_\lambda(t) = 2^{-jQ} \psi(2^{-jQ}t) \) with \( \lambda = 2^{-jQ} \).

• Wavelet transform: \( Wx(t) = \left\{ x \star \phi(t), \ x \star \psi_\lambda(t) \right\}_\lambda \)

• If \( |\phi|^2 + \sum_\lambda |\hat{\psi}_\lambda|^2 = 1 \) then \( W \) is unitary.
For images, dilated and rotated wavelets:

\[ \psi_\lambda(t) = 2^j \psi(2^j r t) \quad \text{with} \quad \lambda = 2^j r \]

- Wavelet transform: \( W x(t) = \{ x \star \phi(t), x \star \psi_\lambda(t) \} \)

- If \( |\phi|^2 + \sum_\lambda |\hat{\psi}_\lambda|^2 = 1 \) then \( W \) is unitary.
Wavelet Mel Spectrum

- Time averaging of wavelet coefficient amplitudes:
  \[ \left\{ |x \ast \psi_\lambda| \ast \phi(t) \right\}_\lambda \]

Wavelet time-frequency

Mel Spectrum time-frequency

Locally translation invariant and stable
Loss of Information by Averaging

\[ |x \ast \psi_\lambda(t)| \quad |f \ast \psi_\lambda| \ast \phi(t) \quad |f \ast \psi_\lambda| \ast \phi(t) \]

\[ T = 23 \text{ ms} \quad T = 23 \text{ ms} \quad T = 370 \text{ ms} \]

Important loss of information for large invariant windows
Recovering Lost Information

- The averaging $|x \ast \psi_{\lambda_1}| \ast \phi$ removes high frequencies:

  ![Diagram showing wavelet transform and Fourier spectrum]

  - Must recover high frequencies: modulation spectrum, SAI
  - Fourier based invariants have instabilities

- Wavelet transform: $|x \ast \psi_{\lambda_1}| \ast \psi_{\lambda_2}(t)$

- Translation invariance by time averaging the amplitude:

  $$\forall \lambda_1, \lambda_2, \quad ||x \ast \psi_{\lambda_1}| \ast \psi_{\lambda_2}| \ast \phi(t) : \text{stable to deformations}$$

- Wavelet scattering along a path $p = (\lambda_1, \lambda_2, \ldots, \lambda_m)$:

  $$S[p]x(t) = | |x \ast \psi_{\lambda_1}| \ast \psi_{\lambda_2} | \ldots | \ast \psi_{\lambda_m} | \ast \phi(t)$$
• Iteration on $Ux = \{x \ast \phi, |x \ast \psi \lambda|\}_\lambda$, contracting.

• Output at all layers: $\{S[p]x\}_{p \in \mathcal{P}}$.

MFSC and SIFT are 1st layer outputs: $S[\lambda_1]x$
• Deep network properties and architecture:

• Fast decay across layers of $\|S[p]x\|$.

• Reduced number of paths with non-negligible output.

• Computational complexity: $O(N \log N)$.
Scattering Stability

- Norm of $Sx = \{S[p]x\}_p : \|Sx\|^2 = \sum_{p \in \mathcal{P}} \|S[p]x\|^2$

**Theorem:** For appropriate wavelets, a scattering is

contracting  \[\|Sx - Sy\| \leq \|x - y\|\]

preserves energy  \[\|Sx\|^2 = \|x\|^2\]

stable to deformations  \[\|Sx - Sx_\tau\| \leq C \sup_t |\tau'(t)| \|x\|\]

invariant to translation when the size of $\phi$ goes to $\infty$. 
Amplitude Modulation

25ms window

\[ S[\lambda_1]x(t) = |x \star \psi_{\lambda_1}| \star \phi(t) \]

512ms window

\[ S[\lambda_1]x(t) = |x \star \psi_{\lambda_1}| \star \phi(t) \]

\[ S[\lambda_1, \lambda_2]x(t) = |x \star \psi_{\lambda_1} \star \psi_{\lambda_2}| \star \phi(t) \text{ for } \lambda_1 = \log(1977) \]

\[ \lambda_1 = \log(\omega_1) \]

\[ \lambda_2 = \log(\omega_2) \]

1977 Hz

18 Hz

Friday, March 30, 2012
Sounds with Same Spectrum

\( X: \) stationary process

\[
S[\lambda_1]X(t) = |X \ast \psi_{\lambda_1}| \ast \phi(t)
\]

\[
S[\lambda_1, \lambda_2]X(t) = ||X \ast \psi_{\lambda_1}| \ast \psi_{\lambda_2}| \ast \phi(t) \text{ for } \lambda_1 = \log(1122)
\]
Image Wavelet Scattering

Images
\[ x \]

\[ \hat{x}(\omega) \]

\[ |x \ast \psi_{\lambda_1} \ast \phi \]

\[ |x \ast \psi_{\lambda_1} \ast \psi_{\lambda_2} \ast \phi \]

window size = image size
Textures with Same Spectrum

$X$: stationary process

<table>
<thead>
<tr>
<th>Textures</th>
<th>Fourier Power Spectrum</th>
<th>Wavelet Scattering</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X$</td>
<td><img src="image1.png" alt="Fourier Spectrum" /></td>
<td>$</td>
</tr>
</tbody>
</table>

$\omega_1$ $\omega_2$

window size = image size

Friday, March 30, 2012
Theorem For appropriate wavelets

\[ Ux = \left\{ x \star \phi, |x \star \psi_\lambda| \right\}_\lambda \]

is invertible and the inverse is continuous.

Inverse scattering: 

Not exactly invertible because the last layer is missing.
Original audio signal $x$

Reconstruction from $Sx$ for a window of 3 s

From order 1 coefficients $S[\lambda_1]x$

From order 2 coefficients $S[\lambda_1, \lambda_2]x$
Digit Classification: MNIST
Digit Classification: MNIST

Wavelet Scattering

$x$

$|x \ast \psi_{\lambda_1} \ast \phi(2^J n)|$

$|x \ast \psi_{\lambda_1} \ast \psi_{\lambda_2} \ast \phi(2^J n)|$

$2^J = 8$: window size

cross-validated
## Digit Classification: MNIST

![MNIST Samples]

### Classification Errors

<table>
<thead>
<tr>
<th>Training size</th>
<th>Conv. Net.</th>
<th>Scattering</th>
</tr>
</thead>
<tbody>
<tr>
<td>300</td>
<td>7.2%</td>
<td>4.4%</td>
</tr>
<tr>
<td>5000</td>
<td>1.5%</td>
<td>1.0%</td>
</tr>
<tr>
<td>20000</td>
<td>0.8%</td>
<td>0.6%</td>
</tr>
<tr>
<td>60000</td>
<td>0.5%</td>
<td>0.4%</td>
</tr>
</tbody>
</table>

LeCun et. al.
Classification of Textures

CUREt database

61 classes

Rotations and illumination variations.
Classification of Textures

$X$

$||X \ast \psi_{\lambda_1}|| \ast \phi$

$||X \ast \psi_{\lambda_1} \ast \psi_{\lambda_2}|| \ast \phi$

window size = image size

cross-validated
Classification of Textures

CUREt database
61 classes

Rotations and illumination variations.

Classification Errors

<table>
<thead>
<tr>
<th>Training per class</th>
<th>Fourier Spectr.</th>
<th>Markov Field</th>
<th>Scattering</th>
</tr>
</thead>
<tbody>
<tr>
<td>46</td>
<td>2.15%</td>
<td>2.46%</td>
<td>0.2%</td>
</tr>
</tbody>
</table>

Varma & Zisserman

Friday, March 30, 2012
Rotation and Affine Invariance

- Scatterings along translation, rotation and affine groups:

\[ x \xrightarrow{\text{Translat. Invar.}} \xrightarrow{\text{Rotation Invar.}} \xrightarrow{\text{Affine Invar.}} Sx \]

UIUC database:

Classification Errors

<table>
<thead>
<tr>
<th>Training</th>
<th>Translation</th>
<th>Transl + Rotation</th>
<th>Affine</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>15 %</td>
<td>3 %</td>
<td>1 %</td>
</tr>
</tbody>
</table>
Conclusion

• *Stable, informative invariants* to think classification.

• Deformation stability need wavelets

• Deep convolution networks provide rich invariants

• Learning data adapted invariants is needed for complex classes: $\Phi_2$

• Understand these invariants: a fascinating problem.

The mathematical tracks still have a long way to go!

Papers/softwares:  [www.cmap.polytechnique.fr/scattering](http://www.cmap.polytechnique.fr/scattering)