Operator-Like Wavelets with Application to Functional Magnetic Resonance Imaging

Ildar Khalidov

Thèse N° 4257 (January 2009)

Thèse présentée à la faculté des sciences et techniques de l'ingénieur
pour l'obtention du grade de docteur ès sciences
et acceptée sur proposition du jury

Prof. Roland Longchamp, *président*
Prof. Michael Unser, *directeur de thèse*
Prof. Christophe Rabut, *rapporteur*
Prof. Pierre Vandergheynst, *rapporteur*
Prof. François Lazeyras, *rapporteur*

École polytechnique fédérale de Lausanne—2009

Cover design by Annette Unser
Printing and binding by Repro-EPFL
Typeset with LaTeX

Abstract

We introduce a new class of wavelets that behave like a given differential operator L. Our construction is inspired by the derivative-like behavior of classical wavelets. Within our framework, the wavelet coefficients of a signal y are the samples of a smoothed version of $\mathrm{L}\{y\}$. For a linear system characterized by an operator equation $\mathrm{L}\{y\} = x$, the operator-like wavelet transform essentially deconvolves the system output y and extracts the "innovation" signal x.

The main contributions of the thesis include:

Exponential-spline wavelets. We consider the system L described by a linear differential equation and build wavelets that mimic the behavior of L. The link between the wavelet and the operator is an exponential B-spline function; its integer shifts span the multiresolution space. The construction that we obtain is non-stationary in the sense that the wavelets and the scaling functions depend on the scale. We propose a generalized version of Mallat's fast filterbank algorithm with scale-dependent filters to efficiently perform the decomposition and reconstruction in the new wavelet basis.

Activelets in fMRI. As a practical biomedical imaging application, we study the problem of activity detection in event-related fMRI. For the differential system that links the measurements and the stimuli, we use a linear approximation of the balloon/windkessel model for the hemodynamic response. The corresponding wavelets (we call them *activelets*) are specially tuned for temporal fMRI signal analysis. We assume that the stimuli are sparse in time and extract the activity-related signal by optimizing a criterion with a sparsity regularization term. We test the method with synthetic fMRI data. We then apply it to a high-resolution fMRI retinotopy dataset to demonstrate its applicability to real data.

Operator-like wavelets. Finally, we generalize the operator-like wavelet construction for a wide class of differential operators L in multiple dimensions. We give conditions that L must satisfy to generate a valid multiresolution analysis.

We show that Matérn and polyharmonic wavelets are particular cases of our construction.

Keywords: operators, splines, wavelets, multiresolution analysis, continuous-discrete signal processing, event-related fMRI

Résumé

Nous introduisons une nouvelle classe d'ondelettes qui imitent un opérateur différentiel donné. Notre construction est inspirée par la similitude de comportement entre les ondelettes classiques et les dérivées. Selon notre concept, les coefficients d'ondelette d'un signal y sont les échantillons d'une version lissée de $\mathrm{L}\{y\}$. Pour un système linéaire caracterisé par l'équation opérationnelle $\mathrm{L}\{y\} = x$, la transformée en ondelettes correspondante déconvolue la sortie y du système et extrait le signal "d'innovation" x.

Les contributions principales de la thèse comprennent :

Ondelettes à base de splines exponentielles. Nous considerons le système L qui est décrit par une équation différentielle linéaire et nous construisons des ondelettes dont le comportement est similaire à celui de L. Le lien entre l'ondelette et l'opérateur est une fonction B-spline exponentielle ; ses décalages par pas entiers génèrent l'espace multirésolution. La construction que l'on obtient est non-stationnaire, dans le sens que les ondelettes et les fonctions d'échelle dépendent de l'échelle. Nous proposons une version généralisée de l'algorithme rapide en bancs de filtres de Mallat avec les filtres qui dépendent de l'échelle, afin de pouvoir effectuer la décomposition et la reconstruction dans la nouvelle base d'une manière efficace.

Activelets en IRM fonctionnelle. Comme application pratique en imagerie biomédicale, nous étudions le problème de détection d'activité en IRM fonctionnelle événementielle. Pour un système différentiel qui lie les mesures avec les stimuli, nous utilisons une approximation linéaire du modèle "ballon/accumulateur" pour la réponse hémodynamique. Les ondelettes correspondantes (que nous appelons "activelets") sont spécialement réglées pour l'analyse temporelle du signal en IRM fonctionnelle. Nous supposons que les stimuli sont clairsemés dans le temps et nous extrayons le signal lié à l'activité en optimisant le critère avec un terme creux de régularisation. Nous testons la méthode avec des données synthétiques d'IRM fonctionnelle. Ensuite, nous l'appliquons à un ensemble de données de rétinotopie par IRM fonctionnelle en haute résolution, afin de démontrer son applicabilité aux données réelles.

Ondelettes basées sur un opérateur. Finalement, nous généralisons la construction des ondelettes basées sur un opérateur pour une grande classe d'opérateurs différentiels L en dimensions multiples. Nous donnons les conditions que L doit satisfaire pour que l'analyse multirésolution générée soit valide. Nous montrons que les ondelettes de Matérn, comme les ondelettes polyharmoniques, sont des cas particuliers de notre construction.

Mots clés : opérateurs, splines, ondelettes, analyse multi-échelle, traitement du signal continu-discret, IRM fonctionnelle événementielle

To Davide and Sara

Contents

Chapter 1

Introduction

1.1 Wavelets In Medical Imaging

The human mind naturally considers things at different levels of detail. For example, our visual system interprets the scene from distinct pixels to meaningful objects as the neuronal impulses propagate from the primary to the higher visual areas. In signal processing, it is the concept of scale that allows control over the level of detail. The wavelet transform, acting as a mathematical eye for signals, provides the multi-scale representation of a signal at a linear computational cost. Since its introduction in the early 1990s, the wavelet transform have been successfully applied to a wide range of problems.

Medical imaging problems in signal processing are challenging, one of the important reasons being the size of the data which can be 3D and time-varying. The quest to reduce healthcare costs results in the attempt to shift the effort from the doctor and/or the imaging device to the signal processing software. Typical characteristics of medical images are low SNR, imaging artifacts, etc. Unlike many other applications, medical data processing is targeted at the enhancement/preservation of relevant (e.g., anatomical) structures in the image, rather than the gain in traditional criteria such as SNR. The need for fast processing algorithms naturally places the spotlight on wavelets.

In imaging tasks, wavelets owe much of their success to the way in which they differentially treat smooth and edge-like components of the signal. The vanishing moment property ensures that, for wavelets of order n, the wavelet coefficients of a polynomial of order up to $n-1$ will be zero. Such a polynomial will be reproduced by the scaling functions at the coarsest scale, resulting in few large coefficients. Because wavelets behave like multiscale derivatives, an edge-like structure would be represented by a sparse set of wavelet coefficients across scale. These two properties explain why the energy of piecewise smooth signals tends to be compacted in a few coefficients in the wavelet domain. Closely related are the improved non-linear approximation properties of the wavelet transform on certain classes of discontinuous signals. Last but not least, wavelet decomposition and reconstruction are implemented with fast filterbank algorithms at a computational cost that is proportional to the data size.

In the sequel, we briefly illustrate the relevance of the properties of wavelets by considering some specific imaging problems.

Vanishing moments and image reconstruction. The way to use wavelets efficiently can be inspired by the image formation model. In computed tomography (CT) image reconstruction, the density function must be recovered from the samples of its Radon transform. The direct inversion formula is unstable. Instead, the filtered backprojection algorithm is used, where a ramp filter is followed by the adjoint of the Radon operator. The ramp filter's impulse response is not differentiable at the origin, which explains the non-local character of the inverse operator. Typically, during a CT exam, we are only interested in a particular region of the body. When the signal is decomposed in the wavelet domain, the vanishing moment property improves the smoothness of the spectrum, and local reconstruction becomes possible from fewer measurements [1]. This reduces both the exposure time for the patient and the cost of the exam itself.

Sparsity and regularization. Image reconstruction from fewer measurements in many modalities such as positron emission tomography (PET) is an ill-posed problem that requires prior knowledge on the imaged object. One way to take into account prior information is to include a regularizing penalty term into the Lagrangian-like cost functional of the problem, as it is done in the variational approach to reconstruction. Due to the energy compaction property of the wavelets, we can assume that the signal of interest has a small support in the wavelet domain. Therefore, a norm that enforces compact energy distribution among the wavelet coefficients is a good penalty function candidate [2].

Wavelet approximation and compression. With the growing amounts of data in medical imaging, compression is an important issue. Good non-linear approximation properties of wavelets, together with their apparent ability to preserve medically-relevant structures, was used for better compression of MRI and CT images [3]. In combined imaging modalities, such as PET-MRI, images

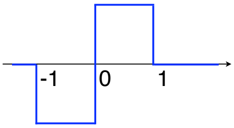


Figure 1.1: Haar wavelet ψ.

from two devices must be co-registered. The multi-scale wavelet representations form a base for robust registration algorithms [4].

Wavelet-domain denoising and detection. In dynamic imaging modalities, such as fMRI, a sequence of images is acquired to study the temporal evolution of the signal. Similar to static imaging, wavelets could be applied on a frame-by-frame basis, e.g., for denoising purposes [5]. In WSPM, the decorrelation property of wavelets in space is used for to improve statistical detection [6].

By reviewing the imaging literature [6,7], it is quite apparent that the use of wavelets has been exploited with good success in the spatial domain. There is much less work with dynamic imaging that applies wavelets along the temporal dimension. This may be explained by the mismatch between the models of the temporal signal generation with the pure derivative nature of traditional wavelets. In this work, we would like to extend the wavelet construction and tune it to the characteristics of temporal data. To motivate our approach, we start by investigating the reasons underlying the favorable behavior of the wavelets for certain classes of signals (e.g., piecewise smooth functions).

1.2 Wavelets And Derivatives

To gain a better understanding of the wavelet transform properties, let us have a closer view at the first-order case. The Haar wavelet function ψ that corresponds to the first-order derivative operator $\mathrm{D} = \frac{\mathrm{d}}{\mathrm{d}t}$ is shown in Fig. 1.1. It satisfies

$$\psi = \mathrm{D}^* \phi,$$

where ϕ is the triangle function (see Fig. 1.2) and $\mathrm{D}^* = -\mathrm{D}$.

Consider the piecewise-constant signal $x(t)$ represented in Fig. 1.3. Let us denote by a_l and t_l the heights and the locations of the jumps in $x(t)$, respectively. If we apply the derivative operator to $x(t)$, we get a stream of

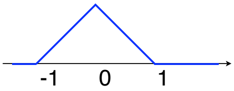


Figure 1.2: Haar multiscale smoothing kernel ϕ.

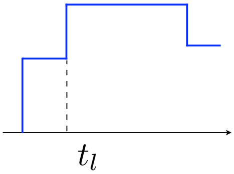


Figure 1.3: Piecewise constant signal $x(t)$.

Dirac impulses $\mathrm{D}\{x\}(t) = \sum_l a_l \delta(t - t_l)$ as shown in Fig. 1.4. At the same time, the wavelet at the scale i acts upon $x(t)$ leading to

$$x(t) * \psi^*(\frac{t}{2^i}) = \mathrm{D}\{x\}(t) * \phi(\frac{t}{2^i}) = \sum_l a_l \phi(\frac{t - t_l}{2^i}). \tag{1.1}$$

The resulting function, which is shown in Fig. 1.5, can be seen as a multiscale approximation of the Dirac train in Fig. 1.4. Its samples yield the wavelet representation, which will be sparse if the distance between two consecutive events $t_{l+1} - t_l$ is larger than the length of the support of $\phi(t/2^i)$.

The vanishing moment property goes hand in hand with this operator-like behavior. Indeed, if we apply the Haar wavelet transform to the signal $x_c(t) =$

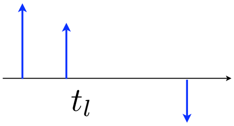


Figure 1.4: The derivative turns $x(t)$ into the Dirac train.

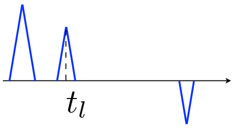


Figure 1.5: The function $x(t) * \psi^*(\frac{t}{2^i})$.

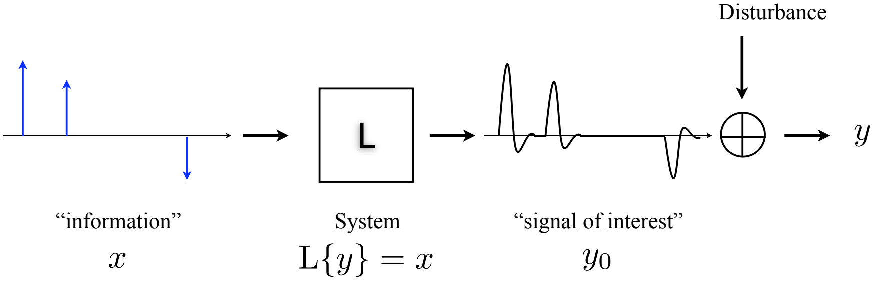


Figure 1.6: Signal formation model.

$const$, we get

$$x_c(t) * \psi^*(\frac{t}{2^i}) = \mathrm{D}\{x_c\}(t) * \phi(\frac{t}{2^i}) = 0.$$

In other words, the Haar wavelet annihilates constant signals; i.e., it has one vanishing moment.

Likewise, N-th order wavelets act as N-iterated derivatives. They annihilate polynomial signals up to order $N - 1$ and provide sparse representations of piecewise polynomials.

1.3 Extension of the piecewise-constant model

Typical biomedical time signals are not piecewise-constant. They contain oscillations and are often ruled by some differential equation. Consider a general signal formation model in Figure 1.6. The signal of interest y_0 is related to an "information" (or "innovation") process x through a linear system

$$\mathrm{L}y_0 = x, \tag{1.2}$$

where L is a differential operator. Additional components that are not described by the model (e.g., noise) are typically superposed onto y_0 and the composite signal y is measured. Suppose that we are interested in the information x. For a well-chosen model, applying the operator L to the measurement y would separate the "disturbance" term $\mathrm{L}(y - y_0)$ from x; e.g., when the support of x is small and the energy of $\mathrm{L}(y - y_0)$ is distributed among many coefficients, the two components can be separated by thresholding. In the classical stochastic framework, x is assumed to be a gaussian white process, while L corresponds the prior information available on the signal power spectrum and essentially is the decorrelating, or whitening, operator for y_0.

We see that the continuous operator L for (1.2) possesses the same properties as the wavelets in discrete image processing. This is not a random observation; in fact, wavelets act as a multiscale version of the edge-detecting Laplacian operator. In the one-dimensional case, the derivative-like behavior of wavelets ensures energy compaction around signal discontinuities, while the orthogonality property means that the noise is distributed uniformly around the coefficients.

1.4 Contributions

In this thesis, we consider the use of wavelets that are matched to the general model (1.2). Here is the list of the main contributions of the present work:

- In the one-dimensional case, we consider a system L characterized by some differential equation with constant coefficients. We derive a new class of exponential-spline wavelets that act as a multiscale version of L. In particular, the exponential polynomials that belong to the kernel of L are annihilated by the new wavelet transform. Furthermore, the wavelet coefficients at each scale correspond to the samples of a smoothed version of the system input. The wavelet representation is non-stationary in the sense that the scaling and wavelet filters are scale-dependent; however, efficient decomposition and reconstruction is still possible. We give explicit expressions of the filters that are used for the fast decomposition and reconstruction routines.

- The operator-like behavior of exponential-spline wavelets suggests their use in biomedical imaging problems that admit differential-equation-based models. In this work, we are interested in the brain activity detection problem in fMRI. FMRI is a non-invasive brain mapping technique that relies on the blood-oxygenation level-dependent (BOLD) signal change in MRI. We rewrite the model for BOLD under the form (1.2) and introduce wavelets (to which we give the name "activelets") that model the underlying operator. The extraction of the activity-related component is performed by means of a sparse-solution search algorithm in the activelet representation.

Unlike traditional data analysis techniques, our method does not require any prior knowledge on experimental timing. The activelet approach lays the foundation to new experimental paradigms, where timing is unknown or imprecise. In the present study, we limit our interest to testing the approach with synthetic and real fMRI datasets where timing is known. As an evaluation scenario for the spatial resolution of the method, we apply it to high-field retinotopy data.

- After having demonstrated the feasibility of our approach in brain localization mapping, we get back to the general model (1.2). We introduce the construction of operator-like wavelets that mimic the behavior of a given differential operator L in multiple dimensions. Starting from the Fourier-domain description of the operator, we define a scale-dependent wavelet function whose shifts span the wavelet space. The apparent simplicity of the concept — we have only one wavelet function per scale — makes it particularly attractive from the mathematical point of view. The fast decomposition and reconstruction algorithms in traditional wavelet processing are easily adapted to the new framework.

1.5 Organization

The cornerstone construction of the thesis — the exponential-spline wavelets — is presented in Chapter 2. In Chapter 3, we give a brief overview of the human brain anatomy and obtain the operator that links the fMRI measurements with the stimulus sequence in neuroscience studies. The activity-related signal extraction problem is formulated in mathematical terms in Chapter 4. In Chapter 5, we consider classical and modern techniques to solve the sparsity problem efficiently. Experimental results are presented and analyzed in Chapter 6; a more subtle retinotopy experiment is described in Chapter 7. Finally, in Chapter 8, we present a general theory of operator-like wavelets that makes our construction applicable to an even wider range of biomedical imaging problems.

Chapter 2

From Differential Equations To The New Wavelet-Like Bases

Abstract — We introduce an approach based on differential operators to construct wavelet-like basis functions. Given a differential operator L with rational transfer function, we obtain elementary building blocks that are shifted replicates of the Green's function of L. We show that these can be used to specify a sequence of embedded spline spaces that admit a hierarchical exponential B-spline representation. The corresponding B-splines are entirely specified by their poles and zeros; they are compactly supported, have an explicit analytical form, and generate multiresolution Riesz bases. Moreover, they satisfy generalized refinement equations with a scale-dependent filter and lead to a representation that is dense in L_2. This allows us to specify a corresponding family of semi-orthogonal exponential spline wavelets, which provides a major extension of earlier polynomial spline constructions. We completely characterize these wavelets and prove that they satisfy the following remarkable properties:

- they are orthogonal across scales and generate Riesz bases at each resolution level;

- they yield unconditional bases of L_2 — either compactly supported (B-spline-type) or with exponential decay (orthogonal or dual-type);

- they have N vanishing exponential moments, where N is the order of the differential operator;
- they behave like multiresolution versions of the operator L from which they are derived;
- their order of approximation is $(N-M)$, where N and M give the number of poles and zeros, respectively.

Last but not least, the new wavelet-like decompositions are as computationally efficient as the classical ones. They are computed using an adapted version of Mallat's filterbank algorithm, where the filters depend on the decomposition level.

This chapter is based on our paper [8].

2.1 Introduction

In recent years, the wavelet transform has emerged as a powerful tool for performing multiresolution signal analysis and processing [9–11]. Wavelets have led to a multitude of applications with a significant impact on image compression, communications, and on other areas of applied mathematics. There has also been an intense activity in wavelet design leading to the construction of a large variety of wavelet bases, the most prominent ones being tailored to special requirements such as orthogonality and short support [12], high number of vanishing moments [13], symmetry and regularity [14], explicit analytical form [15–17], near-optimal time-frequency localization [18], to cite but a few.

One of the key mathematical properties of wavelets is that they behave like multiscale differentiators [9]; i.e., the wavelet coefficients of a signal are the samples of the N-th order derivative of smoothed versions of it. Thus, there is a correspondence between a wavelet with N vanishing moments and the differentiation operator D^N. One of the questions that motivated this work is: What happens if we consider an arbitrary linear differential operator L instead of D^N? Will we be able to construct wavelets that qualitatively behave like L? In this chapter, we show that this is indeed possible for differential operators with arbitrary rational transfer functions.

The derivative-like behavior of wavelets was investigated in some depth in [19]. Specifically, it was shown that it is the regular component of the scaling function (i.e., the polynomial B-spline that lies hidden within) that induces this property. Mathematically, this behavior turns out to be intimately linked to the property that the polynomial B-spline of order N is a localized version (i.e., a linear combination of shifted replicates) of the Green's function of the operator D^N. Thus, a possible way to induce a modified wavelet behavior is to consider generators that are localized versions of the Green's functions of more general differential operators. In the case of ordinary differential operators, the exponential B-splines are natural candidates [20]. In the cardinal setting, it has been shown recently that these B-splines share all important properties of their polynomial counterparts: they are compactly supported, they generate Riesz bases, and they satisfy some general multiresolution-like embedding properties [21, 22]. In this chapter, we go one step further and use these functions to specify an extended family of exponential spline wavelets that satisfy some interesting mathematical properties. While we believe that our proposition to construct wavelets based on operators with rational transfer function is novel, there is a theoretical connection with earlier work on non-stationary wavelets [23–28]. For instance, the present construction falls within the generalized multiresolution framework of de Boor, DeVore and Ron [24]. These authors even briefly considered exponential splines as an illustrative example of their general scheme [24, Section 6]; they established the existence of such multiresolution bases of L_2, but they did not go into much practical details beyond this. Also relevant is the work of Lyche and Schumaker [27], who construct general L-spline wavelets with non-uniform knots on the interval, starting with a time-varying differential operator. There is a link between exponential spline wavelets and some non-stationary subdivision schemes that preserve exponential polynomials [29]. We note, however, that these only correspond to the lowpass synthesis part of the wavelet algorithm and that the filters are interpolating, which is typically not the case here.

This chapter is organized as follows: In Section II, we illustrate our wavelet-construction method by producing an extension of the Haar transform that is matched to the first-order differential operator $(\mathrm{D} - \alpha \mathrm{I})$ with parameter $\alpha \in \mathbb{C}$. In Section III, we consider the general case of a differential operator L and introduce the corresponding exponential B-splines as space-generating functions. We specify the embedding properties of the exponential B-spline spaces and give a new result on their approximation properties that guarantees that the representation is dense in L_2. In Section IV, we use these functions to construct orthonormal and semi-orthogonal wavelets and characterize the corresponding filters. This provides the generalization of the polynomial spline wavelet family [17] for a much larger class of splines. In Section V, we identify the key mathematical properties of the exponential B-spline wavelets. Finally, we discuss the computational aspects of the introduced wavelet transform and consider specific examples.

2.2 An Exponential Extension of the Haar System

To introduce our new wavelet concept, we start with a simple illustrative example and show how piecewise exponentials can be used to construct an extended version of the Haar transform.

2.2.1 E-Spline Multiresolution: First-Order Case

Consider the first-order linear differential operator $\mathrm{L} = \mathrm{D} - \alpha \mathrm{I} : \mathrm{L}f = f' - \alpha f, \alpha \in \mathbb{C}$. As is well known from linear differential equation theory, the Green's function for this operator is $\rho_\alpha(t) = e^{\alpha t} \cdot u(t)$, where $u(t)$ is a step function. We recall that the Green's function ρ is causal and satisfies $\mathrm{L}\rho = \delta$, where δ denotes the Dirac impulse; it is simply the impulse response of the causal inverse operator $\mathrm{H} = \mathrm{L}^{-1}$.

By definition, the exponential spline $s(t)$ associated with the operator L is a function such that

$$\mathrm{L}\{s\}(t) = \sum_{k \in \mathbb{Z}} a[k] \delta(t - t_k),$$

where the $a[k]$'s are arbitrary coefficients, and where the t_k's are called the knots of the spline. In our case where L is a first-order differential operator, the spline $s(t)$, as defined above, is discontinuous at these points. We can integrate this equation by applying the inverse operator $\mathrm{H} = (\mathrm{D} - \alpha \mathrm{I})^{-1}$ to this relation, which yields

$$s(t) = \sum_{k \in \mathbb{Z}} a[k] \rho_\alpha(t - t_k) + p_\alpha(t), \tag{2.1}$$

where the additional term $p_\alpha(t) = p_0 e^{\alpha t}$, with $p_0 \in \mathbb{C}$, is a solution of the homogeneous equation $\mathrm{L}p_\alpha = 0$ to be chosen so that $s(t)$ satisfies specific boundary conditions.

To be able to apply fast filtering algorithms, we restrict ourselves to the case where the knots are equally spaced; i.e., $t_k = Tk, k \in \mathbb{Z}$, where T is the interval between two knots. The corresponding spline space $V_T = \mathrm{span}\{\rho_\alpha(\cdot - Tk)\}$ is in this case T-shift invariant. One can also omit p_α, the element of the null space of L in (2.1), because it can be expressed through the shifts of the Green's function [21].

While the representation of a spline in terms of shifted Green's functions $\rho_\alpha(t - Tk)$ is attractive conceptually, it has the disadvantage of involving basis functions that are not compactly supported. The key idea, which is in the foundation of the present work, is that one can construct an equivalent set of compactly supported basis functions by taking a suitable linear combination of basis functions. It is not difficult to see that the shortest-possible functions in V_T take the form

$$\beta_{\alpha,T}(t) = \rho_\alpha(t) - e^{\alpha T}\rho_\alpha(t - T). \tag{2.2}$$

This function, which is compactly supported in $[0, T]$, is the exponential B-spline of order one. In effect, the Green's function is truncated by substracting its weighted and shifted version (see Figure 2.1).

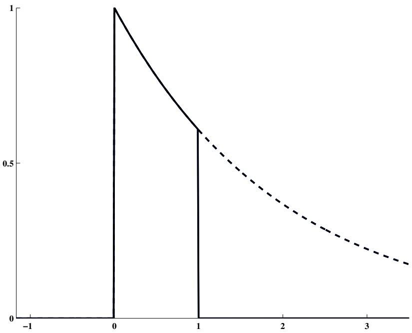


Figure 2.1: Localization of the Green's function, $T = 1$.

The T-integer-shifted B-splines $\{\beta_{\alpha,T}(t - Tk)\}_{k\in\mathbb{Z}}$ are obviously orthogonal to each other, and they form a Riesz basis. Conversely, it is also possible to invert (2.2) and to express the Green's function as the linear combination of B-splines

$$\rho_\alpha(t) = \sum_{k=0}^{\infty} e^{\alpha kT}\beta_{\alpha,T}(t - Tk).$$

The above Green-function reproduction formula, which is easily visualized graphically, implies that all T-shifts of the Green's function ρ_α belong to $\text{span}\{\beta_{\alpha,T}(\cdot - Tk)\}$, and thus

$$V_T = \text{span}\{\beta_{\alpha,T}(\cdot - Tk)\}.$$

The Green's function itself does not depend on the step size T. Thus, given a fixed parameter α, we can construct the Green's function ρ_α and build V_T as a span of T-shifts of ρ_α. On the contrary, the B-splines $\beta_{\alpha,T}(t)$ depend on the scale imposed by T, and this is the price paid for their compact support. Moreover, if we compare the B-spline to a standard scaling function, we can clearly see that the relation between the scales T and $2T$ is no longer a dilation (see Figure 2.2).

So far, we have constructed the spaces V_T. The Green-function representation makes the inclusion $V_{2T} \subset V_T$ obvious, as illustrated in Figure 2.3. It is then a natural step to attempt the construction of a multiresolution-like structure using B-splines at a dyadic scale $T = 2^i$. For this purpose, we localize the Green's functions and consider B-splines as basis functions (see Figure 2.2).

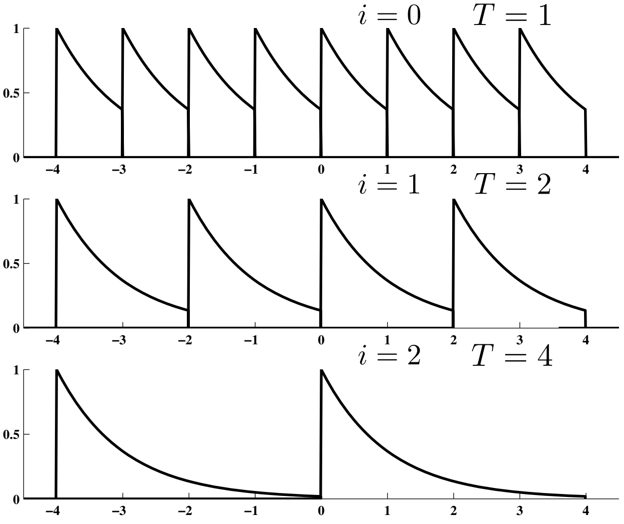


Figure 2.2: Multiresolution analysis: E-spline representation.

What we get is not a multiresolution analysis in the classical sense because it is not dilation that links the B-splines at different scales. Also, as soon as $\rho_\alpha \in L_2$ belongs to each space V_{2^i} (i.e., when $\mathrm{Re}\{\alpha\} < 0$), the intersection $\cap_{i=-\infty}^{\infty} V_{(i)}$ is not empty. In this case, this happens because the Green's function is square-integrable. This means that the sum in the wavelet-space decomposition of L_2

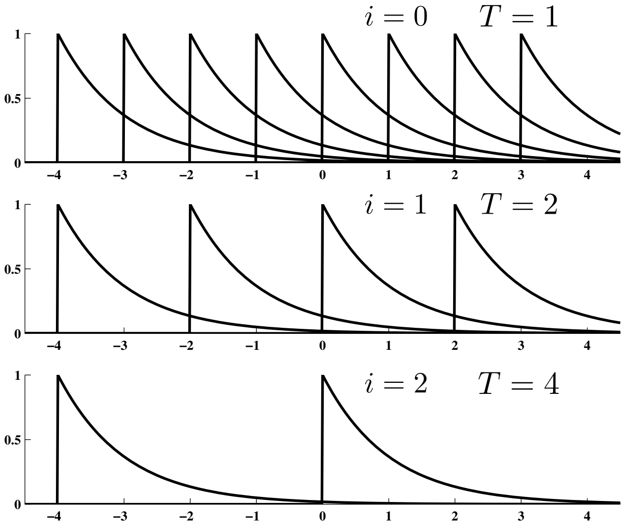


Figure 2.3: Towards a multiresolution analysis: Green-function representation of spline spaces. The basis of V_{2T} contains every other element of the basis of V_T.

must start from a finite scale and include the corresponding lowpass space. One should mention, however, that in practical applications one always keeps lowpass data at the coarsest scale, which makes the empty-intersection property irrelevant.

From now on, we denote $V_{(i)}$ the space V_T with $T = 2^i$, $c_i = ||\beta_{\alpha,2^i}||_{L_2}$, and $\varphi_i(t) = \frac{1}{c_i}\beta_{\alpha,2^i}(t)$, the corresponding normalized generating function. As we see from the Green-function representation of these spaces, $V_{(i+1)}$ is a subset of $V_{(i)}$. This means that we can expand $\varphi_{i+1} \in V_{(i+1)}$ in $\{\varphi_i(\cdot - 2^i k)\}_{k\in\mathbb{Z}}$. Specifically, we have that

$$\varphi_{i+1}(t) = \frac{c_i}{c_{i+1}}(\varphi_i(t) + e^{2^i\alpha}\varphi_i(t - 2^i)),$$

which points out a fundamental difference with conventional wavelet theory: the present filter coefficients exhibit a scale dependence. The Fourier expression for the refinement filter is

$$H_{\alpha,i}(e^{j2^i\omega}) = \frac{\hat{\varphi}_{i+1}(\omega)}{\hat{\varphi}_i(\omega)} = \frac{c_i}{c_{i+1}} \cdot (1 + e^{2^i(\alpha - j\omega)}),$$

which is clearly $(2^{-i} \cdot 2\pi)$-periodic in ω.

The next step is to examine the orthogonal complement $W_{(i+1)}$ such that $V_{(i+1)} \oplus W_{(i+1)} = V_{(i)}$, as it is done in classical wavelet construction. It is not difficult to see that the function $\psi_{i+1}(t) = \frac{c_i}{c_{i+1}}(e^{2^i\alpha^*}\varphi_i(t) - \varphi_i(t - 2^i))$ is a generator of $W_{(i+1)}$. It is clearly included in $V_{(i)}$ and is orthogonal to φ_{i+1}, as justified by

$$\langle\varphi_{i+1}, \psi_{i+1}\rangle = \frac{c_i^2}{c_{i+1}^2} e^{2^i\alpha} \int_0^{2^i} |e^{2\alpha t}| \mathrm{d}t$$

$$-\frac{c_i^2}{c_{i+1}^2} \int_{2^i}^{2^{i+1}} e^{2^i\alpha} \cdot |e^{2\alpha(t-2^i)}| \mathrm{d}t = 0$$

We show this wavelet in Figure 2.4. We observe that the Haar wavelet corresponds to $\alpha = 0$.

There are several important properties that can be seen with this first E-spline wavelet. First, its integral is not necessarily zero, which means that the corresponding filter is not necessarily highpass. Second, the wavelets (and, consequently, their spectra) at different scales are not dilated replica of each other. For instance, when $\alpha = j\omega_0$, the wavelet is complex of constant amplitude and its spectrum is shifted by ω_0 as compared to the Haar case. Finally, we can prove that the extended Haar multiresolution is dense in L_2 (cf. Section III.B). This is the final ingredient that is required to have a wavelet-like basis of L_2.

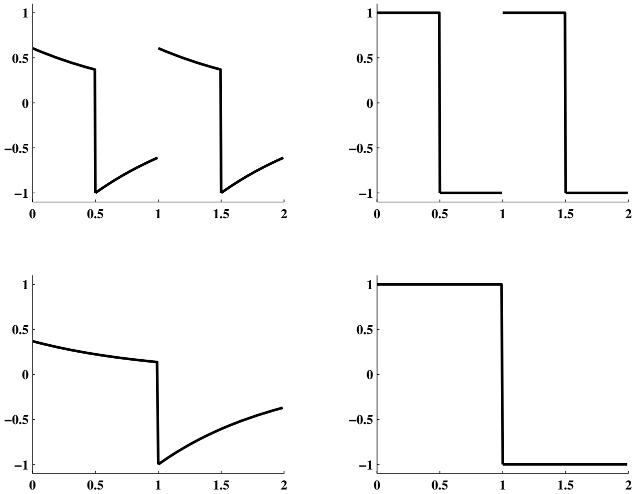


Figure 2.4: Left: First-order E-spline wavelet ψ_i at different scales. Right: Haar wavelet at different scales.

2.3 E-Splines

In this section, we generalize the previous construction by introducing E-splines associated with higher-order operators and by defining the corresponding multiresolution analysis of L_2. We then consider the approximation properties of the constructed multiresolution representation and prove that it is dense in L_2.

2.3.1 E-Spline Multiresolution

We start with the specification of a linear differential system. This system takes an input function $x(t)$ and produces an output $y(t)$; its behavior is generally described by

$$\begin{gathered}\mathrm{D}^N y + a_{N-1}\mathrm{D}^{N-1} y + \cdots + a_0 y = \\ \mathrm{D}^M x + b_{M-1}\mathrm{D}^{M-1} x + \cdots + b_0 x\end{gathered}$$

with $M < N$. This equation can also be written in the equivalent operator form $\mathrm{L}\{y\} = x$. To obtain the transfer function of L, we take the Laplace transform of both sides and solve for $X(s)/Y(s)$, which yields

$$L_{\vec{\alpha},\vec{\gamma}}(s) = \frac{\Pi_{n=1}^{N}(s-\alpha_n)}{\Pi_{m=1}^{M}(s-\gamma_m)} \tag{2.3}$$

with parameter vectors $\vec{\alpha} = (\alpha_1, \ldots, \alpha_N)$ and $\vec{\gamma} = (\gamma_1, \ldots, \gamma_M)$, where $\{\alpha_n\}_{n=1}^N$ and $\{\gamma_m\}_{m=1}^M$ are the roots of the polynomials $s^N + a_{N-1}s^{N-1} + \cdots + a_0$ and $s^M + b_{M-1}s^{M-1} + \cdots + b_0$, respectively. The Green's function of the system is given by

$$\rho_{\vec{\alpha},\vec{\gamma}}(t) = \mathcal{L}^{-1}\left\{\frac{\Pi_{m=1}^{M}(\cdot-\gamma_m)}{\Pi_{n=1}^{N}(\cdot-\alpha_n)}\right\}(t),$$

which can be determined by explicitly computing the inverse Laplace transform. We will therefore refer to the spline-defining parameters, $\{\alpha_n\}_{n=1}^N$ and $\{\gamma_m\}_{m=1}^M$, as the poles and the zeros, respectively.

A generalized E-spline with vector of poles $\vec{\alpha}$, vector of zeros $\vec{\gamma}$ and knots $-\infty < \cdots < t_k < t_{k+1} < \cdots < +\infty$ is a function of the form

$$s(t) = \sum_{k\in\mathbb{Z}} a[k]\rho_{\vec{\alpha},\vec{\gamma}}(t-t_k) + p_{\vec{\alpha},\vec{\gamma}}(t),$$

where $p_{\vec{\alpha},\vec{\gamma}}$ is a linear combination of exponential polynomials from the null space $\mathcal{N}_{\vec{\alpha}}$ of the operator $L_{\vec{\alpha},\vec{\gamma}}$. To make the notation simpler, from now on, we will omit the indices $\vec{\alpha}, \vec{\gamma}$.

As in our illustrative example, we now consider exponential splines on a uniform grid with knots $t_k = Tk, k \in \mathbb{Z}$. In this case [22], all null-space elements $p \in \mathcal{N}_{\vec{\alpha}}$ are reproduced by $\{\rho(\cdot - Tk)\}_{k \in \mathbb{Z}}$. Consequently, from the definition of an E-spline, the shifted versions of the Green's functions form a basis of the space of generalized exponential splines.

Now, we are interested in a localized basis function. If $\vec{\gamma}$ is empty, the Green's function is a convolution of first-order ones, and we use the composition of first-order localization operators (which corresponds to the convolution of their impulse responses) to localize it, as in

$$\Delta_{\vec{\alpha},T} = \Delta_{(\alpha_1,\alpha_2,\ldots,\alpha_n),T} = T(\Delta_{(\alpha_1),T} \cdots \Delta_{(\alpha_n),T})$$

where $\Delta_{(\alpha_i),T} f(t) = \frac{1}{T}(f(t) - e^{\alpha_i T} f(t-T))$. It can be further seen [22] that the same localization operator $\Delta_{\vec{\alpha},T}$ can be used to localize $\rho_{\vec{\alpha},\vec{\gamma}}$ for arbitrary $\vec{\gamma}$, as long as $M < N$.

The function $\beta_T(t) = \beta_{\vec{\alpha},\vec{\gamma},T}(t) = \Delta_{\vec{\alpha},T}\rho(t)$ is called an exponential B-spline; it is supported in $[0, TN)$. Its Fourier transform is given by

$$\hat{\beta}_T(\omega) = \frac{1}{T^{N-1}} \prod_{k=1}^{N} \frac{1 - e^{T(\alpha_k - j\omega)}}{(j\omega - \alpha_k)} \prod_{l=1}^{M} (j\omega - \gamma_l). \tag{2.4}$$

We assume the stability condition $\alpha_l - \alpha_m \neq 2\pi kj/T$ for all distinct pure imaginary roots α_l, α_m. The T-integer shifts of the B-spline then form a Riesz basis [22, Theorem 1]; i.e., they provide a stable signal representation. In addition, one can show that the Green's function can be reconstructed as

$$\rho(t) = \sum_{k=0}^{+\infty} p_T[k] \beta_T(t - Tk),$$

where $p_T[k]$ are some suitable weights [22]. This reproduction formula ensures the completeness of $\{\beta_T(\cdot - Tk)\}_{k \in \mathbb{Z}}$ in the space spanned by the shifted Green's functions. Thus, the exponential B-splines form a stable and complete basis of this space.

To define a corresponding multiresolution analysis, we focus on the dyadic scales $T = 2^i$. The $V_{(i)}$'s are the subspaces of L_2 spanned by 2^i-shifts of the Green's function

$$V_{(i)} = \{s_i = \sum_{k \in \mathbb{Z}} c_k \rho(\cdot - 2^i k)\} \cap L_2,$$

where the c_k's are arbitrary coefficients.

$$\begin{tikzcd}
\boxed{LV_T = \operatorname{span}\{\delta(\cdot - Tk)\}} \arrow[d, shift right=2ex, "H = L^{-1}"'] \\
\boxed{V_T = \operatorname{span}\{\rho(\cdot - Tk)\}} \arrow[u, shift right=2ex, "L"'] \arrow[d, shift right=2ex, "\Delta_T\ \text{(localization)}"'] \\
\boxed{V_T = \operatorname{span}\{\beta_T(\cdot - Tk)\}} \arrow[u, shift right=2ex, "\text{(Green function reproduction)}\ \Delta_T^{-1}"']
\end{tikzcd}$$

Figure 2.5: Construction of exponential-spline spaces.

Due to the equivalence provided by the localization process and by the Green-function reproduction formula, we can also write

$$V_{(i)} = \{s_i(t) = \sum_{k\in\mathbb{Z}} c_k \varphi_i(t - 2^i k) : c \in l_2\},$$

where $\varphi_i(t)$ is the normalized scaling function $\varphi_i(t) = \beta_{2^i}(t)/||\beta_{2^i}||_{L_2}$.

We give in Figure 2.5 a diagram that summarizes the construction of the exponential-spline spaces. The space $V_{(i)}$ is included in $V_{(i-1)}$ by construction. Moreover, as we shall see in the corollary to Theorem 1, the approximation error in $V_{(i)}$ goes to zero as $i \to -\infty$. We therefore have a ladder of spaces satisfying $\overline{\bigcup V_{(i)}} = L_2$. However, in contrast with the common definition of a multiresolution analysis, the scaling function whose shifts form a basis of $V_{(i)}$ is not a dilated version of φ_0 if $i \neq 0$. Instead, at each scale i, we have a specific generating function φ_i. Similarly to the first-order case, the causal Green's function ρ belongs to all spaces $V_{(i)}$ if $\mathrm{Re}\{\alpha_l\} < 0, l = 1, \ldots, N$; else, their intersection is empty as is usually expected (i.e., $\bigcap V_{(i)} = \varnothing$).

2.3.2 Order of Approximation

To measure the quality of the E-spline approximation of a signal f, we estimate the asymptotic behavior of the approximation error as the scale $T = 2^i$ goes to zero (or, equivalently, as $i \to -\infty$). There are results from basic spline theory, applicable to the case where $\vec{\gamma}$ is empty, that imply that the error should decay like T^N as $T \to 0$ [20,30]. An exact asymptotic formula is given in [21, Theorem 2]. Here, we extend this result for the general rational case ($M < N$).

Theorem 1 *Let $L_{\vec{\alpha},\vec{\gamma}}$ be a defining operator for a linear differential system with $\dim(\vec{\alpha}) = N$, $\dim(\vec{\gamma}) = M$. Let $f \in L_2$ be a function such that $\mathrm{D}^N f \in L_2$, and let P_T denote the orthogonal projector into the exponential B-spline space V_T. Then, we have the following asymptotic formula for the approximation error as $T \to 0$:*

$$||f - P_T f||_{L_2} = C_{N,M} \cdot T^{N-M} \cdot ||L_{\vec{\alpha},\vec{\gamma}} f||_{L_2}$$

where $C_{N,M} = \frac{\sqrt{2\zeta(2(N-M))}}{(2\pi)^{N-M}}$ with $\zeta(s) = \sum_{k=1}^{\infty} 1/k^s$ (Riemann's zeta function).

The proof of this theorem is quite technical and is given in Appendix A. The following corollary ensures that the approximation order tends to zero as $T \to 0$ for any $f \in L_2$, as stated in the previous subsection.

Corollary 1 *For any function $f \in L_2$ we have $||f - P_T f|| \to 0$ as $T \to 0$.*

Proof: We use the fact that the Sobolev space W_2^N is dense in L_2. Specifically, we estimate the error from above as $||f - P_T f|| \leq ||f - f_s|| + ||f_s - P_T f_s|| + ||P_T(f_s - f)|| \leq 2||f - f_s|| + ||f_s - P_T f_s||$, where $f_s \in W_2^N$ can be chosen to be arbitrarily close to f. ■

Theorem 1 is a result for the approximation of smooth functions. For rougher functions whose Sobolev degree of smoothness is $r < N - M$, we expect that the error will only decay like $O(T^r)$. A proof for the standard wavelet case can be found in [31].

2.4 Multiresolution Basis Functions

So far, we have constructed the spaces $V_{(i)}$ spanned by the 2^i-shifts of the Green's function and have given an equivalent representation using compactly supported B-splines. In this section, we construct orthonormal and dual basis-generating functions for $V_{(i)}$, and the corresponding wavelets as well.

2.4.1 E-Spline Scaling Functions

We start with a proposition that helps us construct Riesz bases in $V_{(i)}$. First, we define

$$a_i[k] = \langle \varphi_i(\cdot), \varphi_i(\cdot - 2^i k) \rangle \tag{2.5}$$

which is the Gram sequence—or autocorrelation—of the basis $\{\varphi_i(t - 2^i k)\}_{k \in \mathbb{Z}}$. We then have that $A_i(e^{j2^i\omega}) = \sum_{k\in\mathbb{Z}} a_i[k] e^{-j2^i\omega k} = 2^{-i} \sum_{k\in\mathbb{Z}} |\hat{\varphi}_i(\omega + 2\pi k/2^i)|^2$, where $A_i(z)$ is the z-transform of $a_i[k]$. $A_i(z)$ is also referred to as the autocorrelation filter.

Proposition 1 *Let φ_i be an exponential B-spline at scale $T = 2^i$, with exponential parameter vector $\vec{\alpha}$ such that $\alpha_l - \alpha_m \neq j2\pi k/2^i, k \in \mathbb{Z}$ for all distinct purely imaginary roots α_l, α_m. Then $\{\phi_i(\cdot - 2^i k)\}_{k\in\mathbb{Z}}$, with*

$$\phi_i(t) = \sum_{k\in\mathbb{Z}} p_i(k) \varphi_i(t - 2^i k),$$

is a Riesz basis of $V_{(i)}$ if and only if $0 < c_1 \leq |P_i(e^{j\omega})| \leq c_2 < \infty$.

Proof: We know that, if $\vec{\alpha}$ satisfies the conditions of the theorem, then $\{\varphi_i(\cdot - 2^i k)\}_{k\in\mathbb{Z}}$ is a Riesz basis of $V_{(i)}$ [22, Theorem 1]. Considering the autocorrelation filter for ϕ_i, we get

$$\frac{1}{2^i} \sum_{k\in\mathbb{Z}} |\hat{\phi}_i(\omega + \frac{2\pi k}{2^i})|^2 = \frac{1}{2^i} \sum_{k\in\mathbb{Z}} |P_i(e^{j2^i\omega})|^2 \cdot |\hat{\varphi}_i(\omega + \frac{2\pi k}{2^i})|^2$$

$$= |P_i(e^{j2^i\omega})|^2 A_i(e^{j2^i\omega}).$$

Thus, the left-hand side expression is positive and bounded if and only if $|P_i(e^{j\omega})|$ is positive and bounded. ■

The compactly supported basis $\{\varphi_i(\cdot - 2^i k)\}_{k\in\mathbb{Z}}$ that we constructed in the previous section is generally not orthogonal if the order of the exponential B-spline is greater than 1. To construct a dual basis $\{\mathring{\varphi}_i(\cdot - 2^i k)\}$ of the same space $V_{(i)}$, we write in the Fourier domain the condition of biorthonormality between $\{\mathring{\varphi}_i(\cdot - 2^i k)\}$ and $\{\varphi_i(\cdot - 2^i k)\}$, from which we deduce that

$$\hat{\mathring{\varphi}}_i(\omega) = \frac{\hat{\varphi}_i(\omega)}{A_i(e^{j2^i\omega})}. \tag{2.6}$$

The dual basis is useful to project an arbitrary signal $f \in L_2$ onto $V_{(i)}$. This least-squares approximation is computed via the projection formula

$$P_i f(x) = \sum_k \langle f(\cdot), \mathring{\varphi}_i(\cdot - 2^i k)\rangle \varphi_i(\cdot - 2^i k).$$

To build an orthonormal basis $\{\varphi_{o,i}(\cdot - 2^i k)\}_{k\in\mathbb{Z}}$, we orthonormalize $\hat{\varphi}_i(\omega)$ in the Fourier domain, which yields

$$\hat{\varphi}_{o,i}(\omega) = \frac{\hat{\varphi}_i(\omega)}{\sqrt{A_i(e^{j2^i\omega})}}.$$

Since the weighting functions $(A_i(e^{j2^i\omega}))^{-1}$ and $(A_i(e^{j2^i\omega}))^{-\frac{1}{2}}$ are bounded from above and are non-vanishing, we can invoke Proposition 1 which ensures that the constructed dual and orthonormal bases are Riesz bases as well.

As noticed before, $V_{(i+1)} \subset V_{(i)}$. In particular, φ_{i+1} can be decomposed in the basis of $V_{(i)}$, which gives us the scaling relation

$$\varphi_{i+1}(t) = \sum_k h_i[k]\varphi_i(t - 2^i k). \tag{2.7}$$

Taking the Fourier equivalent of this formula and plugging it in (2.4), we find that

$$H_i(e^{j2^i\omega}) = \frac{\hat{\varphi}_{i+1}(\omega)}{\hat{\varphi}_i(\omega)} = 2\frac{c_i}{c_{i+1}} \cdot \prod_{k=1}^{N} \frac{1 + e^{2^i(\alpha_k - j\omega)}}{2},$$

where $c_i = ||\beta_i||_{L_2}$ is a normalizing constant.

We note that this refinement equation was already given in [24] for the standard (non-rational) case. The more general case of a rational operator and of an arbitrary integer dilation factor m (not necessarily a power of two) is considered in [22].

Interestingly, the refinement filter $H_i(z) = 2(c_i/c_{i+1})\cdot\prod_{k=1}^{N}((1 + e^{2^i\alpha_k}z^{-1})/2)$ is now different for each scale; also, it is the same irrespective of the zero vector $\vec{\gamma}$ up to multiplication by a constant. The dual refinement filter $\tilde{H}_i(z)$ is given by

$$\tilde{H}_i(z) = \frac{A_i(z)}{A_{i+1}(z^2)} H_i(z), \tag{2.8}$$

and the dual scaling function $\mathring{\varphi}_i$ satisfies the relation

$$\mathring{\varphi}_{i+1}(t) = \sum_k \tilde{h}_i[k]\mathring{\varphi}_i(t - 2^i k).$$

By substituting (2.7) into the expression for the autocorrelation sequence at scale $i+1$ given by (2.5), it is easy to express $A_{i+1}(z^2)$ in terms of $A_i(z)$ and $H_i(z)$ as

$$\begin{aligned} A_{i+1}(z^2) = \frac{1}{2}(A_i(z)H_i(z)H_i^*(z^{-1}) \\ +A_i(-z)H_i(-z)H_i^*(-z^{-1})). \end{aligned} \tag{2.9}$$

2.4.2 E-Spline Wavelets

The inclusion $V_{(i)} \subset V_{(i-1)}$ allows us to uniquely introduce the orthogonal complements $W_{(i)}$ such that

$$V_{(i)} \oplus W_{(i)} = V_{(i-1)}.$$

The residual space $W_{(i)}$ is the orthogonal complement of $V_{(i)}$ in $V_{(i-1)}$. It plays the same role as in the case of semi-orthonormal wavelet functions, which are orthonormal across scales, but not necessarily within a given scale. Thus, at a given scale $i+1$, we are looking for a wavelet ψ_{i+1} of the form

$$\psi_{i+1}(t) = \sum_k g_i[k]\varphi_i(t - 2^i k) \tag{2.10}$$

that is orthogonal to V_{i+1}. In other words, for all $k \in \mathbb{Z}$, we must have

$$\langle \psi_{i+1}(\cdot), \varphi_{i+1}(\cdot - 2^{i+1}k)\rangle = 0.$$

From (2.10) and (2.7), and after having expressed the z-transform of this orthogonality relation, we obtain

$$G_i(z)H_i^*(z^{-1})A_i(z) + G_i(-z)H_i^*(-z^{-1})A_i(-z) = 0.$$

In contrast to the classical wavelet theory, all filters now depend on the scale. Hence, the solution $G_i(z)$ depends on the scale, too. Its general form is

$$G_i(z) = -zQ_i(z^2)H_i^*(-z^{-1})A_i(-z),$$

where $Q_i(z)$ is a Laurent polynomial in z. We show now how to choose Q_i to get a Riesz basis $\{\psi_i(\cdot - 2^i k)\}_{k\in\mathbb{Z}}$ of W_i, while noting that we cannot directly apply Proposition 1 since φ_{i-1} and ψ_i belong to different functional spaces.

Proposition 2 *$\{\psi_i(\cdot - 2^i k)\}_{k\in\mathbb{Z}}$ is a Riesz basis of W_i if and only if the filter Q_{i-1} is bounded and non-vanishing on the unit circle.*

Proof: The autocorrelation filter is

$$\begin{aligned}
R_i(e^{2^i j\omega}) &= \sum_{k\in\mathbb{Z}} |\hat{\psi}_i(\omega + \frac{2\pi}{2^i}k)|^2 \\
&= G_{i-1}(e^{2^{i-1}j\omega})A_{i-1}(e^{2^{i-1}j\omega}) \\
&\quad + G_{i-1}(e^{j\pi+2^{i-1}j\omega})A_{i-1}(e^{j\pi+2^{i-1}j\omega}) \\
&= A_{i-1}(e^{2^{i-1}j\omega})A_{i-1}(e^{j\pi+2^{i-1}j\omega})|Q_{i-1}(e^{2^{i-1}j\omega})|^2 \\
&\quad \cdot(H_{i-1}(e^{2^{i-1}j\omega})A_{i-1}(e^{2^{i-1}j\omega}) \\
&\quad + H_{i-1}(e^{j\pi+2^{i-1}j\omega})A_{i-1}(e^{j\pi+2^{i-1}j\omega})) \\
&= A_{i-1}(e^{2^{i-1}j\omega})A_{i-1}(e^{j\pi+2^{i-1}j\omega}) \\
&\quad \cdot A_i(e^{j2^i j\omega})|Q_{i-1}(e^{2^{i-1}j\omega})|^2
\end{aligned}$$

Thus, for $\{\psi_i(t-2^i k)\}_{k\in\mathbb{Z}}$ to be a Riesz basis in W_i, there should exist $c_1, c_2 > 0$ such that $c_1 \leq |Q_{i-1}(e^{2^{i-1}j\omega})|^2 < c_2$. ■

In order to perform a hierarchical decomposition in our new wavelet basis, we would like to apply Mallat's filterbank algorithm. Therefore, we build the filterbank shown in Figure 2.6 and impose a perfect reconstruction condition. Following the same construction procedure as before, we obtain the dual wavelet relation

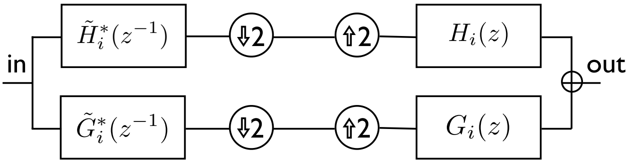


Figure 2.6: Two-channel filterbank used for the wavelet transform.

$$\overset{\circ}{\hat{\psi}}_{i+1}(\omega) = \tilde{G}_i(e^{2^i j\omega})\overset{\circ}{\hat{\varphi}}_i(\omega),$$

where the dual wavelet filter $\tilde{G}_i$ is given by

$$\tilde{G}_i(z) = -z\frac{\tilde{H}_i^*(-z^{-1})}{A_i(-z)Q_i^*(z^{-2})}. \tag{2.11}$$

We typically use $(\tilde{G}_i, \tilde{H}_i)$ on the analysis side and (G_i, H_i) on the synthesis side, so that the synthesis wavelet is compactly supported. However, it is possible to interchange these filter pairs; this corresponds to performing a dual E-spline wavelet transform.

For the orthonormal E-spline wavelet transform, the same filters are used on the analysis and synthesis side. These filters are denoted by $(G_{o,i}, H_{o,i})$ and are given by

$$H_{o,i}(z) = \sqrt{\frac{A_i(z)}{A_{i+1}(z^2)}} H_i(z)$$

$$G_{o,i}(z) = -z^{-1} H^*_{o,i}(-z^{-1}).$$

The constructed scaling functions and wavelets at each scale i belong to the space C^{N-M-2}. Indeed, they all can be expressed as weighted sums of the shifts of $\varphi_{i-1} \in C^{N-M-2}$ [22, Section 3]. The scaling and the wavelet filters are either FIR (B-spline case) or IIR with exponential decay (dual or orthogonal case); thus, the basis functions are either compactly supported, or decay exponentially.

2.5 Properties of E-Splines and E-Spline Wavelets

2.5.1 Reproduction of Exponential Polynomials

The following proposition generalizes the polynomial-reproduction property of the classical wavelet theory:

Property 1 *Suppose that $\vec{\alpha}$ contains a root α_0 of multiplicity m. Then, for $n = 0, \dots, m-1$ the exponential monomial $t^n e^{\alpha_0 t}$ has the B-spline representation*

$$t^n e^{\alpha_0 t} = \sum_{k \in \mathbb{Z}} p_{\vec{\alpha},i,n}[k] \beta_{\vec{\alpha},\vec{\gamma},2^i}(t - 2^i k),$$

where $p_{\vec{\alpha},i,n}[k]$ are suitable coefficients.

For the arbitrary parameter vector $\vec{\alpha}$ consisting of N_d distinct roots of multiplicity $m_k, k = 1, \dots, N_d$, the direct corollary is that the exponential polynomials

$$p_{\vec{\alpha}}(t) = \sum_{k=1}^{N_d} \sum_{n=0}^{m_k - 1} c_k t^n e^{\alpha_k t}$$

that constitute the null space $N_{\vec{\alpha}}$ of the operator L can be reproduced with B-splines. The result follows from Proposition 2 in [21].

2.5.2 Vanishing Exponential Moments

In conventional wavelet theory, the vanishing-moment property of the wavelet is closely related to the ability of the scaling function to reproduce polynomials. This can be generalized as well to our case of rational operators.

Property 2 *For each scale* $i \in \mathbb{Z}$*, shift* $t_0 \in \mathbb{R}$*, and degree* $n = 0, \dots, m_k - 1$*, the analysis wavelet satisfies*

$$\int_{-\infty}^{\infty} t^n e^{\alpha_k t} \mathring{\psi}_i^*(t - t_0) \mathrm{d}t = 0,$$

where m_k *is the multiplicity of* α_k*. In other words, the analysis wavelet* $\mathring{\psi}_i^*$ *has* N *vanishing exponential moments that correspond to the basis functions of the null space of* L*. Equivalently, for each* $p_{\vec{\alpha}}(t) \in N_{\vec{\alpha}}$*, we have*

$$\int_{-\infty}^{\infty} p_{\vec{\alpha}}(t) \mathring{\psi}_i^*(t - t_0) \mathrm{d}t = 0.$$

This proposition becomes obvious if we remember that $W_{(i)}$ is an orthogonal complement of $V_{(i)}$; thus, as long as the null space of L can be reconstructed with the basis functions of $V_{(i)}$, $\mathring{\psi}_i$ is orthogonal to it. In addition, the null space is shift-invariant [22]; i.e., $p_{\vec{\alpha}}(t - t_0) \in N_{\vec{\alpha}}$, which completes the proof.

2.5.3 Operator-Like Wavelets

The following theorem is a key result of the present work. It states that our new wavelets behave qualitatively like the differential operator from which the multiresolution analysis is derived.

Theorem 2 *Let* $\{\psi_{i,k}\}_{i,k \in \mathbb{Z}}$ *be an E-spline wavelet basis of* L_2*. Then, there exists a sequence* $\{\phi_i\}_{i \in \mathbb{Z}}$ *of E-spline scaling functions of order* $2N$ *such that* $\langle f, \mathring{\psi}_i(\cdot - t_0) \rangle = \mathrm{L}\{f * \phi_i\}(t_0)$*; in addition,* $\{\phi_i\}_{i \in \mathbb{Z}}$ *generates a multiresolution analysis of* L_2*. The wavelet coefficients of* f *are therefore the samples of the smoothed versions of* $L_{\vec{\alpha},\vec{\gamma}} f$*.*

Proof: Consider the wavelet coefficient $\langle f, \mathring{\psi}_i(\cdot - t_0) \rangle = \int L(j\omega) \hat{f}(\omega) \frac{\hat{\mathring{\psi}}_i^*(\omega)}{L(j\omega)} \cdot e^{j\omega t_0} \mathrm{d}\omega$. We now define $\hat{\phi}_i(\omega) = \hat{\mathring{\psi}}_i^*(\omega)/L(j\omega)$ and study the behavior of this function. Combining (2.8) and (2.11), we get

$$\hat{\phi}_i(\omega) = \frac{\tilde{G}_{i-1}^*(e^{-j2^{i-1}\omega})}{L(j\omega)} \hat{\mathring{\varphi}}_{i-1}^*(\omega)$$

$$= -e^{-j2^{i-1}\omega}\frac{H_{i-1}(-e^{j2^{i-1}\omega})}{L(j\omega)A_i(e^{j2^i\omega})Q_{i-1}(e^{j2^i\omega})}\hat{\mathring{\varphi}}^*_{i-1}(\omega).$$

In order to further simplify this expression, we notice that $\hat{\varphi}_i(\omega) = \frac{H_i(-e^{j2^i\omega})}{L(j\omega)}$. Hence, taking (2.6) into account, we have

$$\hat{\phi}_i(\omega) = -e^{-j2^{i-1}\omega}\frac{A_{i-1}(e^{j2^{i-1}\omega})}{A_i(e^{j2^i\omega})Q_{i-1}(e^{j2^i\omega})}|\hat{\mathring{\varphi}}_{i-1}(\omega)|^2.$$

In this expression, $|\hat{\mathring{\varphi}}_{i-1}(\omega)|^2$ corresponds to an exponential B-spline with augmented parameters $(\vec{\alpha} : -\vec{\alpha}^*), (\vec{\gamma} : -\vec{\gamma}^*)$. The discrete filters A_{i-1}, A_i, and Q_{i-1}, are bounded and do not vanish on the unit circle. According to Proposition 1, $\{\phi_i(\cdot - 2^{i-1}k)\}_{k\in\mathbb{Z}}$ is a Riesz basis of the space generated by $\mathring{\varphi}_{i-1} * \mathring{\varphi}^T_{i-1}$, which implies that the ϕ_i's generate a multiresolution analysis of L_2. ■

When $\gamma_k \neq 0$ for all k and $\alpha_l \neq 2^{-i} \cdot 2\pi kj$ for $1 \leq l \leq N, k \neq 0$, then the filters ϕ_i are lowpass. The sufficient condition is that φ_i is a valid E-spline scaling function with $\alpha_l = 0$ for at least one l. The validity statement then implies all other requirements. This is exactly the case for the examples in Section VI.

2.6 Implementation and Examples

In this section, we describe the filterbank implementation for the wavelet decomposition and reconstruction based on E-splines. We also discuss possible algorithms for the calculation of the autocorrelation filter. Finally, we show examples of exponential B-splines, illustrating the concepts of this chapter.

2.6.1 Filterbank Implementation

In practice, it is more efficient to work with discrete sequences and filterbanks, rather than with continuous-time signals and projections. In the conventional wavelet theory, this idea leads to Mallat's fast filterbank algorithm [9]. It is easy to see that this algorithm can be applied to the E-spline-wavelet case as well; however, the filters $(G_i, \tilde{G}_i)$ and $(H_i, \tilde{H}_i)$ must be precalculated for each iteration. IIR filters can be implemented recursively (as in [17]) or approximated with FIR filters of sufficient length. A simple alternative is to evaluate these in the Fourier domain using the FFT algorithm [32].

The algorithm first interpolates the given samples $f[k]$ with the exponential B-splines at the initial scale $i = 0$,

$$f_0(t) = \sum_{k \in \mathbb{Z}} c[k] \varphi_0(t-k), \tag{2.12}$$

where $c[k] = (p * f)[k]$, and where $P(z)$ is the interpolation prefilter given by

$$P(z) = \frac{1}{\left(\sum_k \varphi_0(k) z^{-k}\right)}. \tag{2.13}$$

The $c[k]$'s are used to initialize the hierarchical decomposition.

In the unlikely event where $P(z)$ is not stable, which is also sometimes the case with conventional splines, we propose to replace the filter by a generalized quasi-interpolant [33] that is specifically designed to reproduce the exponential polynomials.

The (quasi-)interpolation model (2.12) implies that $f_0 \in V_0$. Thus, the $c[k]$'s given by the initialization procedure are also the coefficients of the projection of f_0 into V_0. Indeed,

$$\langle f_0, \mathring{\varphi}_0(\cdot - k)\rangle = \sum_{k_0 \in \mathbb{Z}} c[k_0] \langle \varphi_0(\cdot - k_0), \mathring{\varphi}_0(\cdot - k)\rangle = c[k],$$

because of the biorthogonality of φ_0 and $\mathring{\varphi}_0$.

For each iteration, we need to know the autocorrelation filter $A_i(z)$ at the current scale i. Then, the wavelet filters $G_i(z), \tilde{G}_i(z), G_{o,i}$, and the scaling filters $H_i(z), \tilde{H}_i(z), H_{o,i}(z)$, can be computed using the explicit expressions given in Section IV. Equation (2.9) allows us to find $A_{i+1}(z)$ from $A_i(z)$. For $A_0(z)$, we recall that the autocorrelation function of the exponential B-spline corresponds to the scaled symmetrical B-spline with parameters $(\vec{\alpha}, -\vec{\alpha}^*), (\vec{\gamma}, -\vec{\gamma}^*)$ [21].

We see that, both for the autocorrelation filter and for the interpolation prefilter, we need to compute the exponential B-spline samples in the time domain. This can be done by evaluating the samples of the Green's function and by applying finite-difference operators to them.

In the first order case, the procedure is especially simple because $A_i(z) = 1$. In that case, $H_i(z) = \tilde{H}_i(z) = H_{\alpha,i}(z) = \frac{1}{\sqrt{1+|e^{2^{i+1}\alpha}|}} \cdot (1 + e^{2^i \alpha} z^{-1})$ and $G_i(z) = \tilde{G}_i(z) = \frac{1}{\sqrt{1+|e^{2^{i+1}\alpha}|}} (e^{2^i \alpha^*} - z^{-1})$, where we observe the simple dependence on the scale. Interestingly, the filters tend to those of the Haar system (sum and difference) as the scale gets finer ($i \to -\infty$).

2.6.2 Computation of the Green-Function Samples

For $\mathrm{Re}\{\alpha_k\} > 0$, the Green's function grows exponentially. To deal with this issue, we use anticausal Green's functions for all k such that $\mathrm{Re}\{\alpha_k\} > 0$, tuning the sign in the partial fraction decomposition. Having decomposed $\hat{\rho}(\omega)$ into partial fractions

$$\hat{\rho}(\omega) = \sum_{n=1}^{N_d} \sum_{l=1}^{m_n} \frac{c_{n,l}}{(j\omega - \alpha_n)^l},$$

we express the components with $\mathrm{Re}\{\alpha_k\} > 0$ as

$$\frac{c_{n,l}}{(j\omega - \alpha_n)^l} = \frac{(-1)^l c_{n,l}}{(-j\omega - (-\alpha_n))^l}.$$

In the time domain, the latter corresponds to anticausal Green's function with parameter α_n.

We further notice that, to calculate N samples of the B-spline, we need $2N+1$ values of the Green's function; thus, $\rho(t)$ should be calculated for $t_k = -N, \ldots, N$.

2.6.3 Computation of the Exponential B-Spline Samples and Filters

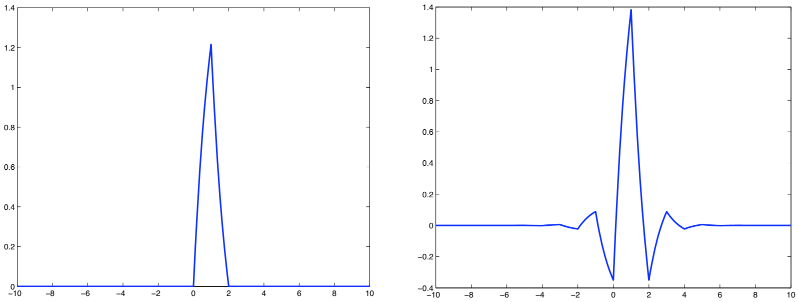
We apply N finite-difference operators $\Delta_{(\alpha_1),2^i}, \ldots, \Delta_{(\alpha_N),2^i}$ to the Green's function samples and obtain N samples of the exponential B-spline. The value $c_i = \sqrt{a_i[0]}$ is equal to the norm of the B-spline and is used for normalization. The autocorrelation filter is obtained by taking the value of the trigonometric polynomial

$$A_i(e^{2^i j\omega}) = \sum_{k=-N}^{N} a_i[k] e^{2^i j\omega k}.$$

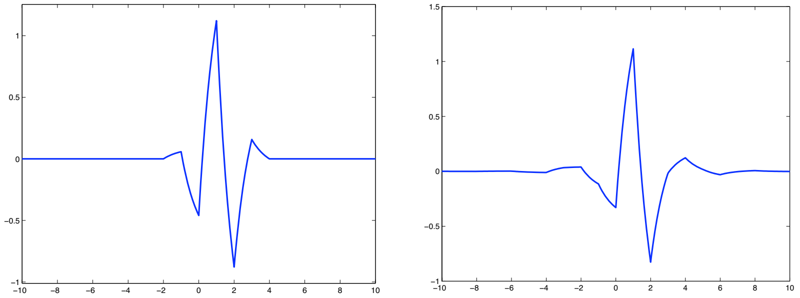
Since the samples of the exponential B-spline are known, the interpolation prefilter can be computed from (2.13).

2.6.4 Examples

For our first example, we choose $\vec{\alpha} = (0, -1)$ and $\vec{\gamma}$ empty. We show the basic and dual scaling functions and the wavelets in Figures 2.7(a)–2.7(d). It can be seen that these functions are non-symmetric, piecewise exponential, and yet continuous because $N - M = 2$. The two synthesis functions φ_0 and ψ_1 are



(a) Scaling function $\varphi_0(t)$ (b) Dual scaling function $\mathring{\varphi}_0(t)$

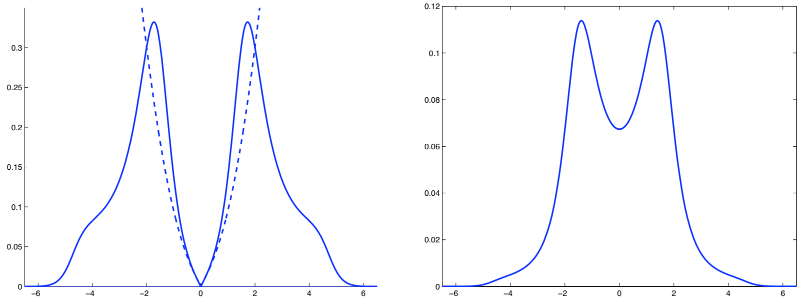


(c) Synthesis wavelet $\psi_1(t)$ (d) Analysis wavelet $\mathring{\psi}_1(t)$

Figure 2.7: Example 1: Basic and dual scaling functions and wavelets; $\vec{\alpha} = (0, -1)$, $\vec{\gamma}$ is empty.

compactly supported, while their dual counterparts $\overset{\circ}{\varphi}_0$ and $\overset{\circ}{\psi}_1$ are exponentially decaying.

The spectrum of the analysis wavelet $\overset{\circ}{\psi}_1$ is displayed in Figure 2.8(a). The frequency response of the corresponding operator $\mathrm{L} = \mathrm{D} + \mathrm{D}^2$ is overlaid with a dashed line. It can be seen that the two plots are well matched around the origin (at the location of the pole $j\omega = \alpha_1 = 0$), which illustrates the differential-operator property (see Theorem 2). We show the spectrum of the smoothing kernel $|\hat{\phi}_1(\omega)| = |\hat{\overset{\circ}{\psi}}{}^*_1(\omega)/L(j\omega)|$ in Figure 2.8(b); it is clearly lowpass and decays rapidly for high frequencies.

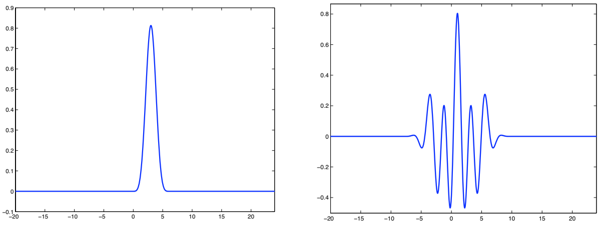


(a) Spectrum of the analysis wavelet $|\hat{\overset{\circ}{\psi}}_1(\omega)|$ and $|L(j\omega)|$. (b) Spectrum $|\hat{\phi}_1(\omega)|$ of the smoothing kernel

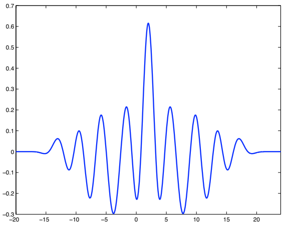
Figure 2.8: Example 1: Spectral behavior of the analysis wavelet and of the smoothing kernel; $\vec{\alpha} = (0, -1)$, $\vec{\gamma}$ is empty

With our second example, we illustrate the spectral behavior of the wavelets and of the scaling functions. For this purpose, we choose the parameters $\vec{\alpha} = (-\frac{5\pi}{8}j, -\frac{5\pi}{8}j, \frac{5\pi}{8}j, \frac{5\pi}{8}j, 0, 0)$ and $\vec{\gamma} = (-5j, 5j)$ to be pure imaginary and to have Hermitian symmetry, which ensures that the time-domain functions are real. In Figure 2.9, we show the scaling function φ_0 and the wavelets ψ_1 and ψ_2 in the time domain; these have a strong oscillatory character, with a larger number of lobes as the scale gets coarser. It is clearly apparent that ψ_2 is not a dilate of ψ_1. Figures 2.7 and 2.9 illustrate the diversity of shapes that can be obtained with E-spline scaling functions and wavelets.

The plots in Figure 2.10 show the effect of the poles $\vec{\alpha}$ and of the zeros $\vec{\gamma}$ on the frequency response. In Figure 2.10(a), the pole $\alpha_1 = -\frac{5\pi}{8}j$ produces a peak of the scaling-function spectrum near $\omega = -\frac{5\pi}{8}j$ and makes it vanish with

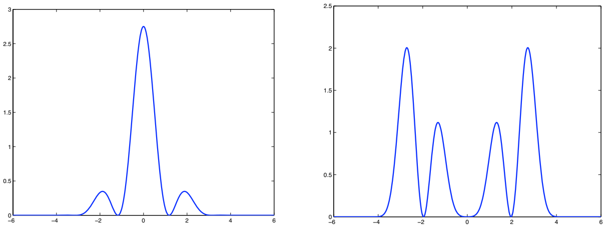


(a) Scaling function $\varphi_0, i = 0$ (b) Wavelet $\psi_1, i = 1$



(c) Wavelet $\psi_2, i = 2$

Figure 2.9: Example 2: The scaling function at the scale $i = 0$ and the wavelets at the scales $i = 1, i = 2$; $\vec{\alpha} = (-\frac{5\pi}{8}j, -\frac{5\pi}{8}j, \frac{5\pi}{8}j, \frac{5\pi}{8}j, 0, 0)$, $\vec{\gamma} = (-5j, 5j)$.



(a) Scaling function spectrum $|\hat{\varphi}_1(\omega)|, i = 1$ (b) Wavelet spectrum $|\hat{\psi}_1(\omega)|, i = 1$

Figure 2.10: Example 2: The spectrum of the scaling function and the spectrum of the wavelet; $\vec{\alpha} = (-\frac{5\pi}{8}j, -\frac{5\pi}{8}j, \frac{5\pi}{8}j, \frac{5\pi}{8}j, 0, 0)$, $\vec{\gamma} = (-5j, 5j)$.

periodicity π. There is also a peak at the origin because of the pole at $\omega = 0$. The zero $\gamma_1 = -5j$ makes the frequency response vanish at $\omega = -5$. In contrast, the wavelet spectrum (Figure 2.10(b)) vanishes at $j\omega = \alpha_l$, as a result of the vanishing exponential moment and of the linear-differential-operator properties of E-spline wavelets.

2.7 Conclusion

We have introduced a new procedure for constructing wavelet-like bases from linear differential operators. Similar to analog filters, the wavelet spaces are characterized by the poles and zeros of the underlying operator. The wavelets come in three different flavors: basic (B-spline), dual, and orthonormal. We have studied the approximation-order properties of the multiresolution analysis. These wavelet bases possess powerful properties, including the ability to kill exponential polynomials, which generalizes the notion of the vanishing moments found in the conventional wavelet theory. More importantly, they essentially behave as multiscale versions of the underlying operator. We recover the polynomial B-spline case by choosing the parameter $\vec{\alpha} = (0, \ldots, 0)$, where the corresponding operator is the N-th derivative.

The constructed wavelets are not dilates of a single function anymore; however, they still can be implemented using a non-stationary version of Mallat's fast filterbank algorithm.

The proposed framework should be of interest for signal-processing applications. It might be well suited to signals that are not predominantly lowpass but that have substantial energy concentrations in some frequency bands. In particular, it offers the possibility of adapting the model for a given class of signals, by selecting roots that fit the natural resonances of the data.

Chapter 3

Imaging Brain Function

Abstract — We would like to apply the new wavelet framework to the studies in brain functional localization in fMRI. Our goal in this chapter is twofold:

- to give a brief overview of the brain anatomy and of the basic principles of neuronal function that are relevant to imaging;
- to identify the differential operator that links the stimulus with the measured fMRI signal. To this end, we give a short introduction to fMRI and present the balloon/windkessel model, which we use as a basis to derive the expression for the linear operator.

For more detailed information, the reader is referred to [34, 35].

3.1 Human Brain: Anatomy And Structure

The human brain is made up of about 100 billions of neuronal cells (neurons) that communicate through a complex network of channels. Each neuron (see Fig. 3.1) has about 10000 communication channels: the receiver channels, called dendrites, and the transmitter channels, which are formed by splitting the end of a relatively long fiber (axon). The cellular bodies have grayish color and form the gray matter, while the axons are surrounded by a myeline sheath of white color and constitute the white matter. Large-scale anatomical regions of the brain are called the brain lobes (see Fig. 3.2).

The communication of neurons is based on action potentials. If a cell receives more excitatory impulses than inhibitory ones above a certain threshold, it

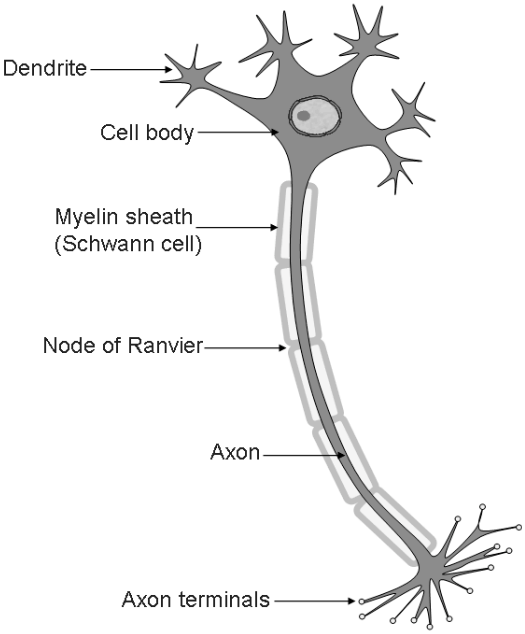


Figure 3.1: A neuronal cell (image taken from: http://www.steve.gb.com/science/nervous_system.html).

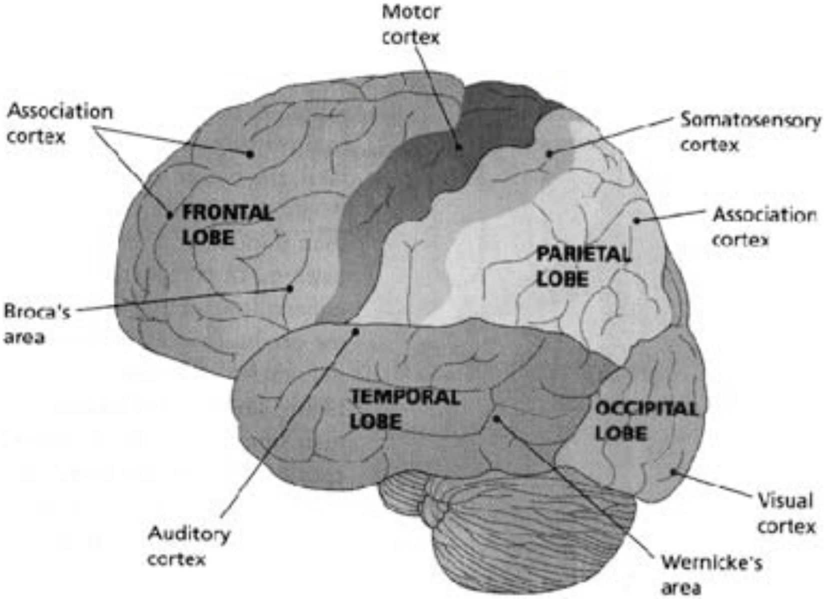


Figure 3.2: Brain anatomy (image taken from http://www.nyas.org).

transmits the charge down the axon — this process is called the firing of a neuron. Firing reduces the free energy, which needs to be restored; through a chain of chemical reactions, the restoration process boils down to oxygen and glucose metabolism. Studies in animals and humans have shown a close correspondence between local neural activity and glucose metabolism [36].

The delivery of glucose and oxygen to the neuronal cells is performed by the cerebral blood flow (CBF). The value of CBF is defined as the rate of delivery of arterial blood to capillary beds of volume of tissue, with an average of 60 mg/g-min. Note that CBF does not include the blood that transits the region through arteries and veins. Already in 1890, it was suggested that local CBF increases in active regions [37]. This hypothesis proved to be true; moreover, the blood vessel density in human brain, to some extent, reflects its functional organization. Although neurovascular coupling still remains an active area of research, many modern techniques rely on measurements of CBF rather than on that of direct electrical activity.

Two another important quantities in hemodynamics are the cerebral metabolic rate of oxygen ($CMRO_2$) and the cerebral blood volume (CBV). $CMRO_2$ grows slightly with the neural activity, but its increase is significantly smaller than the gain in CBF. CBV measures the fraction of volume occupied by blood vessels, and is usually equal to 4% in the resting state. Following the change in CBF, CBV can fluctuate due to dilation or contraction of the vessels.

3.2 Functional Magnetic Resonance Imaging

Brain imaging modalities can be generally classified into two categories: those measuring the electrical activity directly, and those observing the coupled chemical or hemodynamic effects. The first category includes the fully non-invasive techniques like EEG (electroencephalography) and MEG (magnetoencephalography) that suffer from very poor spatial resolution, as well as very invasive single unit recording methods that are only applicable in animal studies. ESM (electrical stimulation mapping) is often used in neurosurgery. The second category includes PET(positron emission tomography)/SPECT(single photon emission computed tomography), OIS (optical imaging of intrinsic signals), OT (optical tomography), NIRS (near-infrared spectroscopy), SSEP (somatosensory evoked potentials) and fMRI (functional magnetic resonance imaging) methods. These techniques are compared in terms of spatiotemporal resolution, invasiveness and number of publications in Fig. (3.3). For a non-invasive technique, the achievable resolution in fMRI is promisingly good, which explains the growing

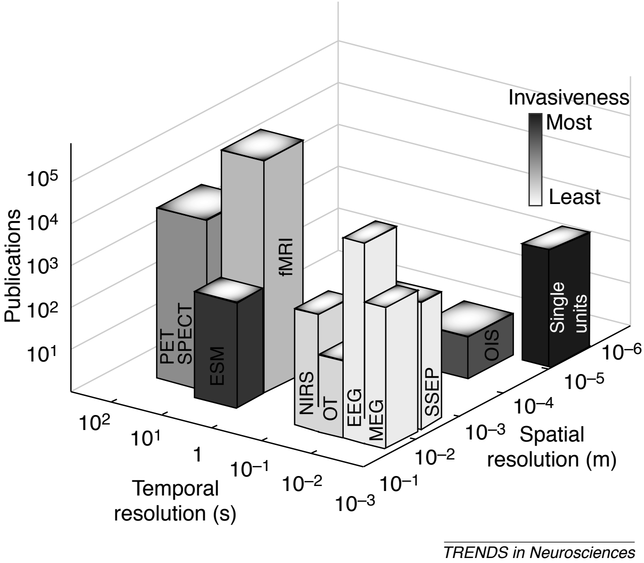


Figure 3.3: Comparing functional imaging modalities: temporal resolution, spatial resolution, invasiveness and number of publications (image taken from [38]).

amount of related publications. In the next subsection, we briefly describe the underlying physical principle of magnetic resonance and the ideas that make it useful for imaging.

3.2.1 Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) has revolutionized medical research. Its popularity can be explained by its non-invasiveness, good spatiotemporal resolution and high SNR in comparison with other imaging methods. The physical principle behind MRI is the nuclear magnetic resonance phenomenon.

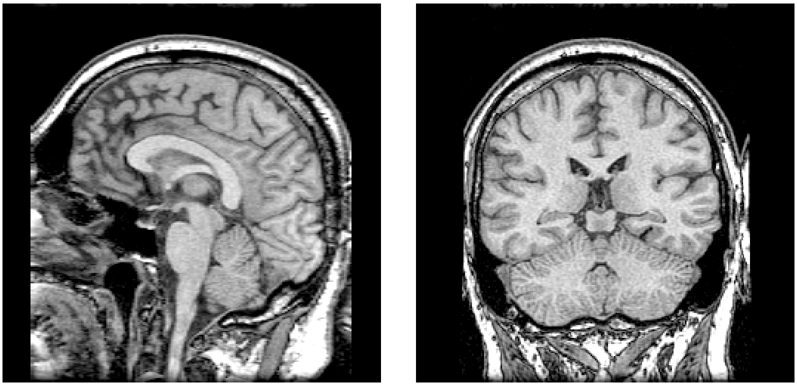
When exposed to a strong static magnetic field B_0, certain atoms (e.g., hydrogen nuclei, which we will be referring to later in this chapter) align their spins parallel with the field vector, resulting in a net bulk magnetization. Following the application of a selective, spatially varying gradient field, the protons in the area (slice) of interest have a fixed resonance frequency ω_0. An excitation radio-frequency (RF) field B_1 is applied; it precesses at ω_0 around B_0 and is designed to induce resonance. As a result, the spins of the particles of interest start precessing around B_0 and the bulk magnetization gets a component that is orthogonal to B_0 and therefore becomes measurable. Spatially-varying gradient fields are used to encode the position of the particle in its precession frequency or phase. In this way, a readout can be performed along a line in the spatial Fourier space (k-space); the spins eventually undergo the relaxation process and return to their equilibrium position, where they are aligned parallel to B_0. The whole process is repeated until there are enough k-space samples to perform a reconstruction. Examples of brain MR image slices are shown in Fig. 3.4.

The whole procedure requires about 1 minute to obtain a 128×128 image with reasonable quality. For dynamic imaging, one has to turn to faster techniques to improve temporal resolution. In 1.5T scanners, the gradient-echo imaging sequence is widely applied. This technique is characterized by a smaller bulk magnetization flip angle during excitation, combined with the use of a reversed dephasing-compensator gradient prior to readout. For higher magnetic fields, an even faster echo-planar imaging (EPI) method uses a single excitation pulse and a swinging readout gradient that constantly refocuses spins that went out of phase, so that the relaxation period is significantly extended and all readouts can be done at once. As a result, a 128×128 image of reasonable quality and sub-millimeter spatial resolution can be obtained in less than a second. The drawback of EPI is its sensitivity to the magnetic field non-uniformities; however, this is also the reason that makes it work well for functional imaging, as described in the next subsection.

Importantly, the SNR of MRI measurements is proportional to scan time and to voxel size, so that an optimal trade-off should be found for the experiment at hand.

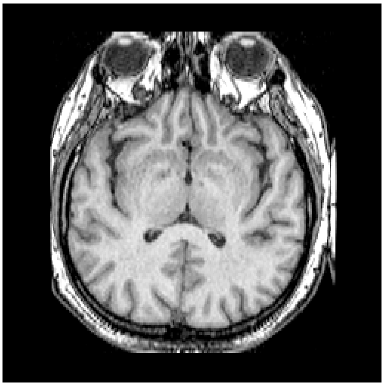
3.2.2 The BOLD effect

It is possible to observe neuronal activity in MRI due to the blood-oxygenation-level-dependent (BOLD) effect that was discovered by Ogawa et al. [39]. The origin of this effect is that hemoglobin is diamagnetic when oxygenated and paramagnetic when deoxygenated. The presence of deoxyhemoglobin alters the local magnetic susceptibility, creating magnetic field distortions within and around the blood vessels and weakening the MR signal. As mentioned in Section 3.1, after the neural activation, the increase in local CBF is much higher than the growth in $CMRO_2$. This implies a smaller level of deoxyhemoglobin and there-



(a) Sagittal slice

(b) Coronal slice



(c) Transaxial slice

Figure 3.4: MRI brain image slices.

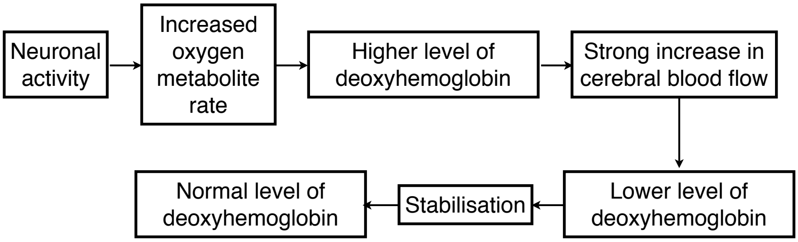


Figure 3.5: The mechanism of the BOLD effect.

fore, stronger MR signal. The overall scheme is presented in Fig. 3.5.

3.2.3 The Hemodynamic Model

Quantitative description of measurements in fMRI requires a model that links the BOLD signal to the stimulus. We adopt the linearized model of Friston et al. [35].

Our hemodynamic model is a combination of the so-called balloon/windkessel model [40, 41] with a model that links synaptic activity and changes in regional blood flow. The auxiliary variable[1] s represents the flow-inducing signal; the flow f_{in} satisfies

$$\dot{f}_{\text{in}} = s.$$

The rate of change of s is a sum of three terms that ensure a behavior similar to experimental observations:

$$\dot{s} = \epsilon u - \frac{s}{\tau_s} - \frac{f_{\text{in}} - 1}{\tau_f}.$$

Here, u is the stimulus function, ϵ is the neuronal efficacy, τ_s and τ_f are parameters that are chosen empirically. The second term ensures the natural decay of the flow-inducing signal, while the third term represents the flow feedback auto-control loop.

As veins are the most distensible vessels, almost all of the CBV change happens in the venous volume v. According to the balloon model, the rate of

---

[1] The meaning of the various variables and parameters are summarized in Table 3.1. Note that all variables are expressed in normalized form; i.e., relative to resting values. Full details can be found in [35].

Table 3.1: List of the variables and constants involved in the hemodynamic model with their typical values.

symbol	meaning	typical value	standard deviation
u	stimulus	–	–
s	flow inducing signal	–	–
f_{in}	blood flow	–	–
v	normalized venous volume	–	–
q	normalized deoxy-hemoglobine content	–	–
ϵ	neuronal efficacy	0.54	0.085
τ_s	signal decay	1.54	0.169
τ_f	autoregulation	2.46	0.212
τ_0	transit time	0.98	0.169
α	balloon stiffness	0.33	0.034
E_0	oxygen extraction fraction	0.34	0.043
V_0	resting blood volume fraction	1	–
k_1	BOLD constant 1	$7E_0$	–
k_2	BOLD constant 2	2	–
k_3	BOLD constant 3	$2E_0 - 0.2$	–

change of v is given by

$$\dot{v} = \frac{1}{\tau_0}(f_{\text{in}} - f_{\text{out}}),$$

where f_{out} is the outflow and τ_0 is the mean transit time. The vessel behavior is described by the windkessel model, so that

$$f_{\text{out}} = v^{\frac{1}{\alpha}}.$$

The BOLD signal is modeled as

$$\text{BOLD}_{\text{non-linear}} = V_0(k_1(1-q) + k_2(1-\frac{q}{v}) + k_3(1-v)), \tag{3.1}$$

where V_0 is the resting blood volume fraction, q is the deoxyhemoglobin voxel content and k_1, k_2 and k_3 are constants. The summation in (3.3) accounts for the extra- and intravascular signals. The variable q is the difference between the delivered and expelled deoxyhemoglobin in the venous compartment:

$$\dot{q} = \frac{1}{\tau_0}(f_{\text{in}}\frac{E(f_{\text{in}}, E_0)}{E_0} - f_{\text{out}}\frac{q}{v}).$$

Here, E_0 is the resting net oxygen extraction fraction by the capillary bed, and $E(f_{\text{in}}, E_0) = 1 - (1-E_0)^{\frac{1}{f_{\text{in}}}}$ is the fraction of oxygen extracted from the inflowing blood.

3.2.4 The Linear Approximation To the Hemodynamic System

Mathematically, our hemodynamic model corresponds to the non-linear state-space definition of a system with four state variables $\{s, f_{\text{in}}, v, q\}$

$$\begin{cases} \dot{s} = \epsilon u - \frac{s}{\tau_s} - \frac{f_{\text{in}}-1}{\tau_f} \\ \dot{f}_{\text{in}} = s \\ \dot{v} = \frac{1}{\tau_0}(f_{\text{in}} - v^{\frac{1}{\alpha}}) \\ \dot{q} = \frac{1}{\tau_0}(f_{\text{in}}\frac{1-(1-E_0)^{\frac{1}{f_{\text{in}}}}}{E_0} - v^{\frac{1}{\alpha}-1}q) \end{cases} \tag{3.2}$$

and one observed quantity (the BOLD signal)

$$\text{BOLD}_{\text{non-linear}} = V_0(k_1(1-q) + k_2(1-\frac{q}{v}) + k_3(1-v)). \tag{3.3}$$

In our work, we use a linear approximation to (3.2), (3.3). In this setting, the hemodynamic system is completely defined by the hemodynamic response function (HRF) $h(t)$ that is the response of the system to the ideal instantaneous excitation $\delta(t)$. We define the variables $\{x_1, x_2, x_3, x_4\} = \{s, 1-f_{\text{in}}, 1-v, 1-q\}$. Linearization of (3.2) around the resting point $\{x_1, x_2, x_3, x_4\} = (0, 0, 0, 0)$ gives

$$\begin{cases} \dot{x}_1 = \epsilon u - \frac{x_1}{\tau_s} + \frac{x_2}{\tau_f} \\ \dot{x}_2 = -x_1 \\ \dot{x}_3 = \frac{1}{\tau_0}(x_2 - \frac{x_3}{\alpha}) \\ \dot{x}_4 = c x_2 - \frac{1-\alpha}{\alpha\tau_0} x_3 - \frac{1}{\tau_0} x_4, \end{cases} \tag{3.4}$$

with $c = \frac{1+(1-E_0)\ln(1-E_0)/E_0}{\tau_0}$. We diagonalize the system by making use of the Gauss method:

$$\begin{cases} (\mathrm{D}^2 + \frac{1}{\tau_s}\mathrm{D} + \frac{1}{\tau_f}\mathrm{I})\{x_2\} = -\epsilon u \\ (\mathrm{D}^2 + \frac{1}{\tau_s}\mathrm{D} + \frac{1}{\tau_f}\mathrm{I})\{x_1\} = \epsilon \mathrm{D} u \\ (\mathrm{D} + \frac{1}{\alpha\tau_0}\mathrm{I})(\mathrm{D}^2 + \frac{1}{\tau_s}\mathrm{D} + \frac{1}{\tau_f}\mathrm{I})\{x_3\} = -\frac{\epsilon}{\tau_0} u \\ (\mathrm{D} + \frac{1}{\tau_0}\mathrm{I})(\mathrm{D} + \frac{1}{\alpha\tau_0}\mathrm{I})(\mathrm{D}^2 + \frac{1}{\tau_s}\mathrm{D} + \frac{1}{\tau_f}\mathrm{I})\{x_4\} = -c\epsilon\mathrm{D}u + (\frac{1-\alpha}{\alpha\tau_0^2} - \frac{c}{\alpha\tau_0})\epsilon u. \end{cases} \tag{3.5}$$

Additionally, the linearized equation for the BOLD signal is

$$\mathrm{BOLD}_{\text{linear}}(t) = V_0((k_1 + k_2)x_4(t) + (k_3 - k_2)x_3(t)).$$

Finally, the HRF $h(t)$ is obtained by setting $u(t) = \delta(t)$. It satisfies the differential equation

$$(\mathrm{D} + \frac{1}{\tau_0}\mathrm{I})(\mathrm{D} + \frac{\alpha}{\tau_0}\mathrm{I})(\mathrm{D}^2 + \frac{1}{\tau_s}\mathrm{D} + \frac{1}{\tau_f}\mathrm{I})\{h\}$$

$$= V_0(k_1 + k_2)(-c\epsilon\mathrm{D}u + (\frac{1-\alpha}{\alpha\tau_0^2} - \frac{c}{\alpha\tau_0})\epsilon u) - V_0(k_3 - k_2)(\frac{\epsilon}{\tau_0}\mathrm{D}u + \frac{\epsilon}{\tau_0^2}u)$$

The right-hand side can be further developed as

$$\frac{V_0\epsilon}{\tau_0}((-(k_1 + k_2)c\tau_0 - k_3 + k_2)\mathrm{D}u + ((k_1 + k_2)(\frac{1-\alpha}{\alpha\tau_0} - \frac{c}{\alpha}) - (k_3 - k_2)\frac{1}{\tau_0})u),$$

yielding the linear differential operator L of the form (2.3) with parameters

$$\begin{aligned} \vec{\alpha} &= \left(-\frac{1}{\tau_0}, -\frac{1}{\alpha\tau_0}, -\frac{1}{2\tau_s}(1 \pm j\sqrt{\frac{4\tau_s^2}{\tau_f} - 1})\right), \\ \vec{\gamma} &= \left(-\frac{(k_1 + k_2)(\frac{1-\alpha}{\alpha\tau_0} - \frac{c}{\alpha}) - (k_3 - k_2)\frac{1}{\tau_0}}{-(k_1 + k_2)c\tau_0 - k_3 + k_2}\right). \end{aligned} \tag{3.6}$$

There is an additional scaling factor $\frac{V_0\epsilon}{\tau_0}(-(k_1+k_2)c\tau_0-k_3+k_2)$.

We will refer to this model later on and eventually design wavelets that are matched to it for the improved detection of brain activity.

Chapter 4

The Signal Processing Problem in fMRI

Abstract — Once the fMRI signal is recorded, it is the task of the imaging software to detect the voxels that were active during a particular cognitive task. We split the detection problem in two steps — (1) extraction of the activity-related signal from the time-course for each voxel and (2) a statistical test on the activity-related signal. In the present work, we concentrate on step (1). The primary goal of this chapter is to state the activity-related signal extraction problem in mathematical terms. We consider two possible solutions:

- the classical linear solution that would be optimal to separate the activity-related signal from noise, if the system were driven by a gaussian random process;

- the non-linear sparse-solution search technique in the wavelet domain that is optimal when the system is driven by the sparse stimulus.

As our fMRI signal formation model corresponds to the second option, we use the non-linear solver, which will be described later on.

4.1 Introduction

FMRI and Time-Course Analysis

Functional Magnetic Resonance Imaging (fMRI) is being used increasingly in modern neuroscience. It allows non-invasive measurements of the evoked neural activity through neurovascular coupling and the blood-oxygenation-level-dependent (BOLD) effect, first observed by Ogawa et al. [39]. Few seconds after the stimulation, a decrease in deoxyhemoglobin (dHb) level in the implicated brain regions occurs, making the local T2* MR signal stronger. The typical spatio-temporal resolution of fast fMRI imaging techniques varies between 1–50 mm^3/voxel and 0.5–5s/volume.

In traditional fMRI, a human subject is scanned and asked to perform a task or exposed to stimuli that are relevant to some cognitive function [34]. There are two general types of experimental design. Within the block-based paradigm, each stimulus lasts for a non-negligible period of time (several scans) and is modelled as a boxcar function. In the event-based framework, a stimulus is short in time and can be mathematically represented with a Dirac delta-function. Event-related fMRI (efMRI) activations are weaker and more variable than block-based responses, which makes them harder to detect. In this chapter, we focus on the event-related paradigm.

Modeling the hemodynamic system has been (and still is) a subject of active research [42]. One popular framework to link neural activation and BOLD response makes use of the so-called balloon model by Buxton and Frank [43]. Essential components of their system include regional blood flow, vessel volume, dHb extraction fraction and the perfusion-inducing signal by the neuronal response to stimulus. Although the underlying differential equations are non-linear, it is common to assume that the hemodynamic system is linear and time-invariant. In this setting, a model for the task-related BOLD response can be obtained as a convolution between the stimulus pattern, describing the task, and the *hemodynamic response function* (HRF). This is the type of model that is used, implicitly or not, in all standard fMRI analysis packages.

The measured functional MR signal is degraded due to various sources: physiological factors such as respiratory and (aliased) cardiac components, subject's movements, scanning artifacts due to field inhomogeneity, and image reconstruction and post-processing. Therefore, traditional fMRI data analysis tries to find evidence for the presence of a hypothetical task-related BOLD response [34]. If such evidence is found (on statistical grounds), voxels are declared as "active". The most popular framework proposes a general linear model (GLM) that contains regressors of interest (e.g., task responses for various conditions) and other variates (baseline, low-frequency drifts, and so on). Given the noise statistics, the parameters (weights of each regressor) are then fitted to the data. The parameters' strength (or their statistical significance) is then evaluated by taking into account the residual error.

Next to confirmatory analysis, researchers have also proposed data-driven methods that do not (or only partially) rely on the prior knowledge. The most popular ones are subspace methods, such as principal components analysis (PCA) [44] and independent component analysis (ICA) [45,46]. These methods have the capability to reveal unmodelled trends in the data. However, manual intervention is often needed to distinguish noise-related from neurophysiological-relevant components. Semi-blind approaches use the knowledge of the stimulus timings. Glover et al. include a calibration trial from which they estimate the HRF [47]. This provides them with a Wiener filter that they apply to subsequent measurements to estimate the neural activity pattern through a deconvolution process. Makni et al. proposed a Bayesian framework for a joint HRF estimation-detection task [48].

In many event-related fMRI experiments, the activity-inducing events are well separated in time from each other. This means that the corresponding BOLD responses do not overlap[1] and that the deconvolved stimulus pattern is sparse in time. Here, we assume the sparsity of the stimulus and develop a new wavelet-based framework that is able to find the activity-related signal component in an fMRI time-course, without prior knowledge on the positions of the activity-inducing time onsets. While current fMRI experiments heavily rely on the knowledge of the timing of the events (typically, a stimulus or a recorded feedback), it should be noted that in some cases this information is imprecise or even unavailable. Possible reasons for this could be related to the type of subject (e.g., feedback in a decision-making task with small children) or because the "task" is implicit (e.g., interictal epileptic discharges in epileptic patients [49]). Our aim is to extract the activity-related signal component from time-courses, which can then be further analyzed depending on the neurological question at hand.

Wavelets and FMRI

Over the past decade, wavelets have become an essential tool in mathematics, engineering, and physics [9]. The wavelet function, when applied to the data,

---

[1] Such paradigms are called slow event-related designs.

behaves as a multi-resolution derivative operator. The order of the derivative operator is directly linked to the wavelet's number of vanishing moments. Therefore, singularities, such as edges, only have a local influence in the decomposition and are well approximated by a few (large) coefficients. This property has been the driving force behind many applications, such as coding [50] and denoising [51]. More recently, this feature is exploited in more general applications (e.g., reconstruction and deconvolution) by the optimizing a sparsity-inducing criterion on the wavelet coefficients, typically ℓ_1 regularization [52]. This principle has also been extended to more general (non-orthogonal) dictionaries such as (redundant) wavelet frames or combined transformations [53].

The wavelet transform has also been applied to fMRI time-series processing. In the spatial domain, the activity maps can be compactly represented by the wavelet decomposition; for an overview, see [6, 7]. In the temporal domain, the transform's decorrelating property can be used advantageously to fit the GLM's parameters in the presence of colored noise, which is the case in fMRI with fast repetition timing [54]. Also, the Hurst exponent—the characterizing parameter of a self-similar process—can be efficiently estimated in the wavelet domain and used to distinguish between healthy subjects and patients suffering from Alzheimer disease [55]. Finally, estimation with the use of penalized partial linear models and classical wavelets has been investigated by Fadili et al. in [56].

Our Contributions

While traditional wavelets, with their derivative-like behavior, offer good energy compaction for piecewise smooth signals, we believe that they are not well suited for the activation-related signal in fMRI. In this work, we introduce a new type of wavelets, named "activelets", driven by a characterization of the hemodynamic system. Basically, we (still) consider the hemodynamic system as linear and stationary and derive the differential operator L that links the hemodynamic response with the stimulus. To that end, we make use of a linearised set of differential equations that model the hemodynamic system [35]. Starting from there, we design the exponential-spline wavelets that essentially invert the system's response, therefore yielding a sparse representation. The essential feature that characterizes activelets from traditional wavelets is that they have a number of (well-chosen) exponential vanishing moments. Short activations then induce a cone of influence in the activelet decomposition of the BOLD signal, the same way as singularities do in the classical wavelet representation. Consequently, the use of a sparsity-pursuing optimization can help finding fMRI "activations", without fixing onset timings nor activity strengths. To that aim, we deploy (iterative) non-linear dictionary search algorithms. As we will see, another advantage of our method is that it improves robustness with respect to the shape of the HRF response; i.e., deviations from the modeled HRF system will not strongly impair the sparsifying properties of the differential operator L. Therefore, ℓ_1 minimization will still allow to recover activity-related components from these signals, a useful property for many event-related experiments.

4.2 FMRI Time-Course Modeling

In the event-related setting, we consider a Dirac impulse train $s(t)$ as the source of (neuronal) activity in the brain. Mathematically, we write the activity-inducing signal

$$s(t) = \sum_k c_k \delta(t - t_k), \tag{4.1}$$

where c_k and t_k denote the activity strengths and activity onsets, respectively; $\delta(t)$ denotes the Dirac impulse. We represent the hemodynamic system that links activity to BOLD signal as

$$x(t) = \mathrm{L}\left\{s\right\}(t), \tag{4.2}$$

where the operator L corresponds to the system of differential equations.

We assume the hemodynamic system to be linear and shift-invariant—an approximation that is common in literature and very reasonable when the events are sufficiently spaced in time. In that case, the activity-related BOLD signal is modeled as a weighted sum of shifted hemodynamic response functions

$$x(t) = \sum_k c_k h(t - t_k), \tag{4.3}$$

where $h(t)$ satisfies the differential equation

$$\mathrm{L}\{h\}(t) = \delta(t), \tag{4.4}$$

and the differential operator L is linear and time-invariant. The function $h(t)$ is a Green's function of L and the impulse response of the causal inverse operator L^{-1}.

In general, the operator L is characterized by its Fourier transform

$$\hat{\mathrm{L}}(\omega) = \frac{\prod_{n=1}^{N}(j\omega - \alpha_n)}{\prod_{m=1}^{M}(j\omega - \gamma_m)}. \tag{4.5}$$

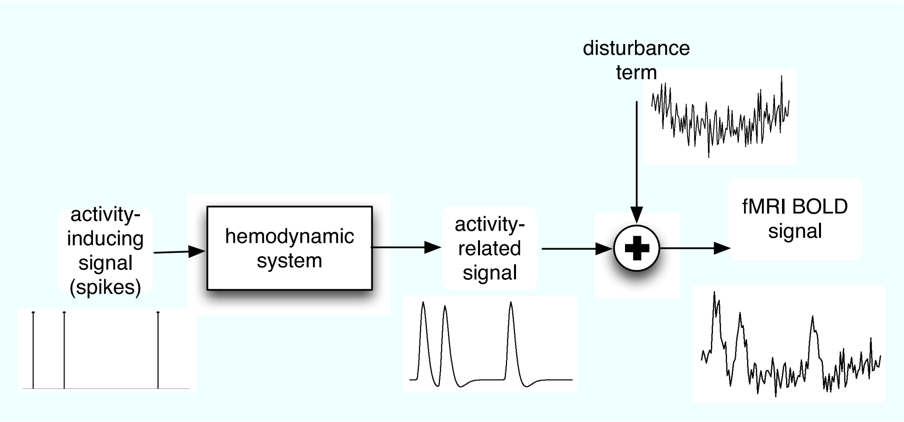


Figure 4.1: The hemodynamic system links neuronal activity to efMRI BOLD response, which is due to neurovascular coupling and a complex interplay between various physiological parameters.

Like in Chapter 2, we group the parameters in $\vec{\alpha} = (\alpha_1, \dots \alpha_N)$ and $\vec{\gamma} = (\gamma_1, \dots, \gamma_M)$.

We identify the operator L and the parameters $\vec{\alpha}, \vec{\gamma}$ by linearizing the balloon/windkessel model described in Chapter 3. This procedure is equivalent to considering the first-order Volterra series approximation [35].

4.3 Problem statement

The problem that we face is to estimate the presence of activity-related signal $x(t)$ in the noisy fMRI measurements:

$$y[k] = x(k) + n[k], \tag{4.6}$$

where $n[k]$ is the disturbance term that includes noise, baseline, drifts, aliased cardiac and respiratory contributions (cf. Fig. 4.1).

Our problem setting is different from a classical "linear model" analysis, which looks for the presence of a *fixed* regressor; i.e., a known linear combination of elementary signals. In our case, we just assume that the response will be of the form of (4.3), without fixing the weights or onset times.

4.4 The Linear Solution

4.4.1 L*L Smoothing Spline Solution

First, let us suppose that the (linear, shift-invariant) hemodynamic system is driven by Gaussian white random process with autocorrelation $\sigma_0\delta(t)$. We measure the noisy samples $y[k] = x(k) + n[k]$, where the fMRI noise $n[k]$ follows an $AR(\rho)$ model — a first-order auto-regressive process with correlation coefficient ρ and associated power spectrum $C_{nn}(\omega) = \sigma^2/|1-\rho e^{-j\omega}|^2$, $\rho = 0.2$. We would like to extract the activity-related component $x(t)$. The best linear estimator $\tilde{x}(t_0) = \sum_k b_{t_0}[k]y[k]$ of $x(t_0)$, given the samples $\{y[k]\}_k$, will then be the one that minimizes the mean-squared error

$$\min_{\tilde{x}(t_0)} E\left[|x(t_0) - \tilde{x}(t_0)|^2\right]. \tag{4.7}$$

We make use of the generalized smoothing spline theory to solve (4.7) [57]. All operators of the form (4.5) with $N > M$ are spline-admissible [58]; hence, so is the operator L associated to the hemodynamic system. Consequently, we can identify an associated exponential B-spline localization filter $\Delta_{\mathrm{L}}(z) = \prod_{k=1}^{N}(1 - e^{\alpha_k}z^{-1})$.

Given the noisy measurements $y[k]$ of the underlying continuous-time stationary process $x(t)$ with whitening operator L, we then know that the best linear MMSE estimator is the L*L smoothing spline, given by [57, Theorem 5]

$$\tilde{x}(t) = \sum_{k\in\mathbb{Z}} (b_{\lambda,\sigma} * y)[k]\varphi(t-k). \tag{4.8}$$

Here, $\varphi(t)$ is the L*L exponential B-spline with parameters $\{\vec{\alpha}, -\vec{\alpha}^*\}$, $\{\vec{\gamma}, -\vec{\gamma}^*\}$, and the digital filter $b_{\lambda,\sigma}$ is given by its z-transform

$$B_{\lambda,\sigma}(z) = \frac{1}{\sum_{k\in\mathbb{Z}}\varphi(k)z^{-k} + \frac{\sigma^2}{\mu_2\sigma_0|1-\rho z^{-1}|^2}\Delta_L(z)\Delta_L(z^{-1})}. \tag{4.9}$$

The smoothing spline solution is the best linear MMSE estimator. If the system input were indeed white Gaussian, it would also be the best possible estimator. This corresponds to an extension of the traditional Wiener filter theory for the case where the input is discrete and the output is continuously defined. If we sample (4.8 at the integers, we recover the classical discrete-time solution.

4.4.2 Random impulses

A more realistic input model for event-related fMRI is a sequence of random impulses. Suppose that the inter-event timings $t_k - t_{k-1}$ are independently and identically distributed (i.i.d.) random variables that follow an exponential distribution with rate parameter $\sigma_0 > 0$ (σ_0 controls the average number of events per unit time). Let the stimulus amplitudes c_k be i.i.d. random variables with known first and second order moments $E\{c_k\} = \mu_1$ and $E\{c_k^2\} = \mu_2$. We assume that c_k are independent of t_k.

Proposition 3 *The random Dirac impulse train process*

$$s(t) = \sum_k c_k \delta(t - t_k). \tag{4.10}$$

is wide-sense stationary with mean $\mu_1\sigma_0$ and autocorrelation function

$$c_{ss}(\tau) = E\{s(t)s(t+\tau)\} = \mu_1^2\sigma_0^2 + \mu_2\sigma_0 \cdot \delta(\tau). \tag{4.11}$$

Proof: Consider the Poisson-like process V_t that has increments c_k at times t_k. For its characteristic function $\chi_{V_t}(\omega) = E\{e^{-j\omega V_t}\}$, we have

$$\chi_{V_{t+\Delta t}}(\omega) = E\{e^{-j\omega V_{t+\Delta t}}\} = E\{e^{-j\omega(V_{t+\Delta t} - V_t + V_t)}\} =$$

$$E\{e^{-j\omega(V_{t+\Delta t} - V_t)}\}E\{e^{-j\omega V_t}\} \approx$$

$$\chi_{V_t}(\omega) \cdot \int_{-\infty}^{\infty} e^{-j\omega a}((1 - \sigma_0\Delta t)\delta(a) + \sigma_0\Delta t \cdot p(a))da =$$

$$\chi_{V_t}(\omega)((1 - \sigma_0\Delta t) + \sigma_0\Delta t\hat{p}(\omega)),$$

where $p(a)$ is the probability density function of the variables c_k. Therefore, $\chi_{V_t}(\omega)$ satisfies

$$\dot{\chi}_{V_t}(\omega) = \sigma_0 \cdot \chi_{V_t}(\omega)(\hat{p}(\omega) - 1),$$

and we find $\chi_{V_t}(\omega) = e^{\sigma_0 t(\hat{p}(\omega)-1)}$. The first and second-order moments of V_t are computed from the derivatives of $\chi_{V_t}(\omega)$ at $\omega = 0$; we have $E\{V_t\} = \sigma_0 t\mu_1$ and $E\{V_t^2\} = \sigma_0 t\mu_2 + (\sigma_0 t)^2\mu_1^2$. Consequently, the autocorrelation function of V_t satisfies

$$c_{V_t,V_t}(t,s) = E\{V_tV_s\} = E\{(V_t - V_s + V_s)V_s\} = E\{V_t - V_s\}E\{V_s\} + E\{V_s^2\} =$$

$$\sigma_0(t-s)\mu_1 \cdot \sigma_0 s\mu_1 + \sigma_0 s\mu_2 + (\sigma_0 s)^2\mu_1^2 = \sigma_0^2 ts\mu_1^2 + \sigma_0 s\mu_2$$

for $t > s$ and, by symmetry, $c_{V_t,V_t}(t,s) = \sigma_0^2 ts\mu_1^2 + \sigma_0 t\mu_2$, for $t < s$.

It is now sufficient to notice that the Dirac impulse train process (4.10) is the derivative of V_t. Its autocorrelation function (4.11) is obtained by taking the derivative of $c_{V_t,V_t}(t,s)$ with respect to t and s. ■

For the power spectrum of $s(t)$, we have $C_{ss}(\omega) = \mathcal{F}\left\{c_{ss}\right\}(\omega) = \mu_1^2\sigma_0^2\delta(\omega) + \mu_2\sigma_0$. By subtracting the mean from (4.10), we get a zero-mean, uncorrelated *innovation signal* $i(t) = s(t) - \mu_1\sigma_0$, with $c_{ii}(\tau) = \mu_2\sigma_0 \cdot \delta(\tau)$.

If the random Dirac impulse train process drives the hemodynamic system, we can rewrite the activity-related BOLD response as

$$\begin{aligned} x(t) &= (s * h)(t) && (4.12)\\ &= \sum_k h(t - t_k) && (4.13)\\ &= (i * h)(t) + \underbrace{\mu_1\sigma_0 \int_{-\infty}^{\infty} h(t)}_{h_0}. && (4.14) \end{aligned}$$

The key observation to make is that the re-centered process $x(t) - h_0$ is regular; i.e., it admits a whitening operator that turns it into an uncorrelated, zero-mean process. Indeed, using the operator L associated with the hemodynamic system, we have $\mathrm{L}\{i * h\}(t) = (i * \mathrm{L}\{h\})(t) = (i * \delta)(t) = i(t)$, which is uncorrelated and zero-mean. Equivalently, we can also say $\mathrm{L}^*\mathrm{L}\left\{c_{(x-h_0)(x-h_0)}\right\} = \mu_2\sigma_0 \cdot \delta(\tau)$.

We can define the re-centered measurements as $y_0[k] = y[k] - h_0$. The corresponding linear MMSE solution would coincide with the smoothing spline solution from the previous subsection. As $s(t)$ is not Gaussian, we can expect non-linear estimators to do better.

The linear solution can be considered as a gold standard for signal processing problems, and thus it is a worthful point of comparison. We implement it, using the causal-anticausal decomposition for the digital filters in (4.8), as described in [57, Appendix II]. Notice that the linear solution can be applied both for traditional B-splines (only zeros at the origin) or the splines associated to the activelet operator.

4.5 Wavelet Representations

In our model, the activity-inducing signal is a Dirac impulse train that passes through a system characterized by the operator L^{-1}. If we applied L to the ideal continuous BOLD signal, we would recover a sparse representation. In practice, given the noisy samples, we would like to have an efficient transform that compacts the activation energy on a few large coefficients.

The classical wavelet transform produces sparse representations of piecewise-smooth functions. Traditional wavelets act as Nth-order derivatives. Whenever the number N of vanishing moments is greater than the smoothness of the function, such a wavelet would, at each point, annihilate the first N terms of the Taylor approximation, essentially leaving significant coefficients only close to the discontinuity points.

In our case, the activaty-related signal $x(t)$ satisfies $\mathrm{L}\{x\}(t) = 0$ on the intervals $t_k < t < t_{k+1}$; i.e., the operator L annihilates $x(t)$ at all points except the activation points t_k. If we choose the parameters $\vec{\alpha}, \vec{\gamma}$ that correspond to the linearized hemodynamic system and construct the operator-like wavelet as described in Chapter 2, our wavelet will behave like L, leaving non-zero coefficients only around t_k. We are now able to estimate activations by identifying the most significant coefficients with the help of sparsity-pursuing search algorithms, which are described in the next chapter.

Chapter 5

Searching For Sparse Solution

Abstract — The activelet transform guarantees a sparse representation for the activity-related signal. The activelet basis can easily be extended to an activelet frame by performing the undecimated activelet transform (UDAT). The frame expansion allows for more flexibility and can lead to even sparser representations. Given the noisy data $y[k]$, $k = 1, \dots, T$, we would like to identify signal components that have a sparse representation in the activelet expansion. This is achieved by imposing an ℓ_1-norm penalty in the transform domain.

5.1 Sparse representations

Let Φ be the $T \times K$ dictionary matrix whose columns include the UDAT basis functions normalized with respect to their ℓ_2 norm. Suppose first that the data y consists only of the activity-related signal. We would like to identify the UDAT functions that provide a sparse representation for y. In the case of an exact fit (i.e., when the noise is absent), this boils down to solving the non-convex program

$$(P_0): \min_w \|w\|_0 \text{ subject to } y = \Phi w, \tag{5.1}$$

where the ℓ_0 quasi-norm of a vector is its number of non-zero components. The optimization problem (5.1) is NP-hard and demands exorbitant computational efforts. A sub-optimal solution can be found by solving the *basis pursuit* problem

$$(P_1) : \inf_{w} \|w\|_1 \text{ subject to } y = \Phi w. \tag{5.2}$$

This formulation can be cast into a classical linear program by introducing the vectors $w_+ \geq 0$ and $w_- \geq 0$ such that $w = w_+ - w_-$:

$$(P_1') : \inf_{w_+ \geq 0, w_- \geq 0} (\sum w_+ + \sum w_-) \text{ subject to } y = (\Phi, -\Phi) \begin{pmatrix} w_+ \\ w_- \end{pmatrix}. \tag{5.3}$$

5.1.1 The Simplex method

The simplex method is a classical solver for problems of type (5.3). It was introduced by Dantzig in the 1940-ies. The linear constraint, together with the variable positivity conditions, defines the solution domain, which is a polygon (simplex). The gradient of the objective functional is obviously non-zero everywhere, meaning that the solution can only be an extremal point (a vertex) of the polygon. Each vertex corresponds to a $T \times T$ submatrix of $(\Phi, -\Phi)$, and adjacent vertices' matrices differ in a single column. Using one $T \times T$ matrix inversion per step, the simplex method is able to determine which of the adjacent vertices defines a direction of criterion decrease, and jump to it. The matrix inversion is implemented as an update from iteration to iteration. Theoretically, the convergence speed is exponential, but in practice, the optimal vertex is usually found in polynomial time.

5.1.2 Interior point methods

The simplex method is prohibitively slow for large-scale problems. The more efficient interior point methods use the Newton solver for (5.3). The variable positivity constraints are eliminated by introducing the *barrier functions* in the optimization criterion. The barrier functions tend to infinity at the borders of the optimization domain and do not allow the Newton iterate to leave it. It can be proved that the solution $(w_+(\mu_b), w_-(\mu_b))$ to the parametrized problem

$$(P_1') : \inf_{w_+, w_- \geq 0} \left(\sum w_+ + \sum w_- - \mu_b \left(\sum_k \log w_{+,k} + \sum_k \log w_{-,k} \right) \right) \tag{5.4}$$

$$\text{subject to } y = (\Phi, -\Phi) \begin{pmatrix} w_+ \\ w_- \end{pmatrix}$$

tends to the solution of (5.3) as $\mu_b \to 0$. Nesterov et al. have shown that the logarithmic barrier function is optimal for the convergence [59]. The most efficient way to tackle (P_1') is to solve it together with its dual problem,

$$(P_{1,d}') : \sup_{z \geq 0} \left(w_d \cdot y + \mu_b \left(\sum_k \log z_k \right) \right) \text{ subject to } z = y - \begin{pmatrix} \Phi^* \\ -\Phi^* \end{pmatrix} w_d. \tag{5.5}$$

Here w_d is the dual variable and z is the *slack* variable, whose role is to complete the negative vector $\begin{pmatrix} \Phi^* \\ -\Phi^* \end{pmatrix} w_d - y$ to zero. By the duality theorem of Karush, Kuhn and Tucker (KKT), the necessary and sufficient condition for $\tilde{w}, \tilde{w_d}$ and $\tilde{z}$ to be the solutions to (5.4) and (5.5), respectively, is

$$\begin{cases} y = (\Phi, -\Phi) \begin{pmatrix} w_+ \\ w_- \end{pmatrix} \\ \tilde{z} = 1 - \begin{pmatrix} \Phi^* \\ -\Phi^* \end{pmatrix} \tilde{w_d} \\ \tilde{w_d}[k] \cdot \tilde{z}[k] = \mu_b, k = 1, \ldots, K. \end{cases} \tag{5.6}$$

The problem thus reduces to a square system of non-linear equations. In practice, instead of solving it for each μ_b, the interior point methods take an initial approximation that satisfies the constraints in (5.3) and perform a single Newton step for each μ_b from a decreasing sequence. Each Newton step boils down to the solution of a dense $K \times K$ system of linear equations. This is where lies most of the computational effort of the interior-point methods.

5.2 Sparse approximations

In case where the measurement vector y contains noise, imposing the constraint $y = \Phi w$ might be not the best choice. Instead, we might want to constrain the discrepancy $\|y - \Phi w\|_2$, while minimizing the sparsity norm. Consider the program

$$(P_{0,\epsilon}) : \min_w \|w\|_0 \text{ subject to } \|W(y - \Phi w)\|_2 \leq \epsilon \tag{5.7}$$

with $\epsilon > 0$ and W a problem-dependent linear weighting operator which may account for prior statistical knowledge. In this chapter, for notational simplicity, we take W to be identity; the case of general W is equivalent if we substitute $y' = W(y)$ and $\Phi' = W\Phi$.

Like in the previous section, this NP-hard problem is replaced by its convexified ℓ_1 form

$$(P_{1,\epsilon}): \min_{w} \|w\|_1 \text{ subject to } \|y - \Phi w\|_2 \leq \epsilon. \tag{5.8}$$

In many practical cases, the solution to $(P_{1,\epsilon})$ approximates well or even coincides with the solution w_0 to $(P_{0,\epsilon})$. It has been shown in [60, Theorem 3.1] that if the noiseless ideal signal x is sufficiently sparse with $\|w\|_0 \leq (\mu_\Phi^{-1} + 1)/4$[1], then w_0 is the unique maximally-sparse representation of x, and stable recovery is possible by minimizing (P_1, ϵ). Furthermore, if $\|w\|_0 < \mu_\Phi^{-1}/2$, and if (P_1, ϵ) is solved using ϵ somewhat larger than the noise level σ_n, then the minimizer has its support included in the support of w_0 [60, Theorem 6.1]. In other words, the solution to (P_1, ϵ) identifies only correct atoms that participate to the sparse signal x and never picks an incorrect atom.

We rewrite $(P_{1,\epsilon})$ in the augmented Lagrangian form

$$(Q_{\lambda,\ell_1}): \min_{w} \ \|y - \Phi w\|_2^2 + \lambda \|w\|_1 . \tag{5.9}$$

(Q_{λ,ℓ_1}) is a (perturbed) linear programming problem known as BPDN [61]. Interior-point methods are the fastest existing solvers for the generic problems of the form (5.9). However, better performance can be achieved if we have specific knowledge about the solution. With our problem setting, we expect the solution to be sparse in the basis Φ. In this case, the two methods presented in this section — Homotopy and Iterated Soft Thresholding — achieve better performance.

5.2.1 The Homotopy method

Unlike the interior point strategy, that optimizes (5.9) for the fixed value of λ, the homotopy method solves (Q_{λ,ℓ_1}) for *all* $\lambda \in [0, \|\Phi^T y\|_\infty]$ [62]. The method starts at $\lambda = \|\Phi y\|^T{}_\infty$ and $w = 0$ and follows the solution path as $\lambda \to 0$ [62]. The key observation is that the solution subset (called active set) is piecewise constant as a function of λ, changing only at critical values of λ; i.e., the solution path is polygonal. The Homotopy algorithm jumps from vertex to vertex of this path and modifies the active set appropriately, either adding or removing an element. Each step takes the computational cost of solving a system of

---

[1] μ_Φ is the mutual coherence of Φ, that is the maximal off-diagonal element in magnitude of its Gram matrix, $\sqrt{\frac{K-T}{T(K-1)}} \leq \mu_\Phi \leq 1$.

linear equations of the size of the active set. In case where the number of atoms in the approximate solution is much less than the number of data points, Homotopy shows a strong improvement in performance [63]. Another advantage of Homotopy is the explicit control of the residual error level; it is useful when the noise level σ_n is known or can be estimated from the data. The iterations are applied until the residual r^k satisfies $\left\|r^k\right\|_2 \leq \sigma_n$.

5.2.2 Iterative Soft Thresholding

If Φ is an orthogonal basis and W is identity, the solution to (Q_{λ,ℓ_1}) can be explicitly obtained by pointwise soft thresholding in the wavelet domain [64]. In the general case, Daubechies et al. suggest an iterative technique that introduces a series of auxiliary functionals that are minimized by soft thresholding [52]. The update iteration is given by:

$$w^{k+1} = \mathrm{ST}_{\mu_k \lambda}\left(w^k + \mu_k \Phi^T (y - \Phi w^k)\right)$$

where $\mathrm{ST}_{\mu_k \lambda}(\cdot)$ is the soft-thresholding operator with threshold $\mu_k \lambda$, and μ_k is a sequence of relaxation parameter satisfying $0 < \inf_k \mu_k \leq \sup_k \mu_k < 2/\left\|\Phi\right\|_2^2$. This iteration is very simple; it only involves fast implicit analysis and reconstruction operators associated to the dictionary Φ. It is also possible to take into account the structure of the basis Φ, yielding very fast algorithms [65]. However, the choice of the regularization parameter λ remains a delicate issue in practice. Interestingly, this approach has been generalized to more general penalty functions; for more details on proximal forward-backward splitting, see [66, 67].

5.3 Inexact heuristic techniques

Another alternative to solve (Q_{λ,ℓ_1}) is to fit sparse models heuristically. A classical strategy of this kind is the greedy stepwise least squares optimization often called Matching Pursuit or Orthogonal Matching Pursuit (OMP) [9] in the signal processing literature. Despite its heuristic and greedy nature, OMP can, in certain circumstances, succeed at identifying the sparsest solution [60]. However, OMP may fail to find the sparsest solution in certain scenarios where ℓ_1-minimization succeeds [61]. Moreover, OMP becomes rapidly burdensome for large-scale problems [68].

The LARS method is a greedy approximation of the homotopy algorithm that never reduces the active set [69]. It has been shown that if x is sufficiently sparse in Φ with incoherent Φ, then LARS selects only correct atoms that belong to the support of w_0 [63]. Moreover, the atoms that are sequentially selected enter the active set with the correct sign.

Replacing the soft thresholding operation with hard thresholding yields sparser representations. In fact, the hard thresholding operation is the solution to (Q_{λ,ℓ_0}) in an orthonormal basis. In our algorithms, we use hard thresholding to get better sparsity.

The LARS method computes the solution by considering one coordinate at a time as a candidate to enter the active set. Inspired by the notion of following the path, an accelerated algorithm (IT-LARS) was proposed by Fadili and Starck [70]. The IT-LARS follows the solution path approximately by identifying groups of atoms at each iteration using a stagewise iterative-thresholding (IT) variant of LARS, where the sequence $\{\lambda^k\}_{k\geq 0}$ is not data-adapted, but allowed to be strictly decreasing. The algorithm is as follows:

Algorithm 1 IT-LARS algorithm

```
$k = 0$, $r^0 = y$, $\lambda_k = \left\|\Phi^T y\right\|_\infty$.
While $\left\|r^k\right\|_2 > \sigma_n$
    • Residual: $r^k = y - \Phi w^k$,
    • Correlation: $c^k = \Phi^T r^k$,
    • Hard Thresholding: $I^k = \{i : |c^k[i]| > \lambda_k\}$,
    • Update direction: $d^k[I^k] = \Phi^+_{I^k} r^k$, and $d^k[\bar{I}^k] = 0$,
    • Update the solution: $w^{k+1} = w^k + \gamma d^k$, $0 < \gamma \leq 1$,
    • $\lambda_{k+1} = g(\lambda_k)$, $k = k + 1$.
Reconstruct $\hat{x}$ from $\hat{w}$.
```

The strictly decreasing function $g(\lambda)$ reflects the update schedule of λ^k, typically exponential or linear. The computation bottleneck of IT-LARS lies in calculating the least-squares projection step to get d^k. Following [68], this is implemented via a conjugate gradient (CG) solver, with each iteration involving multiplications by Φ or Φ^T. It can be shown that the computational complexity of IT-LARS is $O\left(S(l+2)V\right)$, where l is the number of CG iterations (10 were sufficient in our expriments), $S < \|w_0\|_0$ is the number of IT-LARS iterations and V is the computational complexity of an application of the linear operator

Φ^T or its adjoint.

5.3.1 Implementation

In our experiments, we used the IT-LARS and the acceleratied iterated soft thresholding methods which gave the best trade-off between computation time and approximation quality. The search routines involve numerous calls to the decomposition procedure, which can become a bottleneck in the efficiency. However, the pre-computation of the filters makes activelets as fast as traditional wavelets. During this preliminary step, it is also easy to normalize the rows of the decomposition matrix to have their l_2-norm equal to 1, as required by the IT-LARS.

The fast decomposition-reconstruction algorithm for the activelet basis is an adapted version of Mallat's filterbank method [8]. The filtering is performed in the FFT domain. Unlike the least-squares linear solutions, the sparsity algorithm makes use of the adjoint transform for reconstruction (cf. Fig. 5.1). This boils down to using the synthesis filters at the analysis side.

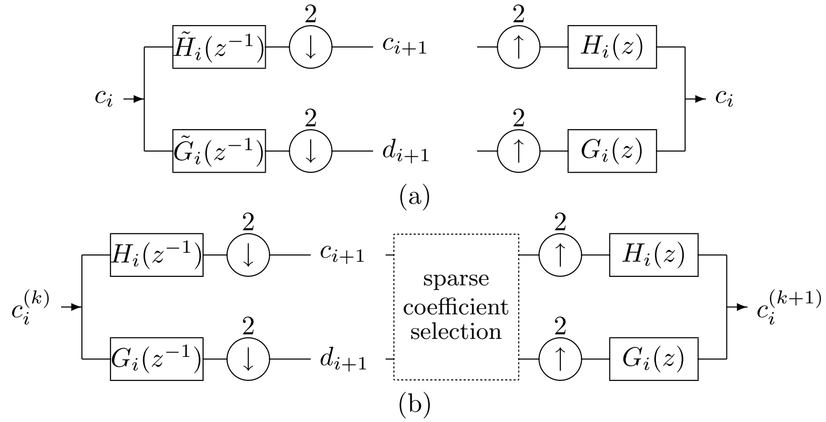


Figure 5.1: (a) Flterbank implementation of the activelet basis decomposition for a single decomposition level. The reconstruction is performed using the dual filters. (b) Filterbank implementation of the activelet decomposition combined with a sparsity iteration. The atoms identified by the sparsity algorithm are reconstructed using the adjoint operation. The (redundant) frame decomposition is done by Mallat's "à trous" algorithm; i.e., the signal is not subsampled at each iteration.

Chapter 6

Experimental Results

The activelets together with the sparse-solution search algorithm make up the core of our method. In this chapter, we test the technique's performance on synthetic and real fMRI datasets. For the synthetic data, we compare our algorithm against its linear alternatives by means of a ROC curve. Real-data experiments show that our algorithm is able to extract the activity-related component from the voxel's time-course.

6.1 Synthetic Data

We first evaluated the performance of our method on a synthetic fMRI data set consisting of 100 sequences generated according to the model (4.3). The stimulus consisted of five short randomly spaced events of length 0.8s; inter-event delays $t_{k+1} - t_k$ followed an exponential distribution with mean interstimulus delay of 40s. Each event was convolved with an HRF, for which the parameters of the underlying balloon model were randomly generated following gaussian distributions; the mean and standard deviation were taken from the experimentally measured histograms in [35], and are listed in the Table 3.1. Correlated noise following the model $AR(0.2)$, with $\sigma = 0.3$, leading to an average SNR of -7dB, and a baseline consisting of a constant and a slowly-varying sinusoid of random amplitude (frequency around 0.01Hz) were added to the data.

We compare the non-linear activelet method with two linear techniques: the first corresponds to the linear filtering signal estimation (MMSE solution) as described in section 4.4, while the second computes the "wavelet Wiener"

Table 6.1: SNR values (mean ± standard deviation, measured in dB) for different estimation methods.

	MMSE	Wavelet Wiener	Non-linear wavelet
Activelets	4.06±0.78	3.76±0.89	6.62±1.66
B-spline Wavelets	3.56±0.77	1.08±0.81	2.27±1.24

solution given by

$$\tilde{x}(t) = \arg\min_{x(t)} ||y[k] - x(k)||_2^2 + \sum_i \lambda_i ||T_i x||_2^2,$$

where $T_i x$ is the i-th scale of the wavelet transform coefficients and the scale-dependent regularisation coefficients λ_i are chosen by an Oracle:

$$\{\lambda_1, \ldots, \lambda_J\} = \arg \min_{\{\lambda_1,\ldots,\lambda_J\}} ||x - \tilde{x}||.$$

The linear filtering and wavelet Wiener techniques also use an Oracle to remove the baseline. Additionally we compare each method with its traditional B-spline counterpart.

In the non-linear activelet method, we propose to use prior information on the second-order statistics of the signal and the noise to select an appropriate weighting operator W in (5.7). The idea is to minimize the residual in the Wiener domain, where the SNR of the measurements is maximized. Specifically, we choose

$$\mathcal{F}(W)(\omega) = \frac{C_{hh}(\omega)}{C_{hh}(\omega) + C_{nn}(\omega)},$$

where $C_{hh}(\omega)$ and $C_{nn}(\omega)$ correspond to the power spectra of the HRF and the AR noise model, respectively.

The computations were done on a 2GHz Intel Core Duo MacBook Pro computer using Matlab 7 of MathWorks, WaveLab 8.02 [71] and SparseLab .100 [72]. In Fig. 6.1, we show a sample synthetic time-course. The results of the applied methods are shown in Fig. 6.2. The corresponding SNR levels are given in Table 6.1.

The best performance is obtained by the non-linear method, but only when the basis functions are well-tuned to the system response, as is the case with the activelets dictionary. Our method looses its advantage when the basis functions are not well matched to the stimulus. Among the linear methods, the MMSE

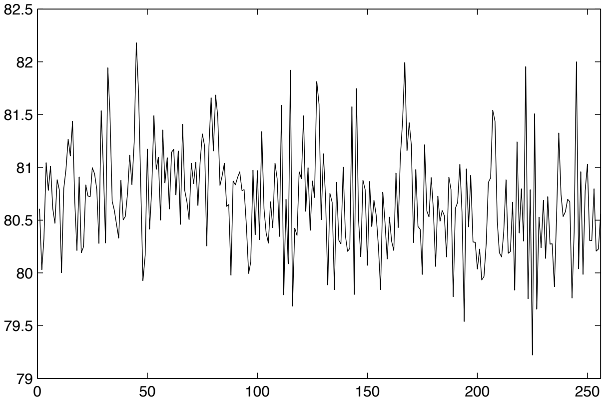


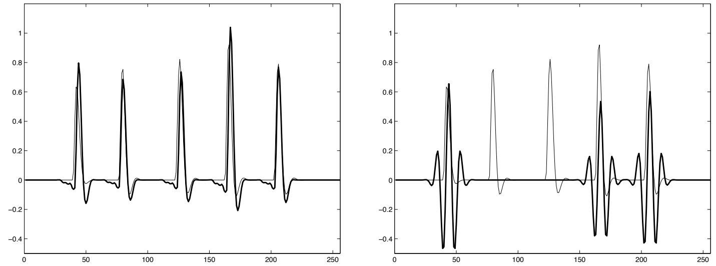
Figure 6.1: A sample synthetic time-course.

estimator performs the best when the model is well chosen. The performance of the Oracle-driven wavelet Wiener is not very adequate even when using the activelets, and gets even worse with standard wavelets.

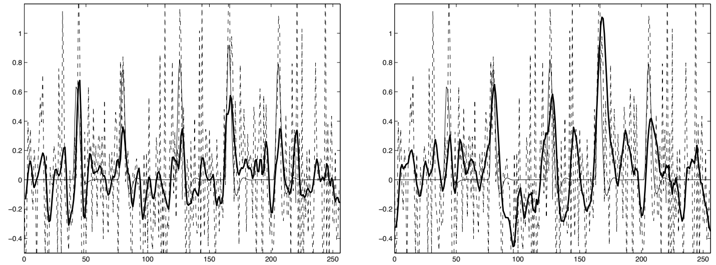
6.2 FMRI Experimental Data: Voxel Study

As a proof of concept, we have applied our method to several real voxels from a traditional fMRI experiment. The subject was scanned in a Siemens Magnetom 3T Scanner. The visual stimulation consisted of 10 flashing checkerboard excitations (duration=500 ms) with varying interstimulus timings, followed by a resting period, during which the subject closed the eyes. 256 scans were performed with TE=30ms, TR=1s and voxel size $2.6 \times 1.8 \times 5$mm. We used the Statistical Parameter Mapping (SPM) Matlab package to do standard preprocessing of the dataset. This operation included realignment, coregistration and gaussian smoothing (FWHM=8mm) of the data. We computed the SPM parameter map for the F-test of the effects-of-interest (see Fig. 6.3) and picked three time-courses that corresponded to a strongly active voxel, a weakly active voxel (around the 5% FWE threshold) and a non-active voxel.

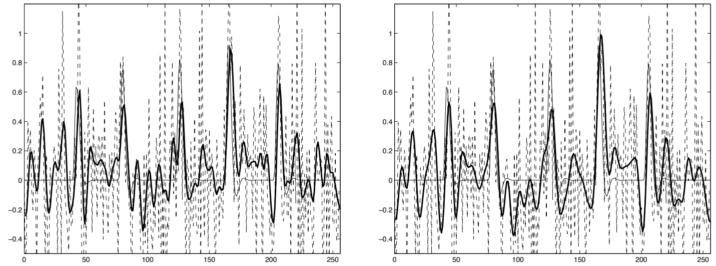
The results of the activelet analysis are presented in Figure 6.4. We show the voxel's time-course and the estimated activity-related signal. The time onset



(a) Activelets with non-linear optimisation (b) B-spline wavelets with non-linear optimisation



(c) Activelet Wiener solution (d) B-spline wavelet Wiener solution



(e) Activelet-spline MMSE solution (f) B-spline MMSE solution

Figure 6.2: Example of synthetic data: Activelets versus B-spline wavelets; non-linear wavelet, linear wavelet and linear MMSE solutions. Thin line: original signal, dotted line: noisy measurements, bold line: estimation.

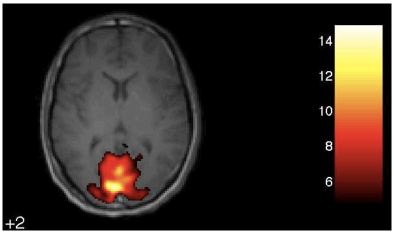


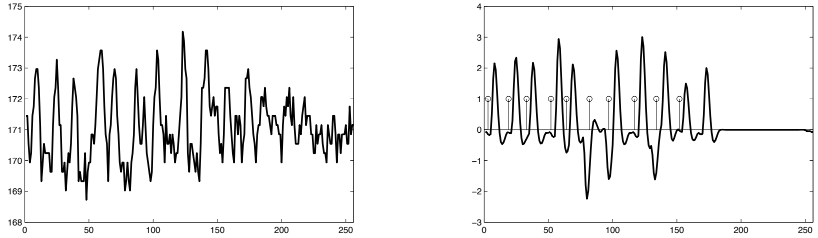
Figure 6.3: Activity map (threshold of 5%, corrected for multiple comparisons) in hot colormap superposed on the corresponding anatomical T_1 slice. A standard regression analysis with known onset times was performed using SPM2.

of each stimulus is indicated by a vertical line. Following an event, the BOLD response is expected to achieve its maximum in 2-3 seconds.

6.3 Discussion

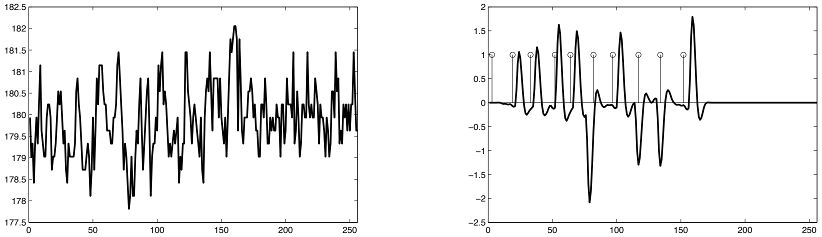
The problem of "blind" activity detection in fMRI is difficult due to the presence of strong disturbance components. Even in traditional fMRI data analysis, where the onset times are known, an important compromise on the flexibility of the model has to be made. As an extreme option, one could test the measurement for the presence of activity-related signal of the form (4.3) with fixed weights and onsets. Obviously, this approach is robust against false positive (FP) detections but loses all temporal resolution. At the same time, the variability of the HRF (which is known not only to vary over space and time for the same subject, but also between subjects) might lead to failures in detecting the activity that is actually present but different from the model; this situation is known as a false negative (FN) response. State-of-art fMRI analysis software (e.g., SPM toolbox for Matlab) employ additional regressors, such as derivatives of the model with respect to its parameters, to account for the BOLD variability [73]. The role of supplementary functions in the GLM boils down to trading some of the FP rate for the improved FN rate. Friman et al. use the "optimal" FN/FP trade-off as a criterion to choose appropriate regressors [74].

When the onset times are not known, the richness of the model becomes too high, and the linear search technique shows low specificity. Even in this case, we observe the superiority of the activelet basis compared to traditional



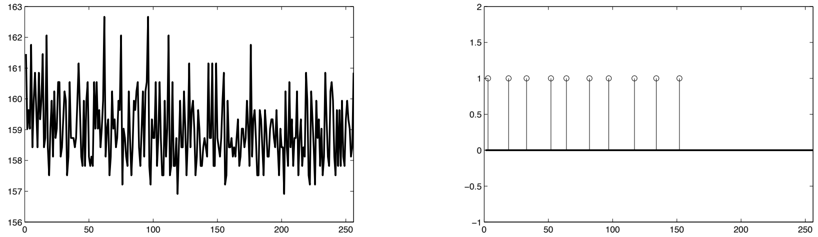
(a) Most active voxel: time-course

(b) Most active voxel: estimated activity-related signal



(c) Weakly active voxel: time-course

(d) Weakly active voxel: estimated activity-related signal



(e) Non-active voxel: time-course

(f) Non-active voxel: estimated activity-related signal

Figure 6.4: FMRI experimental data and the activity-related signal extracted by the activelet method. The vertical lines correspond to onsets times.

wavelets (cf. Table 6.1). At the same time, as we see in Fig. 6.2, there is no visual improvement in the denoised signal.

One key advantage of our approach is the possibility of bypassing the FN/FP trade-off by employing the non-linear search technique. Comparing Fig. 6.2(a) to Fig. 6.2(b), we note that, in this case, the use of the proper (activelet) basis becomes crucial; the B-spline wavelets cannot fit the signal well, and consequently show a very low SNR. Importantly, the non-linear activelet search reveals no activity-related signal in the areas that contain pure noise. This is not the case for the linear methods, meaning that they suffer from a high FP rate.

The improvement in the FN/FP trade-off is best seen on the ROC curve that is plotted in Fig. 6.5. Given the ground truth, we divide the signal into active and non-active intervals. The performance of the result is then measured by the mean-squared error (ϵ_k) for each interval k with respect to the ground truth. After normalization with the true mean signal μ during activation, we obtain the values $1 - \frac{\epsilon_k}{\mu}$, which can be related to the sensitivity (for active intervals) and the specificity (for non-active intervals) within the framework of a statistical decision taken for each interval.

We applied the sparse activelet and the MMSE algorithms for a wide range of regularization parameter λ. Large values of λ force the solution to zero, resulting in no sensitivity but high specificity. As λ decreases, the algorithm eventually fits the measurements. The two markers show the Oracle-driven wavelet Wiener results. All MMSE estimators use an Oracle to remove the baseline. Despite this non-neglible advantage, the results from the proposed activelets method with non-linear optimization are very satisfying. Moreover, the parameters of the HRFs varied randomly while the activelets operator L remained fixed.

The results for fMRI experimental data in Fig. 6.4 further demonstrate the adequacy of our approach. For the most active voxel in Figure 6.4(a)–6.4(b), 9 activations were detected out of 10; the undetected activation coincides with a sudden drop in the baseline BOLD signal. The additional detection (last peak of the thin line) happened right after the subject had closed the eyes. The activity-related signal in the weakly active voxel in Figure 6.4(c)–6.4(d) is almost entirely masked by noise; however, the algorithm still detects 6 activations. Finally, the detection in the non-active voxel is zero.

The usage of more sophisticated search techniques together with the highly redundant undecimated wavelet (activelet) transform leads to increased computation times as compared to linear methods. In our case, one time-course took up to 5 seconds of processing. However, the undecimated wavelet basis has a lot

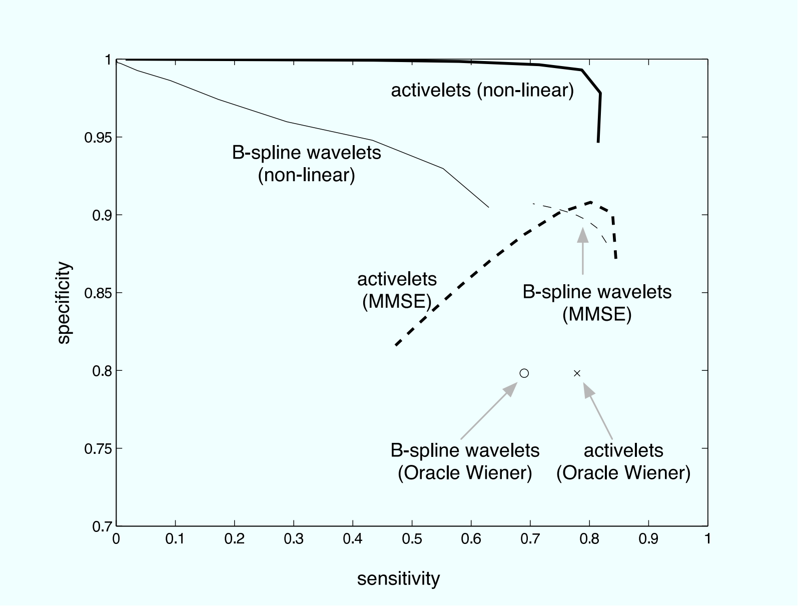


Figure 6.5: ROC curves for the different methods. The importance of both the activelets dictionary and the non-linear optimization are demonstrated; i.e., their combination leads to the best results that outperform the others.

of structure that could be used by the search algorithm. In the next chapter, we use an accelerated method and apply it to a full fMRI dataset.

Chapter 7

Activelets: A Feasibility Study in fMRI Retinotopy

Abstract — The results of the activelet approach on synthetic and individual real voxels are promising. Our next goals are (1) to validate our method on a whole-volume dataset; (2) to find the optimal event length such that the BOLD response is strong enough to be detected. In this chapter, we apply the methodology to a full high-resolution fMRI retinotopy dataset. We show that our algorithm is able to discriminate between active and non-active voxels. This is achieved by introducing a data-driven approach to determine the optimal regularization parameter that corresponds to the best sensitivity-specificity trade-off. We repeat the experiment for different values of the event length and identify the choice that gives optimal detection results.

This work has been done in collaboration with M. Costagli and K. Cheng from the RIKEN Brain Science Institute, Wako-shi, Saitama, Japan.

7.1 Introduction

The real-data experiments in Chapter 6 prove that our method works well on individual fMRI voxels. However, to investigate fully the sensitivity and the specificity of the algorithm in real data, we need to show that the algorithm is able to identify active voxels in the whole volume. The experiment in Chapter 6 is not good enough for this purpose: the activations that happen there are non-local, and it is difficult to characterize the performance of the algorithm on the full-volume scale. In this work, we choose a retinotopy task, which looks ideal to test the method: the size of active area is only a few voxels, and the activity is relatively strong. The experimental timing in fMRI retinotopy is known; therefore, activity maps generated by existing detection methods can be used as a point of comparison.

Apart from the "proof-of-concept" whole-volume activelet study, it is important to investigate the limits of applicability of the technique. In the activelet framework, the assumption is that the wavelet decomposition of the activity-related signal is sparse. In terms of paradigm programming, this means that the stimuli must be short enough so that the Dirac-impulse model is valid, and that the inter-event time must be long compared to the TR. Unfortunately, in real experiments, we can satisfy these conditions only to a limited extent. Indeed, the strength of the neural response (and hence, the SNR) decreases as the stimulus gets shorter. The SNR can be improved by allowing more phase-encoding steps, but this would lead to a longer sampling period and a poorly conditioned reconstruction problem. At the same time, the experiment duration is proportional to the mean inter-stimulus delay; therefore, the latter could not be arbitrarily long. Additionally, long inter-stimulus times may cause loss of attention from the subject.

We therefore chose to investigate the effect of the stimulus length $\Delta\tau_s$ on the performance of the activelet algorithm in fMRI retinotopy. We fix all other experimental parameters and vary $\Delta\tau_s$ within a range of values. The obtained activity maps are checked for consistency with their counterparts produced by SPM.

This chapter is organized as follows. Section 7.2 provides a brief introduction to the organization of the human visual system and to the fMRI retinotopy experiments. In Section 7.3, we give the necessary steps to adapt the algorithm to the whole-volume processing. We describe the experimental setup and show the results in Section 7.4. Finally, we discuss the results and suggest directions for further development of our method.

7.2 FMRI retinotopy

7.2.1 Human visual system

The transduced visual signal coming from the retina passes the lateral geniculate body and gets projected into the occipital lobe of the brain where the primary (striate) visual cortex is located (see Fig. 7.1). Neurons with similar functional

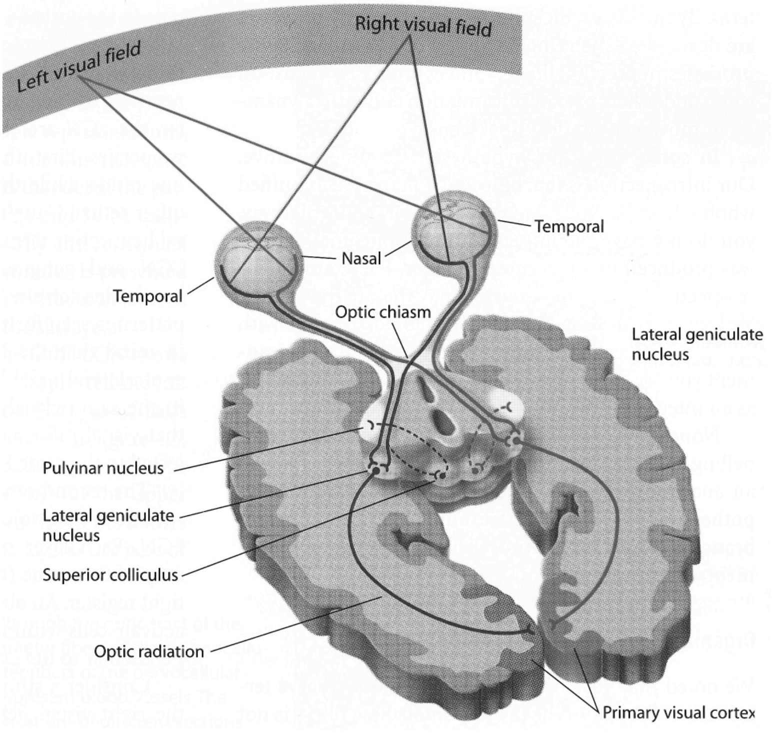


Figure 7.1: Human visual system (image courtesy of K. Cheng).

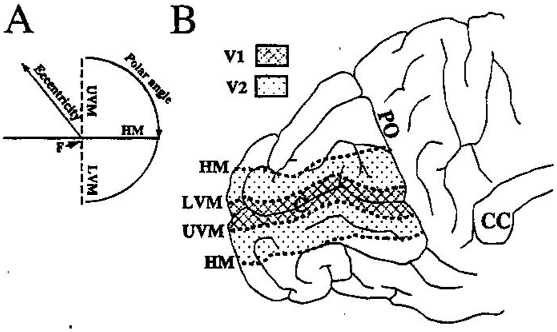


Figure 7.2: Human visual cortex (sagittal slice, one hemisphere is represented). Abbreviations: HM=horizontal meridian, LVM=lower vertical meridian, UVM=upper vertical meridian. Image taken from [77].

properties are grouped together in *columns*. The information corresponding to the same area of the field of view (FOV) from the left and the right eyes gets delivered into adjacently located ocular dominance columns (ODCs). Inouye [75] and Holmes [76] discovered that the mapping of the FOV into the primary visual cortex roughly corresponds to the radial coordinate system. This mapping flips the FOV; i.e., the left hemifield projects onto the right brain hemisphere and vice versa. The horizontal meridian of the FOV is mapped onto the calcarine sulcus. As the FOV radius moves up, the projection moves down, and vice versa. The primary visual cortex is located between the two vertical radii (meridians) with the calcarine sulcus in the middle. Similarly, higher visual areas (V2, V3 etc.) are demarcated by the horizontal and vertical meridians (see Fig. 7.2). Posterior voxels correspond to the fovea (center of the FOV), while anterior ones map onto the periphery.

The goal of retinotopy is to establish a mapping between the regions of FOV and the areas in the brain that are "responsible" for these regions.

7.2.2 Mapping ODCs with fMRI

Good spatial resolution and the non-invasiveness of fMRI have made it a dominant tool in human retinotopic studies. With high-field fMRI, it is presently possible to achieve voxel size as small as 0.5mm×0.5mm by 3-4mm. The ODCs are about 1mm wide, and their mapping was successfully done by Cheng et al. [78]. However, retinotopy experiments are close to the sensitivity limit of modern fMRI, so careful planning is required. The activation signal should be concentrated on as few voxels as possible; therefore, the slice should be taken such that the voxels are aligned with the ODCs, such as to maximize the SNR. In our experiments, the two eyes are stimulated equally, so that the two adjacent ODCs are always active at the same time; this generates clearer maps.

High-field experiments are prone to motion artifacts. We use a fixation apparatus for the head during the experiment; we also compensate for rigid head motion in a post-processing step.

7.3 Activelets for whole-volume studies

We apply the activelet method on the retinotopy dataset to extract the activity-related signal from the measurements. The two main practical challenges are:

1. Handling large number of intracranial voxels in the volume (around 31000 for the dataset in Section 6.2): fast techniques must be exploited and adapted for fMRI time-course processing.

2. Automatic adjustment of the regularization parameter for the sparse-solution search problem.

7.3.1 Data-dependent choice of regularization parameter

The synthetic experiments in Chapter 6 show that the ROC curve for our method gets closest to the ideal point where both sensitivity and specificity are equal to 1. However, to exploit this property, we need to estimate the optimal regularization parameter λ_{opt} that corresponds to the “best” point on the curve, on a voxel-per-voxel basis. Theoretical approaches like the MAD [51] do not give satisfactory results in practice. We suggest a data-driven approach that entirely relies on the information in the timecourse.

Looking at the ROC curve, we notice that it goes almost vertically up, then makes a turn at the optimal point and fades away on the left. Given our measures for FN and FP, this means that activations are progressively fitted as λ decreases until the critical moment when the algorithm starts fitting noise. If we have a sample of pure noise, we can adjust λ to catch this changing point. To this end, we modify the paradigm to include 60 samples of resting state at the end. Our assumption is that the resting state contains only noise and baseline. We perform a dichotomic search for the optimal regularization parameter that doesn't allow fitting any noise, while any smaller value does. For a more efficient implementation, we use only a few iterations during the search; we let the algorithm converge completely only for the final value of λ_{opt}.

It is convenient to add a sensitivity factor μ that multiplies λ_{opt} at the processing step. Tuning of μ allows us to run the algorithm in a more or less conservative mode; its role is qualitatively similar to the notion of confidence level in classical SPM.

7.3.2 Baseline extraction

The processing time of 5 seconds/voxel taken by IT/LARS procedure becomes prohibitive for the whole-volume studies. To circumvent this problem, we use a modification of the iterative thresholding method for wavelet bases that takes advantage of the multiscale structure of the undecimated activelet basis [65]. However, the baseline components cannot be included in the dictionary anymore.

One possible way to deal with this issue would be to estimate the baseline by projecting the measurements on a basis that consists of slowly-varying splines. However, linear baseline estimation techniques leave a lot of space for improvement: at the locations of the activity-related peaks, they tend to fit the baseline around the peak's half-height. Therefore, after baseline subtraction, the peaks are distorted, which results in poor sensitivity of the algorithm. In our procedure, we take a different approach: we extract the baseline from the residue at each iteration of the sparse-solution search routine. At almost no computational cost, this idea allows the algorithm to correct the detected baseline at the locations of the peaks.

7.3.3 Displaying the results

Having the activity-related component extracted from the measurements, it is important to provide a visual representation of the activity map, so that comparison can be done with the reference. A classical strategy would be to measure the l_2 energy of the activity-related signal. However, various factors that are not related to activity (e.g., motion artifacts) may cause strong signal changes that are indistinguishable from BOLD responses. This leads to high energy values for some of the non-active voxels. As an outlier-proof alternative, we use the peak-count measure on the activity-related signal as the activity indicator. Clearly, the reliability of this measure for activity detection grows with the number of events in the experiment.

7.4 The experiment

We studied the occipital cortex of a healthy male subject of 26 years of age. The experiment was conducted on a 4T Varian Unity Inova scanner. A high-resolution 3-D T1-weighted anatomical image was obtained as reference. Rigid head motion was restricted by requiring the subject to use a bite-bar. Heartbeat was monitored with a pulse oximeter, and respiration with a pressure sensor. Both signals were recorded and used at the post-processing step to correct for physiological fluctuations. Attention was controlled by asking the subject to press the button each time the fixation cross changed color. We have used a surface coil to acquire 4 slices of size 128×128 in the occipital cortex.

The stimulus consisted of a flickering horizontal 30°-wide wedge. 2 consecutive experiments were conducted with the stimulus length of 1.5 and 2 seconds. Each experiment contained 7 stimulations with random inter-event interval of 15–20s. A 60-second sample of resting state was acquired at the end.

The data was corrected for motion, heart and respiratory artifacts. From the post-processed data and the stimulus timing, a reference map was computed with the help of SPM (see Figure 7.3).

We applied the accelerated activelet method with data-driven regularization parameter estimation. We compensated for the sensitivity variation of the surface coil by normalizing each voxel to its mean value prior to processing. The maps of the voxel peak-count measure for $\Delta\tau_s = 1.5$s are shown in Figure 7.4 for different values of μ.

Comparing Figure 7.4 with Figure 7.3, we conclude that, for the event length of 1.5s, our algorithm achieves good detection results in the active areas, despite of it not making use of the knowledge on the experimental timing. The graphs of the detected signal in the most active voxel show that higher values of μ produce less outliers at the cost of sensitivity.

For the event length of 2s and the sensitivity level of $\mu = 1.5$, the algorithm is unable to detect the most active voxel, while there are spurious activations present in the image (Figure 7.5). In this case, the Dirac impulse is not a good model for the stimulus anymore.

7.5 Discussion

The results in Figure 7.4 confirm that the activelet method is able to identify active voxels in the retinotopy experiment. The most active voxel (as detected by SPM) is found even at high values of μ. Note that comparison with SPM

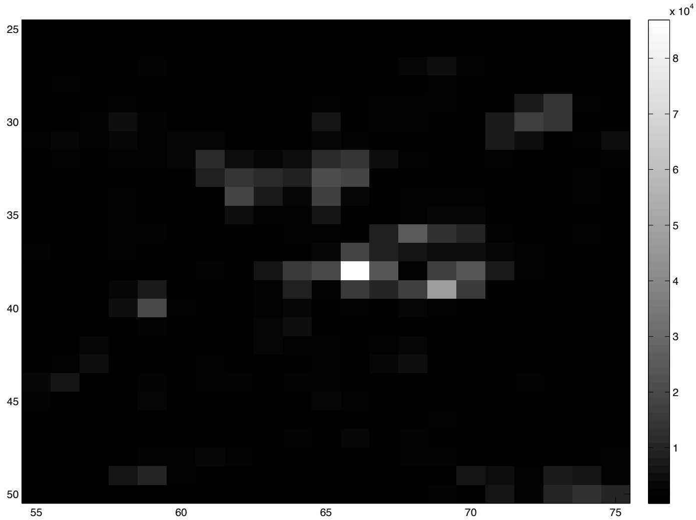
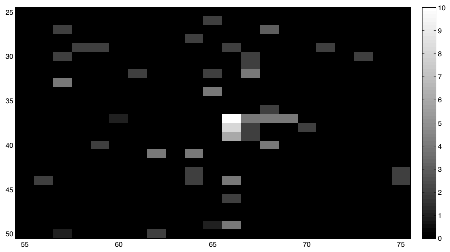
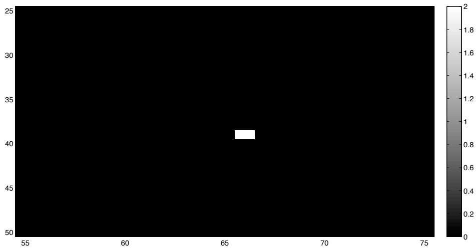


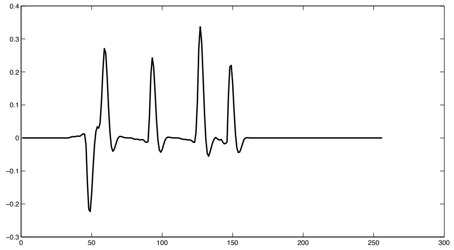
Figure 7.3: Activation map computed with SPM (slice 3, magnification in the area of the most active voxel).



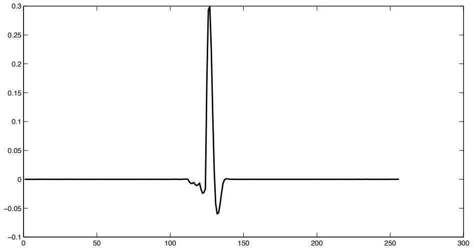
(a) $\mu = 1.75$



(b) $\mu = 2$



(c) Detection in the most active voxel, $\mu = 1.75$



(d) Detection in the most active voxel, $\mu = 2$

Figure 7.4: Peak-count measure for the $\Delta\tau_s = 1.5$s (slice 3, magnification in the area of the most active voxel, each peak increases the measure by 2).

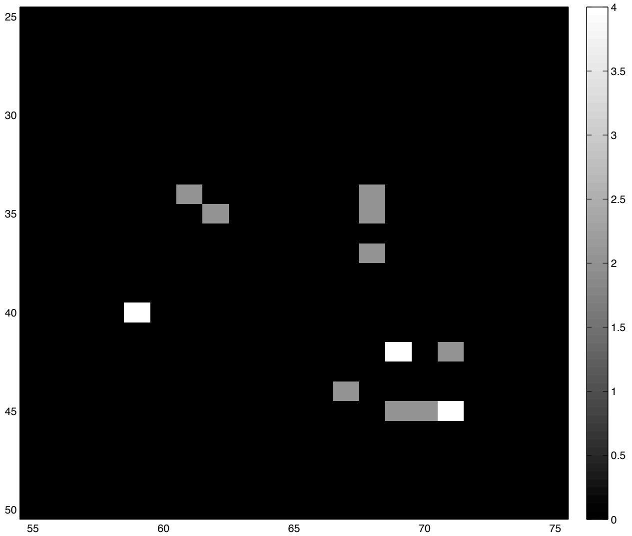


Figure 7.5: Peak-count measure for the $\Delta\tau_s = 2$s and $\mu = 1.5$ (slice 3, magnification in the area of the most active voxel, each peak increases the measure by 2).

should be done with precaution: SPM finds locations where activity of given shape took place at given moments, while our method finds locations where sparse impulses occurred at any time during the experiment.

Let us now discuss the sensitivity limits of the algorithm in reference to the ROC curve in Fig. 6.5. We can think of three relevant "barrier" values for λ:

- λ_{a1} corresponds to the moment when the algorithm starts fitting the activity-related signal. If λ is less than λ_{a1}, then the activity-related signal is partially, or completely, fitted;
- λ_{a2} represents the moment when *all* activations are fitted;
- finally, λ_{noise} is the barrier value such that whenever $\lambda < \lambda_{\text{noise}}$, noise is (at least partially) fitted by the algorithm.

It is clear that $\lambda_{a1} \geq \lambda_{a2}$. In our algorithm, we make the assumption that the noise strength is uniform along the voxel time-course; we take $\lambda_{\text{opt}} = \lambda_{\text{noise}}$. In a non-active voxel, $\lambda_{a1} = \lambda_{a2} = 0$. When the voxel displays activity, three situations are possible:

1. $\lambda_{a1} < \lambda_{\text{noise}}$. In this case, the activity is completely masked by the noise, and the algorithm fails to extract it.
2. $\lambda_{a1} \geq \lambda_{\text{noise}} \geq \lambda_{a2}$. Clearly, the closer λ_{noise} gets to the lower bound, the better detection quality is achievable.
3. $\lambda_{\text{noise}} < \lambda_{a2}$. This is the ideal situation; there exists a λ such that the activity-related signal is fully extracted from the data.

Our preliminary experiments showed that both $\Delta\tau_s = 1$s and $\Delta\tau_s = 3$s correspond to the case 1. With $\Delta\tau_s = 1$s, the energy injected in the system is too low, and detection is impossible due to poor SNR. For the length $\Delta\tau_s = 3$s, the representation of the signal in the activelet basis was not sufficiently sparse, and the activation energy was too spread between wavelet coefficients.

From the results in Fig. 7.4, we conclude that the case $\Delta\tau_s = 1.5$s falls into the second category. We see that we control the "conservativeness" of the algorithm by varying the parameter μ. For $\mu = 2$, only one activation in the most active voxel is detected, while with $\mu = 1.75$ a few false activations appear.

The event length of $\Delta\tau_s = 2$s brings us back to the unfavorable case 1: for $\mu = 1.5$, only spurious activations are present. We conclude that $\Delta\tau_s = 1.5$s is the optimal event length for the activelet method in retinotopy.

The SNR remains the key issue to successful detection. To improve SNR, traditional tools like SPM use spatial smoothing. For certain experiments this could lead to better detection. In retinotopy, the activations are so strongly localized that smoothing distorts the activity-related signal and lowers λ_{a1} more than it does with λ_{noise}. Nevertheless, it is sometimes possible to improve the SNR by taking into account prior information on timing. If the studied activity happens at more or less the same time every time the experiment is repeated, it becomes possible to average the measurements between the sessions, and therefore, achieve higher SNR.

Our algorithm is tuned to the fMRI event-related model, which explains its poor performance in the case of longer events. However, if the events are known to be long, it might be possible to improve the algorithm's performance by adjusting the filter W appropriately. Also, in our future work, we plan to validate the experiment on several subjects.

7.6 Conclusion

We have demonstrated the applicability of the activelet framework to a real full fMRI dataset. Using a retinotopy task as a testing scenario, we have demonstrated that the method is able to detect local activations. We suggested an accelerated version of the algorithm with a data-driven technique to estimate the regularization parameter. Additionally, we have investigated the influence of the event length on the detection performance; the optimal event duration was established for use in future experiments.

In our study, we intendedly concentrated on the signal processing aspects of the problem. An important future direction is setting the neurological framework that would take advantage of the activelet method in detecting activity without prior knowledge on timing.

Chapter 8

Operator-Like Wavelets

Abstract — We have seen that the usage of the problem-specific wavelet basis is crucial to good detection results. In order to extend the class of problems that can profit from wavelet-based techniques, we propose to build new families of wavelets that behave like an arbitrary differential operator. Our extension is general and includes many known wavelet bases. At the same time, the method takes advantage a fast filterbank decomposition-reconstruction algorithm. We give necessary and sufficient conditions for the scale-covariant differential operator to yield admissible basis functions, and we provide examples of new wavelets that can be obtained with our method.

8.1 Introduction

In signal analysis, we are typically interested in particular features, and often those can be detected by continuous operators. These operators can be considered as signal “decorrelators”. For example, edges are well detected by the derivative in 1-D or by the gradient or the Laplacian in 2-D. Specifically, the 1-D derivative turns a step edge into a Dirac delta-function, whose position indicates the location of the discontinuity.

While continuous operators are attractive conceptually, the calculus in the continuous domain is not directly accessible to us. To combine the advantages of the continuous modeling and of the fast discrete computations, signals can be represented with (discrete) coefficients in a (continuous) linear shift-invariant (LSI) space. In this approach, the signal, most often represented by its samples, is projected on the LSI space, and all the following treatment is done on its coefficients. The important question is then the choice of a convenient space that is adapted to the problem at hand.

In our case, we would like the discrete coefficients to correspond to the features of interest; i.e., we would like the projection operation to resemble the action of the continuous operator. For images, traditional wavelets are the perfect solution: acting like multi-scale derivatives, they essentially extract the edge information. At the same time, the wavelet decomposition is very efficient from the computational point of view due to the fast filtering algorithm.

Suppose that the signal of interest y satisfies the operator equation $\mathrm{L}\{y\} = x$, where x is the sparse information (features) of interest and the differential operator L is more general than a pure derivative. Our purpose in this chapter is to construct wavelets that behave like a multiscale version of L; in this case, the wavelet representation of y will be a multiscale representation of the information x.

We restrict ourselves to the class of spline-admissible, scale-covariant differential operators L. We give necessary and sufficient conditions that a spline-admissible operator must satisfy in order to be wavelet-admissible. Importantly, we construct the operator-like wavelet directly from the operator, bypassing the scaling function space. What makes the approach even more attractive is that the wavelet space is generated by the shifts of a single wavelet function. Our work provides a generalization of some known and used constructions including elliptic wavelets [79], polyharmonic spline wavelets [80], Wirtinger-Laplace operator-like wavelets [81] and exponential-spline wavelets [82].

This chapter is organized as follows. In Section 2, we formally define a spline-admissible scale-covariant operator and consider several examples. In Section 3, we construct the multiresolution analysis that corresponds to L. Then, in Section 4, we introduce the operator-like wavelets and study their properties; in particular, we prove that a wavelet-admissible operator L yields a stable wavelet basis at each scale. Finally, we provide examples of the new wavelets and conclude.

8.2 Preliminaries

Consider a linear, shift-invariant operator L that acts on the class of functions $f : \mathbb{R}^d \to \mathbb{C}$, where d is the number of dimensions. The operator L is characterized by its frequency response $\hat{L}(\boldsymbol{\omega})$.

Definition 1 L *is of order* r *if and only if, for all positive* $\rho < r - d/2$*, we have that*

$$\sum_{\mathbf{n}\in\mathbb{Z}^d} \frac{||\boldsymbol{\omega} + 2\pi\mathbf{n}||^{2\rho}}{1 + |\hat{L}(\boldsymbol{\omega} + 2\pi\mathbf{n})|^2} \le C_\rho < \infty.$$

Essentially, Definition 1 requires the frequency response $\hat{L}(\boldsymbol{\omega})$ of the operator to grow at least at the rate of $||\boldsymbol{\omega}||^r$ as $||\boldsymbol{\omega}|| \to \infty$. The associated Sobolev space is defined by

$$W_2^{\mathrm{L}} = \{f \in L_2(\mathbb{R}^d) : \int_{\mathbb{R}^d} |\hat{f}(\boldsymbol{\omega})|^2 (1 + |\hat{L}(\boldsymbol{\omega})|^2) d\boldsymbol{\omega} < \infty\}.$$

Definition 2 L *is spline-admissible of order* r *if and only if the following conditions are satisfied [83]:*

1. L *is a linear shift-invariant operator of order* $r > d/2$*;*
2. L *has a well-defined inverse* L^{-1}*, and the impulse response* $\rho(\mathbf{x})$ *of* L^{-1} *is a function of slow growth. Thus,* ρ *is a Green function of* L *satisfying* $\mathrm{L}\{\rho(\mathbf{x})\} = \delta(\mathbf{x})$*;*
3. *There exists a localization operator* $\hat{\Delta}(e^{j\boldsymbol{\omega}}) = \sum_{\mathbf{k}\in\mathbb{Z}^d} p[\mathbf{k}] e^{-j\langle\boldsymbol{\omega},\mathbf{k}\rangle}$ *with* $p \in \ell_1(\mathbb{Z}^d)$ *such that the corresponding generalized B-spline* $\varphi(\mathbf{x}) = \Delta\{\rho\}(\mathbf{x})$ *satisfies the Riesz basis condition*

$$0 < A \le \sum_{\mathbf{k}\in\mathbb{Z}^d} |\hat{\varphi}(\boldsymbol{\omega} + 2\pi\mathbf{k})|^2 \le B < \infty. \tag{8.1}$$

In particular, this requirement implies that the zeros of $\hat{\Delta}(e^{j\boldsymbol{\omega}})$ *are* $2\pi\mathbf{k}$*-periodized zeros of* $\hat{L}(\boldsymbol{\omega})$*. The condition (8.1) should be checked on a case-by-case basis for each* L *and* Δ*.*

Suppose now that we have a family of linear, shift-invariant operators $\{\mathrm{L}_{\vec{\nu}}\}$, indexed by the parameter vector $\vec{\nu} = (\nu_1, \ldots, \nu_N)$. Let $\mathbf{D}$ be an integer subsampling matrix that corresponds to a similarity transform (i.e., $\mathbf{D} = a\mathbf{R}$ with an orthogonal matrix $\mathbf{R}$ and $a > 0$).

Definition 3 *The operator* $\mathrm{L}_{\vec{\nu}}$ *is called scale-covariant with respect to* $\mathbf{D}$ *if there exists an* $N \times N$ *matrix* $\mathbf{A}$ *such that, for any* $\vec{\nu}$*,*

$$\hat{L}_{\mathbf{A}\vec{\nu}}(\mathbf{D}^T\boldsymbol{\omega}) = c(\mathbf{D}) \cdot \hat{L}_{\vec{\nu}}(\boldsymbol{\omega}),$$

where $c(\mathbf{D})$ *depends only on the choice of the matrix* $\mathbf{D}$. *The equivalent spatial domain condition is, for any smooth function* f,

$$\mathrm{L}_{\mathbf{A}\vec{\nu}}\{f\}(\mathbf{D}^{-1}\mathbf{x}) = c(\mathbf{D}) \cdot \mathrm{L}_{\vec{\nu}}\{f(\mathbf{D}^{-1}\mathbf{x})\}.$$

In other words, the family $\{\mathrm{L}_{\vec{\nu}}\}$ *is invariant with respect to scaling by* $\mathbf{D}$.

In the examples, we will use the unit step function $u(x) = (1 + \mathrm{sign}(x))/2$ and the rectangle function $\mathrm{rect}(x) = u(x + \frac{1}{2}) - u(x - \frac{1}{2})$.

Example 1 *Let* $d = 1$ *and* $\mathrm{L}_\nu = \mathrm{D} - \nu\mathrm{I}$, *where* $\nu < 0$ *is a scalar parameter,* D *is the derivative and* I *the identity operator. Let us verify that* L_ν *is spline-admissible and scale-covariant with respect to* $\mathbf{D} = (2)$. *Indeed,* L_ν *is of order* $r = 1 > \frac{1}{2}$, *and its inverse has an impulse response* $\rho_\nu(x) = e^{\nu x}u(x)$. *With the localization operator* $\hat{\Delta}_\nu(e^{j\omega}) = 1 - e^{j\omega-\nu}$, *we obtain the first-order exponential B-spline*

$$\varphi_\nu(x) = \rho_\nu(x) - e^\nu \rho_\nu(x-1),$$

which is a function of compact support (cf. Chapter 2).

Furthermore, $\hat{L}_\nu(\omega)$ *satisfies* $\hat{L}_{2\nu}(2\omega) = 2j\omega - 2\nu = 2\hat{L}_\nu(\omega)$, *which proves that* L_ν *is scale-covariant with* $\mathbf{A} = (2)$ *and* $c(\mathbf{D}) = 2$.

Example 2 *Let* $\hat{L}_\nu(\boldsymbol{\omega}) = ||\boldsymbol{\omega}||^2 + \nu^2$ *be the Matérn operator with the parameter* $\nu > 0$. *As* $\hat{L}_\nu(\boldsymbol{\omega})$ *does not have poles, no localization operator is needed for its Green's function. Therefore, the operator* L_ν *is spline-admissible and scale-covariant with respect to the quincunx subsampling matrix* $\mathbf{D} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ *with* $\mathbf{A} = (\sqrt{2})$.

The class of scale-covariant operators is large. All scale-invariant rotation-covariant operators [81] are scale-covariant, and therefore included. At the same time, the definition (3) allows us to apply our results to more general operator families, such as the composition of any number of Matérn operators. It also includes differential operators with transfer functions that are ratios of multidimensional polynomials.

For the remaining part of this chapter, we assume that $\{\mathrm{L}_{\vec{\nu}}\}$ is a family of scale-covariant, spline-admissible operators of order r.

8.3 Multiresolution Analysis

Let us fix a scale i. Consider the function $s_i(\mathbf{x})$ of smoothness order r that is characterized by the relation

$$s_i(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^d} c_{\mathbf{k}} \rho_{\vec{\nu}}(\mathbf{x} - \mathbf{D}^i\mathbf{k}).$$

We call this function an $\mathrm{L}_{\vec{\nu}}$-spline with knots $\mathbf{D}^i\mathbf{k}$. Clearly, $s_i(\mathbf{x})$ satisfies

$$\mathrm{L}_{\vec{\nu}}\{s_i\}(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^d} c_{\mathbf{k}} \delta(\mathbf{x} - \mathbf{D}^i\mathbf{k}).$$

Remember that the operator $\mathrm{L}_{\mathbf{A}^i\vec{\nu}}$ admits a generalized B-spline

$$\varphi_{\mathbf{A}^i\vec{\nu}}(\mathbf{x}) = \Delta_{\mathbf{A}^i\vec{\nu}}\{\rho_{\mathbf{A}^i\vec{\nu}}\}(\mathbf{x}).$$

We define $\varphi_i(\mathbf{x}) = \varphi_{\mathbf{A}^i\vec{\nu}}(\mathbf{D}^{-i}\mathbf{x})$; in the Fourier domain,

$$\hat{\varphi}_i(\boldsymbol{\omega}) = |\det(\mathbf{D})|^i \hat{\Delta}_{\mathbf{A}^i\vec{\nu}}(e^{j(\mathbf{D}^T)^i\boldsymbol{\omega}}) \hat{\rho}_{\mathbf{A}^i\vec{\nu}}((\mathbf{D}^T)^i\boldsymbol{\omega}) =$$

$$|\det(\mathbf{D})|^i c(\mathbf{D}) \hat{\Delta}_{\mathbf{A}^i\vec{\nu}}(e^{j(\mathbf{D}^T)^i\boldsymbol{\omega}}) \hat{\rho}_{\vec{\nu}}(\boldsymbol{\omega})$$

by scale covariance. Consequently, $\{\varphi_i(\mathbf{x} - \mathbf{D}^i\mathbf{k})\}_{\mathbf{k}\in\mathbb{Z}^d}$ is a Riesz basis. The multiresolution space V_i is defined as

$$V_i = \{s(\mathbf{x}) : s(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^d} c[\mathbf{k}]\varphi_i(\mathbf{x} - \mathbf{D}^i\mathbf{k}), c[\mathbf{k}] \in l_2(\mathbb{Z}^d)\}.$$

From here on, for notational simplicity, we omit the index $\vec{\nu}$.

By using the same argument as for the 1-D case [83], we can show that $\mathrm{L}^*\mathrm{L}$ is a spline-admissible operator of order $2r > d > d/2$. Its corresponding B-spline, which is given by $\varphi_i(\mathbf{x}) * \varphi_i^*(-\mathbf{x})$, generates a Riesz basis. Consequently, the $\mathrm{L}^*\mathrm{L}$-spline interpolant $\phi_i(\mathbf{x})$, given by

$$\phi_i(\mathbf{x}) \leftrightarrow \hat{\phi}_i(\boldsymbol{\omega}) = |\det(\mathbf{D})|^i \frac{|\hat{\varphi}_i(\boldsymbol{\omega})|^2}{\sum_{\mathbf{k}\in\mathbb{Z}^d} |\hat{\varphi}_i(\boldsymbol{\omega} + (\mathbf{D}^T)^{-i} \cdot 2\pi\mathbf{k})|^2},$$

is well-defined and generates a Riesz basis. Importantly, $\phi_i \in W_2^{\mathrm{L}}$ *does not depend on the specific choice of the localization operator* Δ_i, as we can see from

$$\hat{\phi}_i(\boldsymbol{\omega}) = |\det(\mathbf{D})|^i \frac{|\hat{\Delta}_i(e^{j(\mathbf{D}^T)^i\boldsymbol{\omega}})|^2/|\hat{L}(\boldsymbol{\omega})|^2}{|\hat{\Delta}_i(e^{j(\mathbf{D}^T)^i\boldsymbol{\omega}})|^2 \sum_{\mathbf{k}\in\mathbb{Z}^d} 1/|\hat{L}(\boldsymbol{\omega} + 2 \cdot \mathbf{D}^{-i}\pi\mathbf{k})|^2} =$$

$$\frac{|\det(\mathbf{D})|^i}{1 + |\hat{L}(\boldsymbol{\omega})|^2 \sum_{\mathbf{k}\in\mathbb{Z}^d\backslash\mathbf{0}} 1/|\hat{L}(\boldsymbol{\omega} + (\mathbf{D}^T)^{-i} \cdot 2\pi\mathbf{k})|^2}.$$

The L*L-spline interpolant plays the key role in our wavelet construction, which we describe in the next section.

8.4 Operator-Like Wavelets

We construct the generating wavelet function at scale i as

$$\psi_{i+1} = \mathrm{L}^*\{\phi_i\},$$

where we can apply the operator L^* because $\phi_i \in W_2^{\mathrm{L}}$. The Fourier domain expression for $\hat{\psi}_{i+1}$ has the form

$$\hat{\psi}_{i+1}(\boldsymbol{\omega}) = |\det(\mathbf{D})|^i \frac{\hat{L}^*(\boldsymbol{\omega})}{1 + |\hat{L}(\boldsymbol{\omega})|^2 \sum_{\mathbf{k}\in\mathbb{Z}^d\backslash\mathbf{0}} \frac{1}{|\hat{L}(\boldsymbol{\omega}+(\mathbf{D}^T)^{-i}\cdot 2\pi\mathbf{k})|^2}}. \tag{8.2}$$

Note that for every $\mathbf{p} \in \mathbb{R}^d$ such that $\hat{L}^*(\mathbf{p}) = 0$, $\hat{\psi}_{i+1}(\boldsymbol{\omega})$ vanishes at $\mathbf{p} + (\mathbf{D}^T)^{-i}2\pi\mathbf{k}$. In other words, each pole of L generates a periodic sequence of zeros in the spectrum of the generating wavelet.

The wavelet system is obtained by shifting ψ_{i+1} to all coset points $\mathbf{D}^i\mathbb{Z}^d\backslash\mathbf{D}^{i+1}\mathbb{Z}^d$ of the dilated grid. Remarkably, we have one unique wavelet function even when the dimension d is larger than one; the $|\det(\mathbf{D})|-1$ wavelets $\psi_{i+1}^{(1)}, \ldots, \psi_{i+1}^{(|\det(\mathbf{D})|-1)}$ are shifts of ψ_{i+1}. From here on, we will use the vectors $\mathbf{e}_l$, $l = 0, \ldots, |\det(\mathbf{D})|-1$ to select one of the $|\det(\mathbf{D})|$ cosets, with $\mathbf{e}_0 = \mathbf{0}$ corresponding to the dilated grid.

Let us study the properties of the new wavelet system.

Property 1 *The wavelet function ψ_{i+1} belongs to the approximation space V_i.*

Proof: From (8.2), we have

$$\hat{\psi}_{i+1}(\boldsymbol{\omega}) = |\det(\mathbf{D})|^i \frac{\hat{\Delta}_i^*(e^{j(\mathbf{D}^T)^i\boldsymbol{\omega}})\hat{\varphi}_i(\boldsymbol{\omega})}{|\hat{\Delta}_i(e^{j(\mathbf{D}^T)^i\boldsymbol{\omega}})|^2 \sum_{\mathbf{k}\in\mathbb{Z}^d} 1/|\hat{L}(\boldsymbol{\omega} + 2\pi\mathbf{D}^{-i}\mathbf{k})|^2}.$$

The expression in the denominator is equal to $\sum_{\mathbf{k}\in\mathbb{Z}^d} |\hat{\varphi}(\boldsymbol{\omega} + 2\pi\mathbf{D}^{-i}\mathbf{k})|^2$; it cannot vanish as φ_i generates a Riesz basis. Therefore, ψ_{i+1} is a stable linear combination of the shifts of φ_i. ■

Property 2 *The wavelets* $\{\psi_{i+1}(\mathbf{x}-\mathbf{D}^i\mathbf{k})\}_{k\in\mathbb{Z}^d\backslash\mathbf{D}\mathbb{Z}^d}$ *are orthogonal to the space* V_{i+1}.

Proof: As φ_{i+1} is a linear combination of the functions $\{\rho(\mathbf{x}-\mathbf{D}^{i+1}\mathbf{k})\}_{\mathbf{k}\in\mathbb{Z}^\mathbf{d}}$, it is sufficient to prove that $\langle\rho(\mathbf{x}),\psi_{i+1}(\mathbf{x}-\mathbf{D}^i\mathbf{k})\rangle = 0$ for all $\mathbf{k}\in\mathbb{Z}^d\backslash\mathbf{D}\mathbb{Z}^d$. We use a simple duality argument to transfer the conjugate operator to the left side of the scalar product to obtain

$$\langle\rho(\mathbf{x}),\psi_{i+1}(\mathbf{x}-\mathbf{D}^i\mathbf{k})\rangle = \langle \mathrm{L}\rho(\mathbf{x}+\mathbf{D}^i\mathbf{k}),\phi_i(\mathbf{x})\rangle = \langle\delta(\mathbf{x}+\mathbf{D}^i\mathbf{k}),\phi_i(\mathbf{x})\rangle =$$

$$\langle\delta(\mathbf{x}),\phi_i(\mathbf{x}-\mathbf{D}^i\mathbf{k})\rangle = \phi_i(-\mathbf{D}^i\mathbf{k}) = 0,$$

as ϕ_i is the interpolant and $\mathbf{k}\neq\mathbf{0}$. ■

We conclude that $\{\psi_{i+1}(x-\mathbf{D}^i\mathbf{k})\}_{k\in\mathbb{Z}^d\backslash\mathbf{D}\mathbb{Z}^d,i\in\mathbb{Z}}$ is a semi-orthogonal wavelet system.

A direct implication of our wavelet construction is the following property.

Property 3 *The wavelet function* ψ_{i+1} *behaves like a multiscale version of the underlying operator* L *in the sense that, for any* $f\in W_2^r$, *we have* $f*\psi_{i+1}^T = \mathrm{L}\{f*\phi_i^T\}$.

For a vast majority of operators, ϕ_i is a lowpass filter with its passband varying accordingly to the scale. Therefore, $\{\mathrm{L}\{f*\phi_i^T\}\}_{i\in\mathbb{Z}}$ corresponds to the multiscale representation of $\mathrm{L}\{f\}$.

The next result gives the condition on the operator that is necessary and sufficient for the corresponding operator-like wavelet system to be a stable Riesz basis. To formulate it, we introduce the set $\mathcal{N}=\{\mathbf{p}\in\mathbb{R}^d|\hat{\mathrm{L}}(\mathbf{p})=0\}$ of poles of L. For each scale i, we define

$$\mathcal{N}_i^{(l)} = \mathbf{p}+(\mathbf{D}^T)^{-(i+1)}2\pi\mathbf{e}_l+(\mathbf{D}^T)^{-i}2\pi\mathbb{Z}^d,\quad l=0,\dots,|\det(\mathbf{D})|-1.$$

Property 4 *Let* $i\in\mathbb{Z}$ *be an arbitrary scale. The set of functions* $\Psi=\{\varphi_{i+1}\}\cup\{\psi_{i+1}^{(l)}\}_{l=1}^{|\det(\mathbf{D})|-1}$ *generates a Riesz basis if and only if, for each* $1\le l\le|\det(\mathbf{D})|-1$,

$$\mathcal{N}_i^{(0)}\cap\mathcal{N}_i^{(l)}=\emptyset. \tag{8.3}$$

Proof: **Sufficiency.** Our proof is inspired by the ideas in [84]; we use the theory of finitely-generated shift-invariant spaces developed by de Boor et al [85]. Suppose that (8.3) holds. The set Ψ generates a Riesz basis if and only if, for each $\boldsymbol{\omega}$, the $|\det(\mathbf{D})|$ “fibers”

$$\hat{\mathcal{J}}_{\Psi} = \begin{pmatrix} \hat{\varphi}_{i+1}(\boldsymbol{\omega} + (\mathbf{D}^T)^{-(i+1)}2\pi\mathbf{k}) \\ \hat{\psi}^{(1)}_{i+1}(\boldsymbol{\omega} + (\mathbf{D}^T)^{-(i+1)}2\pi\mathbf{k}) \\ \vdots \\ \hat{\psi}^{(|\det(\mathbf{D})|-1)}_{i+1}(\boldsymbol{\omega} + (\mathbf{D}^T)^{-(i+1)}2\pi\mathbf{k}) \end{pmatrix}_{\mathbf{k}\in\mathbb{Z}^d} \tag{8.4}$$

are linearly independent. We know that the function φ_i generates a Riesz basis, so that the fibers $\hat{\mathcal{J}}_{\Phi}$ that correspond to $\varphi_i(\mathbf{x} - \mathbf{D}^i\mathbf{e}_l)$, $l = 0, \dots, |\det(\mathbf{D})| - 1$ are linearly independent. These fibers are given by

$$\hat{\mathcal{J}}_{\Phi} = \begin{pmatrix} \hat{\varphi}_i(\boldsymbol{\omega} + (\mathbf{D}^T)^{-(i+1)}2\pi\mathbf{k})) \\ \hat{\varphi}_i(\boldsymbol{\omega} + (\mathbf{D}^T)^{-(i+1)}2\pi\mathbf{k}))e^{-j\boldsymbol{\omega}^T\mathbf{D}^i\mathbf{e}_1 + j2\pi\mathbf{k}^T(\mathbf{D}^T)^{-1}\mathbf{e}_1} \\ \vdots \\ \hat{\varphi}_i(\boldsymbol{\omega} + (\mathbf{D}^T)^{-(i+1)}2\pi\mathbf{k}))e^{-j\boldsymbol{\omega}^T\mathbf{D}^i\mathbf{e}_{|\det(\mathbf{D})|-1} + j2\pi\mathbf{k}^T(\mathbf{D}^T)^{-1}\mathbf{e}_{|\det(\mathbf{D})|-1}} \end{pmatrix}_{\mathbf{k}\in\mathbb{Z}^d}. \tag{8.5}$$

Let us introduce the scaling filter

$$H(e^{j(\mathbf{D}^T)^i\boldsymbol{\omega}}) = \frac{\hat{\varphi}_{i+1}(\boldsymbol{\omega})}{\hat{\varphi}_i(\boldsymbol{\omega})} = const \cdot \frac{\hat{\Delta}_{i+1}(e^{j(\mathbf{D}^T)^{i+1}\boldsymbol{\omega}})}{\hat{\Delta}_i(e^{j(\mathbf{D}^T)^i\boldsymbol{\omega}})}$$

and the wavelet filters

$$G_n(e^{j(\mathbf{D}^T)^i\boldsymbol{\omega}}) = \frac{\hat{\psi}^{(n)}_{i+1}(\boldsymbol{\omega})}{\hat{\varphi}_i(\boldsymbol{\omega})} =$$

$$|\det(\mathbf{D})|^i e^{-j\boldsymbol{\omega}^T\mathbf{D}^i\mathbf{e}_n} \frac{\hat{\Delta}^*_i(e^{j(\mathbf{D}^T)^i\boldsymbol{\omega}})}{\sum_{\mathbf{k}\in\mathbb{Z}^d}|\hat{\varphi}_i(\boldsymbol{\omega} + 2\pi\mathbf{D}^{-i}\mathbf{k})|^2}, \quad n = 0, \dots, |\det(\mathbf{D})| - 1.$$

Any filter $P(e^{j(\mathbf{D}^T)^i\boldsymbol{\omega}})$ can be decomposed into polyphase components with respect to $\mathbf{D}$:

$$P(e^{j(\mathbf{D}^T)^i\boldsymbol{\omega}}) = \frac{1}{|\det(\mathbf{D})|} \sum_{n=0}^{|\det(\mathbf{D})|-1} \underbrace{\sum_{l=0}^{|\det(\mathbf{D})|-1} \mathbf{F}[l,n]P^{(l)}(e^{j(\mathbf{D}^T)^i\boldsymbol{\omega}})}_{n\text{-th polyphase component}},$$

where $P^{(l)}(e^{j(\mathbf{D}^T)^i\boldsymbol{\omega}}) = P(e^{j(\mathbf{D}^T)^i\boldsymbol{\omega} - j2\pi(\mathbf{D}^T)^{-1}\mathbf{e}_l})$ is the l-th aliased component and $\mathbf{F}$ is the generalized DFT matrix associated with $\mathbf{D}$ [86]. With these notations, we can express (8.4) through (8.5); the polyphase components get multiplied by the corresponding shifts of the scaling function:

$$\hat{\mathcal{J}}_\Psi = \frac{1}{|\det(\mathbf{D})|} \underbrace{\begin{pmatrix} H^{(0)} & \dots & H^{(|\det(\mathbf{D})|-1)} \\ G_1^{(0)} & \dots & G_1^{(|\det(\mathbf{D})|-1)} \\ & \ddots & \\ G_{|\det(\mathbf{D})|-1}^{(0)} & \dots & G_{|\det(\mathbf{D})|-1}^{(|\det(\mathbf{D})|-1)} \end{pmatrix}}_{\mathbf{T}} \mathbf{F} \hat{\mathcal{J}}_\Phi.$$

We need to show that $\det(\mathbf{T})$ is separated from zero. Under the conditions on the order of L, all entries of $\mathbf{T}$ are continuous, and it is sufficient to prove $\det(\mathbf{T}) \neq 0$. Suppose that for some $\boldsymbol{\omega}$ the determinant $\det T$ is null. Then, there exist $(c_l) \neq \mathbf{0}$ such that

$$\sum_{l=0}^{|\det(\mathbf{D})|-1} c_l G_0^{(l)} \mathbf{F}[l,n]^* = 0, \quad n = 1, \dots, |\det(\mathbf{D})| - 1 \tag{8.6}$$

and

$$\sum_{l=0}^{|\det(\mathbf{D})|-1} c_l H^{(l)} = 0. \tag{8.7}$$

The condition (8.6) means that the vector $[c_l G_0^{(l)}]_{l=0}^{|\det(\mathbf{D})|-1}$ is orthogonal to all columns of the $|\det(\mathbf{D})| \times |\det(\mathbf{D})|$ matrix $\mathbf{F}$ except for the vector $(1, \dots, 1)$ that corresponds to $k = 0$. As $\mathbf{F}$ is orthogonal, we conclude that

$$c_l G_0^{(l)} = c = const.$$

Formally, there are three cases to consider:

1. All $G_0^{(l)} \neq 0$ and $c_l = c/G_0^{(l)}$. Then, $\hat{\Delta}_i(e^{j(\mathbf{D}^T)^i \boldsymbol{\omega} - j2\pi(\mathbf{D}^T)^{-1}\mathbf{e}_l}) \neq 0$. From (8.7), we have

$$\sum_{l=0}^{|\det(\mathbf{D})|-1} \frac{H^{(l)}}{G_0^{(l)}}$$

$$= \hat{\Delta}_{i+1}(e^{j(\mathbf{D}^T)^{i+1}\boldsymbol{\omega}}) \sum_{l=0}^{|\det(\mathbf{D})|-1} \frac{const}{G_0^{(l)} \hat{\Delta}_i(e^{j(\mathbf{D}^T)^i \boldsymbol{\omega} - j2\pi(\mathbf{D}^T)^{-1}\mathbf{e}_l})} = 0.$$

This can only be true if $\hat{\Delta}_{i+1}(e^{j(\mathbf{D}^T)^{i+1}\boldsymbol{\omega}}) = 0$, which implies $H^{(l)} = 0$ for $l = 0, \dots, |\det(\mathbf{D})| - 1$ and $\hat{\varphi}_{i+1}(\boldsymbol{\omega} + j2\pi(\mathbf{D}^T)^{-i-1}\mathbf{k}) = 0$ for all $\mathbf{k} \in \mathbb{Z}^d$. The latter contradicts the fact that $\{\varphi_{i+1}(\mathbf{x} + \mathbf{D}^{i+1}\mathbf{k})\}_{\mathbf{k}\in\mathbb{Z}^d}$ is a Riesz basis.

2. There is a unique l_0 such that $G_0^{(l_0)} = 0$. In this case, $c_l = 0$ for $l \neq l_0$, and from (8.7) we have $H^{(l_0)} = 0$. This means that $\hat{\Delta}_{i+1}(e^{j(\mathbf{D}^T)^{i+1}\boldsymbol{\omega}}) = 0$, and, as $\Delta_i^{(l)} \neq 0$ for $l \neq l_0$, we have $H^{(l)} = 0$ for $l = 0, \dots, |\det(\mathbf{D})| - 1$, which is again impossible.

3. There exist $l_1 \neq l_2$ such that $G_0^{(l_1)} = G_0^{(l_2)} = 0$. Let $\boldsymbol{\omega}_0 = \boldsymbol{\omega} - 2\pi(\mathbf{D}^T)^{-i-1}\mathbf{e}_{l_1}$, and let $\mathbf{e}_l$ be the defining vector for the coset containing $\mathbf{e}_{l_2} - \mathbf{e}_{l_1}$. Clearly, $\hat{\psi}_{i+1}(\boldsymbol{\omega}_0) = 0$ and $\hat{\psi}_{i+1}(\boldsymbol{\omega}_0 - 2\pi(\mathbf{D}^T)^{-i-1}\mathbf{e}_l) = 0$. From the expression (8.2) for $\hat{\psi}_{i+1}(\boldsymbol{\omega})$, we conclude that $\boldsymbol{\omega}_0 \in N_i^{(0)} \cap \mathcal{N}_i^{(l)}$, and (8.3) is violated.

Necessity. Suppose that L violates the necessary condition; i.e., there exist $\mathbf{p}_1, \mathbf{p}_2 \in \mathbb{R}^d$ such that $\hat{L}(\mathbf{p}_1) = \hat{L}(\mathbf{p}_2) = 0$ and $\mathbf{p}_1 = \mathbf{p}_2 + (\mathbf{D}^T)^{-i}2\pi\mathbf{k} + (\mathbf{D}^T)^{-(i+1)}2\pi\mathbf{e}_l$ for some coset $\mathbf{e}_l$ and $k \in \mathbb{Z}^d$. Note that the wavelet fibers in $\hat{\mathcal{J}}_\Psi$ have a particularly simple structure: they are $|\det(\mathbf{D})|$-periodic sequences, multiplied by $\hat{\psi}_{i+1}(\boldsymbol{\omega})$. We have $\hat{\psi}_{i+1}(\mathbf{p}_1) = \hat{\psi}_{i+1}(\mathbf{p}_1 + (\mathbf{D}^T)^{-(i+1)}2\pi\mathbf{e}_l) = 0$. For $\boldsymbol{\omega} = \mathbf{p_1}$, the $|\det(\mathbf{D})| - 1$ fibers will have two zeros on the main period of the periodic sequence at the positions corresponding to $\mathbf{e}_0$ and $\mathbf{e}_l$. The number of fibers associated to wavelets is $|\det(\mathbf{D})| - 1$, and there are at most $|\det(\mathbf{D})| - 2$ non-zeros at each period. Necessarily, the fibers are linearly dependent. ■

Let us study several examples that illustrate the richness of our wavelet construction.

Example 3 *For the first-order differential operator* $\mathrm{L}_\nu = \mathrm{D} - \nu\mathrm{I}$ *from Example 1, the autocorrelation filter is identity. Let* $\mathbf{D} = (m)$, $m > 0$. *We have* $\sum_{\mathbf{k}\in\mathbb{Z}^d} |\hat{\varphi}_i(\omega + 2 \cdot \mathbf{D}^{-i}\pi\mathbf{k})|^2 = 1$ *and the* $\mathrm{L}^*\mathrm{L}$*-interpolant* $\phi_i(x)$ *is a symmetric exponential spline with poles* $\{\nu, -\nu^*\}$. *As* $d = 1$ *and* $m = 2$, *the wavelet space is generated by a single shift of the wavelet function, which is the exponential-spline wavelet (see Figure 8.1).*

Example 4 *With* $m = 3$ *and the derivative operator* $\mathrm{L} = \mathrm{D}$, *we get the wavelet system, generated by two Haar wavelets, shifted to the points* $x = 1$ *and* $x = 2$. *Interestingly, despite the apparent simplicity of this operator, the separable extension to multiple dimensions is not possible — for example, the 2-D operator* $\mathrm{D}_x\mathrm{D}_y$ *fails the necessary stability condition.*

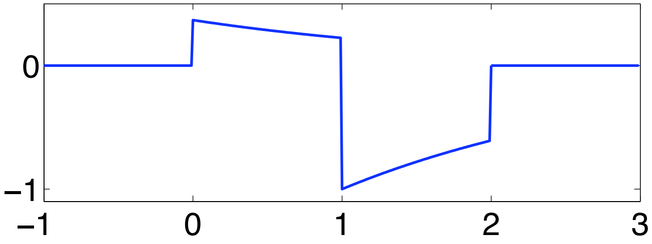


Figure 8.1: First order exponential-spline wavelet corresponds to the choice of $\mathrm{L}_{\nu} = \mathrm{D} - \nu \mathrm{I}$.

Example 5 *We show the wavelet that corresponds to the 2-D Matérn operator* $\hat{L}_{\nu}(\boldsymbol{\omega}) = ||\boldsymbol{\omega}||^2 + \nu^2$ *in Figure 8.2. The function has sharp peaks, which are caused by slow decay of the wavelet spectrum in the Fourier domain.*

Example 6 *As* $\nu \to 0$*, the Matérn operator tends to the scale-invariant Laplacian operator* $\hat{\Delta}(\boldsymbol{\omega}) = ||\boldsymbol{\omega}||^2$*. The corresponding wavelet is shown in Fig. 8.3. In case of iterated Laplacian operator* $\hat{\Delta}^2(\boldsymbol{\omega}) = ||\boldsymbol{\omega}||^4$*, faster Fourier decay of the wavelet spectrum leads to better regularity in space domain (see Fig. 8.4).*

Note that both the Matérn and the Laplace operator-like wavelets can be considered with any admissible subsampling matrix $\mathbf{D}$. In two dimensions, the choice of $\mathbf{D} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$ corresponds to the rectangular grid, while $\mathbf{D} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ characterizes the quincunx subsampling scheme.

8.5 Conclusion

We have constructed wavelet-like bases that behave like the multiresolution version of a given scale-covariant, spline-admissible operator. In our construction, all wavelets are shifts of a single generating function. In the multidimensional setting, the wavelets are not separable. In general, separable wavelets can not be obtained with this non-separable construction.

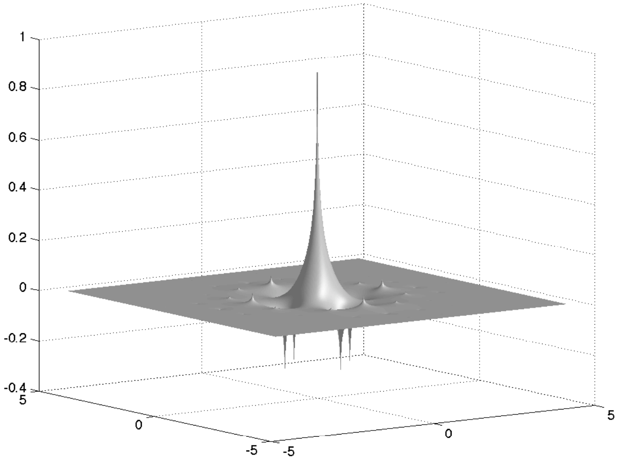


Figure 8.2: Wavelet obtained from the Matérn operator; $\nu = 1$.

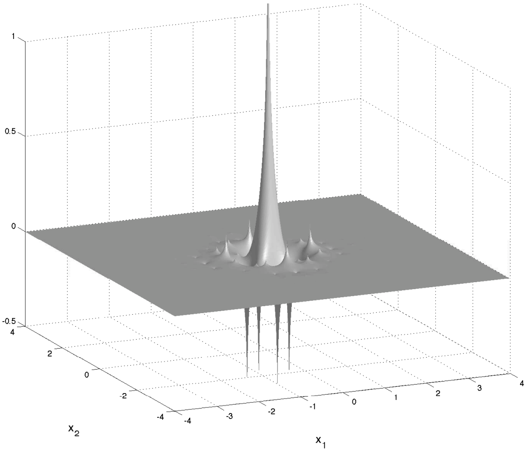


Figure 8.3: Wavelet obtained from the Laplacian operator.

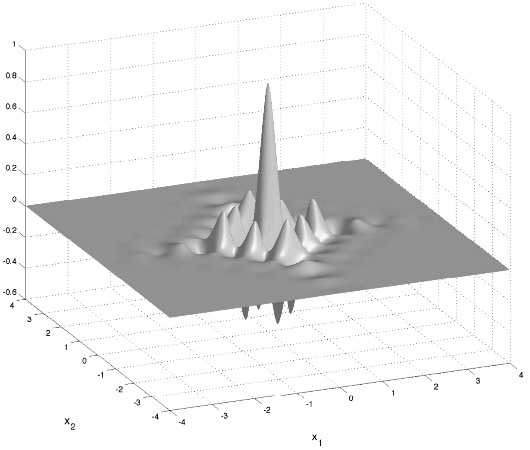


Figure 8.4: Wavelet obtained from the iterated Laplacian operator.

From the computational point of view, our wavelets still admit fast Mallat's filterbank algorithm, although, the filters are generally scale-dependent. Whenever the transform is used intensively, it might be a reasonable option to precompute these filters and store them in memory to accelerate computations.

The completeness of the multiresolution analysis $\cup V_i$ in $L_2(\mathbb{R}^d$ is still a remaining issue. By analogy with the 1-D case, the condition on the operator order $r > d/2$ should be sufficient for the decay of the error rate $||P_i f - f||_2 \rightarrow 0$ as $i \rightarrow -\infty$ for any $f \in W_2^r$, where P_i is the orthogonal projector on V_i.

Our wavelet bases have potential of application in areas of signal processing, where one is dealing with data convolved with a known imaging operator L^{-1}. The wavelets derived from this operator would behave like a multiscale version of L and essentially decorrelate the data, concentrating the object's energy into a small number of coefficients.

Chapter 9

Conclusion

We have suggested a way to construct wavelets in arbitrary dimensions based on a general scale-covariant differential operator. The framework is built in the spirit of discrete-continuous signal processing; it can be useful for problems that are characterized by differential operators. We have demonstrated its potential for temporal signal processing by applying the construction to the problem of brain activity detection in fMRI. The main contributions of this thesis are summarized in the following section.

9.1 Main Contributions

- Starting with a differential operator L, we introduced wavelets that act as a multiscale version of L. For a linear, shift-invariant system characterized by $\mathrm{L}y = x$, our wavelet transform essentially inverted the system response. We were particularly interested in the situation where the input signal x was sparse: in this case, the relevant components of the wavelet transform could be identified by means of a sparse-solution search algorithm. The search time remained reasonable thanks to the fast decomposition and reconstruction routines available within our framework.

- For the one-dimensional system defined by a linear differential equation, we provided the explicit expressions for the scaling functions and wavelets, as well as for scaling and wavelet filters. The classical B-spline wavelets turned out to be part of our construction that corresponded to the pure iterated derivative operator.

- By tuning the differential system to the balloon/windkessel model in fMRI, we obtained activelets — a special type wavelets for temporal fMRI signal processing. Under the assumption of sparse activity (which corresponded to the slow event-related paradigm), we have suggested a framework that allowed estimation of the activity-related signal in the fMRI measurements without knowledge of the actual onset times. We have tested the sensitivity of the algorithm on a synthetic dataset by means of an ROC curve. Unlike standard techniques that use a fixed shape model for BOLD response, our approach works just as well when the corresponding BOLD response exposes variability.
- Finally, we have verified that our algorithm worked on real data: it was able to detect local activations in the primary visual cortex. We have also identified the optimal event length, which came as a compromise between the requirements of short impulse-like events and reasonable SNR.

One more example of a brain imaging application where operator-like wavelets worked well was the reconstruction problem in dynamical PET [87]. We have left this work out of scope of the present thesis and briefly describe it in the following subsection.

9.1.1 Application to Dynamic PET

We assume that the imaged object consists of n tissue compartments that participate in the activity exchange process. The PET signal $y(\tau)$ is modeled by means of a state-space model that has the whole-blood activity and the plasma activity as input. The two input quantities are related to the injection activity through another system that models the tracer concentration dynamics. We approximate the agglomerated system by a linear differential operator L with zero and pole vectors $\vec{\alpha}, \vec{\gamma}$, respectively. The selection of $\vec{\alpha}, \vec{\gamma}$ is a study-dependent process; in [87], we give an example of parameter identification for the $[^{13}N] - NH_3$ kinetic model.

If we approximate the injection events with Dirac delta-functions, the measured PET signal will be a L-generated exponential spline, and

$$\mathrm{L}\{y\}(\tau) = \sum_t \delta(\tau - \tau_t),$$

where τ_t are the injection times. Therefore, the exponential-spline wavelets with the properly-tuned parameters provide a sparse representation of the activity-related signal. In order to extract it, we write the optimization criterion with a regularization term that forces sparsity in the temporal and spatial dimensions:

$$\min_{w} ||y - Rw|| + \lambda ||w||_1,$$

where R is the PET reconstruction operator, and w are the wavelet coefficients. We use traditional B-spline wavelets in space, and exponential-spline wavelets tailored to the model in time. As in the fMRI case, the sparsity constraint is expressed by using the l_1 norm of the coefficients. The solution is obtained with an iterative soft thresholding algorithm, where each iteration is followed by a reprojection on the space of admissible (positive) reconstructions. We verified the approach on a tomographic simulation study, as well as on two real PET data sets [87].

9.2 Future Work

While we have done the whole path from the theoretical concept to the practical data treatment, there is certainly much room left for further investigations. Some possibilities are listed below.

- The operator-like wavelet transform could be applied to other systems characterized by differential operators. It might also be possible to use a separable construction and combine the favorable behavior of operator-like wavelets in time with the good decorrelating properties of traditional wavelets in space.

- In this work, we have tested the activelet method on a single subject, and only visual tasks were used. The next important step to convince the neuroimaging community is to validate the method on many subjects and different tasks.

- A relevant future direction is the development of statistical methods that would produce parameter maps based on the activity-related signal estimated by the activelet framework. The work of Knight and Fu [88] could give some interesting indications in this respect.

- The key advantage of the activelet method is that it does not use any experimental timing information at the processing stage. One can think of scenarios, such as mental chronometry tasks, where precise knowledge of timing is not available. Our algorithm could be used, for example, to identify all locations where activation happened in a specific period of time.

For each task, a particular neurological question should be formulated and an appropriate statistical decision scheme should be established.

Appendix A

Appendices

A.1 Proof of Theorem 1

First, we outline some basic notations used in [21] to be able to perform the estimations. The spectrum of the first-order E-spline $|\hat{\beta}_{\sigma+j\omega_0}(\omega)|$ achieves its maximum $M_{\sigma+j\omega_0}$ at $\omega = \omega_0$. For the N-th order case, we denote $M_{\vec{\alpha}} = \prod_{l=1}^{N} M_{\alpha_l}$.

By referring to the general approximation results of Blu et al. [89], we consider the limit

$$\lim_{T\to 0} \frac{||f - P_T f||_{L_2}^2}{T^{2(N-M)}}$$

$$= \lim_{T\to 0} \int_{-\infty}^{\infty} \sum_{k\neq 0} |\hat{f}(\omega)|^2 \frac{|\hat{\beta}_{T\vec{\alpha},T\vec{\gamma}}(\omega T + 2\pi k)|^2/T^{2N}}{\sum_k |\hat{\beta}_{T\vec{\alpha},T\vec{\gamma}}(\omega T + 2\pi k)|^2/T^{2M}} \mathrm{d}\omega. \tag{A.1}$$

We denote $\Psi_1(k, T, \omega) = (\hat{\beta}_{T\vec{\alpha},T\vec{\gamma}}(\omega T + 2\pi k))/T^N$ and $\Psi_2(k, T, \omega) = \frac{1}{T^M} \cdot \sqrt{\sum_k |\hat{\beta}_{T\vec{\alpha},T\vec{\gamma}}(\omega T + 2\pi k)|^2}$.

To evaluate (A.1), we would like to exchange the $\lim_{T\to 0}$ and $\int \sum$ signs. To be able to apply Lebesgue's theorem, we must first prove that $|\frac{\Psi_1(k,T,\omega)}{\Psi_2(k,T,\omega)} \hat{f}(\omega)|^2$ is bounded by a summable and integrable function that does not depend on T. As we are interested in small values of T only, we assume that

$$T < \min\left\{ \frac{\pi}{\max_l |\gamma_l| + \max_l |\alpha_l|}, \frac{\pi}{3\max_l |\alpha_l|} \right\}. \tag{A.2}$$

Then, for $\Psi_1(k,T,\omega)$, we have that

$$\Psi_1(k,T,\omega) = \prod_{l=1}^{M} \frac{j\omega T - \gamma_l T + 2\pi k}{T(j\omega T - \alpha_l T + 2\pi k)}(1 - e^{T\alpha_l - Tj\omega + 2\pi k})$$

$$\cdot \prod_{l=M+1}^{N} \frac{1 - e^{T\alpha_l - Tj\omega + 2\pi k}}{T(j\omega T - \alpha_l T + 2\pi k)}.$$

We perform the partial fraction decomposition

$$\frac{j\omega T - \gamma_l T + 2\pi k}{(j\omega T - \alpha_l T)(j\omega T - \alpha_l T + 2\pi k)}$$

$$= \frac{(\alpha_l - \gamma_l)T/2\pi k + 1}{(j\omega T - \alpha_l T + 2\pi k)} + \frac{-(\alpha_l - \gamma_l)T/2\pi k}{(j\omega T - \alpha_l T)}.$$

From [21, Appendix A], we know that $|\frac{1-e^{T\alpha_l - Tj\omega+2\pi k}}{(j\omega T-\alpha_l T)(j\omega T-\alpha_l T+2\pi k)}| \leq \frac{M_{\alpha_l}}{\pi|k|}$. Thus, taking into account (A.2), we get the following estimation for Ψ_1:

$$|\Psi_1(k,T,\omega)| \leq \frac{1}{T^N}|L_{T\vec{\alpha}}(j\omega T)| \prod_{l=1}^{M} M_{T\alpha_l}(1 + \frac{|\alpha_l - \gamma_l|T}{\pi|k|})$$

$$\cdot \prod_{l=M+1}^{N} \frac{M_{\alpha_l}}{\pi|k|} \leq |L_{\vec{\alpha}}(j\omega)| \frac{2^M M_{\vec{\alpha}}}{(\pi|k|)^{N-M}}.$$

We now bound Ψ_2 from below. Using a technique similar to the one in [21] for the lower Riesz-bound estimation, we obtain

$$\frac{\Psi_2(k,T,\omega)}{|\prod_{l=1}^{M}(j\omega - \gamma_l)|} \geq \inf_{\omega \in [-\frac{\pi}{T}, \frac{\pi}{T}]} |\hat{\beta}_{T\vec{\alpha}}(\omega T)| \geq \frac{M_{T\vec{\alpha}}}{\pi^N}.$$

Consequently, for $|\frac{\Psi_1}{\Psi_2}\hat{f}|^2$ we get the bound

$$\left|\frac{\Psi_1(k,T,\omega)}{\Psi_2(k,T,\omega)}\hat{f}(\omega)\right|^2 \leq (2\pi)^{2M} \frac{|L_{\vec{\alpha},\vec{\gamma}}(j\omega)\hat{f}(\omega)|^2}{|k|^{2(N-M)}}$$

which is summable over k and integrable over ω under the assumptions of the theorem.

We thus exchange the limit with the integration and summation and calculate

$$\lim_{T\to 0}\left|\frac{\Psi_1(k,T,\omega)}{\Psi_2(k,T,\omega)}\hat{f}(\omega)\right|^2 = \left|\frac{\hat{f}(\omega)\prod_{l=1}^{N}(j\omega-\alpha_l)}{(2\pi k)^{N-M}\prod_{l=1}^{M}(j\omega-\gamma_l)}\right|^2$$

$$= \frac{|L_{\vec{\alpha},\vec{\gamma}}(j\omega)\hat{f}(\omega)|^2}{(2\pi k)^{2(N-M)}},$$

which yields the desired result

$$\lim_{T\to 0}\frac{||f-P_T f||_{L_2}^2}{T^{2(N-M)}} = C_{N,M}^2||L_{\vec{\alpha},\vec{\gamma}}f||_{L_2}^2.$$

■

Bibliography

[1] T. Olson and J. DeStefano, "Wavelet localization of the Radon transform," *IEEE Trans. Signal Process.*, vol. 42, no. 8, pp. 2055–2067, Aug. 1994.

[2] E.D. Kolaczyk, "A wavelet shrinkage approach to tomographic image reconstruction," *J. Am. Stat. Assoc.*, vol. 91, pp. 1079–1990, 1996.

[3] J. Wang and K. Huang, "Medical image compression by using three-dimensional wavelet transformation," *IEEE Transactions on Medical Imaging*, vol. 15, pp. 547–554, 1996.

[4] P. Thévenaz, U.E. Ruttimann, and M. Unser, "A pyramid approach to subpixel registration based on intensity," *IEEE Transactions on Image Processing*, vol. 7, no. 1, pp. 27–41, January 1998.

[5] J. B. Weaver, X. Yansun, Jr. D. M. Healy, and L. D. Cromwell, "Filtering noise from images with wavelet transforms," *Magnetic Resonance in Medicine*, vol. 21, pp. 288–295, 1991.

[6] Dimitri Van De Ville, Thierry Blu, and Michael Unser, "Surfing the brain: An overview of wavelet-based techniques for fMRI data analysis," *IEEE Engineering in Medicine and Biology Magazine*, vol. 25, no. 2, pp. 65–78, 2006.

[7] M. J. Fadili and E. T. Bullmore, "A comparative evaluation of wavelet-based methods for multiple hypothesis testing of brain activation maps," *NeuroImage*, vol. 23, no. 3, pp. 1112–1128, 2004.

[8] I. Khalidov and M. Unser, "From differential equations to the construction of new wavelet-like bases," *IEEE Transactions on Signal Processing*, vol. 54, no. 4, pp. 1256–1267, April 2006.

[9] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, 1998.

[10] G. Strang and T. Nguyen, *Wavelets and Filter Banks*, Wellesley-Cambridge, Wellesley, MA, 1996.

[11] M. Vetterli and J. Kovačević, *Wavelets and Subband Coding*, Prentice Hall, Englewood Cliffs, NJ, 1995.

[12] Ingrid Daubechies, “Orthogonal bases of compactly supported wavelets,” *Comm. Pure Appl. Math.*, vol. 41, pp. 909–996, 1988.

[13] I. Daubechies, “Orthonormal bases of compactly supported wavelets II. Variations on a theme,” *SIAM J. Math. Anal.*, vol. 24, no. 2, pp. 499–519, March 1993.

[14] A. Cohen, I. Daubechies, and J.C. Feauveau, “Bi-orthogonal bases of compactly supported wavelets,” *Comm. Pure Appl. Math.*, vol. 45, pp. 485–560, 1992.

[15] G. Battle, “A block spin construction of ondelettes. Part I: Lemarié functions,” *Comm. Math. Phys.*, vol. 110, pp. 601–615, 1987.

[16] P.-G. Lemarié, “Ondelettes à localisation exponentielle,” *J. Math. Pures. Appl.*, vol. 67, no. 3, pp. 227–236, 1988.

[17] M. Unser, A. Aldroubi, and M. Eden, “A family of polynomial spline wavelet transforms,” *Signal Processing*, vol. 30, no. 2, pp. 141–162, January 1993.

[18] M. Unser, A. Aldroubi, and M. Eden, “On the asymptotic convergence of B-spline wavelets to Gabor functions,” *IEEE Trans. Information Theory*, vol. 38, no. 2, pp. 864–872, 1992.

[19] M. Unser and T. Blu, “Wavelet theory demystified,” *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 470–483, February 2003.

[20] L.L. Schumaker, *Spline Functions: Basic Theory*, Wiley, New York, 1981.

[21] M. Unser and T. Blu, “Cardinal exponential splines: Part I—Theory and filtering algorithms,” *IEEE Trans. Signal Processing*, submitted.

[22] M. Unser, “Cardinal exponential splines: Part II—Think analog, act digital,” *IEEE Trans. Signal Processing*, submitted.

[23] J. Stöckler, "Non-stationary wavelets," *Multivariate approximation: from CAGD to wavelets*, pp. 307–320, 1993.

[24] C. de Boor, R.A. DeVore, and A. Ron, "On the construction of multivariate (pre)wavelets," *Constr. Approx.*, vol. 2, no. 3, pp. 123–166, 1993.

[25] C.K. Chui, J. Stöckler, and J.D. Ward, "Analytic wavelets generated by radial functions," *Adv. Comput. Math.*, vol. 5, pp. 95–123, 1996.

[26] A. Cohen and N. Dyn, "Nonstationary subdivision schemes and multiresolution analysis," *SIAM J. Math. Anal.*, vol. 27, no. 6, pp. 1745–1769, November 1996.

[27] Tom Lyche and Larry L. Schumaker, "L-spline wavelets," in *Wavelets: Theory, Algorithms, and Applications*, Charles K. Chui, Laura Montefusco, and Luigia Puccio, Eds. 1994, pp. 197–212, Academic Press.

[28] Stephan Dahlke, Wolfgang Dahmen, Eberhard Schmitt, and Ilona Weinreich, "Multiresolution analysis and wavelets on S^2 and S^3," *Numer. Funct. Anal. Opt.*, vol. 16(1&2), pp. 19–41, 1995.

[29] N. Dyn, D. Levin, and A. Luzzatto, "Exponentials reproducing subdivision schemes," *Found. Comput. Math.*, vol. 3, pp. 187–206, 2003.

[30] N. Dyn and A. Ron, "Local approximation by certain spaces of exponential polynomials, approximation order of exponential box splines, and related interpolation problems," *Trans. Amer. Math. Soc.*, vol. 319, pp. 381–403, 1990.

[31] M. Unser and T. Blu, "Mathematical properties of the JPEG2000 wavelet filters," *IEEE Transactions on Image Processing*, vol. 12, no. 9, pp. 1080–1090, September 2003.

[32] T. Blu and M. Unser, "The fractional spline wavelet transform: Definition and implementation," in *Proceedings of the Twenty-Fifth IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'00)*, Istanbul, Turkey, June 5-9, 2000, vol. I, pp. 512–515.

[33] C. de Boor, "Quasi-interpolation and approximation power of multivariate splines," in *Computation of Curves and Surfaces*, Dordrecht, 1990, pp. 313–345, Kluwer.

[34] R. Frackowiak, K. Friston, C. Frith, R. Dolan, and J. Mazziotta, *Human Brain Function*, Academic Press, 1997.

[35] K. J. Friston, A. Mechelli, R. Turner, and C. J. Price, “Nonlinear responses in fMRI: The balloon model, Volterra kernels and other hemodynamics,” *NeuroImage*, vol. 12, pp. 466–477, 2000.

[36] Richard B. Buxton, *Introduction to Functional Magnetic Resonance Imaging: Principles and Techniques*, Cambridge University Press, 2001.

[37] C.S. Roy and C.S. Sherrington, “On the regulation of the blood-supply of the brain,” *J. Physiol.*, vol. 11, pp. 85–108, 1890.

[38] N. Pouratian, S.A. Sheth, N.A. Martin, and A.W. Toga, “Shedding light on brain mapping: advances in human optical imaging,” *Trends in Neuroscience*, vol. 26, no. 5, pp. 277–282, May 2003.

[39] S. Ogawa and T. M. Lee, “Magnetic resonance imaging of blood vessels at high fields: in vivo and in vitro measurements and image simulation,” *Magnetic Resonance in Medicine*, vol. 16, pp. 9–18, 1990.

[40] R. B. Buxton, E.C. Wong, and L. R. Frank, “Dynamics of blood flow and oxygenation changes during brain activation: The balloon model,” *Magnetic Resonance in Medicine*, vol. 39, pp. 855–864, June 1998.

[41] J. B. Mandeville, J. J. Marota, C. Ayata, G. Zaharchuk, M. A. Moskowitz, B. R. Rosen, and R. M. Weisskoff, “Evidence of a cerebrovascular postarteriole windkessel with delayed compliance.,” *J Cereb Blood Flow Metab*, vol. 19, no. 6, pp. 679–689, June 1999.

[42] M. E. Raichle, “Behind the scenes of functional brain imaging: A historical and physiological perspective,” *PNAS*, vol. 95, no. 3, pp. 765–772, 1998.

[43] R. B. Buxton and L. R. Frank, “A model for the coupling between cerebral blood flow and oxygen metabolism during neural stimulation,” *Journal of Cerebral Blood Flow and Metabolism*, vol. 17, pp. 64–72, Jan. 1997.

[44] K. J. Friston, C. D. Frith, P. F. Liddle, and R. S. J. Frackowiak, “Functional connectivity: the principal component analysis of large (pet) datasets,” *J. Cereb. Blood Flow Metab.*, vol. 13, pp. 5–14, 1993.

[45] M. McKeown, S. Makeig, G. Brown, T. Jung, S. Kindermann, A. Bell, and T. Sejnowski, "Analysis of fMRI data by blind separation into independent spatial components," *Human Brain Mapping*, vol. 6, pp. 160–188, 1998.

[46] T. Adali and V. D. Calhoun, "Complex ICA of brain imaging data," *IEEE Signal Processing Magazine*, vol. 24, no. 5, pp. 136–139, sep 2007.

[47] G. H. Glover, "Deconvolution of impulse response in event-related BOLD fMRI," *NeuroImage*, vol. 9, no. 4, pp. 416–429, apr 1999.

[48] S. Makni, P. Ciuciu, J. Idier, and J.-B. Poline, "Joint detection-estimation of brain activity in functional MRI: A multichannel deconvolution solution," *IEEE Transactions on Signal Processing*, vol. 53, no. 9, pp. 3488–3502, sep 2005.

[49] A. Salek-Haddadi, B. Diehl, K. Hamandi, M. Merschhemke, A. Liston, K. Friston, J. S. Duncan, D. R. Fish, and L. Lemieux, "Hemodynamic correlates of epileptiform discharges: an EEG-fMRI study of 63 patients with focal epilepsy," *Brain Res.*, vol. 1088, no. 1, pp. 148–166, May 2006.

[50] M. Vetterli and J. Kovacevic, *Wavelets and subband coding*, Open-Access Edition, http://www.waveletsandsubbandcoding.org, 2nd edition, 2007.

[51] David L. Donoho and Iain M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.

[52] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, pp. 1413–1457, 2004.

[53] D. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization," *PNAS*, vol. 100, pp. 2197–2202, Mar. 2003.

[54] M. J. Fadili and E. Bullmore, "Wavelet-generalised least squares: a new BLU estimator of linear regression models with 1/f errors," *NeuroImage*, vol. 15, pp. 217–232, 2002.

[55] Voichita Maxim, Levent Sendur, Jalal Fadili, John Suckling, Rebecca Gould, Rob Howard, and Ed Bullmore, "Fractional Gaussian noise, functional MRI and Alzheimer's disease," *NeuroImage*, vol. 25, no. 1, pp. 141–158, Mar. 2005.

[56] M. J. Fadili and E. Bullmore, "Penalized partially linear models using sparse representations with an application to fMRI time series," *IEEE Transactions on Signal Processing*, vol. 53, no. 9, pp. 9, sep 2005.

[57] M. Unser and T. Blu, "Generalized smoothing splines and the optimal discretization of the Wiener filter," *IEEE Transactions on Signal Processing*, vol. 53, no. 6, pp. 2146–2159, June 2005.

[58] M. Unser, "Cardinal exponential splines: Part II—Think analog, act digital," *IEEE Transactions on Signal Processing*, vol. 53, no. 4, pp. 1439–1449, April 2005.

[59] Yu. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming*, SIAM, 1994.

[60] D. Donoho, M. Elad, and V. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inform. Theory*, vol. 52, pp. 6–18, 2006.

[61] Scott Shaobing Chen, David L. Donoho, and Michael A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1999.

[62] M. R. Osborne, B. Presnell, and B. A. Turlach, "A new approach to variable selection in least squares problems," *IMA J. Numerical Analysis*, vol. 20, pp. 389–403, 2000.

[63] D. Donoho and Y. Tsaig, "Fast solution of ℓ_1 minimization problems when the solution may be sparse," 2006, submitted.

[64] Antonin Chambolle, Ronalda. Devore, Nam yong Lee, and Bradley J. Lucier, "Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage," *IEEE Trans. Image Processing*, vol. 7, pp. 319–335, 1998.

[65] C. Vonesch and M. Unser, "A fast thresholded Landweber algorithm for wavelet-regularized multidimensional deconvolution," *IEEE Trans. Image Process.*, vol. 17, no. 4, pp. 539–549, 2008.

[66] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *SIAM Journal on Multiscale Modeling and Simulation*, vol. 4, no. 4, pp. 1168–1200, 2005.

[67] M.J. Fadili, J.-L. Starck, and F. Murtagh, "Inpainting and zooming using sparse representations," *The Computer Journal*, 2007, in press.

[68] D. Donoho, Y. Tsaig, I. Drori, and J-L. Starck, "Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit," *IEEE Transactions On Information Theory*, 2006, submitted.

[69] B. Efron, T. Hastie, I. M. Johnstone, and R. Tibshirani, "Least angle regression," *Annals of Statistics*, vol. 32, no. 2, pp. 407–499, 2004.

[70] M. J. Fadili and J.-L. Starck, "Sparse representation-based image deconvolution by iterative thresholding," in *Astronomical Data Analysis IV*, F. Murtagh and J.-L. Starck, Eds., Marseille, France, 2006.

[71] Jonathan Buckheit and David Donoho, "Wavelab and reproducible research," 1995, pp. 55–81, Springer-Verlag.

[72] David Donoho, Iddo Drori, Victoria Stodden, and Yaakov Tsaig, "Sparselab," *http://sparselab.stanford.edu/*, December 2005.

[73] K.J. Friston, P. Fletcher, O. Josephs, A.P. Holmes, M.D. Rugg, and R. Turner, "Event-related fMRI: characterizing differential responses," *NeuroImage*, vol. 7, pp. 30–40, 1998.

[74] O. Friman, M. Borga, P. Lundberg, and H. Knutsson, "Adaptive analysis of fMRI data," *NeuroImage*, vol. 19, no. 3, pp. 837–845, 2003.

[75] T. Inouye, *Die Schstorungen bei Schussverletzungen der kortikalen Sehsphare*, W. Englemann, 1909.

[76] G. Holmes and W.T. Lister, "Disturbances of vision from cerebral lesions, with special reference to the cortical representation of the macula," *Brain*, vol. 39, pp. 34–73, 1916.

[77] S. A. Engel, G. H. Glover, and B. A. Wandell, "Retinotopic organization in human visual cortex and the spatial precision of functional MRI," *Cerebral Cortex*, vol. 7, pp. 181–192, 1997.

[78] Kang Cheng, R. Allen Waggoner, and Keiji Tanaka, "Human ocular dominance columns as revealed by high-field functional magnetic resonance imaging," *Neuron*, vol. 32, pp. 359–374, 2001.

[79] C.A. Micchelli, C. Rabut, and F.I. Utreras, “Using the refinement equation for the construction of pre-wavelets III: Elliptic splines,” *Numerical Algorithms*, vol. 1, no. 3, October 1991.

[80] D. Van De Ville, T. Blu, and M. Unser, “Isotropic polyharmonic b-splines: Scaling functions and wavelets,” *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1798–1813, November 2005.

[81] D. Van De Ville and M. Unser, “Complex wavelet bases, steerability, and the marr-like pyramid,” *IEEE Trans. Image Process.*, in press.

[82] I. Khalidov and M. Unser, “From differential equations to the construction of new wavelet-like bases,” *IEEE Trans. Signal Process.*, vol. 54, no. 4, pp. 1256–1267, April 2006.

[83] M. Unser and T. Blu, “Generalized smoothing splines and the optimal discretization of the Wiener filter,” *IEEE Trans. Signal Process.*, vol. 53, no. 6, pp. 2146–2159, June 2005.

[84] Dirong Chen, “On the constructions of pre-wavelets in $L_2(\mathbb{R}^d)$,” *Acta Mathematicae Applicatae Sinica (English Series)*, vol. 14, no. 2, pp. 129–133, 1998.

[85] C. De Boor, R.A. DeVore, and A. Ron, *The structure of finitely generated shift-invariant spaces in* $L_2(\mathbb{R}^d)$, University of Wisconsin-Madison, 1992.

[86] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*, Prentice Hall, 1993.

[87] J. Verhaeghe, D. Van De Ville, I. Khalidov, Y. D’Asseler, I. Lemahieu, and M. Unser, “Dynamic PET reconstruction using wavelet regularization with adapted basis functions,” *IEEE Transactions on Medical Imaging*, vol. 27, no. 7, pp. 943–959, July 2008.

[88] K. Knight and W. Fu, “Asymptotics for Lasso-type estimators,” *Ann. Statist.*, vol. 28, no. 5, pp. 1356–1378, 2000.

[89] T. Blu and M. Unser, “Quantitative Fourier analysis of approximation techniques: Part I—Interpolators and projectors,” *IEEE Trans. Signal Process.*, vol. 47, no. 10, pp. 2783–2795, October 1999.

Acknowledgement

I would like to thank my thesis advisor, Prof. Michael Unser, for the scientific and personal guidance during these years. It is due to his example that I have chosen to become a scientist; it is his balanced strategy between guiding and letting do that taught me to be an independent researcher.

Big thanks to Dr. Dimitri Van De Ville for sharing his unique expertise in engineering and for helping me to see things in a positive way. I am grateful to Prof. Thierry Blu for making me rediscover the beauty of mathematics and to Dr. Philippe Thévenaz for assuring the quality of my work. I would also like to thank Dr. Daniel Sage for his responsiveness with all hardware and software-related questions.

Thanks to Manuelle Borruat for her invaluable support in numerous administrative issues.

I am grateful to the members of the jury — Prof. Lazeyras, Prof. Longchamp, Prof. Rabut and Prof. Vandergheynst — for accepting to read and comment on my thesis.

I thank BIG student members, in particular, Cédric Vonesch, Pouya D. Tafti and François Aguet for their help and cooperation.

The collaboration with the RIKEN BSI Cognitive Brain Mapping group has added important practical ideas to this thesis. I thank Dr. Kang Cheng for making my internship at RIKEN BSI possible. Thanks to CBM members, in particular, Dr. Kang Cheng, Mauro Costagli, Ken'ichi Ueno, Dr. R. Allen Waggoner and Dr. Pei Sun, for sharing their expertise in neuroscience and fMRI.

Special thanks to my friends, Cédric Vonesch and Helene Zach, for the strong support and for the wonderful moments that we have passed together. Big thanks to my friend Mauro Costagli for making my stay in Japan exciting and unforgettable.

I would like to thank my family, my parents and in particular my brother

Vassil for the strong support during these years.

This work would be impossible without my dear friends — Davide Longo and Sara Rüsch, who were always there to help me overcome the hard times and thanks to whom I felt at home in Switzerland. On their example, I am learning to be a better person. I dedicate this thesis to them.

Ildar Khalidov, Lausanne–Genève–Tokyo, October 2008

Curriculum Vitæ

Ildar Khalidov was born in St.-Petersburg, Russia in 1982. He received his M. Sc. (*summa cum laude*) degree in mathematics from the St.-Petersburg State University in 2003. He has been a research assistant at the Biomedical Imaging Group, EPFL, Lausanne (2003–2008) and at the Centre d'Imagerie Biomédicale, Radiology Department, University Hospital of Geneva (2006–2008). He has carried out his PhD at the Doctoral School of EPFL under the supervision of Prof. Michael Unser. He has mainly worked on wavelet design, multiresolution signal approximation and inverse problems in MRSI and fMRI.

His scientific interests include continuous-discrete signal processing, wavelet theory, applied harmonic analysis, inverse problems, MRI and fMRI.