Wavelet Kernels and RKHS

Alain Rakotomamonjy, Xavier Mary and Stéphane Canu
alain.rakotomamonjy@insa-rouen.fr

asi.insa-rouen.fr/˜arakotom

INSA Rouen -Département ASI

Laboratoire PSI

Motivations

- Justify wavelet networks (Zhang, 1992) as a particular case of Regularization Networks (Girosi, 1995)
- Enlarge choice of hypothesis space where one looks for the solution of a learning problem by including wavelet span
- Develop algorithms that adapt the regularization to the scale of data.

Roadmap

1. The context of Learning from examples
2. Building RKHS from Hilbert space
3. Building Wavelet Kernels
4. Examples and Applications
5. Conclusions and Perspectives

The learning problem setting

- Learn the dependency between two sets $\mathcal{X}$ and $\mathcal{Y}$ from examples :

  $$(x_1, y_1) \cdots (x_\ell, y_\ell) \in \mathcal{X} \times \mathcal{Y}$$

  drawn *i.i.d* from $P(x, y)$

- Define a cost function $C(\cdot, \cdot)$ and look for the function f^* that minimizes the risk :

  $$R[f] = \int C(y, f(x)) dP(x, y)$$

- Regularized empirical risk induction principle :

  $$R_{emp}[f] = \frac{1}{\ell} \sum_i C(y_i, f(x_i)) + \Omega(f)$$

  $\Omega(f)$ being some smoothness measure of f.

The learning problem hypothesis

- Minimize R_{emp} over f with $f \in \mathcal{H}$

$\mathcal{H}$ is a functional vector space

- $f(x_i)$ exists and is defined for any $x_i \in \mathcal{X}$

$\mathcal{H}$ is a pointwise defined function set included in $\mathbb{R}^{\mathcal{X}}$

- if two functions of $\mathcal{H}$ are "similar", we wish that their pointwise value for any x are not so much different.

$$\forall x \in \mathcal{X}, \exists M_x \in \mathbb{R}\, s.t\, \forall f, g \in \mathcal{H},\ |f(x) - g(x)| \leq M_x \|f - g\|_{\mathcal{H}}$$

The learning hypothesis in Hilbert Space

- $\mathcal{H}$ is a Hilbert space of function with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$
- The evaluation functional δ_x on $\mathcal{H}$

$$\delta_x : \begin{array}{ccc} \mathcal{H} & \longrightarrow & \mathbb{R} \\ f & \longrightarrow & \delta_x[f] = f(x) \end{array}$$

  exists and is defined for all $x \in \mathcal{X}$
- The evalution functional δ_x is continuous. Hence, there exists $K(x, \cdot) \in \mathcal{H}$ so that :

$$\delta_x[f] = f(x) = \langle K(x, \cdot), f(\cdot) \rangle_{\mathcal{H}}$$

$\mathcal{H}$ is a Reproducing Kernel Hilbert Space

How to build a RKHS from a Hilbert space

- Use a linear application to map a function Hilbert space $\mathcal{B}$ to $\mathbb{R}^{\mathcal{X}}$
- Define a set of function $\Gamma_x(\cdot) \in \mathcal{B}$ with $x \in \mathcal{X}$
- Define the so-called Carleman operator :

$$T : \begin{array}{ccl} \mathcal{B} & \longrightarrow & \mathbb{R}^{\mathcal{X}} \\ f & \longrightarrow & g(\cdot) \text{ so that } g(x) = Tf(x) \triangleq \langle \Gamma_x(\cdot), f(\cdot) \rangle_{\mathcal{B}} \end{array}$$

- Decompose $\mathcal{B} = Ker(T) \oplus \mathcal{M}$ and call S the bijective restriction of T so that :

$$S : \begin{array}{ccl} \mathcal{M} & \longrightarrow & \mathcal{H} = Im(T) \\ f & \longrightarrow & g(\cdot) = Sf = Tf \end{array}$$

How to build a RKHS from a Hilbert space (Ctd)

- if $\mathcal{H}$ is endowed with inner product :

$$\forall g_1, g_2 \in \mathcal{H}, \qquad \langle g_1, g_2 \rangle_{\mathcal{H}} \triangleq \langle S^{-1} g_1, S^{-1} g_2 \rangle_{\mathcal{B}} = \langle f_1, f_2 \rangle_{\mathcal{B}}$$

  then $\mathcal{H}$ is a RKHS

- Idea of proof : Check the continuity of the evaluation functional

$$\begin{aligned} |g(x)| &= \langle \Gamma_x(\cdot), f(\cdot) \rangle_{\mathcal{B}} \\ &\leq \|\Gamma_x\| \|f\|_{\mathcal{B}} \\ &\leq M_x \|g\|_{\mathcal{H}} \end{aligned}$$

- Reproducing Kernel in $\mathcal{H}$:

$$K(x, y) = \langle \Gamma_x(\cdot), \Gamma_y(\cdot) \rangle_{\mathcal{B}}$$

Example of Γ_x and reproducing kernel

- $\mathcal{X} = \mathbb{R}^d$, $\mathcal{B} \subset L_2$ of dimension d with basis elements $\{e_i\}_{i=1 \cdots d}$

- Create $\Gamma_x(\cdot) \in \mathcal{B}$ so that $\forall x \in \mathcal{X}$

$$\Gamma_x(\cdot) = \sum_{i=1}^{d} x_i e_i(\cdot)$$

- Reproducing kernel in $\mathcal{H}$:

$$K(x, y) = \langle \Gamma_x(\cdot), \Gamma_y(\cdot) \rangle_{\mathcal{B}} = \sum_{i=1}^{d} x_i y_i$$

Differences with other ways for building admissible kernels and RKHS

- Constructive algorithm for building RKHS and positive definite kernel from any set of measures function $\Gamma_x(\cdot)$
- $\mathcal{X}$ is not necessarily a compact set (like in Mercer's theorem)
- The RKHS is built without the knowledge of the pd kernel

Mapping $L_2(\mathcal{X})$ to Wavelet RKHS

- Take $\mathcal{B} = L_2(\mathcal{X})$ with $\{\phi_i\}$ be a wavelet basis of L_2 and choose $\Gamma_x(\cdot)$ so that $\mathcal{H} \subset L_2(\mathcal{X})$

- As $\Gamma_x(\cdot)$ is in $L_2(\mathcal{X})$, we have :

$$\Gamma_x(\cdot) = \sum_i \alpha_i(x)\phi_i(\cdot) \qquad \text{with } \forall x \in \mathcal{X}, \sum_i \alpha_i^2(x) < \infty$$

  with $\{\alpha_i(\cdot)\}$ being a set of coefficient depending on the evaluation point x.

- As $\forall i,\ [S\phi_i](x) = \langle \Gamma_x(\cdot), \phi_i(\cdot)\rangle_{L_2} = \alpha_i(x)$, we have $\alpha_i(\cdot) \in \mathcal{H} \subset L_2(\mathcal{X})$. Thus :

$$\alpha_i(x) = \sum_j \alpha_{i,j}\phi_j(x)$$

Mapping $L_2(\mathcal{X})$ to Wavelet RKHS

- $\mathcal{H}$ endowed with $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ is a RKHS.

- Reproducing Kernel in $\mathcal{H}$:

$$\begin{aligned} K(x,y) &= \langle \Gamma_x(\cdot), \Gamma_y(\cdot) \rangle_{L_2} \\ &= \sum_{i,j,n} \alpha_{i,j} \alpha_{i,n} \phi_j(x) \phi_n(y) \end{aligned}$$

- Particular case $\alpha_{i,j} = \delta_{i,j}$:

$$\begin{aligned} K(x,y) &= \sum_{i} \alpha_i^2 \phi_i(x) \phi_i(y) \\ &= \sum_{j,k} \alpha_{j,k}^2 \psi_{j,k}(x) \psi_{j,k}(y) \end{aligned}$$

Practical implementation of wavelet kernel

- Number of wavelet basis in $L_2(\mathbb{R}^d)$ is exponential with respect to the input dimension d.

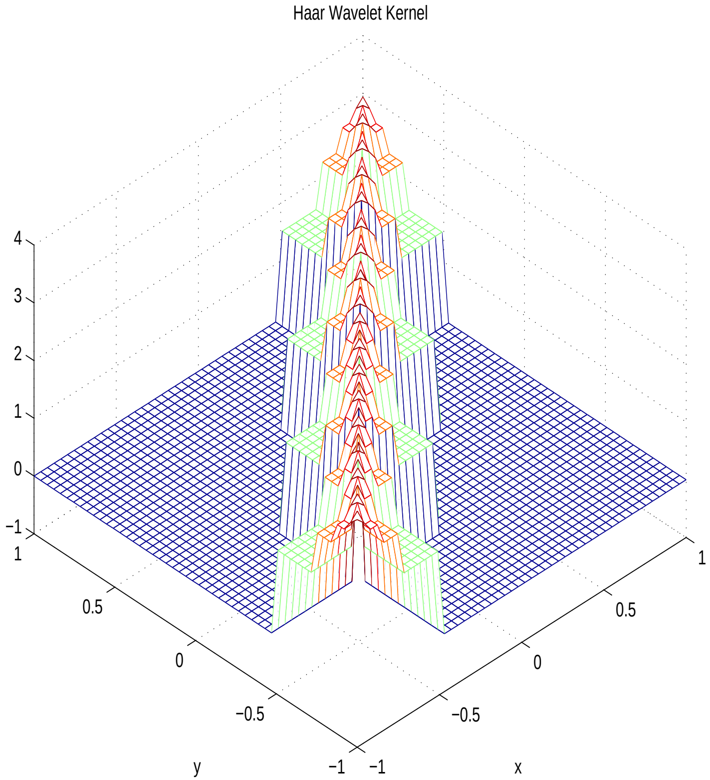
- Trick 1: $\mathcal{H}^d = \otimes_{i=1}^d \mathcal{H}$ and apply Aronszajn's result on tensor-product kernel

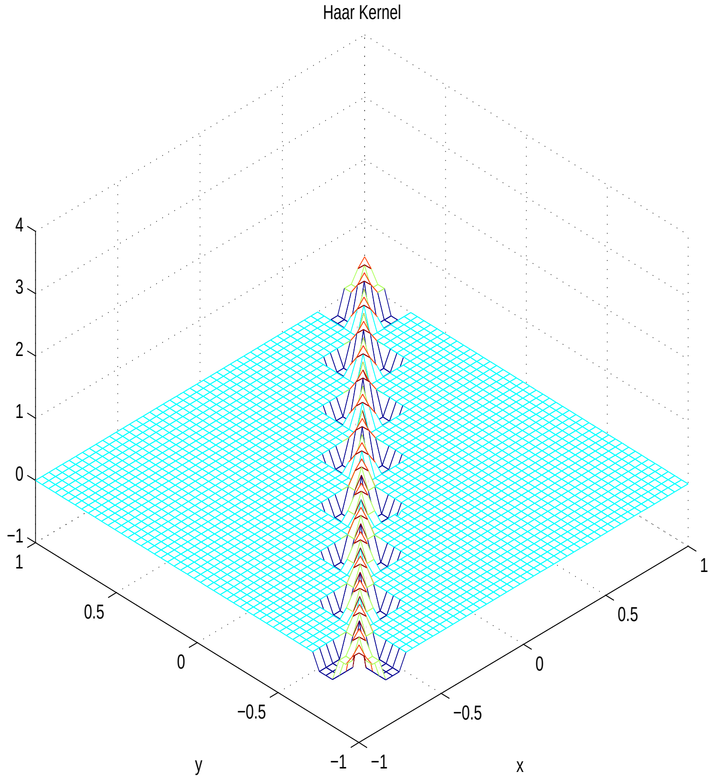
$$K_d(x,y) = \Pi_{m=1}^d K_m(x,y) = \Pi_{m=1}^d \sum_{j,k} \alpha_{j,k}^2 \psi_{j,k}(x_m)\psi_{j,k}(y_m)$$

  where $\psi_{j,k}(\cdot)$ is a 1-dimensional wavelet.

- Trick 2 : If ψ is a compact support wavelet, for any x, only few $\psi_i(x)$ is non-zero

Examples of 1D Haar wavelet kernels

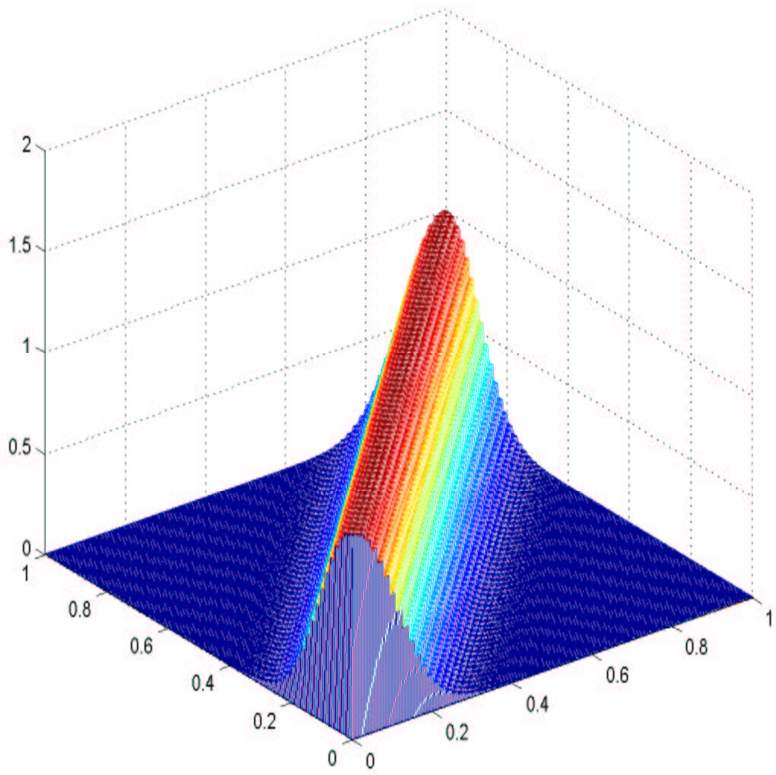




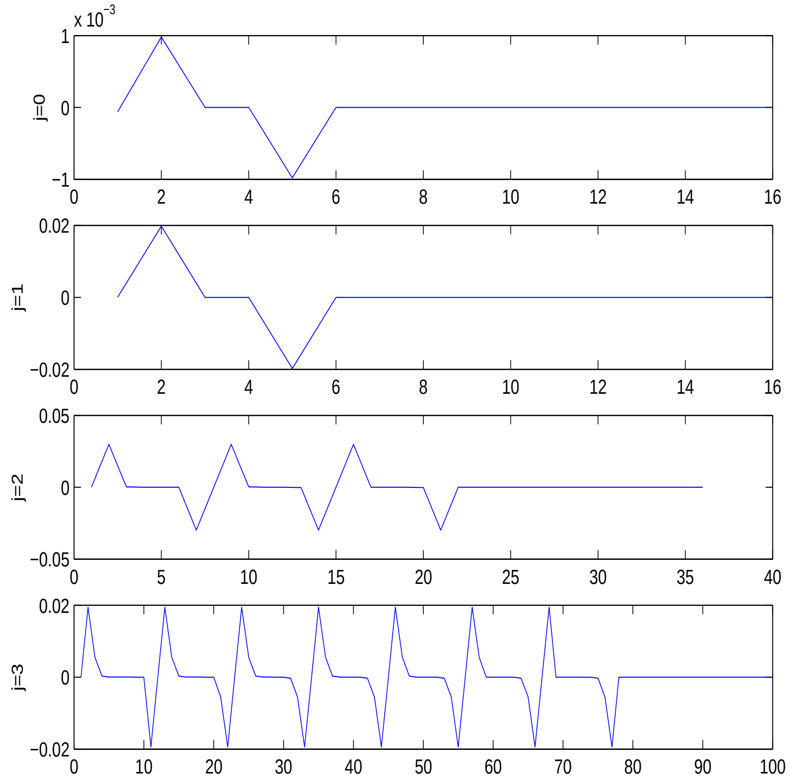
$$\alpha_{j,k}^2 = \begin{cases} \frac{1}{2^j} & j = 0 \ldots 4, k \in \mathbb{Z} \\ 0 & \text{otherwise} \end{cases}$$

$$\alpha_{j,k}^2 = \begin{cases} \frac{1}{2^j} & j = 2 \ldots 4, k \in \mathbb{Z} \\ 0 & \text{otherwise} \end{cases}$$

Example of gaussian-shaped Haar wavelet kernels



Kernel



$\alpha_{j,k}$ coefficients

Does wavelet kernel can approximate other kernels?

Example of performance on benchmarks problem

- Algorithm : SVM
- Kernel : Gaussian and Wavelets
- Gaussian hyperparameters from Raetsch et al (Raetsch,2000).
- No kernel optimization for wavelet kernels.
- Test Error :

Datasets	Features	Wavelet Kernel	Gaussian Kernel
Checkers	2	9.01 ± 4.00	15.59 ± 1.6
Breast Cancer	9	28.81 ± 4.56	26.00 ± 4.7
Diabetis	8	28.04 ± 2.13	23.57 ± 1.7

Conclusions and Perspectives

- Constructive methods for building RKHS
- Wavelet can be used in a learning problem. Among all possible wavelet kernels, how to choose the “best” one?
- Analyse properties of wavelet kernels
- Propose algorithms that takes advantage of wavelet properties