Scattering.m Documentation

Release 0.3

Vincent Lostanlen

Dec 16, 2018

Contents

This is the documentation of the `scattering.m` toolbox. The browser version of this document uses MathJax, a JavaScript display engine for math equations. In order to view these equations, you need to enable Java as well as JavaScript in your browser. Visit whatismybrowser.com to check if both are up-to-date and running. If the problem persists, read the PDF version instead.

- **ref** *manual*

Introduction

1.1 Invariance, Stability, Discriminability

Originated in 2010 by Stéphane Mallat as "Recursive Interferometry" [Mal10], the scattering representation aims at providing robust mid-level features for signal-based machine learning.

It relies on the idea that an ideal representation for classification should verify the three following criteria:

1. **Invariance** to translation: if the signal gets globally translated, the representation should not change.

2. **Stability** to small elastic deformations: if the signal get deformed, the representation should be deformed in a proportional way.

3. **Discriminability**: signals in different classes should be represented differently.

Below is an informal review of the capacities of different signal descriptors in terms of these criteria.

	Invariant	Stable	Discriminative
the signal itself			
moving average			
Fourier transform modulus			
wavelet transform modulus			
constant-Q transform, MFCC			
scattering transform			

Invariance and stability can easily be achieved by averaging the signal over time, at the cost of a poor discriminability. Conversely, the modulus of the wavelet transform (called *scalogram*) is stable and discriminative, yet not invariant to translation. The classical constant-Q transform is equivalent to averaging the scalogram over uniform windows : that is a gain in invariance, yet a loss in discriminability. The rationale behing the scattering transform is to compensate for this averaging by recovering the fast variations in the scalogram as well. This is handled by means of another wavelet transform, so as to guarantee a good property of stability.

1.2 Multivariable scattering

The theoretical framework of the scattering transform is not limited to translations of one-dimensional signals. In fact, it can be formulated for any source of variability as long as it follows an algebraic structure of group [Mal12]. Rotation (for images) and frequency transposition (for sounds) are examples of these sources of variability.

1.3 What's in `scattering.m`

The `scattering.m` MATLAB toolbox intends to provide a pipeline for multi-variable scattering that is very generic and customizable, yet remaining as seamless and efficient as possible.

- Morlet wavelets and auditory Gammatone wavelets

- Tunable quality factor, maximum scale, and number of filters per octave

- Pay only for what you use: a fraction of scattering paths can be explicitly spared

- Logarithmic compression of the wavelet scalogram is opt-in

- FFT-based convolutions with subsampling in the Fourier domain

- Efficient products in the Fourier domain with narrowband wavelets

- Multi-variable scattering: all variables share the same framework

- Automatic padding of the log-frequency axis

- Multiple-support filter banks to avoid extraneous padding

- Spiral scattering by reshaping the log-frequency axis on the fly

- Long signals are processed by constant-sized chunks. Unchunking is automatic

- Signal reconstruction from scattering coefficients (even over multiple variables)

- Efficient formatting into feature vectors for classification

Filter bank specifications

2.1 Length of the input signal (compulsory)

Before running `setup`, it is compulsory so fill the `size` field with

```
opts{1}.time.size = length(signal);
```

It must be a power of 2, which leads to optimally fast Fourier Transforms. If you have `K` signals of the same size `N`, consider stacking them into a `KxN` matrix. This wil automatically vectorize the computation and avoid high-level loop overhead.

Under development is a more general architecture that automates padding to the next power of 2, and adapts to all sizes.

2.2 Amount of invariance to translation

The integer `T` is the amount of invariance to translation that you require. It must also be a power of 2.

A typical value for second-order scattering of audio is `T=8192`, that is 370 ms at a sample rate of 22 kHz. A smaller `T` will not integrate full musical notes or full phonemes ; on the contrary, a bigger `T` will blur different notes/phonemes together.

The number of octaves in the filter bank is equal to `J = log2(T)`. By default, `T` is set equal to `size` which means that the corresponding scattering representation `S` will be fully translation-invariant.

2.3 Quality factor

The quality factor `max_Q` of a band-pass filter is defined as the ratio of its center frequency by its bandwidth. Consequently, for a given center frequency, increasing the quality factor will decrease the bandwidth proportionnally, hence yielding a "sharper" band-pass filter in the frequency domain. This increase in frequency sharpness comes at the

cost of increasing the support of the filter in the time domain, which may prevent the representation to distinguish consecutive events.

All the wavelets in a filter bank share the same quality factor: this is why we refer to it as a constant-Q filter bank. Note that this toolbox also allows variable-Q filter banks in order to cope with time support limitations (see section below). This is why the quality factor is `max_Q`.

Typical values for the first order in audio range from 4 to 16. Typical values for the second order along time are 1 or 2. In the context of multivariable scattering, the value 1 is strongly recommended for any derived variable.

A quality factor of 1, corresponding to the so-called 'dyadic' filter bank, is the default.

2.4 Maximum scale

Note that a potential drawback of the constant-Q filterbank is that the time support of the filters is unbounded at the low frequencies. In audio, it is undesirable that acoustic events more than 100 ms apart fall between the same first-order time bin. To address this issue, this toolbox provides a bound `max_scale` that restricts the time support, at the cost of decreasing locally the quality factor.

For instance, for `max_Q = 12` and a sample rate of 22 kHz, setting `max_scale = 2048` (about 93 ms) will provide constant-Q filters for frequencies above Q/max_scale (about 130 Hz) and constant-bandwidth filters below that limit. Setting `max_scale = Inf` will remove the upper bound on the time support and will guarantee that the quality factor is indeed constant throughout the whole frequency range.

By default, `max_scale` is set to `size`, which means that the time support is only limited by the size of the whole signal.

2.5 Number of filters per octave

The integer `nFilters_per_octave` specified the rational quantization of the `gamma` log-scale variable. In order to cover the whole frequency axis, it is compulsory to have

```
nFilters_per_octave > max_Q
```

The number of filters in the filter bank is equal to `nFilters_per_octave * log2(T)`. Henceforth, note that the computational complexity of the computation is linear in the number of filters per octave of each filter bank.

Wavelets

In order to build a wavelet filter bank, one starts from a single bandpass filter $\psi(t)$ named the *mother wavelet*. Then, one has to contract $\psi(\omega)$ in the Fourier domain by scaling factors $2^\gamma \geq 1$ to derive a bank of band pass filters $\widehat{\psi_\gamma}(\omega)$ for different values of γ:

$$\widehat{\psi_\gamma}(\omega) = \hat{\psi}(2^\gamma\omega).$$

In the time domain, the above is equivalent to

$$\psi_\gamma(t) = 2^{-\gamma}\psi(2^{-\gamma}t).$$

In this section, we review three posssible shapes for the mother wavelet ψ:

1. the Morlet wavelet `morlet_1d`,

2. the Gammatone wavelet `gammatone_1d` and

3. the RLC wavelet (causal with exponential decay) `RLC_1d`.

The default specifications are: Gammatone when transforming over time, Morlet when transforming along chromas, RLC when transforming along octaves.

3.1 Morlet wavelet

The ubiquitous Morlet wavelet, also named Gabor wavelet, is proved to have optimal time-frequency Heisenberg uncertainty (see Mallat's wavelet tour [Mal08], Theorem 2.6). It is defined as the product of a Gaussian bell curve of variance σ by a sine wave of frequency ξ.

$$\psi(t) = \exp\left(-\frac{t^2}{2\sigma^2}\right) \times \exp(2\mathrm{i}\pi\xi t) + \varepsilon(t)$$

The function $\varepsilon(t)$ is a low-frequency corrective term to ensure that the wavelet $\psi(t)$ has zero mean. Remarkably, the Morlet wavelet also has a Gaussian profile in the frequency domain.

$$\hat{\psi}(\omega) = \sigma \exp(-\sigma^2\omega^2) + \hat{\varepsilon}(\omega)$$

Since the Gaussian bell curve is symmetric, the Morlet wavelet transform modulus not sensitive to reversal of the t axis. Yet, our perception of time is strongly asymmetric : therefore, for second-order auditory scattering along time, one should prefer the asymmetric Gammatone wavelet (see below) instead of the Morlet wavelet. The Morlet wavelet is well suited to transforms along log-scales γ.

When performing a joint time-frequency transform or spiral transform, the Morlet wavelet handle *morlet_1d* is the default for the transform along log-scales γ. In many cases, it is sensible to use it for transforms along time as well. Aside from the quality factor, it does not have any specific parameter.

3.2 Gammatone wavelet

Because of their temporal asymmetry and near-optimal uncertainty properties, Gammatone filters are widely used in auditory models. They are defined as the product of a monomial of degree N, an exponential decay of attenuation α, and a sine wave of frequency ξ.

We define a Gammatone wavelet by taking the first derivative of the Gammatone function, and replacing the $\sin(2\pi\xi t)$ by $\exp(2i\pi\xi t)$. By doing this, we ensure that the resulting function has zero mean and is analytic (see Venkitaraman et al. [VAS14]). The expression of the Gammatone wavelet in the time domain is:

$$\psi(t) = \left((-\alpha + i\xi)t^{N-1} + (N-1)t^{N-2}\right)\exp(-\alpha t) \times \exp(2i\pi\xi t)$$

In the Fourier domain:

$$\hat{\psi}(\omega) = \frac{i\omega \times (N-1)!}{(\alpha + i(\omega - 2\pi\xi))^N}$$

Observe that, by this definition, the wavelet modulus $|\psi(t)|$ reaches its maximum *after* $t = 0$. In practice, we translate the resulting function in time in order to match the peak at exactly $t = 0$. We also add a phase term such that the real part also reaches its maximum at exactly $t = 0$.

The integer N, called `gammatone_order` in the specifications, is equal to 4 by default. The bigger the N, the more symmetric (hence "Morlet-like") the wavelet will be. The attenuation parameter α is automatically inferred from the required quality factor, through a tedious closed-form equation.

3.3 RLC wavelet

A RLC circuit consists of a resistor R, an inductor L and a capacitor C. In an underdamped regime, the response of this circuit is a sine wave with an exponentially decaying profile. By setting the phase shift φ to zero and taking the analytic part, we derive an analytic "RLC wavelet" of attenuation α and center frequency ξ.

$$\psi(t) = \begin{cases} \exp(-\alpha t) \times \exp(2i\xi t) \text{ if } t \geq 0 \\ 0 \text{ otherwise} \end{cases}$$

This wavelet is rigorously causal (it is zero for $t < 0$) and has a very fast decay in time, at the cost of an imprecise localization in frequency. These properties makes it adapted to wavelet transform across octaves, in the case of spiral scattering.

As much as the Gammatone wavelet is the product of a Gamma probability density function by a sine wave, the RLC wavelet is the product of a Poisson density function by a sine wave. Consequently, the RLC wavelet could alternatively be named "Poisson wavelet". The attenuation parameter α is automatically inferred from the required quality factor, through the simple equation

$$\alpha = \frac{\xi}{2Q}.$$

RLC wavelets are the default when transforming across octaves in a spiral scattering transform. Aside from the quality factor, it does not have any specific parameter.

Bibliography

[Mal10] Mallat, S. Recursive Interferometric Representations. in European Signal Processing Conference 716–720 (2010).

[Mal12] Mallat, S. Group Invariant Scattering. Communications on Pure and Applied Mathematics, vol. 65, issue 10, pages 1331–1398 (2012).

[Mal08] Mallat, S. A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way. 832 (Academic Press, 2008).

[VAS14] Venkitaraman, A., Adiga, A. & Seelamantula, C. S. Auditory-motivated Gammatone wavelet transform. Signal Processing 94, 608–619 (2014).