Some contributions of the maxiset approach to the study of wavelet-based methods in nonparametric function estimation

Jean-Marc Freyermuth

Statistical laboratory, University of Cambridge

based on joint projects with
F. Autin : Université de Marseille
G. Claeskens KU Leuven
C. Pouet : ecole centrale de Marseille
R. Sachs : Universite Catholique de Louvain la Neuve

Marseille, Feb 2016

Summary

1. Wavelet-based method and information pooling
2. Unidimensional second generation thresholding methods : maxiset approach
3. Multidimensional function estimation
4. Structure detection : asymptotic optimality

Wavelet nested multiscale structure

Definition (Compactly supported periodized wavelet bases)

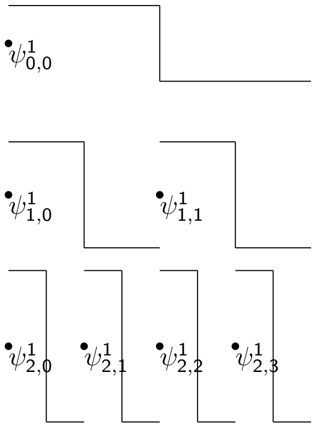
$$B_1 = \left\{ \phi := \psi^0, \psi^1_{j,k}(\cdot) = 2^{j/2}\psi^1\left(2^j \cdot - k\right); \right.$$
$$\left. j \in \mathbb{N}; 0 \leq k < 2^j \right\}.$$

Wlog, consider $f \in L_2[0,1]$,

$$f = \alpha\psi^0 + \sum_{j\geq 0}\sum_{k=0}^{2^j-1} \theta^1_{j,k}\psi^1_{j,k},$$

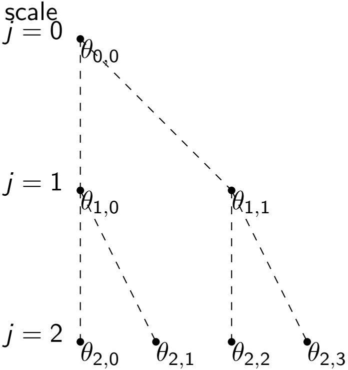
where,

$$\alpha = \left\langle f, \psi^0 \right\rangle$$
$$\theta_{j,k} = \left\langle f, \psi^1_{j,k} \right\rangle$$



Notion of heredity in the coefficient domain

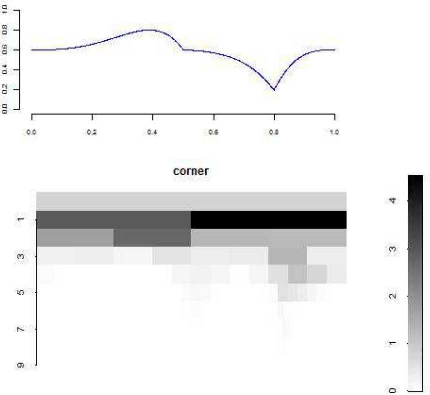
$\mathcal{T}_{j,k}$ the full tree rooted at (j, k)



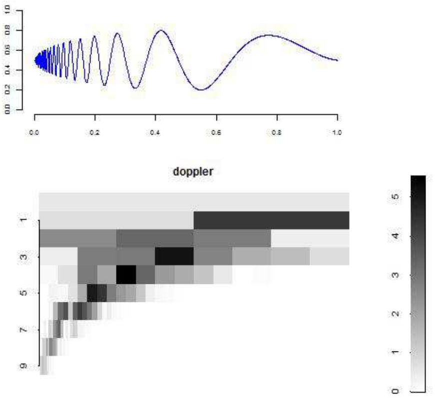
$\theta_{0,0}$							
$\theta_{1,0}$				$\theta_{1,1}$			
$\theta_{2,0}$		$\theta_{2,1}$		$\theta_{2,2}$		$\theta_{2,3}$	
...	...	...	...	...	...	...	...

heredity : $\theta_{j,k}$ is the parents of its two children coefficients $\theta_{j+1,2k}$ and $\theta_{j+1,2k+1}$.

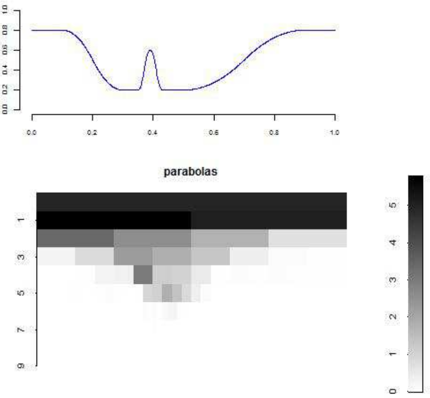
Clusters of large wavelet coefficients



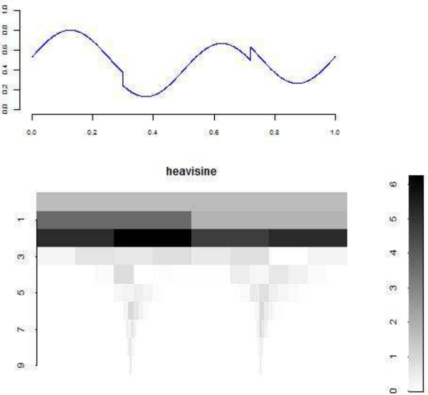
Clusters of large wavelet coefficients



Clusters of large wavelet coefficients



Clusters of large wavelet coefficients



Exploiting the multiscale dependencies

- **Tree approximation** (Cohen et al 2001) is more restrictive than the usual $n-$term approximation but it comes with a little costs in terms of rates of approximation : nevertheless it enables efficient encoding strategies.
- **Multifractal analysis** In practice, multifractal analysis is almost always performed using the coefficients of a continuous or discrete wavelet transform. Jaffard et al introduced a multifractal formalism based on wavelet leaders $\sup_{I' \in 3\mathcal{T}_{j,k}} |\theta_{I'}|$ that allows for accurate estimation of the local hölder exponent,
- **Tree-structured wavelet estimation** : information pooling

- Cohen, A., Dahmen, W., Daubechies, I., Devore, R. (2001). Tree Approximation and Optimal Encoding, Applied and Computational Harmonic Analysis, 11, 192226.
- Jaffard, S., Lashermes, B., Abry, P. (2006). Wavelet leaders in multifractal analysis. In wavelet analysis and applications, Applied and numerical harmonic analysis, 201-246.

Nonparametric function estimation using wavelet thresholding under tree constraint

Consider the nonparametric regression model

$$Y_i = f\left(\frac{i}{n}\right) + \sigma\xi_i,\ \xi_i \sim \mathcal{N}(0,1),\ 1 \leq i \leq n,\ \sigma > 0.$$

- Hard thresholding does not take into account the relations among the wavelet coefficients (influence cones, hereditary structure. . .),

$$\hat{f}^H = \hat{\alpha}\psi^0 + \sum_j \sum_k \hat{\theta}_{j,k}\mathbf{1}_{\{|\hat{\theta}_{j,k}|>\lambda\}}\psi^1_{j,k}.$$

- ⇒ "second generation" wavelet thresholding : Information pooling within some geometric structures in the coefficient domain (horizontal/vertical blocks) to refine the choice of coefficients to Keep/kill,

Many theoretical results about horizontal block thresholding (Cai et al, 1997 ; Hall et al, 1999 ; Cai, 2009) ; not much on TSW.

- Cai, T. (1997), 'On Adaptivity of Blockshrink Wavelet Estimator over Besov Spaces', Technical Report 97-05, Purdue University.
- Hall, P.,Kerkycharian, G., and Picard, D. (1998a), 'Block Threshold Rules for Curve Estimation UsingKernel and Wavelet Methods', Annals of Statistics, 26(3), 922942.
-Hall, P., Kerkycharian, G., and Picard, D. (1998b), 'On the Minimax Optimality for Block Thresholded Wavelet Estimators', Statistica Sinica, 9, 3349.

Nonparametric function estimation using wavelet thresholding under tree constraint

Consider the nonparametric regression model

$$Y_i = f\left(\frac{i}{n}\right) + \sigma\xi_i,\ \xi_i \sim \mathcal{N}(0,1),\ 1 \le i \le n,\ \sigma > 0.$$

- Hard thresholding does not take into account the relations among the wavelet coefficients (influence cones, hereditary structure...),

$$\hat{f}^H = \hat{\alpha}\psi^0 + \sum_j \sum_k \hat{\theta}_{j,k} \mathbf{1}_{\{|\hat{\theta}_{j,k}|>\lambda\}} \psi^1_{j,k}.$$

- $\Rightarrow$ "second generation" wavelet thresholding : Information pooling within some geometric structures in the coefficient domain (horizontal/vertical blocks) to refine the choice of coefficients to Keep/kill,

the tree / hereditary constraint imposes that if a coefficient is used to compute the estimator, all its ancestors have to be included. Tree-Structured Wavelets (TSW).
How to impose tree structure :

- Greedy Tree Approximation (GTA)
- Condensing Sort and Select Algorithm (CSSA)
- Complexity Penalized Sum if Squares (CPSS)

Model and family of estimators

Definition (Sequential Gaussian White noise)

$$\hat{\alpha} = \alpha + \varepsilon\xi \qquad \hat{\theta}_{j,k} = \theta_{jk} + \varepsilon\xi_{j,k}, j \in \mathbb{N}; 0 \leq k < 2^j.$$

where $\xi, \xi_{j,k}$ are i.i.d. $\mathcal{N}(0,1)$ and $\varepsilon \in]0,1[$ (noise level).

Definition ((λ, q)-VBT estimators)

Use a coarse to fine dynamic programming algorithm on the tree. Attributes ℓ_q scores to overlapping vertical blocs of coefficients.
Let $q \in [1, +\infty]$. The estimator $\hat{f}_q$ is defined by :

$$\hat{f}_q = \hat{\alpha}\psi^0 + \sum_{j<j_{\lambda_\varepsilon}} \sum_{k=0}^{2^j-1} \hat{\theta}_{j,k} \, \mathbf{1}\left\{(j,k) \in \tau_q(\hat{\theta}, \lambda_\varepsilon)\right\} \psi^1_{j,k}.$$

Model and family of estimators

Remark :

- $\hat{f}_2$: Oracle approach (Donoho, 1997).
- $\hat{f}_\infty$: hard thresholding + hereditary constraint (Autin, 2008).

Definition (HARD TREE estimator ($\hat{f}_\infty = \hat{f}^{HT}$))

$$\hat{f}^{HT}(\cdot) = \hat{\alpha}\psi^0(\cdot) + \sum_{j<j_{\lambda_\varepsilon}} \sum_{k=0}^{2^j-1} \hat{\theta}_{j,k} \mathbf{1}\left\{ \max_{(j',k')\in\mathcal{T}_{j,k}} \left|\hat{\theta}_{j',k'}\right| > \lambda_\varepsilon \right\} \psi^1_{j,k}(\cdot).$$

- Tomassi, D., Milone, D., Nelson, J.D.B. (2015) Wavelet shrinkage using adaptive structured sparsity constraint. Signal Processing, 106, 73-87

Use the 'recent' Dual Tree Complex Wavelet Transform (Kingsbury, 2001) : nearly shift invariant and has low redundancy.

Tree-structured estimation using a lasso-type algorithm with a mixed norm regularizer that induces structured sparsity of an overcomplete representation (overlapping group lasso) with adaptive weights.

It has been shown that there is more information to retrieve intra-inter scale within DTCWT coefficients domain than DWT.

- Kingsbury, N. (2001). Complex Wavelets for Shift Invariant Analysis and Filtering of Signals. ACHA, 10, 234253

Model and family of estimators

Remark :

- $\hat{f}_2$: Oracle approach (Donoho, 1997).
- $\hat{f}_\infty$: hard thresholding + hereditary constraint (Autin, 2008).

Definition (HARD TREE estimator ($\hat{f}_\infty = \hat{f}^{HT}$))

$$\hat{f}^{HT}(\cdot) = \hat{\alpha}\psi^0(\cdot) + \sum_{j<j_{\lambda_\varepsilon}} \sum_{k=0}^{2^j-1} \hat{\theta}_{j,k} \mathbf{1}\left\{ \max_{(j',k')\in\mathcal{T}_{j,k}} \left|\hat{\theta}_{j',k'}\right| > \lambda_\varepsilon \right\} \psi^1_{j,k}(\cdot).$$

- Tomassi, D., Milone, D., Nelson, J.D.B. (2015) Wavelet shrinkage using adaptive structured sparsity constraint. Signal Processing, 106, 73-87

Use the 'recent' Dual Tree Complex Wavelet Transform (Kingsbury, 2001) : nearly shift invariant and has low redundancy.

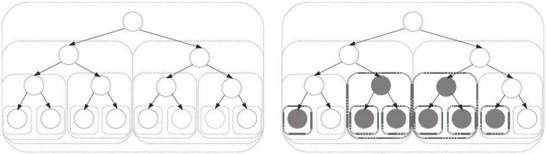


Fig. 2. (a) Grouping scheme for the proposed mixed-norm regularisation and (b) example of estimate induced by the adopted grouping scheme; its support is the complement of the union of the sets pushed to zero during optimisation which in turn sets to zero the coefficients represented by shaded circles.

FIGURE: forest like hierarchical tree structure

MAXISET THEORY

<u>Remark :</u> Minimax approach has some drawbacks :

- choice of $\mathcal{F}$,
- pessimistic approach (at least w.r.t maxiset approach),
- no comparison of *optimal* estimators.

Definition (Maxiset approach (Cohen, Kerkyacharian, Picard (2001)))

The maxiset of an estimator $\hat{f}_\varepsilon$ is the largest functional space where its risk attains a given convergence rate v_ε :

$$MS_2\left(\hat{f}_\varepsilon, v_\varepsilon\right) = \left\{ f : \sup_{0<\varepsilon<1} {v_\varepsilon}^{-1}\mathbb{E}\|\hat{f}_\varepsilon - f\|_2^2 < +\infty \right\}.$$

Remarks :

- Estimator with good *performance* $\Longleftrightarrow$ *large* Maxiset.
- Rate $\nearrow$ Maxiset $\searrow$.
- Usually $v_\varepsilon = v_{\mathcal{F},\varepsilon}$ or $v_{\mathcal{F},\varepsilon}|\ln\varepsilon|^\delta$
- estimator $\leftrightarrow$ maxiset

MAXISET THEORY

Remark : Minimax approach has some drawbacks :

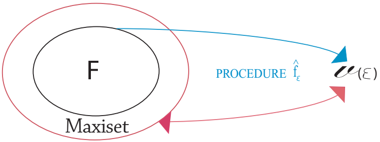
- choice of $\mathcal{F}$,
- pessimistic approach (at least w.r.t maxiset approach),
- no comparison of *optimal* estimators.

Definition (Maxiset approach (Cohen, Kerkyacharian, Picard (2001)))

The maxiset of an estimator $\hat{f}_\varepsilon$ is the largest functional space where its risk attains a given convergence rate v_ε :

$$MS_2\left(\hat{f}_\varepsilon, v_\varepsilon\right) = \left\{ f : \sup_{0<\varepsilon<1} {v_\varepsilon}^{-1}\mathbb{E}\|\hat{f}_\varepsilon - f\|_2^2 < +\infty \right\}.$$

Remarks :



VERTICAL BLOC THRESHOLDING

Theorem (Autin, F. et von Sachs (2011))

Let $s > 0$, $q \in \ [1, +\infty]$, $m \geq 4\sqrt{3}$ *and* $v_\varepsilon = \left(\varepsilon\sqrt{|\ln \varepsilon|}\right)^{\frac{4s}{1+2s}}$.

$$MS_2\left(\hat{f}_q, v_\varepsilon\right) = \mathcal{B}_{2,\infty}^{\frac{s}{1+2s}} \cap \mathcal{W}^V_{\frac{2}{1+2s},q}.$$

$$\mathcal{B}_{2,\infty}^s = \left\{ f; \sup_{J\in\mathbb{N}} 2^{2Js} \sum_{j\geq J} \sum_{k=0}^{2^j-1} |\theta_{j,k}|^2 < \infty \right\},$$

$$\mathcal{W}^V_{r,q} = \left\{ f : \sup_{0<\lambda<1} \lambda^{r-2} \sum_j \sum_k \theta_{jk}^2 \, \mathbf{1}\left\{(j,k) \notin \tau_q(\theta,\lambda)\right\} < +\infty \right\},$$

$$\bigcap$$

$$\mathcal{W}^V_{r,q'} = \left\{ f : \sup_{0<\lambda<1} \lambda^{r-2} \sum_j \sum_k \theta_{jk}^2 \, \mathbf{1}\left\{(j,k) \notin \tau_{q'}(\theta,\lambda)\right\} < +\infty \right\}, q' > q.$$

Autin, F., F, J-M., von Sachs, R. (2011). Ideal denoising within a family of tree-structured wavelet estimators. *Electronic Journal of Statistics*, 5, 829–855.

VERTICAL BLOC THRESHOLDING

Theorem (Autin, F. et von Sachs (2011))

Let $s > 0$, $q \in \ [1, +\infty]$, $m \geq 4\sqrt{3}$ *and* $v_\varepsilon = \left(\varepsilon\sqrt{|\ln \varepsilon|}\right)^{\frac{4s}{1+2s}}$.

$$MS_2\left(\hat{f}_q, v_\varepsilon\right) = \mathcal{B}^{\frac{s}{1+2s}}_{2,\infty} \cap \mathcal{W}^{V}_{\frac{2}{1+2s},q}.$$

Theorem (Autin, F. and von Sachs (2011))

From the maxiset point of view :

1. *the best estimator is :* $\hat{f}_\infty$ *(Hard Tree).*
2. *moreover* $\hat{f}_\infty$ *is better estimator than* $\hat{f}^H$.

Autin, F., F, J-M., von Sachs, R. (2011). Ideal denoising within a family of tree-structured wavelet estimators. *Electronic Journal of Statistics*, 5, 829–855.

NUMERICAL EXPERIMENTS

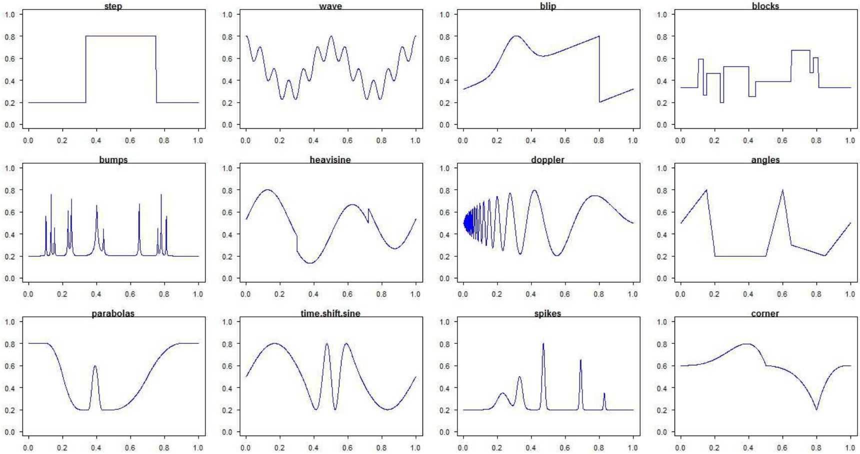


FIGURE: Test functions

NUMERICAL EXPERIMENTS

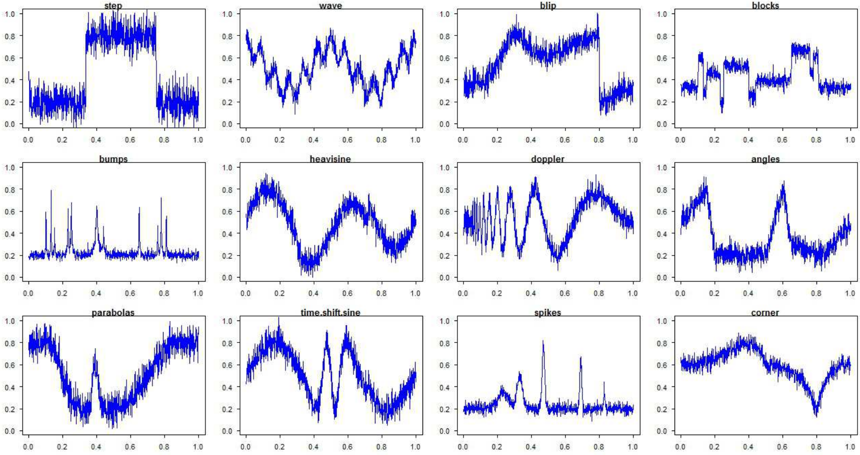


FIGURE: Noisy functions

NUMERICAL EXPERIMENTS

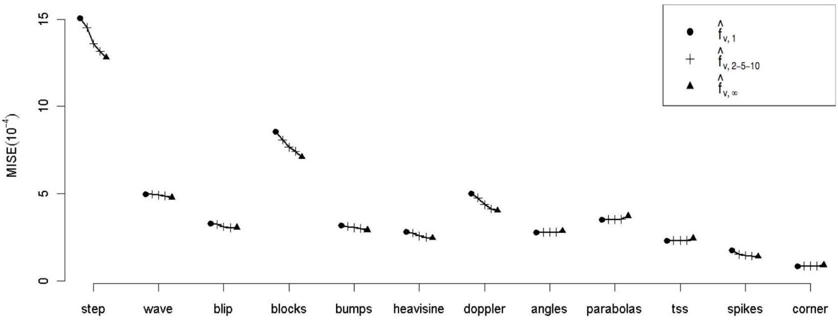


FIGURE: *MISE* estimators $\hat{f}_q$ (Db8, SNR = 10db, N=1024)

Multidimensional function estimation

Multivariate wavelet bases

Motivation : multivariate wavelet bases are well adapted to described functions in $\mathbb{L}_2([0,1]^d)$ in a parsimonious way. Atoms are localized both in time and frequency.

Two popular multidimensional wavelet bases :

1. Standard wavelet basis.
2. Hyperbolic wavelet basis.

Actually, a large regain of interest in :

- Statistics (Neumann (2000), Benhaddou et al. (2013)).
- Approximation theory (Kerkyacharian et al. (2006)).
- Compressive sensing (Duarte, Baraniuk (2012).
- Multifractal analysis (Abry et al. (2014))

- Benhaddou, R., Pensky, M., Picard, D. (2013) Anisotropic denoising in functional deconvolution model with dimension-free convergence rates, EJS, 7, 1686-1715.
- Duarte, M.F., Baraniuk, R. (2012) Kronecker Compressive Sensing. IEEE Transactions on Image Processing, 21(2), 494-504.
- P. Abry, M. Clausel, S. Jaffard, S.G. Roux, B. Vedel, Hyperbolic wavelet transform : an efficient tool for multifractal analysis of anisotropic textures.Revista Matematica Iberoamericana, vol 31(1), pp 313-348 (2015)
- Kerkyacharian, G., Picard, D., Temlyakov, V. (2006) Some inequalities for the tensor product of greedy bases and weight-greedy bases. East journal on approximations, 12(1), 103-118.

Standard and hyperbolic wavelet bases

Definition

From a periodized univariate wavelet basis

$$\mathcal{B}_1 = \left\{ \psi^0, \psi^1_{j,k} : j \in \mathbb{N}, 0 \le k < 2^j \right\},$$

- *the standard wavelet basis is defined as :*
$$\mathcal{I}_d = \left\{ \psi^{\underline{0}}_{\underline{0},\underline{0}}, \psi^{\underline{i}}_{\underline{j},\underline{k}} : \underline{i} \in \{0,1\}^d \setminus \underline{0},\ \underline{j} \in \mathbb{J},\ \underline{k} \in \mathbb{K}_{\underline{j}} \right\},$$
- *the hyperbolic wavelet basis is defined as :*
$$\mathcal{H}_d = \left\{ \psi^{\underline{0}}_{\underline{0},\underline{0}}, \psi^{\underline{i}}_{\underline{j},\underline{k}} : \underline{i} \in \{0,1\}^d \setminus \underline{0},\ \underline{j} \in \mathbb{J}^{\underline{i}},\ \underline{k} \in \mathbb{K}_{\underline{j}} \right\}.$$

$$\begin{aligned}
\psi^{\underline{0}}_{\underline{0},\underline{0}} &= \psi^0 \times \cdots \times \psi^0 \quad \text{and} \quad \psi^{\underline{i}}_{\underline{j},\underline{k}} = \psi^{i_1}_{j_1,k_1} \times \cdots \times \psi^{i_d}_{j_d,k_d}, \\
\mathbb{J} &= \left\{ \underline{j} = (j, \cdots, j) : j \in \mathbb{N} \right\}, \\
\mathbb{J}^{\underline{i}} &= \left\{ \underline{j} = (j_1 i_1, \cdots, j_d i_d) : \forall u, j_u \in \mathbb{N} \right\}, \\
\mathbb{K}_{\underline{j}} &= \left\{ \underline{k} = (k_1, \cdots, k_d) : \forall u, 0 \le k_u < 2^{j_u} \right\}.
\end{aligned}$$

2-D hyperbolic wavelet basis

Let ψ^1 be a 1-D wavelet and ψ^0 its associated scaling function.

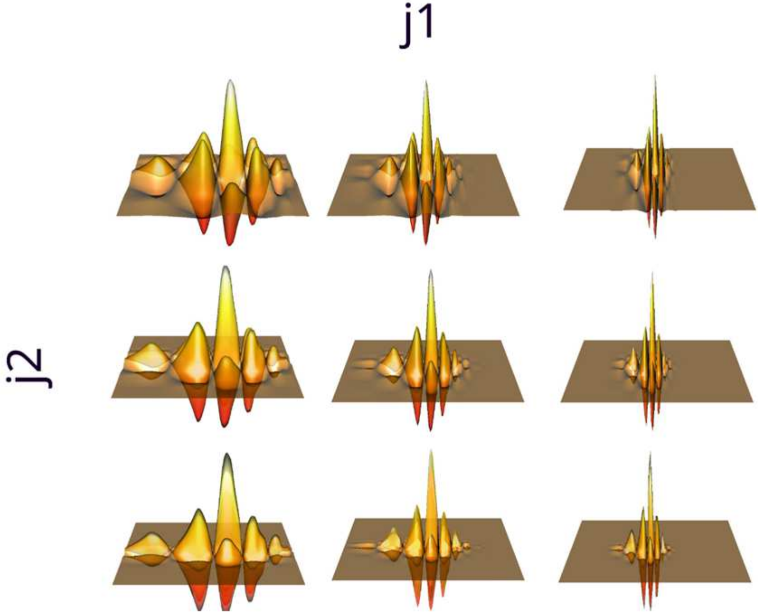
$$\begin{array}{rcll} \psi^{(0,0)}(x,y) & = & \psi^0(x)\,\psi^0(y), & |\underline{i}| = 0 \\ \psi^{(1,0)}_{j_1,k_1}(x,y) & = & \psi^1_{j_1 k_1}(x)\,\psi^0(y), & |\underline{i}| = 1 \\ \psi^{(0,1)}_{j_2,k_2}(x,y) & = & \psi^0(x)\,\psi^1_{j_2 k_2}(y), & |\underline{i}| = 1 \\ \psi^{(1,1)}_{j_1,j_2,k_1,k_2}(x,y) & = & \psi^1_{j_1 k_1}(x)\,\psi^1_{j_2 k_2}(y), & |\underline{i}| = 2 \end{array}$$

The support is a rectangle of size $2^{-j_1} \times 2^{-j_2}$: optimal to adapt to anisotropic smoothness.

Consider $f \in L_2([0,1]^2)$.

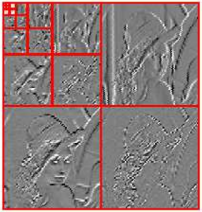
$$\begin{aligned} f &= \underbrace{\text{avg approx}} + \underbrace{\text{horizontal details}} + \underbrace{\text{vertical details}} + \underbrace{\text{diagonal details}} \\ &= \alpha\phi_{00} + \sum_{j_1 \geq 0}\sum_{k_1} \theta^{(1,0)}_{j_1 k_1}\psi^{(1,0)}_{j_1 k_1} + \sum_{j_2 \geq 0}\sum_{k_2} \theta^{(0,1)}_{j_2 k_2}\psi^{(0,1)}_{j_2 k_2} + \sum_{j_1,j_2 \geq 0}\sum_{k_1,k_2} \theta^{(1,1)}_{j_1 j_2 k_1 k_2}\psi^{(1,1)}_{j_1,j_2,k_1,k_2} \\ &= \alpha\phi_{00} + \sum_{\underline{i}=\{h,v,d\}}\sum_{\underline{j}\in\mathbb{J}^{\underline{i}}}\sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} \theta^{\underline{i}}_{\underline{j},\underline{k}}\psi^{\underline{i}}_{\underline{j},\underline{k}} \end{aligned}$$

Hyperbolic wavelet functions

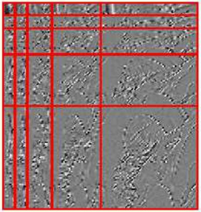


Standard vs Hyperbolic wavelet transforms

lena



Isotropic



Hyperbolic

Wavelet decomposition of a multivariate function

Consider $f \in \mathbb{L}_2([0,1]^d)$:

$$f = \text{avg approx} + \sum_{\underline{i} \in \{0,1\}^d \setminus \underline{0}} \text{details in orientation } \underline{i}$$

$$= \theta_{\underline{0},\underline{0}}^{\underline{0}} \psi_{\underline{0},\underline{0}}^{\underline{0}} + \sum_{\underline{i} \in \{0,1\}^d \setminus \underline{0}} \left(\sum_{\underline{j} \in \mathcal{J}} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} \theta_{\underline{j},\underline{k}}^{\underline{i}} \psi_{\underline{j},\underline{k}}^{\underline{i}} \right),$$

where

- $\mathcal{J} = \mathbb{J}$ for the standard wavelet decomposition,
- $\mathcal{J} = \mathbb{J}^{\underline{i}}$ for the hyperbolic wavelet decomposition.

Remark : in the multivariate wavelet setting

$$\|f\|_2^2 = (\theta_{\underline{0},\underline{0}}^{\underline{0}})^2 + \sum_{\underline{i} \in \{0,1\}^d \setminus \underline{0}} \sum_{\underline{j} \in \mathcal{J}} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} \left(\theta_{\underline{j},\underline{k}}^{\underline{i}} \right)^2 .$$

Besov spaces of smooth functions

Definition (Characterisation of Besov space (Abry et al. (2014)))

The Besov space $\mathcal{B}^{\underline{s}}$ is defined as follows :

$$\mathcal{B}^{\underline{s}} = \left\{ f \in \mathbb{L}_2([0,1]^d) : \sup_{\underline{i} \neq \underline{0}} \sup_{\underline{j} \in \mathbb{J}^{\underline{i}}} \left(\max_{1 \le u \le d} 2^{2 j_u s_u} \right) \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} \left(\theta^{\underline{i}}_{\underline{j},\underline{k}} \right)^2 < +\infty \right\}.$$

- Isotropic Besov spaces : $\underline{s} = (s_1, \cdots, s_d) : \forall (u,v),\ s_u = s_v.$
- Anisotropic Besov spaces : $\underline{s} = (s_1, \cdots, s_d) : \exists (u,v),\ s_u \neq s_v.$

Theorem (Neumann (2000))

Minimax rate over Besov spaces

$$\mathcal{R}^*_\varepsilon(\mathcal{B}^{\underline{s}}) \asymp \varepsilon^{\frac{4\gamma}{1+2\gamma}}.$$

$$\text{where } \gamma = \left(s_1^{-1} + \cdots + s_d^{-1} \right)^{-1} > \frac{1}{2}.$$

Hierarchical structure

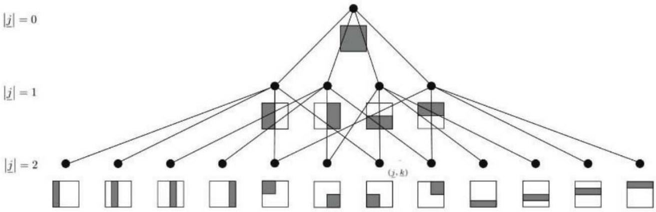


FIGURE: Hierarchical structure $\mathcal{T}^{\underline{i}}$ coming from the MRA ($d = 2, \underline{i} = (1, 1)$)

Proposition (Autin, Claeskens and F.(2014))

Consider $\underline{i} \in \{0, 1\}^d \setminus \underline{0}$. *Each node* $(\underline{j}, \underline{k})$ *of* $\tau_{\infty}^{\underline{i}}$,

1. *has* $2|\underline{i}|$ *children at scales* $|\underline{j}| + 1$ *;*
2. *has at most* $|\underline{i}|$ *ancestors at scales* $|\underline{j}| - 1$ *and a total number of ancestors equals to* $\prod_{u=1}^{d}(j_u + 1)^{i_u}$.

HEREDITARY STRUCTURE

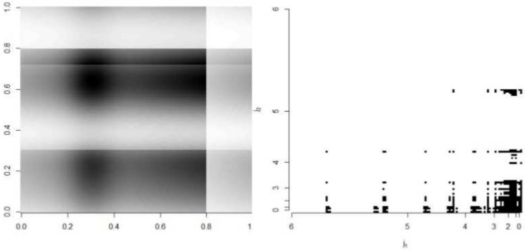


FIGURE: Hereditary structure among θ_{j_1,j_2,k_1,k_2}

Estimation in Hyperbolic wavelet basis

Hyperbolic basis :

$$H_d = \left\{ \psi^{\underline{0}}_{\underline{0},\underline{0}},\ \psi^{\underline{i}}_{\underline{j},\,\underline{k}} : \quad \underline{i} \in \{0,1\}^d \setminus \underline{0};\quad \underline{j} \in \mathbb{J}^{\underline{i}};\ \underline{k} \in \mathbb{K}_{\underline{j}} \right\}.$$

Definition

Hard Thresholding estimator ($c \geq 1/2$) :

$$\hat{f}^H_c \;=\; \hat{\alpha}\phi_{\underline{0},\underline{0}} + \sum_{\underline{i}\neq\underline{0}} \sum_{\underline{j}\in\mathbb{J}^{\underline{i}},\ |\underline{j}|<j_{\lambda_{\varepsilon,c}}} \sum_{\underline{k}} \hat{\theta}^{\underline{i}}_{\underline{j},\underline{k}} \mathbf{1}\left\{ |\hat{\theta}^{\underline{i}}_{\underline{j},\underline{k}}| > \lambda_{\varepsilon,c} \right\} \psi^{\underline{i}}_{\underline{j},\underline{k}}$$

Hard Tree estimator ($c \geq 1/2$) :

$$\hat{f}^{HT}_c = \hat{\alpha}\phi_{\underline{0},\underline{0}} + \sum_{\underline{i}\neq\underline{0}} \sum_{\underline{j}\in\mathbb{J}^{\underline{i}},\ |\underline{j}|<j_{\lambda_{\varepsilon,c}}} \sum_{\underline{k}} \hat{\theta}^{\underline{i}}_{\underline{j},\underline{k}} \mathbf{1}\left\{ \max_{\underline{j}',\underline{k}'\in\mathcal{T}^{\underline{i}}_{\underline{j},\underline{k}}(\lambda_{\varepsilon,c})} |\hat{\theta}^{\underline{i}}_{\underline{j}',\underline{k}'}| > \lambda_{\varepsilon,c} \right\} \psi^{\underline{i}}_{\underline{j},\underline{k}}$$

$$\text{with} \quad \lambda_{\varepsilon,c} \;=\; \varepsilon|\ln\varepsilon|^c \ \text{and}\ 2^{-j_{\lambda_{\varepsilon,c}}} = (\lambda_{\varepsilon,c})^2.$$

RESULTATS MAXISETS

Theorem (Autin, Claeskens et Freyermuth (2014))

Let $\gamma > 0$ $p \geq 2$ *et* $c \geq 1/2$. *Si* $m \geq 4\sqrt{p+1}$, *then :*

$$MS_p\left(\hat{f}^H, (\varepsilon|\ln\varepsilon|^c)^{\frac{2\gamma p}{1+2\gamma}}\right) = \mathcal{A}_{p,\infty}^{\frac{\gamma}{1+2\gamma}} \cap \mathcal{W}^H_{\frac{p}{1+2\gamma}},$$

$$MS_p\left(\hat{f}^{HT}, (\varepsilon|\ln\varepsilon|^c)^{\frac{2\gamma p}{1+2\gamma}}\right) = \mathcal{A}_{p,\infty}^{\frac{\gamma}{1+2\gamma}} \cap \mathcal{W}^{HT}_{\frac{p}{1+2\gamma}} \cap \mathcal{W}^{HT,\star}_{\frac{p}{1+2s},c}.$$

$$\mathcal{A}^r_{p,\infty} = \left\{ f : \sup_{J\in\mathbb{N}} \sum_{\underline{i}\neq 0} \sum_{\underline{j}\in\mathbb{J}^{\underline{i}};\ |\underline{j}|\geq J} 2^{Jrp+|\underline{j}|(p/2-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} \left|\theta^{\underline{i}}_{\underline{j},\underline{k}}\right|^p < +\infty \right\},$$

$$\mathcal{W}^H_r = \left\{ f : \sup_{0<\lambda<1} \lambda^{r-p} \sum_{\underline{i}\neq 0} \sum_{\underline{j}\in\mathbb{J}^{\underline{i}}} 2^{|\underline{j}|(p/2-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} \left|\theta^{\underline{i}}_{\underline{j},\underline{k}}\right|^p \mathbf{1}\left\{\left|\theta^{\underline{i}}_{\underline{j},\underline{k}}\right| \leq \lambda\right\} < +\infty \right\},$$

$$\cap$$

$$W^{HT}_r = \left\{ f : \sup_{0<\lambda<1} \lambda^{r-p} \sum_{\underline{i}\neq 0} \sum_{\underline{j}\in\mathbb{J}^{\underline{i}}} 2^{|\underline{j}|(p/2-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} \left|\theta^{\underline{i}}_{\underline{j},\underline{k}}\right|^p \mathbf{1}\left\{\max_{\underline{j}',\underline{k}'\in\mathcal{T}^{\underline{i}}_{\underline{j},\underline{k}}(\lambda)} |\theta^{\underline{i}}_{\underline{j}',\underline{k}'}| \leq \lambda\right\} < +\infty \right\},$$

$$W^{HT,\star}_{r,c} = \left\{ f : \sup_{0<\lambda<1} \lambda^{r} \log(\lambda^{-1})^{-pc} \sum_{\underline{i}\neq 0} \sum_{\underline{j}\in\mathbb{J}^{\underline{i}}} 2^{|\underline{j}|(p/2-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} \mathbf{1}\left\{\max_{\underline{j}',\underline{k}'\in\mathcal{T}^{\underline{i}}_{\underline{j},\underline{k}}(\lambda)} |\theta^{\underline{i}}_{\underline{j}',\underline{k}'}| > \lambda\right\} < +\infty \right\}.$$

Autin, F., Claeskens, G., Freyermuth, J-M. (2014). Hyperbolic wavelet thresholding rules : the curse of dimensionality through the maxiset approach. *Applied and Computational Harmonic Analysis*, 36, 239-255.

Curse of dimensionality

Constat : Dimension $d \nearrow$ sequence space $\mathcal{W}^{\star}_{r,c} \searrow$.
Is *information pooling* still interesting for $d > 1$?

Theorem (Autin, Claeskens et F. (2014))

Let $p \geq 2$ and $c \geq 1/2$. Then

$$MS_p\left(\hat{f}^H, (\varepsilon|\ln\varepsilon|^c)^{\frac{2\gamma p}{1+2\gamma}}\right) \subset MS_p\left(\hat{f}^{HT}, (\varepsilon|\ln\varepsilon|^c)^{\frac{2\gamma p}{1+2\gamma}}\right),$$

once one of the following properties is satisfied :

1. $d \leq pc$ *(small dimension),*
2. $d > pc$ *(large dimension) and structural constraint* $[H_p(c)]$.

Definition

A function $f \in L_p([0,1])$ satisfies the **structural constraint** $[H_p(c)]$ iff, for all $|\underline{i}| \geq pc$ and all $0 < \lambda < 1$,

$$|\langle f, \psi^{\underline{i}}_{\underline{j},\underline{k}}\rangle| > \lambda \implies \underline{j} \text{ is such that } \max\{j_u;\ 1 \leq u \leq d\} \leq |\ln\lambda|^{\frac{pc}{|\underline{i}|}}.$$

CURSE OF DIMENSIONALITY

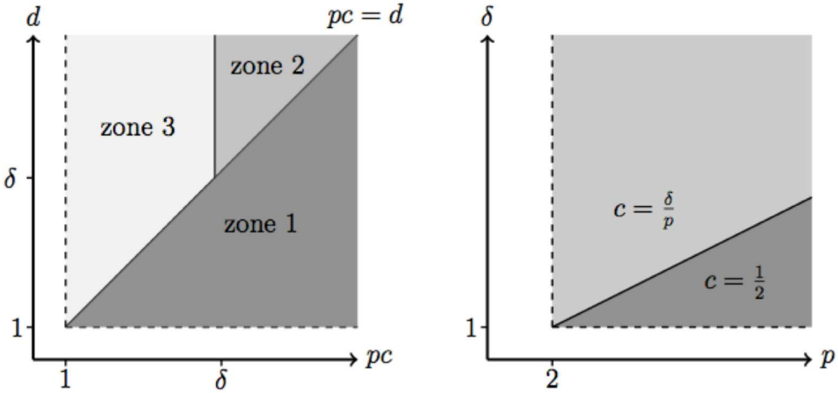


FIGURE: *information pooling*? (Y/N)

Zone 1 : $d \leq pc$. Zone 2 : $d > pc$ and $\delta \leq pc$. Zone 3 : $\delta > pc$.

Conclusion

- Information pooling in signal and image processing has promising future in applications
- still lot of challenges to precisely characterized the performances methods using information pooling
- the maxiset approach is an effective way to asses the performance of methods offering a different perspective to problem

Time-varying spectrum

Following R.Dahlhaus the evolutionary spectrum of locally stationary process can be written as :

$$f(u,\omega) \;=\; \lim_{T\to\infty} f_W([uT],\omega)\,, \qquad u=\frac{t}{T} \;\in\; [0,1]$$

with the *Wigner–Ville* spectrum

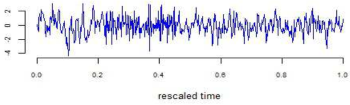
$$f_W(t,\omega) = \sum_s \mathbb{C}\mathrm{ov}\left(X_{t-s/2}\,,X_{t+s/2}\right)\;\exp(-i\omega s)$$

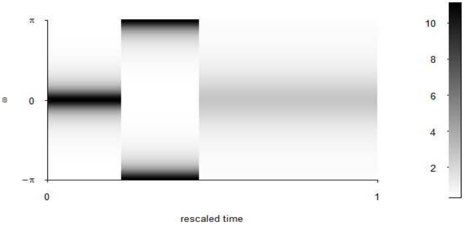
Estimation :
First approaches : Classical periodograms on segments of length N

$$I_N(u,\omega) \;=\; \frac{1}{2\pi N}\left|\sum_{s=0}^{N-1} X_{[uT-N/2+s+1]}\,\exp(-is\omega)\right|^2$$

Drawback : Choice of the length of the blocks N ?

Example : a Piecewise stationary AR process





Time-varying spectrum

Example : a Piecewise stationary AR process

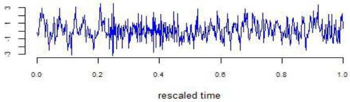
Definition (Preperiodogram/empirical Wigner Ville distribution)

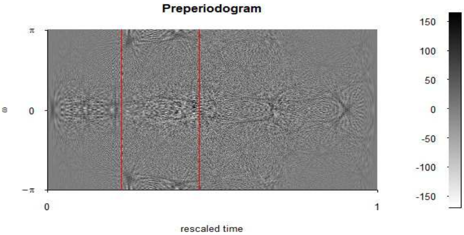
Fourier transform of empirical local autocovariance

Rescaled preperiodogram of a process X

$$I_X(u,\omega) := (2\pi)^{-1} \sum_{s:1\leq uT-s/2,\, uT+s/2\leq T} X_{[uT-s/2],T}\, X_{[uT+s/2],T}\, \exp(-is\omega)$$

$\mathbb{E}\ I_X(u,\omega) \rightarrow f(u,\omega)$ as $T \rightarrow \infty$.





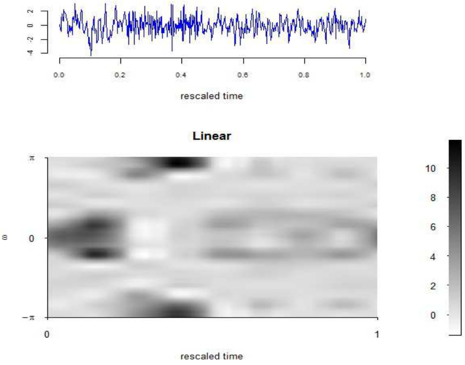
Time-varying spectrum

Smoothing of the preperiodogram

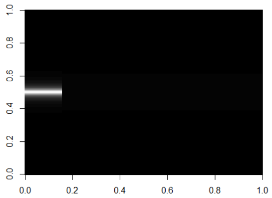
$$\hat{f}(u,\omega) \;:=\; I\,(u,\omega) ** K(u,\omega)$$

where $**$ is the two dimensional convolution operator with a kernel K.

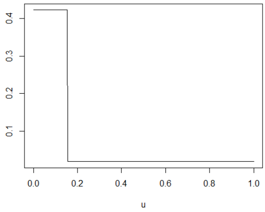
Example : a Piecewise stationary AR process



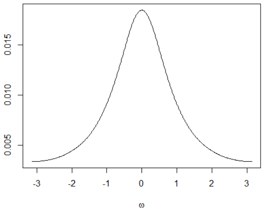
Anisotropy



TimeVaryingspectrum



cut along time



cut along frequency

Structure detection

Sobol decomposition of $f \in L_2\left[0,1\right]^d$

$$f(x_1,\ldots,x_d) = f_0 + \sum_{u=1}^{d} f_u\left(x_u\right) + \sum_{u<v} f_{uv}\left(x_u,x_v\right) + \ldots + f_{1,\ldots,d}\left(x_1,\ldots,x_d\right) \qquad (1)$$

Marginal components (main effect) $f_u : \ [0,1) \longrightarrow \mathbb{R} \ \ (1 \leq u \leq d)$

Dalalyan, A., Ingster, Y., Tsybakov, A. (2014). Statistical inference in compound functional models. *Probability Theory and Related Fields*, 158(3), pp. 512–532.

The atomic dimension δ of f reflects the maximal degree of interaction between the d variables within f.

- Additive structure : $f(\underline{x}) = \sum_{u=1}^{d} f_u(x_u) \ \rightarrow \delta = 1.$
- Sructure with maximal degree of interaction $m < d$:

$$f(\underline{x}) = \sum_{u_1<\cdots<u_m} f_{u_1,\cdots,u_m}(x_{u_1},\cdots,x_{u_m}) \ \rightarrow \delta = m.$$

Structure detection

Sobol decomposition of $f \in L_2[0,1]^d$

$$f(x_1,\ldots,x_d) = f_0 + \sum_{u=1}^{d} f_u(x_u) + \sum_{u<v} f_{uv}(x_u,x_v) + \ldots + f_{1,\ldots,d}(x_1,\ldots,x_d) \quad (1)$$

Marginal components (main effect) $f_u : [0,1) \longrightarrow \mathbb{R} \ (1 \le u \le d)$

Dalalyan, A., Ingster, Y., Tsybakov, A. (2014). Statistical inference in compound functional models. *Probability Theory and Related Fields*, 158(3), pp. 512–532.

The atomic dimension δ of f reflects the maximal degree of interaction between the d variables within f.

Back to Time-Varying spectrum

Question : is the process stationary ?

Multivariate/Multiple nonparameric regression model :

$$Y_{ij} = f(u_i,\omega_j) + \text{error}_{ij}, \ f : [0,1]^2 \to \mathbb{R}$$

i.e., is the variable t_i relevant

$$f(u,\omega) = f_2(\omega) \quad (2)$$

Structure detection

Sobol decomposition of $f \in L_2 [0,1]^d$

$$f(x_1, \ldots, x_d) = f_0 + \sum_{u=1}^{d} f_u(x_u) + \sum_{u<v} f_{uv}(x_u, x_v) + \ldots + f_{1,\ldots,d}(x_1, \ldots, x_d) \quad (1)$$

Marginal components (main effect) $f_u : \ [0,1) \longrightarrow \mathbb{R} \ \ (1 \leq u \leq d)$

Dalalyan, A., Ingster, Y., Tsybakov, A. (2014). Statistical inference in compound functional models. *Probability Theory and Related Fields*, 158(3), pp. 512–532.

The atomic dimension δ of f reflects the maximal degree of interaction between the d variables within f.

Back to Time-Varying spectrum

Question : the process is non stationary but is the behavior similar across the different frequency bands ?

Multivariate/Multiple nonparameric regression model :

$$Y_{ij} = f(u_i, \omega_j) + \text{error}_{ij}, \ \ f : [0,1]^2 \to \mathbb{R}$$

i.e., is there interaction between time and frequency ?

$$f(u, \omega) = f_1(u) + f_2(\omega) \quad (3)$$

Approach

2 characteristic ingredients of our multivariate data :

- general structure (atomic dimension, variable selection)
- anisotropy

Objective : Propose procedures to test for structure of multivariate data with special emphasis of being 'optimal' (or near optimal) under anisotropic smoothness properties.

Method : Find appropriate representation in a wavelet basis → design procedures for detecting some structural characteristics of the estimand :

- maximal degree of interaction of variables (atomic dimension),
- variable selection...

Construct (near) minimax optimal tests

- Abramovich, F. De Feis, I., Sapatinas, T. (2009). Optimal testing for additivity in multiple nonparametric regression. *Ann. Inst. Stat. Math.*, 61, pp. 691–714.
- Comminges, L. Dalalyan, A. (2013). Minimax testing of a composite null hypothesis defined via a quadratic functional in the model of regression *EJS*, 7, pp. 146–190.
- Lepski, O. Pouet, C. (2008). Hypothesis testing under composite function alternative. In *Topics in Stochastic Analysis and nonparametric estimation*, Springer.

The approach

physical domain

d-variate function f = sum of functional components $f_{\underline{i}}$ (related to the Sobol decomposition)

$\Updownarrow$ Hyperbolic wavelet transform

wavelet coefficient domain

coefficients within orientations $\underline{i}$

A : form test statistics within each orientation to detect $f_{\underline{i}}$ (theoretical properties of these elementary test statistics)

B : combine them in appropriate way $\Rightarrow$ test of more general structure (atomic dimension) | Anisotropy

Functional ANOVA components and orientations in the coefficient domain

$$f = \alpha\phi_{00} + \sum_{\underline{i}=\{\underline{i}\neq 0\}} \sum_{\underline{j}\in\mathbb{J}^{\underline{i}}} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} \theta^{\underline{i}}_{\underline{j},\underline{k}} \psi^{\underline{i}}_{\underline{j},\underline{k}}$$

Example variable selection for u :

$$Y = f(u,\omega) + \text{error}, \tag{4}$$

$$\text{with } f(u,\omega) = f_1(u) + f_2(\omega) + f_{12}(u,\omega) \tag{5}$$

In a wavelet bases, relation between functional components and wavelet coefficients of different orientations :

$$\underline{i} = (1,1); \quad \left\langle f; \psi^1_{j_1,k_1}\psi^1_{j_2,k_2} \right\rangle = 0, \forall j_1, j_2, k_1, k_2 \tag{6}$$

$$\underline{i} = (0,1); \quad \left\langle f; \psi^0 \psi^1_{j_2,k_2} \right\rangle \neq 0, \forall j_2, k_2 \tag{7}$$

$$\underline{i} = (1,0); \quad \left\langle f; \psi^1_{j_1,k_1}\psi^0 \right\rangle = 0, \forall j_1, k_1 \tag{8}$$

Functional ANOVA components and orientations in the coefficient domain

$$f = \alpha\phi_{00} + \sum_{\underline{i}=\{\underline{i}\neq 0\}} \sum_{\underline{j}\in\mathbb{J}^{\underline{i}}} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} \theta^{\underline{i}}_{\underline{j},\underline{k}} \psi^{\underline{i}}_{\underline{j},\underline{k}}$$

Example first order interaction, ie. atomic dimension $= 1$:

$$Y = f(u,\omega) + \text{error}, \quad (9)$$

$$\text{with } f(u,\omega) = f_1(u) + f_2(\omega) + f_{12}(u,\omega) \quad (10)$$

In a wavelet bases, relation between functional components and wavelet coefficients of different orientations :

$$\underline{i} = (1,1); \quad \left\langle f; \psi^1_{j_1,k_1}\psi^1_{j_2,k_2} \right\rangle = 0, \forall j_1, j_2, k_1, k_2 \quad (11)$$

$$\underline{i} = (0,1); \quad \left\langle f; \psi^0 \psi^1_{j_2,k_2} \right\rangle \neq 0, \forall j_2, k_2 \quad (12)$$

$$\underline{i} = (1,0); \quad \left\langle f; \psi^1_{j_1,k_1} \psi^0 \right\rangle \neq 0, \forall j_1, k_1 \quad (13)$$

Test coefficient in orientation such that $\{|\underline{i}| > \delta\}$

Wavelet decomposition of a multivariate function

Consider $f \in \mathbb{L}_2([0,1]^d)$:

$$\begin{aligned} f &= \text{avg approx} + \sum_{\underline{i} \in \{0,1\}^d \setminus \underline{0}} \text{details in orientation } \underline{i} \\ &= \theta^{\underline{0}}_{\underline{0},\underline{0}} \psi^{\underline{0}}_{\underline{0},\underline{0}} + \sum_{\underline{i} \in \{0,1\}^d \setminus \underline{0}} \left(\sum_{\underline{j} \in \mathcal{J}} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} \theta^{\underline{i}}_{\underline{j},\underline{k}} \psi^{\underline{i}}_{\underline{j},\underline{k}} \right), \end{aligned}$$

where

- $\mathcal{J} = \mathbb{J}$, $\mathbb{J} = \left\{ \underline{j} = (j, \cdots, j) : j \in \mathbb{N} \right\}$, for the standard wavelet decomposition,
- $\mathcal{J} = \mathbb{J}^{\underline{i}}$ for the hyperbolic wavelet decomposition.

Summary

1. Wavelet-based method and information pooling
2. Unidimensional second generation thresholding methods : maxiset approach
3. Multidimensional function estimation
4. Structure detection : asymptotic optimality

The Model

Model : We observe n^d independent random variables under the multivariate nonparametric regression problem over an equidistant design grid :

$$Y = f\left(\underline{x}\right) + \sigma\xi,\ \xi \sim \mathcal{N}(0,1)$$

where $\underline{x} \in [0,1]^d$. The model is equivalent with the multivariate Gaussian white noise model.

$$dY_\varepsilon(\underline{x}) = f(\underline{x})d\underline{x} + \varepsilon dW(\underline{x}),$$

where $\underline{x} = (x_1, \ldots, x_d) \in [0,1]^d$, $f \in \mathbb{L}_2([0,1]^d)$, $W(\underline{x})$ is the Brownian sheet and ε is the known noise level with the calibration $\varepsilon = \sigma n^{-\frac{d}{2}}$.

Definition (Sequential white noise model)

$$\hat{\theta}^{\underline{i}}_{\underline{j},\underline{k}} = \theta^{\underline{i}}_{\underline{j},\underline{k}} + \varepsilon\xi^{\underline{i}}_{\underline{j},\underline{k}},$$

where $\xi^{\underline{i}}_{\underline{j},\underline{k}}$ are i.i.d. $\mathcal{N}\left(0,1\right)$ and $(\underline{i},\underline{j},\underline{k}) \in \{0,1\}^d \times \mathbb{N}^d \times \mathbb{Z}^d$.

Testing hypotheses

We are interested in testing whether a component in a certain orientation $\underline{i}$ is zero.

Definition (Hypotheses of the testing problem)

$$\mathcal{H}_{\underline{i},0}: \quad f \in \mathcal{N}_{\underline{i}}(R) = \left\{f : \|f\|_2 \leq R,\ f_{\underline{i}} = 0\right\}$$

$$\mathcal{H}_{\underline{i},a}: \quad f \in \mathcal{A}_{\underline{i}}(R, C, \underline{s}, r_\varepsilon) = \left\{f : f \in \mathcal{B}^{\underline{s}}(R),\ \|f_{\underline{i}}\|_2 \geq C r_\varepsilon\right\},$$

where $C > 0$, $(r_\varepsilon)_\varepsilon$ is a decreasing and continuous sequence of real numbers tending to 0 when ε goes to 0 and $\mathcal{B}^{\underline{s}}(R)$ denotes the ball of radius R of the Besov space with $\underline{s}$ as smoothness parameter.

We consider two cases :

- Non adaptive case : $\underline{s}$ is such that s_u are known for all $i_u = 1$.
- Semi adaptive case : $\gamma_{\underline{i}} = \left(i_1 s_1^{-1} + \cdots + i_d s_d^{-1}\right)^{-1}$ is known. Example : Anisotropic Holder ($s_1 = 0.5, s_2 = 0.8$), $\gamma_{(1,0)} = 0.5, \gamma_{(0,1)} = 0.8, \gamma_{(1,1)} = 0.31$

Testing hypotheses

We are interested in testing whether a component in a certain orientation $\underline{i}$ is zero.

Definition (Hypotheses of the testing problem)

$$\mathcal{H}_{\underline{i},0}: \quad f \in \mathcal{N}_{\underline{i}}(R) = \left\{ f : \|f\|_2 \leq R,\ f_{\underline{i}} = 0 \right\}$$

$$\mathcal{H}_{\underline{i},a}: \quad f \in \mathcal{A}_{\underline{i}}(R, C, \underline{s}, r_\varepsilon) = \left\{ f : f \in \mathcal{B}^{\underline{s}}(R),\ \|f_{\underline{i}}\|_2 \geq C r_\varepsilon \right\},$$

where $C > 0$, $(r_\varepsilon)_\varepsilon$ is a decreasing and continuous sequence of real numbers tending to 0 when ε goes to 0 and $\mathcal{B}^{\underline{s}}(R)$ denotes the ball of radius R of the Besov space with $\underline{s}$ as smoothness parameter.

We consider two cases :

- Non adaptive case : $\underline{s}$ is such that s_u are known for all $i_u = 1$.
- Semi adaptive case : $\gamma_{\underline{i}} = \left(i_1 s_1^{-1} + \cdots + i_d s_d^{-1} \right)^{-1}$ is known. Example : Anisotropic Holder ($s_1 = 0.5, s_2 = 0.8$), $\gamma_{(1,0)} = 0.5, \gamma_{(0,1)} = 0.8, \gamma_{(1,1)} = 0.31$
- Adaptive case : no information available

Testing procedure and related errors

Aim at : providing testing procedures which ensure that the sum Error-I + Error-II does not exceed a chosen $\alpha > 0$ and which are optimal or near optimal in the minimax sense (see Ingster (1986)).

Definition

A testing procedure is a random variable Δ *with value in* $\{0, 1\}$ *for which :*

- $\Delta = 1$ *"*$\mathcal{H}_{\underline{i},0}$ *is rejected"*
- $\Delta = 0$ *else.*

Definition

For the testing hypotheses considered and a fixed Δ *:*

- *Error-I :* $$\sup_{f \in \mathcal{N}_{\underline{i}}(R)} \mathbb{P}_f(\Delta = 1).$$
- *Error-II :* $$\sup_{f \in \mathcal{A}_{\underline{i}}(R, C, \underline{s}, r_\varepsilon)} \mathbb{P}_f(\Delta = 0).$$

Minimax results for hypothesis testing : case NA

Theorem (Autin, Claeskens, Freyermuth and Pouet (2015))

1. *there exists a constant* $C' > 0$ *such that for any* $C < C'$ *the testing problem has* *no solution* *for the rate* $r_\varepsilon = \left(\varepsilon^4\right)^{\frac{4\gamma_{\underline{i}}}{1+4\gamma_{\underline{i}}}}$
2. *there exists a constant* $C'' > 0$ *such that for any* $C > C''$ *the testing problem as a* *solution* *for the rate* $r_\varepsilon = \left(\varepsilon^4\right)^{\frac{4\gamma_{\underline{i}}}{1+4\gamma_{\underline{i}}}}$. *Then, the solution of the testing problem is given through the testing procedure :*

$$\Delta_{\underline{i},t} = \left\{ T_{\underline{i}}^{na} > t_{\alpha}^{na} \right\}$$

where the test statistic is the nonnegative random variable

$$T_{\underline{i}}^{na} \;=\; \sum_{\underline{j} \in \mathbb{J}_{na}^{\underline{i}}} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} \left(\hat{\theta}_{\underline{j},\underline{k}}^{\underline{i}}\right)^2 .$$

Minimax results for hypothesis testing : case NA

Definition (calibration of the test $\Delta_{\underline{i}}^{na}$)

- $\mathbb{J}_{na}^{\underline{i}} = \left\{ \underline{j} \in \mathbb{J}^{\underline{i}} : 2^{j_u s_u} < \left(\varepsilon^4\right)^{-\frac{i_u \gamma_{\underline{i}}}{1+4\gamma_{\underline{i}}}}, \forall u \right\}$,
- t_{α}^{na} *is the quantile of order* $1 - \frac{\alpha}{2}$ *of the Chi-squared distribution* $\# \left\{ (\underline{j}, \underline{k}) : \underline{j} \in \mathbb{J}_{na}^{\underline{i}}, \underline{k} \in \mathbb{K}_{\underline{j}} \right\}$ *with degrees of freedom.*

Remark :

- $\Delta_{\underline{i}}^{na}$ is defined through the knowledge of $\underline{s}$.
- Minimax rates of separation depends on the orientation.

Minimax results for hypothesis testing : case SA

Theorem (A., Claeskens, Freyermuth and Pouet (2015))

1. *there exists a constant* $C' > 0$ *such that for any* $C < C'$ *the testing problem has* *no solution* *for the rate* $r_\varepsilon = \left(\varepsilon^4 \log\log \varepsilon^{-1}\right)^{\frac{4\gamma_{\underline{i}}}{1+4\gamma_{\underline{i}}}}$
2. *there exists a constant* $C'' > 0$ *such that for any* $C > C''$ *the testing problem as a* *solution* *for the rate* $r_\varepsilon = \left(\varepsilon^4 \log\log \varepsilon^{-1}\right)^{\frac{4\gamma_{\underline{i}}}{1+4\gamma_{\underline{i}}}}$. *Then, the solution of the testing problem is given through the testing procedure :*

$$\Delta^a_{\underline{i}} = \left\{ T^a_{\underline{i}} > t^a_\alpha \right\}$$

where the test statistic is the nonnegative random variable

$$T^a_{\underline{i}} \;=\; \sum_{\underline{j} \in \mathbb{J}^{\underline{i}}_a} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} \left(\hat{\theta}^{\underline{i}}_{\underline{j},\underline{k}}\right)^2 .$$

Minimax results for hypothesis testing : case SA

Definition (calibration of the test $\Delta^a_{\underline{i}}$)

- $\mathbb{J}^{\underline{i}}_a = \left\{ \underline{j} \in \mathbb{J}^{\underline{i}} : 2^{j_1+\cdots+j_d} < \left(\varepsilon^4 \log\log \varepsilon^{-1}\right)^{-\frac{1}{1+4\gamma_{\underline{i}}}} \right\}$,
- t^a_α *is the quantile of order* $1-\frac{\alpha}{2}$ *of the Chi-squared distribution* $\#\left\{(\underline{j},\underline{k}) : \underline{j} \in \mathbb{J}^{\underline{i}}_a, \underline{k} \in \mathbb{K}_{\underline{j}}\right\}$ *with degrees of freedom.*

remark

- $\Delta^a_{\underline{i}}$ is only defined through the knowledge of $\gamma_{\underline{i}}$.
- There is a loss of rate through the $\log\log$ term.
- At this near optimal rate, $\Delta^a_{\underline{i}}$ is able to detect more functions than the than the previous test for the chosen precision α.

About the questions on structure (selection of variables or atomic dimension) :

1. "Is x_m a true variable for f ?"
   The answer depends on the observation of the testing procedure :

$$\Delta_{sel}(m) = \max\left(\Delta^*_{\underline{i}} : \underline{i} \text{ s.t. } i_m = 1\right).$$

2. "Is the atomic dimension of f smaller than δ ?"
   The answer depends on the observation of the testing procedure :

$$\Delta_{atom}(\delta) = \max\left(\Delta^*_{\underline{i}} : |\underline{i}| \geq \delta\right).$$

Notation : $* \in \{na, a\}$.

Main references

[P1] Abry, P. Clausel, M. Jaffard, S. and Vedel, B. (2014). Hyperbolic wavelet transform : a new tool for multifractal analysis and anisotropic textures. To appear in Revista Matematica Iberoamericana.

[P2] Autin, F., Freyermuth, J.-M. and von Sachs R. (2012). Combining thresholding rules: a new way to improve the performance of wavelet estimators. J. Nonparam. Statist., volume 24, 905-922.

[P3] Autin, F., Freyermuth, J.-M. and von Sachs R. (2014). Block-Threshold-Adapted Estimators via a maxiset approach. Scand. J. Statist., volume 41, 240-258.

[P4] Autin, F., Claeskens, G. and Freyermuth, J.-M. (2014). Hyperbolic wavelet thresholding methods and the curse of dimensionality through the maxiset approach. Appl. Comput. Harmon. Anal., volume 36, 239-255.

[P5] Autin, F., Claeskens, G. and Freyermuth, J.-M. (2015). On the performance of isotropic and hyperbolic wavelet estimators. Avec Claeskens G. et Freyermuth, J.-M.. Submitted.

[P6] Autin, F., Claeskens, G., Freyermuth, J.-M. and Pouet, C. (2015). Minimax optimal testing for structure of multivariate data. Avec Claeskens G. et Freyermuth, J.-M.. Submitted.

[P7] Benhaddou, R., Pensky, M. and Picard, D. (2013). Anisotropic denoising in functional deconvolution model with dimension-free convergence rates. E. J. Statist., volume 7, 1686-1715.

[P8] Comminges, L. and Dalalyan, A. (2013). Minimax testing of a composite null hypothesis defined via a quadratic functional in the model of regression. IEEE Transactions on Signal Processing.

[P9] Duarte, M. and Baraniuk, R. (2012). Kronecker compressive sensing. IEEE Transaction on Image Processing, volume 21, 494-504.

[P10] Kerkyacharian, G., Picard, D. and Temlyakov, V. N. (2006). Some inequalities for the tensor product of greedy bases and weight-greedy bases. East J. Approx., volume 12, 103-118.

[P11] Ingster, Y. (1986). Asymptotically minimax testing of nonparametric hypo- theses. Probab. Theory and Math. Statist., volume 1, 553-574.

[P12] Neumann, N. H. (2000). Multivariate wavelet thresholding in anisotropic function spaces. Statist. Sinica, volume 10, 399-431.

[P13] Sperlich, S., Tjostheim, D. and Yang, L. (2002). Nonparametric estimation and testing of interaction in additive models. Econometric Theory, volume 18, 197-251.