Wavelets, wavelet networks and the conformal group

R. Vilela Mendes
CMAF, University of Lisbon
http://label2.ist.utl.pt/vilela/

April 2016

Contents

- Wavelets: Continuous and discrete
- Wavelet networks. Training and initialization
- Signal transforms: A unified framework
- Wavelets and the conformal group
- Conformal wavelets and higher dimensions
- Appendix: Wavelets on graphs

Fourier and short-time Fourier transforms

- Fourier transform. Let

$$\|f\|^2 = \int_{-\infty}^{\infty} |f(t)|^2 \, dt < \infty$$

Then, the Fourier transform and its inverse are

$$\widehat{f}(\omega) = \int_{-\infty}^{\infty} e^{-i2\pi\omega t} f(t) \, dt$$

$$f(t) = \int_{-\infty}^{\infty} e^{i2\pi\omega t} \widehat{f}(\omega) d\omega$$

- Not appropriate for signals with time-changing spectrum
- Short-time Fourier transform

$$\widetilde{f}(\omega, t) = \int_{-\infty}^{\infty} e^{-i2\pi\omega s} V^*(s-t) f(s) \, ds$$

$$f(t) = \frac{1}{\|W\|^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i2\pi\omega s} V(s-t) \widetilde{f}(\omega, t) \, ds d\omega$$

- Fixed size window. Not satisfactory if the signal modulation is not uniform

Wavelets

- The wavelet transform decomposes the signal $f(t)$ into a set of basis functions

$$W_f(s,\tau) = \int f(t)\,\psi^*_{s,\tau}(t)\,dt$$

the wavelets $\psi_{s,\tau}(t)$ being kernel functions generated from a basic (mother) wavelet $\psi(\tau)$ by means of a translation and a rescaling ($-\infty < \tau < \infty$, $s > 0$):

$$\psi_{s,\tau}(t) = \frac{1}{\sqrt{s}}\,\psi\left(\frac{t-\tau}{s}\right).$$

The parameter τ follows the evolution in time of the signal and the (scaling) parameter s explores different local structures

- For normalized $\psi(t)$ the wavelets $\psi_{s,\tau}(t)$ satisfy the normalization condition

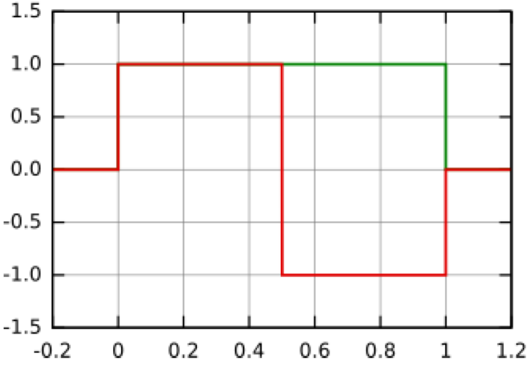
$$\int |\psi_{s,\tau}(t)|^2\,dt = 1$$

Wavelets

The mother wavelet may have different forms, for example:

- Haar wavelet

$$\psi(t) = \begin{cases} 1 & 0 \leq t < 1/2 \\ -1 & 1/2 \leq t < 1 \\ 0 & \text{otherwise} \end{cases}$$



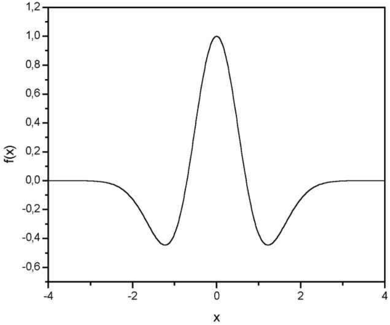
Wavelets

- The Gaussian Fourier-window

$$\psi(t) = \frac{1}{\sqrt{\pi}}\, e^{i\omega_0 t}\, e^{-t^2/2},$$

- The Mexican hat wavelet

$$\psi(t) = (1 - t^2)\, e^{-t^2/2}$$



Wavelets

- Inverse wavelet transform

$$f(t) = N_h^{-1} \int W_f(s,\tau)\, \frac{1}{\sqrt{s}}\, \psi\left(\frac{t-\tau}{s}\right) \frac{d\tau\, ds}{s^2}\,,$$

with

$$N_h = \int \frac{|H(\omega)|^2}{|\omega|}\, d\omega, \qquad H(\omega) = \int \psi(t)\, e^{-i\omega t}\, dt.$$

Therefore, existence of the inverse requires $N_h < \infty$.

- One has the property

$$\int |W_f(s,\tau)|^2\, \frac{d\tau\, ds}{s^2} = N_h \int |f(t)|^2\, dt.$$

Discrete wavelet transform (DWT)

- So far we have dealt with what is called "the continuous wavelet transform" (CWT). Requires continuous parameters s and τ. If, instead one wants to have only a discrete (in τ) set of wavelets to project the signal, further analysis is required.
- Instead of one basic function one requires two, the *scaling function* $\phi(t)$ and the *mother wavelet* $\psi(t)$

  $$\int \phi(t)\, dt = 1; \qquad \int \psi(t)\, dt = 0$$

  DWT is most appropriate in the case of discrete signals

  $$f(n) = \{f(0), f(1), \cdots, f(n-1)\}$$

- From the scaling function and the mother wavelet, rescaled and shifted sequences are constructed

  $$\begin{aligned} \phi_{j,k}(t) &= 2^{j/2}\phi\left(2^j t - k\right) \\ \psi_{j,k}(t) &= 2^{j/2}\psi\left(2^j t - k\right) \end{aligned}$$

Discrete wavelet transform (DWT)

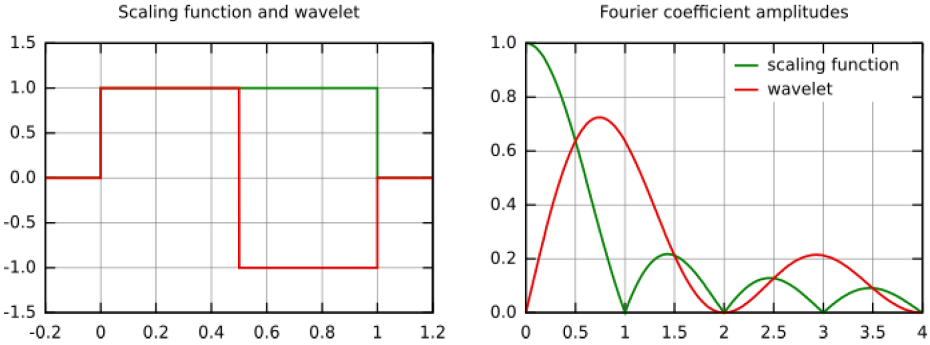
- Requirements:
  (i) The mother wavelet must be orthogonal to its dyadic dilations and integer translations.
  (ii) The scaling function must be orthogonal to its integer translations

$$
\begin{aligned}
V_j &= \text{Span}\left\{\phi_{j,k}(t)\right\} \\
W_j &= \text{Span}\left\{\psi_{j,k}(t)\right\}
\end{aligned}
$$

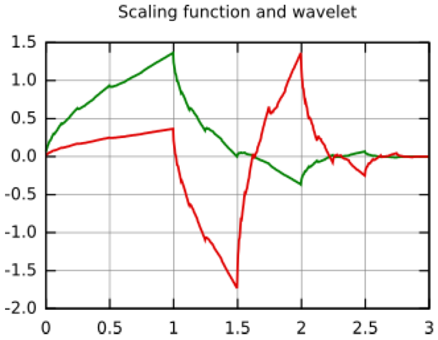
  (iii) $V_j \subset V_{j+1}$
  (iv) $\cap V_j = \{0\}$
  (v) $L^2 = V_0 \oplus W_0 \oplus W_1 \oplus W_2 \oplus \cdots$

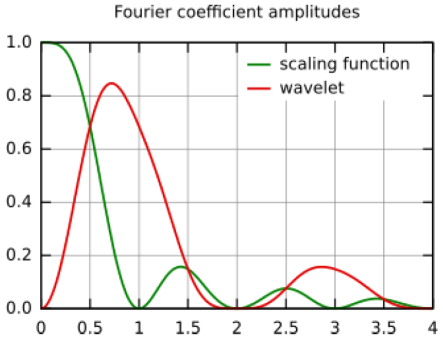
Discrete wavelet transform: Examples

The Haar wavelet



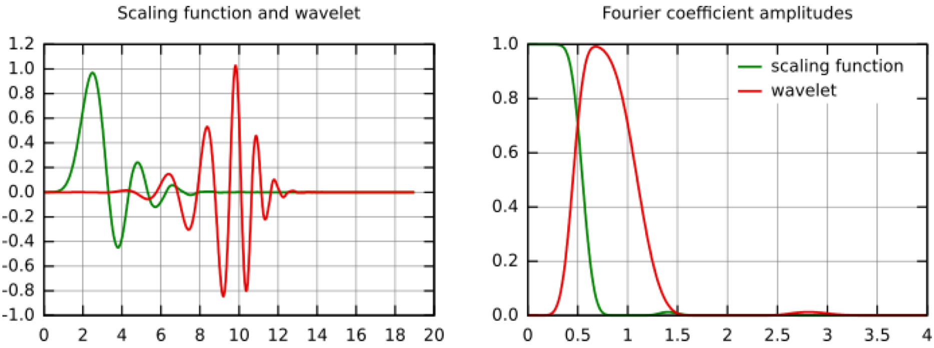
Daubechies 4





Discrete wavelet transform (DWT)

Daubechies 20



Discrete wavelet transform (DWT)

Then

$$f(n) = \frac{1}{\sqrt{N}} \sum_k W_\varphi(j_0, k)\, \varphi_{j_0,k}(n) + \frac{1}{\sqrt{N}} \sum_j \sum_k W_\psi(j, k)\, \psi_{j,k}(n)$$

with

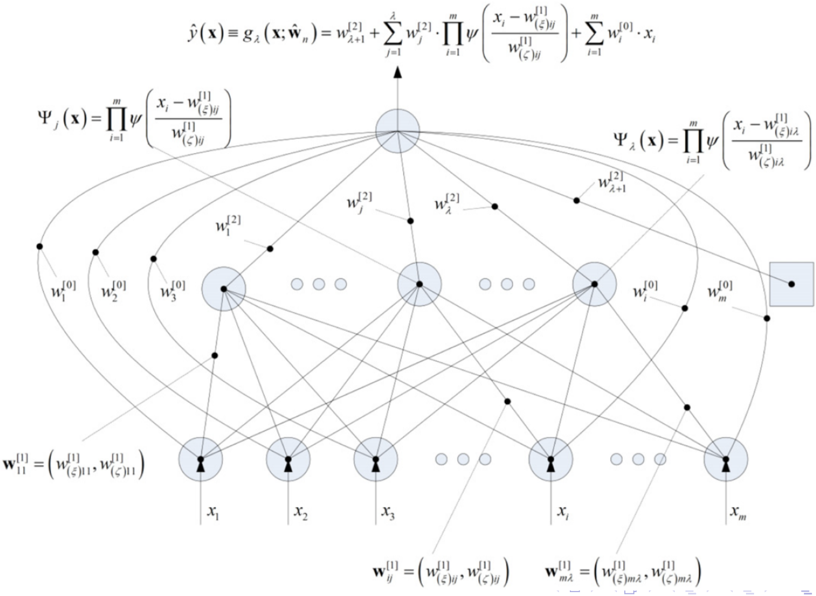
$$W_\varphi(j_0, k) = \frac{1}{\sqrt{N}} \sum_n f(n)\, \varphi_{j_0,k}(n)$$

$$W_\psi(j, k) = \frac{1}{\sqrt{N}} \sum_n f(n)\, \psi_{j,k}(n) \qquad j \geq j_0$$

The coefficients $W_\varphi(j_0, k)$ and $W_\psi(j, k)$ are called *detailed* and *wavelet* coefficients. For practical implementation, the coefficients are obtained by a sequence of filters (FIR's).

Wavelet Networks

- Wavelet analysis (WA) is a valuable tool for signal processing, having been extensively used in time-series analysis, image processing, signal denoising, time-scale decomposition, etc. Its main drawback is the small input dimension, because the construction of the wavelet basis for higher dimensions is computationally very expensive.
- On the other hand, neural networks (NN) handle well enough high dimensional input signals but suffer from the limitation that the objective functions are not necessarily well represented by sigmoid functions.
- Wavelet networks (WN) are a class of networks that combine NN and WA. Are a generalization of radial basis function networks (RBFN), being hidden layer networks that use wavelets for activation instead of the classic sigmoid functions. Adjustable parameters, during the learning stage, are both the connection strenghts and the wavelet parameters (position, scale, 2D-orientation).
- Typically, they are one hidden layer networks. Example: (Neural Networks, 42, 2013, 1-27)


$\hat{y}(\mathbf{x}) \equiv g_\lambda(\mathbf{x}; \hat{\mathbf{w}}_n) = w_{\lambda+1}^{[2]} + \sum_{j=1}^{\lambda} w_j^{[2]} \cdot \prod_{i=1}^{m} \psi\left(\frac{x_i - w_{(\xi)ij}^{[1]}}{w_{(\zeta)ij}^{[1]}}\right) + \sum_{i=1}^{m} w_i^{[0]} \cdot x_i$
$\Psi_j(\mathbf{x}) = \prod_{i=1}^{m} \psi\left(\frac{x_i - w_{(\xi)ij}^{[1]}}{w_{(\zeta)ij}^{[1]}}\right)$
$\Psi_\lambda(\mathbf{x}) = \prod_{i=1}^{m} \psi\left(\frac{x_i - w_{(\xi)i\lambda}^{[1]}}{w_{(\zeta)i\lambda}^{[1]}}\right)$
$w_{\lambda+1}^{[2]}$
$w_j^{[2]}$
$w_\lambda^{[2]}$
$w_1^{[2]}$
$w_1^{[0]}$
$w_2^{[0]}$
$w_3^{[0]}$
$w_i^{[0]}$
$w_m^{[0]}$
$\mathbf{w}_{11}^{[1]} = \left(w_{(\xi)11}^{[1]}, w_{(\zeta)11}^{[1]}\right)$
x_1
x_2
x_3
x_i
x_m
$\mathbf{w}_{ij}^{[1]} = \left(w_{(\xi)ij}^{[1]}, w_{(\zeta)ij}^{[1]}\right)$
$\mathbf{w}_{m\lambda}^{[1]} = \left(w_{(\xi)m\lambda}^{[1]}, w_{(\zeta)m\lambda}^{[1]}\right)$

Wavelet Networks

- Main issues: Training and Initialization
- Network output

$$y(\mathbf{x}) = w_{\lambda+1}^{(2)} + \sum_{j=1}^{\lambda} w_j^{(2)} \bullet \psi_j(\mathbf{x}) + \sum_{i=1}^{m} w_i^{(0)} \bullet x_i$$

Each wavelet is the product of m one-dimensional wavelets

$$\psi_j(\mathbf{x}) = \prod_{i=1}^{m} \psi(z_{ij}); \qquad \psi(z_{ij}) = \left(1 - z_{ij}^2\right) e^{-\frac{1}{2} z_{ij}^2}$$

$$z_{ij} = \frac{x_i - w_{(\tau)ij}^{(1)}}{w_{(s)ij}^{(1)}}$$

The whole set of network parameters is

$$w = \left(w_i^{(0)}, w_{(\tau)ij}^{(1)}, w_{(s)ij}^{(1)}, w_j^{(2)}, w_{\lambda+1}^{(2)}\right)$$

Wavelet Networks

- Training: The weights are trained to minimize the mean quadratic cost function
$$L = \frac{1}{2n} \sum_{p=1}^{n} \left(f_p - y_p\right)^2$$
f_p being the objective function. For each parameter
$$\frac{\partial L}{\partial w} = -\frac{1}{n} \sum_{p=1}^{n} \left(f_p - y_p\right) \frac{\partial y_p}{\partial w}$$
and if the delta rule is used for updating
$$w_{t+1} = w_t - \eta \frac{\partial L}{\partial w_i} + k\left(w_t - w_{t-1}\right)$$
- Initialization: is a critical issue. Several methods, for example:
  - Analyse by WA the input (training) data set.
  - Construct a library of wavelets
  - Remove those that are not in the training data and use the most dominant wavelet parameters as initial values.

Signal transforms: A unified framework

- Signals $f(t)$ considered as vectors $\mid f\rangle$ belonging to a subspace $\mathcal{N}$ of a Hilbert space $\mathcal{H}$ with dual space $\mathcal{N}^*$ (and the canonical identification $\mathcal{N} \subset \mathcal{N}^*$).
- $\{U(\alpha) : \alpha \in I\}$ a family of operators defined on $\mathcal{N}^*$(and a fortriori on $\mathcal{N}$). In many cases, the family of operators $U(\alpha)$ generates a unitary group. However, this is not a necessary condition for the consistency of the formalism, provided completeness conditions are satisfied.
- Three types of transforms are defined. Consider a reference vector $h \in \mathcal{N}^*$ chosen in such a way that the linear span of $\{U(\alpha)h \in \mathcal{N}^* : \alpha \in I\}$ is dense in $\mathcal{N}^*$. Means that, out of the set $\{U(\alpha)h\}$, a complete set of vectors can be chosen to serve as basis. Two of the transforms are:

$$W_f^{(h)}(\alpha) = \langle U(\alpha)\, h \mid f\rangle,$$

$$Q_f(\alpha) = \langle U(\alpha)\, f \mid f\rangle$$

Signal transforms: A unified framework

- If $U(\alpha)$ is a unitary operator generated by

$$B\left(\overrightarrow{\alpha}\right) = \alpha_1 t + i\alpha_2 \frac{d}{dt}$$

  and h is a (generalized) eigenvector of the time-translation operator, $W_f^{(h)}(\alpha)$ is the Fourier transform. With the same $B\left(\overrightarrow{\alpha}\right)$ plus the parity operator, $Q_f(\alpha)$ is the Wigner–Ville transform.

- Similarly, for

$$B\left(\overrightarrow{\alpha}\right) = \alpha_1 D + i\alpha_2 \frac{d}{dt}$$

  where D is the dilation operator $D = -\frac{1}{2}\left(it\frac{d}{dt} + i\frac{d}{dt}t\right)$, $W_f^{(h)}(\alpha)$ is a wavelet transform and $Q_f(\alpha)$ the Bertrand transform.

- Fourier or wavelet are transforms of the $W_f^{(h)}$-type.
  Those of the Q_f-type are *quasidistribution* transforms.

Signal transforms: A unified framework

In general, if $U(\alpha)$ are unitary operators, by Stone's theorem, there are self-adjoint operators $B(\alpha)$ such that

$$W_f^{(h)}(\alpha) = \langle h \mid e^{iB(\alpha)} \mid f \rangle,$$

$$Q_f^{(B)}(\alpha) = \langle f \mid e^{iB(\alpha)} \mid f \rangle.$$

In this case, because $B(\alpha)$ has a real valued spectrum, another transform may be defined, called a *tomogram*. Let, in the unitary operator $U(\alpha) = e^{iB(\alpha)}$, $B(\alpha)$ have the spectral decomposition $B(\alpha) = \int XP(X)\, dX$, where $P(X)$ denotes the projector on the (generalized) eigenvector $\langle X| \in \mathcal{N}^*$ of $B(\alpha)$. The tomogram is

$$M_f^{(B)}(X) = \langle f| P(X) |f\rangle = |\langle X \mid f \rangle|^2 .$$

The tomogram $M_f^{(B)}(X)$ is the squared amplitude of the projection of the signal $|f\rangle \in \mathcal{N}$ on the eigenvector $\langle X| \in \mathcal{N}^*$ of the operator $B(\alpha)$. Therefore it is a **positive** bilinear transform.

Signal transforms: A unified framework

In this general setting a group theoretical interpretation is obtained for linear and bilinear transforms

The wavelets are obtained from the mother wavelet by the application of a affine group operation

$$\begin{aligned} \psi_{s,a}(t) &= U(s,\tau)\psi(t) = e^{\tau\frac{d}{dt}} e^{-\log s\left(t\frac{d}{dt}+\frac{1}{2}\right)} \psi_0(t) \\ &= \frac{1}{\sqrt{s}} \psi\left(\frac{t+\tau}{s}\right). \end{aligned}$$

The operators $\omega = -i\frac{d}{dt}$ and $D = -i\left(t\frac{d}{dt}+\frac{1}{2}\right)$ generate the algebra of the affine group, with commutation relations

$$\begin{aligned} [\omega,\omega] &= [D,D] = 0 \\ [\omega,D] &= -i\omega \end{aligned}$$

The (affine) wavelet transform is the nondiagonal matrix element of a representation of the affine group.

However there is a larger group with first order differential generators, the (one dimensional) **conformal group**.

One dimensional conformal group

- The generators of the algebra of the one-dimensional conformal group are,

$$\omega = -i\tfrac{d}{dt}$$
$$D = -i\left(t\tfrac{d}{dt} + \tfrac{1}{2}\right)$$
$$K = i\left(t^2\tfrac{d}{dt} + t\right)$$

- With commutation relations

$$\begin{aligned}
[\hat{\omega}, \hat{\omega}] &= [D, D] = [K, K] = 0 \\
[\hat{\omega}, D] &= -i\hat{\omega} \\
[\hat{\omega}, K] &= i2D \\
[D, K] &= -iK
\end{aligned}$$

Transforms and the conformal group

The eigenvectors $\zeta^{(F)}(X,t)$, $\zeta^{(F)}(X,t)$ and $\zeta^{(F)}(X,t)$ of the operators of the conformal group may be used to construct signal transforms

$$\begin{aligned}
\omega\zeta^{(F)}(X,t) &= \omega\left(e^{-iXt}\right) = X\left(e^{-iXt}\right) \\
D\zeta^{(D)}(X,t) &= D\left(\frac{1}{\sqrt{|t|}}e^{-iX\log|t|}\right) = X\left(\frac{1}{\sqrt{|t|}}e^{-iX\log|t|}\right) \\
K\zeta^{(C)}(X,t) &= K\left(\frac{1}{t}e^{i\frac{X}{t}}\right) = X\left(\frac{1}{t}e^{i\frac{X}{t}}\right)
\end{aligned}$$

Transforms

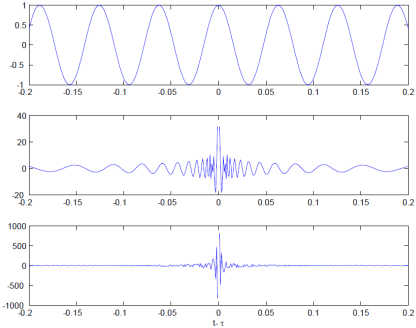
$$\begin{aligned}
F^{(F)}(X) &= \frac{1}{\sqrt{2\pi}}\int dt e^{iXt} f(t) \qquad \text{Fourier} \\
F^{(D)}(X) &= \frac{1}{\sqrt{2\pi}}\int dt \frac{1}{\sqrt{|t|}} e^{iX\log|t|} f(t) \\
F^{(C)}(X) &= \frac{1}{\sqrt{2\pi}}\int dt \frac{1}{t} e^{-i\frac{X}{t}} f(t)
\end{aligned}$$

Transforms and the conformal group

All transforms are invertible because

$$\int dX \zeta^* \left(X, t'\right) \zeta\left(X, t\right) = 2\pi\delta\left(t - t'\right)$$

The analysing power of the $\zeta\left(X, t - \tau\right)$'s



Conformal wavelets

- The (affine) wavelets are linear superpositions of eigenvectors of D translated by $e^{-\tau \frac{d}{dt}}$.
- Likewise, let $\psi_0^{(C)}(t)$ be a superposition of eigenvectors of K

$$\psi_0^{(C)}(t) = \frac{1}{\sqrt{2\pi}} \int dX \frac{1}{t} e^{-i\frac{X}{t}} \widehat{\psi}_0(X)$$

one obtains a set of *conformal wavelets*

$$\psi_{\tau,s}^{(C)}(t) = e^{-\tau \frac{d}{dt}} e^{-s\left(t^2 \frac{d}{dt} + t\right)} \psi_0^{(C)}(t)$$

- and a conformal wavelet transform

$$W_f^{(c)}(s,\tau) = \int f(t)\, \psi_{s,\tau}^{(c)*}(t)\, dt$$

Higher power operators and higher dimensions

- Higher power operators might also be used to construct other (analysing) transforms

$$e^{-\tau \frac{d}{dt}} e^{-s\left(t^n \frac{d}{dt} + \frac{n}{2} t^{n-1}\right)} \psi_0\left(t\right)$$

- For higher dimensions, the conformal group generators in $d \geq 2$, are

$$\begin{aligned}
\omega_k &= -i \frac{\partial}{\partial t_k} \\
D &= -i\left(t \bullet \nabla + \frac{d}{2}\right) \\
R_{j,k} &= i\left(t_j \frac{\partial}{\partial t_k} - t_k \frac{\partial}{\partial t_j}\right) \\
K_j &= i\left(t_j^2 \frac{\partial}{\partial t_j} + t_j\right)
\end{aligned}$$

References

- Q. Zhang and A. Benveniste; *Wavelet Networks*, IEEE Trans. on Neural Networks 3 (1992) 889-898.
- A. K. Alexandridis and A. D. Zapranis; *Wavelet neural networks: A practical guide*, Neural Networks 42 (2013) 1-27.
- G. Kaiser; *A Friendly Guide to Wavelets*, Springer 2011.
- M. A. Man'ko, V. I. Man'ko and R. Vilela Mendes; *Tomograms and other transforms: a unified view*, J. Phys. A: Math. Gen. 34 (2001) 8321-8332.
- R. Vilela Mendes, H. C. Mendes and T. Araújo; *Signal processing on graphs: Transforms and tomograms*, Physica A 450 (2016) 1-17.

Research problems

- Develop the theory of *conformal wavelets* and find applications
- Generalize for graphs and higher dimensions

Wavelet transforms on graphs

Affine wavelets for time series: The wavelet $\psi_{s,a}(t)$ is obtained from the mother wavelet $\psi_0(t)$ by

$$\begin{aligned}
\psi_{s,a}(t) &= U(s,a)\,\psi_0(t) = e^{\log s\left(t\frac{d}{dt}+\frac{1}{2}\right)} e^{a\frac{d}{dt}} \psi_0(t) \\
&= \sqrt{s}\psi_0(st+a).
\end{aligned}$$

For graphs translation is easy to generalize but it is not obvious how to generalize scale transformations. This becomes clearer if we rewrite the wavelet transform in frequency space,

$$\begin{aligned}
W(\tau,s) &= \int dt \psi^*_{s,\tau}(t)\, f(t) \\
&= \int dt \left(e^{\log s\left(t\frac{d}{dt}+\frac{1}{2}\right)} e^{\tau\frac{d}{dt}} \psi^*_0(t) \right) f(t) \\
&= \int d\omega \frac{e^{-i\frac{\omega}{s}a}}{\sqrt{s}} \widehat{\psi_0}^{\,*}\left(\frac{\omega}{s}\right) \widehat{f}(\omega)
\end{aligned}$$

$\widehat{\psi_0}$ and $\widehat{f}$ denoting the Fourier transforms of the mother wavelet and of the signal.

Wavelet transforms on graphs

One sees that the wavelet transform is represented as a sum over the Fourier spectrum Ω with the (frequency) argument of the mother wavelet shifted from ω to $\frac{\omega}{s}$. The mapping $\omega \in \Omega \rightarrow \frac{\omega}{s} \in \Omega$ is a one-to-one onto mapping of the Fourier spectrum Ω into itself. Therefore the natural generalization of the wavelet transform for graphs may be defined as a similar sum, with the spectrum label shift being one of the possible one-to-one onto mappings of the spectrum of the adjacency matrix (or of the Laplacian matrix).

Writing the Fourier-like transform on graphs and its inverse as

$$\begin{aligned} \widehat{f}(\eta) &= \sum_i \chi_\eta(i) f(i) \\ f(i) &= \sum_\eta \widehat{f}(\eta) \chi_\eta(i) \end{aligned} \tag{1}$$

where $\chi_\eta(i)$ is an eigenvector of $\mathbf{A}$ or $\mathbf{L}$ (or a generalized eigenvector or an eigenvector of $\mathbf{A}^T\mathbf{A}$ or $\mathbf{L}^T\mathbf{L}$) and η denotes the spectral label in the spectrum Ω of the matrices.

Wavelet transforms on graphs

With a localized "mother wavelet"

$$\psi^{(k)}(i) = \delta_{k,i}$$

the wavelet-like transform on graphs would be

$$f(a, \widetilde{s}) = \sum_{\eta} \chi_{\widetilde{s}(\eta)}(k+a)\,\widehat{f}(\eta).$$

The mapping $\widetilde{s}(\eta)$ is not $\eta \rightarrow \frac{\eta}{s}$ because in general $\frac{\eta}{s}$ is not in Ω. $\widetilde{s}$ is a mapping in the set $\mathcal{S}$ of the possible one-to-one onto mappings of Ω, $\widetilde{s} \in \mathcal{S}$.
The inverse wavelet transform is

$$\widehat{f}(\eta) = \frac{1}{\#\mathcal{S}} \sum_{a,\widetilde{s}} \chi_{\widetilde{s}(\eta)}(a)\, f\left(a, \tilde{s}\right).$$

$\#\mathcal{S}$ denoting the cardinality of $\mathcal{S}$.

Wavelet transforms on graphs

Hammond, Vandergheynst and Gribonval have also attempted to generalize the notion of wavelet transform to graph signals. However, instead of the sum with the shifted arguments in the spectrum, their construction corresponds to the introduction of a $\eta-$dependent weight on the sum of the 2nd equation in (1), with both the signal component $\widehat{f}(\eta)$ and the eigenvector χ_η associated to the same spectral value η. Therefore their construction is more in the spirit of a Fourier deformation of the signal rather than of a wavelet transform.