Structured anisotropic sparsity priors for non-parametric function estimation

Younes Farouj

N° d'ordre NNT : 2016LYSEI123

THÈSE de DOCTORAT DE L'UNIVERSITÉ DE LYON

Opérée au sein de

L'Institut National des Sciences Appliquées de Lyon

ÉCOLE DOCTORALE N° 160

ÉLECTRONIQUE, ÉLECTROTECHNIQUE, AUTOMATIQUE

Spécialité : Traitement du Signal et de l'Image

Soutenue publiquement le xx novembre 2016 par

Younes Farouj

Structured Anisotropic Sparsity Priors for Non-parametric Function Estimation

---

Devant le jury composé de :

Philippe Carré	Professeur, Université de Poitiers	Rapporteur
Pierre Chainais	Maître de Conférences HDR, École centrale de Lille	Rapporteur
Frédéric Ferraty	Professeur, Univerité de Toulouse	Examinateur
Valérie Perrier	Professeur, Grenoble INP - ENSIMAG	Examinatrice
Dimitri Van de Ville	Professeur, École Polytechnique Fédérale de Lausanne	Examinateur
Maarten Jansen	Professeur associé, Univérsité Libre de Bruxelles	Membre Invité
Marianne Clausel	Maître de Conférences HDR, LJK - Université de Grenoble	Co-encadrante
Philippe Delachartre	Professeur, INSA de Lyon	Directeur de thèse
Laurent Navarro	Maître de Conférences, École des mines de Saint-Étienne	Co-encadrant

Département FEDORA – INSA Lyon - Ecoles Doctorales – Quinquennal 2016-2020

SIGLE	ECOLE DOCTORALE	NOM ET COORDONNEES DU RESPONSABLE
CHIMIE	**CHIMIE DE LYON** **http://www.edchimie-lyon.fr** Sec : Renée EL MELHEM Bat Blaise Pascal 3e etage secretariat@edchimie-lyon.fr Insa : R. GOURDON	M. Stéphane DANIELE Institut de Recherches sur la Catalyse et l'Environnement de Lyon IRCELYON-UMR 5256 Équipe CDFA 2 avenue Albert Einstein 69626 Villeurbanne cedex directeur@edchimie-lyon.fr
E.E.A.	**ELECTRONIQUE, ELECTROTECHNIQUE, AUTOMATIQUE** **http://edeea.ec-lyon.fr** Sec : M.C. HAVGOUDOUKIAN Ecole-Doctorale.eea@ec-lyon.fr	M. Gérard SCORLETTI Ecole Centrale de Lyon 36 avenue Guy de Collongue 69134 ECULLY Tél : 04.72.18 60.97 Fax : 04 78 43 37 17 Gerard.scorletti@ec-lyon.fr
E2M2	**EVOLUTION, ECOSYSTEME, MICROBIOLOGIE, MODELISATION** **http://e2m2.universite-lyon.fr** Sec : Safia AIT CHALAL Bat Darwin - UCB Lyon 1 04.72.43.28.91 Insa : H. CHARLES Safia.ait-chalal@univ-lyon1.fr	Mme Gudrun BORNETTE CNRS UMR 5023 LEHNA Université Claude Bernard Lyon 1 Bât Forel 43 bd du 11 novembre 1918 69622 VILLEURBANNE Cédex Tél : 06.07.53.89.13 e2m2@ univ-lyon1.fr
EDISS	**INTERDISCIPLINAIRE SCIENCES-SANTE** **http://www.ediss-lyon.fr** Sec : Safia AIT CHALAL Hôpital Louis Pradel - Bron 04 72 68 49 09 Insa : M. LAGARDE Safia.ait-chalal@univ-lyon1.fr	Mme Emmanuelle CANET-SOULAS INSERM U1060, CarMeN lab, Univ. Lyon 1 Bâtiment IMBL 11 avenue Jean Capelle INSA de Lyon 696621 Villeurbanne Tél : 04.72.68.49.09 Fax :04 72 68 49 16 Emmanuelle.canet@univ-lyon1.fr
INFOMATHS	**INFORMATIQUE ET MATHEMATIQUES** **http://infomaths.univ-lyon1.fr** Sec :Renée EL MELHEM Bat Blaise Pascal 3e etage infomaths@univ-lyon1.fr	Mme Sylvie CALABRETTO LIRIS – INSA de Lyon Bat Blaise Pascal 7 avenue Jean Capelle 69622 VILLEURBANNE Cedex Tél : 04.72. 43. 80. 46 Fax 04 72 43 16 87 Sylvie.calabretto@insa-lyon.fr
Matériaux	**MATERIAUX DE LYON** **http://ed34.universite-lyon.fr** Sec : M. LABOUNE PM : 71.70 –Fax : 87.12 Bat. Saint Exupéry Ed.materiaux@insa-lyon.fr	M. Jean-Yves BUFFIERE INSA de Lyon MATEIS Bâtiment Saint Exupéry 7 avenue Jean Capelle 69621 VILLEURBANNE Cedex Tél : 04.72.43 71.70 Fax 04 72 43 85 28 Ed.materiaux@insa-lyon.fr
MEGA	**MECANIQUE, ENERGETIQUE, GENIE CIVIL, ACOUSTIQUE** **http://mega.universite-lyon.fr** Sec : M. LABOUNE PM : 71.70 –Fax : 87.12 Bat. Saint Exupéry mega@insa-lyon.fr	M. Philippe BOISSE INSA de Lyon Laboratoire LAMCOS Bâtiment Jacquard 25 bis avenue Jean Capelle 69621 VILLEURBANNE Cedex Tél : 04.72 .43.71.70 Fax : 04 72 43 72 37 Philippe.boisse@insa-lyon.fr
ScSo	**ScSo*** **http://recherche.univ-lyon2.fr/scso/** Sec : Viviane POLSINELLI Brigitte DUBOIS Insa : J.Y. TOUSSAINT viviane.polsinelli@univ-lyon2.fr	Mme Isabelle VON BUELTZINGLOEWEN Université Lyon 2 86 rue Pasteur 69365 LYON Cedex 07 Tél : 04.78.77.23.86 Fax : 04.37.28.04.48

*ScSo : Histoire, Géographie, Aménagement, Urbanisme, Archéologie, Science politique, Sociologie, Anthropologie

Structured Anisotropic Sparsity Priors for Non-parametric Function Estimation

Younes Farouj

Contents

CHAPTER 1

Introduction

The problem of estimating a multivariate function from corrupted observations arises throughout many areas of engineering. For instance, in the particular field of medical signal and image processing, this task has attracted special attention and even triggered new concepts and notions that have found applications in many other fields. This interest is mainly due to the fact that the medical data analysis pipeline is often carried out in challenging conditions, since one has to deal with noise, low contrast and undesirable transformations operated by acquisition systems. On the other hand, the concept of sparsity had a tremendous impact on data reconstruction and restoration in the last two decades. Sparsity stipulates that some signals and images have representations involving only a few non-zero coefficients. This turned out to be verifiable in many practical problems. Imposing sparsity priors on a corrupted input consists in setting to zero small coefficients which results in removing noise components. Two of the most commonly used notions of sparsity are sparse decompositions on some families of orthogonal bases called *wavelets* and sparse discrete gradients obtained through *Total Variation*.

Wavelets

Wavelets have been studied and used for nearly half a century. They have been launched by geo-physicists in the 80′s for analyzing seismic signals. Surprisingly, this analyzing tool, became a major drive in many theoretical fields as well as in solving some practical problems. Some of the first real life applications of wavelets were audio signal analysis and image compression and denoising. Statisticians, attracted by the ability of wavelets in characterizing some classes of functional spaces, started an extensive investigation in the early 90's aiming at providing a theoretical understanding of the performance of wavelet denoising. Particularly, the works of Donoho and Johnstone have set up the ground for wavelet non parametric estimation and the construction of related thresholding algorithms.

Total Variation

A second notion which promotes sparsity is Total Variation (TV). In opposition to wavelets which were motivated by practical issues, TV came from theory to practice when researchers noticed that performing denoising tasks using Tykhonov and L^2 regularization does not preserve strong discontinuities. Rudin, Osher and Fatemi were the first to use TV as a regularization term in a variational framework for image restoration. Since, TV became a classical tool in image processing and computer vision with applications ranging from clustering to segmentation and motion estimation. The non-differentiability of TV also inspired many works on non-smooth convex optimization.

Sparsity in several dimensions

Two concepts of dimensionality are present in sparsity. The first one is the dimension of the data; a function of more than one variable is called a "multivariate function". The second concept appears in the transformation (sparse) domain; coefficients play the role of variables. In both cases dealing with several dimensions has always been regarded with pessimism. This is due to the "curse of dimensionality" which constraints

theoretical performances to the dimension. More precisely the minimax theory quantifies this curse by showing that the dimension appears as a negative exponent in convergence rates. Nonetheless, for estimation purposes with many variables, sparsity can be regarded from different angles. First, the unknown function might not depend on all variables but only on a few of them. Thus, sparsity can be related to **"dimensionality reduction"**. The performances are then linked to the effective dimension which is lower than the dimension of the study space (to which the unknown belongs). Such reductions appear in statistics and machine learning problems such as variable or feature selection problems. Generally, multivariate functions (data) have variables with physical meanings (space, time, spectral, $\cdots$). Such variables are not expected to be inactive and thus dimension reduction cannot be applied. They can, however, have different roles and behaviours. This can be seen as an **"anisotropy"** feature. Each variable can be then treated separately with optimal tools.

Structured Sparsity

Sparsity aims at controlling the cardinality of coefficients in a given representation. This results in elitist denoising procedures such as hard or soft thresholding that act individually on each coefficient regardless of any prior on the structure of non-zero coefficients. However, in many situations, we can expect predefined patterns in sparsity. When analyzing real signals and images, local structures in the function domain results in blocks of wavelet coefficients having the same behaviour; small coefficients are often structured in well-defined neighbourhoods at a fixed scale. Wavelet coefficients also show a natural hierarchical (multi-resolution) structure across scales. The notion of structured sparsity emerged in the recent years to promote meaningful sparse representations by grouping coefficients that have, presumably, the same behaviour. This is often done through structured ℓ_1-norms imposing either disjoint or overlapping groups of variables.

Positioning of the dissertation

The structured sparsity literature focuses on grouping variables in the transformation domain. This is partly a result of the many challenges arising in approximation and learning of very high-dimensional data. Only few works considered imposing structures on variables or dimensions before applying the sparsifying transformation. This is surprising because such structures can be seen simply by visualizing the data. As mentioned before, a natural and easy structure can be revealed by considering anisotropy along the variables. This PhD thesis aims at using this idea to develop sparse representations and regularity priors for multivariate function estimation. In particular, a special attention is given to groupwise anisotropy. This notion enables grouping variables into sub-sets having the same behaviour. We also consider anisotropy as a general concept which goes beyond regularity. Grouped variables can also share the same physical properties or can be observed under the same model.

The present dissertation deals with non-parametric multivariate function estimation; we consider the problem of estimating a multivariate function f from a corrupted observation g

$$g = f + \varepsilon,$$

where ε is a noise component. Each of the three parts of the manuscript handles this problem with a particular prior on f often motivated by challenges arising in medical imaging:

1. In Part I, we consider that f is a function of $d \geq 3$ variables and that these variables can be grouped in sub-ensembles on which f shows a similar behaviour. This behaviour can be related to the regularity of f or/and the physical meaning of the variables.

   Some natural groupings appearing in real applications are

   (i) Space-time: Dynamic data, such as image sequences, show different regularities along spatial and temporal variables. Note also that in many acquisition systems the sequences are acquired frame by frame. As a consequence, many noise models depend only on spatial dimensions.

(ii) Velocity-time: In the case of incompressible flows, the difference between dimensions is also due to the physical properties of the flow; the divergence on spatial variables is null.

(iii) Time-frequency / space-frequency: Time-frequency representations, spectral and hyperspectral data show also different regularities along variables.

2. Part II deals with a case that occurs in ultrasound imaging in which the function f depends on two variables and the variance of the noise component is a function of the unknown f. Ultrasound images represent, often, anisotropic features such as vessels and skin layers.

3. Part III covers another situation that arises in functional magnetic resonance imaging. Similarly to some cases examined in Part I, the function f depends on spatial and temporal variables. Here, an additional difference between the dimensions comes from fact that the function f is observed under the action of a blur operator in the time domain due to the neural system response.

For each of the three problems we suggest a solution that is based on two main ingredients: sparsity and anisotropy. In Part I, we introduce a generalization of the hyperbolic wavelet construction on groups of variables. This allowed us to benefit from both isotropic and anisotropic constructions of wavelets for non-parametric estimation tasks, but also to take into account other characteristics such as divergence properties. An hyperbolic two-dimensional generalization of the wavelet-Fisz methodology is presented in Part II. Theoretically, this construction guarantees optimal estimators in the minimax sens. We demonstrate that for the particular case of wavelet-Fisz thresholding, the two-dimensional hyperbolic wavelet construction outperforms the isotropic wavelet construction also in practice as predicted by theoretical results. This is not always the case for classical wavelet procedures. In Part III, we use a particular TV regularization composed of two terms (spatial and temporal). These two terms can be seen as single structured TV term that offers an isotropic treatment in spatial dimensions and an anisotropic prior in space/time. Moreover, this constructions allowed us to take into account the blur operator in the temporal domain using a particular generalization of TV. Let us, now, give a detailed outline of the different chapters of this manuscript.

Outline

In Chapter 2, we present an overview on wavelets, their different constructions and multivariate extensions, along with some mathematical tools which are going to be useful in the sequel. In particular, we recall related functional spaces, statistical models and elements of non-parametric estimation.

Chapter 3 describes total variation and its use in signal and image processing. We also discuss some basic elements on convex optimization related to total variation minimization.

Chapter 4 offers a new construction of wavelet atoms. First, we show the limits of the standard tensor-product (hyperbolic) construction in image denoising. Then, we motivate the use of such constructions in cases when the regularities along variables are different. This led us to a generalized hyperbolic construction which allows grouping variables with respect to regularity features while performing non-parametric estimation. For thresholding estimators, we show, under usual assumptions on wavelet functions, that the L^2 -loss falls to the dimension of the group of variables with the highest dimension. Moreover, we show that other features besides regularity can be taken thanks to this construction. In particular, divergence-freedom and variance stabilization can be imposed on groups of variables.

Chapter 5 uses the construction presented in Chapter 4 to show how denoising tasks can be performed with respect to the anisotropy of the unknown. We present examples of denoising on spatio-temporal data (image sequences and incompressible flows) and also spectral and hyperspectral data.

In Chapter 6, we give a discussion about the positioning of our work compared to existing works on structured group-sparsity. We also mention some orientations for future works on the wavelet construction presented in this part of the thesis.

In Chapter 7, we consider the following two-dimensional function estimation problem: we want to recover an unknown function α from a noisy observation X, where the noise component has zero mean and a variance function depending on the unknown α. We prove the optimality of hyperbolic wavelet-Fisz hard-thresholding when u belongs to anisotropic Besov balls. This method computes the hyperbolic wavelet transform of the image, before applying a multiscale variance stabilization technique, via a Fisz transformation. This adapts the wavelet coefficients statistics to the wavelet thresholding paradigm. We also describe a data-driven extension of this technique when h is unknown following previous works by Fryzlewicz and Dellouile. The data-driven extension removes the need for any prior knowledge of the noise model parameters by estimating the noise variance using an isotonic Nadaraya-Watson estimator.

In Chapter 8, we use the techniques presented in Chapter 7 to develop an algorithm and its fully data-driven extension for noise reduction in ultrasound imaging. The use of hyperbolic wavelets enables to recover the image while respecting the anisotropic nature of structural details. Experiments on synthetic and real data demonstrate the potential of the proposed algorithm at recovering ultrasound images while preserving tissue details. Furthermore, for this particular wavelet denoising strategy, we show that the results obtained by the hyperbolic construction confirm the theoretical claims. Comparisons with other noise reduction methods show that our method is competitive with the state-of-the art OBNLM filter. Finally, we emphasize the noise model we consider by applying our variance estimation procedure on real images.

Chapter 9 is devoted to some orientations for future works on the wavelet-Fisz and its applications.

Chapter 10 put in review some of the fundamental works on fMRI data analysis before stating the fMRI deconvolution problem that we will consider in the next chapter.

Chapter 11 introduces new strategies for fMRI spatio-temporal deconvolution without anatomical priors. That is, only the membership of the contributing voxels to the gray matter and the global homogeneity of the activation are taken into account. The presented atlas free Total Activation (AfTA) technique is an extension of the Total Activation (TA) framework; we formulate a variational denoising problem involving two regularization terms. These terms express sparsity along variables by taking the temporal and the spatial characteristics. The first term uses a generalized total variation which promotes a block-type structure on the the underlying activity-inducing signal by inverting the hemodynamic response function. The second term is a weighted total variation which favors globally coherent activation patterns within the gray matter while preserving strong discontinuities. Evaluation on synthetic demonstrated the potential of AfTA at recovering the brain activity. Furthermore, we applied this techniques to a real task-evoked fMRI data from an experiment with prolonged resting state periods disturbed by visual stimuli. The results show the ability of proposed technique at retrieving both spontaneous and task-related activities without prior knowledge of the timing of the experimental paradigm nor the triggered regions.

Finally, **??** is devoted to a discussion on a possible extension of the deconvolution method presented in the previous chapter. We consider the problem of estimating both the fMRI signal and the systems response.

CHAPTER 2

Wavelets

Abstract

In this chapter, we present an overview on wavelets, their different constructions and multivariate extensions, along with some mathematical tools which are going to be useful in the sequel. In particular, we recall related functional spaces, statistical models and elements of non-parametric estimation.

2.1 Introduction

The origin of wavelets goes back to the beginning of the 80's when Morlet [1981] had the idea of analyzing seismic signals by integrating them against translated and dilated versions of a mother function called "wavelet". The purpose is to retrieve a localized description of both time and frequency information. Surprisingly, this analyzing tool, which is purely motivated by its application, turned out to be a connection point of various research fields which appear, at first sight, to be disconnected. Thus, wavelets provided a remedy for two distinct pathologies of the Fourier representation:

- Signal processing and quantum physics researchers understood early a major limitation of Fourier analysis: though it gives a complete description in the frequency domain, the temporal information is completely lost.
- While engineers and physicists were busy with solving the localization problem above, mathematicians were still trying to find appropriate bases for some functional spaces arising in applied analysis as Besov and Sobolev spaces.

These two questions have been studied, before wavelets, with two ancestors which are respectively the frames of Gabor [1946] providing a time-frequency representation but not an orthogonal basis of L^2 and the basis constructed by Haar [1910] from shifted and dilated versions of a step function which gives an unconditional orthogonal basis for all L^p spaces with $p \geq 2$ but cannot be a basis for functional spaces of regular (continuous) functions.

The wavelet theory came with a sort of synthesis of these ideas by providing unconditional orthonormal bases for large classes of functional spaces and making the connection between Gabor's time-frequency analysis and Haar's time-scale representation. This connection turned out to be crucial as the time-scale representation allowed sub-band coding practitioners to use pyramidal algorithms for fast implementation. Wavelets are an exciting example of a tool which was discovered for practical purposes and turned out to be of great use in theory. Nowadays, wavelets are standard tools in harmonic analysis, statistic, physics and engineering. In the next section, we briefly present the general framework of time-frequency analysis to motivate the section 2.3 about the multi-resolution theory leading to the construction of wavelet filter. The description of functional spaces through wavelet expansions are given in section 2.4 and then one of the most celebrated applications, which is statistical non-parametric estimation, is given in the last section.

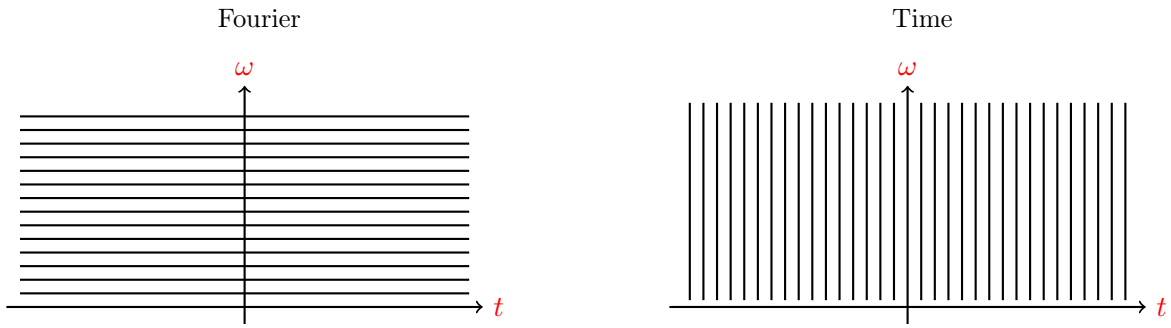


Figure 2.1: Pavings of the time domain and frequency domain representations.

2.2 Time-frequency Analysis

In this section, we expose fundamental elements of time-frequency analysis. In particular, we recall the basic representations and expose their disadvantages which motivated the definition of wavelets.

2.2.1 Fourier Analysis

Fourier analysis of a function $f \in L^2(\mathbb{R})$ consists in representing the function as a sum of elementary signals corresponding to sines and cosines (pure harmonics) and thus complex exponentials

$$f(t) = \frac{1}{2\pi} \int_{\mathbb{R}} \widehat{f}(\omega) e^{i\omega t} d\omega$$

where the Fourier transform is given by

$$\widehat{f}(\omega) = \int_{\mathbb{R}} f(t) e^{-i\omega t} dt$$

The integration against f is done on the complete domain. This implies that a local change in the temporal domain affects all points in the Fourier domain. Thus, the representation produced by the family $\{e^{-i\omega t}\}$ is local in frequency but global in time (cf. Figure 2.1). From a functional analysis point of view, the most notable property of the Fourier representation is the energy preservation given by the Plancherel identity $||f||_2 = ||\widehat{f}||_2$. This allows one to decide from the amplitude of the Fourier coefficients, the membership of f in L^2 spaces. That is, the exponentials family is an unconditional basis for L^2 spaces. However, for other L^p spaces, there is no equivalent of the Plancherel identity. Thus the membership to those spaces cannot be characterized only by knowledge on the amplitude of the Fourier representation. These drawbacks can be related to the fact that the Fourier transform is global: the natural way to address them is to introduce some localization through windowed transforms.

2.2.2 Gabor Analysis

The most straightforward way to introduce some localization is to split the time line into segments $I_j = [a_j, a_{j+1}]$, where a_j is an increasing sequence of real numbers. Then, replace the exponantials in the Fourier transform by the family $\mathbb{1}_{I_j} e^{i\omega t}$ where $j \in \mathbb{Z}$ and $\mathbb{1}_{I_j}$ is the indicator function over the segment I_j. The resulting transform is known as the windowed Fourier transform. It is now possible to have a basis of $L^2(\mathbb{R})$ by periodicity. However, the discontinuity induced by the indicator windows generates functions with Fourier coefficients having a slow decay. Gabor [1946] suggested to replace the indicator function by a family of translated and frequency-modulated smooth functions. The resulting family of functions is of the form

$$g_{\omega_0,t_0} = g(t - t_0) e^{i\omega_0 t}, \quad (\omega_0, t_0) \in \mathbb{R}^2$$

where g is a normalized ($||g||_2^2 = 1$) function called the Gabor window which define the time-frequency support. The width of such a support is however limited by the Heisenberg inequality. This limit can be defined by measuring the width of the window g via the following quadratic deviations

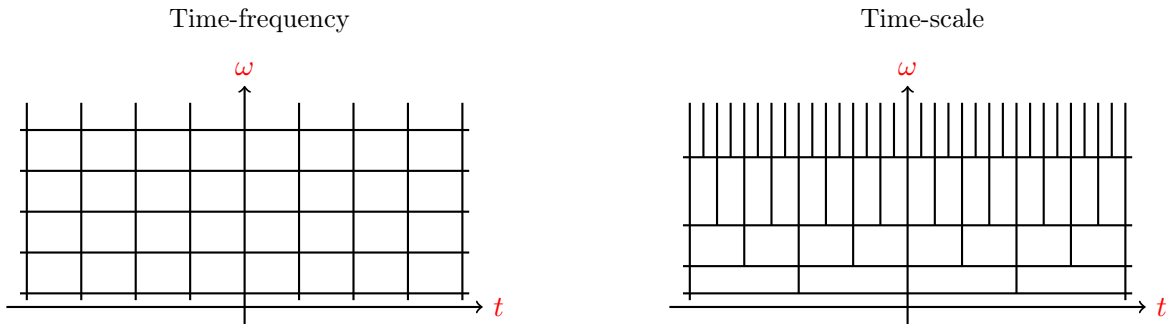


Figure 2.2: Pavings of the time-frequency and time-scale representations.

$$\Delta_x(g) = \Big(\int_{\mathbb{R}} (t - \bar{t}(g))^2 |g(t)|^2 dt\Big)^{\frac{1}{2}},$$

and

$$\Delta_\omega(g) = \Big(\int_{\mathbb{R}} (\omega - \bar{\omega}(\hat{g}))^2 |\hat{g}(t)|^2 dt\Big)^{\frac{1}{2}},$$

where the means $\bar{t}(g)$ and $\bar{\omega}(\hat{g})$ are given by

$$\bar{t}(g) = \int_{\mathbb{R}} t|g(t)|^2 dt \quad ; \quad \bar{\omega}(\hat{g}) = \int_{\mathbb{R}} \omega |\hat{g}(\omega)|^2.$$

The Heisenberg inequality reads

$$\Delta_x(g)\Delta_\omega(g) \geq \frac{1}{2}$$

As a result, the elements of the family $\{g_{\omega_0,t_0}\}$ cannot have an arbitrary small time-frequency support. The minimum $\frac{1}{2}$ is reached when g has a Gaussian form. To represent an arbitrary function as a linear combination of such Gaussian functions, one needs all possible combinations of the parameters (ω_0 and t_0) which is an overly redundant paving of the time-frequency domain. The price to pay in this case, compared to the windowed Fourier transform, is to renounce on having a basis of $L^2(\mathbb{R})$. In fact, to have an orthogonal basis with Gabor systems, one should expect to have a very poor localization either in time or frequency (cf. Bourgain [1988]). This result is known as the strong uncertainty principle due to Balian [1981] which highlights the incompatibility of non-redundancy and time-frequency concentration. It states than any basis of $L^2(\mathbb{R})$ of the Gabor system form is expected to have a window which have an infinite support either in space domain or phase domain:

$$\Delta_x(g)\Delta_\omega(g) = \infty.$$

In order to overcome the strong uncertainty principle, and still have a non-redundant basis, one should go beyond the Gaussian window of the Gabor framework, which is too well-localized, and thus losing some time-frequency concentration. Here again, there are two approaches:

- The time-frequency approach: The paving is performed independently from time width. The function are given as in the Gabor system $\{g_{\omega_0,t_0}\}_{\omega_0,t_0}$.
- The time-scale approach: The temporal support is varying and is inverse proportional to the frequency within the limitations of the uncertainty principle.

The most celebrated examples of the local time-scale approach are wavelets.

2.2.3 Continuous Wavelet Transform (CWT)

Following Morlet's ideas, Grossmann and Morlet [1984] showed that a family constructed from a well chosen function ψ by translations and dilations

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

can be used for the expansion of function in $L^2(\mathbb{R})$. It turned out that such families can form, under some conditions on ψ, orthonormal bases of $L^2(\mathbb{R})$ (cf. Strömberg [1983] and Meyer [1985]). For a function $\psi \in L^2(\mathbb{R})$ with $||\psi||_2^2 = 1$ and null average, the continuous wavelet transform Wf of the function f is given by the following formula:

$$Wf(s,\tau) = \langle f, \psi_{s,\tau}\rangle = \int_{\mathbb{R}} f(t)\overline{\psi_{s,\tau}(t)}\,dt$$

Under the *Calderón's admissibility condition*:

$$C_\psi = \int_0^\infty \frac{|\widehat{\psi}(\omega)|^2}{\omega}\,d\omega < +\infty,$$

one has the so–called reconstruction formula

$$f(t) = \frac{1}{C_\psi}\int_{\mathbb{R}^2} Wf(s,\tau)\psi_{s,\tau}(t)\,d\tau\frac{ds}{s^2}.$$

Note that the wavelet transform is quasi-isometric as

$$||f(t)||_2^2 = \frac{1}{C_\psi}||Wf||_2^2.$$

To have a non-redundant discrete time-scale representation as the one in the previous paragraph, the definition of the discrete wavelet transform is needed.

2.2.4 Discrete Wavelet Transform (DWT)

The discrete wavelet transform (DWT) used by Morlet consists in a discretization of the time and scale parameters in a way that one is proportional to the other $s = k\tau$. The most used choice for the discret grid is the dyadic one. Here again, the starting point is ψ and the wavelet family is constructed from translated and dilated on the grid

$$\{\psi(2^j t - k)\}_{j,k}, \text{ with } (j,k) \in \mathbb{Z}^2.$$

The discrete counterpart of the *Calderón's admissibility condition* is given for each ω belonging to a discretization of the real line, by

$$\sum_{j\in\mathbb{Z}} |\hat{\psi}(2^{-j}\omega)|^2 = 1$$

2.3 Multiresolution Analysis (MRA)

In this section we describe how wavelets are constructed. A classical way to address this description is to do it through the *Multiresolution Analysis* (MRA) framework introduced by Mallat [1989].

2.3.1 Definition of MRA

A *multiresolution analysis* of $L^2(\mathbb{R})$ is an increasing sequence of closed subspaces $(V_j)_{j\in\mathbb{Z}}$,

$$\{0\} \subset \cdots V_{j-1} \subset V_j \subset V_{j+1} \cdots \subset L^2(\mathbb{R})$$

verifying the following properties:

(i) $\bigcap_{j\in\mathbb{Z}} V_j = \{0\}$

(ii) $L^2(\mathbb{R}) = \overline{\bigcup_{j\in\mathbb{Z}} V_j}$.

(iii) $\forall f \in L^2(\mathbb{R}),\ \forall j \in \mathbb{Z} :\ f(x) \in V_j \iff f(2x) \in V_{j+1}$

(iv) $\forall f \in L^2(\mathbb{R}),\ \forall k \in \mathbb{Z} :\ f(x) \in V_0 \iff f(x-k) \in V_0$

(v) $\exists \varphi \in V_0$ such that $\{\varphi(.-k)\}_{k\in\mathbb{Z}}$ forms a frame of the subset V_0.

The assumptions (i) and (ii) ensures completeness through the density of the union of the sub-spaces $(V_j)_{j\in\mathbb{Z}}$ in $L^2(\mathbb{R})$. Assumption (iii) means that the sub-spaces $(V_j)_{j\in\mathbb{Z}}$ are time-scaled versions of each other. Assumption (iv) ensures that V_0 is translation invariant. Assumption (v) requires that V_0 is a *linear span* of shifted versions on $\mathbb{Z}$ of a generating functions φ called the *scaling function*.

Note that assumptions (iii) and (iv) guarantee that

$$\forall f \in L^2(\mathbb{R}), \forall k \in \mathbb{Z} : f(x) \in V_j \iff f(2^j x - k) \in V_{j+1}$$

It also follows, from the assumption (v), that all for all $j \in \mathbb{Z}$, the subspace V_j can be described as a linear span using the family $\{\varphi(2^j. - k)\}_{k\in\mathbb{Z}}$. Finally, the fact that any function of V_0 has to be expressed as an infinite linear combination of shifted versions of φ imposes some decay rate at the infinity on φ.

2.3.2 Wavelet bases

If we suppose that we are given a function φ such that the family $\{\varphi(.-k)\}_{k\in\mathbb{Z}}$ forms an orthonormal basis of V_0, then at each resolution scale j, the subspace V_j has an orthonormal basis $\{\varphi_{j,k}(.) = 2^{j/2}\varphi(2^j.-k)\}_{k\in\mathbb{Z}}$. As a consequence, any function $f \in L^2(\mathbb{R})$ can be approximated at the scale j by an *orthogonal projection* on V_j as follows

$$P_j f(t) = \sum_{k\in\mathbb{Z}} \langle f, \varphi_{j,k}\rangle \varphi_{j,k}.$$

$P_j f$ is an approximation of f at scale 2^{-j}. To get a better approximation at a finest scale $2^{-(j+1)}$, we should be able to evaluate the difference

$$Q_j f = P_{j+1} f - P_j f,$$

or equivalently find $Q_j f$ such that $P_{j+1} f = P_j f + Q_j f$. If we denote by W_j the orthogonal complement of V_j

$$V_j \oplus W_j = V_{j+1}, \tag{2.1}$$

then Q_j is the projection of f on W_j. Wavelets appears as orthonormal bases of the spaces $\{W_j\}_{j\in\mathbb{Z}}$. In the same way as the *scaling function* is used to construct bases for $\{V_j\}_{j\in\mathbb{Z}}$, it is possible to construct bases of W_j from a function ψ called *mother wavelet*. The function ψ is chosen such that $\{\psi(.-k)\}_{k\in\mathbb{Z}}$ forms an orthonormal basis of W_0. The basis $\{\psi_{j,k}\}_{k\in\mathbb{Z}}$ of W_j is then constructed as translated and dilated versions of ψ

$$\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k), \quad \forall j \in \mathbb{Z}, \quad \forall k \in \mathbb{Z},$$

where the factor $2^{j/2}$ is the L^2 normalization. As a result, we have

$$Q_j f = \sum_{k\in\mathbb{Z}} \langle f, \psi_{j,k}\rangle \psi_{j,k}.$$

It implies that any function of $L^2(\mathbb{R})$ can be written as a combination of $\{\psi_{j,k}\}_{(j,k)\in\mathbb{Z}^2}$ in two ways

$$f = \sum_{j,k\in\mathbb{Z}} \langle f, \psi_{j,k}\rangle \psi_{j,k}, \tag{2.2}$$

or

$$f = \sum_{k\in\mathbb{Z}} \langle f, \varphi_{j,k}\rangle \varphi_{j,k} + \sum_{l>j}\sum_{k\in\mathbb{Z}} \langle f, \psi_{l,k}\rangle \psi_{l,k}, \tag{2.3}$$

Each of these two constructions come from

$$V_j \oplus W_j \cdots \oplus W_{l-1} = V_l, \quad j < l. \tag{2.4}$$

with l goes to $+\infty$ which can be easily deduced from 2.1.

The expansion in (2.2) is obtained by letting j goes to $-\infty$, while the expansion in (2.3) is obtained by fixing j. These two expansions correspond respectively to the two following decompositions

$$L^2(\mathbb{R}) = \overline{\bigoplus_{j\in\mathbb{Z}} W_j}$$

and

$$L^2(\mathbb{R}) = V_j \oplus \overline{\bigoplus_{l\geq j} W_l}.$$

Similarly, the union of the bases of the subspaces $\{W_j\}_{j\in\mathbb{Z}}$ results in two types of orthonormal wavelet bases of $L^2(\mathbb{R})$

$$\mathcal{B} = \{\psi_{j,k}\}_{j,k\in\mathbb{Z}},$$

and

$$\mathcal{B}_j = \{\varphi_{j,k}\}_{k\in\mathbb{Z}} \cup \{\psi_{l,k}\}_{l\geq j, k\in\mathbb{Z}}.$$

In the sequel, without lost of generality on the theoretical results, we will consider the second case $\mathcal{B}_j$ at scale $j = 0$. For the sake of readability, we denote $\varphi_{0,k} = \psi_{-1,k}$ for $k \in \mathbb{Z}$. The coefficients $d_{j,k} = \langle f, \psi_{j,k}\rangle$ for $j \geq 1$ and $k \in \mathbb{Z}$ are called *wavelet coefficients*, while the coefficients $c_{0,k} = \langle f, \psi_{-1,k}\rangle$ for $k \in \mathbb{Z}$ are called *approximation coefficients*. The orthogonality of the construction is of great theoretical benefit. For example, it simplifies the study of approximation errors and performances of non-parametric estimation. Moreover, the *wavelet coefficients* of a Gaussian white noise have also a Gaussian distribution. This latter remark will be at the heart of Part I and Part II of this thesis. However, in practice, orthogonal constructions are not flexible. For example, they cannot be symmetric. Bi-orthogonal constructions [Cohen et al., 1992] can be used to overcome this issue. They consist in two dual multiresolution analyses whose scaling functions φ and φ^* satisfy $\langle \varphi, \varphi^*(\cdot - k)\rangle = \delta_{0,k}$. The first one is used for decomposition while the second one is used for reconstruction. The system $(\varphi, \psi, \varphi^*, \psi^*)$ is a basis of $L^2(\mathbb{R})$ and verifies

$$\int_{\mathbb{R}} \psi_{j,k}(t)\psi^*_{j',k'}(t)\, dt = \delta_{j-j'}\delta_{k-k'}; \quad \forall j \geq -1, \forall j' \geq -1, \quad \forall (k,k') \in \mathbb{Z}^2,$$

where δ is the *Kronecker symbol*, and any function of $L^2(\mathbb{R})$ can be expressed as

$$f = \sum_{j\geq -1, k\in\mathbb{Z}} \langle f, \psi_{j,k}\rangle \psi^*_{j,k}. \tag{2.5}$$

2.3.3 Fast Wavelet Transform (FWT)

One of the keys that explain the success of the Fourier Transform in real applications is the *Fast Fourier Transform* (FFT) algorithm. Discovered by Cooley and Tukey [1965], this algorithm allows to compute the Fourier transform of sampled signal of length N in $N\log_2(N)$ operations instead of N^2. When N is dyadic the FFT factorizes the Fourier transfer matrix into $2\log(N)$ sparse matrices. This idea of reduction via a dyadic scheme appears also in wavelet analysis. The Fast Wavelet Transform (FWT) is due to Mallat [1989] who made the connection between the wavelet transform and the pyramidal filtering algorithms (cf. Adelson et al. [1987]). In order to perform the basis change induced from equation (2.1) at each scale j, finite filters h and g are used instead of analytic functions φ and ψ. These filters are defined such as they allow to compute *scaling and wavelet coefficients*, respectively, from fine to coarse scales in the following way

$$c_{j,k} = 2\sum_{k'\in\mathbb{Z}} h[k]\, c_{j+1,2k'-k},$$

$$d_{j,k} = 2\sum_{k'\in\mathbb{Z}} g[k]\, c_{j+1,2k'-k}.$$

This forward transformation corresponds to the *direct sum* iteration (2.4). It can be computed accurately by a scheme of alternating downsampling/filtering in N operations. To have a counter part of the reconstruction formula (2.3), approximations from coarse to fine scales are computed by a cascade filtering

$$c_{j+1,k} = \frac{1}{2}\sum_{k'\in\mathbb{Z}} h[2k'-k]\, c_{j,k'} + g[2k'-k]\, d_{j,k'},$$

until the finest possible scale is reached. Note that in the bi-orthogonal case, h and g should be replaced, in the later equation, by their dual filters h^* and g^* corresponding to the couple (φ^*, ψ^*). This inverse transformation can also be computed by a scheme of alternating upsampling/filtering in N operations.

2.3.4 Multivariate MRA

There is two different paradigms for the construction of multivariate wavelet basis from univariate ones; separable and non-separable constructions. The starting point is a system (φ, ψ) which provides a multiresolution analysis $\{V_j\}_{j\geq 0}$ of $L^2(\mathbb{R})$ which leads to the following decomposition

$$L^2(\mathbb{R}) = \overline{\bigoplus_{l\geq -1} W_l}.$$

with $W_{-1} = V_0$. In order to get a multivariate MRA, we should first remark that

$$L^2(\mathbb{R}^d) = \overline{\bigcup_{l\geq 0} V_l \otimes \cdots \otimes V_l}.$$

Separable MRA

The most natural way to construct a wavelet basis of $L^2(\mathbb{R}^d)$, $d > 1$, is through a *tensor product*. Thus, a basis of $L^2(\mathbb{R}^d)$ can be constructed using the univariate tensor iteration (2.4) at each dimension in the following way

$$\Big(V_l \oplus W_l \cdots \oplus W_{j-1}\Big) \oplus \cdots \oplus \Big(V_l \oplus W_l \cdots \oplus W_{j-1}\Big) = V_j \otimes \cdots \otimes V_j, \quad j > l. \tag{2.6}$$

This corresponds to the following decomposition (letting $l = 0$)

$$L^2(\mathbb{R}^d) = \overline{\bigoplus_{j_1,\cdots,j_d\geq -1} W_{j_1} \otimes \cdots \otimes W_{j_d}} \tag{2.7}$$

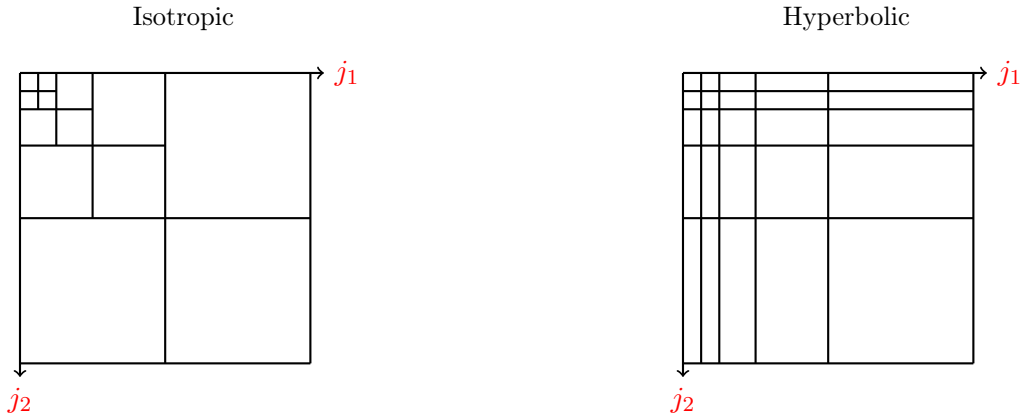


Figure 2.3: Tilings of the wavelet representation in the "isotropic" and "hyperbolic" cases.

Let us denote $\bar{j} = (j_1, \cdots, j_d)$ and $\bar{j} = (k_1, \cdots, k_d)$ The corresponding wavelet basis is called *"Hyperbolic"*[1](cf. DeVore et al. [1998]) and it is obtained using the *tensor product* of univariate wavelets

$$\mathcal{B}_H = \{\psi_{\bar{j},\bar{k}}(\mathbf{x})\}_{\bar{j},\bar{k}} = \{\psi_{j_1,k_1}(x_1) \times \cdots \times \psi_{j_d,k_d}(x_d) \quad : \quad \bar{j} = (j_1, \cdots, j_d) \in (\mathbb{N} \cup \{-1\})^d, \bar{k} = (k_1, \cdots, k_d) \in \mathbb{Z}^d\}. \tag{2.8}$$

Non-separable MRA

Notice that all combinations of the scale parameter $\bar{j}$ are allowed in the separable construction. In particular, scales are mixed and thus the multiresolution structure is lost. To retrieve this structure, the degree of liberty of the scaling parameter should be reduced to one. Let us denote $S_l^0 = V_l$ and $S_l^1 = W_l$ for each $l \in \mathbb{Z}$. The decomposition (2.6) can also be written in the following way

$$\left(S_l^0 \oplus S_l^0 \cdots \oplus S_l^0\right) \oplus \bigoplus_{l<j} \bigoplus_{(i_1,\cdots,i_d)\in\{0,1\}^d\setminus\{(0,\cdots,0)\}} \left(S_l^{i_1} \oplus S_l^{i_2} \cdots \oplus S_l^{i_d}\right) = S_j^0 \otimes \cdots \otimes S_j^0. \tag{2.9}$$

In this case the sub-spaces are arranged by scale, thus providing a multiresolution structure. Note that the approximation sub-spaces are used at all scales $l < j$. A basis of $L^2(\mathbb{R}^d)$ in the non–separable case is then given by

$$L^2(\mathbb{R}^d) = \bigoplus_{j\geq 0} \bigoplus_{(i_1,\cdots,i_d)\in\{0,1\}^d\setminus\{(0,\cdots,0)\}} \left(S_j^{i_1} \oplus S_j^{i_2} \cdots \oplus S_j^{i_d}\right) \tag{2.10}$$

Let us denote $\psi_{j,k}^{(0)} = \varphi_{j,k}$ and $\psi_{j,k}^{(1)} = \psi_{j,k}$. Then, the wavelet basis corresponding to the non-separable MRA is called *"isotropic"* (cf. Mallat [1989] or Daubechies [1992]) and it is given by the collection

$$\mathcal{B}_I = \{\psi_{j,k_1}^{(i_1)}(x_1) \times \cdots \times \psi_{j,k_d}^{(i_d)}(x_d) : j \geq 0, (k_1, \cdots, k_d) \in \mathbb{Z}^d, (i_1, \cdots, i_d) \in \{0,1\}^d\setminus\{(0,\cdots,0)\}\} \tag{2.11}$$

The term *"isotropic"* comes, of course, from the fact that the wavelet product is allowed only for functions at the same scale which provides an isotropic treatment on the different variables. It is not mandatory to impose isotropy to have a non-separable construction which have a multiresolution structure. Roughly speaking (*cf.* [Triebel, 2004]), it is sufficient to impose that the degree of freedom of the scale vector is equal to one. This can also be done imposing that the ratio between two scales j_i and $j_{i'}$, with $i \neq i'$, is constant ($j/j' =$ Constant). The *isotropic* case is obtained if the ratio is equal to one. Otherwise, the obtained construction is called *anisotropic*. Such a construction is not of great interest as only one *anisotropy* is

[1]This wavelets appears under other names names in the literature, such as mixing scales Remenyi et al. [2014] or rectangular Zavadsky [2007].

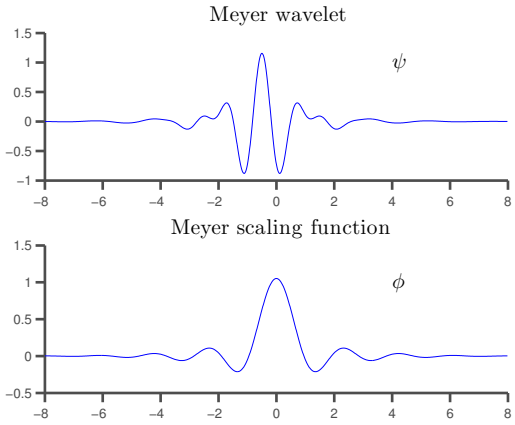


Figure 2.4: Meyer wavelet.

allowed; this prior is too strong in practice. In the sequel we will consider only the *isotropic* and the *hyperbolic* constructions and their properties from a statistical point of view. Figure 2.3 shows the difference in the spatial tilling between the isotropic and the hyperbolic case. Figure 2.4 shows the "Meyer" wavelet and scaling functions, while Figure 2.5 highlights the difference in the supports of the isotropic and hyperbolic two-dimensional Meyer wavelets across scales.

2.4 Related functional spaces

The notion of sparsity has been at the heart of wavelet-based processing since its beginning in early 90's. It relies on the ability of representing a given function using only a few wavelet coefficients. This concept was mainly motivated by compression problems (cf. DeVore et al. [1992a]), but turned out to be useful for denoising (cf. Donoho et al. [1995]) and more recently for recovery in what called the compressed sensing theory (cf. Candès et al. [2006]). This notion can be quantified using functional spaces. This spaces are also crucial to evaluate the statistical performances of wavelet estimators. To emphasize this ideas we start by highlighting the link between frequency/time-scale representations and functional spaces such as Sobolev and Besov. For the sake of readability, we restrict ourselves to the one-dimensional case. Extensions to classical multivariate isotropic spaces are straightforward.

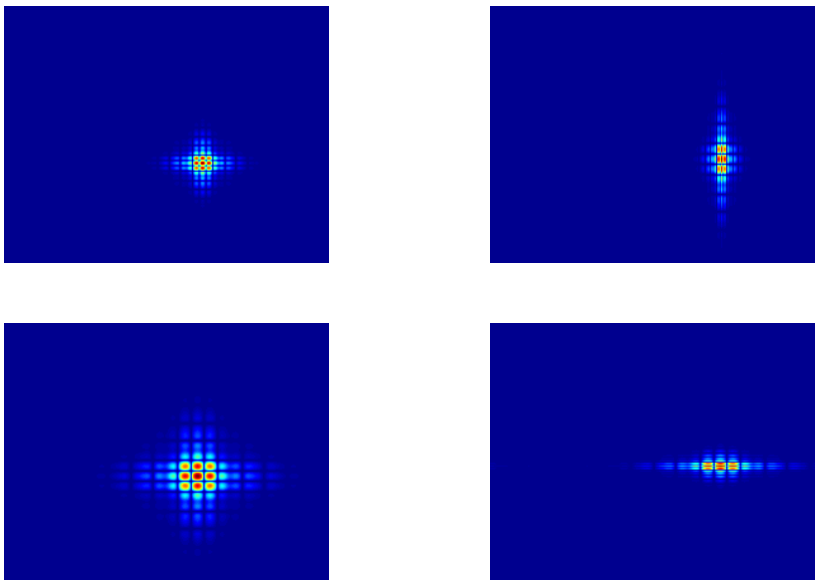
2.4.1 Sobolev Spaces

Sobolev spaces appear naturally in the study of partial differential equations (pde's). Actually, it has been early observed that, for theoretical and practical reasons, spaces of continuous functions $\mathcal{C}^m$ are not well adapted for solutions of pde's (see for example the book of Brezis [1983]). In particular, the summability of the solution and its derivatives appearing in the equation must be verified. Sobolev spaces provides such a characterization

$$W^{s,p}([0,1]) = \left\{ f \in L^p([0,1]) \quad \middle| \quad \forall\, \alpha \leq s, \quad D^\alpha f \in L^p([0,1]) \right\},$$

where $p \in [1, +\infty]$ and D^α is the α-th derivative. Such a definition can be extended to the multivariate case by considering α as multiindex. The case $p = 2$ is for particular interest because the resulting space $H^s([0,1]) = W^{s,2}([0,1])$ forms a Hilbert space. Moreover, as the Fourier transform is well defined for functions in $L^p([0,1])$, the Parseval identity provides an alternative characterization of these spaces

$$H^s([0,1]) = \left\{ f \in L^2([0,1]) \quad \middle| \quad ||f||_{H^s} = \left(\sum_{\ell \in \mathbb{Z}} |\langle f, e^{-i\ell \cdot} \rangle|^2 \left(1 + |\ell|^2\right)^s \right)^{1/2} < \infty \right\},$$



Isotropic Hyperbolic

Figure 2.5: Two-dimensional Meyer wavelet.

In words, functions of $H^s([0,1])$ are described by the decay of their Fourier coefficients. This has an interesting consequence on the approximation properties of the Fourier expansion. In fact, a truncated N-term linear approximation[2] of f through its (discrete) Fourier expansion is given by[3]

$$f_N = \sum_{k=-N/2}^{N/2} \langle f, e^{-2i\pi k.} \rangle e^{2i\pi k.}.$$

Note that

$$|k|^s |\langle f, e^{-2i\pi k.} \rangle| \leq \left(\sum_{\ell \in \mathbb{Z}} |\langle f, e^{-2i\pi \ell.} \rangle|^2 \Big(1 + |\ell|^{2s} \Big) \right)^{1/2}.$$

The right hand side converges whenever f is in $H^s([0,1])$, and thus we have

$$f \in H^s([0,1]) \Longrightarrow |\langle f, e^{-2i\pi k.} \rangle| \leq \frac{C}{|k|^s}.$$

As a consequence, the $N-term$ approximation of f verifies the following error rate (cf. Mallat [2008], Theorem 9.2)

$$||f_N - f||_2^2 \leq \mathcal{O}(N^{-2s}). \tag{2.12}$$

This rate shows the efficiency of Fourier coefficients in representation of functions of $H^s([0,1])$ thanks to the equivalence between smoothness and Fourier coefficients decay. Wavelets provide a similar rate of convergence; if the mother-wavelet ψ and the scaling function ϕ at least s vanishing moments, one has (cf. Mallat [2008], Theorem 9.5)

$$||P_j f - f||_2^2 \leq \mathcal{O}(2^{-2js}), \tag{2.13}$$

with $N = 2^j$. Sobolev spaces gather all spaces of functions that are characterized by global smoothness. In particular, the approximation error (2.12) is valid for continuous functions $\mathcal{C}^\alpha$ and α-Lipschitz functions.

[2] In fact, there are $N+1$ terms by considering the term given by $k = 0$.
[3] The Fourier expansion on $[0,1]$ is calculated through 2π-periodization.

However, in many applications, particularly image processing, one is confronted to functions that are piecewise smooth. Such functions present discontinuities with the result that the Fourier series expansion does not converge anymore. These functions cannot belong to any Sobolev space and cannot be characterized by Fourier representations. In order to fill this gap Besov [1956] introduced spaces of functions that allows singularities.

2.4.2 Besov Spaces

Besov spaces can be seen as an extension of Sobolev spaces. Their mathematical definition requires the definition of the smoothness modulus (cf. DeVore and Popov [1988]). Let us define the first and second order finite difference operators

$$\Delta_h f(x) = f(x+h) - f(x), \quad \forall (x,h) \in \mathbb{R}^2,$$

and its r-th order extension

$$\Delta_h^r(f)(x) = \underbrace{\Delta_h \circ \Delta_h \circ \cdots \Delta_h}_{r}(f)(x).$$

Δ_h^r is decaying to 0 when the step size $h \to 0$. The *smoothness modulus* measures this decay in a given L^p-norm

$$\omega_{r,p}(f,h) = \sup_{|\tilde{h}| \leq h} ||\Delta_{\tilde{h}}^r f||_p.$$

Derivability of order r allows decay rate of order h^r. This order of derivability is also equivalent to say that the function is r-Lipschitz. The role of the L^p-norm is to maintain this regularity by ignoring singularities having regularities smaller than r up to a certain order. This is done thanks to the possibility of calibrating regularities by compensating the loss in the dimension in singularities. This leads to the definition of Besov spaces as given by DeVore and Popov [1988]

Definition 2.4.1. *Let $f \in L^p$, $1 \leq p \leq \infty$. A Besov space of indices $s > 0$, p and $1 \leq p, q \leq \infty$, is denoted by $B^s_{p,q}$ and characterized by*

$$f \in B^s_{p,q} \Longleftrightarrow |f|_{B^s_{p,q}} = \begin{cases} \left(\int_0^{+\infty} \left(h^{-s} \omega_{\lfloor s+1 \rfloor, p}(f,h) \right)^q \frac{dh}{h} \right)^{\frac{1}{q}} < +\infty, & \text{if} \quad q < +\infty \\ \sup_{h>0} \left(h^{-s} \omega_{\lfloor s+1 \rfloor, p}(f,h) \right) < +\infty & \text{if} \quad q = +\infty \end{cases}$$

The parameter q is not of capital importance compared to s and p. It is often ignored or just assumed to be $+\infty$. The application $|.|_{B^s_{p,q}}$ is called the Besov seminorm of indices (s,p,q). The Besov space norm is given by

$$||f||_{B^s_{p,q}} = |f|_{B^s_{p,q}} + ||f||_p.$$

It is interesting to note that Besov Spaces gather functions with Lipschitz and Holder regularities as well as functions in Sobolev spaces. This is due to the universality of the regularity measure in the sense of the *smoothness modulus* which can be seen as a local continuity criteria. For example, when $p = q = 2$, $B^s_{2,2}$ coincides with the Sobolev space H^s and when $s < 1$, $B^s_{p,\infty}$ is the Lipschitz space $Lip(s, L^p)$ and $B^s_{\infty,\infty}$ is the Hölder space of order s.

A fundamental characterization of Besov spaces is obtained by wavelet coefficients. In fact, in the same way that Sobolev spaces are approximation spaces of Fourier coefficients, i.e on which the error rate in (2.12) is obtained, Besov spaces are natural approximation spaces for wavelet coefficients. The following theorem provides a characterization of Besov spaces through wavelets coefficients (cf. DeVore et al. [1992b])

Definition 2.4.2. *Let f be a function of $L^p(\mathbb{R})$ and*

$$c_k = \int \varphi_{0,k}(x) f(x) dx \quad ; \quad d_{j,k} = \int \psi_{j,k}(x) f(x) dx.$$

Then, we have the following characterization of Besov spaces

$$f \in B^s_{p,q} \Longleftrightarrow ||f||_{B^s_{p,q}} = \begin{cases} |c|_p + \left(\sum_{j \geq 0} 2^{jq(s-1/p+1/2)} |d_j|_p^q \right)^{\frac{1}{q}} < +\infty, & \text{if} \quad q < +\infty \\ |c|_p + \sup_{j \geq 0} 2^{jq(s-1/p+1/2)} |d_{j,.}|_p < +\infty & \text{if} \quad q = +\infty, \end{cases}$$

where $c = \{c_k\}_k$ and $d_j = \{d_{j,k}\}_k$. This latter characterization provides an alternative definition Besov norms based on wavelet coefficients. It consists in taking the p-norm of wavelet coefficients at each scale and a q-norm across scales. The smoothness parameter s appears as an exponent which controls the decay of the coefficients at each scale. The linear approximation properties of wavelet are stated in the following theorem (cf. Härdle et al. [2012], Theorem 9.6).

Theorem 2.4.1. *Let $v \in \mathbb{N}$, $0 < s < v$, $1 \leq p, q \leq +\infty$ and (φ, ψ) an orthogonal multi-resolution system such that φ and ψ have v vanishing moments. Then we have*

$$f \in B^s_{p,q} \Longleftrightarrow \begin{cases} \forall j \in \mathbb{N}, \quad ||P_j f - f||_2 \leq \mathcal{O}(2^{-js} u_j), & \text{if} \quad p \geq 2 \\ \forall j \in \mathbb{N}, \quad ||P_j f - f||_2 \leq \mathcal{O}(2^{-j(s+1/2-1/p)} u_j), & \text{if} \quad 1 \leq p < 2, \end{cases}$$

where $(u_j)_{(j \in \mathbb{N})} \in \ell_q(\mathbb{N})$ is a sequence of positive numbers.

This theorem can also be seen as a $N - term$ approximation with $N = 2^j$ in the same spirit as the approximation given by the Fourier transform.

2.4.3 Nonlinear approximation

Approximations such as wavelet projection and truncated Fourier series are called linear in the sense that they depend only on the number N of coefficients and not on the function to be approximated. However, it is clear that for obtaining the best $N - term$ approximation, over a given basis, coefficients should be chosen such as the error is the smallest possible. In other words, large coefficients should be taken into account in the approximation. This is naturally done in the case of Fourier approximation as the largest coefficients are likely to be concentrated around the origin which is not the case for wavelet coefficients. This latter remark motivated the introduction of nonlinear approximation (cf. DeVore [1998]). It consists in choosing the N term with largest amplitudes. This is equivalent to keeping coefficients up to a threshold T. Let $\mathcal{I}_T$ be the set of multiindex corresponding to the coefficients that are kept

$$\mathcal{I}_T = \left\{ (j,k) \in \mathbb{N} \cup \{-1\} \times \mathbb{Z}) \quad \middle| \quad |\langle f, \psi_{j,k} \rangle| \geq T \right\}.$$

Let us set $N = \#(\mathcal{I}_T)$. The nonlinear N-term approximation of f, in the orthogonal case, is given by

$$f_N = \sum_{(j,k) \in \mathcal{I}_T} \langle f, \psi_{j,k} \rangle \psi_{j,k}. \tag{2.14}$$

Besov spaces are still approximation spaces for wavelet coefficients in the nonlinear case. The approximation rates are given by the following theorem (cf. Mallat [2008], Theorem 9.10).

Theorem 2.4.2. *Let $v \in \mathbb{N}$, $0 < s < v$, $1 \leq p, q \leq +\infty$ and (φ, ψ) a orthogonal multi-resolution system such as φ and ψ have v vanishing moments. Then we have*

$$f \in B^s_{p,q} \Longleftrightarrow \forall j \in \mathbb{N}, \quad ||f_N - f||_2 \leq \mathcal{O}(N^{-s} u_j),$$

where $(u_j)_{(j \in \mathbb{N})} \in \ell_q(\mathbb{N})$ is a sequence of positive numbers.

In particular, in the case when $p < 2$, the nonlinear approximation improves the decay of the coefficients and thus the approximation error. It turns out that the case $p = 1$ appears naturally in signal and image modeling, which favours nonlinear approximation.

2.4.4 Link with signal and image processing

The simplest interpretation of a Besov space $B^s_{p,q}$ is to say that it is roughly the space of functions that have s derivatives in L^p which can be distinguished by a finer regularity parameter q. This is, however, not naturally appealing for functions representing real life signals and natural images like the ones we will deal with in this thesis. A more appealing assumption on such functions is their membership to a *Bounded Variation* (BV) space as explained by Meyer [2001]. Such spaces contain functions whose weak gradient is finite measure. This is accomplished by integration against a test function.

Definition 2.4.3. *A function $f : [0,1]^d \longrightarrow \mathbb{R}$ is said to be of bounded variation; $f \in BV([0,1]^d)$, if and only if $f \in L^1(\Omega)$ and*

$$|f|_{BV([0,1]^d)} = \int_{[0,1]^d} |Df| := \sup_{||g||_\infty \leq 1} \left\{ \int_{[0,1]^d} f \, div g \, dx \quad \middle| \quad g \in C^1_c([0,1]^d, \mathbb{R}^d) \right\} < \infty.$$

where *div* refers to the *divergence operator*. Thus, BV spaces are Banach spaces defined by duality and can be equipped by the following norm

$$||f||_{BV([0,1]^d)} = ||f||_{L^p([0,1]^d)} + \int_{[0,1]^d} |Df|.$$

These spaces do not have unconditional bases. They can, however, be linked to Besov spaces by embedding properties. This allows to approximate the $||.||_{BV([0,1]^d)}$. In particular, we have the following embeddings

$$B^1_{1,1}([0,1]^d) \subset BV([0,1]^d) \subset B^1_{1,+\infty}([0,1]^d). \tag{2.15}$$

To understand how these embeddings are obtained, let us consider, for the sake of simplicity, the one dimensional case for which the BV semi-norm is given by

$$|f|_{BV([0,1])} = \int_{[0,1]} |f'(x)| \, dx$$

The embeddings (2.15) are direct consequence of the following theorem (cf. Mallat [2008], Theorem 9.13)

Theorem 2.4.3. *Let f be a function of $L^2([0,1])$ and a wavelet ψ of bounded variation. There exist $A, B > 0$ such that*

$$|f|_{BV([0,1])} \leq B \times \sum_{j \geq -1, k \in \mathbb{Z}} 2^{j/2} |\langle f, \psi_{j,k} \rangle| = B||f||_{B^1_{1,1}([0,1])} \tag{2.16}$$

and

$$|f|_{BV([0,1])} \geq A \times \sup_{j \geq 0} \left\{ \sum_{k \in \mathbb{Z}} 2^{j/2} |\langle f, \psi_{j,k} \rangle| \right\} = B||f||_{B^1_{1,\infty}([0,1])} \tag{2.17}$$

Remark 1. *Inequalities (2.16) and (2.17) are the "tightest" possible in terms of wavelet coefficients; the embedding (2.15) gives the most accurate link between BV and Besov spaces.*

The complete proof can be found in Mallat [2008]. We give a sketch of it as the representation of images through their wavelet coefficients is at the heart of the present thesis. The first inequality is a consequence of the fact that

$$|f|_{BV([0,1])} = \left| \sum_{j \geq -1, k \in \mathbb{Z}} \langle f, \psi_{j,k} \rangle \psi_{j,k} \right|_{BV([0,1])} = \sum_{j \geq -1, k \in \mathbb{Z}} \langle f, \psi_{j,k} \rangle |\psi_{j,k}|_{BV([0,1])}.$$

Moreover, a change of variable gives

$$|\psi_{j,k}|_{BV([0,1])} = 2^{j/2}|\psi|_{BV([0,1])},$$

which leads to the inequality (2.16). The second inequality comes from the fact that the wavelet ψ has at least one vanishing moment; thus, it has a primitive with the same support. this allows and integration by parts for each L^2-product of the wavelet decomposition

$$\langle f, \psi_{j,k}\rangle = 2^{-j/2}\langle f', \Psi_{j,k}\rangle, \quad \forall j \geq -1, \forall k \in \mathbb{Z}, \quad \text{with} \quad \Psi_{j,k}(x) = \Psi(2^j x - k).$$

The inequality (2.17) follows by summation. The embedding (2.15) has an essential consequence in function estimation. In fact, recovering a function $f \in BV([0,1])$ from a corrupted observation g can be done in a variational framework (cf. [Rudin et al., 1992])

$$\tilde{f} = \arg\min_{f\in[0,1]} \left\{ |f|_{BV([0,1])} + \lambda||f-g||_2^2 \right\}.$$

As we now the semi-norm $|.|_{BV([0,1])}$ is well concentrated between two Besov semi-norms -Moreover, it is controlled by the semi-norm $|f|_{B^1_{1,1}([0,1])}$- the latter variational problem can be replaced by

$$\tilde{f} = \arg\min_{f\in[0,1]} \left\{ |f|_{B^1_{1,1}([0,1])} + \lambda||f-g||_2^2 \right\}. \tag{2.18}$$

That is, the relatively abstract BV semi-norm is replaced by $|.|_{B^1_{1,1}([0,1])}$ which is simply a weighted ℓ_1-norm of wavelet coefficients. As a result, the minimization problem (2.18) can be fully characterized in the wavelet domain.

$$\tilde{f} = \arg\min_{f\in[0,1]} \left\{ \sum_{j\geq -1, k\in\mathbb{Z}} |f_{j,k}| + \lambda \sum_{j\geq -1, k\in\mathbb{Z}} |f_{j,k} - g_{j,k}|^2 \right\}, \tag{2.19}$$

where $f_{j,k} = \langle f, \psi_{j,k}\rangle$ and $g_{j,k} = \langle g, \psi_{j,k}\rangle$ are the respective wavelet coefficients of f and g. Note that the L^2 normalizations disappear as they appear on both terms and thus can be factorized. If we assume that the coefficients are uncorrelated, then the problem (2.19) is completely separable; that is, it can be solved for each couple $(j \geq -1, k \in \mathbb{Z})$

$$\tilde{f}_{j,k} = \arg\min_{f_{j,k}} \left\{ |f_{j,k}| + \lambda |f_{j,k} - g_{j,k}|^2 \right\},$$

which solution is the zero of its derivative

$$\text{sgn}(f_{j,k}) + 2\lambda(f_{j,k} - g_{j,k})$$

That is

$$\tilde{f}_{j,k} = \mathcal{T}_{\frac{1}{2\lambda}}(g_{j,k})$$

where $\mathcal{T}_\gamma$, with $\theta > 0$, is the *soft-thresholding* operator

$$\mathcal{T}_\theta(u) := \begin{cases} u - \theta\dfrac{u}{|u|}, & \text{if} \quad |u| > \theta \\ 0, & \text{if} \quad |u| \leq \theta \end{cases} \tag{2.20}$$

This operator is the base of the celebrated *wavelet shrinkage* framework introduced by Donoho and Johnstone [1994] for non-parametric estimation.

2.5 Nonparametric function estimation

2.5.1 Generalities

Non-parametric estimation aims at recovering, predicting or estimating a function from an observation under very general assumptions on regularity, such as the membership to a particular functional space. We define some notions which are going to be useful in the sequel.

Statistical modeling

The unknown function might refer different observations of interest, depending on the considered statistical model. The first example is density estimation where the observable quantities

$$(X_1, \cdots, X_n) \tag{2.21}$$

are n independent and identically distributed (*i.i.d.*) random variables with unknown density f. The second example concerns continuous stochastic process $\{X(t); \quad t \in [0,1]\}$ defined by

$$dX(t) = f(t)dt + (n)^{-1/2}dW(t), \tag{2.22}$$

where $f : [0,1] \to \mathbb{R}$ is a unknown function playing the role of the drift and dW is a Wiener process. The third example arises when the unknown function is sampled data that are corrupted by additive Gaussian noise

$$X_t = f(t/n) + z_t, \tag{2.23}$$

where $(z_1, \cdots, z_n)$ are *i.i.d.* random variables with zero mean and unit variance. This latter case is of particular interest because many problems of denoising in signal and image processing are given under model (2.23).

L^p-risk

A function which takes the observations as variables is called *"estimator"*. The risk evaluates the error of the estimation when reconstructing the original function.

Definition 2.5.1. *Let $\widehat{f}_n : [0,1]^d \to \mathbb{R}$ be an estimator of a function f. The L^p-risk of $\widehat{f}_n$ is defined as*

$$r_p(\widehat{f}_n) = \mathbb{E}_f(||\widehat{f}_n - f||_p^p),$$

where $\mathbb{E}_f$ is the expectation with respect to the probability law of the observations.

Different values of p allow different characterizations of the local behavior of the estimator. The most commonly used value is $p = 2$ for which the risk is simply the mean squared error. The risk does not only allow to evaluate the performances of a given estimator but also the construction of optimal estimators through the minimax framework.

Minimax risk

The minimax framework was introduced by Von Neumann in the $1920's$ for the study of optimal strategies in two-player zero-sum game theory. Its main purpose is minimizing the possible loss for a worst case scenario. An important breakthrough in statistics was the extension of Von Neumann's ideas to classical problems in statistical decision theory by Wald [1945a,b, 1947, 1949]. In particular, for function estimation, we have the following definition over a given functional space $\mathcal{F}^\alpha$ with a regularity (multi-) parameter α.

Definition 2.5.2. *The minimax risk over $\mathcal{F}^\alpha$ is defined by*

$$R_{n,p}(\mathcal{F}^\alpha) = \inf_{\widehat{f}_n} \sup_{\mathcal{F}^\alpha} \mathbb{E}_f(||\widehat{f}_n - f||_p^p),$$

The study of the *minimax risk* relies on studying its asymptotic convergence through *lower bounds* and *upper bounds*.

Definition 2.5.3. *We call a lower bound any sequence u_n of positive numbers for which a exists $c > 0$ such that*

$$R_{n,p}(\mathcal{F}^\alpha) \geq c\, u_n$$

Definition 2.5.4. *We call an upper bound any sequence u_n of positive numbers for which a exists $c > 0$ such that*

$$R_{n,p}(\mathcal{F}^\alpha) \leq c\, u_n$$

Definition 2.5.5. *We call the minimax rate of convergence any sequence u_n of positive numbers for which a exist $c_1, c_2 > 0$ such that*

$$c_1\, u_n \leq R_{n,p}(\mathcal{F}^\alpha) \leq c_2\, u_n$$

Definition 2.5.6. *Let $\widehat{f}_n$ be an estimator of f. $\widehat{f}_n$ is called optimal, if it achieves the minimax rate of convergence u_n, in the sense that*

$$\sup_{\mathcal{F}^\alpha} \mathbb{E}_f(||\widehat{f}_n - f||_p^p) = \mathcal{O}(u_n).$$

The *minimax rate of convergence* allows to know exactly the best performance that can be achieved for a given statistical model on a given functional space with respect to a given risk. The first examples of minimax bounds where established by Farrell [1972] for density estimation at fixed point and not on global functional spaces. *minimax rate of convergence* for general L^p errors, with $1 \leq p \leq \infty$, are due to Bretagnolle and Huber [1979] Bretagnolle and Huber. The majority of these works considered Hölder regularities on which the authors obtained rates of the form $n^{-\alpha/(2\alpha+1)}$ up to a logarithmic factor, where α is the Hölder regularity parameter, or their multivariate extensions which are of the form $n^{-\alpha/(2\alpha+d)}$, where d is the dimension of the observations. Nemirovskii et al. [1985] found the same rates when the considered functional class is a Sobolev space under regression models. Finally, these rates where again obtained for Besov spaces by Kerkyacharian and Picard [1992]. A remarkable result appears in the work of Nemirovskii [1985], where it is stated that linear estimators, such as Kernel-based smoothing, splines smoothing or projections methods, cannot achieve these rate of convergence. The rise of wavelets had set the ground for the development of nonlinear estimators which overcome this limitation.

2.5.2 Wavelet estimation

Wavelet based function estimation appeared shortly after the emergence of wavelets. The first works concerns linear methods based on projection for density estimation (cf. Doukhan and León [1990] and Kerkyacharian and Picard [1992]). In most cases, the results obtained for density estimation can be naturally extended to regression and Wigner process models. Nonlinear methods, which are based on thresholding, were introduced in a series of papers by Donoho and Johnstone [1994, 1995], then developed and synthesized by Donoho et al. [1995, 1996]. We present the most remarkable results on the performances of linear and nonlinear estimation which are valid for problems (2.21), (2.22) and (2.23). Any of these problems can be written in the wavelet domain

$$Y_{j,k} = d_{j,k} + z_{j,k}, \quad j = 0, \cdots, J-1, \quad k = 1, \cdots, 2^j,$$

with $J = \log_2 n$. The $Y_{j,k}$ are wavelet coefficients of the observation X_t, $d_{j,k}$ are wavelet coefficients of the unknown f and $z_{j,k}$ are the wavelet coefficients of the noise component z_t. The coefficients corresponding to $(-1, 1)$ are approximation coefficients. By considering that the unknown belongs to a given functional space, wavelet coefficients are expected to be more or less sparse. Wavelet estimation aims at estimating $d_{j,k}$ from $Y_{j,k}$ by putting some coefficients to zero. Thus, one needs to fix the set of indices that are kept $\mathcal{I}_n$. The estimated coefficients $\hat{d}_{j,k}$ are given as

$$\hat{d}_{j,k} = \mathcal{T}(d_{j,k}) \mathbb{1}_{\mathcal{I}_n}(j,k), \tag{2.24}$$

where $\mathbb{1}_{\mathcal{I}_n}(j,k)$ is the indicator function; equal to one if $(j,k) \in \mathcal{I}_n$ and zero otherwise. If The estimated coefficients are used to construct an estimation of f

$$\hat{f} = \sum_{j=1}^{J} \sum_{k=1}^{2^j} \hat{d}_{j,k} \psi_{j,k}$$

When $\{z_t\}_{t=1}^n$ are i.i.d Gaussian, so are their wavelet coefficients[4]. This property is crucial for the construction of wavelet estimators. It facilitates the study of theoretical performances and the choice of thresholding parameters in practice.

Linear estimation

Linear estimation relies on the approximation properties of projection operators. It uses the fact that the finest scale in the decomposition can be calibrated to detect features with a given smoothness. Let f be a function observed under models (2.21), (2.22) or (2.23). A linear estimator -by truncation- is constructed by taking $\hat{d}_{j,k}$ as $Y_{j,k}$ up to a certain scale J^* and ignoring the others. This is equivalent to taking $\mathcal{T}(d_{j,k}) = d_{j,k}$ and

$$\mathcal{I}_n = \Big\{ (j,k) \quad : \quad j = 0, \cdots, J^* - 1, \quad k = 1, \cdots, 2^j \text{ with } J^* < J \Big\}.$$

In the sequel, we note f_n^L the linear estimator. The choice of J^* is related to the *a priori* regularity of f and the variance of the noise which we assumed to unit. In particular, we have the following theorem (cf. Kerkyacharian and Picard [1993])

Theorem 2.5.1. *Assume that $f \in B^s_{p,q}$ with $2 \leq p \leq \infty$ and $1 \leq q \leq \infty$. Choose $J^* \simeq n^{\frac{1}{2s+1}}$, Then*

$$\inf_{\widehat{f}_n^L} \sup_{B^s_{p,q}} \mathbb{E}(||\widehat{f}_n^L - f||_2^2) = \mathcal{O}(n^{-\frac{-2s}{2s+1}})$$

Remark 2. *The rate of convergence obtained in the last theorem is minimax. It can be obtained also for other L^p errors by replacing $\mathcal{O}(n^{-\frac{-2s}{2s+1}})$ by $\mathcal{O}(n^{-\frac{-2sp}{2s+1}})$. Note however that the theorem is valid only for $p \geq 2$. In fact, for $p \leq 2$ linear estimators cannot achieve minimax rates of convergence, not even up to logarithmic terms. Moreover, the choice of the critical scale J^* depends on the regularity of the unknown which is not of practical interest.*

Nonlinear estimation

The nonlinear estimation relies more on the sparsity of wavelet coefficients than the scale. We take $\mathcal{T}(d_{j,k}) = d_{j,k}$ and

$$\mathcal{I}_n = \Big\{ (j,k) \quad : \quad j = 0, \cdots, J - 1, \quad k = 1, \cdots, 2^j \quad \Big| \quad |Y_{j,k}| > t_n \Big\},$$

where t_n is the threshold. The optimal value of such threshold- in a *minimax* sense- was given in Donoho and Johnstone as $t_n = \sqrt{2 \log n}$. In the case when the standard deviation is not unit but a positive $\sigma \neq 1$ the $t_n = \sigma\sqrt{2 \log n}$. Such a thresholding procedure is called *hard thresholding*. Note $\widehat{f}_n^H$ the obtained estimator. Its performances are given by the following theorem (cf. Donoho et al. [1995])

Theorem 2.5.2. *Assume that $f \in B^s_{p,q}$ with $1 \leq p \leq \infty$ and $1 \leq q \leq \infty$. Choose $t_n = \sqrt{2 \log n}$, Then*

$$\inf_{\widehat{f}_n^H} \sup_{B^s_{p,q}} \mathbb{E}(||\widehat{f}_n^H - f||_2^2) = \mathcal{O}((n/\log n)^{-\frac{-2s}{2s+1}})$$

[4] Under the assumption that the wavelet system is orthogonal.

As in the linear case, it is possible to extend this theorem to more general L^p errors. Note that the *minimax rates of convergence* is obtained up to a logarithmic term. Such a rate is called *nearly minimax*. It is however obtained also for $1 \leq p < 2$ without prior knowledge on the regularity parameter s. This theorem is still valid for an estimator f_n^S constructed by *soft thresholding* (cf Donoho and Johnstone [1995]), where

$$\mathcal{I}_n = \left\{ (j,k) \quad : \quad j = 0, \cdots, J-1, \quad k = 1, \cdots, 2^j \right\},$$

and $\mathcal{T}(d_{j,k}) = \mathcal{T}_{t_n}(d_{j,k})$ is the *soft thresholding* operator defined in (2.20).

2.5.3 Beyond Gaussian situations

In regression problems, the assumption that the noise components $\{z_t\}_{t=1}^n$ is crucial for wavelet estimation. As mentioned before, this hypothesis is crucial for the theory and the practice of wavelet nonparametric estimation. However, one might have to deal with noise components which are not necessarily Gaussian or i.i.d. In such situations, wavelet estimators need to be adapted to the noise distribution. Two seminal works in this direction are attributed to Johnstone and Silverman [1997] who constructed scale dependent thresholding techniques for correlated noise suppression and Neumann [1996] for stationary non-Gaussian time series. We refer to Neumann and von Sachs [1995] for a review on such adaptations. Particular cases, which will be of interest for us in the sequel, are heteroscedastic regression models (cf. Efromovich and Pinsker [1996]). These models are characterized by a time-dependent variance

$$X_t = f(t/n) + \sigma(t) z_t,$$

where $\sigma : [0,1] \to \mathbb{R}^+$. For such models, adaptive wavelet estimation techniques were developed by Cai and Wang [2008] which can estimate locally the value of σ and achieve, up to a logarithmic term, the optimal *rates of convergence* as in the i.i.d. Gaussian case. Now consider that f has values in $[0,1]$ and that σ depends not on t but on $f(t/n)$. The corresponding heteroscedastic model reads as follows

$$X_t = f(t/n) + \sigma(f(t/n)) z_t. \tag{2.25}$$

The problem of estimating f, under model (2.25), using wavelet thresholding was studied first by Fryzlewicz and Nason [2004] for the particular case of Poisson estimation. In this case, the observable quantities X_t are independent variables with Poisson distribution. The mean and the variance are linked via the relation $\mathrm{Var}(X_t) = \sigma(\mathbb{E}(X_t))$ with $\sigma(u) = u$. Their proposed methodology, called *Haar-Fisz*, uses a local means pre-estimation of the unknown f to stabilize the variance. Fryzlewicz [2008] generalized these technique to any non-increasing variance function σ. The obtained *wavelet-Fisz* methodology attains *optimal rates of convergence* as in the i.i.d Gaussian case. Moreover, the methodology was extended to cases when σ is unknown, still with *optimal rates of convergence*. It is interesting to note that the large class of multiplicative models are given in the form can be written in the form (2.25). These models are given as

$$X_t = f(t/n) z_t, \tag{2.26}$$

Such situations arise, for instance, in periodogram estimation which allows spectral density estimation (cf. Fryzlewicz et al. [2008]). Models of the form (2.26) are often handled using *logarithmic transformations* which allow to transform multiplicative models to additive models. The main drawback of *logarithmic* is the fact that they can hide or cancel features with high energy. Moreover, studying the performances of the log-transformed functions is not straightforward. Another way to write (2.26) is of the form (2.25). In fact, equation (2.26) is equivalent to

$$X_t = f(t/n) + f(t/n)\tilde{z}_t, \tag{2.27}$$

where $\tilde{z}_t = z_t - 1$. The *wavelet-Fisz* will be exposed in details later, as it the base of the methodology proposed in Part 1 of the present thesis. In particular the optimal of the estimation in anisotropic multivariate setting is studied.

2.5.4 Multivariate wavelet estimation

Wavelet nonparametric estimation can be extended to the multivariate case considering wavelet bases from a multivariate MRA. The multivariate case appears, first and naturally when dealing with images. Nason and Silverman [1994] showed how wavelet estimators can be constructed in such cases. Performances and optimality results appeared first in Tribouley [1995] for multivariate Besov classes for the density estimation problem. Delyon and Juditsky [1996] generalized the results for nonparametric estimation. In all these works, authors used non-separable MRA constructions and considered classical (isotropic) Besov spaces where the regularity is the same along all variables. However, Neumann and von Sachs [1997] and Neumann [2000] showed that, when the unknown belongs to anisotropic Besov spaces, hyperbolic wavelets, which are construted upon a a separable MRA construction achieves optimal *rates of convergence*. Recently, Autin et al. [2015] showed that non-separable wavelet cannot achieve optimal performances when the unknown has an anisotropic regularity. They used, the more recent technique of *maxiset* unstead of the *minimax* introduced by Cohen et al. [2001]. This technique considers a given *rates of convergence* and seeks for the largest class on which this rate is obtained. Anisotropy wavelet processing is at the heart of the present work. Part I consider functions in anisotropic two-dimensional spaces while in Part II, we study the potential of considering isotropy and anisotropy and high dimensions.

Bibliography

Edward H Adelson, Eero Simoncelli, and Rajesh Hingorani. Orthogonal pyramid transforms for image coding. In *1987 Cambridge Symposium*, pages 50–58. International Society for Optics and Photonics, 1987.

F. Autin, G. Claeskens, and J.M. Freyermuth. Asymptotic performance of projection estimators in standard and hyperbolic wavelet bases. *To appear in Electronic journal of statistics*, 2015.

R. Balian. Un principe d'incertitude fort en théorie du signal ou en mécanique quantique. In *C.R. Acad.Sci.Paris*, volume 292, pages 1357–1362, 1981.

OV Besov. On a certain family of functional spaces, imbedding and continuation. In *Dokl. Akad. Nauk SSSR*, volume 126, pages 1163–1165, 1956.

J. Bourgain. A remark on the uncertainty principle for hilbertian basis. *J. Funct. Anal.*, 79:136–143, 1988.

J. Bretagnolle and C. Huber. Estimation des densités: risque minimax. 1979.

H Brezis. Analyse fonctionnelle, théorie et applications. *Mason Paris*, 1983.

T Tony Cai and Lie Wang. Adaptive variance function estimation in heteroscedastic nonparametric regression. *The Annals of Statistics*, 36(5):2025–2054, 2008.

Emmanuel J Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *Information Theory, IEEE Transactions on*, 52 (2):489–509, 2006.

Albert Cohen, Ingrid Daubechies, and J-C Feauveau. Biorthogonal bases of compactly supported wavelets. *Communications on pure and applied mathematics*, 45(5):485–560, 1992.

Albert Cohen, Ronald DeVore, Gerard Kerkyacharian, and Dominique Picard. Maximal spaces with given rate of convergence for thresholding algorithms. *Applied and Computational Harmonic Analysis*, 11(2): 167–191, 2001.

J.W. Cooley and J.W. Tukey. An algorithm for the machine computation of complex fourier series. *Math. Comp.*, 19:297–301, 1965.

Ingrid Daubechies. *Ten lectures on wavelets*, volume 61. SIAM, 1992.

B. Delyon and A. Juditsky. On minimax wavelet estimators. *Journal of Applied and Computational Harmonic Analysis*, 3:215–228, 1996.

R.A. DeVore, B. Jawerth, and B.J. Lucier. Image compression through wavelet transform coding. *Information Theory, IEEE Transactions on*, 38(2):719–746, 1992a.

RA DeVore, SV Konyagin, and VN Temlyakov. Hyperbolic wavelet approximation. *Constructive Approximation*, 14(1):1–26, 1998.

Ronald A DeVore. Nonlinear approximation. *Acta numerica*, 7:51–150, 1998.

Ronald A DeVore and Vasil A Popov. Interpolation of besov spaces. *Transactions of the American Mathematical Society*, 305(1):397–414, 1988.

Ronald A DeVore, Björn Jawerth, and Vasil Popov. Compression of wavelet decompositions. *American Journal of Mathematics*, 114(4):737–785, 1992b.

David L. Donoho and Iain M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81: 425–455, 1994.

David L Donoho and Iain M Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the american statistical association*, 90(432):1200–1224, 1995.

David L Donoho, Iain M Johnstone, Gérard Kerkyacharian, and Dominique Picard. Wavelet shrinkage: asymptopia? *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 301–369, 1995.

David L Donoho, Iain M Johnstone, Gérard Kerkyacharian, and Dominique Picard. Density estimation by wavelet thresholding. *The Annals of Statistics*, pages 508–539, 1996.

Paul Doukhan and J León. Déviation quadratique d'estimateurs de densité par projections orthogonales. *Comptes Rendus Acad. Sci. Paris,(A)*, 310:425–430, 1990.

Sam Efromovich and Mark Pinsker. Sharp-optimal and adaptive estimation for heteroscedastic nonparametric regression. *Statistica Sinica*, pages 925–942, 1996.

RH Farrell. On the best obtainable asymptotic rates of convergence in estimation of a density function at a point. *The Annals of Mathematical Statistics*, pages 170–180, 1972.

P. Fryzlewicz. Data-driven wavelet-fisz methodology for nonparametric function estimation. *Electronic journal of statistics*, 2:863–896, 2008.

Piotr Fryzlewicz, Guy P Nason, and Rainer Von Sachs. A wavelet-fisz approach to spectrum estimation. *Journal of time series analysis*, 29(5):868–880, 2008.

P.Z. Fryzlewicz and G.P. Nason. A haar-fisz algorithm for poisson intensity estimation. *Journal of Computational and Graphical Statistics*, 13 (3):621–638, 2004.

D. Gabor. Theory of communication. *Journal of IEE*, 93:429–457, 1946.

Alexander Grossmann and Jean Morlet. Decomposition of hardy functions into square integrable wavelets of constant shape. *SIAM journal on mathematical analysis*, 15(4):723–736, 1984.

Alfred Haar. Zur theorie der orthogonalen funktionensysteme. *Mathematische Annalen*, 69(3):331–371, 1910.

Wolfgang Härdle, Gerard Kerkyacharian, Dominique Picard, and Alexander Tsybakov. *Wavelets, approximation, and statistical applications*, volume 129. Springer Science & Business Media, 2012.

I. Johnstone and B. W. Silverman. Wavelet methods for data with correlated noise. *Journal of the Royal Statistical Society, Series B*, 59:319–351, 1997.

Gérard Kerkyacharian and Dominique Picard. Density estimation in besov spaces. *Statistics & Probability Letters*, 13(1):15–24, 1992.

Gérard Kerkyacharian and Dominique Picard. Density estimation by kernel and wavelets methods: optimality of besov spaces. *Statistics & Probability Letters*, 18(4):327–336, 1993.

S. G. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 11(7):674–693, Jul 1989.

Stephane Mallat. *A wavelet tour of signal processing: the sparse way*. Academic press, 2008.

Y Meyer. Oscillating paterns in image processing and in some nonlinear evolution equation, AMS, Boston, MA, USA, The Fifteenth Dean Jacqueline B. *Lewis Memorial Lectures*, 2001.

Yves Meyer. Principe d'incertitude, bases hilbertiennes et algebres d'operateurs. *Séminaire Bourbaki*, 28: 209–223, 1985.

J. Morlet. In *Proc. 51st Annual International Meeting of the Society of Exploration Geophysicists, Los Angeles*, 1981.

G. Nason and B. W. Silverman. The discrete wavelet transform in s. *Journal of Computational and Graphical Statistics*, 3:163–191, 1994.

A.S. Nemirovskii. Nonparametric estimation of smooth regression functions. *Soviet J. of Computer and Systems Sciences*, 23:1–11, 1985.

A.S. Nemirovskii, B.T. Polyak, and A.B. Tsybakov. Rate of convergence of nonparametric estimators of maximum-likelihood type. *Problems of Information Transmission*, 21:258–272, 1985.

M. Neumann. Spectral density estimation via nonlinear wavelet methods for stationary non-gaussian time series. *Journal of Time Series Analysis*, 17:601–633, 1996.

M. Neumann and R. von Sachs. Wavelet thresholding: beyond the gaussian iid situation. *in Antoniadis & Oppenheim, Wavelets and statistics. Lecture Notes in Statistics*, pages 301–329, 1995.

M. Neumann and R. von Sachs. Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra. *Annals of Statistics*, 25:38–76, 1997.

Michael H. Neumann. Multivariate wavelet thresholding in anisotropic function spaces. *Statistica Sinica*, 10: 399–431, 2000.

Norbert Remenyi, Orietta Nicolis, Guy Nason, and Brani Vidakovic. Image denoising with 2d scale-mixing complex wavelet transforms. *Image Processing, IEEE Transactions on*, 23(12):5165–5174, 2014.

L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

J Strömberg. A modified franklin system and higher-order systems of r" as unconditional bases for hardy spaces. In *Conf. in Harmonic Analysis in honor of A. Zygmund, Wadsworth Math. Series*, volume 2, pages 475–493, 1983.

K. Tribouley. Practical estimation of multivariate densities using wavelet methods. *Statistica Neerlandica*, 49:41–62, 1995.

Hans Triebel. Wavelet bases in anisotropic function spaces. In *Function Spaces, Differential Operators and Nonlinear Analysis*, volume 18, page 529–550, 2004.

Abraham Wald. Generalization of a theorem by v. neumann concerning zero sum two person games. *Annals of Mathematics*, pages 281–286, 1945a.

Abraham Wald. Statistical decision functions which minimize the maximum risk. *Annals of Mathematics*, pages 265–280, 1945b.

Abraham Wald. Foundations of a general theory of sequential decision functions. *Econometrica, Journal of the Econometric Society*, pages 279–313, 1947.

Abraham Wald. Statistical decision functions. *Ann. Math. Statist.*, 20(2):165–205, 06 1949. doi: 10.1214/aoms/1177730030. URL `http://dx.doi.org/10.1214/aoms/1177730030`.

Vyacheslav Zavadsky. Image approximation by rectangular wavelet transform. *Journal of Mathematical Imaging and Vision*, 27(2):129–138, 2007.

CHAPTER 3

Total Variation & Elements of convex optimization

Abstract

In this chapter, we present an overview on total variation and its use in signal and image processing. We also discuss some basic elements on convex optimization related to total variation minimization. The topics addressed in this chapter are going to be useful in chapter 10

3.1 Introduction

Though the wavelet framework presented in the previous chapter had a major impact on signal and image denoising, the assumptions on the observation remains relatively restricted. In particular, the data is supposed to be corrupted only by noise and this noise should be (asymptotically) Gaussian. Total Variation (TV) regularization, introduced by Rudin et al. [2003], provides a rather flexible tool for recovery tasks. Similarly to wavelets techniques, it is also efficient in preserving structural details. The general procedure aims at minimizing a functional combining a TV term and a data-fidelity term which tends at bounding the error due to noise without a necessary knowledge of its statistics. The solution to this problem is not straightforward because of the non-differentiability of the TV semi-norm. This deceptive complication arises in all minimization problems evolving a ℓ_1 regularization which provoked a big fuss in the signal and image community yielding to a consequent number of works on the subject. We start by describing ℓ_1-regularization and related optimization algorithms, then we address total variation. We note that our review here will not exhaustive neither be theory-oriented as we only aim at using this tools for a practical problem related to functional Magnetic resonance imaging (fMRI) deconvolution.

3.2 Sparsity promoting ℓ_1-regularization

In the sparsity priors literature, the ℓ_1 norm appeared intuitively as a convex relaxation of the ℓ_0 quasi-norm which counts the number of nonzero entries of its argument. Suppose that we are are observing a signal

$$g = f + \varepsilon, \tag{3.1}$$

and that f has a sparse representation under the action of an operator Φ. Reconstructing f can be achieved by imposing sparsity within a distance δ from the observation. This gives the following problem[1]

$$\min||\Phi\, f||_0 \text{ subject to } ||g - f||_2 \leq \delta. \tag{3.3}$$

[1] This is called an *analysis* prior. One could also consider a *synthesis* prior in which the argument inside the norm is directly the sparse vector $w = \Phi\, f$

$$\min||w||_0 \text{ subject to } ||g - \Phi^{-1}w||_2 \leq \delta, \tag{3.2}$$

where Φ^{-1} is the inverse of Φ.

Finding a solution to (3.3) is combinatorially complex . In fact, it requires testing all possibilities for the vector Φf. To overcome this problem the ℓ_0 quasi-norm is replaced by the ℓ_1 norm which favors large coefficients

$$\min||\Phi f||_1 \text{ subject to } ||g - f||_2 \leq \delta. \tag{3.4}$$

The advantage of this formulation comes from the nature of the ℓ_1 norm which is the closest convex function to the ℓ_0 quasi-norm. This means that this problem can be solved using classical tools from convex programming. Further, it is possible to consider this problem in a variational way by combining the norm and the constraints

$$\min_f \lambda||\Phi f||_1 + \frac{1}{2}||g - f||_2, \tag{3.5}$$

where the parameter λ aims at controlling the trade-off between the sparsity of the term Φf and the deviation from the observation g.

3.3 Related optimization algorithms

Minimization problems of the form (3.5) rises two challenges. First, the presence of two terms and second the non-differentiability of the regularization term. The differentiabily issue is addressed using *proximity operator* of Moreau [1965].

3.3.1 Proximity operators

Definition 3.3.1. *Let φ be a lower semicontinious convex function. The proximity operator $\text{prox}_{\varphi(\mathbf{x})}$ is defined as*

$$prox_{\varphi(\mathbf{x})}(y) = \arg\min_y \left\{||x - y||_2^2 + \gamma\varphi(y)\right\}.$$

One of the keys behind the success of ℓ_1-regularization is the simplicity of the *proximity operator* of the ℓ_1-norm given by

$$prox_{\gamma|.|_1}(\mathbf{x}) = \arg\min_y \left\{||x - y||_2^2 + \gamma|y|_1\right\}.$$

with $\gamma \in \mathbb{R}$. An easy computation shows that this operator coincides with *soft-thresholding operator* $\mathcal{T}_\gamma$ defined by its point-wise components

$$\mathcal{T}_\gamma(\mathbf{x})_k = \text{sgn}(x_k)\max\{|x_k| - \gamma, 0\}$$

with $x = \{x_1, \cdots, x_n\}$ and $y = \{y_1, \cdots, y_n\}$. This operator can be used to solve (3.5)for a large class of operators Ψ using *proximal splitting* algorithms. These algorithms include, for instance, the forward-backward algorithm [Mercier, 1979, Lions and Mercier, 1979, Chen and Rockafellar, 1997, Combettes and Wajs, 2005], the Douglas-Rachford [Eckstein and Bertsekas, 1992] scheme or the generic scheme of Condat [2013] which brings together many *proximal splitting* algorithms as particular cases. We, first, describe in the sequel the classical Forward-Backward algorithm, then its generalized version with an arbitrary number of regularization terms.

3.3.2 The Forward-Backward algorithm

Consider the following problem

$$\min_{\mathbf{x}} F(\mathbf{x}) + G(\mathbf{x}), \tag{3.6}$$

where F has a Lipschitz continuous gradient and G is simple[2] The forward-backward splitting algorithm reads

Algorithm 1 Forward-Backward algorithm for solving (3.6)

Input: (1) Corrupted data $\mathbf{y}$, (2) a gradient descent stepsize $\mu = 1/L$ where L is the Lipschitz constant of F and (3) update constant τ
Output: Estimate $\widetilde{\mathbf{x}}$
1: $\widetilde{\mathbf{x}}^{(0)} = \mathbf{y}$; $\mathbf{r}^{(1)} = \tilde{\mathbf{x}}^{(0)}$
2: **for** $k = 1 : k_{max}$ **do**
3: $\quad \mathbf{r}^{(k+1)} = prox_{\mu G}(\widetilde{\mathbf{x}}^{(k)} - \mu \nabla F(\widetilde{\mathbf{x}}^{(k)}))$
4: $\quad \widetilde{\mathbf{x}}^{(k+1)} = \widetilde{\mathbf{x}}^{(k)} + \tau(\mathbf{r}_{(k+1)} - \widetilde{\mathbf{x}}^{(k)})$
5: **end for**

3.3.3 The Generalized Forward-Backward algorithm

Now, we consider a similar problem to (3.6) which involve several simple functions $\{G\}_{i=1}^{n}$

$$\min_{\mathbf{x}} F(\mathbf{x}) + \sum_{i=1}^{n} G_i(\mathbf{x}), \tag{3.7}$$

The generalized forward-backward splitting algorithm for solving this problem was introduced by Raguet et al. [2013]. It computes a weighted average of the functions $\{\mathbf{x}_i\}_{i=1}^{n}$ verifying

$$\min_{\mathbf{x}} F(\mathbf{x}) + G_i(\mathbf{x}), \tag{3.8}$$

The algorithm is given as follows

Algorithm 2 Generalized Forward-Backward algorithm for solving (3.8)

Input: (1) Corrupted data $\mathbf{y}$, (2) a gradient descent stepsize $\mu = 1/L$ where L is the Lipschitz constant of F, (3) update constant τ and (4) $\omega_i \in [0, 1]$ with $\sum_{i=1}^{n} \omega_i = 1$.
Output: Estimate $\widetilde{\mathbf{x}}$
1: $\widetilde{\mathbf{x}}^{(0)} = \mathbf{y}$; $\mathbf{r}_i^{(1)} = \tilde{\mathbf{x}}^{(0)}$
2: **for** $k = 1 : k_{max}$ **do**
3: $\quad$ **for** $i = 1 : n$ **do**
4: $\quad\quad \mathbf{r}_i^{(k+1)} = \mathbf{r}_i^{(k)} + \tau\Big(prox_{\frac{\mu}{\omega_i} G_i}(2\widetilde{\mathbf{x}}^{(k)} - \mathbf{r}_i^{(k)} - \mu \nabla F(\widetilde{\mathbf{x}}^{(k)}) - \widetilde{\mathbf{x}}^{(k)})\Big)$
5: $\quad\quad \widetilde{\mathbf{x}}^{(k+1)} = \sum_{i=1}^{n} \omega_i \mathbf{r}_i^{(k+1)}$
6: $\quad$ **end for**
7: **end for**

3.4 Total Variation

The classical TV of a one-dimensional continuous[3] function $s(t)$ defined on an interval $[a, b]$, is given as the L_1-norm of its continuous derivative $D\{s\}$

$$\int_a^b |D\{s\}(t)| dt. \tag{3.9}$$

Its discrete counterpart is defined through the finite difference operator Δ

[2] Its *proximity operator* is easy to compute.
[3] Or more generally functions in Sobolev spaces H^1.

$$TV\,(s) = \sum_{t\in\mathbb{Z}} |\Delta\{s\}[t]| = \sum_{t=1}^{n} |s[t] - s[t-1]| \tag{3.10}$$

TV-regularization is performed via the following ℓ_1-minimization problem

$$\widehat{f} = \arg\min_{f} \left\{ ||g - f||_2^2 + \lambda\, TV(f) \right\}. \tag{3.11}$$

The forward-backward algorithm, in this case, lies on two steps which are a gradient decent for data-fidelity and a proximal operator for TV correction. Its accelerated version consists in updating the relaxation parameter for a faster convergence. In part III, we will use an accelerated version of this algorithm in the spirit of the fast iterative soft thresholding algorithm (FISTA) of Beck and Teboulle [2009]. It requires to set a gradient descent step $\mu < \dfrac{1}{||\Delta^T||_2^2}$, where Δ^T is the discrete divergence operator.

Algorithm 3 FISTA for TV

Input: $g, \lambda, \mu,\ k_{max},\ f^{(1)},\ r_1$
Output: Estimate $\widehat{f}$

$\tilde{f}^{(0)} = \tilde{f}^{(1)} = f^{(1)}$; $h^{(1)} = f^{(1)}$
for $k = 1 : k_{max}$ **do**
  $\tilde{f}^{(k)} = \tilde{f}^{(k)} - \mu\Delta\left\{g - \Delta^T\left\{h^{(k)}\right\}\right\}$
  $f^{(k+1)} = \tilde{f}^{(k)} - \mathcal{T}_\lambda\left(\tilde{f}^{(k)}\right)$
  $r_{k+1} = \dfrac{1 + \sqrt{1 + 4r_k^2}}{2}$
  $h^{(k+1)} = \tilde{f}^{(k)} + \dfrac{r_k - 1}{r_{k+1}}\left(\tilde{f}^{(k)} - \tilde{f}^{(k-1)}\right)$
end for
$\widehat{f} = g - \Delta^T\left\{f^{(k+1)}\right\}$

TV had a large success in image processing because it imposes small bounded-variation; reducing the magnitude of the finite differences imposes regularity on the variations of the image. This is a vaguely accepted prior for images. One-dimensional signals, on the other hand, often show richer variations. A particular situation in which a signal can be recovered accurately using TV is when it is composed mainly of blocks; i.e Δf is mainly *spikes*. This is, for example, the case for the neural activation. However, in the case of fMRI time courses, one is observing signals which are convolved versions of the neural activation. This problem motivated the introduction of the generalized TV by Karahanoglu et al. [2011] which is at the heart of the work presented in Part III.

Bibliography

Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009.

George HG Chen and R Tyrrell Rockafellar. Convergence rates in forward–backward splitting. *SIAM Journal on Optimization*, 7(2):421–444, 1997.

Patrick L Combettes and Valérie R Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Modeling & Simulation*, 4(4):1168–1200, 2005.

Laurent Condat. A primal–dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms. *Journal of Optimization Theory and Applications*, 158(2):460–479, 2013.

Jonathan Eckstein and Dimitri P Bertsekas. On the douglas—rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55(1-3):293–318, 1992.

Fikret Işık Karahanoglu, İlker Bayram, and Dimitri Van De Ville. A signal processing approach to generalized 1-d total variation. *IEEE Transactions on Signal Processing*, 59(11):5265–5274, 2011.

Pierre-Louis Lions and Bertrand Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM Journal on Numerical Analysis*, 16(6):964–979, 1979.

Bertrand Mercier. *Topics in finite element solution of elliptic problems*. Springer, 1979.

Jean-Jacques Moreau. Proximité et dualité dans un espace hilbertien. *Bulletin de la Société mathématique de France*, 93:273–299, 1965.

Hugo Raguet, Jalal Fadili, and Gabriel Peyré. A generalized forward-backward splitting. *SIAM Journal on Imaging Sciences*, 6(3):1199–1226, 2013.

L. Rudin, P. L. Lions, and S. Osher. *Multiplicative denoising and deblurring: theory and algorithms*. Geometric Level Set Methods in Imaging, Vision, and Graphics. S. Osher and N. Paragios, Eds., 2003.

Part I

Generalized Hyperbolic Crossing

CHAPTER 4

Variable Groupwise Structured Wavelets

Abstract

A generalized hyperbolic cross construction of wavelet atoms is presented. First, we show the limits of the standard tensor-product (hyperbolic) construction in image denoising. Then, we motivate the use of such constructions in cases when the regularities along variables are different. This led us to a generalized hyperbolic construction which allows grouping variables with respect to regularity features while performing non-parametric estimation. For thresholding estimators, we show, under usual assumptions on wavelet functions, that the L^2 -loss falls to the dimension of the group of variables with the highest dimension. Moreover, we show that other features besides regularity can be taken thanks to this construction. In particular, divergence-freedom and variance stabilization can be imposed on groups of variables.

4.1 Introduction and Motivation

We consider the problem of recovering an unknown function from a noisy observation. This task is common in various problems related to image processing. We are interested in the following *Additive White Gaussian noise* (AWGN) model

$$f_\varepsilon = f + \varepsilon\xi \tag{4.1}$$

where f_ε is the observed data, $\varepsilon \in (0, \infty)$ the noise level and $\xi \sim \mathcal{N}(0, 1)$ is a white noise. The goal is to recover f from its noisy observation f_ε. We are interested in cases where the multidimensional function f depends on variables with different physical meaning along the different coordinate axes (e.g spatial, temporal, spectral, hyperspectral,...).

Non parametric methods for denoising based on wavelets expansions of the function f_ε have been widely developed in the two last decades since the seminal work of Donoho and Johnstone [Donoho and Johnstone, 1994] defining the celebrated wavelet shrinkage procedure (*cf.* chapter 2). Thereafter, these wavelet denoising techniques have been extended to deal with images or volumes of images. In such situations, variables live in the same (physical) space. Thus, it is natural to use the standard (isotropic) multidimensional wavelet bases. However, in many applications one is confronted to data in which the variables have different natures and hence, the signal or image of interest is likely to have different properties according to these different variables. We give here some examples:

Spectral, Multispectral and Hyperspectral data A first example is the evolutionary spectra of non-stationary processes which is a bidimensional function, of frequency and time. In this estimation context emerges the theory of anisotropic function estimation based on the thresholding of hyperbolic wavelet

coefficients of the preperiodogram [Neumann and Von Sachs, 1997]. A similar phenomenon can be observed in multispectral and hyperspectral imaging; The 3D stack of images have completely different regularities in spectral and spatial variables as the data consists of the same image with different spectral bands and wavelengths [Chang, 2003, Shaw and Burke, 2003].

Image and volume sequences An image or volume temporal sequence in which voxel intensities are preserved over time is governed by the following conservation equation

$$\partial_t I + \mathbf{u}.\nabla_{\mathbf{x}} I = 0 \,, \tag{4.2}$$

where I refers to the sequence and $\mathbf{u}$ is the velocity field of the voxels. When the unknown is the velocity $\mathbf{u}$, Equation (4.2) is the so-called optical flow equation and plays a major role in computer vision since it is the reference model for motion estimation [Horn and Schunck, 1981]. From a PDE's point of view this equation can be seen as a simple advection equation on the scalar field I assuming that the flow is incompressible. Solutions of such time–dependent partial differential equations have different degrees of smoothness in space and time and they are not well characterized in isotropic spaces [DeVore et al., 2008].

Velocity fields Now, consider that the velocity field $\mathbf{u}$ is given. This situation occurs for example in blood flow measurement provided by Phase Contrast MRI [Markl et al., 2003] or Ultrasound Doppler [Garcia et al., 2010]. The output $\mathbf{u}$ is computed from phase transitions. It usually suffers from low velocity–to–noise ratio (VNR) and a denoising step is often needed. The velocity field $\mathbf{u}$ can also be computed solving some PDE modeling the underlying motion. For example, a blood flow verifies variants of Navier-Stokes equations [Quarteroni et al., 2002]. In this case, the difference between the temporal and the spatial directions is related to their physical nature in the sense that the divergence over spatial dimensions is null.

Partial Data-dependent noise intensity In many applications, the observed data cannot be properly modelled by the equation (4.1). In particular the noise intensity can be proportional to the underlying unknown function. As a consequence wavelet thresholding methods need to be adapted via variance stabilization techniques. Such data–dependent noises models are often purely related to spatial domain because of the imaging systems. Hereafter the variance stabilization technique will be only applied on the spatial variables.

In this chapter we describe a novel construction of a so–called structured wavelet basis that can faithfully represent multivariate functions presenting different characteristics along the different coordinate axis. Our construction extend the so–called hyperbolic wavelet basis introduced in [DeVore et al., 1998]. It has been shown that the hyperbolic wavelet basis is well adapted for functions having different degrees of smoothness along the directions [Neumann, 2000]. In particular, interesting recent results [Autin et al., 2015] brought to light that classical wavelet estimators cannot achieve optimal reconstructions of functions with different regularities on the different dimensions. We present a *structured wavelets* construction which consider sparsity assumptions on groups of variables and not only on single variables using a generalization of the hyperbolic construction of wavelets. We describe a relevant type of functional classes which are described by these structured wavelets basis, and we demonstrate that the minimax lower bound of the L^2-loss of the estimator is driven by the dimension of the largest group, breaking, "*partially*", the curse of dimensionality. We show that this generalization do not only allow to take regularity features into account but also physical characteristic such as divergence-freedom. We also show that the obtained construction can handle situations in which the noise characteristics are data-dependent but only on a part of the variables.

There is no reason for the standard multidimensional wavelet basis to be systematically outperformed by the hyperbolic wavelet basis. If the theoretical justification does not exists yet, from an empirical point of view we will explain the advantages of the primer.

The rest of the chapter is organized as follows. In Section 4.2, we give a reminder on wavelets and their extensions in multivariate cases. The proposed wavelet construction is exposed and motivated in Section 4.3,

along with a derivation of the lower bound of the L^2-loss for a particular functional space. Extensive numerical experiments and comparisons are presented in the next chapter.

4.2 Limits of usual wavelet estimation procedures

First, for the sake of clarity and ease of reading, we recall some definitions from chapter 2. We begin by introducing one–dimensional wavelet bases, then we describe two different extensions in higher dimensions.

One dimensional wavelet bases are defined from a one-dimensional function ψ, called mother wavelet and its dilated and translated versions $\psi_{j,k}(.) = 2^{j/2}\psi(2^j. - k)$ with $(j,k) \in \mathbb{N} \times \mathbb{Z}$ and a scaling function φ defined with its dilated and translated versions $\varphi_{j,k}(.) = 2^{j/2}\varphi(2^j. - k)$. It is then well–known that $\{\varphi_{0,k}\}_{k\in\mathbb{Z}} \cup \{\psi_{j,k}\}_{j\leq 0, k\in\mathbb{Z}}$ forms an orthogonal basis of $L^2(\mathbb{R})$. In the sequel we shall denote $\psi_{-1,k} = \varphi_{0,k}$.

In the multivariate setting, one defines multidimensional wavelets as

$$\psi_{j_1,\cdots,j_d,k_1,\cdots,k_d}(\mathbf{x}) = \psi_{j_1,k_1}(x_1) \otimes \cdots \otimes \psi_{j_d,k_d}(x_d). \tag{4.3}$$

Two possible multidimensional extensions of wavelet bases come from this definition. The first one consists in taking all possible combinations of the multiindices $\mathbf{j} = (j_1, \cdots, j_d) \in (\mathbb{N} \cup \{-1\})^d$, $\mathbf{k} = (k_1, \cdots, k_d) \in \mathbb{Z}^d$ leading to the so–called *"hyperbolic"*[1] wavelet basis [DeVore et al., 1998] $\mathcal{B}_{hyp,d} = \{\psi_{\mathbf{j},\mathbf{k}}\}_{\mathbf{j}\in(\mathbb{N}\cup\{-1\})^d,\mathbf{k}\in\mathbb{Z}^d}$. The second possibility is to fix the directional dilation indices to be the same along each dimension $j = j_1 = \cdots = j_d$. Then one first define multidimensional wavelets $\psi^{(i)}$ for any $i \in \{0,1\}^d \setminus \{(0,\cdots,0)\}$ as

$$\psi^{(i)} = \psi^{(i_1)} \otimes \cdots \otimes \psi^{(i_d)} \text{ with } \psi^{(0)} = \phi,\ \psi^{(1)} = \psi \tag{4.4}$$

and for any $j \geq -1$, $\mathbf{k} \in \mathbb{Z}^d$ and $i \in \{0,1\}^d \setminus \{(0,\cdots,0)\}$, one set

$$\Psi^{(i)}_{j,\mathbf{k}}(x) = 2^{jd/2}\psi^{(i)}(2^j x - k)$$

We also set for any $\mathbf{k} \in \mathbb{Z}^d$, $\Psi^{(0,\cdots,0)}_{-1,k}(x) = [\phi \otimes \cdots \otimes \phi](x - \mathbf{k})$. The family $\mathcal{B}_{iso,d} = \{\Psi^{(i)}_{\mathbf{k}}, j, \mathbf{k}\}$ is then an orthonormal wavelet basis of $L^2(\mathbf{R}^d)$, which is said to be *"isotropic"* [Daubechies, 1992]. See Figure 4.1 for a comparison of hyperbolic and classical wavelet decompositions on a two-dimensional example.

In the sequel, we denote as I the indices of a particular wavelet or scaling function. The set of all combinations of I will be referred as $\mathcal{I}$. If not specified $\mathcal{I}$ can refer both to standard or hyperbolic constructions. In the hyperbolic case one has

$$\mathcal{I} = \{I = (\mathbf{j},\mathbf{k}) \text{ with } \mathbf{j} = (j_1, \cdots, j_d) \in (\mathbb{N} \cup \{-1\})^d,\ \mathbf{k} = (k_1, \cdots, k_d) \in \mathbb{Z}^d\}$$

whereas in the isotropic case

$$\mathcal{I} = \{I = (\mathbf{i}, j, \mathbf{k}) \text{ with } i \in \{0,1\}^d \setminus \{(0,\cdots,0)\}, j \in \mathbb{N} \cup \{-1\},\ \mathbf{k} = (k_1, \cdots, k_d) \in \mathbb{Z}^d\}$$
$$\cup \{I = ((0,\cdots,0), -1, \mathbf{k})\mathbf{k} = (k_1, \cdots, k_d) \in \mathbb{Z}^d\}$$

The wavelet decomposition of a function $f \in L^2([0,1]^d)$ is then given by

$$f = \sum_{I\in\mathcal{I}} \beta_I(f)\Psi_I, \tag{4.5}$$

where

$$\beta_I(f) = \int_{(0,1)^d} f(x)\Psi_I(x)\,dx. \tag{4.6}$$

[1] This wavelets appears under other names names in the literature, such as mixing scales [Remenyi et al., 2014] or rectangular [Zavadsky, 2007].

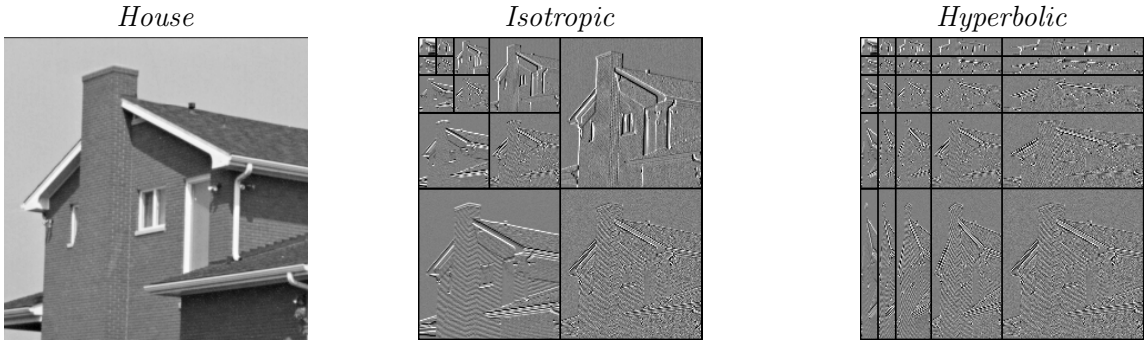


Figure 4.1: Wavelets decomposition in isotropic and hyperbolic settings

Denoising methods based on thresholding the wavelet coefficients has proven its effectiveness since the seminal work of Donoho and Johnstone [Donoho and Johnstone, 1994]. Under model (4.1), it is possible to recover the unknown function f from its noisy observation f_ε by different procedures on the wavelet empirical wavelet coefficients $\beta_I(f_\varepsilon)$. This results in the following estimator

$$\hat{f} = \sum_{I \in \mathcal{I}_\varepsilon} \beta_I(f_\varepsilon)\Psi_I,$$

where $\mathcal{I}_\varepsilon \subset \mathcal{I}$ is to the set of multiindices corresponding to the coefficients that are kept in the reconstruction. The choice of $\mathcal{I}_\varepsilon$ can be performed following different rules. In the sequel we shall consider the most classical one, the so called hard thresholding rule where

$$\mathcal{I}_\varepsilon = \{I \in \mathcal{I} \text{ s.t } |\mathbf{j}| \leq j(t_\varepsilon) \text{ and } |\beta_I(f_\varepsilon)| > t_\varepsilon\},$$

where t_ε is a threshold calibrated in function of the level of noise ε : $t_\varepsilon = 4\sqrt{2}\varepsilon\sqrt{\log \varepsilon^{-1}}$. The integer $j(t_\varepsilon)$ is chosen such that $2^{-j(t_\varepsilon)} \leq t_\varepsilon^2 < 2^{1-t_\varepsilon}$. The notation $|\mathbf{j}|$ allow to encompass both the hyperbolic and the isotropic case in the definition of the hard thresholding estimator, since we set $|\mathbf{j}| = j_1 + \cdots + j_d$ in the case of a multidimensional index $\mathbf{j} = (j_1, \cdots, j_d)$ whereas one has $|j| = j$ in the case of an unidimensional scaling index j. Depending on the considered wavelet basis (isotropic or hyperbolic), one has then two possible estimations procedures which can be derived from the hard thresholding rule : the *isotropic* hard thresholding estimator and the *hyperbolic* one.

We want to compare the numerical performances of these two procedures according to the anisotropic nature of the data. From the theoretical point of view, this question has already been addressed and appears in the work of Neumann and Von Sachs [Neumann and Von Sachs, 1997] where the extension of wavelet thresholding techniques to multivariate anisotropic scenarios is introduced. It is shown that one has better performances for hyperbolic wavelets whenever the unknown function has some anisotropic features. Moreover, due to the adaptive nature of the *hard thresholding* and the fact that isotropy is a particular case of anisotropy, it has been proved recently that even if the considered data are isotropic, hyperbolic wavelet thresholding gives the same theoretical guarantees as isotropic wavelets [Autin et al., 2015].

However, empirical evidences contradict this result. We give here a motivational experiment using the sequence *"Ayiko"*. In Figure 4.2, we consider a spatial frame (i.e, we fix time). The denoising experiment uses the *hard thresholding* rule. Isotropic wavelets are slightly better in terms of PSNR. Moreover, undesirable axis-aligned artifacts are visible on the reconstruction from noisy data in the hyperbolic settings. This results are in contradiction with theoretical estimations. A possible explanation for this phenomenon is the biased definition of anisotropic functional spaces which consider only axis-aligned regularities, and the fact that the used image, and natural images, in general do not have strong differences in terms of smoothness along the vertical and the horizontal directions, that is are not strongly anisotropic.

In Figure 4.3, a temporal cross section of the sequence is considered (i.e, we fix one of the spatial dimensions). The resulting $2D$ images (see Figure 4.3) has different regularities in time and space and it then is highly anisotropic. In particular, the resulting image is highly regular in the temporal dimension[2]. In

[2]Of course, this is not a generalization. It is also possible to have some irregularity in the temporal dimension, for example, when new objects are appearing in the scene.

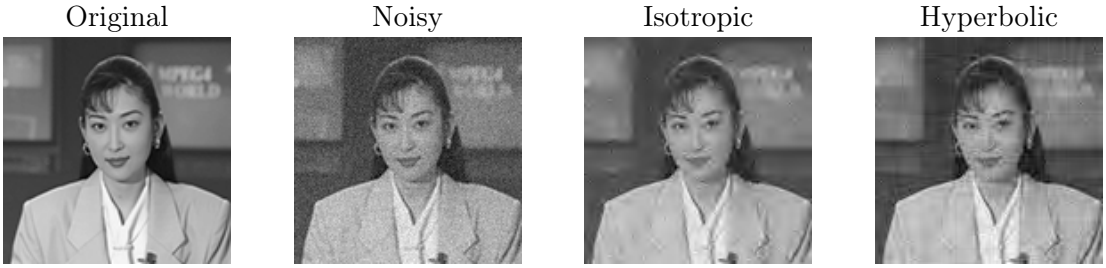


Figure 4.2: Denoising of one frame of the Ayiko sequence.

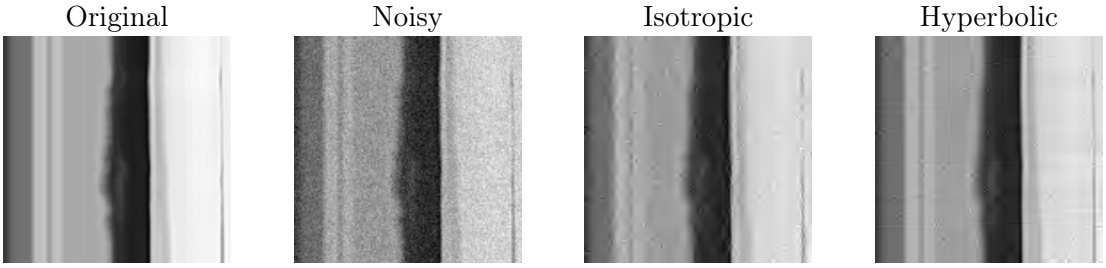


Figure 4.3: Denoising of a temporal cross section of the sequence Ayiko

this case, the hyperbolic wavelets have clearly superior performance. To have a better understanding of this phenomenon, the degree of anisotropy can be seen as maximal when the degree of interaction between the variables is small. This degree is called the atomic dimension. Simple examples where the atomic dimension is one, are additive models. For example in the two dimensional case, these functions are of the form

$$f(x_1, x_2) = f_1(x_1) + f_2(x_2),\ (x_1, x_2) \in \mathbb{R}^2. \tag{4.7}$$

Such functions are known to allow rates of convergence corresponding to the one-dimensional case [Autin et al., 2015]. Figure 4.4 shows an example of denoising under the model (4.7) where f_1 and f_2 are respectively the standard test functions, Doppler and Blocks [Antoniadis et al., 2001]. The results show that the hyperbolic setting outperforms the isotropic setting in this case.

To the best of our knowledge, existing research focuses on cases when the data are fully anisotropic, that is all the regularities are different according each direction. In this work, we investigate the cases when variables can be grouped in sub-ensembles with the same physical meaning. This is motivated by the conclusions of the numerical results in this section and by the fact that functions having a group behavior arises in many applications such as spatio-temporal and multi-spectral data. This naturally suggests the use of wavelet atoms that consider both features. In the next section, we discuss the construction of such wavelets, which are an instance of the so-called composite wavelets [Guo et al., 2004] and the definition of associated estimation procedures.

4.3 Estimation procedures based on tensored wavelet basis

4.3.1 The tensored wavelet basis and associated estimation procedures

We now introduce the structured wavelet basis [Farouj et al., 2015, 2017], which arise as a case of composite wavelets defined in [Guo et al., 2004]. The starting point are N multidimensional wavelet bases $\{\Psi_{I_1,1}, I_1 \in \mathcal{I}_1\}, \cdots, \{\Psi_{I_N,N}, I_N \in \mathcal{I}_N\}$ of $L^2(\mathbb{R}^{d_1}), \cdots, L^2(\mathbb{R}^{d_N})$ respectively. For each $\mathbf{I} = (I_1, \cdots, I_N) \in \mathcal{I}_1 \times \cdots \times \mathcal{I}_N$, one then set

$$\Psi_{\mathbf{I}} = \Psi_{I_1,1} \otimes \cdots \otimes \Psi_{I_N,N}$$

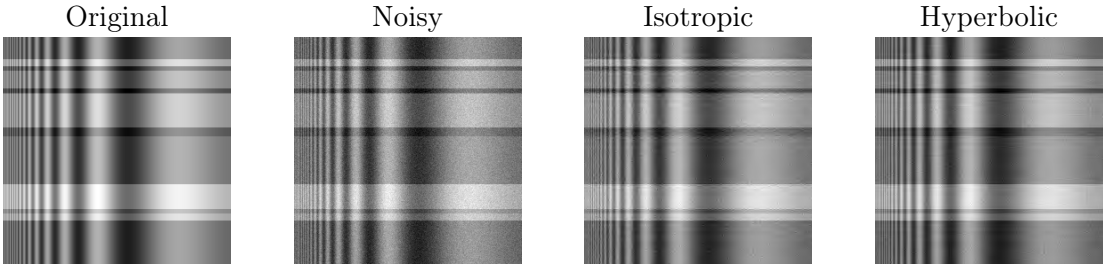


Figure 4.4: Denoising of an additive model.

The family $\{\Psi_{\mathbf{I}}, \mathbf{I} \in \mathcal{I}_1 \times \cdots \times \mathcal{I}_N\}$ is then a basis of $L^2(\mathbb{R}^d)$ with $d = d_1 + \cdots + d_N$. Note that if $N = d$, we recover anhyperbolic wavelet basis as a special case whereas in the case $N = 1$ this construction corresponds to isotropic wavelet basis.

Any function $f \in L^2(\mathbb{R}^d)$ can then be decomposed in

$$f = \sum_{\mathbf{I}} \beta_{\mathbf{I}} \Psi_{\mathbf{I}}$$

One then define the hard thresholding estimator as

$$\widehat{f} = \sum_{\mathbf{I} \in \mathcal{I}_\varepsilon} \beta_{\mathbf{I}}(f)\, \Psi_{\mathbf{I}}$$

with $\mathcal{I}_\varepsilon = \{\mathbf{I} \in \mathcal{I}_1 \times \cdots \times \mathcal{I}_N,\ |\mathbf{j}| = j_1 + \cdots + j_N \leq j(\lambda_\varepsilon)\}$ where $j(\lambda_\varepsilon)$ is such that $2^{-j(\lambda_\varepsilon)} \leq \lambda_\varepsilon^2 < 2^{1-j(\lambda_\varepsilon)}$ with $\lambda_\varepsilon = m\varepsilon\sqrt{\log(\varepsilon^{-1})}$.

The next section aims at proving the optimality of this procedure in the minimax sense.

4.3.2 Minimax results

To state and prove our minimax results about the hard thresholding procedure in the tensored wavelet basis, we first need to introduce some appropriate functional spaces, which appear as approximation spaces for our structured wavelet bases. These spaces extend the space of functions with dominating mixed derivatives introduced in [Schmeisser, 2007] as approximation spaces for the hyperbolic wavelet basis

For a vector $\alpha = (\alpha_1, \cdots \alpha_d) \in \mathbb{N}^d$ and function $f \in L^2([0,1]^d)$, let us define as usual its partial derivatives in the distribution sense

$$D^\alpha f = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}},$$

where $|\alpha| = \alpha_1 + \cdots + \alpha_d$.

Given $\mathbf{R} = (R_1, \cdots, R_N) \in \mathbb{N}^N$, we now define the following functional spaces

$$\mathcal{F}^{\mathbf{R}}_{d,N} = \{f \in L^2((0,1)^d)\, \|f\|_{\mathcal{F}^{\mathbf{R}}_{d,N}} = \sum_{i=1}^{N} \sum_{\mathbf{r}=(\mathbf{r}_1,\cdots,\mathbf{r}_N) \in \mathbb{N}^{d_1} \times \cdots \times \mathbb{N}^{d_N}, |\mathbf{r}_i| \leq R_i} \|D^{\mathbf{r}_1} \cdots D^{\mathbf{r}_N} f\|_{L^2((0,1)^d)} < \infty\}$$

We equip these spaces with the norm

$$\|f\|_{\mathcal{F}^{\mathbf{R}}_{d,N}} = \|f\|_{L^2((0,1)^d)} + \sum_{i=1}^{N} \sum_{\mathbf{r}=(\mathbf{r}_1,\cdots,\mathbf{r}_N) \in \mathbb{N}^{d_1} \times \cdots \times \mathbb{N}^{d_N}, |\mathbf{r}_i| \leq R_i} \|D^{\mathbf{r}_1} \cdots D^{\mathbf{r}_N} f\|_{L^2((0,1)^d)}\ .$$

We then denote $\mathcal{F}^{\mathbf{R}}_{d,N}(K)$ the ball of $\mathcal{F}^{\mathbf{R}}_{d,N}$ of radius K.

The space $\mathcal{F}^{\mathbf{R}}_{d,N}$ contains functions with dominating mixed derivatives which have for any $\ell \in \{1, \cdots, N\}$, d_ℓ variables having the same regularity order R_ℓ. Two special cases of interest arise from this definition:

- If $\forall i = 1, \cdots, N$ we have $\mathbf{r}_\ell \in \mathbb{R}$, that is if $N = d$, we find the so–called spaces with dominating mixed derivatives defined in [Schmeisser, 2007]

$$\mathcal{F}_N^{\mathbf{R}} = \{f \in L^2((0,1)^N) | \sum_{(r_1,\cdots,r_N), r_i \leq R_i} \|D^{r_1} \cdots D^{r_N} f\|_{L^2((0,1)^N)} < \infty\}$$

- If $N = 1$ and $d_1 = d$, we recover the classical isotropic Sobolev spaces.

$$H^R = \{f \in L^2((0,1)^d) | \sum_{r_1 \in \mathbb{N}^d, |r_1| \leq R} \|D^{r_1} f\|_{L^2((0,1)^d)} < \infty\}$$

We can now state our main result giving the rate of convergence of the procedure defined in Section 4.3.1 and proving its optimality in the minimax sense :

Theorem 4.3.1. *Assume that* $R_1 = \cdots = R_N = R$. *Let* J_ε *be such as* $\varepsilon^2 2^{J_\varepsilon d_{max}} J_\varepsilon = 2^{-2J_\varepsilon R}$, *with* $d_{\max} = \max(d_i)$. *Then*

$$\inf_{\widehat{f}} \sup_{f \in \mathcal{F}_R^{d,N}(K)} \mathbb{E}\|\widehat{f} - f\|_2^2 = \left(\varepsilon^{4R/(2R+d_{\max})}\right) \left(|\log(\varepsilon)|^{N-1}\right) .$$

Note that this rate of convergence shows a lower dimension which is different from one in the mimimax estimate and thus breaking "partially" the *curse of dimensionality*. The lower bound, in this case, is of the order of $\varepsilon^{2R/(R+d_{max})}$, while it is $\varepsilon^{2R/(R+d)}$ in the isotropic case and of $\varepsilon^{2R/(R+1)}$ in the fully anisotropic case. It is interesting to note that the order of the logarithmic term is related to the degrees of freedom of the scales vector instead of the usual dimension.

4.4 Extension of the method to other settings

The construction (4.3) allows to consider different families of wavelets along the different dimensions. One could for example consider a representation of a piecewise stationary spectrum with Haar wavelets along the time and a smooth wavelet along the spatial frequency. The hard thresholding can also be modified considering possibly random threshold depending not only on the level of noise but also on the index $\mathbf{I}$. It will be the case when considering stabilisation of variance approaches. Here, we extend the approach introduced in Section 4.3 by pointing out other situations where specific types of wavelets and estimation procedure are needed.

4.4.1 Denoising of spatio temporal incompressible flows

We consider the problem of denoising spatio-temporal vector fields with null divergence in space such as incompressible flows. Denoising such type of data gains a lot of attention with the emergence of Phase contrast MRI imaging [Markl et al., 2003]. A velocity field is a function of $(\mathbf{x}, t)$ where $\mathbf{x}$ is the spatial variable of dimension $d \geq 2$. The wavelet paradigm for such data consists in two types of algorithms based on a decomposing the flow in a divergence free wavelet frame. The classical spatial denoising [Markl et al., 2003] do not consider the time variable and a thresholding is done on the divergence-free wavelet transform of each spatial stack. A more interesting procedure was introduced recently by Bostan *et al.* [Bostan et al., 2013, 2015] which considers spatio-temporal regularization. The algorithm considers two regularization terms in a variational framework

$$\hat{f} = arg\,min\{||W_{\mathcal{B}_{div}} f||_1 + \mathcal{R}^t(f) + \frac{1}{2}||g_\varepsilon - f||_2^2\}, \tag{4.8}$$

where $W_{\mathcal{B}_{div}} f$ denotes the vector of the coefficients resulting from the decomposition of f in free-divergence wavelet basis $\mathcal{B}_{div}$ and $\mathcal{R}^t(f)$ is a term aiming at regularizing f in the temporal dimension on which incompressibility is not maintained. Many choices for $\mathcal{R}^t(f)$ are possible. The presence of two norms makes the minimization process challenging. In particular, the direct use of simple *soft-thresholding* algorithms is not

possible. Moreover, the convexity is lost with the result of the non-uniqueness of a minimum. If $\mathcal{R}^t(f)$ is not differentiable, as in the case of the ℓ_1-norm, the problem needs to be divided into two sub-problems by considering one norm at the time in an iterative fashion.

Here, we propose an alternative approach which consists in using one single norm which sparsifies the complete vector. This can be achieved by the the structured wavelet construction. Let us define the following spatio tempral wavelets :

$$\Psi = \psi^{div}_{\mathbf{x}}(\mathbf{x}) \otimes \psi_t(t) \tag{4.9}$$

where ψ^{div} is a divergence-free spatial wavelet and ψ_t is a one-dimensional temporal wavelet. With our construction, we can replace the estimation procedure (4.8) by a simpler one as a thresholding on the wavelet coefficients of the flow of interest in the structured wavelet basis. We also propose to consider other possible bases. For example, if temporal discontinuities are not allowed, ψ_t can be replaced by a *Fourier* basis: family of complex exponentials $\{e^{int}\}$.

4.4.2 Partial data-dependent noise

Another advantage of the tensor construction given is its ability to deal with the case where the noise dependss on the data. Such situations occurs, for example, in dynamic imaging when the noise has a specific spatial characteristic due to imaging system which is not preserved in time. We are, of course interested in cases where the noise can be removed via known thresholding methods. Consider, for example, the following noise model

$$f_\varepsilon = f + F(f)\xi \tag{4.10}$$

where F is a non-decreasing function and ξ a white noise. Many noise models are of the form (4.10). Examples are speckle noise in Ultrasound imaging [Loupas et al., 1989] and Poisson noise in Photon imaging systems [Nowak and Baraniuk, 1999, Fryzlewicz and Nason, 2004]. When an image sequence is considered, each image of the sequence follows model (4.10). For each image, this model can be solved by wavelet techniques after a *Gaussianizing* process on the wavelet coefficients [Fadili et al., 2003, Fryzlewicz, 2008]. The corresponding transform is called the Wavelet-Fisz transform and consists in stabilizing the variance by dividing wavelet coefficients by local estimations of the noise standard deviation. Now consider the problem of denoising a sequence in which each image follows model (4.10). Using a construction similar to (4.9) comes naturally. Here, instead of a divergence free wavelet transformation, a wavelet-Fisz transformation is considered; the spatio-temporal coefficients are stabilized in the spatial dimension before applying a hard thresholding procedure. We refer to Part II for a full presentation on the wavelet-Fisz methodology. The next chapter is devoted to an exhaustive experimental study on the performance of structured wavelet construction in practice.

4.5 Appendix

4.5.1 Preliminary results on structured wavelets and function spaces

In what follows, we shall need a structured wavelet characterization of the functional spaces $\mathcal{F}^{\mathbf{R}}_N$. We have the following result, which extend Theorem 3.1 of [Schmeisser and Sickel, 2004]

Theorem 4.5.1. *Assume that the structured wavelet basis is sufficiently regular. Then $f \in \mathcal{F}^{\mathbf{R}}_{N,d}$ iff*

$$\|f\|^2_{hyp,\mathcal{F}^{\mathbf{R}}_{N,d}} = \sum_{\mathbf{i}} \sum_{\mathbf{j}} \left(1 + 2^{2(\sum_\ell j_\ell R_\ell)}\right) \sum_{\mathbf{k}\in\mathbb{Z}^N} |\beta_{\mathbf{I}}|^2 < \infty \, .$$

In addition, the two norms $\|\cdot\|_{\mathcal{F}^{\mathbf{R}}_{N,d}}$ and $\|\cdot\|_{hyp,\mathcal{F}^{\mathbf{R}}_{N,d}}$ are equivalent.

These functional spaces can be related in a rather classical way to their weak counterpart. More precisely, let us define the following weak space :

$$W_{N,d}(r) = \{f \in L^2, \sup_{\lambda>0} \lambda^{r-2} \sum_{\mathbf{I}} |\beta_{\mathbf{I}}(f)|^2 1_{|\beta_{\mathbf{I}}(f)|<\lambda} < \infty\} .$$

As in Lemma 2.2 of [Kerkyacharian and Picard, 2000], we can prove that we have an alternative definition of these spaces :

Proposition 4.5.2. *Let $r \in (0,2)$. Then*

$$W_{N,d}(r) = \{f \in L^2, \sup_{\lambda>0} \lambda^{r} \sum_{\mathbf{I}} 1_{|\beta_{\mathbf{I}}(f)|>\lambda} < \infty\} .$$

Proof
Let us first prove that if $f \in W_{N,d}(r)$ then $\sup_{\lambda>0} \lambda^r \sum_{\mathbf{I}} 1_{|\beta_{\mathbf{I}}(f)|>\lambda} < \infty$. Indeed,

$$\begin{aligned} \sum_{\mathbf{I}} 1_{|\beta_{\mathbf{I}}(f)|>\lambda} &= \sum_{\mathbf{I}} \sum_{\ell\geq 0} 1_{2^\ell\lambda\leq|\beta_{\mathbf{I}}(f)|<2^{\ell+1}\lambda} \\ &\leq \sum_{\ell\geq 0} \sum_{\mathbf{I}} (2^\ell\lambda)^{-2} |\beta_{\mathbf{I}}(f)| 1_{|\beta_{\mathbf{I}}(f)|<2^{\ell+1}\lambda} \end{aligned}$$

We now use the assumption that $f \in W_{N,d}(r)$ which implies that

$$\sum_{\mathbf{I}} |\beta_{\mathbf{I}}(f)|^2 1_{|\beta_{\mathbf{I}}(f)|<2^{\ell+1}\lambda} \leq (2^{\ell+1}\lambda)^{2-r}$$

Hence

$$\sum_{\mathbf{I}} 1_{|\beta_{\mathbf{I}}(f)|>\lambda} \leq \sum_{\ell\geq 0} (2^\ell\lambda)^{-2} (2^{\ell+1}\lambda)^{2-r} \leq 2^{2-r}\lambda^{-r} \left[\sum_{\ell\geq 0} 2^{-\ell r}\right]$$

Since $r > 0$, the last sum converges and we get the first inclusion.

Conversely, let us assume that $\sup_{\lambda>0} \lambda^r \sum_{\mathbf{I}} 1_{|\beta_{\mathbf{I}}|>\lambda} < \infty$ and let us prove that $f \in W_{N,d}(r)$. Indeed,

$$\begin{aligned} \sum_{\mathbf{I}} |\beta_{\mathbf{I}}|^2 1_{|\beta_{\mathbf{I}}|<\lambda} &= \sum_{\mathbf{I}} \sum_{\ell\geq 0} |\beta_{\mathbf{I}}(f)|^2 1_{2^{-\ell-1}\lambda<|\beta_{\mathbf{I}}(f)|<2^{-\ell}\lambda} \\ &\leq \sum_{\ell\geq 0} \sum_{\mathbf{I}} (2^{-\ell}\lambda)^2 1_{2^{-\ell-1}\lambda<|\beta_{\mathbf{I}}(f)|<2^{-\ell}\lambda} \end{aligned}$$

We now use the assumption $\sup_{\lambda>0} \lambda^r \sum_{\mathbf{I}} 1_{|\beta_{\mathbf{I}}(f)|>\lambda} < \infty$ and deduce that

$$\begin{aligned} \sum_{\mathbf{I}} |\beta_{\mathbf{I}}(f)|^2 1_{|\beta_{\mathbf{I}}|<\lambda} &\leq \sum_{\ell\geq 0} (2^{-\ell}\lambda)^2 (2^{-\ell-1}\lambda)^{-r} \\ &\leq \lambda^{2-r} \left[\sum_{\ell\geq 0} 2^{-\ell(2-r)}\right] \end{aligned}$$

Since $2 - r > 0$ the last sum converges and we can conclude.

The following proposition gives some details about the embeddings between the spaces $\mathcal{F}^{\mathbf{R}}_{N,d}$ and $W_{N,d}(r)$:

Proposition 4.5.3. *Assume that $R_1 = \cdots = R_N = R$. Set $r = 2d_{\max}/(d_{\max} + 2R)$. Then*

$$\mathcal{F}^{\mathbf{R}}_{N,d} \subset W_{N,d,\log^{N-1}}(r) .$$

Proof
Let $f \in \mathcal{F}^{\mathbf{R}}_{N,d}$. Define J_λ such that $2^{-J_\lambda} \leq \lambda^{\frac{2}{2R+d_{\max}}} \leq 2^{-J_\lambda+1}$. Observe that

$$\begin{aligned}
\sum_{\mathbf{I}} |\beta_{\mathbf{I}}(f)|^2 1_{|\beta_{\mathbf{I}}(f)|\leq\lambda} &\leq \sum_{|\mathbf{j}|\leq J_\lambda}\sum_{\mathbf{k}} |\beta_{\mathbf{I}}(f)|^2 1_{|\beta_{\mathbf{I}}(f)|\leq\lambda} + \sum_{|\mathbf{j}|> J_\lambda}\sum_{\mathbf{k}} |\beta_{\mathbf{I}}(f)|^2 1_{|\beta_{\mathbf{I}}(f)|\leq\lambda} \\
&\leq \lambda^2 \sum_{|\mathbf{j}|\leq J_\lambda}\sum_{\mathbf{k}} 1 + \sum_{\ell> J_\lambda}\sum_{|\mathbf{j}|=\ell} 2^{-2R|\mathbf{j}|} \\
&\leq \lambda^2 \sum_{|\mathbf{j}|\leq J_\lambda} 2^{j_1 d_1+\cdots+j_N d_N} + \sum_{\ell> J_\lambda} 2^{-2R\ell}\ell^{N-1} \\
&\leq \lambda^2 \sum_{\ell\leq J_\lambda}\sum_{|\mathbf{j}|=\ell} 2^{|\mathbf{j}|d_{\max}} + \sum_{\ell> J_{\lambda,R}} 2^{-2R\ell}\ell^{N-1} \\
&\leq \lambda^2 2^{J_\lambda d_{\max}} J_\lambda^{N-1} + 2^{-2RJ_\lambda} J_\lambda^{N-1}
\end{aligned}$$

where in the two last inequalities we used the assumption $f \in \mathcal{F}^{\mathbf{R}}_{N,d}$ and Theorem 4.5.1. The conclusion comes from the definition of the index $J_{\lambda,N}$ and of that of the space $W_{N,d}(r)$ for $r = 2d_{\max}/(d_{\max}+2R)$.

We can now use all these results to deduce the minimax results stated in Section 4.3.2.

4.5.2 Proof of upper bound

We fix some function $f \in \mathcal{F}^{\mathbf{R}}_{N,d}$ and bound as usual the quadratic risk $\mathbb{E}\|f-\widehat{f}\|^2_{L^2}$:

$$\begin{aligned}
\mathbb{E}\|f_\varepsilon-\widehat{f}\|^2_{L^2} &\leq \sum_{\mathbf{i}}\sum_{j_1+\cdots+j_N>J_\varepsilon} [\beta_I(f)]^2 \\
&\quad + \sum_{\mathbf{i}}\sum_{j_1+\cdots+j_N\leq J_\varepsilon} \beta_I(f)^2\mathbb{E}[1_{|\beta_I(f_\varepsilon)|\leq\lambda_\varepsilon}] \\
&\quad + \sum_{\mathbf{i}}\sum_{j_1+\cdots+j_N\leq J_\varepsilon} \mathbb{E}[|\beta_I(f_\varepsilon)-\beta_I(f)|^2 1_{|\beta_I(f_\varepsilon)|>\lambda_\varepsilon}]
\end{aligned}$$

We first bound the sum $\sum_{\mathbf{i}}\sum_{j_1+\cdots+j_N>J_\varepsilon}[\beta_I(f)]^2$ using the assumption $f \in \mathcal{F}^R_{N,d}$ and Theorem 4.5.1. Since $R_1 = \cdots = R_N = R$ and $j_1+\cdots+j_N > J_\varepsilon$, one deduces that

$$\sum_{\mathbf{i}}\sum_{j_1+\cdots+j_N>J_\varepsilon} [\beta_I^2] \leq C 2^{-2RJ_\varepsilon} J_\varepsilon^{N-1} .$$

We now bound the sum $\sum_{\mathbf{i}}\sum_{j_1+\cdots+j_N\leq J_\varepsilon}\sum_k \beta_I(f)^2\mathbb{E}[1_{|\beta_I(f_\varepsilon)|\leq\lambda_\varepsilon}]$. Observe that if $|\beta_I(f_\varepsilon)| \leq \lambda_\varepsilon$, either $|\beta_I(f)| \leq 2\lambda_\varepsilon$ either $|\beta_I(f_\varepsilon)-\beta_I(f)| > \lambda_\varepsilon$. It implies that

$$\begin{aligned}
\sum_{\mathbf{i}}\sum_{j_1+\cdots+j_N\leq J_\varepsilon}\sum_k \beta_I(f)^2\mathbb{E}[1_{|\beta_I(f_\varepsilon)|\leq\lambda_\varepsilon}] &\leq \sum_{\mathbf{i}}\sum_{j_1+\cdots+j_N\leq J_\varepsilon}\sum_k \beta_I(f)^2\mathbb{E}[1_{|\beta_I(f)|\leq\lambda_\varepsilon}] \\
&\quad + \sum_{\mathbf{i}}\sum_{j_1+\cdots+j_N\leq J_\varepsilon}\sum_k \beta_I(f)^2\mathbb{E}[1_{|\beta_I(f_\varepsilon)-\beta_I(f)|>\lambda_\varepsilon}]
\end{aligned}$$

The sum $\sum_{\mathbf{i}}\sum_{j_1+\cdots+j_N\leq J_\varepsilon}\beta_I(f)^2\mathbb{E}[1_{|\beta_I(f)|\leq\lambda_\varepsilon}]$ is deterministic and equals

$$\sum_{\mathbf{i}}\sum_{j_1+\cdots+j_N\leq J_\varepsilon} \beta_I(f)^2 1_{|\beta_I(f)|\leq\lambda_\varepsilon}.$$

It can then be bounded using the embedding proved in Proposition 4.5.3 and the definition of the spaces $W_{N,d}(r)$.

The bound of the sum $\sum_i\sum_{j_1+\cdots+j_N\leq J_\varepsilon}\sum_k \beta_I(f)^2\mathbb{E}1_{|\beta_I(f_\varepsilon)-\beta_I(f)|>\lambda_\varepsilon}$ comes from the Gaussian assumption on the noise and the classical concentration inequality $\mathbb{P}(|Z|>\lambda) \leq Ce^{-\lambda^2/2}$ valid for any standard

Gaussian random variable Z which implies that

$$\begin{aligned}\sum_{\mathbf{i}}\sum_{j_1+\cdots+j_N\leq J_\varepsilon}\beta_I(f)^2\mathbb{E}[1_{|\beta_I(f_\varepsilon)-\beta_I(f)|>\lambda_\varepsilon}] &= \sum_{\mathbf{i}}\sum_{j_1+\cdots+j_N\leq J_\varepsilon}\beta_I(f)^2\mathbb{P}[|(\beta_I(f_\varepsilon)-\beta_I(f))/\varepsilon|>\lambda_\varepsilon/\varepsilon]\\ &\leq \varepsilon\sum_{\mathbf{i}}\sum_{j_1+\cdots+j_N\leq J_\varepsilon}\beta_I(f)^2 = C\varepsilon^{m^2/2}\end{aligned}$$

We then deduce that

$$\sum_{\mathbf{i}}\sum_{j_1+\cdots+j_N\leq J_\varepsilon}\sum_k\beta_I(f)^2\mathbb{E}1_{|\beta_I(f_\varepsilon)|\leq\lambda_\varepsilon}\leq C\varepsilon^{m^2/2}+\lambda_\varepsilon^{4R/(2R+d_{\max})}\leq C\lambda_\varepsilon^{4R/(2R+d_{\max})}$$

as soon as $m>2\sqrt{2}$.

Finally the bound of the sum $\sum_i\sum_{j_1+\cdots+j_N\leq J_\varepsilon}\sum_k\mathbb{E}|\beta_I(f_\varepsilon)-\beta_I(f)|^2 1_{|\beta_I(f_\varepsilon)|>\lambda_\varepsilon}$ shall follow from Propositions 4.5.2 and 4.5.3. Indeed, one has by the Cauchy Schwartz inequality and as in the bound of the last sum

$$\begin{aligned}&\sum_{i}\sum_{j_1+\cdots+j_N\leq J_\varepsilon}\sum_k\mathbb{E}|\beta_I(f_\varepsilon)-\beta_I(f)|^2 1_{|\beta_I(f_\varepsilon)|>\lambda_\varepsilon}\\ \leq\ & \varepsilon^2\sum 1_{|\beta_I(f)|>\lambda_\varepsilon}+2^{J_\varepsilon/2}\varepsilon^2\sum\mathbb{P}^{1/2}[|\beta_I(f_\varepsilon)-\beta_I(f)|>\lambda_\varepsilon/2]\\ \leq\ & C\lambda_\varepsilon^{4R/(2R+d_{\max})}|\log(\lambda_\varepsilon)|^{N-1}+2^{J_\varepsilon/2}\varepsilon^{m^2/16}\leq C\lambda_\varepsilon^{4R/(2R+d_{\max})}|\log(\lambda_\varepsilon)|^{N-1}\end{aligned}$$

as soon as $m>8$. The last display is deduced from Proposition 4.5.3 and the classical concentration inequality for standard Gaussian random variables. Gathering the bound of the three sums $\sum_{\mathbf{i}}\sum_{j_1+\cdots+j_N\leq J_\varepsilon}\beta_I(f)^2\mathbb{E}[1_{|\beta_I(f)|\leq\lambda_\varepsilon}]$, $\sum_i\sum_{j_1+\cdots+j_N\leq J_\varepsilon}\sum_k\beta_I(f)^2\mathbb{E}1_{|\beta_I(f_\varepsilon)-\beta_I(f)|>\lambda_\varepsilon}$ and $\sum_i\sum_{j_1+\cdots+j_N\leq J_\varepsilon}\sum_k\mathbb{E}|\beta_I(f_\varepsilon)-\beta_I(f)|^2 1_{|\beta_I(f_\varepsilon)|>\lambda_\varepsilon}$, we get the upper bound stated in Theorem 4.3.1.

4.5.3 Proof of lower bound

The derivation of a lower bound for wavelet function estimation on $\mathcal{F}^{\mathbf{R}}_{N,d}(K)$ relies on a classical procedure for constructing minimax lower bounds in non-parametric estimation known as the Assouad's method. We give here a version of this lemma when functional spaces and quadratic L^2 distances are considered (see Lemma 2 in [Yu, 1997] and Lemma 10.2 in [Härdle et al., 2012]).

Lemma 4.5.4 (Assouad's lemma)**.** *Let V be a functional space containing a set of functions $\{g_\tau\}_\tau$ with $\tau\in\{0,1\}^m$. For each couple τ and τ', we write $\tau\sim\tau'$ if τ and τ' differs in only one coordinate, and $\tau\underset{k}{\sim}\tau'$ if it is the k^{th} coordinate. If we assume that for any k*

$$\inf_{\tau\underset{k}{\sim}\tau'}\mathbb{E}\|g_\tau-g_{\tau'}\|_{L^2}^2\geq\delta$$

Then, any estimator $\widehat{f}$ of a function $f\in V$ verifies

$$\max_{g_\tau}\mathbb{E}\|\widehat{f}-g_\tau\|_{L^2}^2\geq m\frac{\delta}{2}\min_{\tau\sim\tau'}\{\Lambda(P_\tau,P_{\tau'})\},$$

where P_τ is the probability measure associated to g_τ and the affinity Λ is given by

$$\Lambda(P_\tau,P_{\tau'})=1-\frac{|P_\tau-P_{\tau'}|_1}{2}.$$

In order to use this lemma for constructing the lower bound, we need first to reduce the problem to a parametric family of the form $\{g_\tau\}_\tau$. The value of m depends on a fixed scale which calibrates the risk in order to have an optimal convergence. One expects that such scale defines the limit between finest scales for which the coefficients are small and so dominated by the smoothness (i.e encode noise on some coefficients) and coarse scales. As coarse scales do not contribute to the risk, the error is driven by the error made on

the hardest scale of fine scales which we denote J_ε. We are given $C_0 > 0$ not depending on $\varepsilon > 0$ and J_ε. Thereafter, we define the following family $\{g_\tau\}_\tau$ depending on ε, C_0. For any $\tau = (\tau_I) \in \{0,1\}^m$ with $m = \#\{I \quad s.t \quad |\mathbf{j}| = J_\varepsilon\}$, we set

$$g_\tau = \sum_{\mathbf{i}} \sum_{\mathbf{j}, j_1+\cdots+j_N=J_\varepsilon} \beta_{I,\tau}\psi_I \text{ with } \beta_{I,\tau} = \begin{cases} 0 & \text{if } \tau_I = 0 \\ C_0\varepsilon & \text{otherwise.} \end{cases},$$

We first check that this family belongs to the class $V = \mathcal{F}^{\mathbf{R}}_{N,d}(K)$:

Lemma 4.5.5. *Assume that $\varepsilon^2 2^{J_\varepsilon d_{max}} J_\varepsilon \asymp 2^{-2J_\varepsilon R}$, with $d_{max} = max(d_i)$. Then we have*

$$\{g_\tau\}_\tau \subseteq \mathcal{F}^{\mathbf{R}}_{N,d}(K)$$

Proof

It directly comes from the definition of the class $\mathcal{F}^{\mathbf{R}}_{N,d}(K)$ and from Theorem 4.5.1.

All that remains, now, is to apply lemma 4.5.4 for the class $\mathcal{F}^{\mathbf{R}}_{N,d}(K)$ and the parametric family $\{g_\tau\}_\tau$. First, note that for the family $\{g_\tau\}_\tau$, we have

$$\delta = C_0^2\varepsilon^2.$$

Note also that the hypercube dimension m is given by the cardinality of the coefficients at the scale J_ε.

$$m = \#\{I \quad s.t \quad |\mathbf{j}| = J_\varepsilon\} = J_\varepsilon^{N-1} 2^{(\sum_{i=1}^n d_i j_i)}, \tag{4.11}$$

Hence,

$$\begin{aligned} \max_{g_\tau} \mathbb{E}\|\widehat{f} - g_\tau\|_{L^2}^2 &\geq C_0^2\varepsilon^2 J_\varepsilon^{N-1} 2^{(\sum_{i=1}^n d_i j_i)} \min_{\tau\sim\tau'}\{\Lambda(P_\tau, P_{\tau'})\}, \\ &\geq C_0^2\varepsilon^2 J_\varepsilon^{N-1} 2^{J_\varepsilon d_{\max}} \min_{\tau\sim\tau'}\{\Lambda(P_\tau, P_{\tau'})\}, \end{aligned}$$

As usual, the calibration $\varepsilon^2 2^{J_\varepsilon d_{max}} J_\varepsilon \asymp 2^{-2J_\varepsilon R}$ imposes that J_ε is of the same order as $\log(\varepsilon^{-1})$ which yields to

$$2^{J_\varepsilon} = \left(\varepsilon^2[\log(\varepsilon^{-1})]^{N-1}\right)^{-1/(2R+d_{\max})}. \tag{4.12}$$

Thus

$$\max_{g_\tau} \mathbb{E}\|\widehat{f} - g_\tau\|_{L^2}^2 \geq C_0^2\varepsilon^2\left(\varepsilon^2 \log(\varepsilon^{-1})]^{N-1}\right)^{-d_{\max}/(2R+d_{\max})} \min_{\tau\sim\tau'}\{\Lambda(P_\tau, P_{\tau'})\}.$$

Finally, by definition the affinity Λ takes only positive values, which gives

$$\max_{g_\tau} \mathbb{E}\|\widehat{f} - g_\tau\|_{L^2}^2 \geq C\left(\varepsilon^2 \log(\varepsilon^{-1})]^{N-1}\right)^{2R/(2R+d_{max})}.$$

with $C = C_0^2 \min_{\tau\sim\tau'}\{\Lambda(P_\tau, P_{\tau'})\}$, which ends the proof.

Bibliography

Anestis Antoniadis, Jeremie Bigot, and Theofanis Sapatinas. Wavelet estimators in nonparametric regression: a comparative simulation study. *Journal of Statistical Software*, 6:pp–1, 2001.

F. Autin, G. Claeskens, and J.M. Freyermuth. Asymptotic performance of projection estimators in standard and hyperbolic wavelet bases. *To appear in Electronic journal of statistics*, 2015.

Emrah Bostan, Orestis Vardoulis, Davide Piccini, Pouya D Tafti, Nikolaos Stergiopulos, and Michael Unser. Spatio-temporal regularization of flow-fields. In *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on*, pages 836–839. Ieee, 2013.

Emrah Bostan, Michael Unser, and John Paul Ward. Divergence-free wavelet frames. *Signal Processing Letters, IEEE*, 22(8):1142–1146, 2015.

Chein-I Chang. *Hyperspectral imaging: techniques for spectral detection and classification*, volume 1. Springer Science & Business Media, 2003.

Ingrid Daubechies. *Ten lectures on wavelets*, volume 61. SIAM, 1992.

RA DeVore, SV Konyagin, and VN Temlyakov. Hyperbolic wavelet approximation. *Constructive Approximation*, 14(1):1–26, 1998.

R.A. DeVore, G. Petrova, and P. Wojtaszczyk. Anisotropic smoothness spaces via level sets. *Communications on Pure and Applied Mathematics*, 61(9):1264–1297, September 2008.

David L. Donoho and Iain M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81: 425–455, 1994.

Jalal M Fadili, Jérôme Mathieu, Barbara Romaniuk, and Michel Desvignes. Bayesian wavelet-based poisson intensity estimation of images using the fisz transformation. In *International conference on image and signal processing*, volume 1, pages 242–253, 2003.

Younes Farouj, Laurent Navarro, Marianne Clausel, and Philippe Delacharte. Débruitage de séquence d'images dynamiques par ondelettes espace-temps hyperboliques. In *In XXVème colloque GRETSI*, 2015.

Younes Farouj, Jean-Marc Freyermuth, Laurent Navarro, Marianne Clausel, and Philippe Delacharte. Generalized hyperbolic-crossing wavelets. *In preparation*, 2017.

P. Fryzlewicz. Data-driven wavelet-fisz methodology for nonparametric function estimation. *Electronic journal of statistics*, 2:863–896, 2008.

Piotr Fryzlewicz and Guy P Nason. A haar-fisz algorithm for poisson intensity estimation. *Journal of computational and graphical statistics*, 13(3):621–638, 2004.

Damien Garcia, Juan C del Álamo, David Tanné, Raquel Yotti, Cristina Cortina, Éric Bertrand, José Carlos Antoranz, Esther Pérez-David, Régis Rieu, Francisco Fernández-Avilés, et al. Two-dimensional intraventricular flow mapping by digital processing conventional color-doppler echocardiography images. *Medical Imaging, IEEE Transactions on*, 29(10):1701–1713, 2010.

Kanghui Guo, Demetrio Labate, Wang-Q Lim, Guido Weiss, and Edward Wilson. Wavelets with composite dilations. *Electronic research announcements of the American Mathematical Society*, 10(9):78–87, 2004.

Wolfgang Härdle, Gerard Kerkyacharian, Dominique Picard, and Alexander Tsybakov. *Wavelets, approximation, and statistical applications*, volume 129. Springer Science & Business Media, 2012.

B. K. P. Horn and B. G. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

Gérard Kerkyacharian and Dominique Picard. Thresholding algorithms, maxisets and well-concentrated bases. *Test*, 9(2):283–344, 2000.

T. Loupas, W.N. McDicken, and P.L. Allan. An adaptive weighted median filter for speckle suppression in medical ultrasonic images. *Circuits and Systems*, 36:129–135, 1989.

Michael Markl, Frandics P Chan, Marcus T Alley, Kris L Wedding, Mary T Draney, Chris J Elkins, David W Parker, Ryan Wicker, Charles A Taylor, Robert J Herfkens, et al. Time-resolved three-dimensional phase-contrast mri. *Journal of Magnetic Resonance Imaging*, 17(4):499–506, 2003.

Michael H. Neumann. Multivariate wavelet thresholding in anisotropic function spaces. *Statistica Sinica*, 10: 399–431, 2000.

Michael H Neumann and Rainer Von Sachs. Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra. *The Annals of Statistics*, 25(1):38–76, 1997.

Robert D Nowak and Richard G Baraniuk. Wavelet-domain filtering for photon imaging systems. *Image Processing, IEEE Transactions on*, 8(5):666–678, 1999.

Alfio Quarteroni, Alessandro Veneziani, and Paolo Zunino. Mathematical and numerical modeling of solute dynamics in blood flow and arterial walls. *SIAM Journal on Numerical Analysis*, 39(5):1488–1511, 2002.

Norbert Remenyi, Orietta Nicolis, Guy Nason, and Brani Vidakovic. Image denoising with 2d scale-mixing complex wavelet transforms. *Image Processing, IEEE Transactions on*, 23(12):5165–5174, 2014.

Hans-Jürgen Schmeisser. Recent developments in the theory of function spaces with dominating mixed smoothness. *Nonlinear Analysis, Function Spaces and Applications*, pages 145–204, 2007.

Hans-Jürgen Schmeisser and Winfried Sickel. Spaces of functions of mixed smoothness and approximation from hyperbolic crosses. *Journal of Approximation Theory*, 128(2):115 – 150, 2004. ISSN 0021-9045. doi: http://dx.doi.org/10.1016/j.jat.2004.04.007. URL `http://www.sciencedirect.com/science/article/pii/S0021904504000693`.

Gary A Shaw and Hsiao-hua K Burke. Spectral imaging for remote sensing. *Lincoln Laboratory Journal*, 14 (1):3–28, 2003.

Bin Yu. Assouad, fano, and le cam. In *Festschrift for Lucien Le Cam*, pages 423–435. Springer, 1997.

Vyacheslav Zavadsky. Image approximation by rectangular wavelet transform. *Journal of Mathematical Imaging and Vision*, 27(2):129–138, 2007.

CHAPTER 5

Experiments

Abstract

In this chapter, we present some experiments to illustrate the effectiveness of the structured variable grouping presented earlier. We start with the simple problem of sequence denoising under model (4.1), then we consider the denoising of hyperspectral data, velocity flows and finally denoising under partially data-dependent noise models.

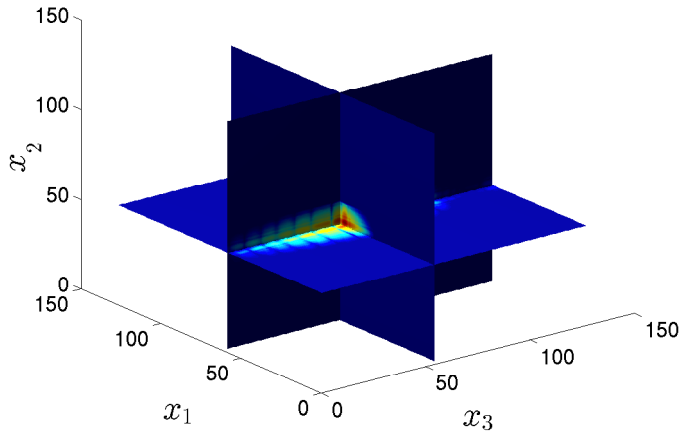
We aim at showing the merits of using structured wavelets presented above compared to the classical constructions. Therefore we compare our results with those obtained by wavelet thresholding which do not take all variables into account (2D-Wavelets) and isotropic multivariate wavelet thresholding techniques (3D-Wavelets) which acts in the same manner on all variables. We also use simple universal hard thresholding rules. Except the velocity flows, all data have a $[0-255]$ grey-values range. At each noise level, we compared results of the different methods using the classical *Peak Signal to Noise Ratio* (PSNR) as a criteria. In the case of data-dependent noise, we also show the difference between the true image and the denoised result of every method. This is known in the literature as the *method noise* [Buades et al., 2005]. For the wavelet filters, *Daubechies* wavelets with 6 vanishing moments are used in the different directions for the 2D-Wavelets and the Stuctured wavelets, while the 3D-Wavelets are complex isotropic filters [Selesnick and Li, 2003].

In each case, we first define sets of variables that we shall group. On these variables the appropriate wavelet transform is taken (isotropic, divergence-free) and the appropriate estimator is defined (isotropic hard thresholding, hyperbolic hard thresholding, Fisz, etc$\cdots$). The rest of variables are taken into account by the standard tensor product. It is interesting to note that the complexity of the construction in dimension d is $O(N^d)$ as for classical *isotropic* wavelets, no matter what is the order of the variables. Figure 5.1 shows a particular case construction for $d = 3$ in which isotropy in $\{x_1, x_2\}$ is considered. This latter case will be of special interest, as x_3 will refer to the time variable. It can be seen that the obtained atoms have an isotropic support along x_1 and x_2 while it is different along x_3. Let us now detail the different cases that we considered :

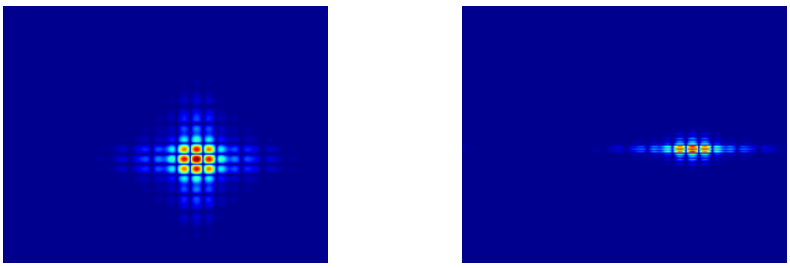
5.1 Image Sequence denoising

The utility of considering isotropy in space and anisotropy in time was discussed and motivated in Section 4.2. We also want to show the merit of the spatio-temporal treatment in general. We consider two images for our synthetic experiments. The first sequence *"Ayiko"* available on the web [1]. The second sequence *"Heart"* is generated by the ASSESS software [Clarysse et al., 2011]. This software applies a simple kinematic model of the heart deformation on an initial image. The result is a non-corrupted realistic simulation of a cardiac MRI sequence. We corrupted the two sequences with respect to noise model (4.7) by picking a noise level $\varepsilon \in \{5, 10, 15, 20\}$. PSNR results are given in 5.1. Results obtained by the structured wavelets are clearly

[1] `http://see.xidian.edu.cn/vipsl/database_Video.html`



The 3D atom



$x_1 - x_2$ plane $x_1 - x_3$ plane

Figure 5.1: Example of a Structured Wavelet atoms.

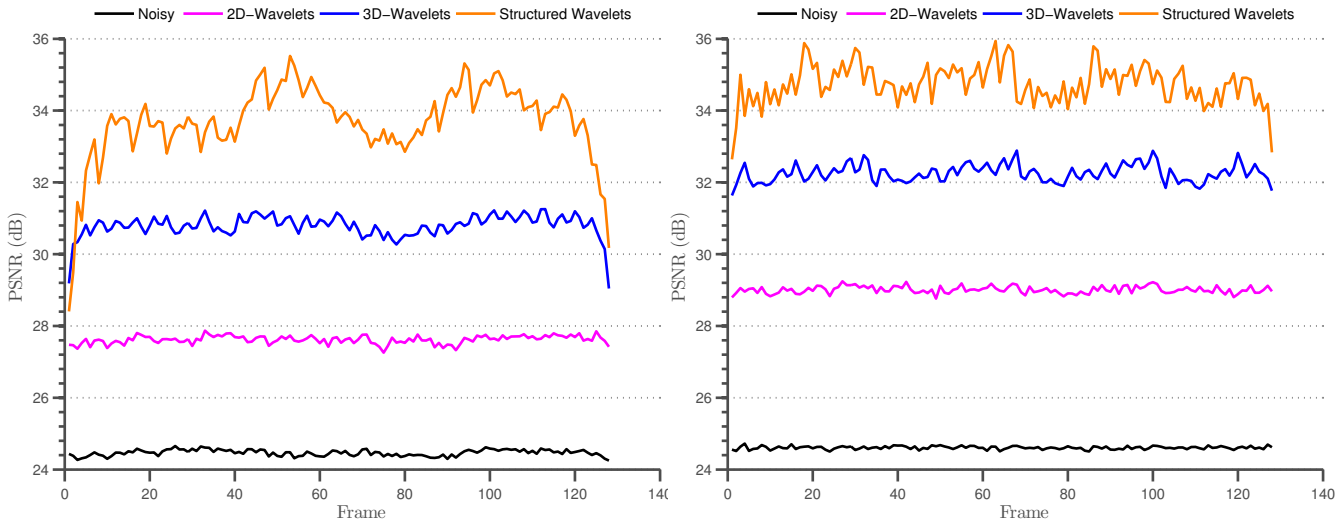
superior to those obtained by 2D-Wavelets and 3D-Wavelets at all noise levels. The benefit of taking the temporal dimension into account can be observed as 3D-Wavelets give better results than 2D-Wavelets. As the visualization is done frame by frame, we show in figure the temporal evolution of the PSNR for each method. It can be observed that, in the case of the 2D-Wavelet approach it imposes a local treatment in time, as a consequence there is no temporal evolution of the PSNR. It can be observed that the results of 3D-Wavelets and structured wavelets demonstrate a time-dependent behaviour. In particular, the PSNR starts and finishes with low values as there is not enough information in the temporal domain and has higher values in the rest of the sequence. In the case of structured wavelets some sudden drops in the PSNR are observed due to temporal discontinuities. Note that an interesting phenomenon is observed in the case of the *"Heart"* sequence as the periodicity of the cardiac cycles can be seen in the PSNR results. This is due to the variations in deformation and discontinuities between cycles as the heart motion changes direction introducing a strong temporal discontinuity.

Original Noisy ($\sigma = 10$) 2D-wavelets 3D-Wavelets Structured Wavelets

Figure 5.2: Results of various methods applied to the 20^{th} image of the sequence *"Ayiko"*. Quantitative evaluation is given in Table 5.1.

Original | Noisy ($\sigma = 10$) | 2D-wavelets | 3D-Wavelets | Structured Wavelets

Figure 5.3: Results of various methods applied to the 20^{th} image of the sequence *"Heart"*. Quantitative evaluation is given in Table 5.1.



"Ayiko" *"Heart"*

Figure 5.4: PSNR evolution for different methods applied to the sequences *"Ayiko"* and *"Heart"*.

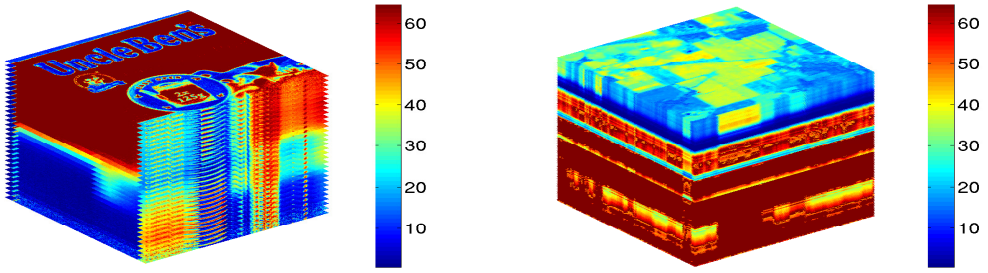
	"Ayiko"			*"Heart"*		
σ	2D	3D	Structured	2D	3D	Structured
5	35.05	38.12	**40.48**	35.60	38.54	**41.23**
10	30.33	33.62	**36.16**	31.41	34.53	**37.12**
25	27.75	30.92	**33.68**	29.00	32.24	**34.70**
20	26.03	29.11	**31.88**	27.33	30.70	**32.96**

Table 5.1: Quantitative comparison (PSNR) for different methods applied to *"Ayiko"* and *"Heart"* at different noise levels.

5.2 Spectral Denoising

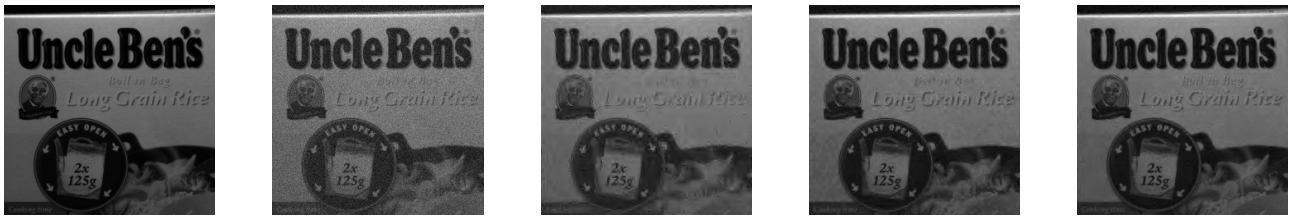
We considered also two examples of hyperspectral and multispectral data. The first sample *"Uncle Bens"* is taken from the Multispectral database[2]. It consists of 31 images with a resolution reflectance ranging from 400nm to 700nm. The second sample *"Indian Pine"* from the AVIRIS hyperspectral sensor database[3]. It contains 220 images, each one representing a specific wavelength. The multispectral data was modified to have a dyadic size in the spectral dimension for the wavelet transform. Note that the Structured Wavelet approach does not require to have the same size in all dimension as the 3D-Wavelet approach. We show the results only for the original 31 samples.

[2] `http://www2.cmp.uea.ac.uk/Research/compvis/MultiSpectralDB.htm`
[3] `https://purr.purdue.edu/publications/1947`



"Uncle Bens" *"Indian Pines"*

Figure 5.5: 3D Stacks : *"Indian Pines"* and *"Uncle Bens"*.



Original Noisy ($\sigma = 10$) 2D-wavelets 3D-Wavelets Structured Wavelets

Figure 5.6: Results of various methods applied to the 20^{th} image of the stack *"Uncle Bens"*. Quantitative evaluation is given in Table 5.2.

Original Noisy ($\sigma = 10$) 2D-wavelets 3D-Wavelets Structured Wavelets

Figure 5.7: Results of various methods applied to the image 10^{th} image of the stack *"Indian Pines"*. Quantitative evaluation is given in Table 5.2.

	"Uncle Bens"			*"Indian Pines"*		
σ	2D	3D	Structured	2D	3D	Structured
5	36.14	37.78	**39.41**	35.26	35.64	**37.38**
10	32.26	34.24	**36.24**	31.81	32.22	**34.36**
15	30.11	32.17	**34.16**	30.08	30.55	**32.70**
20	28.67	30.65	**32.53**	28.94	29.36	**31.42**

Table 5.2: Quantitative comparison (PSNR) for different methods applied to *"Uncle Bens"* and *"Indian Pines"* at different noise levels.

5.3 Incompressible flows Denoising

The experiments were done on a synthetic velocity map which was initiated by a null divergence vector flow (u, v) where the horizontal and vertical component are given respectively by $u_0(x, y) = \sin(2\pi x)^2 \sin(4\pi y)$ and $v_0(x, y) = -\sin(2\pi y)^2 \sin(4\pi x)$. The temporal evolution was governed by a rotation matrix $u(x, y, t + 1) = u(x, y, t) + hu(x, y, t)$ and $v(x, y) = v(x, y, t) - hv(x, y, t)$ where h is fixed. The result is a velocity map in form of 4 vertices with increasing velocity (e.g Figure 5.9). We corrupted this flow with gaussian noise where $\sigma = 0.1, 0.3, 0.5$. We tested spatial two-dimensional divergence free wavelets and spatio-temporal

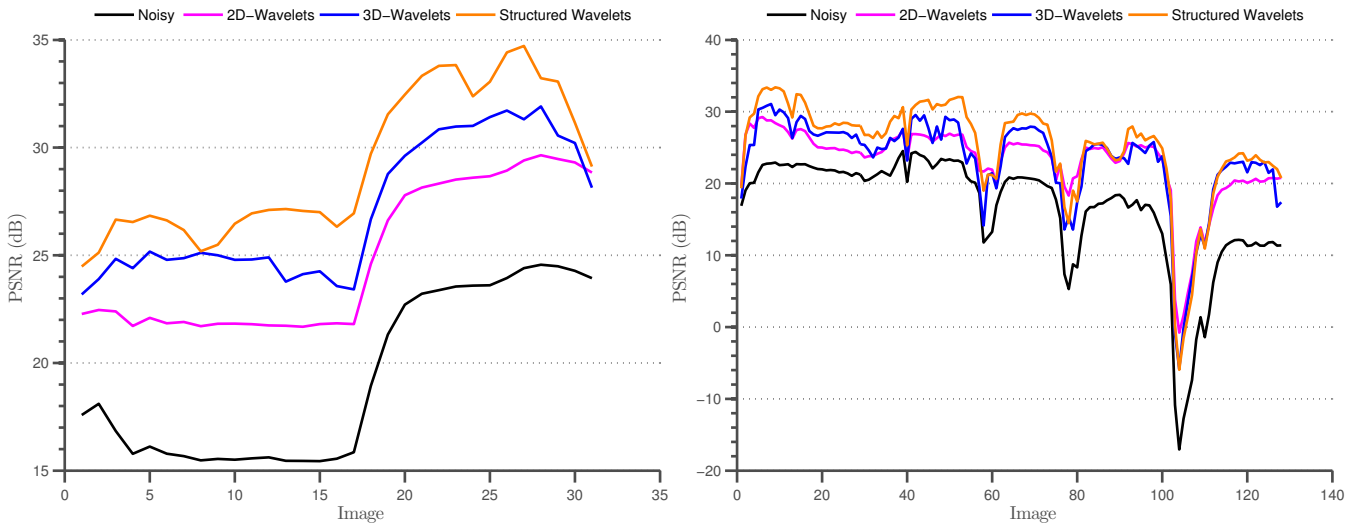


Figure 5.8: PSNR evolution for different methods applied to the data *"Uncle Bens"* and *"Indian Pines"*.

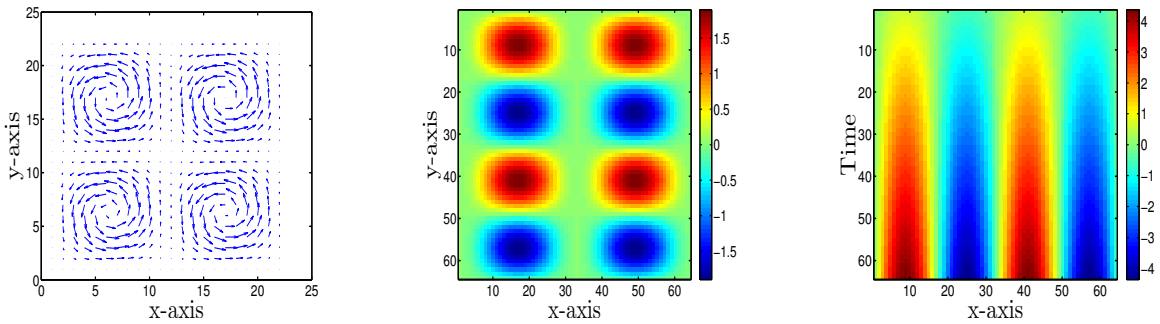


Figure 5.9: 2D+t Flow data : (a) Vector field at the 10^{th} time step. (b) Horizontal velocity at the 10^{th} time step. (c) Temporal evolution of the horizontal velocity for a fixed vertical plan.

structured divergence free wavelets. Here, the 3D construction is not appropriate as divergence freedom cannot be imposed in time. We used divergence free wavelets with periodic boundary condition introduced by Harouna and Perrier [Kadri Harouna and Perrier, 2015]. A visual comparison of the results is given in 5.11. The vector flow obtained using the spatio-temporal structured wavelets are visually smoother and still preserve divergence freedom. As the velocity is increasing and the the noise variance is constant the PSNR is increasing within time time Figure 5.12. Note that the spatio-temporal approach fails in some time steps because of the discontinuities but its overall performance is better than the one of the spatial approach as it can be seen in Table 5.3.

	Horizontal velocity		Vertical velocity	
σ	2D	Structured	2D	Structured
$0,1$	46.64	**49.00**	46.65	**48.90**
$0,3$	39.65	**43.83**	39.51	**43.53**
$0,5$	35.49	**39.65**	35.47	**39.36**

Table 5.3: Quantitative comparison (PSNR) for different methods for horizontal and vertical velocity at different noise levels.

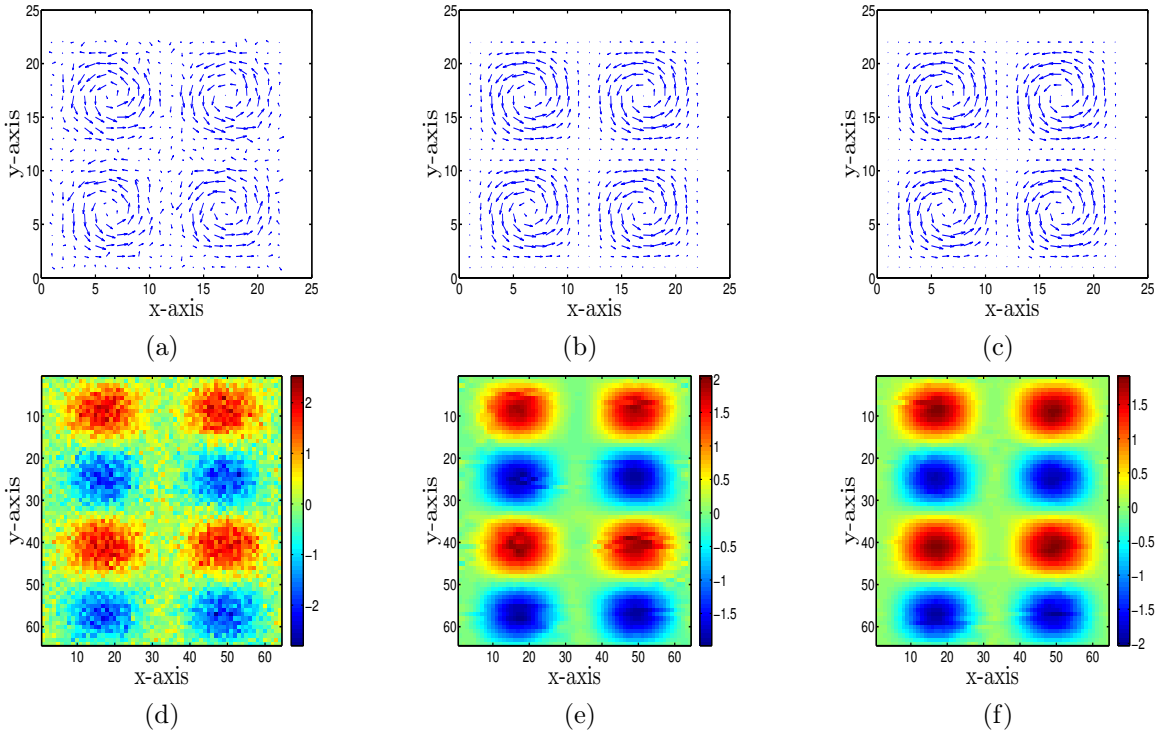


Figure 5.10: Example of results obtained by the two methods for $\sigma = 0.3$ at the 10^{th} time step. (a) Noisy flow, (b) 2D-Wavelets reconstruction of the flow, (c)Structured Wavelets reconstruction of the flow, (d) Noisy horizontal velocity, (e) 2D-Wavelets reconstruction of the horizontal velocity., (f) Stuctured Wavelets reconstruction of the horizontal velocity.

5.4 Mixed models Denoising

Here we consider the model (4.10), we have a particular interest in models arising from ultrasound imaging [Loupas et al., 1989] in which $h(f) = \sqrt{f}$. Our experiments are performed with respect to this model for $\sigma = \{1, 2, 3, 4\}$. We used spatial isotropic wavelets with Fisz variance stabilization [Fryzlewicz, 2008]. We compared the spatial approach to the spatio-temporal approach. Note that again classical $3D$ wavelets are not applicable for this problem. Visual comparisons for the sequences *"Ayiko"* and *"Heart"* are given in Figure 5.13 and Figure 5.14. Note that the spatio-temporal approach provided by the structured wavelet construction provides a considerable visual improvement compared to the spatial approach. In the PSNR evolution in Figure 5.16, we can observe again the temporal structure linked to the heart motion and the superiority of structured wavelets locally and and globally (e.g Table 5.4).

	Test *"Ayiko"*		Kidney *"Heart"*	
σ	2D	Structured	2D	Structured
1	28.82	**29.66**	32.93	**34.31**
2	27.13	**29.26**	30.44	**33.28**
3	25.64	**28.82**	28.48	**32.27**
4	24.42	**28.39**	26.90	**31.34**

Table 5.4: Quantitative comparison (PSNR) for different methods applied to *"Ayiko"* and *"Heart"* at different noise levels.

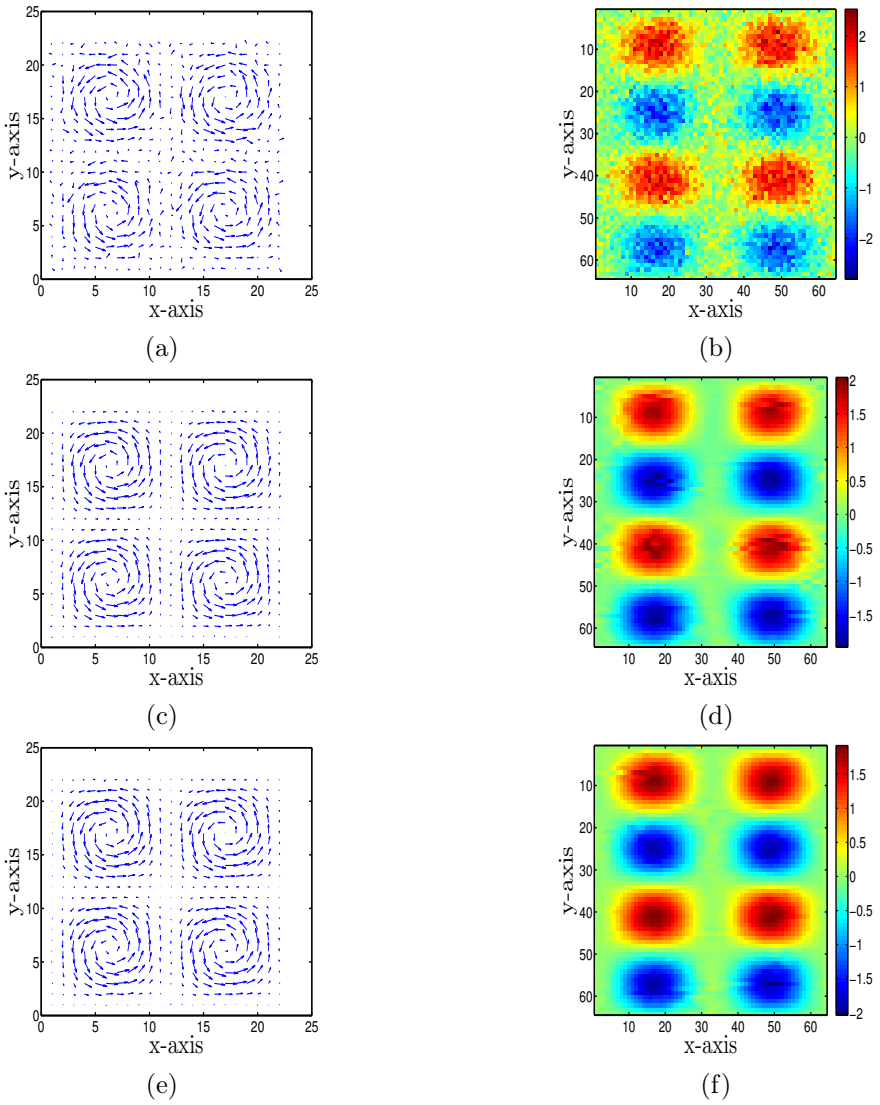
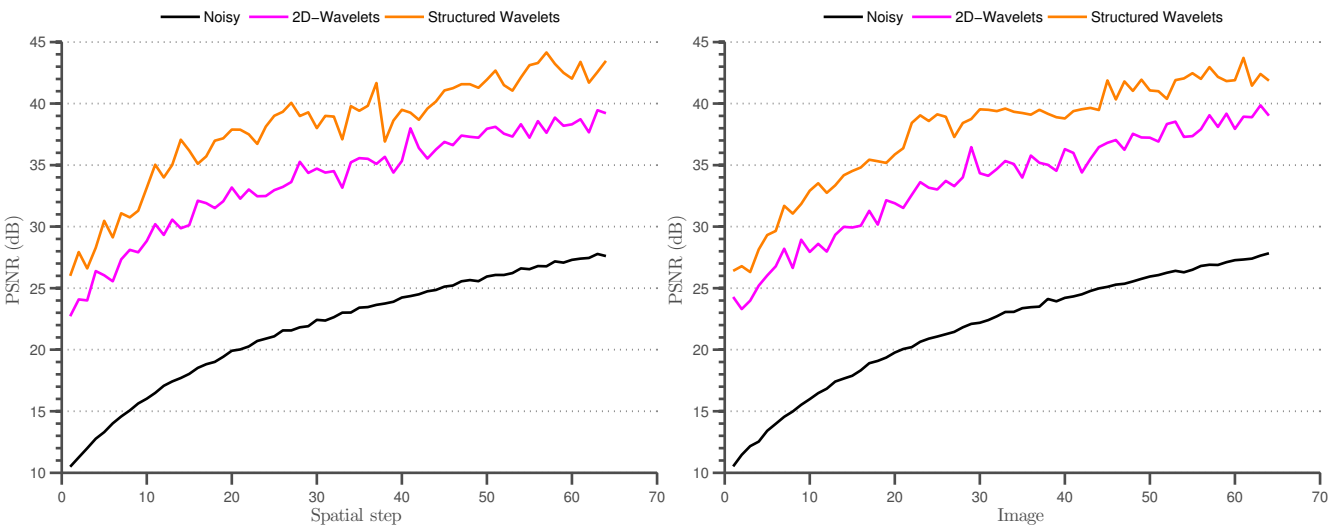


Figure 5.11: Example of results obtained by the two methods for $\sigma = 0.3$ at the 10^{th} time step. (a) Noisy flow, (b) Noisy horizontal velocity, (c) 2D-Wavelets reconstruction of the flow, (d) 2D-Wavelets reconstruction of the horizontal velocity, (c)Structured Wavelets reconstruction of the flow, (d) Stuctured Wavelets reconstruction of the horizontal velocity.

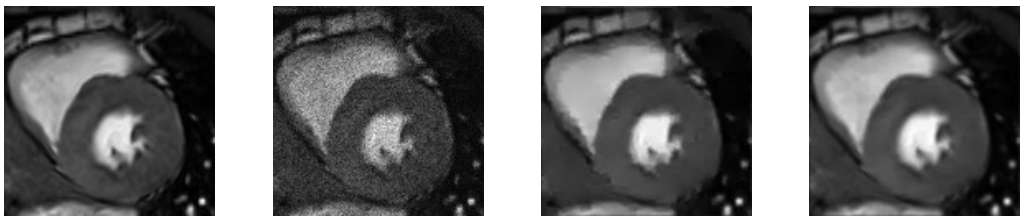


Horizontal velocity Vertical velocity

Figure 5.12: PSNR evolution for the two methods.

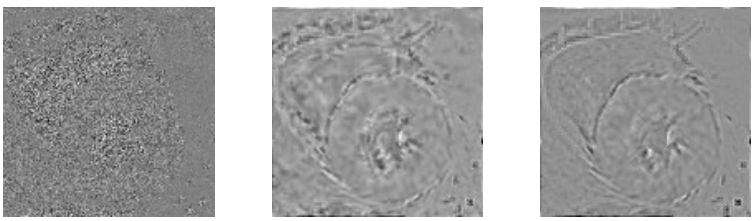
Original Noisy ($\sigma = 2$) 2D-wavelets Structured Wavelets

Figure 5.13: Example of results obtained by the two methods for the 20^{th} image of the sequence *"Ayiko"* with $\sigma = 0.3$.



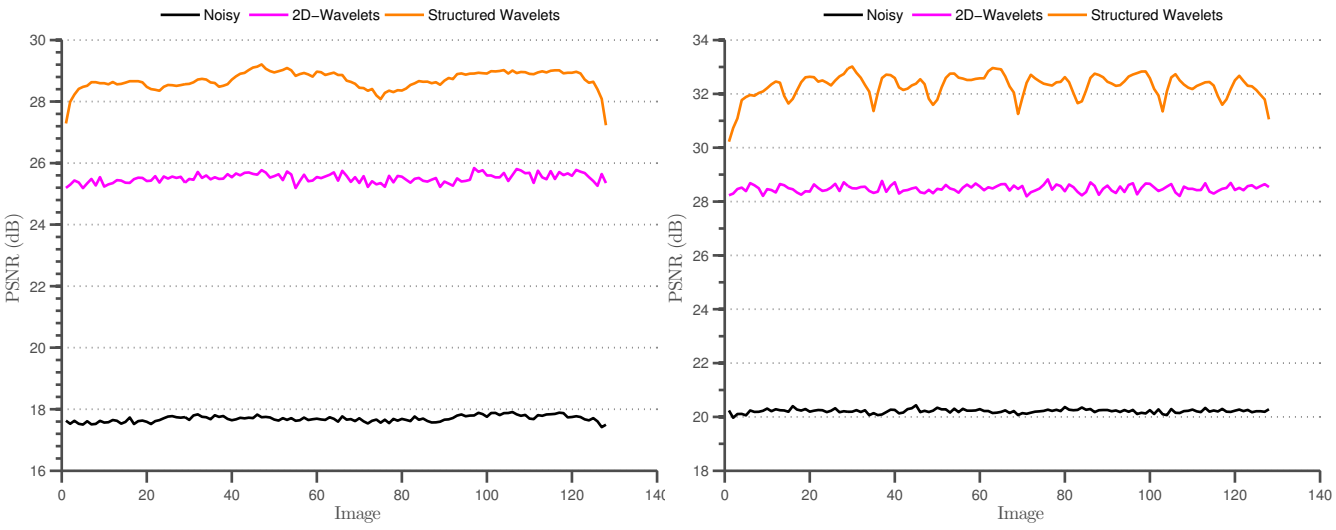
Original Noisy ($\sigma = 2$) 2D-wavelets Structured Wavelets

Figure 5.14: Example of results obtained by the two methods for the 20^{th} image of the sequence *"Heart"* with $\sigma = 0.3$.



Original 2D-wavelets Structured Wavelets

Figure 5.15: Method noise: various methods applied to the 20^{th} image of the sequence *"Heart"*. Quantitative evaluation is given in Table 5.1.



"Ayiko" *"Heart"*

Figure 5.16: PSNR evolution for different methods applied to the sequences *"Ayiko"* and *"Heart"*.

Bibliography

Antoni Buades, Bartomeu Coll, and Jean-Michel Morel. A non-local algorithm for image denoising. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 60–65, 2005.

P. Clarysse, J. Tafazzoli, P. Delachartre, and P. Croisille. Simulation based evaluation of cardiac motion estimation methods in tagged-mr image sequences. *Journal of Cardiovascular Magnetic Resonance*, 13 (Suppl 1):P360, 2011. ISSN 1532-429X. doi: 10.1186/1532-429X-13-S1-P360. URL `http://jcmr-online.com/content/13/S1/P360`.

P. Fryzlewicz. Data-driven wavelet-fisz methodology for nonparametric function estimation. *Electronic journal of statistics*, 2:863–896, 2008.

Souleymane Kadri Harouna and Valérie Perrier. Divergence-free wavelet projection method for incompressible viscous flow on the square. *Multiscale Modeling & Simulation*, 13(1):399–422, 2015.

T. Loupas, W.N. McDicken, and P.L. Allan. An adaptive weighted median filter for speckle suppression in medical ultrasonic images. *Circuits and Systems*, 36:129–135, 1989.

Ivan W Selesnick and Ke Yong Li. Video denoising using 2d and 3d dual-tree complex wavelet transforms. In *Optical Science and Technology, SPIE's 48th Annual Meeting*, pages 607–618. International Society for Optics and Photonics, 2003.

CHAPTER 6

Discussion & Perspectives

Abstract

We give in this chapter a discussion about the positioning of our work compared to existing works on structured group-sparsity. We also mention some orientations for future works on the wavelet construction presented in this part of the thesis.

6.1 Some related work on structured group-sparsity

First, we want to note that a similar construction to the one presented in the paragraph 4.3.1 was proposed in [Duarte and Baraniuk, 2012] for compressed sensing which is also motivated by video acquisition and hyperspectral imaging. There, the purpose was to study the efficiency of considering Kronecker products of different bases for multivariate signals with different regularities. In particular, the authors studied the sparsifying properties and conditions for which these bases can be used in the compressed sensing theory. Structured sparsity have proven to be useful in denoising, existing structures consider groupe-wise sparsity as in *Block Thresholding* [Cai, 2002], Hierarchical sparsity as in *Tree Thresholding* [Baraniuk, 1999] or combinations the two [Autin et al., 2012]. Imposing sparsity in a variational framework can be done via minimizing a ℓ_1*-regularized least squares functional*. This is known as the Lasso problem [Tibshirani, 1996]. Structuring the regularization term in groups is known as the group-lasso [Meier et al., 2008] and has other extensions such as the Fused Lasso [Tibshirani et al., 2005] and the Group Fused Lasso [Alaíz et al., 2013]. All these paradigms consider structures directly on the sparse representation (i.e after transformation). In our work, we aimed at showing the merit of considering the group selection on variables before transformation. As mentioned before, many works in the statistical literature demonstrated the advantages of using hyperbolic wavelets in multivariate analysis. Overcoming the curse of dimensionality has been pointed first in [Neumann, 2000] with recent contributions in [Autin et al., 2014]. In all these works, anisotropy was considered on single variables and the authors aimed at dropping the error to dimension one.

6.2 Methodological extensions

The minimax results of Chapter 4 were obtained assuming that the same regularity parameter along he different groups of variables. As mentioned by [Neumann, 2000], it wouldbe also interesting to consider the different regularity parameters. This is, actually, more appealing when considering problems with different regularities such as the experiments presented in Chapter 5. One can predict that the hardest regularity (smallest parameter) appears in the rate of convergence. Still, this is an open question, particularly, how this parameter will be combined with the dimension of the largest group.

For Functional Data Analysis within the *functional analysis of variance* (FANOVA) framework [Abramovich et al., 2004, Abramovich and Angelini, 2006], the structured wavelet construction might give a new ways of gathering variables presenting similar behaviour, especially, in very high dimensions.

Similarly to considering divergence-free wavelets on some variables, it is also possible to consider operator-like wavelets [Van De Ville et al., 2005, Khalidov et al., 2007]. This type of wavelets can be used, for instance, to invert convolution operators for joint deconvolution/denoising [Khalidov et al., 2011]. The structured construction can deal with problems in which the data is blurred only on some of the variables.

6.3 Real applications

As mentioned in Chapter 5, many of our experiments are motivated by real applications. Trying the proposed methods on real data still the ultimate goal. Applications such as Phase-contrast MRI velocity enhancement and US Doppler imaging are of great interest and enjoyed a lot of attention, recently.

Bibliography

Felix Abramovich and Claudia Angelini. Testing in mixed-effects fanova models. *Journal of statistical planning and inference*, 136(12):4326–4348, 2006.

Felix Abramovich, Anestis Antoniadis, Theofanis Sapatinas, and Brani Vidakovic. Optimal testing in a fixed-effects functional analysis of variance model. *International Journal of Wavelets, Multiresolution and Information Processing*, 2(04):323–349, 2004.

Carlos M Alaíz, Álvaro Barbero, and José R Dorronsoro. Group fused lasso. In *Artificial Neural Networks and Machine Learning–ICANN 2013*, pages 66–73. Springer, 2013.

F. Autin, G. Claeskens, and J.M. Freyermuth. Hyperbolic wavelet thresholding methods and the curse of dimensionality through the maxiset approach. *Appl. Comput. Harmon. Anal.*, 36:239–255, 2014.

Florent Autin, J-M Freyermuth, and Rainer von Sachs. Combining thresholding rules: a new way to improve the performance of wavelet estimators. *Journal of Nonparametric Statistics*, 24(4):905–922, 2012.

Richard G Baraniuk. Optimal tree approximation with wavelets. In *SPIE's International Symposium on Optical Science, Engineering, and Instrumentation*, pages 196–207. International Society for Optics and Photonics, 1999.

T Tony Cai. On block thresholding in wavelet regression: Adaptivity, block size, and threshold level. *Statistica Sinica*, pages 1241–1273, 2002.

M. F. Duarte and R. G. Baraniuk. *Image Processing, IEEE Transactions on*, 21(2):494–504, Feb 2012.

Ildar Khalidov, Dimitri Van De Ville, Thierry Blu, and Michael Unser. Construction of wavelet bases that mimic the behaviour of some given operator. In *Optical Engineering+ Applications*, pages 67010S–67010S. International Society for Optics and Photonics, 2007.

Ildar Khalidov, Jalal Fadili, François Lazeyras, Dimitri Van De Ville, and Michael Unser. Activelets: Wavelets for sparse representation of hemodynamic responses. *Signal Processing*, 91(12):2810–2821, 2011.

Lukas Meier, Sara Van De Geer, and Peter Bühlmann. The group lasso for logistic regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1):53–71, 2008.

Michael H. Neumann. Multivariate wavelet thresholding in anisotropic function spaces. *Statistica Sinica*, 10: 399–431, 2000.

Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

Robert Tibshirani, Michael Saunders, Saharon Rosset, Ji Zhu, and Keith Knight. Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(1): 91–108, 2005.

Dimitri Van De Ville, Thierry Blu, Brigitte Forster, and Michael Unser. Semi-orthogonal wavelets that behave like fractional differentiators. In *Optics & Photonics 2005*, pages 59140C–59140C. International Society for Optics and Photonics, 2005.

Part II

Fisz-wavelet thresholding in two-dimensional anisotropic settings

CHAPTER 7

The methodology

Abstract

In this chapter, we consider the following two-dimensional function estimation problem: we want to recover an unknown function α from a noisy observation X, where the noise component has zero mean and a variance function depending on the unknown α. We prove the optimality of hyperbolic wavelet-Fisz hard-thresholding when u belongs to anisotropic Besov balls. This method computes the hyperbolic wavelet transform of the image, before applying a multiscale variance stabilization technique, via a Fisz transformation. This adapts the wavelet coefficients statistics to the wavelet thresholding paradigm. We also describe a data-driven extension of this technique when h is unknown following previous works by Fryzlewicz and Dellouile. The data-driven extension removes the need for any prior knowledge of the noise model parameters by estimating the noise variance using an isotonic Nadaraya-Watson estimator.

7.1 Introduction

Consider the following regression model

$$X_{t_1,t_2} = \alpha(t_1/n_1, t_2/n_2) + \varepsilon_{t_1,t_2} \quad t_1 = 1, \cdots, n_1 \text{ and } t_2 = 1, \cdots, n_2, \tag{7.1}$$

where ε_{t_1,t_2}'s are random variables with $\mathbb{E}(\varepsilon_{t_1,t_2}) = 0$. The case where $\mathrm{Var}(\varepsilon_{t_1,t_2})$ is constant results in additive white Gaussian noise models such as the ones studied in the previous part of this thesis. When $\mathrm{Var}(\varepsilon_{t_1,t_2})$ is not constant the regression model is called *heteroscedastic*. Such models were less studied in the wavelet thresholding literature. In fact, the empirical wavelet coefficients of the noise component are non longer a independent, identically distributed Gaussian sequence. This makes the construction of an elitist empirical program such as thresholding difficult. In many applications, however, the Gaussianity of the observation is not insured. This has triggered various extensions of the wavelet paradigm depending on the nature of the problem. For instance, in Poisson intensity estimation, ε_{t_1,t_2} are centered Poisson and one has $\mathrm{Var}(X_{t_1,t_2}) = \mathbb{E}(X_{t_1,t_2})$, which is not constant as in the Gaussian case. For such scenarios, Fryzlewicz and Nason [2004] developed a methodology which they coined Haar-Fisz. The idea behind this technique is a normalization inspired from the work of Fisz [1955] which turns out to have an exact closed form when expressed by the scaling or wavelet coefficient of the Haar-wavelet[1]. It was extended to *Poisson* intensity estimation in images Fadili et al. [2003]. Jansen [2006] generalized the Haar-Fisz technique to wavelet-Fisz procedures with arbitrary wavelet families. Finally, Fryzlewicz [2008] synthesized the wavelet-Fisz methodology by

[1] Hence the name Haar-Fisz.

- Showing its utility in non-parametric regression under (7.1) for a large class of variance functions which allows to cover other problems than Poisson intensity estimation such as spectral density estimation.
- Obtaining the minimax rate of convergence for the hard thresholding procedure when the unknown belongs to one-dimensional Besov balls.
- Developing a data-driven extension based on a previous work by Fryzlewicz and Delouille [2005] and showing the consistency of the minimax rate of convergence in this case.

In this chapter, we extend the main theorem of Fryzlewicz [2008] to two-dimensional anisotropic Besov balls using the hyperbolic wavelet transform. Our main motivation for studying this problem is an application in ultrasound imaging which we describe exhaustively in the next chapter. As in many medical imaging techniques, ultrasound images are rich of anisotropic features such as vessels and skin layers. Moreover, for the particular case of wavelet-Fisz denoising, the normalization step needs an accurate local means estimation which can take benefit from filter with anisotropic supports. We keep more discussion about the motivations and the empirical remarks for the next chapter and we focus here on the methodological results.

7.2 Model & Assumptions

In the sequel, we will consider that (7.1) holds where $\{X_{t_1,t_2}\}$ are assumed to be non-negative and independent with

$$\mathbb{E}(X_{t_1,t_2}) = \alpha(t_1/n_1, t_2/n_2) \text{ and } \mathrm{Var}(X_{t_1,t_2}) = \mathrm{Var}(\varepsilon_{t_1,t_2}) = h[\alpha(t_1/n_1, t_2/n_2)] \,. \tag{7.2}$$

We recall that the purpose is to estimate α, assuming that the function h is known. We define, following Neumann and Von Sachs [1997], a discrete version of the definition of BV spaces through TV on $[0,1]^2$.

Definition 7.2.1. *A function $f : [0,1]^2 \longrightarrow \mathbb{R}$ is said to be of bounded variation -belongs to $BV([0,1]^2)$- if :*

$$TV_{[0,1]^2}(f) = \sum_{t_1=1}^{n_1}\sum_{t_2=1}^{n_2}\left|f\left(\frac{t_1}{n_1},\frac{t_2}{n_2}\right) - f\left(\frac{t_1-1}{n_1},\frac{t_2}{n_2}\right) - f\left(\frac{t_1}{n_1},\frac{t_2-1}{n_2}\right) + f\left(\frac{t_1-1}{n_1},\frac{t_2-1}{n_2}\right)\right| < \infty \,.$$

Now, we set the ground for the main result of this chapter by fixing some assumptions on the two quantities contributing to the observation: The unknown α, the variance function h and the noise component ε_{t_1,t_2}.

Assumption 1. *Let the unknown function be such as $\alpha : [0,1]^2 \longrightarrow \mathbb{R}$. We denote $\underline{\alpha} = \inf_{(x,y)\in[0,1]^2} \alpha(x,y)$ and $\overline{\alpha} = \sup_{(x,y)\in[0,1]^2} \alpha(x,y)$. We assume*

(i) $\alpha \in BV([0,1]^2)$.

(ii) $\sup_{u\in[0,1]} TV\alpha(u,\cdot) < \infty$ *and* $\sup_{v\in[0,1]} TV\alpha(\cdot,v) < \infty$

(iii) $0 < \underline{\alpha} \leq \overline{\alpha} < \infty$.

Assumption 2. *The variance function h is defined on $[0,\infty)$ and take values in the same domain. We denote $\underline{h} = \inf_{u\in[\underline{\alpha},\overline{\alpha}]} h(u)$ and $\overline{h} = \sup_{u\in[\underline{\alpha},\overline{\alpha}]} h(u)$. We assume*

(i) $0 < \underline{h} \leq \overline{h} < \infty$.

(ii) h *is non decreasing*

(iii) h *is Lipschitz continuous of order* 1 *on* $[\underline{\alpha},\overline{\alpha}]$ *with constant* L

(iv) There exist $\delta, \overline{\delta}, L_\delta > 0$ *such that* h^δ *is Holder continuous with exponent* $\overline{\delta}$ *and constant* L_δ.

Assumption 3. *There exists some $K > 0$ such that for any $\ell \geq 0$ and all t*

$$\mathbb{E}[|\varepsilon_{t_1,t_2}|^{\ell}] \leq (\ell!)^{1+\gamma} K^{\ell-2} h[\alpha(t_1/n_1, t_2/n_2)] \, .$$

Assumption 1.(i) and **Assumption 1.(ii)** describe the smoothness of α. Note that BV spaces and Besov spaces are related through embedding theorems and motivates the use of wavelet thresholding (*cf.* Chap2). **Assumption 1.(iii)** are natural conditions for a large class of variance functions depending on α. **Assumption 2.(i)** and **Assumption 2.(ii)** are also natural for any variance function while **Assumption 2.(iii)** and **Assumption 2.(iv)** are easily verifiable for many distributions of interest having a variance function h. In particular, Poisson distribution for which $h(u) = u$, or more generally, distributions for which $h(u) = u^{\gamma}$ with $\gamma > 0$. **Assumption 3** allows the use of the wavelet paradigm in non-Gaussian settings. It can be seen as an asymptotic Gaussianity assumption.

Before establishing the theoretical results, we give an overview of the Hyperbolic wavelet-Fisz program for non-parametric estimation under model (7.1).

7.3 Overview of the program

We start by reformulating the model described by equation (7.1) in the wavelet domain by projecting the equation on hyperbolic wavelet basis of $L^2([0,1]^2)$ generated, here again, from a system (ψ, ϕ) . This yields to

$$\widetilde{\mu}_{\mathbf{j},\mathbf{k}} = \mu_{\mathbf{j},\mathbf{k}} + \varepsilon_{\mathbf{j},\mathbf{k}} \, , \tag{7.3}$$

with $\mathbf{j} = (j_1, j_2), \mathbf{k} = (k_1, k_2), j_1 = 0, \cdots, J_1 - 1, j_2 = 0, \cdots, J_2 - 1, k_1 = 1, \cdots, 2^{J_1}, k_2 = 1, \cdots, 2^{J_2}$ where $J_1 = \log_2(n_1), J_2 = \log_2(n_2)$. The $\varepsilon_{\mathbf{j},\mathbf{k}}$ are then independent random variables and we want to recover $\mu_{\mathbf{j},\mathbf{k}}$ from $\widetilde{\mu}_{\mathbf{j},\mathbf{k}}$.

As usual, we separate the paving induced by the couple $(\mathbf{j}, \mathbf{k})$ into two regions in order to calibrate the scale. We fix[2] $\epsilon \in (0,1)$, set $J = J_1 + J_2$ and define J^* such that $J^* = (1-\varepsilon)J$. We then define

$$\mathcal{I}_{\Delta} = \{(j_1, j_2, k_1, k_2),\ 2^{j_1+j_2} \leq 2^{J^*}\} \, .$$

Only coefficients corresponding to $\mathcal{I}_{\Delta}$ are maintained for the synthesis. We perform a thresholding operation on the reminding coefficients, following the classical universal threshold of Donoho and Johnstone [1994] conceived for the case when the errors (ε_{t_1,t_2}) have a Gaussian distribution

$$\lambda_{univ} = \{2\mathrm{Var}(\varepsilon_{\mathbf{j},\mathbf{k}}) \log(\mathrm{Card}(\mathcal{I}_{\Delta}))\}^{1/2}$$

This threshold does not depend on the couple $(\mathbf{j}, \mathbf{k})$. For our case, the noise is non-Gaussian. It does, however, have an asymptotic Gaussianity property given by **Assumption 3**. Moreover, we have (by independence)

$$\mathrm{Var}(\varepsilon_{\mathbf{j},\mathbf{k}}) = \mathrm{Var}\left(\sum_{\mathbf{t}=(t_1,t_2)} \psi_{\mathbf{j},\mathbf{k}-\mathbf{t}} \varepsilon_{\mathbf{t}}\right) = \sum_{\mathbf{t}=(t_1,t_2)} \psi^2_{\mathbf{j},\mathbf{k}-\mathbf{t}} h\{\alpha(t_1/n_1, t_2/n_2)\}.$$

Computing this variance requires pre-estimating the unknown α in the supports of the wavelet basis elements. This can simply be a local means estimate $\kappa_{\mathbf{j},\mathbf{k}}$ for all $(\mathbf{j}, \mathbf{k})$. Thus

$$\widehat{\alpha}(t_1/n_1, t_2/n_2) = \sum_{\mathbf{q}=(q_1,q_2)} \kappa_{\mathbf{j},\mathbf{k}-\mathbf{q}} X_{\mathbf{q}}$$

The properties that $\kappa_{\mathbf{j},\mathbf{k}}$ should verify will be stated in the next section. We can now derive an adaptive threshold depending on the scale and position

[2] The choice of this parameter depends on the regularity of α and will be clearer from the results of the next section.

$$\lambda_{j,k} = h^{1/2} \left(\sum_{\mathbf{q}=(q_1,q_2)} \kappa_{\mathbf{j},\mathbf{k}-\mathbf{q}} X_{\mathbf{q}} \right) \times \{2 \log \mathrm{Card}(\mathcal{I}_\Delta)\}^{1/2} . \tag{7.4}$$

for all $(\mathbf{j},\mathbf{k}) \in \mathcal{I}_\Delta$, coefficients $\widehat{\mu}_{\mathbf{j},\mathbf{k}}$'s that are smaller than $\lambda_{j,k}$ are set to zero[3]. the function reconstructed from the sequence $\widehat{\mu}_{\mathbf{j},\mathbf{k}}$ is the estimator $\widetilde{\alpha}$ of α.

7.4 Mean-square convergence rate

First, we state the assumptions on the constants $\kappa_{\mathbf{j},\mathbf{k}}$.

Assumption 4. *The constant $\kappa_{\mathbf{j},\mathbf{k}} \geq 0$ are such that*

$$\sum_{\mathbf{k}} \kappa_{\mathbf{j},\mathbf{k}} = 1 ,$$

$$\sum_{\mathbf{k}} \kappa^2_{\mathbf{j},\mathbf{k}} \leq C 2^{j_1 - J_1} 2^{j_2 - J_2} ,$$

$$\max_{\mathbf{k}} \kappa_{\mathbf{j},\mathbf{k}} = O(2^{j_1 - J_1} 2^{j_2 - J_2}) ,$$

and

$$\mathrm{supp} \kappa_{\mathbf{j},\cdot} = \mathrm{supp} \psi_{\mathbf{j},\cdot} .$$

for all $(\mathbf{j},\mathbf{k}) \in \mathcal{I}_\Delta$

The set of these constants at a fixed scale $\kappa_{\mathbf{j},\cdot}$ can be seen as local filter. This is of very practical interest as we will see in the subsequent chapter. In fact, a filter which verifies these assumptions is simply given by the scaling operation.

Definition 7.4.1. *Let $p \geq 1$ and $s_1, s_1 > 1/p$. We define the Besov ball*

$$b^{s_1,s_2}_{p,p}(C) = \{\nu = (\nu_{\mathbf{j},\mathbf{k}})_{\mathbf{j},\mathbf{k}}, \sum_{j_1,j_2 \geq 0} 2^{p(j_1\sigma_1 + j_2\sigma_2)} \|\nu_{\mathbf{j},\mathbf{k}}\|_{\ell^p}\}$$

where $\nu_{\mathbf{j},\mathbf{k}} = n_1^{-1/2} n_2^{-1/2} \mu_{\mathbf{j},\mathbf{k}}$, with $\sigma_1 = s_1 + 1/2 - 1/p$, $\sigma_2 = s_2 + 1/2 - 1/p$ and $\|\nu_{\mathbf{j},\mathbf{k}}\|_{\ell^p} = \left(\sum_{k_1,k_2} |\nu_{\mathbf{j},\mathbf{k}}|^p \right)^{1/p}$.

The function α is said to belong to $\mathcal{F}^{(s_1,s_2)}_p$ if and only if $\nu_{\mathbf{j},\mathbf{k}} = n_1^{-1/2} n_2^{-1/2} \mu_{\mathbf{j},\mathbf{k}}$ belongs to $b^{s_1,s_2}_{p,p}(C)$. We can now give the main result of this chapter which is the following mean-square rate of convergence.

Theorem 7.4.1. *Set $\theta(s_1,s_2) = 2s_1s_2/(2s_1s_2 + s_1 + s_2)$. Let $\alpha \in \mathcal{F}^{(s_1,s_2)}_p$ and assume that Assumptions 1–4 hold and the functions ϕ and ψ have at least m vanishing moments. Then*

$$\sup_{\alpha \in \mathcal{F}^{(s_1,s_2)}_p} \mathbb{E}[\|\widetilde{\alpha} - \alpha\|_2^2] = O((\log(n_1 n_2)/(n_1 n_2))^{2\theta(s_1,s_2)})$$

Remark 3. *The rate of convergence given in Theorem 7.4.1 is the best possible over $\mathcal{F}^{(s_1,s_2)}_p$.*

[3] Or by soft-thresholding

7.5 Data-driven extension

In some real applications such as ultrasound imaging, the function h is either completely or partially unknown. Fryzlewicz and Delouille [2005] developed extensions of the wavelet-Fisz algorithm which adapt to models with unknown variance. This was applied, for instance, to the variance stabilization and normalization of one-color microarray data Motakis et al. [2006][4]. Fryzlewicz [2008] has put this in a theorotical persepective by showing that it is possible to conserve an optimal rate of convergence even when h is unknown. We give here our version for the two-dimensional case without prooving its optimality. We expect, however, that as in the one-dimensional case, one should be able to obtain again the rate given in Theroem 7.4.1.

Before estimating h, we need, again, a preliminary estimation of $\alpha(t_1/n_1, t_2/n_2)$. To address this problem, any filter of low computational cost can be used on the noisy image to obtain a pre-estimation $\bar{u}$. We can for example apply a simple mean filter of size $M_1 \times M_2$ to the data

$$\widehat{\alpha}_{t_1,t_2} = \frac{1}{(2M_1+1)(2M_2+1)} \sum_{p_1=t_1-M_1}^{t_1+M_1} \sum_{p_2=t_2-M_2}^{t_2+M_2} X_{p_1,p_2}. \tag{7.5}$$

An estimation of the noise component η is then given by the empirical residuals

$$\widehat{\varepsilon}_{t_1,t_2} = X_{t_1,t_2} - \widehat{\alpha}_{t_1,t_2}$$

To estimate the variance, a kernel smoothing technique is applied to the highly oscillating squared residuals $\widehat{\varepsilon}^2_{t_1,t_2}$. For any vector w with values belonging to $[\underline{h}, \overline{h}]$, the variance estimator predict the value of h on w. First we define

$$W_{n\mathbf{t}}(w) = \frac{1}{n_1 n_2 b} K\left(\frac{\alpha(t_1/n_1, t_2/n_2) - w}{b}\right)$$

with b the bandwidth of the kernel K. We also define

$$\widehat{W}_{n\mathbf{t}}(w) = \frac{1}{n_1 n_2 b} K\left(\frac{\widehat{\alpha}(t_1/n_1, t_2/n_2) - w}{b}\right)$$

The variance estimator is given by

$$\widehat{h}(w) = \frac{\sum_{t_1,t_2} \widehat{W}_{n\mathbf{t}}(w) \widehat{\varepsilon}^2_{t_1,t_2}}{\sum_{t_1,t_2} \widehat{W}_{n\mathbf{t}}(w)}.$$

Thus, the counterpart of the threshold (7.4) is given by

$$\lambda_{j,k} = \widehat{h}^{1/2} \left(\sum_{\mathbf{q}=(q_1,q_2)} \kappa_{\mathbf{j},\mathbf{k}-\mathbf{q}} X_{\mathbf{q}} \right) \times \left\{ 2 \log \mathrm{Card}(\mathcal{I}_\Delta) \right\}^{1/2}. \tag{7.6}$$

The Kernel smoothing step is obviously the exact same as the one from the dimensional case studied by Fryzlewicz and Delouille [2005] and Fryzlewicz [2008]. Actually, this step depends only on h which is, in both cases, a function of $[0,1]$ with values in $\mathbb{R}^+$. The theoretical and empirical statements of this chapter are tested in practice in Chap. We give there further details on the implementation of the Hyperbolic Fisz wavelet program and its data-driven counterpart. We also discuss the performances of the hyperbolic construction compared to the Gaussian noise settings.

[4] An R software package (DDHFm) for this routine is available on the web: `https://cran.r-project.org/web/packages/DDHFm/index.html`

7.6 Appendix

7.6.1 Proof of Theorem 7.4.1

Set $\sigma_{\mathbf{j},\mathbf{k}} = Var(\varepsilon_{\mathbf{j},\mathbf{k}}) = \sum_{\mathbf{t}} \psi^2_{\mathbf{k}-\mathbf{t}} h[\alpha(t_1/n_1, t_2/n_2)]$ and $T = 2^{J_1+J_2}$. Let $\tau > 0$. Recall that

$$\mathcal{I}_T(\tau) = \{(j_1, j_2, k_1, k_2),\ 2^{j_1+j_2} \leq T^{1-\tau}\}\ .$$

Define

$$\mathcal{I}^0_T(\tau) = \{(j_1, j_2, k_1, k_2) \in \mathcal{I}_T,\ (j_1, j_2) \neq (-1, -1)\}$$

We first recall the result for non–random threshold.

Theorem 7.6.1. *Set*

$$\gamma(s_1, s_2, p) = \frac{2s_1s_2 + s_1 + s_2 - 2(s_1 + s_2)/\widetilde{p}}{s_1 + s_2}\ ,$$

with $\widetilde{p} = \min(p, 2)$. *Let* $\tau > 0$ *s.t.*

$$(1 - \tau)\gamma(s_1, s_2, p) \geq \theta(s_1, s_2)\ .$$

Assume that we are given non–random thresholds s.t. for $\gamma_T \to 1$ *and some* $C > 0$

$$\gamma_T \sigma_I \sqrt{2\log(\#\mathcal{I}^0_T(\tau))} \leq \lambda_{I,T} \leq C\sqrt{\log(T)}$$

for any $I \in \mathcal{I}^*_T$, *s.t.* $\mathcal{I}^*_T \subset \mathcal{I}^0_T(\tau)$ *and* $\#(\mathcal{I}^0_T(\tau) \setminus \mathcal{I}^*_T) = O(T^{1-\theta(s_1,s_2)})$ *and that Assumptions 1,3 and 4 hold. Then*

$$\sup_{\alpha \in \mathcal{F}_p^{s_1,s_2}} \mathbb{E}[\|\widetilde{\alpha} - \alpha\|_2^2] = O((\log(T)/T)^{\theta(s_1,s_2)})$$

Remark 4. *Since* $\#(\mathcal{I}^0_T(\tau) \setminus \mathcal{I}^{(LL)}) \sim (J_1 + J_2)2^{J_1 - J_1^* + J_2 - J_2^*}$, *if we set* $J_1^* = (1 - \tau)J_1$ *and* $J_2^* = (1 - \tau)J_2$, *the condition* $\#(\mathcal{I}^0_T(\tau) \setminus \mathcal{I}^*_T) = O(T^{1-\theta(s_1,s_2)})$ *is satisfied if* $1 - \tau > \theta(s_1, s_2)$.

Remark 5. *This Theorem is the bidimensional counterpart of Theorem 4 of Fryzlewicz [2008]. The quantity* $2s/(2s + 1)$ *is replaced by* $\theta(s_1, s_2)$.

Proof. This is Theorem 3.2 of Neumann and Von Sachs [1997]. □

We now consider the case of random threshold. We are given $\mathcal{I}^*_T \subset \mathcal{I}^0_T$ such that $\#(\mathcal{I}^0_T \setminus \mathcal{I}^*_T) = O(T^{1-\theta(s_1,s_2)})$. We then consider random thresholds $\lambda_{I,T}$ with $I \in \mathcal{I}^*_T$. We set

$$\underline{\lambda}_I = \gamma_T h^{1/2} \left[\sum_{\mathbf{q}=(q_1,q_2)} \kappa_{\mathbf{j},\mathbf{k}-\mathbf{q}} \alpha\left(\frac{q_1}{n_1}, \frac{q_2}{n_2}\right) \right] \left(2\log(\#\mathcal{I}^0_T)\right)^{1/2} \text{ and } \overline{\lambda}_I = C\sqrt{\log(T)}\ ,$$

Observe that we have the following lemma

Lemma 7.6.2. *Assume that* $C \geq (2\overline{h})^{1/2}$ *then for any* $\mathbf{j}, \mathbf{k}$, $\sup_{\mathbf{j},\mathbf{k}} \underline{\lambda}_{\mathbf{j},\mathbf{k}} \leq \overline{\lambda}_{\mathbf{j},\mathbf{k}}$.

We now make the two following assumptions :

$$\sum_{I \in \mathcal{I}^*_T} \mathbb{P}(\lambda_{I,T} < \underline{\lambda}_I) = O(T^{-\nu}) \text{ with } \nu > \theta(s_1, s_2)\ . \tag{7.7}$$

$$\sum_{I \in \mathcal{I}^*_T} \mathbb{P}(\lambda_{I,T} > \overline{\lambda}_I) = O(T^{-\theta(s_1,s_2)})\ . \tag{7.8}$$

Theorem 7.6.3. *Let $\tau > 0$ s.t. $(1-\tau)\gamma(s_1,s_2,p) \geq \theta(s_1,s_2)$. Assume that Assumptions 1,3 and 4 hold and that we are given random threshold $(\lambda_{I,T})_I$ satisfying (7.7) and (7.8). In addition assume that*

$$0 < \tau < \frac{\nu - \theta(s_1,s_2)}{\nu+1}\,.$$

Then

$$\sup_{\alpha \in \mathcal{F}_p^{(s_1,s_2)}} \mathbb{E}[\|\widetilde{\alpha} - \alpha\|_2^2] = O((\log(T)/T)^{\theta(s_1,s_2)})$$

Proof. We extend Theorem 6.1 of Neumann [1996]. Observe that

$$(\delta(\widetilde{\mu}_I, \lambda_{I,T}) - \mu_I)^2 \leq \begin{cases} (\widetilde{\mu}_I - \mu_I)^2 + (\delta(\widetilde{\mu}_I, \underline{\lambda}_I) - \mu_I)^2 \text{ if } \lambda_{I,T} < \underline{\lambda}_I \\ (\delta(\widetilde{\mu}_I, \underline{\lambda}_I) - \mu_I)^2 + (\delta(\widetilde{\mu}_I, \overline{\lambda}_I) - \mu_I)^2 \text{ if } \underline{\lambda}_I < \lambda_{I,T} < \overline{\lambda}_I \\ (\delta(\widetilde{\mu}_I, \overline{\lambda}_I) - \mu_I)^2 + (\mu_I)^2 \text{ if } \lambda_{I,T} > \overline{\lambda}_I \end{cases}$$

One now uses that

$$\begin{aligned} \sup_{\alpha \in \mathcal{F}_p^{(s_1,s_2)}} \mathbb{E}[\|\widetilde{\alpha} - \alpha\|_2^2] \quad \leq \quad & \sup_{\alpha \in \mathcal{F}_p^{(s_1,s_2)}} \mathbb{E}[1_{\lambda_{I,T} < \underline{\lambda}_I} \|\widetilde{\alpha} - \alpha\|_2^2] + \sup_{\alpha \in \mathcal{F}_p^{(s_1,s_2)}} \mathbb{E}[1_{\underline{\lambda}_I < \lambda_{I,T} < \overline{\lambda}_I} \|\widetilde{\alpha} - \alpha\|_2^2] \\ & + \sup_{\alpha \in \mathcal{F}_p^{(s_1,s_2)}} \mathbb{E}[1_{\lambda_{I,T} > \overline{\lambda}_I} \|\widetilde{\alpha} - \alpha\|_2^2] \end{aligned}$$

The term $\sup_{\alpha \in \mathcal{F}_p^{(s_1,s_2)}} \mathbb{E}[1_{\underline{\lambda}_I < \lambda_I < \overline{\lambda}_I} \|\widetilde{\alpha} - \alpha\|_2^2]$ is bounded thanks to Theorem 7.6.1. To bound

$$\sup_{\alpha \in \mathcal{F}^{(m_1,m_2)}} \mathbb{E}[1_{\lambda_I < \underline{\lambda}_I} \|\widetilde{\alpha} - \alpha\|_2^2],$$

we see that

$$\mathbb{E}[1_{\lambda_{I,T} < \underline{\lambda}_I} \|\widetilde{\alpha} - \alpha\|_2^2] \leq \sum_I \mathbb{E}[1_{\lambda_{I,T} < \underline{\lambda}_I} (\widetilde{\mu}_I - \mu_I)^2] + \sum_I \mathbb{E}[(\delta(\widetilde{\mu}_I, \underline{\lambda}_I) - \mu_I)^2]$$

By Theorem 7.6.1, the sum $\sum_I \mathbb{E}[(\delta(\widetilde{\mu}_I, \underline{\lambda}_I) - \mu_I)^2]$ is bounded. To bound $\sum_I \mathbb{E}[1_{\lambda_I < \underline{\lambda}_I} (\widetilde{\mu}_I - \mu_I)^2]$ we use Hölder inequality with $p = 1/\tau$, we get that

$$\mathbb{E}[1_{\lambda_I < \underline{\lambda}_I} (\widetilde{\mu}_I - \mu_I)^2] \leq \mathbb{E}[1_{\lambda_I < \underline{\lambda}_I}]^{1-\tau} \mathbb{E}[(\widetilde{\mu}_I - \mu_I)^{2/\tau}]^{\tau} = \mathbb{P}[\lambda_I < \underline{\lambda}_I]^{1-\delta} \mathbb{E}[(\widetilde{\mu}_I - \mu_I)^{2/\tau}]^{\tau}$$

Since by Assumption 3, $\mathbb{E}[(\widetilde{\mu}_I - \mu_I)^{2/\tau}] = \mathbb{E}[|\varepsilon_I|^{2/\tau}]$ is uniformly bounded, one has

$$\begin{aligned} \sum_I \mathbb{E}[1_{\lambda_I < \underline{\lambda}_I} (\widetilde{\mu}_I - \mu_I)^2] \quad & \leq \quad C \sum_I \left(\mathbb{P}[\lambda_I < \underline{\lambda}_I]^{1-\tau}\right) \\ & \leq \quad C \left(\sum_I \mathbb{P}[\lambda_I < \underline{\lambda}_I]\right)^{1-\tau} (\#\mathcal{I}_T)^{\tau} \\ & \leq \quad T^{-\nu_1(1-\tau)+\tau} \end{aligned}$$

If $-\nu(1-\tau) + \tau < -\theta(s_1,s_2)$, that is $0 < \tau < (\nu - \theta(s_1,s_2))/(\nu+1)$, one has $T^{-\nu(1-\tau)+\tau} = o\left(T^{-\theta(s_1,s_2)}\right)$ and the sum $\sum_I \mathbb{E}[1_{\lambda_I < \underline{\lambda}_I} (\widetilde{\mu}_I - \mu_I)^2]$ is negligible.

We now bound $\sup_{\alpha \in \mathcal{F}_p^{(s_1,s_2)}} \mathbb{E}[1_{\lambda_{I,T} > \overline{\lambda}_I} \|\widetilde{\alpha} - \alpha\|_2^2]$. Observe that

$$\mathbb{E}[1_{\lambda_{I,T} > \overline{\lambda}_I} \|\widetilde{\alpha} - \alpha\|_2^2] \leq \sum_I \mathbb{E}[1_{\lambda_{I,T} > \overline{\lambda}_I} (\mu_I)^2] + \sum_I \mathbb{E}[(\delta(\widetilde{\mu}_I, \overline{\lambda}_I) - \mu_I)^2]$$

By Theorem 7.6.1, the sum $\sum_I \mathbb{E}[(\delta(\widetilde{\mu}_I, \overline{\lambda}_I) - \mu_I)^2]$ is bounded. In addition

$$\sum_I \mathbb{E}[1_{\lambda_{I,T} > \overline{\lambda}_I} (\mu_I)^2] \leq \sum_I (\mu_I)^2 \mathbb{P}[\lambda_{I,T} > \overline{\lambda}_I] \leq T^{-\theta(s_1,s_2)} [\sum_I \mu_I^2]$$

the last inequality coming from Assumption (7.8). □

To finish we check that our Fisz threshold satisfy the conditions above and use the following proposition (which is similar to Lemma 2 of Fryzlewicz [2008]). To do so we first prove the following large deviation result

Lemma 7.6.4. *Assume that the constants $\kappa_{\mathbf{j},\tau}$ satisfy Assumption 5. Suppose Assumptions 2 and 3 hold. Then for n_1, n_2 sufficiently large, any $\nu > 0$ and any $(j_1, j_2) \in \mathcal{I}_T(\tau)$*

$$\mathbb{P}\left(\left|\sum_{\mathbf{q}} \kappa_{\mathbf{j},\mathbf{k}-\mathbf{q}}[X_{\mathbf{q}} - \alpha(q_1/n_1, q_2/n_2)]\right| > \delta\right) \leq \delta^{\nu}(n_1 n_2)^{-\nu}$$

Proof. For any $\mathbf{j} = (j_1, j_2)$, $\mathbf{k} = (k_1, k_2)$ and $\mathbf{q} = (q_1, q_2)$ set

$$Z_{\mathbf{j},\mathbf{k},\mathbf{q}} = \frac{\kappa_{\mathbf{j},\mathbf{k}-\mathbf{q}}\left[X_{\mathbf{q}} - \alpha\left(\frac{q_1}{n_1}, \frac{q_2}{n_2}\right)\right]}{\left(\sum_{\mathbf{q}} \kappa^2_{\mathbf{j},\mathbf{k}-\mathbf{q}} \mathbb{E}\left[\varepsilon^2_{\mathbf{q}}\right]\right)^{1/2}}$$

Observe that for $(\mathbf{j}, \mathbf{k})$ fixed the random variables $Z_{\mathbf{j},\mathbf{k},\mathbf{q}}$ are independent, centered and satisfy

$$\sup_{\mathbf{q}} \mathbb{E} \exp(|Z_{\mathbf{j},\mathbf{k},\mathbf{q}}|^p) < \infty$$

for any $p < 1/(1+\gamma)$. Observe that this supremum only depends on the constants $K, \Lambda, \overline{h}$ involved in Assumption 3.

By Assumption 5 and Assumption 2 which implies that h is bounded from above, for any $\mathbf{j}, \mathbf{k}$, one has $\sum_{\mathbf{q}} \kappa^2_{\mathbf{j},\mathbf{k}-\mathbf{q}} \mathbb{E}\left[\varepsilon^2_{\mathbf{q}}\right] = O(2^{j_1 - J_1 + j_2 - J_2})$. Hence by Theorem 7.6.6 recalled in Appendix applied to $Z_{\mathbf{j},\mathbf{k},\mathbf{q}}$ and $y = \delta / \left(\sum_{\mathbf{q}} \kappa^2_{\mathbf{j},\mathbf{k}-\mathbf{q}} \mathbb{E}\left[\varepsilon^2_{\mathbf{q}}\right]\right)^{1/2}$, $n = 2^{j_1 + j_2}$, we obtain that

$$\begin{aligned}
& \mathbb{P}\left(\left|\sum_{\mathbf{q}} \kappa_{\mathbf{j},\mathbf{k}-\mathbf{q}}[X_{\mathbf{q}} - \alpha(q_1/n_1, q_2/n_2)]\right| > \delta\right) \\
= \; & \mathbb{P}\left(\left|\sum_{\mathbf{q}} Z_{\mathbf{j},\mathbf{k},\mathbf{q}}\right| > \delta / \left(\sum_{\mathbf{q}} \kappa^2_{\mathbf{j},\mathbf{k}-\mathbf{q}} \mathbb{E}\left[\varepsilon^2_{\mathbf{q}}\right]\right)^{1/2}\right) \\
\leq \; & \exp(-c_1 \delta^2 \cdot 2^{2(J_1+J_2)-j_1-j_2}) + 2^{j_1+j_2} \exp(-c_2 \delta^p 2^{p(J_1+J_2-j_1-j_2)})
\end{aligned}$$

Since $J_1, J_2 \to \infty$ and $(j_1, j_2) \in \mathcal{I}_T(\tau)$, one easily deduces that for J_1, J_2 sufficiently large and for any $\nu > 0$ one has

$$\mathbb{P}\left(\left|\sum_{\mathbf{q}} \kappa_{\mathbf{j},\mathbf{k}-\mathbf{q}}[X_{\mathbf{q}} - \alpha(q_1/n_1, q_2/n_2)]\right| > \delta\right) \leq \delta^{\nu} 2^{-\nu(J_1+J_2-j_1-j_2)}$$

□

Proposition 7.6.5. *Assume that Assumptions 2,3 and 5 hold. Then, for some well–chosen sequence $\gamma_T \to 1$ from below, the two inequalities (7.7) and (7.8) both hold for any $\nu > 0$.*

Proof. Set

$$\widehat{Y}_{\mathbf{j},\mathbf{k}} = \sum_{\mathbf{q}} \kappa_{\mathbf{j},\mathbf{k}-\mathbf{q}} X_{\mathbf{q}}, \; Y_{\mathbf{j},\mathbf{k}} = \sum_{\mathbf{q}} \kappa_{\mathbf{j},\mathbf{k}-\mathbf{q}} \alpha(q_1/n_1, q_2/n_2),$$

and define $C_T = (2 \log \#\mathcal{I}^0_T)^{1/2} / (2 \log \#\mathcal{I}_T)^{1/2} \to 1^-$ when $T \to \infty$. One then has

$$\begin{aligned}
\mathbb{P}(\lambda_{I,T} < \underline{\lambda}_I) &= \mathbb{P}(h^{1/2}(\widehat{Y}_{\mathbf{j},\mathbf{k}}) < C_T \gamma_T h^{1/2}(Y_{\mathbf{j},\mathbf{k}})) \\
&= \mathbb{P}(h^{\widetilde{\delta}}(\widehat{Y}_{\mathbf{j},\mathbf{k}}) < [C_T \gamma_T]^{2\widetilde{\delta}} h^{\widetilde{\delta}}(Y_{\mathbf{j},\mathbf{k}}))
\end{aligned}$$

We now use the fact that $h^{\widetilde{\delta}}$ is $\overline{\delta}$–Hölder with constant $\widetilde{H}$. We then deduce that

$$\begin{aligned}\mathbb{P}(\lambda_{I,T} < \underline{\lambda}_I) &= \mathbb{P}(h^{1/2}(\widehat{Y}_{\mathbf{j},\mathbf{k}}) < C_T\gamma_T h^{1/2}(Y_{\mathbf{j},\mathbf{k}})) \\ &\leq \mathbb{P}(|\widehat{Y}_{\mathbf{j},\mathbf{k}} - Y_{\mathbf{j},\mathbf{k}}| > \nu_T)\end{aligned}$$

with

$$\Lambda_T = [1 - C_T\gamma_T]^{2\widetilde{\delta}/\overline{\delta}}\widetilde{H}^{-1/\overline{\delta}}\underline{h}^{\widetilde{\delta}/\overline{\delta}}$$

To conclude use Lemma 7.6.4. One then deduces that

$$\mathbb{P}(\lambda_{I,T} < \underline{\lambda}_I) \leq \Lambda_T^{-\nu} 2^{-\nu(J_1 - j_1 + J_2 - j_2)}$$

provided that Λ_T tends logarithmically to 0. We then get inequality (7.7). The proof of inequality (7.8) is similar. □

7.6.2 The Bernstein inequality in the independent case for sub–exponential random variables

We recall know some known results concerning the Bernstein–type inequalities. Let us consider a sequence $X_1, X_2, \cdots$ of centered real valued random variables defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, and set $S_n = X_1 + X_2 + \cdots + X_n$. The classical Bernstein inequality for independent random variables is given by the following

Theorem 7.6.6. *Suppose that the random variables $X_1, X_2, \cdots$ are independent, centered and satisfy*

$$\mathbb{E}|X_i|^\ell \leq \ell!\sigma_i^2 K^{\ell-2}/2 \text{ for all } \ell \geq 2\ . \tag{7.9}$$

Set $\Sigma_n^2 = \sigma_1^2 + \cdots + \sigma_n^2$ and $S_n = \sum_{i=1}^n X_i$. Then

$$\mathbb{P}[S_n > \sqrt{2\Sigma_n^2 x} + Kx] \leq \exp(-x)\ . \tag{7.10}$$

Bibliography

David L. Donoho and Iain M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81: 425–455, 1994.

Jalal M. Fadili, Jérôme Mathieu, Barbara Romaniuk, and Michel Desvignes. Bayesian wavelet-based Poisson intensity estimation of images using the Fisz transformation. In *Proc.International Conference on Image and Signal Processing, Agadir, Morocco*, pages 242–253, 2003.

M. Fisz. The limiting distribution of a function of two independent random variables and its statistical application. *Colloquium Mathematicum*, 3:138–146, 1955.

P. Fryzlewicz. Data-driven wavelet-Fisz methodology for nonparametric function estimation. *Electronic journal of statistics*, 2:863–896, 2008.

Piotr Fryzlewicz and Veronique Delouille. A data-driven Haar-Fisz transform for multiscale variance stabilization. In *IEEE/SP 13th Workshop on Statistical Signal Processing, 2005*, pages 539–544. IEEE, 2005.

P.Z. Fryzlewicz and G.P. Nason. A Haar-Fisz algorithm for Poisson intensity estimation. *Journal of Computational and Graphical Statistics*, 13 (3):621–638, 2004.

Maarten Jansen. Multiscale poisson data smoothing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):27–48, 2006.

ES Motakis, Guy P Nason, Piotr Fryzlewicz, and GA Rutter. Variance stabilization and normalization for one-color microarray data using a data-driven multiscale approach. *Bioinformatics*, 22(20):2547–2553, 2006.

M.H. Neumann. Spectral density estimation via non linear wavelet methods for stationary non–gaussian time series. *Journal of Time Series Analysis*, 17(6):601–633, 1996.

M.H. Neumann and R. Von Sachs. Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra. *The Annals of Statistics*, 25(1):38–76, 1997.

CHAPTER 8

Application to Ultrasound Image denoising

Abstract

We use the techniques presented in chapter 7 to develop an algorithm and its fully data-driven extension for noise reduction in ultrasound imaging. The use of hyperbolic wavelets enables to recover the image while respecting the anisotropic nature of structural details. Experiments on synthetic and real data demonstrate the potential of the proposed algorithm at recovering ultrasound images while preserving tissue details. Furthermore, for this particular wavelet denoising strategy, we show that the results obtained by the hyperbolic construction confirm the theoretical claims. Comparisons with other noise reduction methods show that our method is competitive with the state-of-the art OBNLM filter. Finally, we emphasize the noise model we consider by applying our variance estimation procedure on real images.

8.1 Introduction

ULTRASOUND (US) imaging has been a well-established diagnostic tool in various medical applications for many years. This technology remains one of the least expensive and safest among medical imaging modalities. Nevertheless, the examination and the interpretation of ultrasound images is particularly challenging. This is mainly due to the presence of a particular type of noise called “speckle", which can also be found in similar imaging systems such as Synthetic-Aperture-Radar (SAR) and laser imaging. In ultrasound imaging, acquired signals are adjusted, inside the scanner, prior to display, by a non-linear processing transformation called *log-compression* (*cf.* Kaplan and Ma [1994]). This process aims at enhancing backscatterers in order to facilitate visual understanding. In this chapter, we develop a novel methodology to recover ultrasonic images using a relevant signal-dependent noise model(*cf.* Loupas et al. [1989]) that takes into account the modification of noise characteristics due to the *log-compression*. Adaptations of non-local and variational techniques to this model have already been described in the literature respectively by Coupé et al. [2009] then Jin and Yang [2011].

Although these methods lead to good results in terms of signal-to-noise ratio, they still produce displeasing visual quality, mostly characterized by attenuated sharp edges. In this chapter we present a different strategy which belongs to the wavelet denoising approaches (Donoho and Johnstone [1994]). As in the majority of denoising approaches, the wavelet denoising paradigm relies on the constraining assumption that the noise is additive white Gaussian noise (AWGN). To go beyond this case, we adapt a multiscale variance stabilization technique introduced by Fryzlewicz [2008] in order to make the distribution of wavelet coefficients asymptotically Gaussian with the same variance. We extend this method to hyperbolic wavelets and show how variance stabilization can be easily performed using the low frequency outputs from the wavelet transform at different scales. The motivation behind the use of hyperbolic wavelets is their capacity to provide better estimators than the standard wavelet-tensor construction whenever images contain anisotropic features (see,

for example, the works of Neumann [2000], Autin et al. [2015] and Remenyi et al. [2014]). The notion of anisotropy has been promoted in many works related to ultrasound image denoising (e.g. see Yu and Acton [2002], Krissian et al. [2007] and Ramos-Llorden et al. [2015]) as it often occurs due to the presence of features such as skin layers and vessels. Our algorithm consists of the following steps: (1) compute the wavelet transform of the image, (2) estimate local means by the approximation coefficients of the wavelet transform at each scale, (3) evaluate the variance function for each local mean component. (4) compute the Fisz-transformation of the wavelet coefficients: each coefficient is divided by the estimated local variance in order to stabilize them. (5) hard thresholding: keep the coefficients obtained in step (1) which Fisz-transformed versions have magnitudes larger than a given.

Finally, we show how our approach can be performed in a blind mode, that is, without any prior knowledge of the noise variance. This involves the use of a mean filter for a pre-estimation of the image. The variance function is then estimated using a Nadaraya-Watson estimator.

To validate our methods we present numerical experiments based on synthetic and real data, and a comparative study with the state of the art non-local and variational algorithms. We demonstrate that our data-driven approach performs nearly as well as in situations in which the noise variance is known. Moreover, the measured variance on real ultrasound images confirms the relevance of the noise model we are considering.

The rest of the chapter is organized as follows. A brief overview of different US noise models and dedicated denoising techniques is presented in Section 8.2. In section 8.3, we describe the novel wavelet based algorithm [Farouj et al., 2016] based on the previous chapter. Finally, we provide extensive experimental results and comparisons in Section 8.4.

8.2 Image formation and Related Work

Medical US imaging consists in sending a collection of ultrasound waves from a probe (an array of transducers) inside the body. These waves propagate through different tissues, and get reflected back by the scatterers to the transducers. The echoes are converted back into electrical impulses, and then beam-formed, to give the so-called Radio-Frequency (RF) signals. These signals are then analyzed in order to retrieve the depth and the strength of the echos and thus forming the US image from the amplitudes and the locations of the scatterers. Before display, the RF signals are post-processed. The high frequency carrier is suppressed via a demodulation step (envelope detection). The dynamic range of the obtained signals is, however, too large for the human visual perception. In order to overcome this issue, a process called logarithmic compression is used to enhance backscatterers.

US speckle noise results from the coherent accumulation of individual scattered beams from tissue inhomogeneities. It can be shown that the sum of contributions of these scatterers within a resolution cell is normally distributed (*cf.* Goodman [1975]). Novel techniques emerging in the general image processing community have been continuously adapted to deal with US speckle noise removal. Hereafter, we propose an overview of the main models and techniques we are concerned with.

8.2.1 Multiplicative Noise

An important challenge in developing novel methods for denoising ultrasound images is to find an adequate noise model. One can derive a natural noise model from the knowledge of the statistics of the echo signals. It can be shown that after the demodulation step, the distribution of the magnitude image is no longer Gaussian but rather a Rayleigh distribution [Wagner et al., 1983]. This understanding gave rise to multiplicative noise models similar to those used in SAR imaging. Many filters have been proposed for such a model, including the seminal works by Lee [1980], *et al.* Frost et al. [1982] and Kuan et al. [1985]. Anisotropic diffusion filters of Perona and Malik [1990] have also been successful in US imaging. These include adaptations to account for speckle noise statistics as in the Speckle Reducing Anisotropic Diffusion (SRAD) proposed by Yu and Acton [2002] and its oriented version (OSRAD) of Krissian et al. [2007], and more recently, memory-driven filters of Ramos-Llorden et al. [2015].

8.2.2 Additive Noise

Multiplicative noise models do not take into consideration the logarithmic compression leading to the final US images visualized on scanners. A simple solution is to assume that the signal and the noise are totally distinct. Thus the logarithmic compression step transforms the multiplicative noise into an additive signal-independent model

$$v = u + \varepsilon, \tag{8.1}$$

where v is the observation, u is the unknown image and ε is a random noise component. Wavelet based methods have been considered to deal with such a model depending on the nature of ε. For example, Zong et al. [1998] assumed that ε is a zero-mean Gaussian white noise which leads to AWGN models that are perfectly suited for the classical wavelet thresholding approaches [Donoho and Johnstone, 1994]. Achim et al. [2001] showed that under model (8.1), the wavelet coefficients of the noise component ε, after logarithmic transformations, have non-Gaussian statistics which can be described by some alpha-stable distributions [Samorodnitsky and Taqqu, 1994] and customized the wavelet thresholding for such a situation.

8.2.3 Hybrid Noise

The main drawback of model (8.1) is that it does not take into account the assumption that the noise level is proportional to the underlying image intensity. This assumption is widely used and accepted in Echography. For example, it is the key idea behind motion estimation via speckle tracking [Suhling et al., 2005]. It turns out that the logarithmic compression makes the statistics of ultrasound images deviate from the Rayleigh distribution Tuthill et al. [1988]. For instance, a Fisher-Tippett distribution was used in the work of Slabaugh et al. [2006] to distinguish between tissues in segmentation tasks. A relevant model for ultrasound noise suppression was presented in [Loupas et al., 1989] and assumes that the variance of the noise component is no longer constant but respects the following equation

$$v = u + u^{\gamma}\varepsilon, \tag{8.2}$$

where ε is a zero-mean Gaussian white noise $\varepsilon \sim \mathcal{N}(0, \sigma^2)$, with $\sigma \in (0, \infty)$, and $\gamma > 0$. Model (8.2) seems to be more appropriate as it preserves the signal dependency and has shown been to be effective for speckle modelling [Loupas et al., 1989], as well as motion estimation in US image sequences [Tenbrinck et al., 2013]. This model has the advantage of being general and flexible. Thus, the parameter γ can be adapted to catch the image statistics depending on the post-processing inside the scanner. In this chapter, we develop an appropriate wavelet thresholding method assuming that model (8.2) holds true. Adaptations of two other classical paradigms in denoising, beside wavelet methods, have been studied for model (8.2) [Coupé et al., 2009, Jin and Yang, 2011]. We recall these paradigms:

Nonlocal methods. The non local point of view was initially developed by Buades et al. [2005] leading to the famous N-L means filter. It can be seen as a type of smart average filtering that uses the fact that similar pixels are not necessarily neighbours. Given two pixels, the similarity measure is the Euclidean distance between patches within their respective neighbourhoods. Note that the Euclidean distance is more appropriate in white noise cases [Buades et al., 2005]. In the case of the noise model (8.2), Coupé et al. [2009] presented the OBNLM algorithm in which the *Pearson* distance is used along with an optimized version of the N-L means filter.

Variational methods. The variational approach relies on the minimization of a functional involving a data-fidelity term and a regularity assumption. A common assumption is that images belong to BV (Bounded Variation) spaces, and so Total Variation (TV) is often used for the regularization term [Rudin et al., 1992]. For AWGN models, the fidelity term is simply given as the Euclidean distance between the unknown image and the corrupted one. Adaptations to model (8.2) consists in dividing the fidelity term by the unknown image to the power γ, which is also the standard deviation of the noise. The reader is referred to papers by Rudin et al. [2003] and by Jin for the case $\gamma = 1$ and Jin and Yang [2011] for the case $\gamma = 0.5$. However, the functional cannot be minimized via simple primal dual algorithms [Condat, 2014] as in the AWGN case, a gradient descent is required.

Variance stabilization methods in the image processing literature focuses mainly on Poisson or Poisson-Gaussian noise models, arising in fluorescence microscopy. The Anscombe transform is often used for such a task (see, for example, the works of Makitalo and Foi [2011], Mäkitalo and Foi [2014] and Boulanger et al. [2010]). Zhang et al. [2007] used a multiscale procedure to tackle this problem relying also on a local normalization of wavelet coefficients. To the best of our knowledge, variance stabilization for the noise model (8.2) has never been considered beyond the one-dimensional case. In the following section we propose a technique to adapt the wavelet-based methods to such model. Moreover, we present a data-driven algorithm which solves the problem without prior knowledge of the parameters σ and γ of model (8.2).

8.3 Method

Hyperbolic wavelet bases are unconditional bases for functions in $L^2([0,1]^2)$. They provide sparse representations so that the simple hard thresholding procedure which consists in keeping only coefficients with magnitude larger than a given threshold; setting the others to zero, provide estimators with very good theoretical and practical performances [Neumann, 2000, Autin et al., 2015].

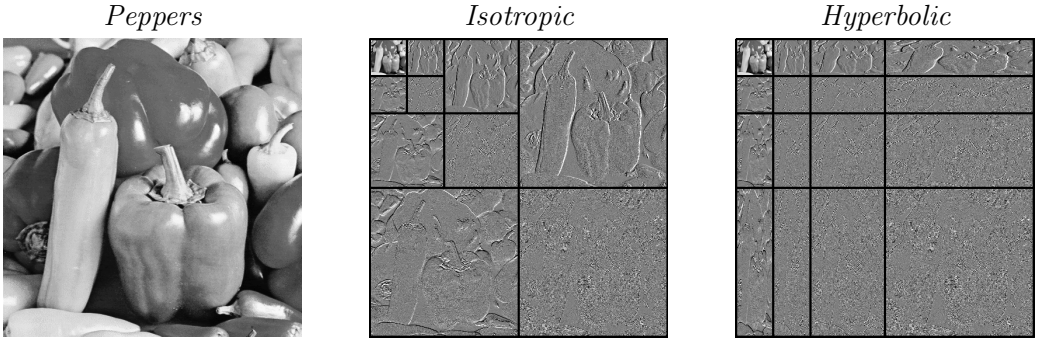


Figure 8.1: Wavelet decomposition in isotropic and hyperbolic settings

8.3.1 Notations

We begin with the hyperbolic wavelet transform (HWT). The starting point is a one-dimensional function ψ, called the mother wavelet, to which one can associate dilated and translated versions $\psi_{j,k}(.) = 2^{j/2}\psi(2^j. - k)$ with $j \geq 0$ and $k \geq 0$. In the same manner, a scaling function φ is defined, along with its dilated and translated versions $\varphi_{j,k}(.) = 2^{j/2}\varphi(2^j. - k)$. Then, the 2D hyperbolic wavelet basis of $L^2([0,1]^2)$ is given by

$$\begin{aligned}
\psi_{j_1,j_2,k_1,k_2}(x_1,x_2) &= \psi_{j_1,k_1}(x_1)\psi_{j_2,k_2}(x_2),\\
\psi_{0,j_2,k_1,k_2}(x_1,x_2) &= \varphi_{0,k_1}(x_1)\psi_{j_2,k_2}(x_2),\\
\psi_{j_1,0,k_1,k_2}(x_1,x_2) &= \psi_{j_1,k_1}(x_1)\varphi_{0,k_2}(x_2),\\
\psi_{0,0,k_1,k_2}(x_1,x_2) &= \varphi_{0,k_1}(x_1)\varphi_{0,k_2}(x_2),
\end{aligned} \tag{8.3}$$

for all $(j_1, j_2) \in \mathbb{N} \times \mathbb{N}$ and $(k_1, k_2) \in \mathbb{Z}^2$. This construction differs from that of the classical two dimensional Discrete Wavelet Transform (DWT), in the sense that different dilation factors are used in each dimension. In the case of the standard 2D DWT, only the cases $j_1 = j_2$ are allowed; therefore the resulting atoms are isotropic.

Let us note $I = \{\underline{j} = (j_1, j_2) \in \mathbb{N}^2$ and $\underline{k} = (k_1, k_2) \in \mathbb{Z}^2\}$ The projection of a function f of $L^2([0,1]^2)$ onto the HWT basis gives a set of hyperbolic wavelet coefficients $\{d_{\underline{j},\underline{k}}\}_{(\underline{j},\underline{k})\in I}$ where

$$d_{\underline{j},\underline{k}}(f) = \langle \psi_{\underline{j},\underline{k}}, f\rangle. \tag{8.4}$$

The set $\{d_{\underline{0},\underline{k}}\}$, where $\underline{0} = (0,0)$, represent the approximation coefficients. In finite discrete settings, a maximal scale is fixed at $J = \log_2(N)$ for an image of size $N \times N$. As the HWT can be seen as a tensor

product of one-dimensional wavelet transforms, the numerical implementation of the HWT can be achieved by applying two successive 1D DWT on each of the two dimensions. Figure 8.1 highlights the difference of the scale-space tilling in the standard and hyperbolic settings. This image will be used in the sequel.

8.3.2 Wavelet denoising

When the noisy observation v verifies model (8.1), the very simple, but powerful, procedure of wavelet thresholding mentioned earlier (cf. chapter 2)can be used. In the wavelet domain, the additive model (8.1) reads

$$d_{\underline{j},\underline{k}}(v) = d_{\underline{j},\underline{k}}(u) + d_{\underline{j},\underline{k}}(\varepsilon), \tag{8.5}$$

with $(\underline{j},\underline{k}) \in I$. The hard thresholding estimator $\hat{u}_\sigma$ is given by

$$\hat{u}_\sigma = \sum_{(\underline{j},\underline{k})\in I_\sigma} d_{\underline{j},\underline{k}}(v)\psi_{\underline{j},\underline{k}}, \tag{8.6}$$

where $I_\sigma = \{(\underline{j},\underline{k}) \in I, \text{ such that } |d_{\underline{j},\underline{k}}(v)| > t(\sigma)\}$ and $t(\sigma)$ is the threshold parameter. Moreover, one of the distinctive features of this procedure is the existence of a *universal threshold* given by

$$\begin{aligned} t(\sigma) &= \Big(2\log(\mathrm{Card}(I))\mathrm{Var}(d_{\underline{j},\underline{k}}(\varepsilon))\Big)^{1/2}, \\ &= \sigma\Big(2\log(N^2)\Big)^{1/2}. \end{aligned} \tag{8.7}$$

In image restoration, we often model the unknown image as an element of anisotropic function spaces,i.e, the regularity parameters are allowed to be different along the different dimensions. This notion of anisotropy is at the heart of multivariate function estimation [Neumann, 2000]. Hyperbolic wavelets are well suited to such situations [Roux et al., 2013]. It has recently been shown by Remenyi et al. [2014], that mixing scales when constructing wavelets, as in (8.3), makes thresholding techniques comparable to state-of-the art denoising algorithms. The choice of the threshold (8.7) is crucial and relies on the fact that the wavelet coefficients are Gaussian and independent. In the next section we show how, in the case of the ultrasound noise model (8.2), this obstacle can be overcome via a wavelet-based variance stabilization technique of chapter 7.

8.3.3 The Proposed Wavelet-Fisz (WF) approach

In (8.2), the noise component is of the form

$$\eta = u^\gamma \varepsilon, \tag{8.8}$$

thus, its variance depends on the unknown image. In order to obtain an adaptive image-dependent threshold we apply the procedure given in chapter 7.

Lemma 8.3.1. *Let $\{\psi_{\underline{j},\underline{k}}\}_{(\underline{j},\underline{k})\in I}$ be a normalized wavelet basis, that is such that $\|\psi\|_2^2 = 1$. Let $u_{\underline{j},\underline{k}}$ denote the restriction of u to the support of the function $\psi_{\underline{j},\underline{k}}$. Assume that, we are given for each $(\underline{j},\underline{k}) \in I$, a constant function $\bar{u}_{\underline{j},\underline{k}}$, converging to $u_{\underline{j},\underline{k}}$, as $j_1, j_2 \to \infty$. Then we have*

$$\Big(\frac{d_{\underline{j},\underline{k}}(\eta)}{\bar{u}^\gamma_{\underline{j},\underline{k}}}\Big)_{\underline{j},\underline{k}} \xrightarrow{d} \mathcal{N}(0,\sigma), \quad \text{as } j_1, j_2 \to \infty\,. \tag{8.9}$$

Since the noise is assumed to be a centered Gaussian random variable, the vector $\Big(\frac{d_{\underline{j},\underline{k}}(\eta)}{\bar{u}^\gamma_{\underline{j},\underline{k}}}\Big)_{\underline{j},\underline{k}}$ is normal with zero mean. In Appendix 8.6, we derive the asymptotic variance when $j_1, j_2 \to \infty$ given in (8.9). Convergence and optimality results were given in the previous chapter. The idea of applying the Gaussianizing

routine to wavelets coefficients was first introduced for *Poisson* intensity estimation by Fryzlewicz and Nason [2004], following a general framework introduced by Fisz [1955]. It was later extended to *Poisson* intensity estimation in images by Fadili et al. [2003]. An approximation $\bar{u}_{\underline{j},\underline{k}}$ of the unknown image u needs to be computed in the support of the function $\psi_{\underline{j},\underline{k}}$. A key point here is our use of the low frequency outputs of the wavelet transform at each scale as local means pre-estimations. These outputs are given by scaling coefficients:

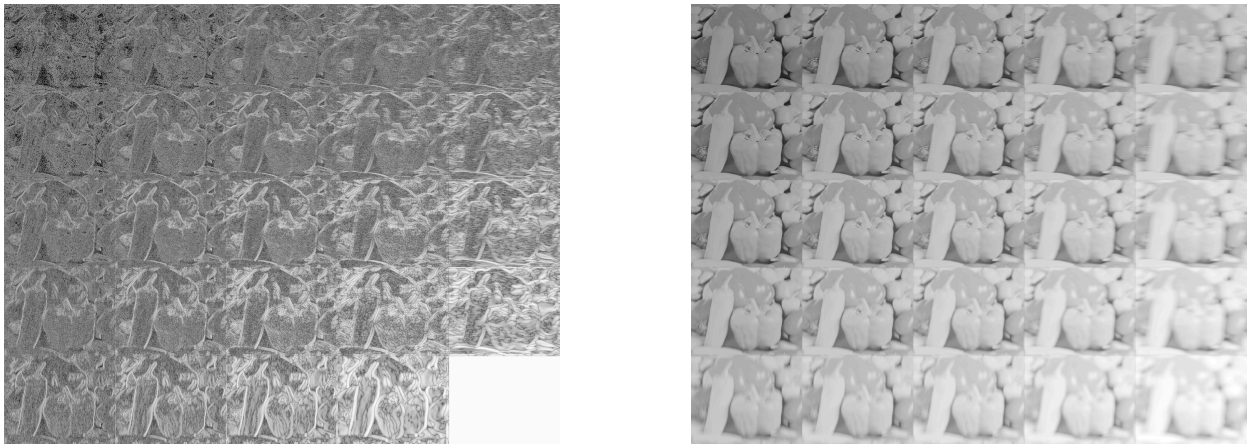
$$c_{\underline{j},\underline{k}}(f) = \langle \varphi_{\underline{j},\underline{k}}, f \rangle, \tag{8.10}$$

where

$$\varphi_{j_1,j_2,k_1,k_2}(x_1,x_2) = \varphi_{j_1,k_1}(x_1)\varphi_{j_2,k_2}(x_2). \tag{8.11}$$

The support of the function $\varphi_{\underline{j},\underline{k}}$ decreases as the value $|\underline{j}| = j_1 + j_2$ increases. As a consequence of the law of large numbers, the local means approximation (8.10) becomes less accurate. This issue has limited consequences since, following Fryzlewicz [2008], we consider only the coarsest scales up to a certain level $|\underline{j}| \leq J_{max}$. Not much information is lost since the finest scales consist of high-frequency components which are essentially noise. Using the lemma 8.3.1, we can now define a new set for the construction of the nonlinear estimator (8.6) given by

$$\tilde{I_\sigma} = \{(\underline{j},\underline{k}) \in I, \text{ s.t } |\underline{j}| \leq J_{max}; \frac{|d_{\underline{j},\underline{k}}(v)|}{\sigma\, c_{\underline{j},\underline{k}}(v)^\gamma} > t(1)\}. \tag{8.12}$$



(a) (b)

Figure 8.2: Outputs of the hyperbolic NDWT of the “Peppers" image: (a) The set of wavelet coefficients $\{d_{\underline{j},\underline{k}}\}_{\underline{j},\underline{k}}$ and (b) the set of approximation coefficients $\{c_{\underline{j},\underline{k}}\}_{\underline{j},\underline{k}}$

Implementation: The presented WF technique can be performed using the non-decimated wavelets transform (NDWT) introduced in [Coifman and Donoho, 1995]. The wavelets coefficients magnitudes (8.4) and the approximation coefficients (8.10) for the NDWT are presented in Figure 8.2. It has been shown that the denoising methods based on NDWT outperforms those based on traditional (decimated) wavelets in terms of mean-squared error (MSE) and signal-to-noise ratio (SNR) [Nason and Silverman, 1995]. This is mainly due to its translation invariance. However, the non-decimated wavelet coefficients are, in general, correlated even if the noise is uncorrelated. The choice of relevant wavelet coefficients becomes a correlated multiple hypothesis testing problem. Thus, the choice of the threshold (8.7) can lead to non-optimal results. In practice, one can consider the non-decimated wavelet coefficients as separate packets of uncorrelated coefficients [Nason, 2010]. The universal threshold can then be applied to each packet. The pseudo code for the routine is given in Algorithm 4.

The key step in this algorithm is the stabilization technique leading to the set $(s_{\underline{j},\underline{k}})_{(\underline{j},\underline{k})\in I}$. Figure 8.3 shows how the wavelet coefficients are stabilized after the WF procedure (with $\sigma = 2$ and $\gamma = 0.5$). We rescaled the wavelet coefficients magnitudes between 0 and 1 and we fit a normal distribution. The *Liver*

Algorithm 4 WF algorithm

Input: $f, \sigma, \gamma, J_{max}$
Output: Estimate $\tilde{u}$

```
[d_{j,k}, c_{j,k}] ⟵ NDWT(f)
for each couple (j, k) do
    if |j| > J_max then d_{j,k} = 0
    else
        p_{j,k} = σ × (c_{j,k})^γ
        s_{j,k} = |d_{j,k}| / p_{j,k}
        if s_{j,k} < t(1) then d_{j,k} = 0
        end if
    end if
end for
ũ = INDWT(d_{j,k})
```

image represents a section of a human liver along with the portal vein. The diagonal details of the wavelet transform at the first thresholding scale are examined. At fine scales, the wavelet transformation retrieves, mainly, the noise component. We can clearly see that the distribution of the wavelet coefficients deviates from the Gaussian distribution. This phenomenon can be explained by model (8.2) as the noise is perturbed by the image statistics. In fact, it was observed that the statistics of the wavelet coefficients of an image are more likely to follow distributions with heavier tails than a *Gaussian* one, such as *Exponential* and *Laplacian* distributions [Mallat, 1999, Jaffard, 2004]. Note that the non-Gaussianity of the wavelet coefficients distribution in US images was first observed by Achim et al. [2001]. In this latter work, the authors assumed that the noise has an alpha-stable distribution.

8.3.4 Fully data-driven extension

Besides the fact that there is no conventional noise model in ultrasound imaging, different authors may use different parametrizations for a given noise model. In particular, for our model of interest, different values for the parameters σ and γ are given in [Coupé et al., 2009, Jin and Yang, 2011] and Rudin et al. [2003]. A point of debate is whether a large value should be used for γ and a small one for σ or vice versa. We sidestep the problem by estimating the standard deviation of the noise directly from the data. Here, we follow the work of Fryzlewicz and Delouille [2005] who developed extensions of the wavelet-Fisz algorithm which adapt to models with unknown variance. This was applied, for instance, to the variance stabilization and normalization of one-color microarray data by Motakis et al. [2006].

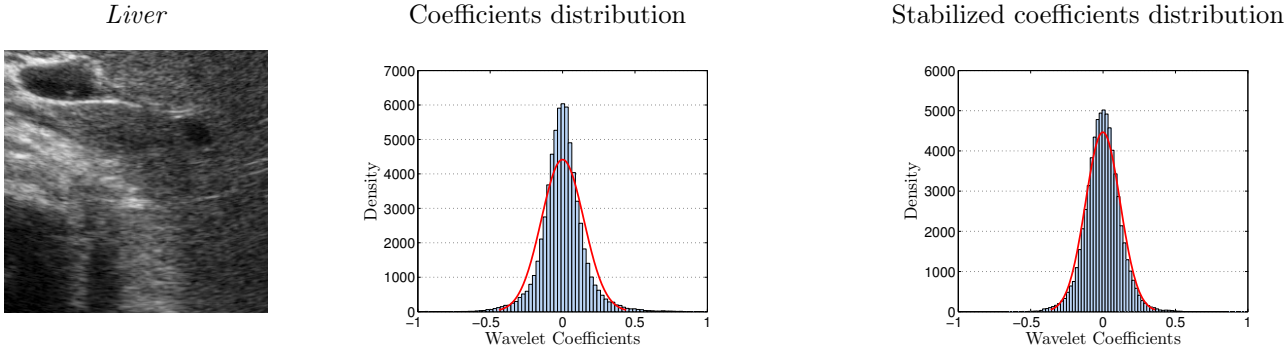


Figure 8.3: Wavelet decomposition of US images: Statistics of the diagonal details at the finest scale

Standard deviation estimation

To address this problem, any filter of low computational cost can be used on the noisy image to obtain a pre-estimation $\bar{u}$. We applied a simple mean filter of size M to the image. An estimation of the noise component η is then given by the residual

$$\hat{\eta}(\bar{u}) = v - \bar{u} \tag{8.13}$$

To estimate the variance, a kernel smoothing technique is applied to the highly oscillating squared residuals $\hat{\eta}^2$. For any vector w with values belonging to $[min(\bar{u}), max(\bar{u})]$, the variance estimator of w is given as

$$h(w) = \widehat{\mathrm{Var}(\eta(w))} = \frac{\langle \hat{W}_b(w), \hat{\eta}^2 \rangle}{\hat{W}_b(w)}, \tag{8.14}$$

where W is defined as

$$\hat{W}_b(w) = \frac{1}{N^2 b} K\Big(\frac{\bar{u} - w}{b}\Big), \tag{8.15}$$

with b the bandwidth of the kernel K. This regression technique is called the Nadaraya-Watson estimation. Under the assumption that the variance of the noise is a positive power of the image intensity, as suggested by model (8.2), it is natural to constrain the estimator of the variance to be non-decreasing. This can be done using the so-called isotonic regression [Fryzlewicz et al., 2007], which consists in finding the closest non-decreasing function, in term of least mean square error, using a "pool-adjacent-violators" algorithm [Mair et al., 2009]. We present an example of such a routine on a corrupted 512×512 *Peppers* image. Our choice of the image *Peppers* is motivated by the fact that it has many variations in grey values, resulting in an interval of intensities well covered by the vector w. The global regularity of the image is the principle criteria for the choice of the the size M of the average filter. In fact, a compromise must be done; M should be chosen as the largest possible with respect to the homogeneity of the image. One is obliged to use small values for M if the image has many discontinuities. We investigated various choices for the size M of the average filter. We found that a value of $M = 12$ gives a reliable pre-estimation of the image. In general we recommend the use of this value for images with a moderate number of discontinuities such as the *Peppers* image. Naturally, M depends also on the resolution of the image; this dependence is expected to be linear. The bandwidth b has less influence on the estimator than M. This is due to the regression step which corrects remaining oscillations. This was also pointed by Fryzlewicz Fryzlewicz [2008]. A value of $b = 3$ was found out to be stable. Through the rest of the chapter, we fix this value and tune only M. The results of two experiments with different values of γ and σ are given in Figure 8.4. As the image pixel values range from 0 to 255, we simply choose w to be a uniform discretization of $[0, 255]$. The results confirm the reliability of the standard deviation estimator $h^{1/2}$ in comparison to the ground truth.

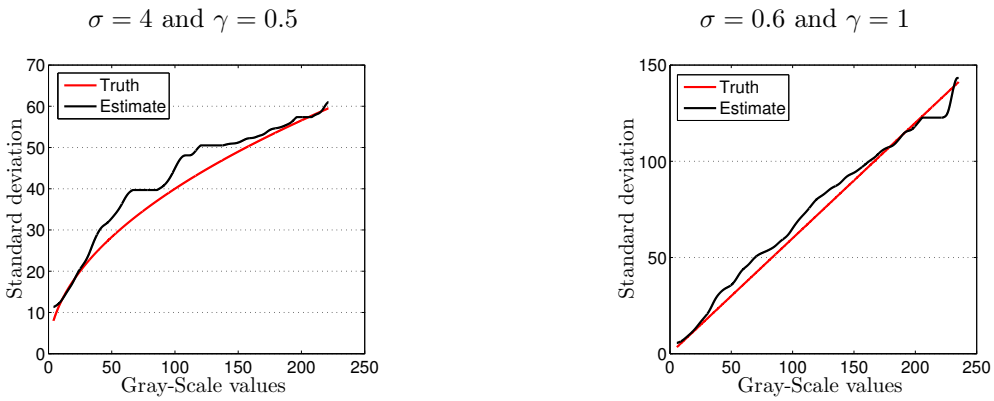


Figure 8.4: Standard deviation estimation from experiments on the "*Peppers*" image for different values of σ and γ.

Blind denoising

We describe the adaptation of the WF algorithm to the fully data-driven methodology. For our noise model, the standard deviation estimator presented in 8.3.4 gives the following approximation

$$h^{1/2}(w) \approx \sigma w^{\gamma}, \tag{8.16}$$

hence, we have a similar result to the one given in lemma 8.3.1

$$\left(\frac{d_{\underline{j},\underline{k}}(\eta)}{h^{1/2}(c_{\underline{j},\underline{k}})}\right)_{\underline{j},\underline{k}} \xrightarrow{d} \mathcal{N}(0,1), \tag{8.17}$$

In Algorithm 4, the parameters σ and γ appear in the auxiliary step 5 for the computation of the variance in the supports of the wavelets. In order to have a data-driven version of this algorithm, the knowledge σ and γ should not be required as inputs. Equation (8.17) suggests to replace step 5 in Algorithm 4 by

$$p_{\underline{j},\underline{k}} = h^{1/2}(c_{\underline{j},\underline{k}}), \tag{8.18}$$

8.4 Experiments and Discussion

In this section we present some experiments to evaluate the performance of our WF method. In order to distinguish the different contributions of our work, we divide this section into two parts. First, we compare the performance of the WF method for both Isotropic (IWF) and Hyperbolic (HWF) constructions, and then we show the potential of the data-driven extension.

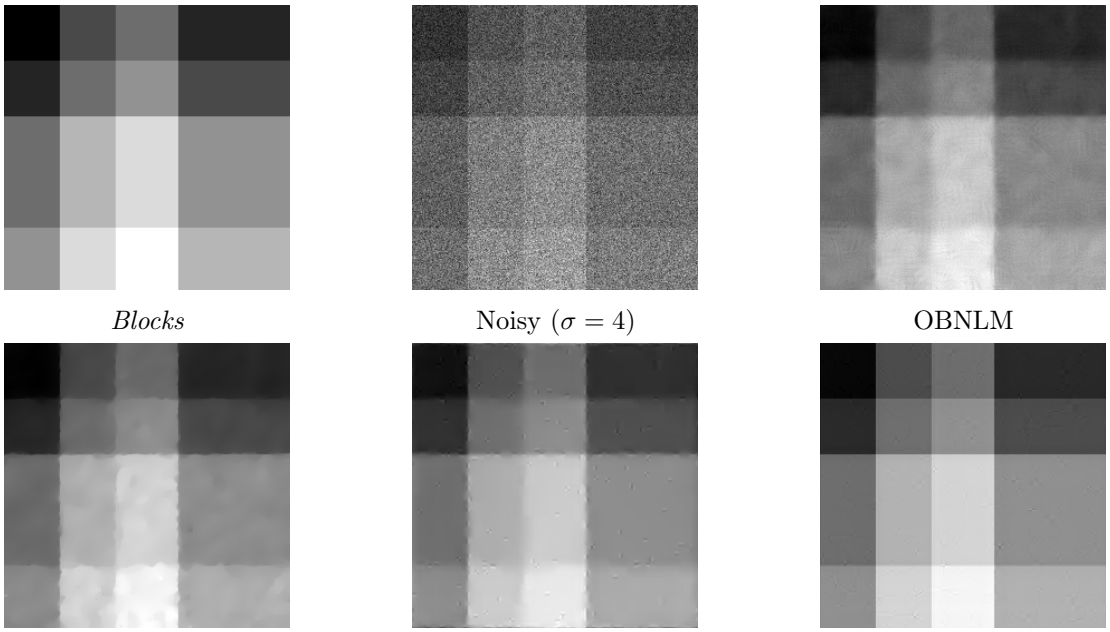
8.4.1 The WF method

We compare our results to those obtained using two other approaches that considers the noise model given by (8.2). The OBNLM *filter* has proven to be very effective in speckle noise reduction and to perform better than classical filters [Coupé et al., 2009]. The variational approach of Jin and Yang [2011] is an adaptation of the well-established TV denoising to model (8.2). The criteria used for the comparisons were the classical *Peak Signal to Noise Ratio* (PSNR), and the *Structural Similarity Index Measure* (SSIM) [Wang et al., 2004] which allowed evaluation of the tissue structure preservation. As the great majority of ultrasound imaging is not concerned with functional studies, the preservation of morphological information while performing denoising is normally more important than preserving the true measured pixel intensity. We also show the difference between the true image and the denoised result of every method. This is known in the literature as the *method noise* [Buades et al., 2005]. One expect to retrieve more noise in areas of high pixel intensities according to model (8.2). The OBNLM *filter* is available on the web[1]. The parameters α and M controlling the number of blocks and the size of the search window are fixed at 3 and 6 as in the original paper, and the filtering parameter h was optimized for different levels of noise. The variational algorithm was implemented with the gradient descent step fixed at 0.2 as suggested by the authors. We use *Haar* wavelets for the WF method. The scaling function associated to these wavelets behave like a simple mean filter which gives a reliable set of approximation coefficients $\{c_{\underline{j},\underline{k}}\}_{\underline{j},\underline{k}}$. These wavelets are also efficient at preserving discontinuities. We are aware, however, that these wavelets do not lead to optimal results in terms of PSNR and its is possible to improve the results using wavelets from other families such as Daubechies or Coiflets. In all experiments, the coefficients corresponding to the first finest scale are truncated.

Experiments on synthetic data

Our two experiments were performed by adding synthetic noise to clean images. We fix $\gamma = 0.5$, as in the papers of Coupé et al. [2009] and Jin and Yang [2011]. The first image *Blocks* aims only at demonstrating the ability of hyperbolic wavelets to deal with highly anisotropic images. In fact, this image is an additive model when the regularities in the two space dimensions are distinct. This is a highly anisotropic case where

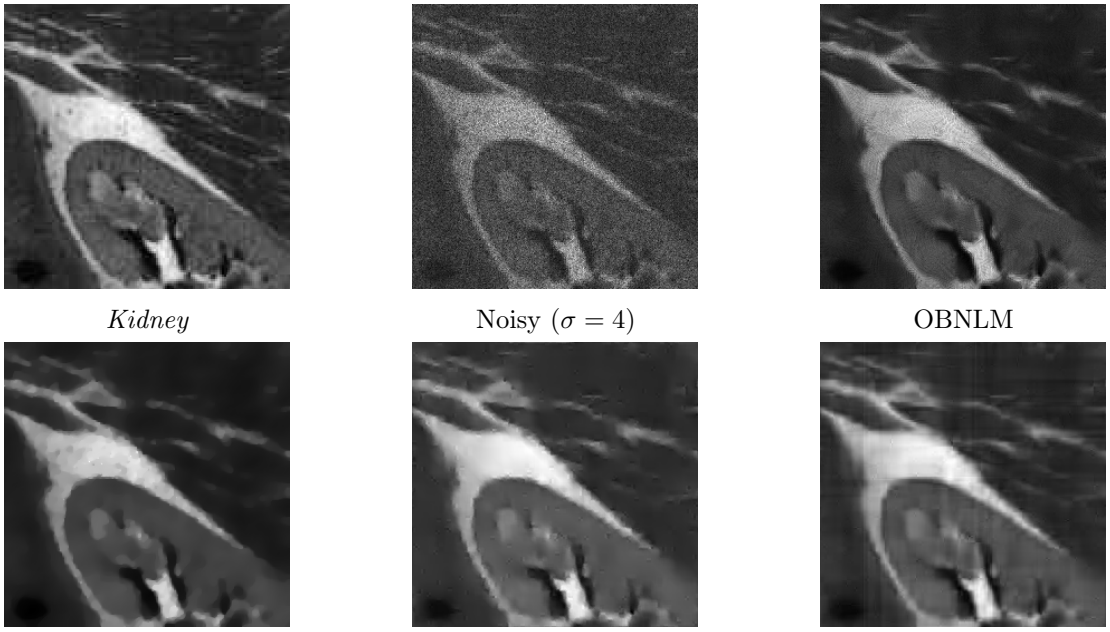
[1] `https://sites.google.com/site/pierrickcoupe`



Blocks | Noisy ($\sigma = 4$) | OBNLM

Variational | IWF | HWF

Figure 8.5: Results of various methods applied to the image *Blocks*. Quantitative evaluation is given in Table 8.1.



Kidney | Noisy ($\sigma = 4$) | OBNLM

Variational | IWF | HWF

Figure 8.6: Results of various methods applied to the image *Kidney*. Quantitative evaluation is given in Table 8.1.

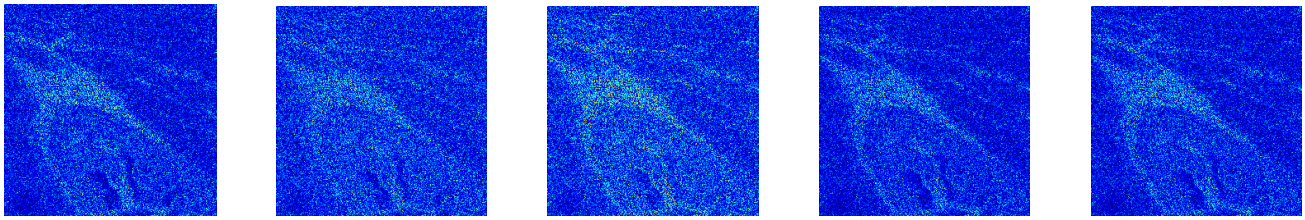
the hyperbolic setting gives optimal results [Autin et al., 2015]. The second *Kidney* image is a CT image taken from FIELD II's website[2]. This latter example is challenging to denoise because of the presence of many gray level variations.

Table 8.1 outlines the performances of the different methods with their optimal parameters calibration. In the case of the image *Blocks*, the contribution of the hyperbolic setting is clearly visible in terms of PSNR,

[2] `http://field-ii.dk`

	Blocks			*Kidney*		
	PSNR (dB)	SSIM	Parameters	PSNR (dB)	SSIM	Parameters
Noisy($\sigma = 2$)	22.83	0.135	–	24.83	0.430	–
OBNLM	35.13	0.917	h=1.5	30.05	0.845	h=1
Variational	37.98	0.972	n_{iter}=180	28.72	0.814	n_{iter}=160
IWF	35.78	0.958	–	29.04	0.837	–
HWF	**49.65**	**0.993**	–	**30.24**	**0.866**	–
Noisy($\sigma = 3$)	20.95	0.071	–	22.30	0.272	–
OBNLM	32.86	0.836	h=2	**28.71**	0.752	h=1
Variational	35.57	0.955	n_{iter}=260	27.91	0.799	n_{iter}=180
IWF	33.75	0.929	–	27.39	0.791	–
HWF	**46.65**	**0.987**	–	28.20	**0.822**	–
Noisy($\sigma = 4$)	19.31	0.044	–	20.83	0.187	–
OBNLM	31.38	0.739	h=2.5	27.81	0.765	h=2
Variational	34.41	0.946	n_{iter}=350	**28.03**	0.782	n_{iter}=210
IWF	32.40	0.909	–	26.63	0.764	–
HWF	**43.04**	**0.973**	–	27.31	**0.791**	–

Table 8.1: Quantitative comparison (PSNR & SSIM), and optimal parameters for different methods applied to the *Blocks* and *Kidney* images with different noise levels.



Original OBNLM Variational IWF HWF

Figure 8.7: The method noise [Buades et al., 2005] of the various approaches applied to the *Kidney* image ($\sigma = 3$).

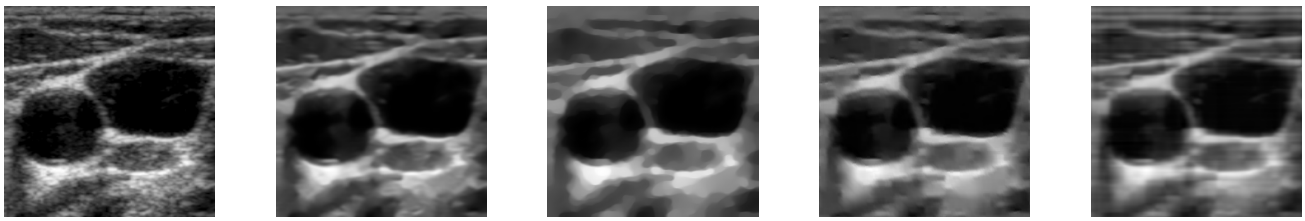


Image OBNLM Variational IWF HWF

Figure 8.8: Visual evaluation of various methods applied to the *Carotid-Thyroid* image.

SSIM and visual quality (e.g. Figure 8.5). The hyperbolic Wavelet-Fisz thresholding gives the best results for all noise levels. The variational approach also gives good results because the image is a piece-wise constant. However, it suffers from over-blurring effects around the edges. Artifacts due to the patching process are clearly visible in the OBNLM *filter* results. For the image *Kidney*, our approach fails at outperforming the OBNLM *filter* and the variational method in terms of PSNR, but remains competitive. This can be explained by the different philosophy of wavelet thresholding methods more oriented toward the complete elimination of the noise rather than the minimizing the MSE [Donoho and Johnstone, 1994]. The OBNLM approach

gives good results when the noise level is low. Conversely, the variational method was more appreciable for high noise levels. As can be observed in Figure 8.6, the wavelet approaches efficiently preserved the structure. Unfortunately, we note the presence of artifacts related to the supports of the wavelet basis. This is a common inconvenience of wavelet thresholding methods. The *method noise* presented in Figure 8.7 shows the structure of the suppressed noise for each method. It can be observed that, for all methods, the removed noise component has higher values in areas of high intensities in coherence with the noise model (8.2). Moreover, the proposed method gives a good compromise between efficient suppression of the noise and preservation of structures; in the sense that the different regions are easily distinguishable. A major advantage of wavelet thresholding is its adaptability; the threshold comes directly from the knowledge of σ and γ. Tuning the OBNLM *filter* is less straightforward, as the algorithm parameters are not directly expressed in terms of the model parameters.

Experiments on real data

We propose to evaluate our algorithm on some samples from real US imaging. Unfortunately, the blind extension of the SSIM presented by Kong et al. [2013] was not suitable here as the noise is signal-dependent. Therefore comparisons and parameters tuning were entirely based on the visual quality of the resulting image. We applied the different set of parameters in Table 8.1 and choose those giving the best results. The first test concerns the *Carotid-Thyroid* image. In this image shows a human carotid artery. Speckle can be seen on the left due to blood flow, while the thyroid gland is visible on the right. Denoising such images may be a pre-processing step in segmentation of the thyroid gland. An enhanced image also eases the tracking of the carotid artery wall in dynamic imaging. The second test concerns Cranial US. This technique is mostly used for babies, before the cranial bones have closed, as US waves cannot pass through the skull. An example of its use is obtaining information on complications in premature birth.

Figure 8.8 shows results obtained using the different algorithms applied to the *Carotid-Thyroid* image. The results obtained using the variational method are clearly over-blurred; this is due to the piece-wise constancy constraint of total variation. The OBNLM *filter* achieved good results , although, there was some visible partitioning in the image. The proposed method gives results with well-defined structures thanks the local treatment of the wavelets paradigm. Moreover, in the hyperbolic case, one can see that the horizontal structures are nicely recovered. The main artifact with the proposed method is the occurrence of wavelet basis atoms in the final image. Figure 8.9 reveals how these artifacts can be drastically reduced when the hyperbolic wavelet is used. In the image obtained using the IWF procedure, small regions representing the supports of the *Haar* basis can be seen. These are similar to the artifacts related to patching that occurred using the Nonlocal methods. An improved result is obtained using the Hyperbolic settings, even though some lines are still visible.

Image IWF HWF

Figure 8.9: Visual comparison between IWF and HWF for the *Brain* image

8.4.2 The Data-driven WF method

In this section, our experiments reveal the potential of the data-driven extension of our algorithm. The Table 8.2 presents a comparison of the results obtained using HWF and data-driven HWF (dHWF) for the "*Peppers*" image studied in Figure 8.4. As expected, there is a loss, proportional to the noise level, in the PSNR and the SSIM up to 0.5 dB and 4% respectively. We believe this loss is acceptable, especially when the noise level is not very high.

	PSNR (dB)			SSIM		
σ	2	3	4	2	3	4
Noisy	20.21	16.65	14.16	0.35	0.23	0.16
HWF	29.25	27.65	26.44	0.78	0.75	0.73
dHWF	29.09	27.40	26.17	0.76	0.72	0.69

Table 8.2: Denoising of "*Peppers*" image: Quantitative Comparison (PSNR & SSIM) of the HWF and its fully data-driven version for different noise levels.

We applied this data-driven technique to the 256×256 *Liver* image studied in Figure 8.3. This image has a few discontinuities, allowing the use of a large window for the mean filter. Here, we use a window of size $M = 8$. The experiments were performed using a PC DELL Latitude E6430 with an Intel Core i7-3740QM CPU, 2.7 GHZ processor and 8 GB of RAM under Fedora 20, using MATLAB v.8.2.0.701, 64-bit. The recorded running time for HWT was 40.72 s while it was 54.51 s for dHWT. The difference in timings is due to the different routines of the variance estimation step. Results are given in Figure 8.10. A first interesting result is the "non constant slope" of the estimated standard deviation. This points out that model (8.1) cannot be used. We suspect that this function is proportional to the power of the image, thereby presenting an image processing based evidence of the noise model (8.2) relevance directly from the data. It was also noted that the set of wavelet coefficients was well stabilized. We compared the image obtained in the data-driven mode to the one obtained by using an exhaustive search for the parameter σ with $\gamma = \{0.5; 1\}$. Results demonstrated that the data-driven result is satisfying and less blurred.

8.5 Conclusion

In this paper, we described a novel approach for denoising ultrasound images based on wavelet thresholding, variance stabilization and the use of hyperbolic wavelet basis. Quantitative and visual results show the potential of the proposed method and the merit of the hyperbolic settings. A data-driven extension of our method is also presented. When applied to real data, this extension provides evidence of the relevance of the noise model. We also believe that a method free of tuning requirements is very desirable, especially for physicians. The extension to three-dimensional wavelets can be used for two purposes; 3D denoising or (2D+t) dynamic US denoising. While the 3D case is straightforward, the dynamic US case must be handled carefully as the noise variance depends only on the spatial dimension. Thus, the variance stabilization and the local means approximation should be performed only on the spatial variables. We are currently addressing this issue.

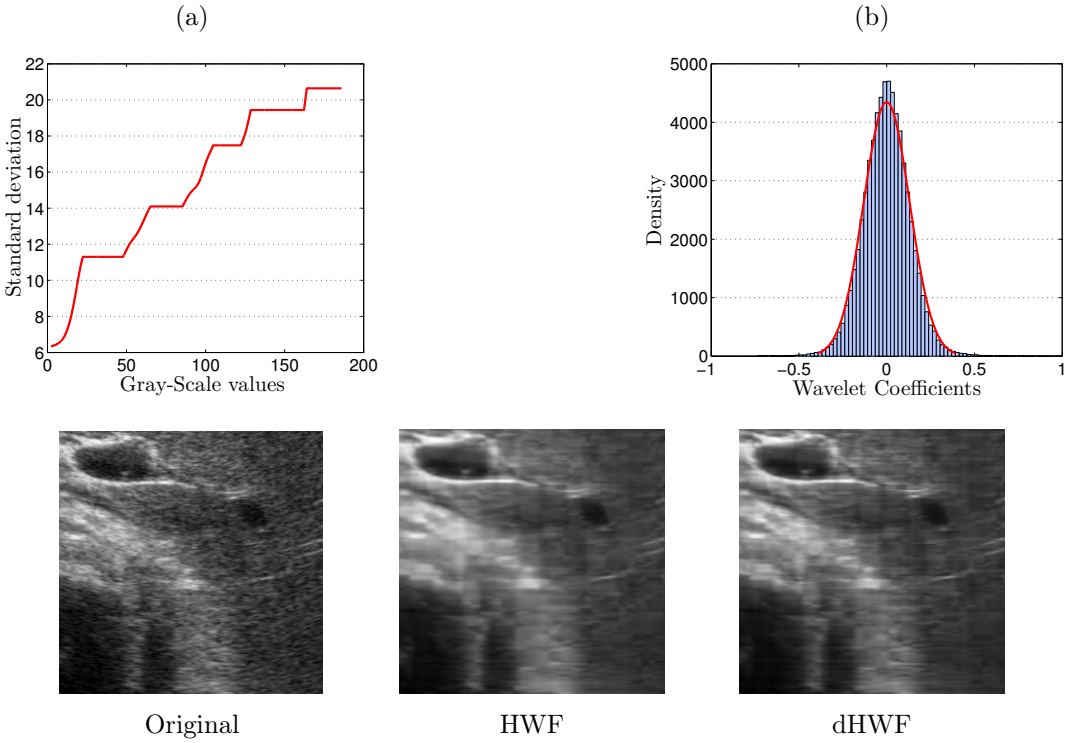


Figure 8.10: Experiments on the *Liver* image: (a) Estimated standard deviation, (b) Blind stabilized Coefficients.

8.6 Variance derivation

We recall that $\eta = u^\gamma \varepsilon$, and we note $t = (t_1, t_2)$

$$\begin{aligned}
\mathrm{Var}\Big(\frac{d_{\underline{j},\underline{k}}(\eta)}{\bar{u}^\gamma_{\underline{j},\underline{k}}}\Big) &= \frac{1}{\bar{u}^{2\gamma}_{\underline{j},\underline{k}}}\mathrm{Var}\Big(\sum_t \psi_{\underline{j},\underline{k}}(t)u^\gamma(t)\varepsilon(t)\Big),\\
&= \frac{\sigma^2}{\bar{u}^{2\gamma}_{\underline{j},\underline{k}}}\sum_t \psi^2_{\underline{j},\underline{k}}(t)u^{2\gamma}(t),\\
&= \frac{\sigma^2}{\bar{u}^{2\gamma}_{\underline{j},\underline{k}}}\sum_t \psi^2_{\underline{j},\underline{k}}(t)u^{2\gamma}_{\underline{j},\underline{k}}(t).\\
&= \sigma^2\sum_t \frac{u^{2\gamma}_{\underline{j},\underline{k}}(t)}{\bar{u}^{2\gamma}_{\underline{j},\underline{k}}}\psi^2_{\underline{j},\underline{k}}(t).
\end{aligned}$$

Finally, since when $j_1, j_2 \to \infty$, $\bar{u}_{\underline{j},\underline{k}}$ converges to $u_{\underline{j},\underline{k}}$, then for each t in the support of $\psi_{\underline{j},\underline{k}}$, we have

$$\lim_{\bar{u}_{\underline{j},\underline{k}}\to u_{\underline{j},\underline{k}}} \frac{u^{2\gamma}_{\underline{j},\underline{k}}(t)}{\bar{u}^{2\gamma}_{\underline{j},\underline{k}}} = 1.$$

Thus

$$\mathrm{Var}\Big(\frac{d_{\underline{j},\underline{k}}(\eta)}{\bar{u}^\gamma_{\underline{j},\underline{k}}}\Big) = \sigma^2\sum_t \psi^2_{\underline{j},\underline{k}}(t) = \sigma^2||\psi_{\underline{j},\underline{k}}||_2^2 = \sigma^2.$$

8.7 Discussion

From a practical point of view, the two-dimensional hyperbolic construction seems to be more convincing in the wavelet-Fisz settings than in the classical Gaussian noise settings. This can be explained by the accuracy of the anisotropic multiscale variance stabilization compared to the isotropic one. Note, also, that it is possible to perform wavelet-Fisz denoising within a variational framework using the ℓ_1-norm of the wavelet-Fisz coefficients as a regularizer. This allows to take into account other types of deterioration due to ultrasound imaging systems such as blurring, by constructing appropriate data-fidelity terms. In the following chapter we give some perspectives and orientations related to the wavelet-Fisz methodology.

Bibliography

Alin Achim, Anastasios Bezerianos, and Panagiotis Tsakalides. Novel Bayesian multiscale method for speckle removal in medical ultrasound images. *IEEE Trans. Med. Imaging*, 20(8):772–783, 2001.

F. Autin, G. Claeskens, and J.M. Freyermuth. Asymptotic performance of projection estimators in standard and hyperbolic wavelet bases. *Electronic journal of statistics*, 9(2):1852–1883, 2015.

Jérôme Boulanger, Charles Kervrann, Patrick Bouthemy, Peter Elbau, Jean-Baptiste Sibarita, and Jean Salamero. Patch-based nonlocal functional for denoising fluorescence microscopy image sequences. *IEEE transactions on medical imaging*, 29(2):442–454, 2010.

Antoni Buades, Bartomeu Coll, and Jean-Michel Morel. A non-local algorithm for image denoising. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 60–65, 2005.

R. R. Coifman and D. L. Donoho. Translation-invariant denoising. *in Antoniadis & Oppenheim, Wavelets and statistics. Lecture Notes in Statistics*, pages 125–150, 1995.

Laurent Condat. A generic proximal algorithm for convex optimization: application to total variation minimization. *Signal Processing Letters, IEEE*, 21(8):985–989, 2014.

P. Coupé, P. Hellier, C. Kervrann, and C. Barillot. Nonlocal means-based speckle filtering for ultrasound images. *IEEE Transactions on Image Processing*, 18(7):2221–2229, 2009.

David L. Donoho and Iain M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81: 425–455, 1994.

Jalal M. Fadili, Jérôme Mathieu, Barbara Romaniuk, and Michel Desvignes. Bayesian wavelet-based Poisson intensity estimation of images using the Fisz transformation. In *Proc.International Conference on Image and Signal Processing, Agadir, Morocco*, pages 242–253, 2003.

Younes Farouj, Jean-Marc Freyermuth, Laurent Navarro, Marianne Clausel, and Philippe Delacharte. Hyperbolic wavelet-fisz denoising for a model arising in ultrasound imaging. *IEEE Trans. Computational Imag. (in minor revision)*, 2016.

M. Fisz. The limiting distribution of a function of two independent random variables and its statistical application. *Colloquium Mathematicum*, 3:138–146, 1955.

Victor S. Frost, Josephine Abbott Stiles, K. S. Shanmugan, and Julian C. Holtzman. A model for radar images and its application to adaptive digital filtering of multiplicative noise. *IEEE Trans. Pattern Anal. Mach. Intell.*, 4(2):157–166, 1982.

P. Fryzlewicz. Data-driven wavelet-Fisz methodology for nonparametric function estimation. *Electronic journal of statistics*, 2:863–896, 2008.

Piotr Fryzlewicz and Veronique Delouille. A data-driven Haar-Fisz transform for multiscale variance stabilization. In *IEEE/SP 13th Workshop on Statistical Signal Processing, 2005*, pages 539–544. IEEE, 2005.

Piotr Fryzlewicz, Véronique Delouille, and Guy P Nason. GOES-8 X-ray sensor variance stabilization using the multiscale data-driven Haar-Fisz transform. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 56(1):99–116, 2007.

P.Z. Fryzlewicz and G.P. Nason. A Haar-Fisz algorithm for Poisson intensity estimation. *Journal of Computational and Graphical Statistics*, 13 (3):621–638, 2004.

Joseph W Goodman. Statistical properties of laser speckle patterns. In *Laser speckle and related phenomena*, pages 9–75. Springer, 1975.

Stéphane Jaffard. Beyond Besov spaces part 1: Distributions of wavelet coefficients. *Journal of Fourier Analysis and Applications*, 10(3):221–246, 2004.

Zhengmeng Jin and Xiaoping Yang. A variational model to remove the multiplicative noise in ultrasound images. *Journal of Mathematical Imaging and Vision*, 39(1):62–74, 2011.

D. Kaplan and Q. Ma. On the statistical characteristics of the log-compressed Rayleigh signals: theoretical formulation and experimental results. *J. Acoust. Soc. Am.*, 95:1396–1400, 1994.

Xiangfei Kong, Kuan Li, Qingxiong Yang, Liu Wenyin, and Ming-Hsuan Yang. A new image quality metric for image auto-denoising. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 2888–2895. IEEE, 2013.

K. Krissian, C. F. Westin, R. Kikinis, and K. G. Vosburgh. Oriented speckle reducing anisotropic diffusion. *Image Processing, IEEE Transactions on*, 16:1412–1424, 2007.

Darwin T. Kuan, Alexander A. Sawchuk, Timothy C. Strand, and Pierre Chavel. Adaptive noise smoothing filter for images with signal-dependent noise. *IEEE Trans. Pattern Anal. Mach. Intell.*, 7(2):165–177, 1985.

Jong-Sen Lee. Digital image enhancement and noise filtering by use of local statistics. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2(2):165–168, 1980.

T. Loupas, W.N. McDicken, and P.L. Allan. An adaptive weighted median filter for speckle suppression in medical ultrasonic images. *Circuits and Systems*, 36:129–135, 1989.

Patrick Mair, Kurt Hornik, and Jan de Leeuw. Isotone optimization in r: pool-adjacent-violators algorithm (pava) and active set methods. *Journal of statistical software*, 32(5):1–24, 2009.

Markku Makitalo and Alessandro Foi. A closed-form approximation of the exact unbiased inverse of the anscombe variance-stabilizing transformation. *IEEE transactions on image processing*, 20(9):2697–2698, 2011.

Markku Mäkitalo and Alessandro Foi. Noise parameter mismatch in variance stabilization, with an application to poisson–gaussian noise estimation. *IEEE Transactions on Image Processing*, 23(12):5348–5359, 2014.

S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 1999. ISBN 978-0-521-11913-9.

ES Motakis, Guy P Nason, Piotr Fryzlewicz, and GA Rutter. Variance stabilization and normalization for one-color microarray data using a data-driven multiscale approach. *Bioinformatics*, 22(20):2547–2553, 2006.

G. P. Nason and B. W. Silverman. The stationary wavelet transform and some statistical applications. In *Lecture Notes in Statistics*, pages 281–300. Springer-Verlag, 1995.

Guy Nason. *Wavelet methods in statistics with R*. Springer Science & Business Media, 2010.

Michael H. Neumann. Multivariate wavelet thresholding in anisotropic function spaces. *Statistica Sinica*, 10: 399–431, 2000.

P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.*, 12(7):629–639, 1990.

Gabriel Ramos-Llorden, Gonzalo Vegas-Sanchez-Ferrero, Marcos Martin-Fernandez, Carlos Alberola-López, and Santiago Aja-Fernández. Anisotropic diffusion filter with memory based on speckle statistics for ultrasound images. *IEEE Transactions on image processing*, 24(1):345–358, 2015. doi: http://dx.doi.org/10.1109/TIP.2014.2371244.

Norbert Remenyi, Orietta Nicolis, Guy Nason, and Brani Vidakovic. Image denoising with 2d scale-mixing complex wavelet transforms. *Image Processing, IEEE Transactions on*, 23(12):5165–5174, 2014.

Stephane G Roux, Marianne Clausel, Beatrice Vedel, Stéphane Jaffard, and Patrice Abry. Self-similar anisotropic texture analysis: The hyperbolic wavelet transform contribution. *Image Processing, IEEE Transactions on*, 22(11):4353–4363, 2013.

L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

L. Rudin, P. L. Lions, and S. Osher. *Multiplicative denoising and deblurring: theory and algorithms*. Geometric Level Set Methods in Imaging, Vision, and Graphics. S. Osher and N. Paragios, Eds., 2003.

Gennady Samorodnitsky and Murad S. Taqqu. *Stable non-Gaussian random processes : stochastic models with infinite variance*. Stochastic modeling. Chapman & Hall, New York, 1994.

Greg Slabaugh, Gozde Unal, Tong Fang, and Michael Wels. Ultrasound-specific segmentation via decorrelation and statistical region-based active contours. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 1, pages 45–53. IEEE, 2006.

Michael Suhling, Muthuvel Arigovindan, Christian Jansen, Patrick Hunziker, and Michael Unser. Myocardial motion analysis from b-mode echocardiograms. *Image Processing, IEEE Transactions on*, 14(4):525–536, 2005.

Daniel Tenbrinck, Sönke Schmid, Xiaoyi Jiang, Klaus P. Schäfers, and Jörg Stypmann. Histogram-based optical flow for motion estimation in ultrasound imaging. *Journal of Mathematical Imaging and Vision*, 47(1-2):138–150, 2013.

T.A. Tuthill, R.H. Sperry, and K.J. Parker. Deviation from Rayleigh statistics in ultrasonic speckle. *Ultrason. Imag.*, pages 81–90, 1988.

Robert F Wagner, Stephen W Smith, John M Sandrik, and Hector Lopez. Statistics of speckle in ultrasound b-scans. *IEEE Transactions on sonics and ultrasonics*, 30(3):156–163, 1983.

Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *Image Processing, IEEE Transactions on*, 13(4):600–612, 2004.

Yongjian Yu and Scott T. Acton. Speckle reducing anisotropic diffusion. *Image Processing, IEEE Transactions on*, 11(11):1260–1270, 2002.

Bo Zhang, MJ Fadili, J-L Starck, and J-C Olivo-Marin. Multiscale variance-stabilizing transform for mixed-poisson-gaussian processes and its applications in bioimaging. In *2007 IEEE International Conference on Image Processing*, volume 6, pages VI–233. IEEE, 2007.

X. Zong, A. F. Laine, and E. A. Geiser. Speckle reduction and contrast enhancement of echocardiograms via multiscale nonlinear processing. *IEEE Trans. Med. Imag.*, page 532–540, Aug. 1998.

CHAPTER 9

Perspectives

Abstract

We give in this chapter some orientations for future works on the wavelet-Fisz and its applications.

9.1 Theoretical extensions

A natural extension of the theoretical results of Chapter 7 is the derivation of rates of convergences for the data-driven wavelet-Fisz method. We expect that, as in the one dimensional case of Fryzlewicz [2008], the rate of convergence will be optimal. The proof should evolve of showing that the order of error made on the pre-estimation $\widehat{\alpha}$ is negligible compared to the order of the error made on $\widetilde{\alpha}$. More precisely, a Bernstein-type inequality such as the one given in theorem 7.6.6 must be established for the error $\sum_{t_1,t_2} |\alpha_{t_1,t_2} - \widehat{\alpha}_{t_1,t_2}|$.

Another interesting extension of the results of Fryzlewicz [2008] -in one or several dimensions- is to derive rates of convergence in the presence of weakly-depended noise components [Neumann, 2013, Gannaz and Wintenberger, 2010]. This raises the awkward question of how to extend concentration inequalities and large deviation results to such settings.

9.2 The wavelet-Fisz-Galerkin method

Consider that the observation obeys the following linear model

$$y_\varepsilon = K\, f + \varepsilon, \tag{9.1}$$

where K is a linear operator. When $\mathrm{Var}(\varepsilon)$ is constant and for certain classes of compact operators K, the wavelet-Galerkin method can be used for function estimation. This method was developed by Cohen et al. [2004] for one-dimensional function estimation. It relies on the fact that many compact operators with fast decay have sparse approximations in the wavelet domain [Beylkin et al., 1991]. The general Galerkin method consists in projecting (9.1) on a finite dimensional subspace E of $L^2(\mathbb{R})$. The linear estimator $\widetilde{f} \in E$ is, then, the solution to

$$\langle K\widetilde{f}, v\rangle = \langle y_\varepsilon, v\rangle \quad \text{for all} \quad v \in E \tag{9.2}$$

In the wavelet-Galerkin framework, the subspace E is considered to be the linear span W_J. This yields to a system of the form

$$K_J F_J = Y_{\varepsilon,J} \tag{9.3}$$

where K_J is the so-called Galerkin *stiffness matrix*. Finding F_J requires here a simple matrix inversion of the *stiffness matrix*. This can also be done in a nonlinear fashion by adding a thresholding operator $\mathcal{T}_\varepsilon$ on $Y_{\varepsilon,J}$. The system to solve becomes

$$K_J F_J = \mathcal{I}_\varepsilon(Y_{\varepsilon,J}). \tag{9.4}$$

Projection on wavelet spaces is particularly interesting for shift-variant operators[1]. For instance, Chang et al. [2000] used the wavelet characterization of compact operators to model foveation in images, while Malgouyres [2002] used it for image-deblurring.

For our application in the previous chapter, a convolution operator can be considered to model ultasound images while taking into account the blurring inherent to the acquisition [Shin et al., 2009]. Moreover, the point-spread function (PSF) considered in ultrasound is varies with the depth of the image (shift-variant) with different characteristics along axial and lateral directions (anisotropic) which encourages the use of hyperbolic wavelets. In fact, in the same manner as for functions, anisotropic operators have better approximation in hyperbolic (tensor-product) wavelets [Reich, 2010]. In order to do joint deblurring-denoising in ultrasound, an idea would be to include the Fisz variance stabilization step in the linear system (9.4). If we denote by $\mathcal{F}$ the stabilization operation and $\mathcal{F}^{-1}$ its inverse, then one can solve

$$K_J F_J = \mathcal{F}^{-1}(\mathcal{T}_\varepsilon(\mathcal{F}(Y_{\varepsilon,J}))). \tag{9.5}$$

This can provide a nice Galerkin wavelet-Fisz framework for solving problems requiring both variance stabilization and the inversion of an linear operator.

9.3 Wigner-Ville distribution smoothing

In this section we discuss a still open problem of adapting the wavelet-Fisz methodology for the problem smoothing Wigner-Ville distributions. This problem was pointed to us by R. von Sachs and it is well suited for the hyperbolic framework because of the different regularities along the two (space/time) dimensions

The study of non-stationary signals through their Wigner-Ville (WV) distributions [Martin and Flandrin, 1985] is a fundamental time-frequency (TF) analysis tool which proved its utility over many years in numerous applications ranging from speech processing, music, geophysics and bio-engineering. Besides its ability to provide a compact characterization of the TF plan compared to classical spectrograms or scalograms, it also enjoys many interesting mathematical properties such as energy conservation. For a given function $x(t)$ of $L^2(\mathbb{R})$, the WV is given by

$$W_x(t,f) = \int_{-\infty}^{\infty} x(t+\tau/2)x^*(t-\tau/2)e^{-2j\pi f\tau}\,d\tau.$$

The main drawback of this representation comes from its quadratic nature. In fact, cross interference terms appears whenever the signal is multi-component. In an additional model, one expects

$$W_{x+z}(t,f) = W_x(t,f) + W_z(t,f) + 2Re\{W_{x,z}(t,f)\},$$

with

$$W_{x,z}(t,f) = \int_{-\infty}^{\infty} x(t+\tau/2)z^*(t-\tau/2)e^{-2j\pi f\tau}\,d\tau.$$

This limitation makes the interpretation very challenging. When the second component is a random process, the denoising problem of retrieving $W_x(t,f)$ from $W_{x+z}(t,f)$ is particularly challenging as the modeling of the noise component in the WV distribution is not straightforward. The dominating paradigm used to tackle this problem is linear smoothing; convolution using a kernel function K

$$C_x(t,f) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} W_x(u,v)K(t-u,f-v)\,dudv.$$

The resulting smoothed WD distributions form the so-called Cohen's class [Cohen, 1989]. Such linear denoising techniques often lead to over-smoothing with the result that some of the original signal components

[1] In opposition to the Fourier transform which diagonlize shit-invariant operators.

might be suppressed. This motivated the use of non-linear techniques such as the one proposed by Baraniuk [1994]. Such techniques use the simple process of thresholding (*Soft* or *Hard*) in the wavelet domain to suppress the noise component. However, these methods are justified when the noise is white Gaussian, which is clearly not the case here; the heteroscedastic nature of the noise require an adaptation of the denoising technique. To enhance the relationship between this problem noise and the wavelet-Fisz methodology, we derive the noise:

In the sequel z will refer to a Gaussian white noise $z \sim \mathcal{N}(0, \sigma)$ with $\sigma \in (0, \infty)$. The noise component in the WV distribution is then given by $W_z(t, f) + 2Re\{W_{x,z}(t, f)\}$. The first term $W_z(t, f)$ is the L^2-product of two random variables, while the second term is the L^2-product of a Gaussian white random variable and a deterministic function. One expect that the variance of $W_z(t, f)$ is of the order of σ^2, while the variance of $Re\{W_{x,z}(t, f)\}$ is of the order of $\sigma^2 x^2$. When the signal-to-noise ratio (SNR) is higher than 1 everywhere, the second term is always dominating. Hence, an alternative noise model is given by

$$W_x(t, f) = W_x(t, f) + 2Re\{W_{x,z}(t, f)\}. \tag{9.6}$$

Applying the wavelet-Fisz methodology to the model (9.6) is not straightforward because of the ambiguous relationship between $W_x(t, f)$ and $Re\{W_{x,z}(t, f)\}$. This however deserves more investigation.

Bibliography

Richard G Baraniuk. Wigner-ville spectrum estimation via wavelet soft-thresholding. In *IEEE-SP International Symposium on Time-frequency and Time-scale Analysis*, pages 452–455, 1994.

Gregory Beylkin, Ronald Coifman, and Vladimir Rokhlin. Fast wavelet transforms and numerical algorithms i. *Communications on pure and applied mathematics*, 44(2):141–183, 1991.

Ee-Chien Chang, Stephane Mallat, and Chee Yap. Wavelet foveation. *Applied and Computational Harmonic Analysis*, 9(3):312–335, 2000.

Albert Cohen, Marc Hoffmann, and Markus Reiss. Adaptive wavelet galerkin methods for linear inverse problems. *SIAM Journal on Numerical Analysis*, 42(4):1479–1501, 2004.

Leon Cohen. Time-frequency distributions-a review. *Proceedings of the IEEE*, 77(7):941–981, 1989.

P. Fryzlewicz. Data-driven wavelet-Fisz methodology for nonparametric function estimation. *Electronic journal of statistics*, 2:863–896, 2008.

Irène Gannaz and Olivier Wintenberger. Adaptive density estimation under weak dependence. *ESAIM: Probability and Statistics*, 14:151–172, 2010.

François Malgouyres. A framework for image deblurring using wavelet packet bases. *Applied and Computational Harmonic Analysis*, 12(3):309–331, 2002.

Wolfgang Martin and Patrick Flandrin. Wigner-ville spectral analysis of nonstationary processes. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 33(6):1461–1470, 1985.

Michael H Neumann. A central limit theorem for triangular arrays of weakly dependent random variables, with applications in statistics. *ESAIM: Probability and Statistics*, 17:120–134, 2013.

Nils Reich. Wavelet compression of anisotropic integrodifferential operators on sparse tensor product spaces. *ESAIM: Mathematical Modelling and Numerical Analysis*, 44(1):33–73, 2010.

Ho-Chul Shin, Richard Prager, James Ng, Henry Gomersall, Nick Kingsbury, Graham Treece, and Andrew Gee. Sensitivity to point-spread function parameters in medical ultrasound image deconvolution. *Ultrasonics*, 49(3):344–357, 2009.

Part III

Towards Fully Data-driven fMRI spatio-temporal deconvolution

CHAPTER 10

fMRI: Review on Data analysis techniques & statement of the deconvolution problem

Abstract

fMRI offers the possibility to retrieve brain activity through the BOLD signal. Analyzing such signals is, however, carried out in challenging conditions as one has to deal with highly noisy time courses. While traditional BOLD signal processing is based on linear regressors that need a pre-modeling of the signal dynamics, the quest for exploring spontaneous activity gave rise to spatio-temporal deconvolution techniques which are free of timing or duration priors on the activity. On the other hand, with the increasing interest in the study of resting state networks and the high variability of the anatomical brain structure across subjects, it is also natural to develop methods that are not promoting particular spatial partitioning in the brain and to keep these methods as data-driven as possible. In this chapter, We put in review some of the fundamental works on fMRI data analysis before stating the problem of fMRI deconvolution we will consider in the next chapter.

10.1 Introduction

Since its appearance in early 1990's, functional magnetic resonance imaging (fMRI) has remained the key tool for the understanding of human brain by giving new insights on its functioning when responding to certain tasks or during rest. This technique measures, in a non-invasive way, the fluctuations in blood oxygenation. This results in the so-called blood oxygenation level-dependent (BOLD) signals; the concentrations of blood oxygenation are expected to increase in regions evolving in the neuronal activity. The aim of fMRI analysis is to detect sufficient evidence of the presence of activation in the BOLD signal. Traditional detection algorithms are conceived to fit the time courses of each voxel to the experimental paradigm. The voxels are supposed to be active when the corresponding time courses demonstrate significant correlation with the timing and the duration of the stimuli. A familiar strategy in this sense is the use of the general linear model (GLM) [Friston et al., 1998] for the regression. Here, the BOLD signal is modeled as the response of a linear shift-invariant system - having the hemodynamic response function (HRF) as an impulse response - to the experimental excitation. The obtained statistical parametric map (SPM) is then thresholded, using a multiple hypothesis testing procedure, in order to decide which tests (i.e voxels) are task related. Although such regression techniques have been predominating in fMRI analysis schemes, they are limited when it comes to studying spontaneous activity (e.g resting state activity). In fact, the absence of explicit tasks makes it hard to find meaningful inputs to drive the fitting. This has motivated the introduction of new techniques that do not require an explicit modeling of the BOLD signal dynamics. Methods such as the fuzzy clustering algorithms [Baumgartner et al., 2000, Fadili et al., 2000], wavelet filtering [Wink and Roerdink, 2004], seed correlation analysis [Biswal et al., 1995], principal component analysis PCA, or independent

component analysis (ICA) [Beckmann et al., 2005] aim at retrieving the main signal components from the corrupted BOLD signal. On the other hand, temporal correlation analysis (TCA) [Liu et al., 2000] and its extensions [Morgan et al., 2008] are inspired from dimension reduction. However, these methods have two main drawbacks: (1) they do not incorporate the prior knowledge about the hemodynamics or about the regularity of the activity-driven signal, (2) they rely either on spatial or temporal coherence and do not take into account the whole the spatio-temporal structure of the data. The first issue can be handled, naturally, by thinking of the activity-inducing signal reconstruction as a temporal deconvolution problem. A seminal work in this sense is due to Glover [1999] who introduced Wiener deconvolution filtering for task-related fMRI. Recently, this was customized for resting-state fMRI by Wu et al. [2013]. Still, these methods are not optimized for the specific fMRI application. Countering this limitation has been following general trends in signal processing. First, the emergence of variational techniques inspired spatial regularization terms for the activation maps with techniques such as the regularised GLM of Flandin and Penny [2007]. This techniques are still, however, task-related as it relies on a GLM. Then, the fuss about l_1 regularization in the 2000's also had its influence on this problem. Gaudes et al. [2011] and Gaudes et al. [2013] use l_1 regularization terms directly on the signal to promote spikes of activation. Khalidov et al. [2007, 2011] use a nice generalization of wavelets called "activelets" which acts like a differential operator and thus enables to take the HRF into account when sparsifying the signal. Following this tendency, Karahanoğlu et al. [2013] presented total activation, an attractive combination of the Generalized Total Variation of Karahanoğlu et al. [2011] allowing hemodynamic modelling and smoothness priors inside brain anatomical regions. Although, this work is a big step towards data driven deconvolution while leveraging information on the spatio-temporal structure and the hemodynamics of fMRI data, it still uses anatomical atlases as priors for the spatial coherence of the activation. This prior can be misleading for three main reasons. First, there is no conventional functional atlasing for the human brain; a bewildering variety of conventions on anatomical atlases in available in the literature. Second, these anatomical atlases are, anyway, created with the philosophy that they cluster regions that have the same functional behavior. This might be a pessimistic constraint for capturing the richness and complexity of brain activity, particular in resting-state fMRI. Finally, anatomical atlases are still generic and are far from being subject specific. Pursuing this reasoning, we suggest in the present work the use of strategies that are not driven by anatomical priors. More precisely, we show the merit of using a spatial regularization based on a weighted total variation. The weights can be either a binary matrix having non-zero entries for voxels that are likely to belong to the gray matter, or can be computed from a similarity measure derived from a posterior probability map [Friston and Penny, 2003]. Combining this with the generalized total variation for temporal deconvolution results in a minimization problem evolving the sum of two ℓ_1-norms. We tackle this problem by using the generalized forward-backward algorithm to find a global minima. Besides the fact that the proposed method does not require any information on timing, duration or pre-defined spatial structure, it also has a lower complexity than the original TA thanks to the global definition of total variation. In fact, there is no need it for a loop on anatomical regions and TV minimization algorithms can be performed with simple vector-wise operations.

The rest of the chapter is organized as follows. In the next section we recall the BOLD signal modeling for the sake of having a self-contained presentation which motivates the methodological part. Section II is devoted to the description of the AfTA algorithm; we write the variational problem to be solved and we describe the associated GFB algorithm and how it can be accelerated using FISTA. Finally, we test the algorithm on a both a synthetic phantom and real data from a visual stimuli experiment.

10.2 fMRI modeling

10.2.1 BOLD Signal modeling

The purpose here is to retrieve underlying activation patterns from corrupted fMRI time courses. Such time courses are given, for each voxel $i = \{i_1, i_2, i_3\}$ at the time t, as a noisy version y of the measured blood oxygen level dependent (BOLD) response $\mathbf{x}$ (activity related signal)

$$\mathbf{y}(i,t) = \mathbf{x}(i,t) + \varepsilon(i,t)$$

where $\{\varepsilon(i,t)\}_{(i,t)}$ are random noise and nuisance components (fluctuations , signal drift, residual errors from motion correction, etc $\cdots$). On the other hand, the BOLD response describes the neural activation under the action of a causal linear (linearized) system S. If $\mathbf{s}$ denotes the activity inducing signal, we have

$$\mathbf{x}(i,t) = S\left(\mathbf{s}(i,t)\right);$$

The system S can be fully characterized by its impulse response function-here the so called hemodynamic response function (HRF)- and can be written:

$$\mathbf{x}\left(i,t\right) = \mathbf{s}\left(i,t\right) * \mathbf{h}(t)$$

In the sequel, we consider that the following assumptions are verified:

(i) The neural activation is assumed to follow a boxcar model.

(ii) The neural activation turns instantly on when stimulus is presented.

(iii) The HRF is obtained directly through biophysical equations.

The assumptions (i) and (ii) guarantees that the activity inducing signal can be modeled as spikes or blocks. The assumption (iii) insures the possibility to derive a linear differential operator which inverts the HRF.

10.2.2 HRF modeling

Many modes for the HRF have been proposed in the literature. For example, The classical *canonical* HRF [Friston et al., 1998], given in Figure 10.1, is derived mathematically using a combination of two gamma functions. The first-order Volterra series approximation of the Balloon model [Friston et al., 2000] is an alternative description which is based on biophysical equations. In this last description, the convolution with the function h can be characterized by a transfer function with a gain $G \in \mathbb{R}$, one zero $\gamma \in \mathbb{C}$ and four poles $(\alpha_1, \alpha_2, \alpha_3, \alpha_4) \in \mathbb{C}^4$ (*cf.* [Khalidov et al., 2011])

$$\mathrm{HRF}(s) = G\frac{s-\gamma}{\prod_{i=1}^{4}(s-\alpha_i)}, . \tag{10.1}$$

This also equivalent to applying a differential operator H verifying

$$\mathrm{H}\{\mathbf{s}\} = \mathbf{s} * \mathbf{h}, \tag{10.2}$$

with

$$\mathrm{H} = G\frac{D-\gamma I}{\prod_{i=1}^{4}(D-\alpha_i I)}, \tag{10.3}$$

where D and I are, respectively, the derivative and the identity operators.

10.2.3 Spatial modeling

The human brain has a very complex and impressive structure. Billions of neurons are exhibiting activations while interacting with each other. The main neuronal activity is presumed to occur in the gray matter. It has been early understood that the brain activity has a macroscale behaviour; groups of neighbour neurons are activating at the same time and are related to the same tasks. This triggered an entire branch of research aiming at mapping the human brain by providing structural or functional atlases. This atlases are in general found using task-related fMRI. As we mentioned before, we want to develop a fMRI deconvolution method which requires a minimal knowledge on the structural behaviour in activation. We will use only deterministic and probabilistic membership to gray matter as a prior.

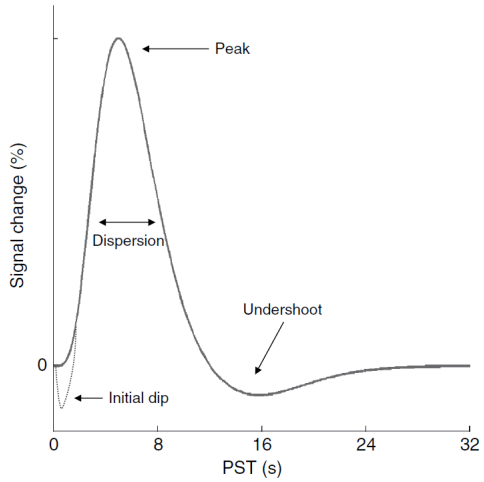


Figure 10.1: Canonical HRF.

Bibliography

R Baumgartner, L Ryner, W Richter, R Summers, M Jarmasz, and R Somorjai. Comparison of two exploratory data analysis methods for fmri: fuzzy clustering vs. principal component analysis. *Magnetic Resonance Imaging*, 18(1):89–94, 2000.

Christian F Beckmann, Marilena DeLuca, Joseph T Devlin, and Stephen M Smith. Investigations into resting-state connectivity using independent component analysis. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 360(1457):1001–1013, 2005.

Bharat Biswal, F Zerrin Yetkin, Victor M Haughton, and James S Hyde. Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magnetic resonance in medicine*, 34(4):537–541, 1995.

MJ Fadili, Su Ruan, Daniel Bloyet, and B Mazoyer. A multistep unsupervised fuzzy clustering analysis of fmri time series. *Human brain mapping*, 10(4):160–178, 2000.

Guillaume Flandin and William D Penny. Bayesian fmri data analysis with sparse spatial basis function priors. *NeuroImage*, 34(3):1108–1125, 2007.

Karl J Friston, P Fletcher, Oliver Josephs, ANDREW Holmes, MD Rugg, and Robert Turner. Event-related fmri: characterizing differential responses. *Neuroimage*, 7(1):30–40, 1998.

Karl J Friston, Andrea Mechelli, Robert Turner, and Cathy J Price. Nonlinear responses in fmri: the balloon model, volterra kernels, and other hemodynamics. *NeuroImage*, 12(4):466–477, 2000.

KJ Friston and W Penny. Posterior probability maps and spms. *Neuroimage*, 19(3):1240–1249, 2003.

C Caballero Gaudes, Natalia Petridou, Susan T Francis, Ian L Dryden, and Penny A Gowland. Paradigm free mapping with sparse regression automatically detects single-trial functional magnetic resonance imaging blood oxygenation level dependent responses. *Human brain mapping*, 34(3):501–518, 2013.

Cesar Caballero Gaudes, Natalia Petridou, Ian L Dryden, Li Bai, Susan T Francis, and Penny A Gowland. Detection and characterization of single-trial fmri bold responses: Paradigm free mapping. *Human brain mapping*, 32(9):1400–1418, 2011.

Gary H Glover. Deconvolution of impulse response in event-related bold fmri 1. *Neuroimage*, 9(4):416–429, 1999.

Fikret Işık Karahanoğlu, İlker Bayram, and Dimitri Van De Ville. A signal processing approach to generalized 1-d total variation. *IEEE Transactions on Signal Processing*, 59(11):5265–5274, 2011.

Fikret Işık Karahanoğlu, César Caballero-Gaudes, François Lazeyras, and Dimitri Van De Ville. Total activation: fmri deconvolution through spatio-temporal regularization. *Neuroimage*, 73:121–134, 2013.

Ildar Khalidov, Dimitri Van De Ville, Jalal Fadili, and Michael Unser. Activelets and sparsity: a new way to detect brain activation from fmri data. In *Optical Engineering+ Applications*, pages 67010Y–67010Y. International Society for Optics and Photonics, 2007.

Ildar Khalidov, Jalal Fadili, François Lazeyras, Dimitri Van De Ville, and Michael Unser. Activelets: Wavelets for sparse representation of hemodynamic responses. *Signal Processing*, 91(12):2810–2821, 2011.

Yijun Liu, Jia-Hong Gao, Ho-Ling Liu, and Peter T Fox. The temporal response of the brain after eating revealed by functional mri. *Nature*, 405(6790):1058–1062, 2000.

Victoria L Morgan, Yong Li, Bassel Abou-Khalil, and John C Gore. Development of 2dtca for the detection of irregular, transient bold activity. *Human brain mapping*, 29(1):57–69, 2008.

Alle Meije Wink and Jos BTM Roerdink. Denoising functional mr images: a comparison of wavelet denoising and gaussian smoothing. *IEEE transactions on medical imaging*, 23(3):374–387, 2004.

Guo-Rong Wu, Wei Liao, Sebastiano Stramaglia, Ju-Rong Ding, Huafu Chen, and Daniele Marinazzo. A blind deconvolution approach to recover effective connectivity brain networks from resting state fmri data. *Medical image analysis*, 17(3):365–374, 2013.

CHAPTER 11

fMRI deconvolution without functional priors

Abstract

The work presented in this chapter introduces new strategies for fMRI spatio-temporal deconvolution without anatomical priors. That is, only the membership of the contributing voxels to the gray matter and the global homogeneity of the activation are taken into account. The presented Atlas free Total Activation (AfTA) technique is an extension of the Total Activation (TA) framework; we formulate a variational denoising problem involving two regularization terms. These terms express sparsity along variables by taking the temporal and the spatial characteristics. The first term uses a generalized total variation which promotes a block-type structure on the the underlying activity-inducing signal by inverting the hemodynamic response function. The second term is a weighted total variation which favors globally coherent activation patterns within the gray matter while preserving strong discontinuities. Evaluation on synthetic data demonstrated the potential of AfTA at recovering brain activity-like signals. Furthermore, we applied this techniques to a real task-evoked fMRI data from an experiment with prolonged resting state periods disturbed by visual stimuli. The results show the ability of proposed technique at retrieving both spontaneous and task-related activities without prior knowledge of the timing of the experimental paradigm nor the triggered regions.

11.1 Generalized Total Variation (GTV) of Karahanoğlu et al. [2011]

We start by introducing the concept of Generalized total variation (GTV).

11.1.1 Concept

Karahanoğlu et al. [2011] gave a new version of the definition (3.10) from Chapter 3 which generalize the TV concept to a larger class of operator beyond finite differences. Consider a N-th order differential operator L which can be characterized by its poles and zeros $\gamma_i \in \mathbb{C}$, with $i = \{0, \cdots, M\}$ and $\alpha_i \in \mathbb{C}$ with $i = \{1, \cdots, N\}$

$$L = \prod_{i=1}^{N}(D - \alpha_i I)\left(\prod_{i=1}^{M}(D - \gamma_i I)\right)^{-1} \tag{11.1}$$

where I is the identity operator and D is the continuous derivation operator. Note that this definition includes, as a special case, the classical TV when $N = 1$, $\alpha_1 = 0$ and $M = 0$. Let Δ_L denotes the discrete version of L, the discrete GTV is given by

$$GTV_L(s) = \sum_{t\in\mathbb{Z}} |\Delta_L\{s\}[t]|. \tag{11.2}$$

Remarks on the discrete construction Δ_L can be found in [Karahanoğlu et al., 2011]. Denoising an observation g using the GTV consists in finding

$$\widehat{f} = \arg\min_f \left\{ ||g-f||_2^2 + \lambda\, GTV_L(f) \right\}. \tag{11.3}$$

11.1.2 Forward-backward algorithm for GTV denoising

The FISTA algorithm given in Algorithm 3 can be customized for (11.3), by choosing μ such as $\mu > ||\Delta_L^T||_2^2$ where $\Delta_L^T[t] = \Delta_L^*[-t]$ is the adjoint of Δ_L. The algorithm reads

Algorithm 5 FISTA for GTV

Input: g, λ, μ, k_{max}, $f^{(1)}$, r_1
Output: Estimate $\widehat{f}$

$\tilde{f}^{(0)} = \tilde{f}^{(1)} = f^{(1)}$; $h^{(1)} = f^{(1)}$
for $k = 1 : k_{max}$ **do**
  $\tilde{f}^{(k)} = \tilde{f}^{(k)} - \mu\Delta_L\left\{g - \Delta_L^T\left\{h^{(k)}\right\}\right\}$
  $f^{(k+1)} = \tilde{f}^{(k)} - \mathcal{T}_\lambda\left(\tilde{f}^{(k)}\right)$
  $r_{k+1} = \dfrac{1+\sqrt{1+4r_k^2}}{2}$
  $h^{(k+1)} = \tilde{f}^{(k)} + \dfrac{r_k - 1}{r_{k+1}}\left(\tilde{f}^{(k)} - \tilde{f}^{(k-1)}\right)$
end for
$\widehat{f} = g - \Delta_L^T\left\{f^{(k+1)}\right\}$

Note that by the isometry of the *Fourier* transform, an estimation of a lower bound of μ can be obtained from the z-transform of Δ_L given by $\mathcal{F}(\Delta_L)(z) := \sum_t \Delta_L[n] z^{-n}$

$$\mu > \frac{1}{\sup_w |\mathcal{F}(\Delta_L)(e^{j\omega})|^2} = \sup_w \frac{\prod_{i=1}^M |1 - e^{\gamma_i} e^{-j\omega}|^2}{\prod_{i=1}^N |1 - e^{\alpha_i} e^{-j\omega}|^2} \tag{11.4}$$

We describe, now, the AfTA method for fMRI deconvolution.

11.2 Atlas-free Total Activation

Given a fMRI time courses, AfTA aims at recovering the activity related signal $\mathbf{x}$ from the observation $\mathbf{y}$ via a spatio-temporal regularization. In the sequel $\mathcal{R}_T$ and $\mathcal{R}_S$ will refer respectively to the the temporal and spatial regularization terms. Similarly to the work of Karahanoğlu et al. [2013], the variational formulation of AfTA reads

$$\widehat{\mathbf{x}} = \arg\min_x \left\{ \frac{1}{2}||\mathbf{y}-\mathbf{x}||_2^2 + \mathcal{R}_T(\mathbf{x}) + \mathcal{R}_S(\mathbf{x}) \right\}. \tag{11.5}$$

11.2.1 Temporal regularization $\mathcal{R}_T$

As explained in the previous chapter, the unknown activity related signal $\mathbf{x}$ is under the action of the operator H given in (11.6). The temporal regularization exploits the GTV to invert H. Let H^{-1} denotes the pseudo-inverse of H. Under some assumptions, H^{-1} can have a closed form expression. In fact

$$\mathrm{H}\{\mathbf{s}\} = \mathbf{s} * \mathbf{h} = \mathbf{x}.$$

where $\mathbf{s}$ and h denote respectively the activity inducing signal and the HRF. Thus, $\mathrm{H}^{-1} = \mathbf{g} * \mathbf{x}$, where

$$\begin{cases} \widehat{\mathbf{g}}(\omega) = \dfrac{1}{\widehat{\mathbf{h}}(\omega)} & \text{if} \quad \widehat{\mathbf{h}}(\omega) \neq 0 \\ 0 & \text{if} \quad \widehat{\mathbf{h}}(\omega) = 0 \end{cases}$$

One expect that the effect of H^{-1} cancels the effect of H

$$\begin{aligned} \mathrm{H}^{-1}\{\mathbf{x}\} &= \mathrm{H}^{-1}\{\mathrm{H}\{\mathbf{s}\}\} \\ &= \mathrm{H}^{-1}\{\mathbf{h} * \mathbf{s}\} \\ &= \mathbf{g} * \mathbf{h} * \mathbf{s} \\ &= \mathbf{s} + \mathbf{s}_{null} \end{aligned}$$

with $\mathbf{s}_{null} \in \mathrm{Null(H)}$, the *null space* set of the operator H. Moreover, from the expression of the transfer function (10.1), we have the Fourier-domain counterpart of $\mathbf{h}$

$$\widehat{\mathbf{h}}(\omega) = \frac{j\omega - \gamma}{\prod_{i=1}^{4}(j\omega - \alpha_i)}, .$$

Whenever[1] $\gamma \neq j\omega$, one can guarantee a stable inversion

$$\widehat{\mathbf{g}}(\omega) = \frac{\prod_{i=1}^{4}(j\omega - \alpha_i)}{j\omega - \gamma}.$$

We conclude that

$$\mathrm{H}^{-1} = G\frac{\prod_{i=1}^{4}(D - \alpha_i I)}{D - \gamma I}. \tag{11.6}$$

We are now able to use H^{-1} within the GTV framework to invert the effect of the HRF.

Block-type priors

As mentioned in the previous chapter, we consider that the activity-inducing signal $\mathbf{s}$ follows a box-car model (block-type); its derivative is a sparse innovation signal composed of Dirac pulses

$$D\{\mathbf{s}\}[t] = \sum_{k=1}^{K} \alpha_k \delta[t - \tau_k], \tag{11.7}$$

with $K > 1$, $\alpha_k \in \mathbb{C}$ and $\tau_k > 0$. The term $\mathcal{R}_T$ uses the GTV framework to promote the sparsity prior (11.7) by controlling its ℓ_1-norm. Notice that

$$D\{\mathbf{s}\} = D\,\mathrm{H}^{-1}\{\mathbf{x}\},$$

and so, the operator L of (11.2) can be constructed here by adding a pole in the expression of H^{-1}

$$L = D\,\mathrm{H}^{-1} = G\frac{\prod_{i=1}^{4}(D - \alpha_i I)}{(D - \gamma I)(D - I)} \tag{11.8}$$

Finally, the temporal regularization is given by

[1] Whenever the zero of the system lies within the unit circle in the z (Z-transform) plane. This is the case for the HRF [Khalidov et al., 2011, Karahanoğlu et al., 2013]

$$\mathcal{R}_T(\mathbf{x}) = \sum_{v\in\mathbb{Z}^3} \lambda_T(v) \sum_{t\in\mathbb{Z}} |\Delta_L\{\mathbf{x}\}[v,t]|, \tag{11.9}$$

with L given by (11.8) and $\lambda_T(v)$ is the regularization parameter for voxel v.

11.2.2 Spatial regularization $\mathcal{R}_S$

TV can be used to tackle other tasks besides signal and image restoration. In particular, classical TV has shown to be very effective for segmentation and clustering tasks. It seems valuable to use such a regularization as a spatial prior for our problem since we expect local coherence inside evoked regions with possibly sharp variations between these regions. In fact, TV regularization might identify atlas-like clusters of activations without any functional priors. Therefore, we propose the following spatial regularization term

$$\mathcal{R}_S(\mathbf{x}) = \sum_{t\in\mathbb{Z}} \lambda_S(t) \sum_{v\in\mathbb{Z}^3} \left(\sum_{u\in N_v} w_{uv} \Big(\mathbf{x}[v,t] - \mathbf{x}[u,t]\Big)^2 \right)^{1/2}, \tag{11.10}$$

where $\lambda_S(t)$ is the spatial regularization parameter at timepoint t, N_v is the set of voxels which contributes to the finite differences around v and $W = \{w_{uv}\}$ is the weight matrix. The expression in 11.10 is simply a multivariate weighted TV. This general definition allows more flexibility on the contributing voxels. In particular, N_v will refer to voxels belonging to the gray matter that are neighbors to v within a three-dimensional Cartesian lattice. That is $\#(N_v)$ is not constant ($\min_v \#(N_v) = 0$ and $\max_v \#(N_v) = 6$). On the other hand, we propose two choices for the weight matrix W. The first one is simply an identity matrix; voxels have the same importance. The second choice tends to imposing a fading effect. In fact, it is often observed in fMRI data that evoked regions show higher activation in the center with vanishing effects towards the edge [Logan and Rowe, 2004]. We would like to take this feature into account. We use grey matter probability maps (GM-PM): A map $\mathcal{P}$ giving for each voxel v the probability $\mathcal{P}(v)$ that v belongs to the gray matter. This maps will drive the coherence of the activation. If two voxels v and u have probabilities which are close to each other the weight w_{uv} should be large and vice-versa. We construct the weight matrix via the following similarity measure

$$w_{uv} = \exp\left(- \frac{|\mathcal{P}(v) - \mathcal{P}(u)|^2}{\sigma} \right), \tag{11.11}$$

where σ controls the tolerance to the similarity.

11.2.3 Dedicated minimization algorithm

Both $\mathcal{R}_T$ and $\mathcal{R}_S$ are non-differentiable, but simple; their *proximal operators* are easy to compute. The Forward-Backward proximal splitting of Raguet et al. [2013] can be applied here to find a solution to (11.5). As explained in Chapter 3 this consists of a weighted average of the couple $(\widehat{\mathbf{x}_t}, \widehat{\mathbf{x}_s})$ solutions to the two following variational problems

$$\begin{cases} \widehat{\mathbf{x}_t} = \arg\min_{\mathbf{x}} \left\{ \frac{1}{2}||\mathbf{y} - \mathbf{x}||_2^2 + \mathcal{R}_T(\mathbf{x}) \right\}, \\ \widehat{\mathbf{x}_s} = \arg\min_{\mathbf{x}} \left\{ \frac{1}{2}||\mathbf{y} - \mathbf{x}||_2^2 + \mathcal{R}_S(\mathbf{x}) \right\}. \end{cases} \tag{11.12}$$

11.3 Experiments

This section illustrates experimentally the performances of the AfTA algorithm. First, we start with synthetic data then we show results on real fMRI series. The spatial and temporal regularization parameters should provide a compromise between the data-fidelity and the regularization costs. As in [Karahanoğlu et al., 2013], the temporal regularization parameter was estimated, for each voxel, from the median absolute deviation of

Algorithm 6 Generalized Forward-Backward algorithm for solving (11.5)

Input: Corrupted data $\mathbf{y}$, $(\omega_t, \omega_s) \in [0,1]^2$ with $\omega_1 + \omega_2 = 1$ and the operator L of (11.8)
Output: Estimate $\widetilde{\mathbf{x}}$

1: **for** $k = 1 : k_{max}$ **do**
2: $\quad \mathbf{x}_t^k = \arg\min_{\mathbf{x}} \left\{ \frac{1}{2}||\mathbf{y} - \mathbf{x}||_2^2 + \mathcal{R}_T(\mathbf{x}) \right\}$, by the FISTA for GTV (*cf.* Section 11.1.2)
3: $\quad \mathbf{x}_s^k = \arg\min_{\mathbf{x}} \left\{ \frac{1}{2}||\mathbf{y} - \mathbf{x}||_2^2 + \mathcal{R}_S(\mathbf{x}) \right\}$, by the FISTA for TV (*cf.* Section 3.4)
4: $\quad \mathbf{x}^k = \omega_t \mathbf{x}_t^k + \omega_s \mathbf{x}_s^k$
5: **end for**
6: $\widetilde{\mathbf{x}} = \mathbf{x}^{k_{max}}$

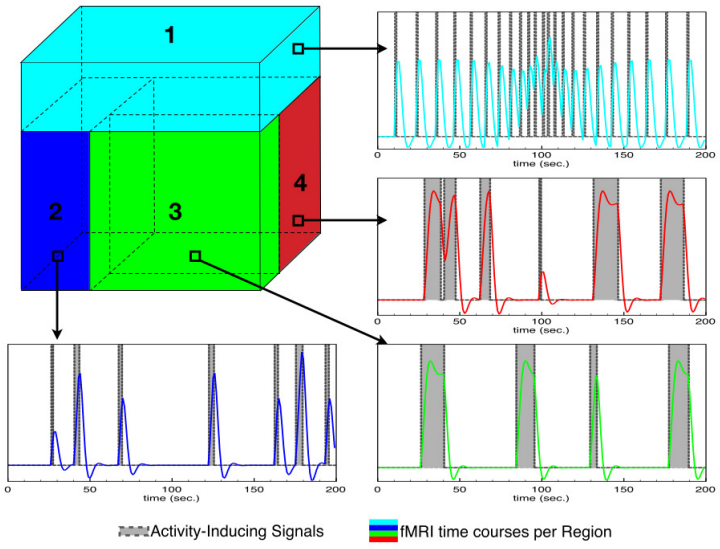


Figure 11.1: Software phantom: Courtesy of Karahanoğlu et al. [2013].

fine-scale wavelet coefficients (Daubechies with 3 vanishing moments). For each iteration n of the FISTA algorithm, the temporal regularization parameter λ_T is updated as proposed by Chambolle [2004]

$$\lambda_T^{(n+1)} = \frac{N\widehat{\varepsilon}}{||\mathbf{y} - \mathbf{x}||_2^2} \lambda_T^{(n)}$$

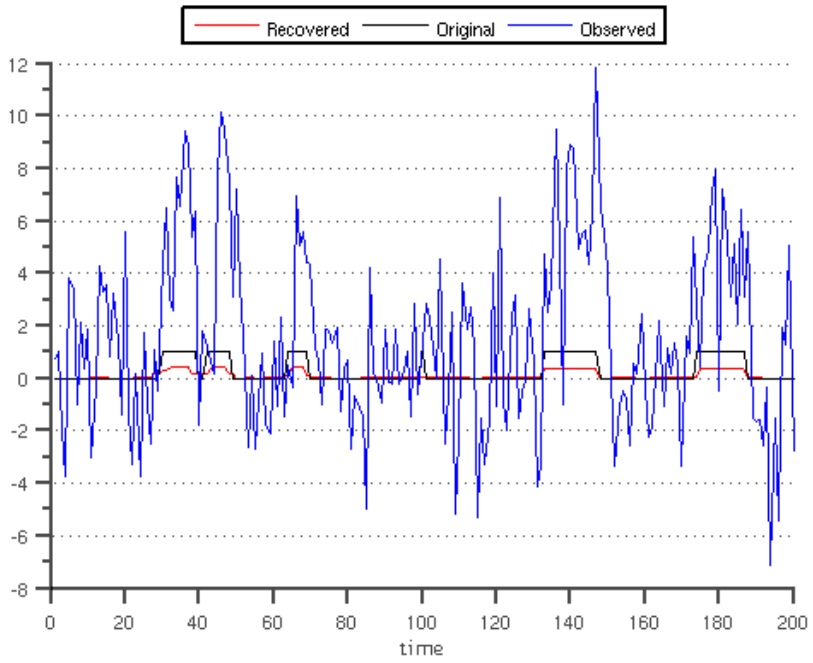
Where N is the number of time points and $\widehat{\varepsilon}$ is the estimated noise. For the spatial regularization term, the parameter was estimated empirically. We give the values we used depending on the experiment.

11.3.1 Synthetic data experiment

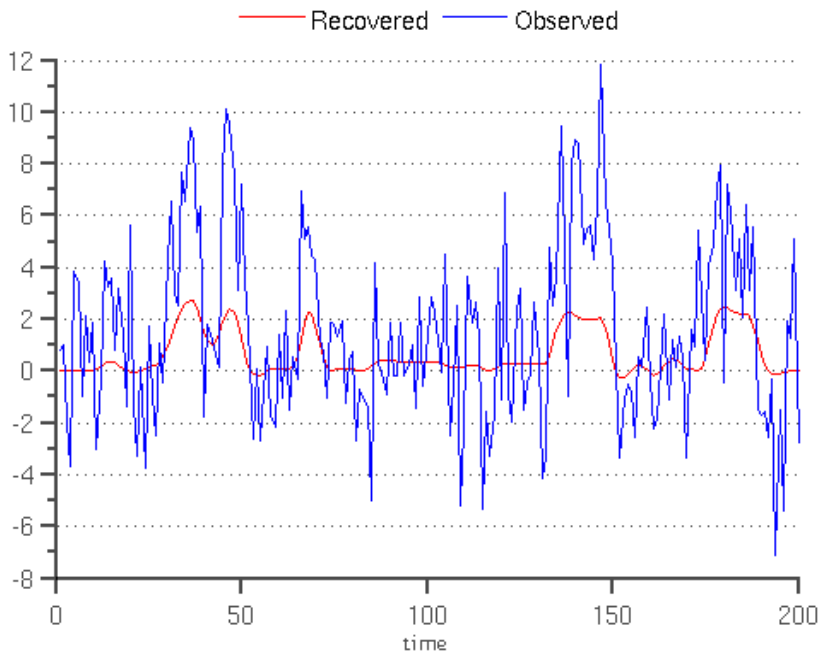
We used the software phantom from [Karahanoğlu et al., 2013] to evaluate the AfTA algorithm with no weights. The phantom, given in Figure 11.1, contains 4 regions in a $10 \times 10 \times 10$ cube. The noisy data was generated by convolving the 10^3 signals with the HRF and adding a i.i.d Gaussian noise such that the signal-to-noise (SNR) ratio is 1 dB.

We show, now, some of the results we obtained using AfTA without weights. The spatial regularization was empirically fixed at 5 and gives a good compromise between the two regularization terms. Figure 11.2 and Figure 11.3 show an example of reconstruction from a voxel in region 4 of the phantom. The activation blocks are nicely recovered. Moreover the innovation signal is very sparse as it is expected. Such innovation signals have been used, recently, for the development of innovation co-activation patterns (iCAP's) [Karahanoğlu and Van De Ville, 2015] aiming at discovering spatially and temporally overlapping networks in the brain.

Figure 11.4 highlights the spatial structure of the noisy and the recovered cube of signals at different time points. Note that the TV regularization enables to retrieve the different regions without any prior on their positioning.

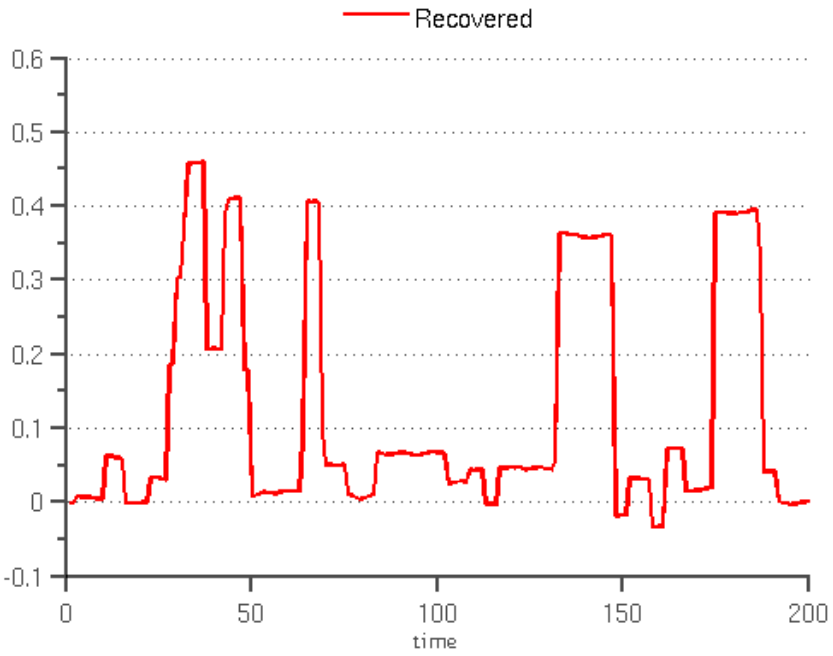


Activity inducing signal $\mathrm{H}^{-1}\{\mathbf{x}\}$

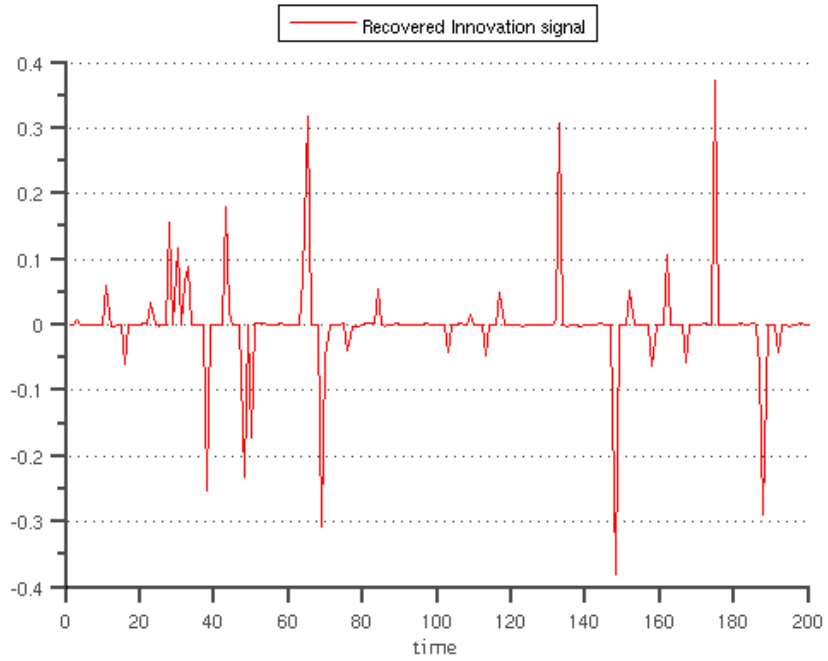


Activity related signal $\mathbf{x}$

Figure 11.2: Example of reconstruction (Region 4); the activity inducing and activity related signals.

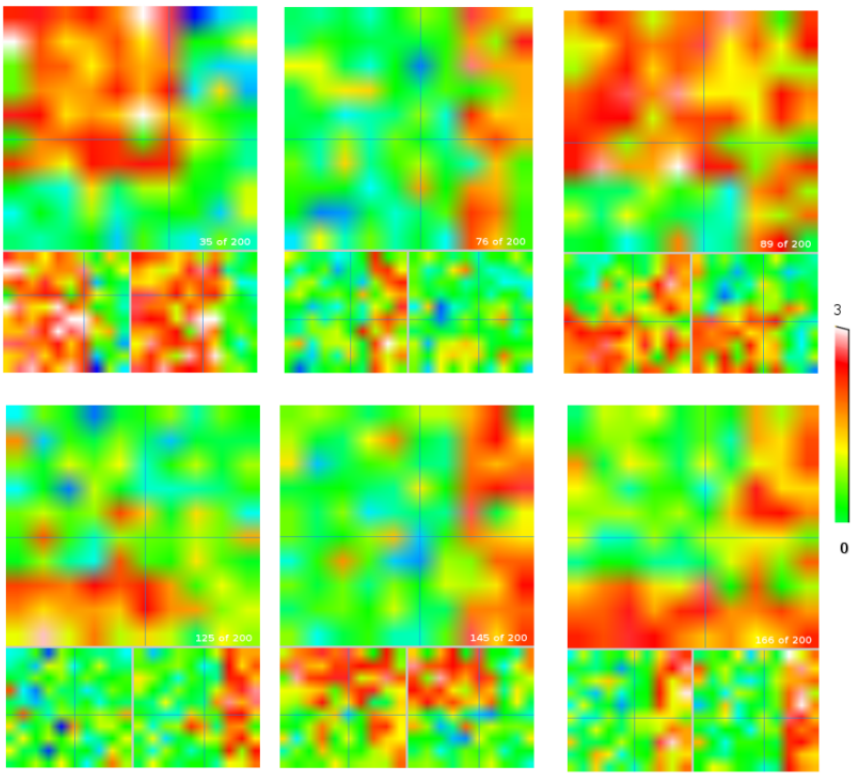


Activity inducing signal $\mathrm{H}^{-1}\{\mathbf{x}\}$

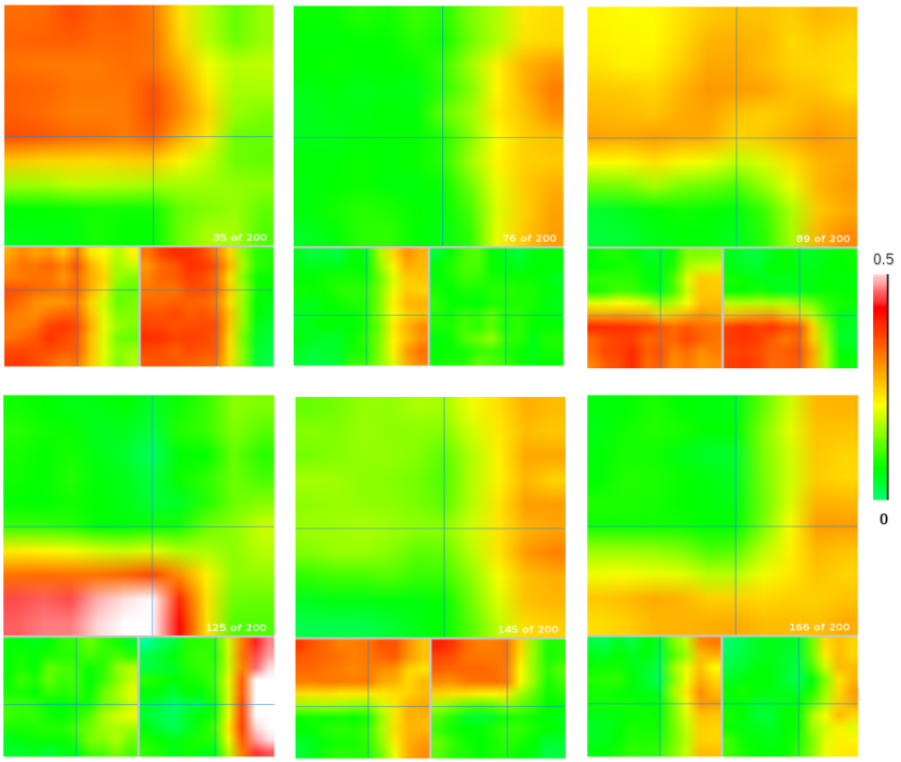


innovation signal $D\,\mathrm{H}^{-1}\{\mathbf{x}\}$

Figure 11.3: Example of reconstruction (Region 4); activity inducing signal ans its derivative.



Noisy

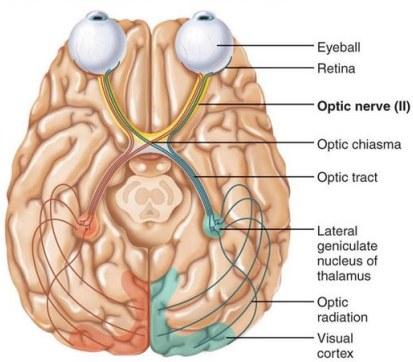


Recovered

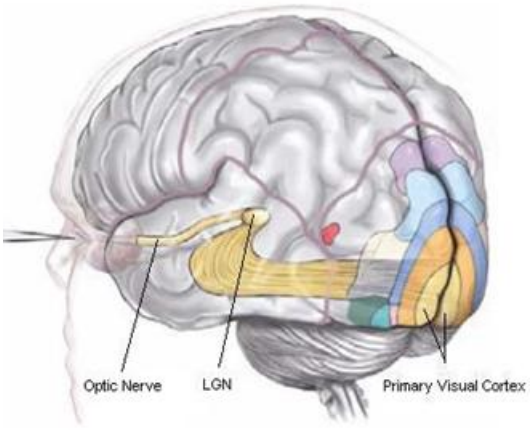
Figure 11.4: Noisy & recovered activity related signal $\mathbf{x}$ at different time points.

11.3.2 Real data

Let us describe the experimental setup used in this subsection. The experimental data was acquired on a subject during resting state periods disturbed by 10 visual stimulation of flickering checkerboard of duration $1s$. Figure 11.5 highlights the position of the primary visual cortex in the brain. We expect this area to show activation whenever the visual stimulation is on. We wish to show the difference between the regularization without weights and the one using the probability map.



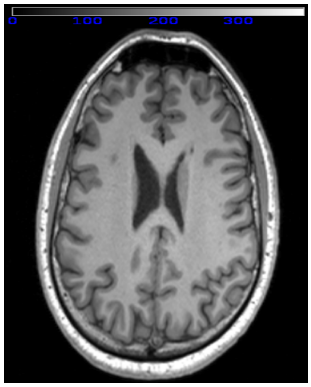
(Credits: Mike Murtaugh).



(Credits: Tom Kinniburgh / Nottingham)

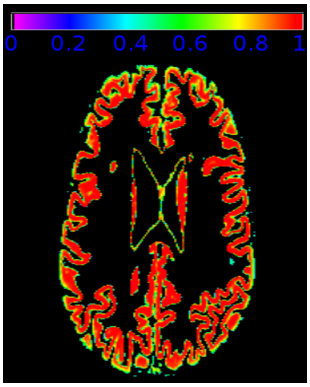
Figure 11.5: Primary visual cortex

In Figure 11.6, we can see an fMRI image from the data we used along with a gray matter mask and the corresponding probability map. When no weights are used a conventional TV is computed on the graph given by the gray matter mask. When the probability map is used, weights are computed for each finite difference component of TV as given in (11.11). In our experiments we found a value of $\sigma = 2$ gives a sufficient control of the tolerance to the similarity between voxels. The spatial regularization parameter λ_S was also empericcaly fixed at 4 for the Gray matter driven strategy and 5 for the probability map driven strategy.



T1 MRI image

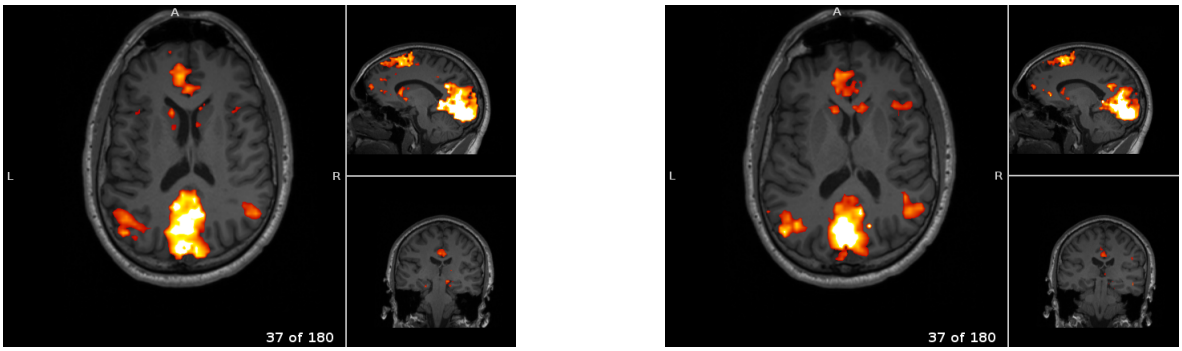
GM mask



Probability map of GM

Figure 11.6: MRI data

We start by highlighting the difference that one can expect between the two strategies. Figure 11.7 shows the spatial structure of the retrieved signal in a given time-course. When only the gray matter mask is used, the results are piece-wise constant as expected from TV. This is one of the drawbacks of TV which made it not very desirable in image processing. On the other hand, the weights provided by the probability map give more smoothness and spatial coherence to the solution. As expected in fMRI the activation is higher in the center and vanishes when getting closer to the edges.



Gray matter mask Probability map

Figure 11.7: Structure of the retrieved activity related signals with and without probability maps.

We concentrate now on the probability map driven regularization. Figure 11.8 represents the 10 activations due to the visual stimuli. We can note that activations in the primary visual cortex are found without any prior information about the experimental paradigm. Figure 11.9 shows the mean of the reconstructed temporal time courses (activity related signal) in a small window of $12mm \times 12mm \times 12mm$ which is likely to be in the primary visual cortex. We can see the 10 activations of the visual stimuli.

11.3.3 Discussion

The work of Karahanoğlu et al. [2013] offered a new technique for reliable fMRI deconvolution without any priors on the experimental paradigm. Here, we went further by proposing a method which does impose any spatial structure. A nice application would be to retrieve new functional atlases by applying the AfTA to large sets of data and learning structures and networks that appear often. In the next chapter, we describe one of our perspectives on the subject which aims at making the method more data-driven by adapting it to the studied subject.

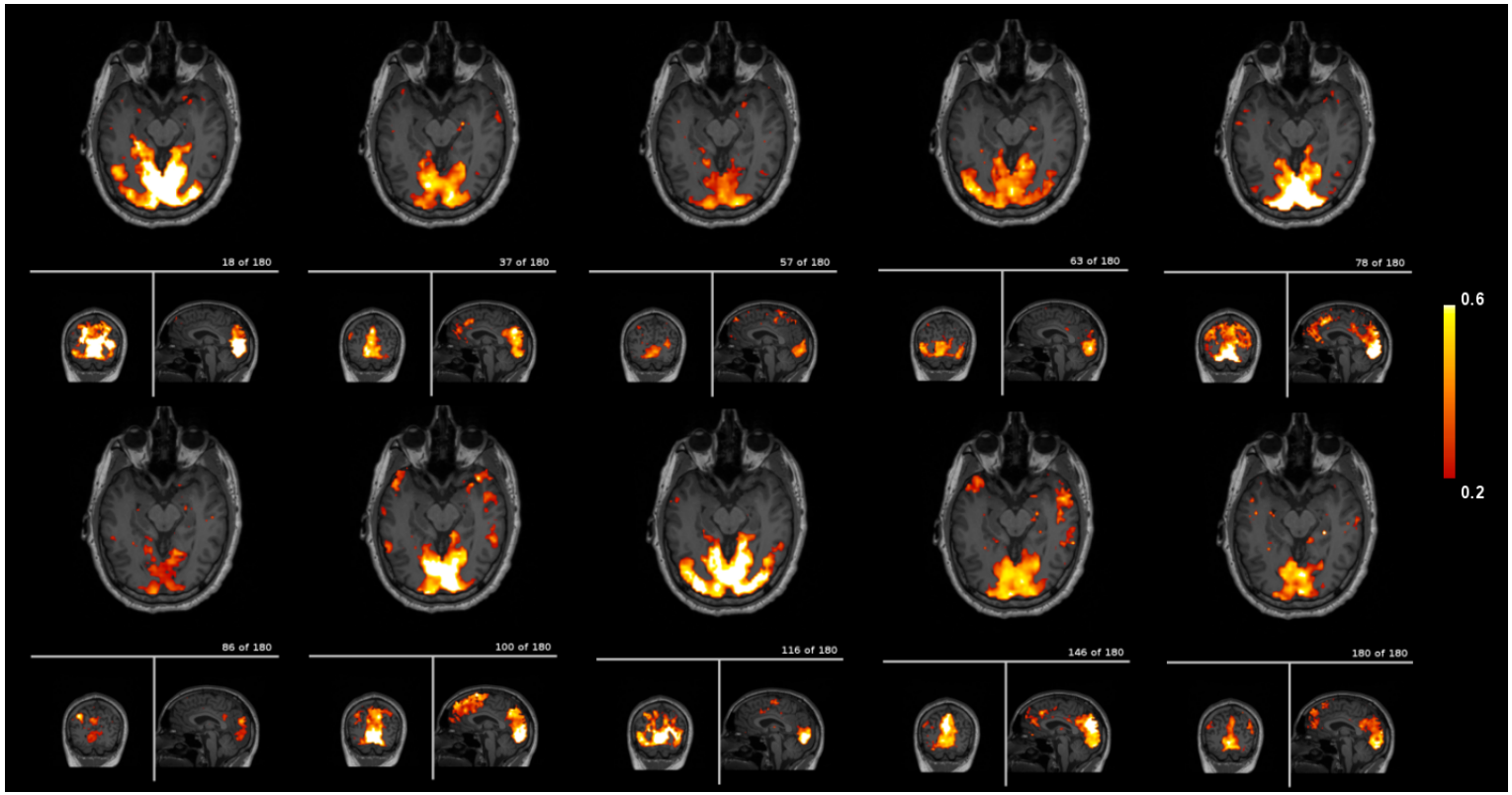
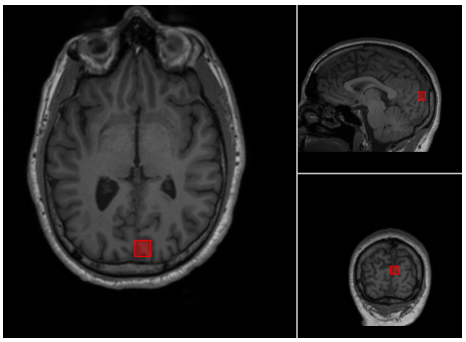
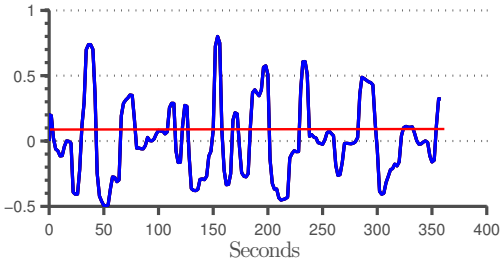


Figure 11.8: Activations referring to the visual stimuli at 10 different time points: recovered activity related signal using probability maps.



The position of the window (in red.



The mean activity related signal inside the window. The red line distinguish the upper 10 activations.

Figure 11.9: Mean of activation $\mathbf{x}$ in a small window.

Bibliography

Antonin Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical imaging and vision*, 20(1-2):89–97, 2004.

Fikret Işik Karahanoğlu and Dimitri Van De Ville. Transient brain activity disentangles fmri resting-state dynamics in terms of spatially and temporally overlapping networks. *Nature communications*, 6, 2015.

Fikret Işık Karahanoğlu, İlker Bayram, and Dimitri Van De Ville. A signal processing approach to generalized 1-d total variation. *IEEE Transactions on Signal Processing*, 59(11):5265–5274, 2011.

Fikret Işık Karahanoğlu, César Caballero-Gaudes, François Lazeyras, and Dimitri Van De Ville. Total activation: fmri deconvolution through spatio-temporal regularization. *Neuroimage*, 73:121–134, 2013.

Ildar Khalidov, Jalal Fadili, François Lazeyras, Dimitri Van De Ville, and Michael Unser. Activelets: Wavelets for sparse representation of hemodynamic responses. *Signal Processing*, 91(12):2810–2821, 2011.

Brent R Logan and Daniel B Rowe. An evaluation of thresholding techniques in fmri analysis. *NeuroImage*, 22(1):95–108, 2004.

Hugo Raguet, Jalal Fadili, and Gabriel Peyré. A generalized forward-backward splitting. *SIAM Journal on Imaging Sciences*, 6(3):1199–1226, 2013.

CHAPTER 12

Perspectives: Joint Signal-HRF estimation

Abstract

This chapter is devoted to a discussion on a possible extension of the deconvolution method presented in the previous chapter. We consider the problem of estimating both the fMRI signal and the HRF.

The AfTA technique provides a new tool for fMRI deconvolution without spatial nor temporal priors. Still, although this techniques is data-driven, it is not model-free. For instance, the HRF is predefined in advance. This is a strong assumptions which is violated in real data. The shape and magnitude of HRFs vary across subjects and brain regions [Handwerker et al., 2004]. The variation across subject is due to neural activity differences. The differences across brain regions are, presumably, due to variations in the vasculature of different regions. This affects the amplitude, the time-to-peak and the width of the HRF [Lindquist et al., 2009]. Many authors have considered the problem of estimating the HRF directly from the data [Seghouane and Shah, 2012] or joint detection estimation as in [Chaari et al., 2012, Pedregosa et al., 2015] and [Vincent et al., 2010]. We would like to use the relation between the time-to-peak/width of the HRF and its first and second derivative (*cf.* Figure 12.1). We propose to use a second-order Taylor expansion of the HRF as a model and estimate the expansion's coefficients. This can be done easily and elegantly for differential operators defined by impulse responses.

12.1 Taylor-Expansion correction

The K-th order Taylor expansions of a $R \geq K$ times differentiable function f is given by

$$f_{Taylor}(z) = \sum_{k=1}^{K} t_k f^{(k)}(z),$$

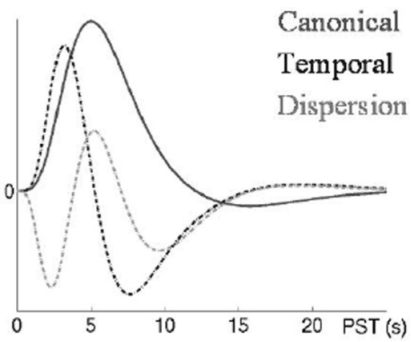


Figure 12.1: Canonical HRF and its first (temporal) and second (dispersion) derivative. Courtesy: [Tachtsidis et al., 2010].

where $\{t_k\}$ are Taylor coefficients. Thus, its Laplace transform reads

$$\begin{aligned}\mathcal{L}f_{Taylor}(z) &= \mathcal{L}f(z)\left(\sum_{\ell=1}^{L} t_\ell z^\ell\right), \\ &= \mathcal{L}f(z)\left(\prod_{i=1}^{L}(s-\beta_i)\right),\end{aligned} \tag{12.1}$$

where $\{\beta_i\}_{i=1}^{L}$ are the roots of the polynomial $\sum_{\ell=1}^{L} t_\ell z^\ell$. Now, consider f to be the rational response of a given operator

$$f = \frac{\prod_{i=1}^{N}(s-\alpha_i)}{\left(\prod_{i=1}^{M}(s-\gamma_i)\right)}$$

Then, we conclude

$$f_{Taylor} = \frac{\prod_{i=1}^{N}(s-\alpha_i)\prod_{i=1}^{N}(s-\beta_i)}{\left(\prod_{i=1}^{M}(s-\gamma_i)\right)}.$$

This means that the differential operator counterpart of the Taylor expansion is given by

$$L_{Taylor} = \frac{\prod_{i=1}^{N}(D-\alpha_i)\prod_{i=1}^{N}(D-\beta_i)}{\left(\prod_{i=1}^{M}(D-\gamma_i)\right)}. \tag{12.2}$$

12.1.1 Application to the HRF

This can be applied to the HRF operator H defined in the previous chapter

$$\mathrm{H} = G\frac{D-\gamma I}{\prod_{i=1}^{4}(D-\alpha_i I)},$$

The second-order Taylor expansion can be derived as in (12.2)

$$\mathrm{H}_{Taylor} = G\frac{(D-\gamma I)\prod_{i=1}^{2}(D-\beta_i)}{\prod_{i=1}^{4}(D-\alpha_i I)},$$

where $(\beta_1, \beta_2) \in \mathbb{C}^2$ are zeros of $\alpha_1 + \alpha_2\, s + \alpha_3\, s^2$ and $(\alpha_1, \alpha_2, \alpha_3) \in \mathbb{C}^3$ are the Taylor coefficients.

12.2 HRF Taylor coefficients estimation

Once an the activity related signal $\mathbf{x}$ is estimated by AfTA, the Taylor coefficients $(\alpha_1, \alpha_2, \alpha_3) \in \mathbb{C}^3$ can be estimated by solving a dictionary learning-like problem

$$(\widehat{\alpha_1}, \widehat{\alpha_2}, \widehat{\alpha_3}) = \arg\min_{(\alpha_1,\alpha_2,\alpha_3)} \left\{\frac{1}{2}||\,\mathrm{H}_{Taylor}\{\widehat{\mathbf{x}}_t\} - \mathbf{y}||_2^2\right\},$$

where $\widehat{\mathbf{x}}_t$ is the solution to the AfTA algorithm. One can also think of adding a regularity term to force spatial coherency.

12.3 Joint-estimation

The joint estimation perspective will consist in solving the following problem

$$(x, \alpha, \beta) = \begin{cases} \text{AfTA with } \mathrm{H}_{Taylor} \\ (\widehat{\alpha_1}, \widehat{\alpha_2}, \widehat{\alpha_3}) = \arg \min_{(\alpha_1,\alpha_2,\alpha_3)} \left\{ \frac{1}{2} || \, \mathrm{H}_{Taylor}\{\widehat{\mathbf{x}}_t\} - \mathbf{y}||_2^2 \right\}, \end{cases}$$

This problem can be solved by alternating the two problems. The convexity is clearly lost. One expect, however, that the solution will be stabilized when an accurate estimation of the HRF is obtained.

Bibliography

Lotfi Chaari, Florence Forbes, Thomas Vincent, and Philippe Ciuciu. Hemodynamic-informed parcellation of fmri data in a joint detection estimation framework. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 180–188. Springer, 2012.

Daniel A Handwerker, John M Ollinger, and Mark D'Esposito. Variation of bold hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *Neuroimage*, 21(4):1639–1651, 2004.

Martin A Lindquist, Ji Meng Loh, Lauren Y Atlas, and Tor D Wager. Modeling the hemodynamic response function in fmri: efficiency, bias and mis-modeling. *Neuroimage*, 45(1):S187–S198, 2009.

Fabian Pedregosa, Michael Eickenberg, Philippe Ciuciu, Bertrand Thirion, and Alexandre Gramfort. Data-driven hrf estimation for encoding and decoding models. *NeuroImage*, 104:209–220, 2015.

Abd-Krim Seghouane and Adnan Shah. Hrf estimation in fmri data with an unknown drift matrix by iterative minimization of the kullback–leibler divergence. *IEEE transactions on medical imaging*, 31(2):192–206, 2012.

Ilias Tachtsidis, Peck H Koh, Charlotte Stubbs, and Clare E Elwell. Functional optical topography analysis using statistical parametric mapping (spm) methodology with and without physiological confounds. In *Oxygen Transport to Tissue XXXI*, pages 237–243. Springer, 2010.

Thomas Vincent, Laurent Risser, and Philippe Ciuciu. Spatially adaptive mixture modeling for analysis of fmri time series. *IEEE Transactions on Medical Imaging*, 29(4):1059–1074, 2010.

CHAPTER 13

Concluding remarks

Consider the following statistical problem: we want to recover a multivariate function f from a noisy observation g such as

$$g = f + \varepsilon$$

where ε is the noise component. The main objective of the present PhD thesis was to construct sparsity priors that fits the best the behavior of the unknown function f on each variable or group of variables using two main ingredients which are sparsity and anisotropy.

For this concluding remarks, instead of giving a (probably) redundant summary of each of the three contributions, we present the main results in their real chronological order (part II, part I then part III) showing how the different ideas came to perspective.

The hyperbolic wavelet construction imposed itself as a potential guiding principle for the present dissertation. Recent theoretical results by J-M Freyermuth and his colleagues on the statistical performances of hyperbolic wavelets were convincing and we wanted to demonstrate their potential in (medical) image processing even if it is not always a good idea to define a tool and then try to make it useful. This theoretical results mainly concern the convergence rates of thresholding wavelet estimators. They stipulate that hyperbolic wavelet thresholding estimators achieve the best possible (minimax) convergence rates on anisotropic functional spaces. Moreover, because of the fact that isotropy is a particular case of aniotropy and thanks to the the adaptive nature of wavelet thresholding techniques, the hyperbolic construction should also give minimax convergence rates in isotropic functional spaces analogously to the isotopic wavelet construction. However, our first experiments on natural images were disheartening. In the absence of strong anisotropy and/or axis-aligned discontinuities, classical wavelets still outperform the hyperbolic wavelets in contradiction with the theoretical results. A possible explanation for this phenomenon is the biased definition of anisotropic functional spaces which consider only axis-aligned regularities, and the fact that, in general, natural images do not have strong differences in terms of smoothness along the vertical and the horizontal directions.

The work presented in part II was a first attempt to find an appropriate use of hyperbolic wavelets through a particular application which is ultrasound image denoising. Here, we were triggered by two observations

- Ultrasound images are more anisotropic than natural images because of the presence of features such as vessels and skin layers.
- The noise model in ultrasound differs from the traditional Gaussian noise model for which the classical wavelet thresholding procedures were conceived.

The second point raised also a question about the possibility of lifting the minimax results to cases where the variance of the noise is a function of the unknown image. The answer to this question is positive and was studied in chapter 7. The minimax convergence rate was obtained for hyperbolic wavelet-Fisz procedures

in two-dimensional anisotropic functional spaces. This result is a generalization of a theorem by Fryzlewicz that was obtained for the one-dimensional case. The application to ultrasound images also motivated the conception of the data-driven hyperbolic wavelet-Fisz methodology. This extension obviates the need for any prior knowledge of the noise model parameters by estimating the noise variance using an isotonic Nadaraya-Watson estimator. We also believe that a method free of tuning requirements is highly desirable, especially for physicians. Empirical results were beyond our expectations. In fact, experiments on real and synthetic images presented in chapter 8 revealed a major advantage of the hyperbolic construction; the artifacts due to the variance stabilization step were less visible resulting in better performances. This artifacts appear as the supports of the scaling functions that give the local means estimations used to perform wavelet-Fisz thresholding. When considering a hyperbolic wavelet construction, different values of the scaling parameter along the directions are allowed. As result, with the result that the supports of the scaling functions do not accumulate from scale to scale. For the particular ultrasound application, the proposed algorithm is competitive with the state-of-the-art techniques.

In many theoretical studies concerning the performances of hyperbolic wavelets, authors used images with predefined oriented patterns along one of the two directions. The same patterns can also be seen in a temporal slice of an image sequence (when one of the spatial dimension is fixed). This observation was behind the work presented in part I. A natural construction of wavelet atoms comes to mind; if the wavelet is isotropic in spatial dimensions and hyperbolic on each of the space-time planes, we should be able to benefit from the empirical performances of the two constructions. This lead to the generalized hyperbolic construction of wavelets introduced in chapter 4. The study of the theoretical performances of these wavelets motivated the introduction of some functional spaces which are isotropic not on the entire dimension but only within groups of variable. The convergence rate of hard-thresholding estimators using the new construction was obtained in chapter 4. This rate breaks "partially" the curse of dimensionality. The rate is not driven by the number of all variables but by the cardinality of the group with the largest number of variables. Many examples of application of these wavelets were presented in chapter 5. Though the two straightforward applications were image sequence and spectral/hyperspectral data denoising, we rapidly figured out that the advantages of the presented construction go beyond regularity features. One can easily construct atoms that are imposing null divergence on a group of variables. It is also possible to stabilize wavelet coefficients as in the wavelet-Fisz methodology before considering variables on which the stabilization is not needed.

In contrast to the first two parts, part III was meant to address a specific problem. We wanted to recover the brain activity, from functional magnetic resonance imaging data, without using spatial priors on the regularity of the activation. This works subscribe as an extension of the total activation (TA) methodology pioneered by I. Kaharanoglu, D. Van de Ville and their colleagues. Surprisingly[1], this problem turned out to have a common feature with the problems studied in part I. The characteristics of the data are different along the spatial and temporal dimensions. In particular, the signals are blurred in time. This, actually, brought to our minds another situation where the construction given in part I is useful, which is when we want the wavelet to act as an differential operator on some variables and not on others. Following the TA methodology, the sparsity prior, here, was on the spatial and temporal gradients through the notion of total variation (TV). In chapter 11, we introduced new strategies for fMRI spatio-temporal deconvolution using an anisotropic TV regularization which also allows to invert the effect of the temporal blur operator. The results demonstrated that this construction retrieves a spatially coherent activity and reveals networks in the brain without any prior. This is another step towards data-driven fMRI deconvolution which is promising for the retrieval of spontaneous activity and resting-state networks in the brain.

Through the different parts of the thesis, we presented many perspectives that are interesting from our point of view. We recall some of them

- The minimax results of part I were obtained assuming that the same regularity parameter along he different groups of variables. It would be also interesting to consider the different regularity parameters.

[1] Also luckily.

This is, actually, more appealing when considering problems with different regularities such as the experiments presented in chapter 5.

- As mentioned in chapter 5, many of our experiments are motivated by real applications. Trying the proposed wavelet construction on real data still the ultimate goal. Applications such as Phase-contrast MRI velocity enhancement and US Doppler imaging denoising are of great interest and enjoyed a lot of attention, recently.

- A natural extensions of the theoretical results about the hyperbolic wavelet-Fisz methodology is the derivation of rates of convergences for the data-driven extension. We expect that, as in the one dimensional case, the rate of convergence will remain optimal.

- It is also possible to use wavelets to represent faithfully some classes of spatially varying operators while solving inverse problems. This is known as Galerkin wavelet method. Hyperbolic wavelets are expected to outperform classical wavelets when the studied operator is anisotropic. A Galerkin wavelet-Fisz framework can be derived for solving problems requiring both variance stabilization and the inversion of an linear operator.

- Using the wavelet-Fisz methodology for Wigner-Ville distribution smoothing remains a challenge. well suited for the hyperbolic framework because of the different regularities along the two (space/time) dimensions.

- The main perspective of part III is to perform joint estimation of the activity and the hemodynamic response function. This is can be done through the Taylor expansion of the latter.

Personal Bibliography

Journal papers

Younes Farouj, Jean-Marc Freyermuth, Laurent Navarro, Marianne Clausel, and Philippe Delacharte. Hyperbolic wavelet-fisz denoising for a model arising in ultrasound imaging. *IEEE Trans. Computational Imag. (in minor revision)*, 2016.

Younes Farouj, Jean-Marc Freyermuth, Laurent Navarro, Marianne Clausel, and Philippe Delacharte. Generalized hyperbolic-crossing wavelets. *In preparation*, 2017.

Conference papers

Y Farouj, L Navarro, M Clausel, and P Delachartre. A variational shearlet-based model for aortic stent detection. In *2014 12th International Conference on Signal Processing (ICSP)*, pages 1052–1056. IEEE, 2014a.

Younes Farouj, Liang Wang, Patrick Clarysse, Laurent Navarro, Marianne Clausel, and Philippe Delachartre. Cardiac motion analysis using wavelet projections from tagged mr sequences. In *2014 IEEE International Conference on Image Processing (ICIP)*, pages 189–193. IEEE, 2014b.

Younes Farouj, Laurent Navarro, Marianne Clausel, and Philippe Delacharte. Débruitage de séquence d'images dynamiques par ondelettes espace-temps hyperboliques. In *In XXVème colloque GRETSI*, 2015.