Functions of Bounded Variation, Wavelets, and Applications to Image Processing

Anthony Morgante

A Thesis

in

The Department

of

Mathematics and Statistics

Presented in Partial Fulfillment of the Requirements

for the Degree of Masters of Science (Mathematics and Statistics) at

Concordia University

Montreal, Quebec, Canada

May 2006

Your file Votre référence
ISBN: 978-0-494-20730-7
Our file Notre référence
ISBN: 978-0-494-20730-7

ABSTRACT

Functions of Bounded Variation, Wavelets, and Applications to Image Processing

Anthony Morgante

A common problem in image processing is to decompose an observed image f into a sum $u+v$, where u represents the more vital features of the image, i.e. the objects, and v represents the textured areas and any noise that may be present. The benefit of such a decomposition is that the "u" component represents a compressed and noise reduced version of the original image f.

The space BV of functions of bounded variation has been known to work very well as a model space for the objects in an image because indicator functions of sets whose boundary is finite in length belong to BV. This thesis is aimed at investigating the mathematical properties of the space BV while looking at a very well known "u+v" model, called the ROF model, in which it is assumed that $u \in BV$.

More recent work has shown that the optimal pair (u, v) to many decomposition problems can be obtained by expanding a given image f into a wavelet basis and performing simple operations on the wavelet coefficients. This thesis will provide a detailed introduction to the theory of orthonormal wavelets, giving some important examples of their effectiveness, as well as showing comparisons of wavelet bases with classical Fourier series.

Contents

1 Notation

C_c^m	; space of continuously m-differentiable functions with compact support
C_0^m	; space of continuously m-differentiable functions that vanish at infinity
$\|f\|_p$	; L^p norm of f; $(\int \|f(x)\|^p dx)^{1/p}$
$\mathcal{H}^n$	; n-dimensional Hausdorff measure, p. 6
$\mathcal{L}^n$	; n-dimensional Lebesgue measure, p. 7
$\hat{f}$	; Fourier transform of f, p. 28
$\mathcal{S}(\mathbb{R}^n)$	; Schwartz class of smooth functions with rapid decay, p. 4
$\mathcal{S}'(\mathbb{R}^n)$	; Space of tempered distributions
$\mu_1 \ll \mu_2$	; μ_1 is absolutely continuous with respect to μ_2
$\mu_1 \perp \mu_2$	; μ_1 and μ_2 are mutually singular
$\|Df\|$	; variation measure of f, p. 4
$\|f\|_{BV}$	; BV norm/semi-norm of f, pp. 9,16
$\|f\|_*$	; norm of f in the space G, p. 18
BV	; space of functions of bounded variation, p. 3
$W^{k,p}$	; Sobolev spaces, p. 3
$B_p^{s,q}$	; Besov spaces, p. 53
$\mathrm{div}\vec{\varphi}$	; Divergence of the vector $\vec{\varphi}$
$\vec{\nabla} f$	; Gradient of f

2 Introduction

Over recent years, many models have been introduced to address the issues of restoration and compression of black and white digital images. A wide class of these models can be categorized as "$u + v$" models. What is consistent in all "$u + v$" models is that, given an observed image f which may possibly contain some noisy data, one would like to decompose f as a sum $f = u+v$ where the function u is a well structured function in some smoothness space which is ideal for compression and represents a good approximation of the original image, while the function v contains the noisy data and some less interesting features of the image, such as textures, that one is willing to disregard in order to free up memory. A well known "$u+v$" model is the model introduced by Leonid Rudin, Stanley Osher, and Emad Fatemi in [17]. This model is referred to as the ROF model. The assumptions of this model are that the observed image $f(x) = f(x_1, x_2)$ is an $L^2(\mathbb{R}^2)$ function that can be written as a sum $f = u + v$ where u belongs to the space $BV(\mathbb{R}^2)$ and represents the objects of the image while v is an $L^2(\mathbb{R}^2)$ function and contains the textures and noise of the image.

With the ROF model and with many similar "u+v" models, one attempts to minimize a certain energy. In the past, it has been customary to solve such minimization problems using PDE methods, but more recently, wavelet based methods have been employed. When wavelets are used, a given image f is expanded into a wavelet basis and a simple operation, referred to as wavelet thresholding, is performed on each of the wavelet coefficients of f. As for the construction of wavelets, most orthonormal wavelet bases can be constructed by means of what is called a multiresolution analysis (MRA). Furthermore, when the wavelet bases that are formed by an MRA have additional properties such as sufficient decay and smoothness, many function spaces, such as Besov spaces, can be characterized by size conditions on their wavelet coefficients. Consequently, thresholding is a stable operation for many of these function spaces.

The breakdown of this thesis is as follows; section three will give an account on the space BV which plays an important role in the modelling of objects in images when the ROF model is used; in section four the ROF model, along with some important theorems regarding the optimal "$u + v$" decomposition, will be described, as well as some interesting examples of images and textures and how they are treated by the ROF model; section five will give some background information about wavelet analysis and construction of orthonormal wavelets by means of a multiresolution analysis; in section six Besov spaces will be introduced; and in section seven we will establish a connection between wavelet analysis and image processing, in particular, the ROF model, and some comparisons between Fourier and wavelet series will be made.

With regards to originality, the author has filled in many details throughout this thesis which were not included by the respective references. In section 3.1, details have been added to Example 3.1 and the author has added a proof for Theorem 3.6, which was stated without proof in [13]. The example given in 3.2 is a concrete realization of the type of set mentioned in [13]. In section 4, Lemma 4.1, which was stated in [13], has been proved, while details have been added for the proofs of Lemma 4.2 and Theorems 4.2 and 4.3, which were all proved in [13]. The example given in section 4.2 is original and some details for the example concerning Meyer's wavelet in section 5.2 (which was found in [8]) were added as well. Lastly, details were added to the example concerning the minimization problem in section 7.3, which was originally given in [7].

3 Functions of Bounded Variation

This first section gives an overview of some of the basic properties of functions of bounded variation, which is denoted the space BV. This space has become the standard for the modelling of images, as indicator functions of sets whose boundaries are finite in length belong to BV. This statement can actually be improved significantly as we will see when discussing the reduced boundary of a set. Indicator functions are important in imaging because their boundaries represent edges, which are the main features of the objects in an image.

3.1 Properties of the space BV

We begin this section by defining the space BV and some other well known spaces.

Definition 3.1 *A real-valued function $f \in L^1(U)$ is said to have bounded variation in $U \subset \mathbb{R}^n$, and is written $f \in BV(U)$, if*

$$\sup\left\{\int_U f \operatorname{div}\vec{\varphi}\, dx \ ; \ \vec{\varphi} \in C_c^1(U;\mathbb{R}^n), \|\vec{\varphi}\|_\infty \leq 1\right\} < \infty. \tag{3.1}$$

Definition 3.2 *Given $U \subset \mathbb{R}^n$ and a non-negative integer k, a function f belongs to the Sobolev Space $W^{k,p}(U)$ if*

$$f \in L^p(U) \qquad \textit{and} \qquad \frac{\partial^\alpha f}{\partial x^\alpha} \in L^p(U)$$

for any α such that $|\alpha| \leq k$, where $|\alpha| = \alpha_1 + ... + \alpha_n$ and the derivatives,

$$\frac{\partial^\alpha f}{\partial x^\alpha} = \frac{\partial^{\alpha_1} ... \partial^{\alpha_n} f}{\partial x_1 ... \partial x_n},$$

are taken in the weak sense. When endowed with the norm

$$\|f\|_{W^{k,p}} = \|f\|_p + \sum_{|\alpha|\leq k} \left\|\frac{\partial^\alpha f}{\partial^\alpha x}\right\|_p,$$

the space $W^{k,p}$ is a Banach space. A function f belongs to the homogeneous Sobolev space $\dot{W}^{k,p}$ if the semi-norm,

$$\|f\|_{\dot{W}^{k,p}} = \sum_{|\alpha|=k} \left\|\frac{\partial^\alpha f}{\partial^\alpha x}\right\|_p,$$

is finite.

Remark. The meaning of differentiability in the "weak sense" in the above definition is the following: A function $g \in L^1(\mathbb{R}^n)$ is the weak derivative of $f \in L^1(\mathbb{R}^n)$ if

$$\int_{\mathbb{R}^n} f(x)\varphi'(x)dx = -\int_{\mathbb{R}^n} g(x)\varphi(x)dx$$

for all $\varphi \in C_c^\infty(\mathbb{R}^n)$.

Definition 3.3 *A function $f \in C^\infty(\mathbb{R}^n)$ belongs to the Schwartz class $\mathcal{S}(\mathbb{R}^n)$ if there are constants $C_{\alpha,\beta}$ such that*

$$\sup_{x\in\mathbb{R}^n} \left| x^\alpha \frac{\partial^\beta f}{\partial x^\beta} \right| \leq C_{\alpha,\beta}$$

for each multi-indices α and β. The dual space of tempered distributions will be denoted by $\mathcal{S}'(\mathbb{R}^n)$.

We will soon see the relationship between BV and the Sobolev spaces. The space BV, as this next theorem illustrates, contains the functions whose weak first partial derivatives are Radon measures.

Theorem 3.1 *Let $f \in BV(U)$. Then there exists a positive Radon measure μ on U and a μ-measurable function $\vec{\sigma} : U \longrightarrow \mathbb{R}^n$ such that*

(i) $|\vec{\sigma}(x)| = 1$ *μ a.e.*

(ii) $\int_U f \mathrm{div}\vec{\varphi}\, dx = -\int_U \vec{\varphi}\cdot\vec{\sigma}\, d\mu$

The proof of Theorem 3.1 will not be given here but can be found in [10]. We will denote the measure μ by

$$\mu = \|Df\|$$

and we call this measure the *variation measure* of f. The total variation of f, $\|Df\|(U)$, is given by (3.1) and moreover, for any V compactly contained in U, we have

$$\|Df\|(V) = \sup\left\{ \int_V f\mathrm{div}\vec{\varphi}\, dx \; ; \; \vec{\varphi} \in C_c^1(V;\mathbb{R}^n), \|\vec{\varphi}\|_\infty \leq 1 \right\}.$$

We must also associate with this measure, the vector-valued measure $[Df]$ which is given by

$$d[Df] = \vec{\sigma}\, d\|Df\|\,.$$

Example 3.1

Let $f \in W^{1,1}(U)$. Then given $\vec{\varphi} \in C_c^1(U;\mathbb{R}^n)$ with $\|\vec{\varphi}\|_\infty \leq 1$ we have,

$$\int_U f \mathrm{div}\vec{\varphi}\, dx = -\int_U \vec{\nabla} f \cdot \vec{\varphi}\, dx \leq \int_U |\vec{\nabla} f| dx < \infty, \tag{3.2}$$

where $\vec{\nabla} f$ denotes the gradient of f:

$$\vec{\nabla} f = (\frac{\partial f}{\partial x_1}, ..., \frac{\partial f}{\partial x_n}).$$

This implies $f \in BV(U)$. Thus we have

$$W^{1,1}(U) \subset BV(U).$$

Furthermore,

$$\|Df\|\,(\mathbf{R}^n) = \int_{\mathbf{R}^n} |\nabla f| dx. \tag{3.3}$$

It is sufficient to prove (3.3) for $f \in C_c^\infty(\mathbb{R}^n)$ since $C_c^\infty(\mathbb{R}^n)$ is dense in $W^{1,1}(\mathbb{R}^n)$ and the general case can be obtained by using a sequence $f_k \in C_c^\infty(\mathbb{R}^n)$ converging to $f \in W^{1,1}(\mathbb{R}^n)$. Then, given $f \in C_c^\infty(\mathbb{R}^n)$, it suffices to find a sequence $\vec{\varphi}_k \in C_c^1(\mathbb{R}^n;\mathbb{R}^n)$ with the following two properties; $\vec{\varphi}_k$ converges, in the vector-valued L^2 sense, to the vector-valued function $\vec{g} = (g_1, ..., g_n)$ defined by

$$g_i = \begin{cases} -\dfrac{\partial_{x_i} f}{|\vec{\nabla} f|} & \text{if } |\vec{\nabla} f|(x) \neq 0 \\ \\ 0 & \text{if } |\vec{\nabla} f|(x) = 0 \end{cases}$$

and $\|\vec{\varphi}_k\|_\infty \leq \|\vec{g}\|_\infty$ for each $k \in \mathbb{Z}$ since $\|\vec{g}\|_\infty = 1$. For finding such a sequence $\vec{\varphi}_k$, we can use mollifiers, i.e. consider a nonnegative function $\eta \in C_c^\infty(\mathbb{R}^n)$ such that

$$\int_{\mathbb{R}^n} \eta(x) dx = 1 \tag{3.4}$$

and

$$\eta(0) = 1.$$

Then define

$$\eta_\epsilon(x) = \frac{1}{\epsilon^n}\eta(x/\epsilon)$$

and $\vec{\varphi}_\epsilon = (\varphi_1^\epsilon, \varphi_2^\epsilon, ..., \varphi_n^\epsilon)$, where

$$\varphi_i^\epsilon(x) = \int_{\mathbb{R}^n} g_i(x)\eta_\epsilon(x-y)dy$$

for each $i = 1, ..., n$. Then for each fixed $\epsilon > 0$, $\vec{\varphi}_\epsilon$ belongs to $C_c^\infty(\mathbb{R}^n;\mathbb{R}^n)$ and, by (3.4), $\|\vec{\varphi}_\epsilon\|_\infty \leq \|\vec{g}\|_\infty$. To prove (3.3), we then have

$$\left|\int(\vec{\varphi}_\epsilon - \vec{g})\cdot\vec{\nabla}f dx\right| \leq (\int|\vec{\varphi}_\epsilon - \vec{g}|^2dx)^{1/2}(\int|\vec{\nabla}f|^2dx)^{1/2}$$

where the first integral on the right goes to zero as $\epsilon \to 0^+$ and the second integral is finite since $f \in C_c^\infty(\mathbb{R}^n)$. Thus

$$\lim_{\epsilon\to0^+}\int f\mathrm{div}\vec{\varphi}_\epsilon dx = -\lim_{\epsilon\to0^+}\int\vec{\varphi}_\epsilon\cdot\vec{\nabla}f dx = -\int\vec{g}\cdot\vec{\nabla}f dx = \int|\vec{\nabla}f|dx.$$

Coupling the above equality with (3.2), we have proved (3.3).

Example 3.2

Let E be a bounded open set in $\mathbb{R}^n$ such that ∂E is a C^2 boundary with $\mathcal{H}^{n-1}(\partial E) < \infty$, where $\mathcal{H}^n$ denotes the n-dimensional Hausdorff measure. Then by the divergence theorem, for $\vec{\varphi} \in C_c^1(\mathbb{R}^n;\mathbb{R}^n)$ such that $\|\vec{\varphi}\|_\infty \leq 1$,

$$\int_{\mathbb{R}^n}\chi_E\ \mathrm{div}\vec{\varphi}\ dx = \int_E \mathrm{div}\vec{\varphi}\ dx = \int_{\partial E}\vec{\varphi}\cdot\vec{\nu}\ d\mathcal{H}^{n-1},$$

where $\vec{\nu}$ is the unit normal along ∂E. Then since E is bounded and ∂E is C^2, we can find a compact set K such that $E \subset K$ and a vector field $\vec{\varphi} \in C_c^1(\mathbb{R}^n;\mathbb{R}^n)$ such that $\vec{\varphi}(x) = \vec{\nu}(x)$ for $x \in \partial E$ and $\vec{\varphi}(x) = 0$ for $x \notin K$. Hence,

$$\begin{aligned}\sup\int_E \mathrm{div}\vec{\varphi}\ dx &= \int_{\partial E}|\vec{\nu}|^2\ d\mathcal{H}^{n-1}\\ &= \mathcal{H}^{n-1}(\partial E)\\ &< \infty,\end{aligned}$$

where the supremum is taken over all $\vec{\varphi} \in C_c^1(\mathbb{R}^n;\mathbb{R}^n)$ with $\|\vec{\varphi}\|_\infty \leq 1$. Then we have shown

$$\|D\chi_E\|(\mathbb{R}^n) = \mathcal{H}^{n-1}(\partial E).$$

Remark. A set E whose characteristic function belongs to BV, such as the one seen in Example 3.2, is referred to as a set with *finite perimeter.* In this case, the variation measure $\|D\chi_E\|$ will be denoted by

$$\|\partial E\|$$

and the function $\vec{\sigma}$ given in Theorem 3.1 will be denoted by ν_E.

Lebesgue Decomposition

For any $f \in BV(U)$, if we write

$$d\mu_k = \sigma_k d\|Df\| \quad \text{for } k = 1, ..., n$$

where $\vec{\sigma} = (\sigma_1, ..., \sigma_n)$, then by Lebesgue's Decomposition Theorem (see [18]), we have

$$\mu_k = \mu_k^a + \mu_k^s$$

such that

$$\mu_k^a \ll \mathcal{L}^n \quad \text{and} \quad \mu_k^s \perp \mathcal{L}^n.$$

Here $\mathcal{L}^n$ denotes the n-dimensional Lebesgue measure. Furthermore, for each k, there exists a unique function $h_k \in L^1(U)$ such that

$$d\mu_k^a = h_k \, d\mathcal{L}^n.$$

We can then write $h_k = \partial f / \partial x_k$ and

$$Df = (\frac{\partial f}{\partial x_1}, ..., \frac{\partial f}{\partial x_n}), \quad [Df]^a = (\mu_1^a, ..., \mu_n^a), \quad [Df]^s = (\mu_1^s, ..., \mu_n^s),$$

so that $[Df] = [Df]^a + [Df]^s$ and $d[Df]^a = Df\ d\mathcal{L}^n$. Thus

$$d[Df] = Df\ d\mathcal{L}^n + d[Df]^s. \tag{3.5}$$

Remark. We should mention that a function belongs to the Sobolev space $W^{1,1}$ if and only if the singular measure $[Df]^s$ is zero. To see this, suppose first that $[Df]^s = 0$. Then by the Lebesgue decomposition above, the weak first partial derivatives of f all belong to $L^1(\mathbb{R}^n)$ which implies $\vec{\nabla} f \in L^1(\mathbb{R}^n)$. Conversely, suppose $f \in W^{1,1}$. Then from Example 3.1, we have

$$d[Df] = \vec{\sigma} d \|Df\| = \vec{\nabla} f d\mathcal{L}^n.$$

However, by (3.5), $d[Df] = \vec{\nabla} f d\mathcal{L}^n + d[Df]^s$ implying $[Df]^s = 0$. This remark is also telling us that for such a set E as given in Example 3.2, $\chi_E \notin W^{1,1}$. Of course, the weak first partial derivatives of $\chi_E(x)$ for $x \in \partial E$ correspond to a delta function δ on ∂E which is in fact a singular measure.

We would now like to define weak-$*$ convergence of BV functions. In order to do so, it will be helpful to view BV as a dual of some Banach space. As it is mentioned in [13], BV is the dual of the space Γ of all tempered distributions f that can be written as $f = \mathrm{div}\vec{g}$, where $\vec{g} = (g_1, ..., g_n)$ and $g_1, ..., g_n \in C_0(\mathbb{R}^n)$. We then have the following definition.

Definition 3.4 *A sequence of functions f_k converge to f in the weak-$*$ topology of $BV(U)$ if for any $\vec{\varphi} \in C_c^1(U;\mathbb{R}^n)$ with $\|\vec{\varphi}\|_\infty \leq 1$,*

$$\lim_{k\to\infty} \int_U f_k(x)\mathrm{div}\vec{\varphi}(x)dx = \int_U f(x)\mathrm{div}\vec{\varphi}(x)dx.$$

Theorem 3.2 (Lower Semicontinuity) *Let $\{f_k\} \subset BV(U)$ such that $f_k \to f$ in the weak-$*$ topology of $BV(U)$. Then*

$$\|Df\|(U) \leq \liminf_{k\to\infty} \|Df_k\|(U).$$

Proof. Given any $\vec{\varphi} \in C_c^1(U;\mathbb{R}^n)$ with $\|\vec{\varphi}\|_\infty \leq 1$ we have,

$$\begin{aligned} \int_U f \mathrm{div}\vec{\varphi}\, dx &= \lim_{k\to\infty} \int_U f_k \mathrm{div}\vec{\varphi}\, dx \\ &= -\lim_{k\to\infty} \int_U \vec{\varphi}\cdot d[Df_k] \\ &\leq \liminf_{k\to\infty} \|Df_k\|(U), \end{aligned}$$

where the last inequality follows since $\int_U \vec{\varphi}\cdot d[Df_k] \leq \|Df_k\|(U)$ by definition. Taking the supremum over all such $\vec{\varphi}$ we have

$$\|Df\|(U) \leq \liminf_{k\to\infty} \|Df_k\|(U).$$

Theorem 3.3 *The space $BV(U)$ is a Banach space when endowed with the norm*

$$\|f\|_{BV} = \|f\|_1 + \|Df\|(U).$$

Proof. The norm properties of $\|\cdot\|_{BV}$ are obvious and are left to the reader. We will need to show completeness, i.e. any Cauchy sequence f_k in $BV(U)$, converges to some $f \in BV(U)$ with convergence in the BV-norm. Given such a sequence f_k and an $\epsilon > 0$, then for k, m sufficiently large

$$\|f_k - f_m\|_{BV} < \epsilon \implies \|f_k - f_m\|_1 < \epsilon$$

and since $L^1(U)$ is a Banach space this implies there is a function $f \in L^1(U)$ such that $\|f - f_k\|_1 \to 0$. Since f_k is a Cauchy sequence then the $\|Df_k\|(U)$ are uniformly bounded and since L^1 convergence implies convergence in the weak-$*$ topology of BV, Theorem 3.2 tells us that $f \in BV(U)$. We still must show that $\|f - f_k\|_{BV} \to 0$. Since we already know $f_k \to f$ in $L^1(U)$, then all we need to show is that $\|D(f_k - f)\|(U) \to 0$. Since

$$\|f_k - f_m\|_{BV} < \epsilon \implies \|D(f_k - f_m)\|(U) < \epsilon$$

and for fixed k we know

$$\|(f_k - f_m) - (f_k - f)\|_1 \to 0 \qquad \text{as } m \to \infty,$$

then using the weak-$*$ convergence of $(f_k - f_m)$ to $(f_k - f)$, Theorem 3.2 yields

$$\|D(f_k - f)\|(U) \leq \liminf_{m\to\infty} \|D(f_k - f_m)\|(U) < \epsilon$$

for k chosen sufficiently large. This proof was due to [11].

Remark. By considering only the space of functions f such that the seminorm $\|Df\|(U) < \infty$ we obtain a homogeneous version of the space BV which we can denote by $\dot{BV}$. This space consists of functions modulo constants. We then have the embedding $\dot{W}^{1,1} \subset \dot{BV}$ which follows directly from Example 3.1.

Theorem 3.4 (Smooth approximation of BV) *Let $f \in BV(U)$. Then there exists functions $f_k \in BV(U) \cap C^\infty(U)$ such that*
(a) $\|Df_k\|(U) \to \|Df\|(U)$ as $k \to \infty$
(b) $f_k \to f$ in $L^1(U)$.

The proof of this theorem is obtained by the use of mollifiers and the details can be found in [10]. Another useful theorem found in [10] is the following.

Theorem 3.5 *For the functions f_k and f given in Theorem 3.4, define the vector-valued Radon measures μ_k by*

$$d\mu_k = \vec{\nabla} f_k dx.$$

Then $\mu_k \rightharpoonup [Df]$ weakly in the sense of vector-valued Radon measures.

Remark. We should note that the assertion $\|D(f_k - f)\|(U) \to 0$ does not hold for Theorem 3.4. Indeed, as it is discussed in [13], the closure of the Schwartz class $\mathcal{S}(\mathbb{R}^n)$ in $\dot{BV}$ is not all of $\dot{BV}$ but is the subspace $\dot{W}^{1,1}$. We now state this theorem formally and provide a proof.

Theorem 3.6 *The closure of $S(\mathbb{R}^n)$ in $\dot{BV}$ is the space $\dot{W}^{1,1}$.*

Proof. Let f_k be a sequence belonging to $\mathcal{S}(\mathbb{R}^n)$ and suppose f_k converges to some f in $\dot{BV}$, i.e.,

$$\lim_{k\to\infty} \|D(f_k - f)\| = 0. \tag{3.6}$$

Then this implies the f_k are a Cauchy sequence in $\dot{BV}$ as well. Therefore since

$$\begin{aligned}
\|D(f_k - f_m)\| &= \sup\left\{ \int_{\mathbb{R}^n} (f_k - f_m)(x)\operatorname{div}\vec{\varphi}(x)dx;\ \ \vec{\varphi} \in C_c^1(\mathbb{R}^n;\mathbb{R}^n),\ \ \|\vec{\varphi}\|_\infty \le 1 \right\} \\
&= \sup\left\{ \int_{\mathbb{R}^n} \vec{\nabla}(f_k - f_m)(x)\cdot\vec{\varphi}(x)dx;\ \ \vec{\varphi} \in C_c^1(\mathbb{R}^n;\mathbb{R}^n),\ \ \|\vec{\varphi}\|_\infty \le 1 \right\} \\
&= \int_{\mathbb{R}^n} |\vec{\nabla}(f_k - f_m)|(x)dx,
\end{aligned}$$

we have

$$\lim_{k,m\to\infty} \int_{\mathbb{R}^n} |\vec{\nabla}(f_k - f_m)|(x)dx = 0.$$

The integral on the left is equivalent to the $L^1(\mathbb{R}^n;\mathbb{R}^n)$ norm of $\vec{\nabla}(f_k - f_m)$ given by

$$\left\|\vec{\nabla}(f_k - f_m)\right\|_1 = \sum_{i=1}^{n} \left\|\frac{\partial(f_k - f_m)}{\partial x_i}\right\|_1.$$

Thus for each $i \in \{1, ..., n\}$, by the completeness of $L^1(\mathbb{R}^n)$ there exists a function $g_i \in L^1(\mathbb{R}^n)$ such that

$$\lim_{k\to\infty} \left\|\frac{\partial f_k}{\partial x_i} - g_i\right\|_1 = 0.$$

Let $\vec{g} = (g_1, ..., g_n)$ and take any $\vec{\varphi} \in C_c^1(\mathbb{R}^n;\mathbb{R}^n)$ such that $\|\vec{\varphi}\|_\infty \le 1$. Then we have

$$\begin{aligned}
\int_{\mathbb{R}^n} \vec{g}\cdot\vec{\varphi}\ dx &= \lim_{k\to\infty} \int_{\mathbb{R}^n} \vec{\nabla} f_k \cdot \vec{\varphi}\ dx \\
&= \lim_{k\to\infty} \int_{\mathbb{R}^n} f_k \operatorname{div}\vec{\varphi}\ dx \\
&= \int_{\mathbb{R}^n} f \operatorname{div}\vec{\varphi}\ dx.
\end{aligned}$$

Thus $\vec{g} = \vec{\nabla} f$ in the weak sense which implies $\vec{\nabla} f \in L^1(\mathbb{R}^n;\mathbb{R}^n)$.

Theorem 3.7 *There exists a constant C such that*

$$\|f\|_{L^{n/n-1}(\mathbb{R}^n)} \le C\,\|Df\|\,(\mathbb{R}^n).$$

Proof. Let $f \in BV(\mathbb{R}^n)$. Then from Theorem 3.4, we can find a sequence of smooth functions f_k such that

$$\lim_{k\to\infty} \|f_k - f\|_1 = 0$$

and

$$\|Df\|(\mathbb{R}^n) = \lim_{k\to\infty} \|Df_k\|(\mathbb{R}^n). \tag{3.7}$$

The L^1 convergence of f_k to f implies convergence $\mathcal{L}^n$ a.e. Then by Fatou's Lemma,

$$\|f\|_{L^{n/n-1}(\mathbb{R}^n)} \leq \liminf_{k\to\infty} \|f_k\|_{L^{n/n-1}(\mathbb{R}^n)}. \tag{3.8}$$

The Gagliardo-Nirenberg-Sobolev inequality says there exists a constant C such that for all k,

$$\|f_k\|_{L^{n/n-1}(\mathbb{R}^n)} \leq C \|Df_k\|_1, \tag{3.9}$$

and since $\|Df_k\|_1 = \|Df_k\|(\mathbb{R}^n)$, we have by (3.7),(3.8), and (3.9) that

$$\|f\|_{L^{n/n-1}(\mathbb{R}^n)} \leq C \|Df\|(\mathbb{R}^n).$$

This theorem will be of particular interested in the sections of this paper regarding image analysis as for the 2-dimensional case it gives the continuous embedding of $BV(\mathbb{R}^2) \subset L^2(\mathbb{R}^2)$.

3.2 The Reduced Boundary: Construction of a Swiss Cheese set

In Example 3.2 we observed that if we are given a bounded open set $E \subset \mathbb{R}^n$ with smooth boundary ∂E, then $\|\partial E\|(\mathbb{R}^n) = \mathcal{H}^{n-1}(\partial E)$. In general, however, $\|\partial E\|(\mathbb{R}^n) \leq \mathcal{H}^{n-1}(\partial E)$ for an arbitrary open set E. As mentioned by Y. Meyer in [13], there is a subset $\partial^* E$ of ∂E, referred to as the *reduced boundary* of E, such that $\|\partial E\|(\mathbb{R}^n) = \mathcal{H}^{n-1}(\partial^* E)$ for any open set E. This result was proved by Ennio De Giorgi, who is also responsible for the definition of $\partial^* E$. Recalling the definition of $\vec{\nu}_E$ given in the remark that followed Example 3.2, we now define the reduced boundary.

Definition 3.5 *Given an open set $E \subset \mathbb{R}^n$ and $x \in \partial E$. We say $x \in \partial^* E$, the reduced boundary of E, if the following conditions are satisfied:*

$$\|\partial E\|\,(B(x,r)) > 0 \quad for\ all\ r > 0, \tag{3.10}$$

$$\lim_{r\to 0} \frac{1}{\|\partial E\|\,(B(x,r))} \int_{B(x,r)} \vec{\nu}_E d\,\|\partial E\| = \vec{\nu}_E(x), \tag{3.11}$$

and

$$|\vec{\nu}_E(x)| = 1. \tag{3.12}$$

In [13], Meyer suggests the existence of a set E, referred to as a "swiss cheese" set, such that $\partial^* E$ is much smaller than ∂E with respect to the measure $\mathcal{H}^{n-1}$. More precisely, that $\|\partial E\|\,(\mathbb{R}^n)$ is finite while $\mathcal{H}^{n-1}(\partial E) = \infty$.

We wish to construct such a set E in 2-dimensions. This set will be the union of a countable number of open squares centered at the vertices of a dyadic grid. Furthermore E will be dense in the unit square C. The term "swiss cheese" is used to describe this type of set because, as one can imagine it, has many holes in it.

To begin the construction, let W_1 be the open square of side length 9^{-1} centered at $(2^{-1}, 2^{-1})$. Then for $n \geq 2$ let $\widetilde{Q}^n_{j,k}$ be the open square of side length 9^{-n} centered at $(j2^{-n}, k2^{-n})$ and let

$$W_n = \bigcup_{j,\ k=1}^{2^n-1} Q^n_{j,k} \ \text{ where } \ Q^n_{j,k} = \begin{cases} \widetilde{Q}^n_{j,k} & \text{if } \widetilde{Q}^n_{j,k} \cap W_{n-1} = \phi \\ \\ \phi & \text{otherwise.} \end{cases}$$

Then we let

$$E_n = \bigcup_{i=1}^{n} W_i \ \text{ and } \ E = \bigcup_{i=1}^{\infty} W_i.$$

We can now show some interesting properties of the set E. Our first claim is that E is dense in the unit square C. To see this, define the set of dyadic points by

$$D_n = \{(j2^{-n}, k2^{-n});\ j,k = 1, ..., 2^n - 1\}$$

and

$$D = \bigcup_{n=1}^{\infty} D_n.$$

Then given a ball $B \subset C \setminus D_n$ of radius r we see that $r \leq 2^{-n+1/2}$ so that $C \setminus D$ contains no ball of positive radius, implying that D is dense in C. Then by the construction of E, given any point $x \in D$, for every n there exists a point $y_n \in E_n$ such that

$$|x - y_n| \leq \sqrt{2}\,9^{-n},$$

so that $D \subset \overline{E}$ which implies $\overline{D} \subset \overline{E}$. Combining that with the fact that $\overline{D} = C$ and, from our construction, that $E \subset C$ we have $\overline{E} = C$ and we are finished.

Our second claim is that $\mathcal{L}^2(\partial E) > 0$, where $\mathcal{L}^2$ is the 2-dimensional Lebesgue measure and ∂E is the boundary of E in C. To prove this claim we first need to show an estimate for $\mathcal{L}^2(E)$. At the *nth* stage of construction of E we note that at most 8^{n-1} squares are created, each of side length 9^{-n} (and hence area 9^{-2n}). Thus

$$\mathcal{L}^2(E) \leq \sum_{n=1}^{\infty} 8^{n-1} 9^{-2n}.$$

Since E is dense in C we have $\partial E = C \setminus E$ and therefore

$$\mathcal{L}^2(\partial E) = \mathcal{L}^2(C \setminus E) = \mathcal{L}^2(C) - \mathcal{L}^2(E) \geq 1 - \sum_{n=1}^{\infty} 8^{n-1} 9^{-2n} > 0.$$

An immediate consequence of this result is that $\mathcal{H}^1(\partial E) = \infty$, since $\mathcal{L}^2(\partial E) > 0$ implies $\mathcal{H}^2(\partial E) > 0$.

Our final claim is that $\|\partial E\|(\mathbb{R}^2) < \infty$. This will follow easily if, for any square $Q \subset C$ we have $\|\partial Q\|(\mathbb{R}^2) = \mathcal{H}^1(\partial Q) = 4s$, where s is the side length of Q. This holds true since from Example 3.2, we can certainly approximate a square by a set with smooth boundary (think of a square that is rounded at the vertices), whose perimeter is arbitrarily close to the perimeter of the square. By the construction of E and since the $Q^n_{j,k}$ are all disjoint,

$$\chi_E(x) = \lim_{N \to \infty} \sum_{n=1}^{N} \sum_{j,k=1}^{2^n - 1} \chi_{Q^n_{j,k}}(x).$$

Then by Theorem 3.2 we have

$$
\begin{aligned}
\|\partial E\|(\mathbb{R}^2) &\le \liminf_{N\to\infty} \sum_{n=1}^{N} \sum_{j,k=1}^{2^n-1} \left\|\partial Q_{jk}^n\right\|(\mathbb{R}^2) \\
&\le \sum_{n=1}^{\infty} \sum_{j,k=1}^{2^n-1} \left\|\partial Q_{jk}^n\right\|(\mathbb{R}^2) \\
&= \sum_{n=1}^{\infty} \sum_{j,k=1}^{2^n-1} \mathcal{H}^1(\partial Q_{jk}^n) \\
&\le \sum_{n=1}^{\infty} 4(8)^{n-1} 9^{-n} \\
&< \infty.
\end{aligned}
$$

4 Image Processing: The ROF model

4.1 The Optimal Solution to the ROF Problem

In this section the mathematical properties of the ROF model will be explored and the optimal decomposition of f into $u + v$ will be studied. We begin with the splitting algorithm.

Definition 4.1 *For a positive parameter λ and a given function $f \in L^2(\mathbb{R}^2)$, the ROF model selects the decomposition $f = u + v$ that solves the variational problem given by*

$$\inf\{J(u); f = u + v\}, \tag{4.1}$$

where

$$J(u) = \|u\|_{BV} + \lambda \|v\|_2^2. \tag{4.2}$$

Remark. With an obvious abuse of notation, in this section and all of those to follow, the term $\|f\|_{BV}$ is used to denote the semi-norm $\|Df\|(\mathbb{R}^2)$ given in Definition 3.1. Although we obtain from this semi-norm the homogeneous version of the space BV, denoted $\dot{BV}$, which is a function space modulo constants, we a priori have that the u component not only belongs to $\dot{BV}$, but to $L^2(\mathbb{R}^2)$ as well since $u = f - v$ and $f, v \in L^2(\mathbb{R}^2)$. Thus the only constant function allowed is $f = 0$ and hence $\|\cdot\|_{BV}$ is in fact a norm.

Theorem 4.1 *There exists a unique solution (u,v) to the minimization problem given in Definition 4.1*

Proof. Here we follow the proof of existence given in [14]. Let $\{u_j\} \subset BV$ be a minimizing sequence of the functional $J(u)$, i.e. $J(u_j) \to \alpha = \inf J(u)$. Also for each j, let $v_j = f - u_j$ and fix $\epsilon > 0$. Then we have

$$\begin{aligned} J(\frac{u_j + u_k}{2}) &= \left\|\frac{u_j + u_k}{2}\right\|_{BV} + \lambda \left\|f - \frac{u_j + u_k}{2}\right\|_2^2 \\ &\leq \frac{1}{2}\|u_j\|_{BV} + \frac{1}{2}\|u_k\|_{BV} + \lambda \left\|\frac{v_j}{2} + \frac{v_k}{2}\right\|_2^2 \\ &= \frac{1}{2}\|u_j\|_{BV} + \frac{1}{2}\|u_k\|_{BV} + \lambda(\frac{1}{2}\|v_j\|_2^2 + \frac{1}{2}\|v_k\|_2^2 - \frac{1}{4}\|v_j - v_k\|_2^2) \\ &= \frac{1}{2}(J(u_j) + J(u_k)) - \frac{\lambda}{4}\|v_j - v_k\|_2^2. \end{aligned}$$

For j and k sufficiently large we then have

$$J(\frac{u_j + u_k}{2}) \leq \alpha + \epsilon - \frac{\lambda}{4}\|v_j - v_k\|_2^2$$

and since $\alpha \leq J(\frac{u_j+u_k}{2})$ it follows that

$$\|u_j - u_k\|_2 = \|v_j - v_k\|_2 \leq 2\sqrt{\frac{\epsilon}{\lambda}}.$$

Hence u_j converges strongly to some $\tilde{u}$ in L^2. Then for any $\vec{\varphi} \in C_c^1(\mathbb{R}^2; \mathbb{R}^2)$,

$$\begin{aligned} \left|\int_{\mathbb{R}^2} (u_j - \tilde{u})\mathrm{div}\vec{\varphi}\right| &\leq \|u_j - \tilde{u}\|_2 \|\mathrm{div}\vec{\varphi}\|_2 \\ &= C\|u_j - \tilde{u}\|_2 \to 0 \end{aligned}$$

as $j \to 0$. Thus L^2 convergence implies weak-$*$ convergence and by the lower semicontinuity property (Theorem 3.2) we have

$$\|\tilde{u}\|_{BV} \leq \liminf_{j\to\infty} \|u_j\|_{BV}.$$

Hence

$$\alpha = \inf J(u) \leq J(\tilde{u}) \leq \liminf_{j\to\infty} J(u_j) = \alpha,$$

implying that $\alpha = J(\tilde{u})$ so that $\tilde{u}$ is a minimizer of the functional J.

We prove uniqueness by contradiction. Suppose $\inf J(u) = J(u_1) = J(u_2)$ where $u_1 \neq u_2$. Then since $\|\cdot\|_{BV}$ and $\|\cdot\|_2^2$ are convex and strictly convex functionals respectively, we have

$$\begin{aligned}
J(\frac{u_1+u_2}{2}) &= \left\|\frac{u_1}{2}+\frac{u_2}{2}\right\|_{BV} + \lambda \left\|f-(\frac{u_1}{2}+\frac{u_2}{2})\right\|_2^2 \\
&= \left\|\frac{u_1}{2}+\frac{u_2}{2}\right\|_{BV} + \lambda \left\|\frac{f-u_1}{2}+\frac{f-u_2}{2}\right\|_2^2 \\
&< \frac{1}{2}(\|u_1\|_{BV} + \lambda\|f-u_1\|_2^2) + \frac{1}{2}(\|u_2\|_{BV} + \lambda\|f-u_2\|_2^2) \\
&= \frac{1}{2}J(u_1) + \frac{1}{2}J(u_2) \\
&= \inf J(u)
\end{aligned}$$

which is a contradiction.

The most commonly used method for solving the minimization problem given in Definition 4.1 is done by a PDE approach. Solving the ROF problem is equivalent to finding a solution to the Euler equation:

$$u = f + \frac{1}{2\lambda}\mathrm{div}(\frac{\vec{\nabla} u}{|\vec{\nabla} u|}).$$

In practice, the optimal u is obtained by a sequence $u_n \in \dot{W}^{1,1}$ and the term $|\vec{\nabla} u|$ is replaced by $|\vec{\nabla} u| + \beta^2$ for some small constant β in order to avoid division by zero. For more information on this approach and how the optimal solution is obtained, the reader is referred to [1]. An example of an image decomposed in this way is given in Figure 1 below (images taken from [3]).

For analyzing the properties of the optimal pair (u, v) which minimizes the ROF functional from a mathematical standpoint, Meyer introduced an important function space in [13] which models the textured components of the ROF model. The definition of this space now follows.

Definition 4.2 *Let G be the space of all tempered distributions $f \in \mathcal{S}'(\mathbb{R}^2)$ which can be written as*

$$f(x_1, x_2) = \partial_{x_1} g_1(x_1, x_2) + \partial_{x_2} g_2(x_1, x_2),$$

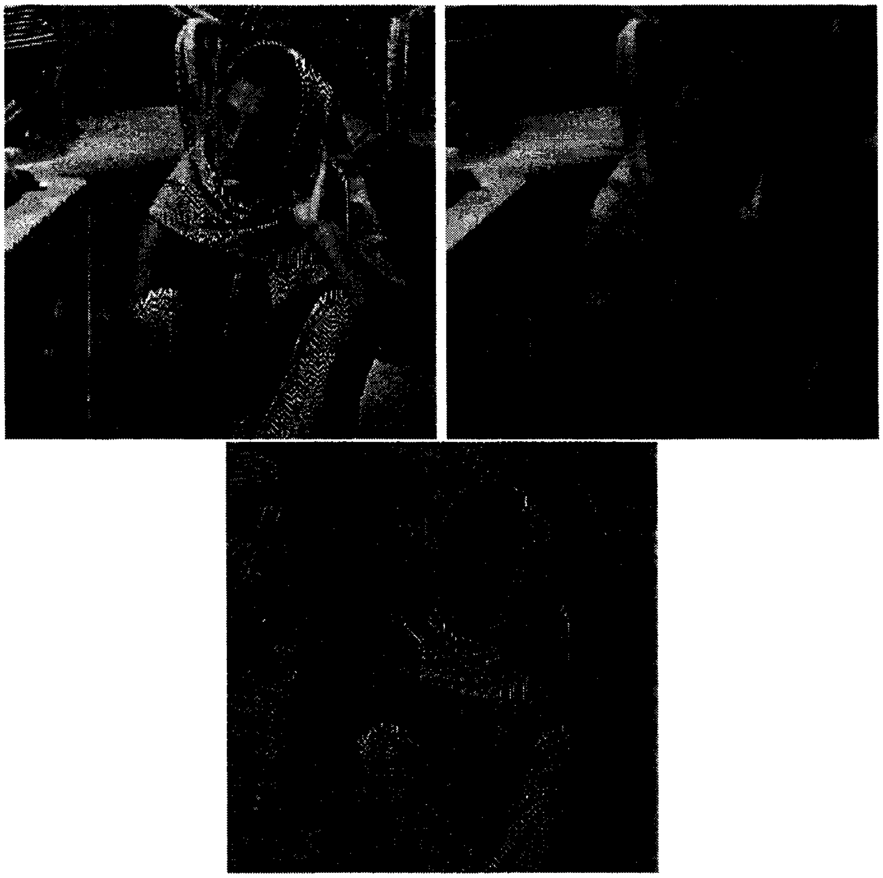

Figure 1: The ROF decomposition of an image with parameter $\lambda = 43$; (top left) original image $f(x)$; (top right) $u(x)$; (bottom) $v(x) + 150$

where $g_1, g_2 \in L^\infty(\mathbb{R}^2)$. *The associated norm in* G, *which we denote by* $\|\cdot\|_*$, *is given by*

$$\|f\|_* = \inf \left\| \sqrt{{g_1}^2 + {g_2}^2} \right\|_\infty ,$$

where the infimum is taken over all possible representations of f *as* $\mathrm{div}\vec{g}$ *where* $\vec{g} = (g_1, g_2)$.

The definition of G leads to the following lemma.

Lemma 4.1 *The space* G *is the dual space of the homogeneous Sobolev space* $\dot{W}^{1,1}(\mathbb{R}^2)$.

Proof. First we show that any $f \in G$ is a bounded linear functional on $\dot{W}^{1,1}$. Given $f \in G$ and any $\varphi \in C_c^\infty(\mathbb{R}^2)$,

$$\int f(x)\varphi(x)dx = -\int \vec{g}(x) \cdot \vec{\nabla}\varphi(x)dx,$$

where $\vec{g} \in L^\infty(\mathbb{R}^2;\mathbb{R}^2)$. Then

$$\begin{aligned}
|\int f(x)\varphi(x)dx| &\leq \int |\vec{g}||\vec{\nabla}\varphi|dx \\
&\leq \left\|\sqrt{g_1^2+g_2^2}\right\|_\infty \left\|\sqrt{(\frac{\partial\varphi}{\partial x_1})^2+(\frac{\partial\varphi}{\partial x_2})^2}\right\|_1 \\
&\approx \left\|\sqrt{g_1^2+g_2^2}\right\|_\infty \|\varphi\|_{\dot{W}^{1,1}}.
\end{aligned}$$

Since $C_c^\infty(\mathbb{R}^2)$ is dense in $\dot{W}^{1,1}(\mathbb{R}^2)$, it is obvious that the above inequality holds for $\varphi \in \dot{W}^{1,1}(\mathbb{R}^2)$.

For the converse, we need to show that if we are given $\Lambda \in (\dot{W}^{1,1})^*$, then for any $\varphi \in \dot{W}^{1,1}$, we can write

$$\Lambda\varphi = \int f\varphi,$$

where $f \in G$. For this purpose, consider the subspace of $L^1(\mathbb{R}^2;\mathbb{R}^2)$ given by

$$M = \{\vec{\psi} \in C_c^\infty(\mathbb{R}^2;\mathbb{R}^2) : \ \vec{\nabla} \times \vec{\psi} = 0\}.$$

Then for any $\vec{\psi} \in M$ there exists a unique function $\varphi \in C_c^\infty(\mathbb{R}^2)$ such that $\vec{\nabla}\varphi = \vec{\psi}$ (uniqueness is achieved by setting $\varphi = 0$ on $\mathbb{R}^2\backslash K$ where K is the compact set in which $\vec{\psi}$ is supported). Then we can define $\tilde{\Lambda}$ to be the linear functional on M given by

$$\tilde{\Lambda}\vec{\psi} = \Lambda\varphi.$$

Moreover, $\tilde{\Lambda}$ is a bounded linear functional on M since

$$\begin{aligned}
|\tilde{\Lambda}\vec{\psi}| = |\Lambda\varphi| &\leq C\|\varphi\|_{\dot{W}^{1,1}} \\
&\approx \|\partial_{x_1}\varphi\|_1 + \|\partial_{x_2}\varphi\|_1 \\
&\approx \left\||\vec{\psi}|\right\|_1.
\end{aligned}$$

The Hahn-Banach Theorem allows us to extend $\tilde{\Lambda}$ from M to all of $L^1(\mathbb{R}^2;\mathbb{R}^2)$. Thus there is a unique $\vec{g} \in L^\infty(\mathbb{R}^2;\mathbb{R}^2)$ such that

$$\tilde{\Lambda}\vec{\psi} = \int \vec{g}\cdot\vec{\psi}$$

for all $\vec{\psi} \in L^1(\mathbb{R}^2;\mathbb{R}^2)$. Then for any $\varphi \in \dot{W}^{1,1}(\mathbb{R}^2)$ we have

$$\begin{aligned} \Lambda\varphi = \tilde{\Lambda}(\vec{\nabla}\varphi) &= \int \vec{g}\cdot\vec{\nabla}\varphi \\ &= -\int f\varphi, \end{aligned}$$

where $f = \mathrm{div}\vec{g}$. Thus $f \in G$ and the proof is complete.

This tells us that the space G is "almost" the dual space of BV. Indeed, from Theorem 3.6, the closure in $\|\cdot\|_{BV}$ of the Schwartz class of functions $\mathcal{S}(\mathbb{R}^2)$ is the space $\dot{W}^{1,1}$. Also, it should be noted that Theorem 3.4 allows the approximation of $\|\cdot\|_{BV}$ by $\|\cdot\|_{\dot{W}^{1,1}}$, and in many instances this is sufficient. Another useful lemma is the following.

Lemma 4.2 *For any $f, g \in L^2(\mathbb{R}^2)$,*

$$\left|\int f(x)g(x)dx\right| \leq \|f\|_{BV}\,\|g\|_* . \tag{4.3}$$

Proof. It is obvious that the lemma holds for $f \in \dot{W}^{1,1}$ since $G = (\dot{W}^{1,1})^*$ and $\|f\|_{BV} = \|f\|_{\dot{W}^{1,1}}$. Then for $f \in BV$, by Theorems 3.4 and 3.5, we can approximate f by functions $f_m \in BV \cap C^\infty \subset \dot{W}^{1,1}$ such that $\|f_m\|_{BV} \to \|f\|_{BV}$ and $\int f_m \mathrm{div}\vec{\varphi} \to \int f\mathrm{div}\vec{\varphi}$. Then for each m

$$\left|\int f_m g\right| \leq \|f_m\|_{BV}\,\|g\|_* ,$$

and since $\|f_m\|_{BV} \to \|f\|_{BV}$ we need only show that $\int f_m g \to \int fg$. Since $g \in L^2(\mathbb{R}^2)$ we can find a sequence $\vec{\varphi}_k \in C_c^1(\mathbb{R}^2;\mathbb{R}^2)$ such that $\|\mathrm{div}\vec{\varphi}_k - g\|_2 \to 0$ as $k \to \infty$. To see this, we consider the solution u of Poisson's equation $\Delta u = g$. By regularity we can take $u \in W^{2,2}(\mathbb{R}^2)$, so there exists a sequence $\psi_k \subset C_c^\infty(\mathbb{R}^2)$ such that $\psi_k \to u$ in $W^{2,2}$. Let $\vec{\varphi}_k = \vec{\nabla}\psi_k$. Then $\vec{\varphi}_k \in C_c^\infty(\mathbb{R}^2;\mathbb{R}^2)$ and $\mathrm{div}\vec{\varphi}_k = \mathrm{div}\vec{\nabla}\psi_k = \Delta\psi_k \to \Delta u = g$ in L^2. Using these approximations for f and g and fixing $\epsilon > 0$ we have

$$\begin{aligned} \left|\int (f_m g - fg)\right| &= \left|\int (f_m g - f_m\mathrm{div}\vec{\varphi}_k + f_m\mathrm{div}\vec{\varphi}_k - f\mathrm{div}\vec{\varphi}_k + f\mathrm{div}\vec{\varphi}_k - fg)\right| \\ &\leq \left|\int f_m(g - \mathrm{div}\vec{\varphi}_k)\right| + \left|\int (f_m - f)\mathrm{div}\vec{\varphi}_k\right| + \left|\int f(\mathrm{div}\vec{\varphi}_k - g)\right| \\ &\leq (\|f_m\|_2 + \|f\|_2)\,\|\mathrm{div}\vec{\varphi}_k - g\|_2 + \left|\int (f_m - f)\mathrm{div}\vec{\varphi}_k\right| \\ &\leq C(\|f_m\|_{BV} + \|f\|_{BV})\,\|\mathrm{div}\vec{\varphi}_k - g\|_2 + \left|\int (f_m - f)\mathrm{div}\vec{\varphi}_k\right|, \end{aligned}$$

where the constant C is given by the embedding theorem for BV (Theorem 3.7). For some m_0 we have $\|f_m\|_{BV} \leq 2\|f\|_{BV}$ for all $m \geq m_0$. Also there is some k_0 so that $\|\text{div}\vec{\varphi}_k - g\|_2 < \epsilon/6C\,\|f\|_{BV}$ for all $k \geq k_0$. Then for any fixed k larger than k_0, by Theorem 3.5, there is an m_1 such that

$$\left|\int (f_m - f)\text{div}\vec{\varphi}_k\right| < \frac{\epsilon}{2}$$

whenever $m \geq m_1$. Finally for $m \geq M = max\{m_0, m_1\}$ and $k \geq k_0$,

$$\left|\int (f_m g - fg)\right| < 3C\,\|f\|_{BV} \cdot \frac{\epsilon}{6C\,\|f\|_{BV}} + \frac{\epsilon}{2} = \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

The optimal solution (u, v) to the ROF model can now be characterized by the following two theorems, both of which were proved in [13]. The first of these theorems tells us that an image with a small enough norm in G is considered too small to be an object and, consequently, is recognized as a texture. Letting $\lambda > 0$ be the free parameter given in the ROF decomposition, the theorem formally reads as follows.

Theorem 4.2 *If f is such that $\|f\|_* \leq \frac{1}{2\lambda}$ then the ROF decomposition of f is given by $u = 0$ and $v = f$.*

Proof. The ROF model gives $u = 0$ iff for any $h \in BV$ we have

$$\|h\|_{BV} + \lambda\,\|f - h\|_2^2 \geq \lambda\,\|f\|_2^2. \tag{4.4}$$

Expanding the squared L^2 norm on the left we have

$$\begin{aligned}
\|h\|_{BV} + \lambda\,\|f - h\|_2^2 &= \|h\|_{BV} + \lambda \int (f - h)^2 \\
&= \|h\|_{BV} + \lambda \int (f^2 - 2fh + h^2) \\
&= \|h\|_{BV} + \lambda(\|f\|_2^2 + \|h\|_2^2) - 2\lambda \int fh.
\end{aligned}$$

Hence (4.4) is equivalent to

$$\|h\|_{BV} + \lambda\,\|h\|_2^2 \geq 2\lambda \int fh. \tag{4.5}$$

Fixing $\epsilon > 0$ and replacing h by ϵh we have

$$\epsilon \|h\|_{BV} + \lambda\epsilon^2 \|h\|_2^2 \geq 2\lambda\epsilon \int fh. \tag{4.6}$$

Dividing by ϵ and letting $\epsilon \to 0$, we have

$$|\int f(x)h(x)dx| \leq \frac{1}{2\lambda} \|h\|_{BV}, \tag{4.7}$$

which is also equivalent to (4.4). By taking $h \in \dot{W}^{1,1}$, we see that (4.7) implies $\|f\|_* \leq \frac{1}{2\lambda}$. For the converse, if $\|f\|_* \leq \frac{1}{2\lambda}$, then since we are assuming $f \in L^2(\mathbb{R}^2)$, Lemma 4.2 shows that (4.7) is satisfied, which completes the proof.

Theorem 4.3 *Let f, u, and v belong to $L^2(\mathbb{R}^2)$. If $\|f\|_* > \frac{1}{2\lambda}$ then the optimal ROF decomposition $f = u + v$ satisfies*

$$\|v\|_* = \frac{1}{2\lambda}$$

and

$$\int u(x)v(x)dx = \frac{1}{2\lambda} \|u\|_{BV}.$$

Proof. First we have (u, v) as the solution if and only if for any $h \in BV$ and any scalar $\epsilon \in \mathbb{R}$,

$$\|u + \epsilon h\|_{BV} + \lambda \|v - \epsilon h\|_2^2 \geq \|u\|_{BV} + \lambda \|v\|_2^2. \tag{4.8}$$

Then we have

$$|\epsilon| \|h\|_{BV} + \lambda \|v - \epsilon h\|_2^2 \geq \lambda \|v\|_2^2.$$

Hence we can proceed as in the proof of Theorem 4.2 to obtain

$$\|v\|_* \leq \frac{1}{2\lambda}. \tag{4.9}$$

Lemma 4.2 then tells us that

$$\int u(x)v(x)dx \leq \frac{1}{2\lambda} \|u\|_{BV}. \tag{4.10}$$

To prove inequality in the other direction take $h = u$ and $-1 < \epsilon < 0$ in (4.8). Then after expanding the squared L^2 norm on the left we have,

$$|1 + \epsilon| \|u\|_{BV} + \lambda \|v\|_2^2 + \epsilon^2\lambda \|u\|_2^2 - 2\epsilon\lambda \int u(x)v(x)dx \geq \|u\|_{BV} + \lambda \|v\|_2^2.$$

Noting that $|1+\epsilon| = 1+\epsilon$, we can cancel out terms from each side of the above inequality. Then dividing both sides by ϵ we obtain

$$\|u\|_{BV} + \lambda\epsilon \|u\|_2^2 \leq 2\lambda \int u(x)v(x)dx.$$

Taking $\epsilon \longrightarrow 0^-$ and combining with (4.10), the desired result

$$\int u(x)v(x)dx = \frac{1}{2\lambda} \|u\|_{BV}, \tag{4.11}$$

is obtained. But (4.9) implies that $\|v\|_* = 1/2\lambda$ since otherwise equality would not be reached in (4.11). Conversely, given the pair (u, v) satisfying

$$\|v\|_* = \frac{1}{2\lambda} \qquad \text{and} \qquad \int u(x)v(x)dx = \frac{1}{2\lambda} \|u\|_{BV},$$

then for any $h \in BV(\mathbb{R}^2)$ and $\epsilon \in \mathbb{R}$,

$$\begin{aligned}
\|u+\epsilon h\|_{BV} + \lambda \|v-\epsilon h\|_2^2 &\geq 2\lambda \int (u(x)+\epsilon h(x))v(x)dx + \lambda \|v\|_2^2 \\
&\quad - 2\lambda \int v(x)h(x)dx + \lambda\epsilon^2 \|h\|_2^2 \\
&= 2\lambda \int u(x)v(x)dx + \lambda \|v\|_2^2 + \lambda\epsilon^2 \|h\|_2^2 \\
&= \|u\|_{BV} + \lambda \|v\|_2^2 + \lambda\epsilon^2 \|h\|_2^2 \\
&\geq \|u\|_{BV} + \lambda \|v\|_2^2.
\end{aligned}$$

The first inequality follows from $\|v\|_* = 1/2\lambda$ and Lemma 4.2, since v is assumed to belong to $L^2(\mathbb{R}^2)$. Hence Theorem 4.3 is now proved.

Example 4.1

The ROF model suffers from at least one drawback as we will now see. Let $f = \chi_B$ where $B = B_0(R)$ is the ball of radius R centered at 0. We would hope to get $u = \chi_B$ and $v = 0$ but that will not be the case. We first show, as in [13], that $\|\chi_B\|_* = R/2$. From Lemma 4.2 and Example 3.2,

$$\pi R^2 = \int \chi_B(x)dx = \int \chi_B(x)\chi_B(x)dx \leq \|\chi_B\|_{BV} \|\chi_B\|_* = 2\pi R \|\chi_B\|_*.$$

Thus $\|\chi_B\|_* \geq R/2$. To prove inequality in the other direction consider the vector $\vec{g} = (g_1, g_2)$ given by $g_i(x) = x_i\omega(|x|)$, where

$$\omega(r) = \begin{cases} 1/2 & \text{if } 0 \leq r \leq R \\ \\ R^2/2r^2 & \text{if } r > R. \end{cases}$$

Then one can easily compute that $\text{div}\vec{g} = \chi_B$. Furthermore, for $0 \leq r \leq R$,

$$|\vec{g}| = \sqrt{g_1^2 + g_2^2} = r/2 \leq R/2,$$

and for $r > R$,

$$|\vec{g}| = \sqrt{\frac{(x_1^2 + x_2^2)R^4}{4(x_1^2 + x_2^2)^2}} = \frac{R^2}{2r} < \frac{R}{2}.$$

Hence $\||\vec{g}|\|_\infty \leq R/2$ which implies $\|\chi_B\|_* \leq R/2$. Next we let $f = c\chi_B$ for some constant $c > 0$. Then if $c \leq 1/R\lambda$ the ROF decomposition of f is given as

$$u = 0, \quad v = c\chi_B. \tag{4.12}$$

If $c \geq 1/R\lambda$ then the ROF algorithm yields

$$u = (c - (\lambda R)^{-1})\chi_B, \quad v = (\lambda R)^{-1}\chi_B. \tag{4.13}$$

The first case is easily proved since for $c \leq 1/R\lambda$, $\|c\chi_B\|_* = cR/2 \leq 1/2\lambda$ and Theorem 4.2 tells us that the decomposition given by (4.12) is the correct one. As for the second case we see that

$$\|v\|_* = \|(\lambda R)^{-1}\chi_B\|_* = \frac{1}{\lambda R}\frac{R}{2} = \frac{1}{2\lambda}$$

and

$$\begin{aligned} \int u(x)v(x)dx &= \int (\frac{c}{\lambda R} - \frac{1}{(\lambda R)^2})\chi_B(x)dx \\ &= \frac{c\pi R}{\lambda} - \frac{\pi}{\lambda^2} \\ &= (c - \frac{1}{\lambda R})2\pi R \cdot \frac{1}{2\lambda} \\ &= \|u\|_{BV}\,\|v\|_*. \end{aligned}$$

Hence, since $u + v = f$, it follows by Theorem 4.3 that the optimal decomposition is as given in (4.13).

4.2 Textures and the "v" Component

We would now like to discuss the v component of the ROF model which is supposed to be modelling textures and noise. As far as noise is concerned, a statistical analysis is required which we will not discuss here. Our concern is whether the v component is actually modelling textures. Textures are referred to in [16] as "small scale repeated details". It is in this regard that oscillating patterns are a good model for representing textures.

Consider the following example. Suppose we are given a function

$$f(x) = \cos(N \cdot x)\chi_E(x)$$

which is modelling the textured patterns of an image defined on some bounded set E. Here $N = (N_1, N_2)$ and we assume that $|N_1|$ and $|N_2|$ are large. Then we can write

$$f(x) = \mathrm{div}\vec{g},$$

where $\vec{g} = (g_1, g_2)$ with

$$g_i = \frac{1}{2N_i}\sin(N \cdot x)\chi_E.$$

Then we have

$$\| |\vec{g}| \|_\infty = \frac{1}{2}\sqrt{\frac{1}{N_1^2} + \frac{1}{N_2^2}}$$

and consequently

$$\|f\|_* \leq \frac{1}{2}\sqrt{\frac{1}{N_1^2} + \frac{1}{N_2^2}},$$

where $\|\cdot\|_*$ is the norm in the space G. As a result of Theorem 4.2, the ROF algorithm applied to f alone gives $u = 0$ and $v = f$ (provided $|N_1|, |N_2|$ are sufficiently large). The issue of whether the ROF model will be able to separate a sum $h + f$ (where f is as above and h is a simple object such as a disk), yielding $u = h$ and $v = f$ will be discussed shortly.

The example of a texture provided above falls under a class of more general functions which are known to have a small norm in G. The following definition will be useful in describing such functions.

Definition 4.3 *A positive measure μ is called a Guy David measure if there exists a constant $C > 0$ such that for any ball B_R of radius R,*

$$\mu(B_R) \leq CR.$$

Recalling the measure $\|Df\|$ given in section 3 for $f \in BV$, Meyer proved the following theorem [13].

Theorem 4.4 *Let $f \in L^\infty(\mathbb{R}^2)$ and let $\|Df\|$ be a Guy David measure. Then there exists a constant C_0 such that*

$$\|f_\omega\|_* \leq \frac{C_0}{|\omega|},$$

where $f_\omega(x) = e^{i\omega \cdot x} f(x)$.

We now wish to understand more precisely how the ROF algorithm splits an image f into a sum $u + v$. Suppose we a priori have $f = g + h$ where g is some simple object and h is a texture satisfying $\|h\|_* < \epsilon$ for some small ϵ. We would hope that the ROF model would yield $u = g$ and $v = h$. However, this seems hardly possible since Theorems 4.2 and 4.3 tell us that the ROF model applied to g alone gives $u = \tilde{g}$ and $v = g - \tilde{g}$ for some new function $\tilde{g}$. As suggested in [14], the best we can hope for is that $f = g + h$ be decomposed as $u = \tilde{g}$ and $v = g - \tilde{g} + h$. Although, in general, this is not the case, the L^2-error between u and $\tilde{g}$ is small whenever $\|h\|_*$ is small. Meyer proved the following.

Theorem 4.5 *Let $f_1, f_2 \in L^2(\mathbb{R}^2)$ and let $f_j = u_j + v_j$ be the ROF decomposition of f_j, $j \in \{1, 2\}$ with λ the chosen value of the free parameter. Then we have*

$$\|u_2 - u_1\|_2 \leq 13(\|f_1\|_2 + \|f_2\|_2)\sqrt{\lambda \|f_2 - f_1\|_*}.$$

This theorem was proved in [14] in a more general setting on Hilbert spaces. We see that when $f_1 = g + h$ with $\|h\|_* < \epsilon$ and $f_2 = g$ then

$$\|u_2 - u_1\|_2 \leq 13(\|f_1\|_2 + \|f_2\|_2)\sqrt{\lambda\epsilon}.$$

5 Orthonormal Wavelets and Multiresolution Analysis

Wavelet analysis has become a very popular field of mathematics over the last 20 or so years and has very strong ties to Fourier analysis, including roots in time-frequency analysis and digital image processing. Although there are several types of wavelets, the focus of this section will be on orthonormal wavelets, which we will eventually see, give alternative methods of solving minimization problems such as the ROF problem given by (4.1). A short introduction to wavelets in a more general setting, however, will be given first as motivation. A common approach to explaining wavelets often begins with a discussion of time-frequency analysis. This leads us to define the Fourier transform of a function.

Definition 5.1 *Given any function $f \in L^1(\mathbb{R})$, we define its Fourier transform, denoted $\mathcal{F}f$ or $\hat{f}$, by*

$$\hat{f}(\xi) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{-ix\xi} dx.$$

The definition can then be extended to any $f \in L^2(\mathbb{R})$ by defining the integral above as a limit

$$\hat{f}(\xi) = \lim_{N \to \infty} \frac{1}{\sqrt{2\pi}} \int_{-N}^{N} f(x) e^{-ix\xi} dx.$$

If we think of the function f as a signal evolving in time (where x represents time), then the Fourier transform gives a representation of the "frequency content" of f (where ξ represents frequency). Viewing a signal from the frequency domain can often give information that is not present in the time domain. However, if we are interested in analyzing some local properties of f from $\hat{f}$, the Fourier transform can pose a problem. This is due to the fact that for any fixed ξ, the function $e^{-ix\xi}$ is supported globally as a function of x and as a result all $x \in \mathbb{R}$ are needed to compute $\hat{f}(\xi)$ from $f(x)$. One way of fixing this problem is by first multiplying f by some function g which is well localized in time (usually having compact support) and relatively smooth, and then taking its

Fourier transform:

$$(Wf)(\xi,t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f(x)g(x-t)e^{-ix\xi}dx.$$

The mapping W is called a windowed Fourier transform (WFT) and the function g is referred to as a window function. We can now obtain the frequency information of f locally in the support of $g(\cdot - t)$ and using different values of t will allow us to shift the window over $\mathbb{R}$. In this way, "time-localization" is being achieved.

Wavelets offer a different method for achieving "time-localization". The Continuous Wavelet Transform (CWT) of a function $f \in L^2(\mathbb{R})$ is defined as

$$W_{(a,b)}(f) = \frac{1}{\sqrt{|a|}}\int_{-\infty}^{\infty} f(x)\psi\left(\frac{x-b}{a}\right)dx,$$

where $a \in \mathbb{R}\backslash\{0\}$ and $b \in \mathbb{R}$. We are assuming ψ is real valued, otherwise ψ would be replaced by its complex conjugate ψ^* when defining the CWT. The terms $W_{(a,b)}(f)$ are called the continuous coefficients of f. The requirement on ψ is that it must satisfy the admissibility condition:

$$C_\psi = 2\pi\int_{-\infty}^{\infty} |\hat{\psi}(\xi)|^2|\xi|^{-1}d\xi < \infty, \tag{5.1}$$

which implies that $\hat{\psi}(0) = 0$ (i.e. $\int \psi(x)dx = 0$) when $\psi \in L^1(\mathbb{R})$ since $\hat{\psi}$ would be continuous. Any ψ satisfying (5.1) is called a "mother wavelet". Other properties of ψ, such as fast decay (compact support is most preferable) and some sort of smoothness conditions, are usually imposed as well. Of course, any $\psi \in \mathcal{S}(\mathbb{R})$ such that $\hat{\psi}(0) = 0$ will satisfy (5.1). The term "wavelet" comes from the fact that ψ is necessarily oscillating, resembling a wave, and also is well localized, hence a small wave. A well known example of a mother wavelet is the Mexican Hat function,

$$\psi(x) = \frac{2}{\sqrt{3}}\pi^{-1/4}(1-x^2)e^{-x^2/2},$$

which is a normalized second derivative of the Gaussian function. This function is also known as the second Hermite function.

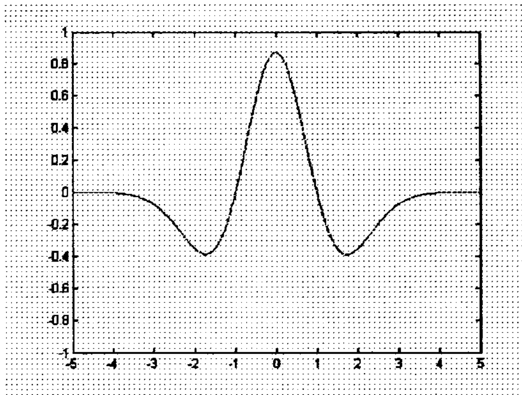

Figure 2: The Mexican Hat function (image courtesy of [19])

Any $f \in L^2(\mathbb{R})$ can be reconstructed by using its CWT in the following formula:

$$f = C_\psi^{-1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \langle f, \psi_{a,b} \rangle \psi_{a,b} \frac{dadb}{a^2}, \tag{5.2}$$

where here we have the notation $\psi_{a,b}(\cdot) = |a|^{-1/2}\psi(\frac{\cdot - b}{a})$ and $\langle \; , \; \rangle$ denotes the L^2 inner product. Equation (5.2) is called Calderón's reproducing identity and is also sometimes referred to as the "resolution of the identity" formula. A proof can be found in [8].

We would like to compare the WFT and the CWT. Both the CWT and WFT are similar in the fact that they are achieving time-localization, however, the CWT has the added benefit that the window sizes can be changed as we change the values of the parameter a. This property is often referred to as a "zooming" effect. A prime example of its usefulness is for analyzing singularities; for small values of a, the $\psi_{a,b}$ are "zooming" in and therefore can detect singularities much better. On the other hand, the window given in the WFT has the width of its support fixed and therefore does not have this "zooming" capability.

The formula given by (5.2) naturally raises the question of whether f can be reconstructed if only a discrete number of the $W_{(a,b)}(f)$ are known. Although there is not a formula analagous to (5.2) for the discrete case, it is possible to reconstruct f using the Discrete Wavelet Transform (DWT),

$$W_{j,k}(f) = a_0^{j/2} \int_{-\infty}^{\infty} f(x)\psi(a_0^j x - kb_0)dx,$$

under certain conditions. Here $a_0 > 1$ and $b_0 > 0$ are fixed, while $j, k \in \mathbb{Z}$. We now denote $\psi_{j,k}$ as

$$\psi_{j,k}(x) = a_0^{j/2}\psi(a_0^j x - kb_0).$$

The condition that the $\psi_{j,k}$ should satisfy in order to allow reconstruction of a function $f \in L^2(\mathbb{R})$ is that there must exist constants A and B, with $0 < A \leq B < \infty$, such that

$$A \|f\|_2^2 \leq \sum_{j,k\in\mathbb{Z}} |\langle f, \psi_{j,k}\rangle|^2 \leq B \|f\|_2^2. \tag{5.3}$$

A function ψ satisfying (5.3) for some fixed a_0, b_0 is said to generate a *frame*. Moreover, when $A = B$, ψ is said to generate a *tight frame*. When $A = B = 1$, (5.3) reduces to Parseval's identity and the $\{\psi_{j,k}\}$ are necessarily orthogonal. For the sake of simplicity, we will assume that the functions being considered take only real values to avoid using complex conjugates in the inner product above. To see that the condition given by (5.3) is sufficient for the reconstruction of a function $f \in L^2(\mathbb{R})$, consider the linear mapping T on $L^2(\mathbb{R})$ defined by

$$Tf = \sum_{j,k\in\mathbb{Z}} \langle f, \psi_{j,k}\rangle \psi_{j,k}, \qquad f \in L^2(\mathbb{R}).$$

We can see that T is bounded since, for any $f \in L^2(\mathbb{R})$,

$$\begin{aligned}
\|Tf\|_2 &= \sup_{\|g\|_2\leq 1} \langle Tf, g\rangle \\
&= \sup_{\|g\|_2\leq 1} \sum_{j,k\in\mathbb{Z}} \langle f, \psi_{j,k}\rangle\langle \psi_{j,k}, g\rangle \\
&\leq \sup_{\|g\|_2\leq 1} (\sum_{j,k\in\mathbb{Z}} |\langle f, \psi_{j,k}\rangle|^2)^{1/2} (\sum_{j,k\in\mathbb{Z}} |\langle \psi_{j,k}, g\rangle|^2)^{1/2} \\
&\leq \sup_{\|g\|_2\leq 1} B \|f\|_2 \|g\|_2 = B \|f\|
\end{aligned}$$

We can also see T is a one-to-one mapping since

$$\langle Tf, f\rangle = \sum_{j,k\in\mathbb{Z}} |\langle f, \psi_{j,k}\rangle|^2$$

and the lower bound in (5.3) implies that the kernel of T, ker $T = \{0\}$. The range of T, $R(T)$, is closed in $L^2(\mathbb{R})$. To see this, let $g_n = Tf_n$ be a sequence in $R(T)$ such that g_n

goes to some g in L^2. Then given $\epsilon > 0$, we can choose m, n large enough such that

$$A \|f_n - f_m\|_2^2 \leq \langle Tf_n - Tf_m, f_n - f_m \rangle \leq \|Tf_n - Tf_m\|_2 \|f_n - f_m\|_2 < \epsilon \|f_n - f_m\|_2$$

and thus

$$\|f_n - f_m\|_2 < \epsilon / A.$$

Since the sequence f_n is a Cauchy sequence in L^2 then f_n converges to some f in L^2, i.e. g_n converges to $g = Tf$. Also for any $g \in R(T)^\perp$, we have by definition,

$$\langle Tf, g \rangle = 0 \quad \forall f \in L^2(\mathbb{R}),$$

which can be rewritten as

$$\sum_{j,k \in \mathbf{Z}} \langle f, \psi_{j,k} \rangle \langle \psi_{j,k}, g \rangle = 0 \quad \forall f \in L^2(\mathbb{R}).$$

Rearranging the above equation, we have

$$\langle f, \sum_{j,k \in \mathbf{Z}} \langle g, \psi_{j,k} \rangle \psi_{j,k} \rangle = 0 \quad \forall f \in L^2(\mathbb{R})$$

or equivalently,

$$\langle f, Tg \rangle = 0 \quad \forall f \in L^2(\mathbb{R}),$$

which implies that $Tg = 0$. Since we have already established that T is one-to-one, we have $g = 0$. Thus $R(T)^\perp = \{0\}$. Combining this with the fact that $R(T)$ is closed, this implies that T maps $L^2(\mathbb{R})$ onto itself. Therefore, since T is one-to-one and onto, its inverse T^{-1} is well defined and bounded. Its boundedness can be seen using the lower bound in (5.3) once more; if $Tf = g$ then

$$A \|f\|_2^2 \leq \langle Tf, f \rangle \leq \|Tf\|_2 \|f\|_2,$$

and replacing f by $T^{-1}g$ we have

$$A \left\|T^{-1} g\right\|_2 \leq \|g\|_2$$

so that

$$\left\|T^{-1}\right\|_2 \leq A^{-1}.$$

Finally, the reconstruction of a function $f \in L^2(\mathbb{R})$ is given by

$$\begin{aligned} f = T^{-1}Tf &= T^{-1} \sum_{j,k\in\mathbb{Z}} \langle f, \psi_{j,k}\rangle \psi_{j,k} \\ &= \sum_{j,k\in\mathbb{Z}} \langle f, \psi_{j,k}\rangle T^{-1}\psi_{j,k} \\ &= \sum_{j,k\in\mathbb{Z}} \langle f, \psi_{j,k}\rangle \tilde{\psi}_{j,k}, \end{aligned}$$

where the functions $\tilde{\psi}_{j,k} = T^{-1}\psi_{j,k}$ are commonly referred to as dual wavelets.

We should remark that in the case of the $\{\psi_{j,k}\}$ generating a frame, we do not necessarily obtain a basis of $L^2(\mathbb{R})$ (i.e. the $\{\psi_{j,k}\}$ may be linearly dependant). Such redundancy of the functions $\psi_{j,k}$ is often very useful in practical applications but is not the focus of this thesis. The rest of this section will deal with orthonormal wavelet bases. For the choices of the constants $a_0 = 2$ and $b_0 = 1$ there is a remarkable method of constructing ψ such that $\psi_{j,k}$, $j, k \in \mathbb{Z}$ constitutes an orthonormal basis (o.n.b.) of $L^2(\mathbb{R})$. This method is done by means of a *multiresolution analysis* which will now be defined.

5.1 Multiresolution Analysis

Definition 5.2 *A multiresolution analysis (MRA) of $L^2(\mathbb{R})$ is a sequence of closed subspaces $\{V_j\}_{j\in\mathbb{Z}}$ with the following properties:*

$$... \subset V_{-1} \subset V_0 \subset V_1 \subset ... \subset L^2(\mathbb{R}) \tag{5.4}$$

$$\bigcap_{j\in\mathbb{Z}} V_j = \{0\} \tag{5.5}$$

$$\overline{\bigcup_{j\in\mathbb{Z}} V_j} = L^2(\mathbb{R}) \tag{5.6}$$

$$f(x) \in V_j \Longleftrightarrow f(2x) \in V_{j+1} \tag{5.7}$$

$$\textit{There exists a function } \phi \in V_0 \textit{ such that} \tag{5.8}$$
$$\{\phi(x-k)\}_{k\in\mathbb{Z}} \textit{ is an orthonormal basis of } V_0.$$

Since $V_j \subset V_{j+1}$ for each j, we can define W_j as the orthogonal complement of V_j in V_{j+1}. That is, $V_{j+1} = V_j \oplus W_j$. Then we have

$$V_{j+1} = V_j \oplus W_j = V_{j-1} \oplus W_{j-1} \oplus W_j = \ldots = \bigoplus_{k=-\infty}^{j} W_k.$$

By (5.4) and (5.6), letting $j \to \infty$ we have,

$$L^2(\mathbb{R}) = \overline{\bigoplus_{k=-\infty}^{\infty} W_k}. \tag{5.9}$$

From (5.7) and (5.8), if we define

$$\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k),$$

then for fixed j, the set $\{\phi_{j,k};\ k \in \mathbb{Z}\}$ forms an o.n.b. of V_j. Furthermore, since $\phi(x) \in V_0 \subset V_1$ and $\{\phi_{1,k} = \sqrt{2}\phi(2\cdot -k) :\ k \in \mathbb{Z}\}$ is an o.n.b. of V_1, there exists a sequence $\{c_k\}$ belonging to $l^2(\mathbb{Z})$ such that

$$\phi(x) = \sum_{k\in\mathbb{Z}} c_k \phi_{1,k}(x). \tag{5.10}$$

The equation given by (5.10), which will be met again throughout this thesis, is commonly referred to as the *dilation equation.* If we can find a function ψ such that $\{\psi(x-k);\ k \in \mathbb{Z}\}$ forms an o.n.b. of W_0 then by defining

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k),$$

the set $\{\psi_{j,k};\ k \in \mathbb{Z}\}$ forms an o.n.b. of W_j. This follows by (5.7) , (5.8), and the fact that $W_j = V_{j+1} - V_j$ for each $j \in \mathbb{Z}$. Hence, by (5.9), $\{\psi_{j,k};\ j,k \in \mathbb{Z}\}$ forms an o.n.b. of

$L^2(\mathbb{R})$. The functions ϕ and ψ given above are often referred to as the scaling function and mother wavelet of the MRA, respectively.

A View from the Frequency Domain

For the construction of a mother wavelet ψ, it is often beneficial to consider the properties that ϕ and ψ must satisfy in terms of their Fourier transforms, $\hat{\phi}$ and $\hat{\psi}$. It should be noted that the results now given are not the author's own. Many graduate texts on wavelets contain these results and some good references are [4], [6], and [8].

By taking the Fourier transform of the dilation equation in (5.10), we obtain

$$\hat{\phi}(\xi) = \frac{1}{\sqrt{2}} \sum_{k \in \mathbb{Z}} c_k e^{-ik\xi/2} \hat{\phi}(\xi/2), \tag{5.11}$$

where the series converges in L^2. We can then rewrite (5.11) as

$$\hat{\phi}(\xi) = m_\phi(\xi/2)\hat{\phi}(\xi/2), \tag{5.12}$$

where

$$m_\phi(\xi) = \frac{1}{\sqrt{2}} \sum_{k \in \mathbb{Z}} c_k e^{-ik\xi} \tag{5.13}$$

is a 2π-periodic function belonging to $L^2([0, 2\pi])$ since

$$m_\phi(\xi + 2\pi) = \frac{1}{\sqrt{2}} \sum_{k \in \mathbb{Z}} c_k e^{-ik(\xi+2\pi)} = \frac{1}{\sqrt{2}} \sum_{k \in \mathbb{Z}} c_k e^{-ik\xi} e^{-2ik\pi} = m_\phi(\xi)$$

and by Parseval's identity,

$$\begin{aligned} \int_0^{2\pi} |m_\phi(\xi)|^2 d\xi &= \int_0^{2\pi} |\frac{1}{\sqrt{2}} \sum_{k \in \mathbb{Z}} c_k e^{-ik\xi}|^2 d\xi \\ &= \frac{2\pi}{\sqrt{2}} (\sum_{k \in \mathbb{Z}} |c_k|^2) d\xi \\ &= \frac{2\pi}{\sqrt{2}} \|\phi\|_2^2 < \infty. \end{aligned}$$

Since the set $\{\phi(x-k);\ k \in \mathbb{Z}\}$ is orthonormal we have

$$\begin{aligned}
\delta_{k,0} &= \int_{-\infty}^{\infty} \phi(x)\overline{\phi(x-k)}dx \\
&= \int_{-\infty}^{\infty} \hat{\phi}(\xi)\overline{\hat{\phi}(\xi)e^{-ik\xi}}d\xi \\
&= \int_{-\infty}^{\infty} |\hat{\phi}(\xi)|^2 e^{ik\xi}d\xi \\
&= \sum_{m\in\mathbb{Z}} \int_{2m\pi}^{2(m+1)\pi} |\hat{\phi}(\xi)|^2 e^{ik\xi}d\xi \\
&= \sum_{m\in\mathbb{Z}} \int_{0}^{2\pi} |\hat{\phi}(\xi+2m\pi)|^2 e^{ik\xi}d\xi, \qquad (5.14)
\end{aligned}$$

where the second equality follows from Plancherel's Formula and where $\delta_{k,l}$ is the Kronecker delta defined by

$$\delta_{k,l} = \begin{cases} 1 & if\ k = l \\ 0 & if\ k \neq l. \end{cases}$$

We would like to interchange the integral and the summation given in (5.14). This is justified if we consider the separate cases when $k = 0$ and when $k \neq 0$. In either case, let

$$f_m(\xi) = |\hat{\phi}(\xi+2m\pi)|^2 e^{ik\xi}.$$

Then for $k = 0$, since $f_m \geq 0$ for all $m \in \mathbb{Z}$, it follows from a corollary of Lebesgue's Monotone Convergence Theorem (Theorem 1.27 in [18]) that

$$\int_0^{2\pi} \sum_{m\in\mathbb{Z}} f_m(\xi)d\xi = \sum_{m\in\mathbb{Z}} \int_0^{2\pi} f_m(\xi)d\xi. \qquad (5.15)$$

For $k \neq 0$,

$$\begin{aligned}
\sum_{m\in\mathbb{Z}} \int_0^{2\pi} |f_m(\xi)| &= \sum_{m\in\mathbb{Z}} \int_0^{2\pi} |\hat{\phi}(\xi+2m\pi)|^2 d\xi \\
&= \sum_{-\infty}^{\infty} |\hat{\phi}(\xi)|^2 d\xi = \|\phi\|_2^2 < \infty.
\end{aligned}$$

A corollary of Lebesgue's Dominated Convergence Theorem (Theorem 1.38 in [18]) then tells us that (5.15) is true for almost every ξ. Thus we have,

$$\delta_{k,0} = \int_0^{2\pi} (\sum_{m\in\mathbb{Z}} |\hat{\phi}(\xi+2m\pi)|^2) e^{ik\xi}d\xi. \qquad (5.16)$$

Notice from (5.16) that we can calculate the Fourier coefficients h_k of the 2π-periodic function $\sum_{m\in\mathbf{Z}} |\hat{\phi}(\xi+2m\pi)|^2$ as $h_k = 0$ for all $k \neq 0$ and $h_0 = 1/2\pi$. That is,

$$\sum_{m\in\mathbf{Z}} |\hat{\phi}(\xi+2m\pi)|^2 = \frac{1}{2\pi} \quad \text{a.e.} \tag{5.17}$$

We can now show a nice property that the function m_ϕ satisfies. Substituting (5.12) into (5.17) and by a change of variables ($\omega = \xi/2$) we have

$$\begin{aligned}\frac{1}{2\pi} &= \sum_{m\in\mathbf{Z}} |\hat{\phi}(2\omega+2m\pi)|^2 \\ &= \sum_{m\in\mathbf{Z}} |m_\phi(\omega+m\pi)|^2 |\hat{\phi}(\omega+m\pi)|^2.\end{aligned}$$

Splitting the sum into even and odd integers, we obtain

$$\begin{aligned}\frac{1}{2\pi} &= \sum_{m\in\mathbf{Z}} |m_\phi(\omega+2m\pi)|^2 |\hat{\phi}(\omega+2m\pi)|^2 \\ &\quad + \sum_{m\in\mathbf{Z}} |m_\phi(\omega+(2m+1)\pi)|^2 |\hat{\phi}(\omega+(2m+1)\pi)|^2 \\ &= |m_\phi(\omega)|^2 \sum_{k\in\mathbf{Z}} |\hat{\phi}(\omega+2m\pi)|^2 \\ &\quad + |m_\phi(\omega+\pi)|^2 \sum_{k\in\mathbf{Z}} |\hat{\phi}(\omega+(2m+1)\pi)|^2.\end{aligned} \tag{5.18}$$

The second equality comes from the fact that m_ϕ is 2π-periodic. Since (5.17) holds for a.e. ξ, then it holds for a.e. $\xi+\pi$ and hence we have

$$\sum_{m\in\mathbf{Z}} |\hat{\phi}(\xi+2m\pi)|^2 = \sum_{m\in\mathbf{Z}} |\hat{\phi}(\xi+(2m+1)\pi)|^2 = \frac{1}{2\pi} \quad \text{a.e.}$$

Then (5.18) simplifies to

$$|m_\phi(\omega)|^2 + |m_\phi(\omega+\pi)|^2 = 1 \quad \text{a.e.} \tag{5.19}$$

Equation (5.19) will be used shortly.

In order to find our mother wavelet ψ, we now would like to describe the properties of a function f belonging to W_0. Since $f \in W_0$ if and only if $f \in V_1$ and $f \perp V_0$, we have

$$f = \sum_{k\in\mathbf{Z}} a_k\, \phi_{1,k}, \quad \text{where } a_k = \langle f,\, \phi_{1,k}\rangle \tag{5.20}$$

and

$$\langle f,\ \phi_{0,k}\rangle = 0 \tag{5.21}$$

for all $k \in \mathbb{Z}$. Writing (5.20) in the frequency domain, we have

$$\hat{f}(\xi) = \frac{1}{\sqrt{2}} \sum_{k\in\mathbb{Z}} a_k e^{-ik\xi/2} \hat{\phi}(\xi/2) = m_f(\xi/2)\hat{\phi}(\xi/2), \tag{5.22}$$

where m_f is the 2π-periodic $L^2([0, 2\pi])$ function

$$m_f(\xi) = \frac{1}{\sqrt{2}} \sum_k a_k e^{-ik\xi}.$$

We can rewrite (5.21) as

$$\begin{aligned} 0 &= \int_{\mathbb{R}} \hat{f}(\xi)\overline{\hat{\phi}(\xi)}e^{ik\xi}d\xi \\ &= \sum_{m\in\mathbb{Z}} \int_{2m\pi}^{(2m+1)\pi} f(\xi)\overline{\hat{\phi}(\xi)}e^{ik\xi}d\xi \\ &= \int_0^{2\pi} \left(\sum_{m\in\mathbb{Z}} \hat{f}(\xi + 2m\pi)\overline{\hat{\phi}(\xi + 2m\pi)}\right) e^{ik\xi}d\xi, \end{aligned}$$

so that the Fourier coefficients of $\sum_{m\in\mathbb{Z}} \hat{f}(\xi + 2m\pi)\overline{\hat{\phi}(\xi + 2m\pi)}$ are all zero. Thus replacing ξ by 2ξ, we have

$$\sum_{m\in\mathbb{Z}} \hat{f}(2\xi + 2m\pi)\overline{\hat{\phi}(2\xi + 2m\pi)} = 0 \text{ a.e.}$$

Rewriting the left side of the above equation using (5.12) and (5.22), we have

$$\sum_{m\in\mathbb{Z}} m_f(\xi + m\pi)\hat{\phi}(\xi + m\pi)\overline{m_\phi(\xi + m\pi)\hat{\phi}(\xi + m\pi)} = 0,$$

or equivalently,

$$\sum_{m\in\mathbb{Z}} m_f(\xi + m\pi)|\hat{\phi}(\xi + m\pi)|^2\overline{m_\phi(\xi + m\pi)}. \tag{5.23}$$

Breaking the sum in (5.23) into sums over even and odd $m \in \mathbb{Z}$, we obtain

$$\begin{aligned} &\sum_{m\in\mathbb{Z}} m_f(\xi + (2m+1)\pi)|\hat{\phi}(\xi + (2m+1)\pi)|^2\overline{m_\phi(\xi + (2m+1)\pi)} \\ &\qquad + \sum_{m\in\mathbb{Z}} m_f(\xi + 2m\pi)|\hat{\phi}(\xi + 2m\pi)|^2\overline{m_\phi(\xi + 2m\pi)} = 0 \end{aligned}$$

and, by the periodicity of both m_f and m_ϕ, we have

$$m_f(\xi+\pi)\overline{m_\phi(\xi+\pi)}\sum_{m\in\mathbf{Z}}|\hat{\phi}(\xi+(2m+1)\pi)|^2$$
$$+m_f(\xi)\overline{m_\phi(\xi)}\sum_{m\in\mathbf{Z}}|\hat{\phi}(\xi+2m\pi)|^2=0.$$

From (5.17), each of the sums above equals $1/2\pi$ and hence we obtain

$$m_f(\xi+\pi)\overline{m_\phi(\xi+\pi)}+m_f(\xi)\overline{m_\phi(\xi)}=0 \quad \text{a.e.} \tag{5.24}$$

Since m_f and m_ϕ are both 2π-periodic we can certainly write

$$m_f(\xi)=\beta(\xi)\overline{m_\phi(\xi+\pi)} \quad \text{a.e.,} \tag{5.25}$$

where β is 2π-periodic and since (5.19) tells us that for almost every ξ, $\overline{m_\phi(\xi+\pi)}$ and $\overline{m_\phi(\xi)}$ cannot both equal zero, then (5.24) implies that

$$\beta(\xi)+\beta(\xi+\pi)=0 \quad \text{a.e.,} \tag{5.26}$$

which can be rewritten as

$$\beta(\xi)=e^{i\xi}v(2\xi)$$

for some 2π-periodic function v. Thus any function $f\in W_0$ can be written as

$$\hat{f}(\xi)=e^{i\xi/2}\overline{m_\phi(\xi/2+\pi)}v(\xi)\hat{\phi}(\xi/2), \tag{5.27}$$

where v is a 2π-periodic function depending on f. This suggests that the function ψ that we seek may be defined by the condition

$$\hat{\psi}(\xi)=e^{i\xi/2}\overline{m_\phi(\xi/2+\pi)}\hat{\phi}(\xi/2) \tag{5.28}$$

since this would imply

$$\hat{f}(\xi)=v(\xi)\hat{\psi}(\xi)=\Big(\sum_{k\in\mathbf{Z}}v_ke^{-ik\xi}\Big)\hat{\psi}(\xi), \tag{5.29}$$

or equivalently,

$$f(x)=\sum_{k\in\mathbf{Z}}v_k\psi(x-k).$$

We still need to prove that $\{\psi(x-k);\ k \in \mathbb{Z}\}$ is an o.n.b of W_0. As was the case for proving orthonormality of the $\phi(\cdot - k)$, it suffices to show

$$\sum_{m\in\mathbf{Z}} |\hat{\psi}(\xi + 2\pi m)|^2 = 1/2\pi \quad \text{a.e.}$$

to prove that the $\psi_{0,k}$ are orthonormal. This follows easily since

$$\begin{aligned}
\sum_{m\in\mathbf{Z}} |\hat{\psi}(\xi+2\pi m)|^2 &= \sum_{m\in\mathbf{Z}} |m_\phi(\xi/2+m\pi+\pi)|^2|\hat{\phi}(\xi/2+m\pi)|^2 \\
&= |m_\phi(\xi/2+\pi)|^2 \sum_{m\in\mathbf{Z}} |\hat{\phi}(\xi/2+2m\pi)|^2 \\
&\quad + |m_\phi(\xi/2)|^2 \sum_{m\in\mathbf{Z}} |\hat{\phi}(\xi/2+2m\pi+\pi)|^2 \\
&= 1/2\pi\left(|m_\phi(\xi/2)|^2 + |m_\phi(\xi/2+\pi)|^2\right) \quad \text{a.e.} \\
&= 1/2\pi \quad \text{a.e.}
\end{aligned}$$

It only remains to prove that $\{\psi(x-k);\ k \in \mathbb{Z}\}$ is a basis for W_0. For this purpose we return to (5.29). It then suffices to show that

$$\sum_{k\in\mathbf{Z}} |v_k|^2 < \infty,$$

or equivalently, that v is 2π-periodic and belonging to $L^2([0,2\pi])$. The periodicity of v was already given and hence we only need to show the latter property. First note that

$$\begin{aligned}
\int_0^{2\pi} |v(\xi)|^2 d\xi &= \int_0^{2\pi} |e^{-i\xi/2}\beta(\xi/2)|^2 d\xi \\
&= \int_0^{\pi} |\beta(\xi)|^2 d\xi \\
&= \int_0^{\pi} |\beta(\xi)|^2\left(|(m_\phi(\xi)|^2 + |(m_\phi(\xi+\pi)|^2\right)d\xi,
\end{aligned}$$

where the last equality follows from (5.19). Notice also from (5.26) and the periodicity of m_ϕ that

$$\begin{aligned}
\int_0^{\pi} |\beta(\xi)|^2|m_\phi(\xi)|^2 d\xi &= \int_0^{\pi} |-\beta(\xi+\pi)|^2|m_\phi(\xi+2\pi)|^2 d\xi \\
&= \int_\pi^{2\pi} |\beta(\xi)|^2|m_\phi(\xi+\pi)|^2 d\xi,
\end{aligned}$$

and hence

$$\int_0^{2\pi} |v(\xi)|^2 d\xi = \int_0^{2\pi} |\beta(\xi)|^2 |m_\phi(\xi+\pi)|^2 d\xi$$
$$= \int_0^{2\pi} |m_f(\xi)|^2 d\xi < \infty,$$

where the second equality follows from (5.25) and the last inequality from the fact that $m_f \in L^2([0,2\pi])$. This gives us the desired result. Thus we have shown that (5.28) is sufficient for defining a mother wavelet. This is expressed by the following theorem.

Theorem 5.1 *Let $\{V_j\}$ be an MRA with scaling function ϕ that satisfies (5.8). Then the function ψ given by*

$$\hat{\psi}(\xi) = e^{i\xi/2}\overline{m_\phi(\xi/2+\pi)}\hat{\phi}(\xi/2)$$

or equivalently

$$\psi(x) = \sqrt{2}\sum_{k\in\mathbb{Z}}(-1)^{k-1}\overline{c_{-k-1}}\phi(2x-k),$$

where m_ϕ and c_k are given by (5.13) and (5.10), respectively, has the property that $\{\psi_{j,k}; j,k \in \mathbb{Z}\}$ is an o.n.b. of $L^2(\mathbb{R})$.

Remark. The mother wavelet ψ is not uniquely determined by the associated MRA. In fact, if ψ is given as in Theorem 5.1, then any ψ_0 satisfying

$$\hat{\psi}_0(\xi) = \sigma(\xi)\hat{\psi}(\xi),$$

with σ being 2π-periodic such that $|\sigma(\xi)| = 1$ a.e., will constitute a mother wavelet as well.

Some of the conditions that must be satisfied for obtaining an MRA often follow from simpler conditions which may very well be easier to prove. An example is given by this next theorem which can be found in [4] along with a proof (see Exercises 7.5).

Theorem 5.2 *Let $\{V_j\}_{j\in\mathbb{Z}}$ be a sequence of closed subspaces of $L^2(\mathbb{R})$ satisfying (5.4), (5.7), and (5.8). Then* *(a)* $\bigcap_{j\in\mathbf{Z}} V_j = \{0\}$,

(b) $W = \overline{\bigcup_{j\in\mathbf{Z}} V_j}$ *is invariant under translation,*

(c) If $\hat{\phi}$ *is continuous at 0 and* $\hat{\phi}(0) = 1$*, then* $\overline{\bigcup_{j\in\mathbf{Z}} V_j} = L^2(\mathbb{R})$.

5.2 Some Examples of Wavelets

The Haar Wavelet

The first ever construction of orthonormal wavelets was due to Alfred Haar, who proved that the function

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

is a mother wavelet. Although Haar's proof (in 1910) was long before the concept of a multiresolution analysis, these wavelets are the simplest example of wavelets generated from an MRA. One starts with the scaling function $\phi(x) = \chi_{[0,1)}(x)$. Then it is obvious that the spaces V_j are defined by

$$V_j = \left\{ f : \ f \text{ is constant on } \left[2^{-j}k, 2^{-j}(k+1)\right) \ , k \in \mathbb{Z} \right\},$$

and properties (5.4) and (5.7) follow immediately. As for property (5.5), for any interval (a, b) such that $(a, b) \cap \{0\} = \phi$, there exists a negative integer j with $|j|$ large enough such that $(a, b) \subset [0, 2^{-j})$ or $(a, b) \subset [-2^{-j}, 0)$, thus any function $f \in \bigcap_{j\in\mathbf{Z}} V_j$ must be of the form

$$f = \begin{cases} c_1 & \text{if } \ 0 \le x < \infty \\ c_2 & \text{if } -\infty < x < 0. \end{cases}$$

The only such function belonging to $L^2(\mathbb{R})$ is $f = 0$. This proves (5.5) is satisfied. The condition given by (5.6) is obviously satisfied if we note that any L^2 function can be approximated with arbitrary precision by step functions and that for any step function $s(x)$ and $\epsilon > 0$, one can choose j large enough so that there is a function $h \in V_j$ such that $\|s - h\|_2 < \epsilon$. Computing the coefficients of the dilation equation (5.10), we have

$$c_0 = c_1 = 1/\sqrt{2}$$

and

$$c_k = 0$$

for all $k \neq 0, 1$. Then Theorem 5.1 tells us that

$$\psi_0(x) = \phi(2x+1) - \phi(2x+2) = \begin{cases} -1 & \text{if } -1 \leq x < -1/2 \\ 1 & \text{if } -1/2 \leq x < 0 \\ 0 & \text{otherwise} \end{cases}$$

is a mother wavelet. Although this is not exactly the mother wavelet of the Haar system, the remark following Theorem 5.1 tells us that the function ψ such that

$$\hat{\psi}(\xi) = -e^{-i\xi}\hat{\psi}_0(\xi),$$

is also a mother wavelet. This ψ is precisely the Haar mother wavelet.

The Meyer Wavelet

Meyer's wavelet ψ is defined by

$$\hat{\psi}(\xi) = \begin{cases} (2\pi)^{-1/2} e^{i\xi/2} \sin[\frac{\pi}{2} u(\frac{3}{2\pi}|\xi| - 1)], & \frac{2\pi}{3} \leq |\xi| \leq \frac{4\pi}{3} \\ (2\pi)^{-1/2} e^{i\xi/2} \cos[\frac{\pi}{2} u(\frac{3}{4\pi}|\xi| - 1)], & \frac{4\pi}{3} \leq |\xi| \leq \frac{8\pi}{3} \\ 0 & \textit{otherwise}, \end{cases} \tag{5.30}$$

where the function u is a C^N or C^∞ function satisfying the properties

$$u(x) = \begin{cases} 0 \;\; if \;\; x \leq 0 \\ 1 \;\; if \;\; x \geq 1 \end{cases} \tag{5.31}$$

and

$$u(x) + u(1-x) = 1. \tag{5.32}$$

An example of such a function (in C^3) is the function given by

$$u(x) = x^4(35 - 84x + 70x^2 - 20x^3)$$

for $0 \leq x \leq 1$ (taken from [8]). It is obvious from the definition of $\hat{\psi}$ that if $u \in C^N$, then so is $\hat{\psi}$.

We would like to construct ψ from the scaling function ϕ given by

$$\hat{\phi}(\xi) = \begin{cases} (2\pi)^{-1/2}, & |\xi| \leq \frac{2\pi}{3} \\ (2\pi)^{-1/2}\cos[\frac{\pi}{2}u(\frac{3}{2\pi}|\xi| - 1)], & \frac{2\pi}{3} \leq |\xi| \leq \frac{4\pi}{3} \\ 0 & otherwise. \end{cases}$$

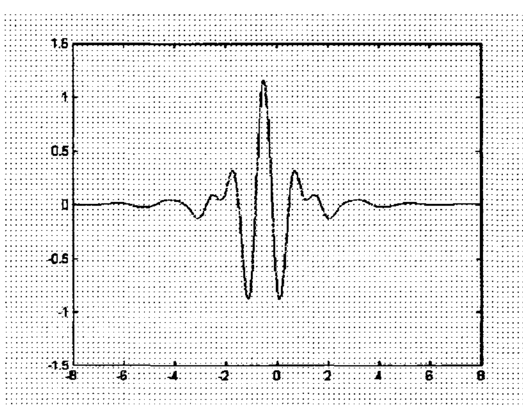

Figure 3: The Meyer Wavelet (image courtesy of [19])

Our first observation is that the $\phi(\cdot - k)$ are orthonormal. To see this we note that for any $\xi \in \mathbb{R}$ there exists exactly one or exactly two $k \in \mathbb{Z}$ such that

$$\hat{\phi}(\xi + 2\pi k) \neq 0. \tag{5.33}$$

If there is one such k satisfying (5.33), which we denote $\tilde{k}$, then we must have $|\xi+2\pi\tilde{k}| \leq 2\pi/3$ otherwise either $\tilde{k}-1$ or $\tilde{k}+1$ would also satisfy (5.33). Hence in this case,

$$\sum_{k\in\mathbf{Z}} |\hat{\phi}(\xi+2\pi k)|^2 = |\hat{\phi}(\xi+2\pi\tilde{k})|^2 = 1/2\pi.$$

If there are two integers satisfying (5.33), it is obvious that we can denote them by $\tilde{k}$ and $\tilde{k}+1$, so that $-4\pi/3 \leq \xi+2\pi\tilde{k} \leq -2\pi/3$ and $2\pi/3 \leq \xi+2\pi(\tilde{k}+1) \leq 4\pi/3$. Then

$$\begin{aligned}
\sum_{k\in\mathbf{Z}} |\hat{\phi}(\xi+2\pi k)|^2 &= |\hat{\phi}(\xi+2\pi\tilde{k})|^2 + |\hat{\phi}(\xi+2\pi(\tilde{k}+1))|^2 \\
&= \frac{1}{2\pi}\big(\cos^2\big[\frac{\pi}{2}u(-\frac{3}{2\pi}(\xi+2\pi\tilde{k})-1)\big] \\
&\qquad + \cos^2\big[\frac{\pi}{2}u(\frac{3}{2\pi}(\xi+2\pi(\tilde{k}+1))-1)\big]\big) \\
&= \frac{1}{2\pi}\big(\cos^2\big[\frac{\pi}{2}(1-u(v(\xi)))\big] + \cos^2\big[\frac{\pi}{2}u(v(\xi))\big]\big),
\end{aligned}$$

where $v(\xi) = \frac{3}{2\pi}(\xi+2\pi(\tilde{k}+1))-1$, and the last inequality follows from using (5.32) in the first term. Hence by applying the trigonometric identity $\cos(a-b) = \cos a\cos b + \sin a \sin b$ to the first term we obtain

$$\sum_{k\in\mathbf{Z}} |\hat{\phi}(\xi+2\pi k)|^2 = \frac{1}{2\pi}.$$

We now let V_0 be the closed subspace spanned by $\{\phi_{0,k}(x); k\in\mathbb{Z}\}$ and V_j the closed subspace spanned by $\{\phi_{j,k}(x); k\in\mathbb{Z}\}$. It is obvious that (5.4) is satisfied if and only if $\phi\in V_1$ or equivalently, as we have already seen, if there exists a function m_ϕ which is 2π-periodic belonging to $L^2([0,2\pi])$ such that

$$\hat{\phi}(\xi) = m_\phi(\xi/2)\hat{\phi}(\xi/2).$$

Choosing $m_\phi(\xi) = \sqrt{2\pi}\sum_{k\in\mathbf{Z}}\hat{\phi}(2\xi+4\pi k)$, we see that

$$\begin{aligned}
m_\phi(\xi/2)\hat{\phi}(\xi/2) &= \sqrt{2\pi}\sum_{k\in\mathbf{Z}}\hat{\phi}(\xi+4\pi k)\hat{\phi}(\xi/2) \\
&= \sqrt{2\pi}\hat{\phi}(\xi)\hat{\phi}(\xi/2) \\
&= \hat{\phi}(\xi),
\end{aligned}$$

where the second equality follows since $|\xi/2| \leq 8\pi/3$ and $|\xi + 4\pi k| \leq 8\pi/3$ if and only if $k = 0$, and the third equality follows since $\xi \in \text{supp } \hat{\phi}$ implies $|\xi/2| \leq 2\pi/3$ and hence $\sqrt{2\pi}\hat{\phi}(\xi/2) = 1$. Property (5.7) is trivial from the dilation equation and hence, Theorem 5.2(c) proves (5.6), and consequently that $\{V_j\}$ constitute an MRA. Lastly, to see ψ, as defined in (5.30), is the mother wavelet produced by the MRA,

$$
\begin{aligned}
\hat{\psi}(\xi) &= e^{i\xi/2}\overline{m_\phi(\xi/2+\pi)}\hat{\phi}(\xi/2) \\
&= \sqrt{2\pi}e^{i\xi/2}\sum_{k\in\mathbb{Z}}\hat{\phi}(\xi+2\pi+4\pi k)\hat{\phi}(\xi/2) \\
&= \sqrt{2\pi}e^{i\xi/2}[\hat{\phi}(\xi+2\pi)+\hat{\phi}(\xi-2\pi)]\hat{\phi}(\xi/2),
\end{aligned}
$$

where the last equality follows since $\xi/2 \in \text{supp } \hat{\phi}$ implies $\hat{\phi}(\xi + 2\pi + 4\pi k) = 0$ for all $k \neq 0, -1$. Then for $|\xi| > 8\pi/3$ we have $\hat{\phi}(\xi/2) = 0$ which implies $\hat{\psi}(\xi) = 0$. Also for $|\xi| \leq 2\pi/3$ both $\hat{\phi}(\xi \pm 2\pi) = 0$ which implies $\hat{\psi}(\xi) = 0$. For $4\pi/3 \leq \xi \leq 8\pi/3$,

$$
\hat{\phi}(\xi/2) = (2\pi)^{-1/2}\cos\big[\frac{\pi}{2}u(\frac{3}{4\pi}|\xi|-1)\big]
$$

and

$$
\begin{aligned}
\hat{\phi}(\xi+2\pi)+\hat{\phi}(\xi-2\pi) &= (2\pi)^{-1/2}\big(\cos\big[\frac{\pi}{2}u(\frac{3}{2\pi}(\xi+2\pi)-1)\big]+1\big) \\
&= (2\pi)^{-1/2}\big(\cos(\frac{\pi}{2})+1\big) \\
&= (2\pi)^{-1/2},
\end{aligned}
$$

where the second equality follows from (5.31). When $-8\pi/3 \leq \xi \leq 4\pi/3$ the same results hold and will be left for the reader. Then for $4\pi/3 \leq |\xi| \leq 8\pi/3$ we have shown

$$
\begin{aligned}
\hat{\psi}(\xi) &= \sqrt{2\pi}e^{i\xi/2}[\hat{\phi}(\xi+2\pi)+\hat{\phi}(\xi-2\pi)]\hat{\phi}(\xi/2) \\
&= \sqrt{2\pi}e^{i\xi/2}(2\pi)^{-1/2}(2\pi)^{-1/2}\cos\big[\frac{\pi}{2}u(\frac{3}{4\pi}|\xi|-1)\big] \\
&= (2\pi)^{-1/2}e^{i\xi/2}\cos\big[\frac{\pi}{2}u(\frac{3}{4\pi}|\xi|-1)\big],
\end{aligned}
$$

which is consistent with (5.30). The cases when $-4\pi/3 \leq \xi \leq -2\pi/3$ and when $2\pi/3 \leq \xi \leq 4\pi/3$ are also very similar and we will only consider the latter one. For such a ξ, it is immediate that $\hat{\phi}(\xi/2) = (2\pi)^{-1/2}$, and since $\xi + 2\pi \geq 8\pi/3$ it follows that $\hat{\phi}(\xi+2\pi) = 0$. Also

$$
\begin{aligned}
\hat{\phi}(\xi - 2\pi) &= (2\pi)^{-1/2} \cos\big[\frac{\pi}{2} u(-\frac{3}{2\pi}(\xi - 2\pi) - 1)\big] \\
&= (2\pi)^{-1/2} \cos\big[\frac{\pi}{2}(1 - u(\frac{3}{2\pi}\xi - 1)\big] \\
&= (2\pi)^{-1/2} \sin\big[\frac{\pi}{2} u(\frac{3}{2\pi}\xi - 1)\big].
\end{aligned}
$$

We then have, for $2\pi/3 \leq |\xi| \leq 4\pi/3$,

$$
\hat{\psi}(\xi) = (2\pi)^{-1/2} e^{i\xi/2} \sin[\frac{\pi}{2} u(\frac{3}{2\pi}|\xi| - 1)].
$$

Thus we have proved that ψ is as defined in (5.30).

We can now take a look at the properties of ψ. Since $\hat{\psi}$ has compact support, integrating by parts N times (if u belongs to C^N), we have

$$
\begin{aligned}
\int_{\mathbb{R}} \hat{\psi}^{(N)}(\xi) e^{ix\xi} d\xi &= (-ix)^N \int_{\mathbb{R}} \hat{\psi}(\xi) e^{ix\xi} d\xi \\
&= \sqrt{2\pi}(-ix)^N \psi(x),
\end{aligned}
$$

which implies that

$$
\psi(x) = O(\frac{1}{1 + |x|^N}).
$$

Similarly, if $u \in C^\infty$ then ψ decays faster than any polynomial, i.e. for any $N \in \mathbb{N}$

$$
|\psi(x)||x|^N \to 0 \quad \text{as} \quad |x| \to \infty.
$$

From the Riemann-Lebesgue Lemma (see [6]), we know that $f \in L^1$ implies $\hat{f} \in C_0$ or similarly $\hat{f} \in L^1$ implies $f \in C_0$. Then, since $\hat{\psi} \in C_c$, it is obvious that for any $k \in \mathbb{N}$,

$$
(-i\xi)^k \hat{\psi}(\xi) = \widehat{\frac{\partial^k \psi}{\partial \xi^k}}(\xi)
$$

also belongs to $C_c \subset L^1$. Hence $\frac{\partial^k \psi}{\partial \xi^k} \in C_0$, which implies $\psi \in C^k$. Since k was arbitrary we have $\psi \in C^\infty$.

The Meyer wavelet seems, in many ways, to be more useful then the Haar wavelet we have seen. Although Meyer's wavelet does not have compact support as does the Haar wavelet, it still has nice decay for the function u chosen sufficiently smooth. Furthermore, we see that the Meyer wavelet belongs to C^∞ while the Haar wavelet has a jump discontinuity. These properties lead to the question of whether we can find wavelets with compact support that are also relatively smooth. This brings us to the construction of Daubechies' wavelets.

Daubechies' Wavelets

One of the greatest achievements in wavelet analysis was that of Ingrid Daubechies, who was able to construct a family of wavelets with compact support, such that as the length of their support increases so does their regularity. The full construction of these wavelets is beyond the scope of this paper and the reader may refer to [8] for a detailed exposition. We will however give a general description as to how these wavelets are constructed. Instead of attempting to construct the wavelet ψ by starting with the scaling function ϕ or the MRA $\{V_j\}$, one starts with the function m_ϕ. The easiest way to guarantee that a mother wavelet ψ is compactly supported is by choosing the scaling function ϕ to have compact support. It then follows that the coefficients,

$$c_k = \sqrt{2}\int_{\mathbf{R}} \phi(x)\phi(2x-k)dx,$$

are non-zero for only a finite number of k. It obviously would follow then that ψ is a finite sum of functions with compact support and hence is compactly supported itself. The function

$$m_\phi(\xi) = \frac{1}{\sqrt{2}}\sum_{k\in\mathbf{Z}} c_k e^{-ik\xi} \tag{5.34}$$

is a finite sum and hence a trigonometric polynomial. Daubechies then was able to prove that by setting, for each integer $N \geq 0$,

$$m_{\phi,N}(\xi) = \left(\frac{1+e^{-i\xi}}{2}\right)^N P_N(\xi),$$

where

$$P_N(\xi) = \sum_{k=0}^{N-1}\binom{N-1+k}{k}\sin^{2k}(\xi/2),$$

the functions ϕ_N and ψ_N defined by

$$\hat{\phi}_N(\xi) = m_{\phi,N}(\xi/2)\hat{\phi}(\xi/2)$$

and

$$\hat{\psi}_N(\xi) = e^{i\xi/2}\overline{m_{\phi,N}(\xi/2+\pi)}\hat{\phi}_N(\xi/2)$$

would constitute a scaling function and mother wavelet of an MRA respectively. More importantly, both ϕ_N and ψ_N are supported in an interval of length $2N-1$ and become smoother as N increases. More precisely, ϕ_N and ψ_N both belong to $C^{r(N)}$ where

$$\lim_{N\to+\infty}\frac{r(N)}{N} = \sigma \approx 1/5.$$

That is to say, for example, if ψ is to be 20 times differentiable, then the length of its support must be approximately 200. We should also note that when $N = 1$, Daubechies' wavelet reduces to the Haar wavelet. For each N, the wavelet ψ_N is often called the Daubechies $2N$ wavelet because there are $2N$ non-zero coefficients c_k in (5.34).

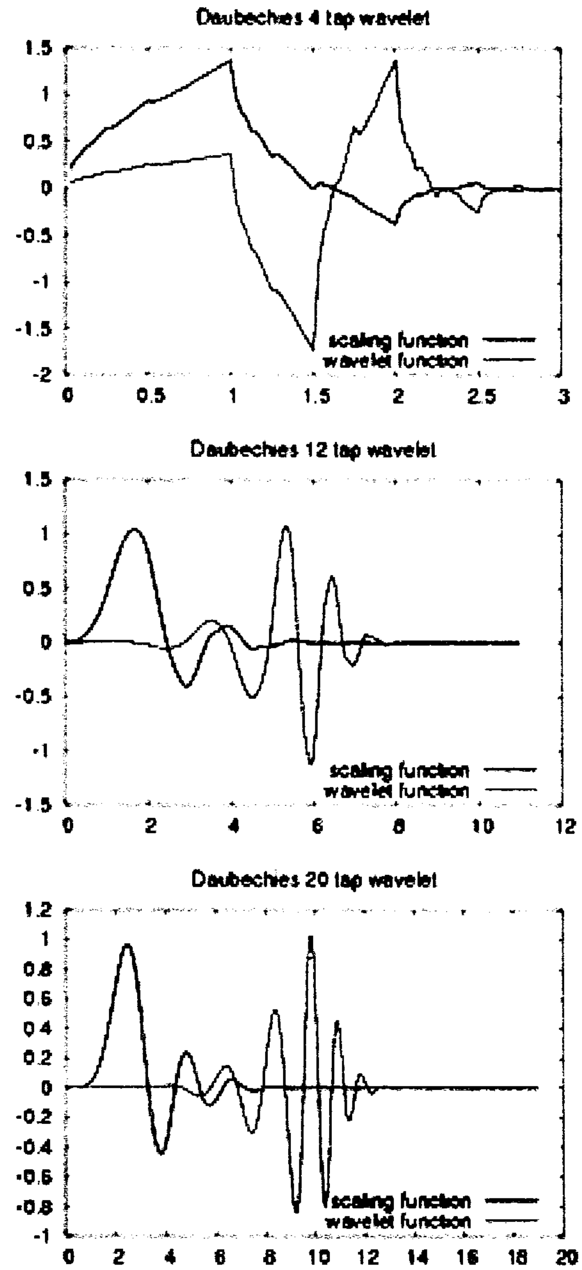

Figure 4: (from top to bottom) Daubechies wavelets and scaling functions ψ_N and ϕ_N for $N = 2, 6, 10$ (images courtesy of [19])

5.3 Multidimensional Wavelets

The notion of an MRA can easily be extended to $L^2(\mathbb{R}^n)$ by a tensor product construction involving the one dimensional wavelet ψ and scaling function ϕ. Although there are more elegant constructions of multidimensional wavelets, the tensor product construction here will be sufficient for the purpose of this paper. To simplify notation, we let $\phi = \psi^0$ and $\psi = \psi^1$. Then the MRA of $L^2(\mathbb{R}^n)$ is the sequence of closed subspaces $\{V_j'\} \subset L^2(\mathbb{R}^n)$ where $V_j' = \bigotimes^n V_j$ and $\{V_j\}$ is the MRA corresponding to the one dimensional scaling function ψ^0. Then the scaling function associated with $\{V_j'\}$ is given by

$$\phi(x) = \psi^0(x_1)\psi^0(x_2)...\psi^0(x_n).$$

There will be $2^n - 1$ mother wavelets needed and they are given as follows. Let Ω be the set of vertices of the unit cube in $\mathbb{R}^n$ and $\Omega' := \Omega\backslash\{0\}$ Then the mother wavelets are given by

$$\psi^\omega(x) = \psi^{\omega_1}(x_1)\psi^{\omega_2}(x_2)...\psi^{\omega_n}(x_n),$$

where $\omega = (\omega_1, \omega_2, ..., \omega_n) \in \Omega'$. Letting $\psi^\omega_{j,k}(x) = 2^{nj}\psi^\omega(2^j x - k)$ we have that $\{\psi^\omega_{j,k}; \omega \in \Omega', j, k \in \mathbb{Z}\}$ forms an o.n.b. of $L^2(\mathbb{R}^n)$, i.e., for any $f \in L^2(\mathbb{R}^n)$ we have

$$f = \sum_{\omega\in\Omega'} \sum_{j,k\in\mathbb{Z}} \langle f, \psi^\omega_{j,k}\rangle \psi^\omega_{j,k}.$$

Notation. To simplify notation for wavelet bases of $L^2(\mathbb{R}^n)$, throughout this thesis we let $\Lambda = \mathbb{Z} \times \mathbb{Z}^2 \times \{1, ..., 2^n - 1\}$ and for the analyzing wavelets $\{\psi^1, \psi^2, ..., \psi^{2^n-1}\}$, we define ψ_λ, $\lambda = (j, k, m) \in \Lambda$, as

$$\psi_\lambda = \psi^m_{j,k}.$$

We now look at an example of 2-dimensional wavelets as they are obviously very important in image processing.

Example: The 2-D Haar Wavelets

We begin with the 1-D Haar scaling function ψ^0 and wavelet function ψ^1 given by

$\psi^0(x) = \chi_{[0,1)}$ and

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

For each j, the subspace $V_j = \text{span}\{\psi^0_{j,k}(x); k \in \mathbb{Z}\}$ is the set of functions which are constant on the intervals $I_{j,k} = [2^{-j}k, 2^{-j}(k+1))$. Then for the construction of the 2-D MRA we have $\phi(x) = \chi_{[0,1)}\chi_{[0,1)} = \chi_E$ where $E = [0,1)^2$ is the unit square and hence for each j, the subspace V_j' is the set of functions which are constant on the cubes $E_{j,k} = [2^{-j}k_1, 2^{-j}(k_1+1)] \times [2^{-j}k_2, 2^{-j}(k_2+1))$ where $k = (k_1, k_2) \in \mathbb{Z}^2$. The three mother wavelets are then given by

$$\psi_1(x) = \psi^0(x_1)\psi^1(x_2) = \begin{cases} 1 & \text{if } 0 \leq x_1 < 1,\ 0 \leq x_2 < 1/2 \\ -1 & \text{if } 0 \leq x_1 < 1,\ 1/2 \leq x_2 < 1 \\ 0 & \text{otherwise,} \end{cases}$$

$$\psi_2(x) = \psi^1(x_1)\psi^0(x_2) = \begin{cases} 1 & \text{if } 0 \leq x_1 < 1/2,\ 0 \leq x_2 < 1 \\ -1 & \text{if } 1/2 \leq x_1 < 1,\ 0 \leq x_2 < 1 \\ 0 & \text{otherwise,} \end{cases}$$

and

$$\psi_3(x) = \psi^1(x_1)\psi^1(x_2) = \begin{cases} 1 & \text{if } 0 \leq x_1, x_2 < 1/2 \text{ or } 1/2 \leq x_1, x_2 < 1 \\ 0 & \text{if } x \in \mathbb{R}^n \backslash [0,1)^2 \\ -1 & \text{otherwise.} \end{cases}$$

6 Besov Spaces

The Besov spaces $B_p^{s,q}$ are generalizations of other spaces such as the Sobolev and Hölder spaces. There are various ways of defining the spaces $B_p^{s,q}$ such as in the context of Littlewood-Paley analysis or in terms of wavelet coefficients. For our purposes we will be using the definition found in [15], which gives these spaces in terms of the wavelet expansions of their functions. Before we do so, the following definition will be needed.

Definition 6.1 *A multiresolution analysis $\{V_j;\ j \in \mathbb{Z}\}$ is called r-regular ($r \in \mathbb{N}$) if the associated scaling function ϕ satisfies*

$$|\partial^\alpha \phi(x)| \leq C_m(1+|x|)^{-m}$$

for each integer $m \in \mathbb{N}$ and for every multi-index $\alpha = (\alpha_1, ..., \alpha_n)$ such that $|\alpha| \leq r$ (where $\partial^\alpha = \partial^{\alpha_1}/\partial x_1 ... \partial^{\alpha_n}/\partial x_n$ and $|\alpha| = \alpha_1 + ... + \alpha_n$).

For any $p, q \in [1, \infty]$, and any $s \in \mathbb{R}$, let ϕ be the scaling function and $\psi_1, ..., \psi_{2^n-1}$ the mother wavelets of an r-regular MRA $\{V_j\}$ of $L^2(\mathbb{R}^n)$ such that $|s| \leq r$. Then setting $\Lambda_j = \{j\} \times \mathbb{Z}^n \times \{1, ..., 2^n - 1\}$ and $\psi_\lambda(x) = 2^{jn/2}\psi_m(2^j x - k)$ for $\lambda = (j, k, m)$, $j \in \mathbb{Z}$, $k \in \mathbb{Z}^n$, $m \in \{1, ..., 2^n - 1\}$, the generalized functions $f \in B_p^{s,q}(\mathbb{R}^n)$ are those that can be written as

$$f(x) = \sum_{k \in \mathbb{Z}^n} \tilde{c}(k)\phi(x-k) + \sum_{j \geq 0} \sum_{\lambda \in \Lambda_j} c(\lambda)\psi_\lambda(x),$$

with

$$\left(\sum_{k \in \mathbb{Z}^n} |\tilde{c}(k)|^p \right)^{1/p} < \infty,$$

and for $j \geq 0$,

$$2^{js} 2^{nj(1/2-1/p)} \left(\sum_{\lambda \in \Lambda_j} |c(\lambda)|^p \right)^{1/p} = \epsilon_j$$

with $\epsilon_j \in l^q(\mathbb{Z})$. Similarly, a generalized function f belonging to the homogeneous Besov space $\dot{B}_p^{s,q}$ is characterized by

$$f(x) = \sum_{j \in \mathbb{Z}} \sum_{\lambda \in \Lambda_j} c(\lambda)\psi_\lambda(x),$$

where for any $j \in \mathbb{Z}$,

$$2^{js}2^{nj(1/2-1/p)}\left(\sum_{\lambda\in\Lambda_j}|c(\lambda)|^p\right)^{1/p} = \epsilon_j$$

with $\epsilon_j \in l^q(\mathbb{Z})$. We should point out that, as it is given in [15], a function belongs to a particular Besov space if and only if the above conditions hold on the wavelet coefficients of their functions. This means that if are defining a Besov space in an independent way, such as by means of a Littlewood-Paley analysis, then we get an equivalent space when defining the space in terms of wavelet coefficients, which is independent of the choice of wavelet basis provided the analyzing wavelet satisfies the forementioned conditions.

For the most part, we will only be concerned with some special cases of the homogeneous spaces $\dot{B}_p^{s,q}$. From [15], the dual of the space $\dot{B}_p^{s,q}$ where $s \in \mathbb{R}$, $p, q \in [1, \infty)$ is the space $\dot{B}_{p'}^{s',q'}$ where $s' = -s$ and $1/p + 1/p' = 1/q + 1/q' = 1$. We now treat some specific exanples. Suppose we have an MRA consisting of C^2 or smoother wavelets ψ_λ with compact support (which implies the MRA is 2-regular). Then it is easily verified that the space $\dot{B}_1^{1,1}(\mathbb{R}^2)$ is the space of generalized functions

$$f(x) = \sum_{j\in\mathbb{Z}}\sum_{\lambda\in\Lambda_j} c(\lambda)\psi_\lambda(x),$$

such that

$$\sum_{j\in\mathbb{Z}}\sum_{\lambda\in\Lambda_j}|c(\lambda)| < \infty,$$

i.e. the coefficients of f belong to $l^1(\mathbb{Z}^3 \times F)$ where $F = \{1, 2, 3\}$. The dual space of $\dot{B}_1^{1,1}(\mathbb{R}^2)$ is $\dot{B}_\infty^{-1,\infty}(\mathbb{R}^2)$, which is easily verified to be the space of all generalized functions whose wavelet coefficients belong to $l^\infty(\mathbb{Z}^3 \times F)$. The l^1 and l^∞ norms of the coefficients are equivalent to the usual Besov norms of $\dot{B}_1^{1,1}(\mathbb{R}^2)$ and $\dot{B}_\infty^{-1,\infty}(\mathbb{R}^2)$, respectively.

Remark. We have been referring to any $f \in \dot{B}_p^{s,q}(\mathbb{R}^n)$ as being a "generalized" function. By this we mean that $f \in \mathcal{S}'(\mathbb{R}^n)$, however this is not entirely true. When $s = n/p$ and $q > 1$, or $s > n/p$, then the space $\dot{B}_p^{s,q}(\mathbb{R}^n)$ becomes a space of tempered distributions modulo polynomials of degree $\leq m$, where m is the integer part of $s - n/p$. These issues will not be of any concern to us.

We end this section by referring [5] to a theorem showing an important relationship between the homogeneous Besov spaces and the homogeneous Sobolev spaces. This theorem will be revisited later.

Theorem 6.1 *For $\alpha \in \mathbb{R}$ and $1 \leq p \leq \infty$ we have the inclusion*

$$\dot{B}_p^{\alpha,1} \subset \dot{W}^{\alpha,p}.$$

Furthermore, this embedding is a continuous one.

We should point out that we have not discussed the meaning of $\dot{W}^{\alpha,p}$ when α takes noninteger values. These spaces however, referred to as potential spaces, will not be of concern to us and hence we will not bother defining them here.

7 Wavelets and Image Processing

7.1 Wavelets in Practice: The Fast Wavelet Transform

The Fast Wavelet Transform (FWT), analagous to the Fast Fourier Transform, is an algorithm used to find wavelet coefficients in a recursive fashion. We use here the FWT implemented in [9]. This algorithm will be shown in one dimension but can certainly be done in higher dimensions as well. In numerical applications, one starts with a function f (which we will assume belongs to $L^2(\mathbb{R})$). At this point f is approximated by a function S_j belonging to the approximation space V_j, where j is chosen large enough so that the L^2 error between f and S_j is as small as desired. We now have replaced f by a function

$$S_j = \sum_{k\in\mathbb{Z}} s(j,k)\phi_{j,k},$$

where the coefficients $\{s(j,k)\}_{k\in\mathbb{Z}} \subset l^2(\mathbb{Z})$ are obtained from f in some suitable way which we do not discuss for the time being.

We would like to find the coefficients in the corresponding wavelet representation of S_j. Noting that $S_j \in V_j$ and denoting the orthogonal projections of $L^2(\mathbb{R})$ onto V_j and W_j by P_j and Q_j respectively, we have

$$\begin{aligned} S_j = P_jS_j &= P_0S_j + \sum_{i=1}^{j} P_iS_j - P_{i-1}S_j \\ &= P_0S_j + \sum_{i=1}^{j} Q_{i-1}S_j \\ &= P_0S_j + \sum_{i=1}^{j}\sum_{k\in\mathbb{Z}} d(i,k)\psi_{i,k}, \end{aligned}$$

where $d(i,k)$ are the wavelet coefficients. We now show an iteration for finding the coefficients $\{s(j-1,k)\}_{k\in\mathbb{Z}}$ of $P_{j-1}S_j$. We have

$$s(j-1,k) = \langle S_j, \phi_{j-1,k}\rangle = \int_{\mathbb{R}} \left(\sum_{m\in\mathbb{Z}} s(j,m)\phi_{j,m}\right)\phi_{j-1,k},$$

but from the dilation equation (5.10) we know that

$$\begin{aligned}
\phi_{j-1,k}(x) &= 2^{(j-1)/2}\phi(2^{j-1}x-k) \\
&= 2^{j/2}\sum_{i\in\mathbb{Z}} c(i)\phi(2^jx-2k-i) \\
&= \sum_{i\in\mathbb{Z}} c(i)\phi_{j,2k+i}(x),
\end{aligned}$$

where $c(i)$ are the coefficients given in (5.10). Hence

$$\begin{aligned}
s(j-1,k) &= \int_{\mathbf{R}}\Big(\sum_{m\in\mathbb{Z}} s(j,m)\phi_{j,m}\Big)\Big(\sum_{i\in\mathbb{Z}} c(i)\phi_{j,2k+i}\Big) \\
&= \sum_{m\in\mathbb{Z}} c(m-2k)s(j,m).
\end{aligned}$$

Hence the sequence $s_{j-1} := \{s(j-1,k)\}_{k\in\mathbb{Z}}$ is obtained from the sequence $s_j := \{s(j,k)\}_{k\in\mathbb{Z}}$ by the matrix multiplication

$$s_{j-1} = As_j, \quad \text{where} \ \ A = (a_{k,m}) \ \ \text{and} \ \ a_{k,m} = c(m-2k).$$

We also wish to obtain the coefficients $d(j-1,k)$, $k\in\mathbb{Z}$, of $Q_{j-1}S_j$ from the coefficients $s(j,k)$, $k\in\mathbb{Z}$. To do so, we see that

$$\begin{aligned}
d(j-1,k) &= \langle S_j, \psi_{j-1,k}\rangle \\
&= \int_{\mathbf{R}}\Big(\sum_{m\in\mathbb{Z}} s(j,m)\phi_{j,m}\Big)\Big(\sum_{l\in\mathbb{Z}} b(l)\phi_{j,2k+l}\Big) \\
&= \sum_{m\in\mathbb{Z}} s(j,m)b(m-2k),
\end{aligned}$$

where $b(i) = (-1)^{i-1}\overline{c(-i-1)}$, and the second equality follows from Theorem 5.1. Hence we obtain the sequence $d_{j-1} := \{d(j-1,k)\}_{k\in\mathbb{Z}}$ by the matrix multiplication

$$d_{j-1} = Bs_j, \quad \text{where} \ \ B = (b_{k,m}) \ \ \text{and} \ \ b_{k,m} = b(m-2k).$$

So we see that the coefficients s_{j-i} and d_{j-i} are obtained iteratively from the coefficients s_{j-i+1} by multiplication with A and B, respectively.

We would also like to show how to reconstruct S_j given that we know the coefficients of P_0S_j and Q_kS_j for $k=0,1,...,j-1$. For this purpose we need to be able to rewrite an element $S = \sum_{k\in\mathbf{Z}} s(j,k)\phi_{j,k}$ belonging to V_j as an element $S = \sum_{k\in\mathbf{Z}} s(j+1,k)\phi_{j+1,k}$ in V_{j+1}. From the dilation equation (5.10), we have

$$S = \sum_{k\in\mathbf{Z}} s(k)\left[\sum_{l\in\mathbf{Z}} \phi_{j+1,2k+l}\right] = \sum_{i\in\mathbf{Z}}\left[\sum_{k\in\mathbf{Z}} s(k)c(i-2k)\right]\phi_{j+1.i}.$$

Thus the sequence $s_{j+1} := \{s(j+1)\}$ is obtained from $s_j := \{s(j)\}$ by multiplication with the transpose A^* of A:

$$s_{j+1} = A^* s_j \quad \text{where} \quad A^* = (a^*_{k,m}),\ a^*_{k,m} = c(k-2m).$$

A similar calculation tells us how to rewrite a function $S = \sum_{k\in\mathbf{Z}} d(j,k)\psi_{j,k}$ as $S = \sum_{k\in\mathbf{Z}} s'(j+1,k)\phi_{j+1,k}$. We have multiplication by the transpose B^* of B:

$$s'_{j+1} = B^* d_j \quad \text{where} \quad B^* = (b^*_{k,m}),\ b^*_{k,m} = b(k-2m).$$

This algorithm thus allows us to rewrite

$$S_j = P_0 S_j + \sum_{i=1}^{j} Q_{i-1} S_j$$

as

$$S_j = \sum_{k\in\mathbf{Z}} s(j,k)\phi_{j,k}.$$

7.2 Image Compression: Some Basic Terminology

To begin, we will need to know some backround information about image processing and some basic terminology that is used. First of all, a black and white image is represented by its grey-level $f(x) = f(x_1, x_2)$, which describes the magnitute of light intensity at the point (x_1, x_2). That is, we have $0 \le f(x) \le 255$ where $f(x) = 0$ means the image is black at x and $f(x) = 255$ means that the image is white at x. It is also assumed that the energy of the image is finite (i.e. $f \in L^2(\mathbb{R}^2)$). In real life application, an image f is no longer a continuum of points but is represented by a discrete set of points corresponding to a fine grid, say $N \times N$, where N is a large integer. Then f is now represented by a matrix $(f_{m,n})$ where $m, n \in \{0, 1, ..., N-1\}$. This discretization of an image, which is referred to as **sampling**, can be done simultaneously with the process of representating the image by a wavelet basis. To be more precise, let ϕ be the scaling function of an MRA of $L^2(\mathbb{R}^2)$. Ideally we would like to choose ϕ to be smooth (say C^2) and having compact support. Then choosing n large enough so that the desired accuracy rate is obtained, we write

$$f \approx \sum_{k\in\mathbb{Z}^2} a_k \phi_{n,k}, \tag{7.1}$$

where sampling is now done by choosing the a_k in some numerical way from f. If f has compact support then the series in (7.1) should obviously be finite, otherwise the series may be truncated since it is assumed that $f \in L^2(\mathbb{R}^2)$. As for choosing values of the coefficients a_k, a typical choice is to let $a_k = f(2^{-n}k)$ since $2^{-n}k$ most likely corresponds to the support of $\phi_{n,k}$. Once the values for a_k are known, one can apply the FWT described in section 7.1 to obtain the coefficients $\tilde{c}(k)$ and $c(\lambda)$ in the wavelet expansion of f:

$$f = \sum_{k\in\mathbb{Z}^2} \tilde{c}(k)\phi_{0,k} + \sum_{\lambda\in\Lambda} c(\lambda)\psi_\lambda$$

where $\Lambda = \mathbb{Z} \times \mathbb{Z}^2 \times \{1,2,3\}$ and the first series, which contains very few terms, is of little cost to accept into the expansion. The image can then be compressed substantially through quantization and thresholding. To perform **quantization**, coefficients that are close to some fixed coefficients are replaced by those fixed coefficients. For example, partitioning the real line by intervals $I_n = [a_n, b_n)$ each of length 2ϵ one could apply a quantization operator, given by

$$Q(x) = \frac{b_n - a_n}{2} \quad \text{for } x \in I_n,$$

to each of the coefficients $c(j,k)$. The process of **thresholding** is done by replacing coefficients smaller than some carefully chosen parameter $\tau > 0$ (referred to as the threshold) by zero. In a hard threshold all other coefficients remain unchanged, while in a soft threshold, they are moved closer to zero. More precisely, the hard and soft thresholding operators are defined as

$$\Theta_\tau^h(x) = \begin{cases} x & if\ |x| \geq \tau \\ 0 & if\ |x| < \tau \end{cases}$$

and

$$\Theta_\tau^s(x) = \begin{cases} x - \tau(sign(x)) & if\ |x| \geq \tau \\ 0 & if\ |x| < \tau. \end{cases}$$

respectively.

7.3 Wavelet Based Methods for Solving Minimization Problems

A very useful application of wavelet analysis can be seen by solving minimization problems such as that given by (4.1). Before dealing with the ROF model, let us consider a similar problem of minimizing the functional

$$J(u) = \|u\|_{\dot{B}_1^{1,1}} + \lambda \|f - u\|_2^2 \tag{7.2}$$

in the 2-dimensional case. We have seen that if a 2-regular wavelet basis with mother wavelet ψ is used then the coefficients c_γ of a function in $\dot{B}_1^{1,1}$ belong to $l^1(\Lambda)$, (where Λ denotes the set of indices for the wavelet basis) and the l^1 norm of the coefficients is equivalent to the usual Besov norm. Instead of solving the problem given by (7.2) we can solve the equivalent discrete problem of finding a sequence $\{u_\gamma\}$ in l^1 which minimizes

$$\tilde{J}(u) = \sum_{\gamma \in \Lambda} |u_\gamma| + \lambda |f_\gamma - u_\gamma|^2, \tag{7.3}$$

where f_γ are the wavelet coefficients of f. The exact solution to (7.3) is given by a soft thresholding with a threshold of $1/2\lambda$. To see this, we minimize (7.3) for each index γ, i.e. we minimize the function

$$h(u_\gamma) = |u_\gamma| + \lambda |f_\gamma - u_\gamma|^2$$

for each γ. It is obvious that if $f_\gamma \geq 0$ then so is the solution $\tilde{u}_\gamma$ which minimizes h, otherwise $h(-\tilde{u}_\gamma) \leq h(\tilde{u}_\gamma)$ which is a contradiction. A similar argument shows that $f_\gamma \leq 0$ implies $\tilde{u}_\gamma \leq 0$. With this fact in mind, suppose $f_\gamma \geq 0$. Then for $u_\gamma \in [0, \infty)$ we have $h'(u_\gamma) = 0$ if and only if $u_\gamma = f_\gamma - 1/2\lambda$. For $0 \leq f_\gamma < 1/2\lambda$ we have $f_\gamma - 1/2\lambda < 0$ and hence $\inf h(u_\gamma) = h(0)$ for $u_\gamma \in [0, \infty)$. Hence for $f_\gamma \geq 0$, the solution which minimizes $h(u_\gamma)$ is given by

$$\tilde{u}_\gamma = \max\{0, f_\gamma - 1/2\lambda\}.$$

A similar calculation shows that if $f_\gamma \leq 0$ then $\tilde{u}_\gamma$ is given by

$$\tilde{u}_\gamma = -\max\{0, |f_\gamma| - 1/2\lambda\}.$$

Combining these last two results, we have, for any $f_\gamma \in \mathbb{R}$,

$$\tilde{u}_\gamma = \operatorname{sign}(f_\gamma) \max\{0, |f_\gamma| - 1/2\lambda\},$$

which is presicely the result one would get when applying a soft thresholding with threshold $1/2\lambda$.

Such a nice algorithm for minimizing (7.2) may lead one to suggest using this functional instead of the ROF functional to decompose an image into a $u+v$ sum. However, by Theorem 6.1 we have $\dot{B}_1^{1,1} \subset \dot{W}^{1,1}$ and hence characteristic functions of sets with smooth boundary do not belong to $\dot{B}_1^{1,1}$. As a consequence of this last remark, the edges of objects, which correspond to the boundaries of characteristic functions, will not be well preserved in the u component. Minimization of the ROF functional by wavelet thresholding will be discussed in section 7.6.

7.4 Thresholding: Wavelets vs. Fourier Series

We would now like to address the issues of preformance of wavelet series expansions versus that of Fourier series expansions for representing images. We will soon see that wavelets are the better choice. The first problem we would like to shed light on is whether or not thresholding is a stable operation when performed on both wavelet and Fourier expansions of functions. If wavelets having compact support are used, then they have the benefit that changes made to a particular coefficient $c(j,k)$ will only effect the function in the support of $\psi(2^jx-k)$. However, the sine and cosine basis functions used in a Fourier series are supported globally, and hence any change to a single coefficient will effect the function on the entire domain. Meyer proved an interesting result in [13]. Given a hard thresholding operator Θ_ϵ with threshold ϵ and any Hölder space C^α with $\alpha < 1/2$, one can find a 2π-periodic function $f \in C^\alpha$ such that

$$\|\Theta_\epsilon(f)\|_\infty \to \infty \quad \text{as} \quad \epsilon \to 0^+,$$

where the thresholding is applied to the Fourier coefficients of f. To see that this problem does not occur when thresholding the wavelet coefficients of a function in C^α, we note, as it is given in [15], that the space C^α is exactly the space $B_\infty^{\alpha,\infty}$ and we have seen that a function f belongs to the latter space if and only if

$$f(x) = \sum_{k\in\mathbf{Z}^n} \tilde{c}(k)\phi(x-k) + \sum_{j\geq 0}\sum_{\lambda\in\Lambda_j} c(\lambda)\psi_\lambda(x),$$

such that

$$\sup_{k\in\mathbf{Z}^n} |\tilde{c}(k)| \leq C_0 < \infty \tag{7.4}$$

and

$$\sup_{j\geq 0} 2^{j\alpha}2^{nj/2} \sup_{\lambda\in\Lambda_j} |c(\lambda)| \leq C_1 < \infty, \tag{7.5}$$

where $\Lambda_j = \{j\} \times \mathbb{Z}^n \times \{1, ..., 2^n - 1\}$ and the wavelets $\{\psi_1, ..., \psi_{2^n-1}\}$ are chosen to be r-regular where $|r| \geq 1$. Summing the infimum over all the constants C_0 and the infimum over all constants C_1 which satisfy (7.4) and (7.5) respectively, one obtains an equivalent norm to the usual norm in C^α. Hence if the wavelet coefficients $c(\lambda)$ of f were replaced by $\Theta_\epsilon(c(\lambda))$ then they would still satisfy (7.5) and the resulting function f_ϵ would satisfy

$$\|f_\epsilon\|_{C^\alpha} \leq C\,\|f\|_{C^\alpha},$$

with the constant C not depending on ϵ. We remark that the coefficients $\tilde{c}(k)$, $k \in \mathbb{Z}^n$, are left untouched when applying wavelet thresholding to functions belonging to nonhomogeneous spaces such as the one just described.

7.5 Expansions of BV Functions: Wavelets vs. Fourier Series

When dealing with functions of bounded variation, we will show that it is beneficial to use wavelet series as opposed to Fourier series. The Fourier coefficients of a BV function can decay as poorly as $O(n^{-1/2})$ if logarithmic factors are ignored. An example of such a function in 2-dimensions is given by

$$f(x) = |x|^{-1}(\log|x|)^{-2}\varphi(|x|),$$

where $\varphi \in C^\infty$ with $\varphi(x) = 0$ for $|x| > 1/2$ and $\varphi(x) = 1$ for $|x| < 1/4$. First of all, we show that this function does in fact belong to BV. Since f is radial we write in polar co-ordinates,

$$f(r) = r^{-1}(\log r)^{-2}\varphi(r),$$

and differentiating we have

$$f'(r) = \frac{\varphi'(r) r \log r - \varphi(r)\log r + 2\varphi(r)}{r^2(\log r)^3}.$$

Hence

$$\begin{aligned}
\iint_{\mathbb{R}^2} |\vec{\nabla} f(x)| dx &= \int_0^{2\pi}\int_0^\infty |f'(r)| r dr d\theta \\
&= 2\pi \int_{1/4}^{1/2} |f'(r)| r dr + 2\pi \int_0^{1/4} \frac{\log r - 2}{r(\log r)^3} dr.
\end{aligned}$$

The integral

$$\int_{1/4}^{1/2} |f'(r)| r dr$$

is obviously finite. As for the second integral we have

$$\begin{aligned}
\int_0^{1/4} \frac{\log r - 2}{r(\log r)^3} dr &= \int_0^{1/4} \frac{1}{r(\log r)^2} dr - \int_0^{1/4} \frac{2}{r(\log r)^3} dr \\
&= \frac{-1}{\log r}\Big|_0^{1/4} + \frac{1}{(\log r)^2}\Big|_0^{1/4} \\
&= \frac{1}{(\log 4)} + \frac{1}{(\log 4)^2} < \infty.
\end{aligned}$$

Thus we have $f \in \dot{W}^{1,1} \subset BV$. According to [13], the Fourier coefficients $c(k)$ of f can be estimated by

$$\begin{aligned}
c(k) &\approx C|k|^{-1}(\log|k|)^{-2} \\
&= \frac{4C}{\sqrt{k_1^2+k_2^2}[\log(k_1^2+k_2^2)]^2},
\end{aligned}$$

which implies that the nonincreasing rearrangement of the $|c(k)|$ denoted by c_n^*, decay as $n^{-1/2}(\log n)^{-2}$. As far as wavelet bases are concerned, we are guaranteed a better decay on coefficients than what we have just seen. Given an o.n.b. $2^j\psi(2^jx-k)$, $j \in \mathbb{Z}$, $k \in \mathbb{Z}^2$, $\psi \in \{\psi_1, \psi_2, \psi_3\}$ of $L^2(\mathbb{R}^2)$ constructed by usual tensor product from a one-dimensional wavelet ψ, where ψ has compact support and belongs to a Hölder space C^r for some $r > 0$, then we have the following theorem.

Theorem 7.1 *For any $f \in BV(\mathbb{R}^2)$, the wavelet coefficients $c_\lambda = \langle f, \psi_\lambda \rangle$, $\lambda \in \Lambda = \mathbb{Z} \times \mathbb{Z}^2 \times \{1,2,3\}$ belong to weak $l^1(\Lambda)$.*

This theorem was first proved in [7] for the Haar wavelets and was later extended to the general case in [13]. This theorem says that if the $|c_\lambda|$ are rearranged in nonincreasing order, then these sorted coefficients satisfy

$$c_n^* \le \frac{C}{n}$$

for some constant C.

7.6 The ROF Model: Revisited

We would now like to attempt to use wavelet methods in order to minimize the ROF functional:

$$J(u) = \|u\|_{BV} + \lambda \|v\|_2^2 ,$$

where $f = u + v$. We have already seen that when replacing the BV-norm in the above problem with the norm in the smaller space $\dot{B}_1^{1,1}$, an exact solution (up to a constant that gives the equivalence of the discrete and continuous norms) is given by a soft thresholding of level $1/2\lambda$ on the wavelet coefficients of f.

Let us first recall that $\|v\|_* \leq 1/2\lambda$, where $\|\cdot\|_*$ denotes the norm in the space G. This is a direct consequence of Theorems 4.2 and 4.3. On the other hand, by Lemma 4.1 and Theorem 6.1, we have the following estimate for the G-norm:

$$\|v\|_{\dot{B}_\infty^{-1,\infty}} \leq C_0 \|v\|_* \tag{7.6}$$

for some constant $C_0 > 0$. This is a very appealing statement since we know that the norm of a function f in the Besov space $\dot{B}_\infty^{-1,\infty}$ is equivalent to the l^∞ norm on the wavelet coefficients of f. Thus if v_γ, $\gamma \in \Lambda = \mathbb{Z} \times \mathbb{Z}^2 \times \{1, 2, 3\}$, are the wavelet coefficients of the v component of the ROF model then

$$\sup_{\gamma \in \Lambda} |v_\gamma| \leq \frac{C}{2\lambda}$$

for some $C > 0$. By using either a hard or soft thresholding at level $\epsilon = \epsilon_\lambda = C/2\lambda$ the v component of the ROF model will therefore be completely removed. We will denote the resulting function from this thresholding by $\hat{u}$ and also let $\hat{v} = f - \hat{u}$. The error of approximating u by $\hat{u}$ can be measured in the L^2 norm. To see this, we first denote the wavelet coefficients of u by u_γ. Since $BV(\mathbb{R}^2) \subset L^2(\mathbb{R}^2)$, we have

$$\sum_{\gamma \in \Lambda} |u_\gamma|^2 < \infty.$$

Then since $|\theta_\epsilon(u_\gamma)| \leq |u_\gamma|$, where θ_ϵ is a thresholding operator (either hard or soft), it is obvious that $\hat{u}$ also belongs to L^2. Thus the L^2 error can be estimated by

$$\|u - \hat{u}\|_2 \leq C(\sum_{\gamma\in\Lambda} |u_\gamma - \hat{u}_\gamma|^2)^{1/2},$$

where $\hat{u}_\gamma = \theta_\epsilon(u_\gamma)$. Let us denote by u_n^* the nonincreasing rearrangement of the sequence $|u_\gamma|$, $\gamma \in \Lambda$. Then if θ_ϵ represents a hard threshold operator, and if N is the least integer such that

$$u_N^* \leq \frac{C_0}{N} \leq \epsilon,$$

then the above estimate simplifies to

$$\begin{aligned}\sum_{\gamma\in\Lambda} |u_\gamma - \hat{u}_\gamma|^2 &= \sum_{n=N}^{\infty} |c_n^*|^2 \\ &\leq C_0^2 \sum_{n=N}^{\infty} \frac{1}{n^2}.\end{aligned}$$

Similarly when θ_ϵ is a soft threshold operator,

$$\begin{aligned}\sum_{\gamma\in\Lambda} |u_\gamma - \hat{u}_\gamma|^2 &= \sum_{|c_\gamma|>\epsilon} |u_\gamma - \hat{u}_\gamma|^2 + \sum_{|c_\gamma|\leq\epsilon} |u_\gamma - \hat{u}_\gamma|^2 \\ &\leq \sum_{n=1}^{N-1} |\epsilon|^2 + C_0^2 \sum_{n=N}^{\infty} \frac{1}{n^2} \\ &\leq \sum_{n=1}^{N-1} \frac{C_0^2}{(N-1)^2} + C_0^2 \sum_{n=N}^{\infty} \frac{1}{n^2} \\ &\leq \frac{C_0}{N-1} + C_0^2 \sum_{n=N}^{\infty} \frac{1}{n^2}.\end{aligned}$$

Although this is fine for measuring the error of $u - \hat{u}$ in L^2, unfortunately it does not say anything about this error in the BV norm. This is due to the fact that the wavelet coefficients of a BV function are not characterized by their coefficients belonging to weak l^1, i.e., the converse to Theorem 7.1 is not true in general. This issue leads to the formulation of the following problem.

Conjecture 7.1 *If $f(x) = \sum_{\gamma\in\Lambda} c_\gamma \psi_\gamma(x)$ is the wavelet expansion of a function f belonging to BV and if θ_ϵ is a hard or soft thresholding operator of level ϵ then*

$$f_\epsilon(x) = \sum_{\gamma\in\Lambda} \theta_\epsilon(c_\gamma)\psi_\gamma(x)$$

still belongs to BV with a norm not depending on ϵ.

Although this has not been proved for the general case, it was however proved for the Haar wavelets by Cohen et al. in [7]. More precicely, they proved that

$$\|f_\epsilon\|_{BV} \leq C \|f\|_{BV},$$

where

$$C = 10 + 28\sqrt{2}\left[18\sqrt{3} + 36(480\sqrt{5} + 168\sqrt{3})\right].$$

References

[1] Acar, R. and Vogel, C. *Analysis of bounded variation penalty methods.* Inverse Problems, v.10, pp.1217-1229, 1994.

[2] Ali, S.T., Antoine, J.P., Gazeau, J.P. *Coherent States, Wavelets, and Their Generalizations.* Springer-Verlag, New York, 2000.

[3] Aujol, Jean-Francois. *SAR Images Restoration.* ⟨http://www.sop.inria.fr/ariana/DEMOS/demosar/node3.html⟩. June, 2003.

[4] Bachman et al. *Fourier and Wavelet Analysis.* Springer-Verlag, New York, 2000.

[5] Bergh, J. and Löfström, J. *Interpolation Spaces: An Introduction.* Springer-Verlag, 1976.

[6] Chui, Charles K. *An Introduction to Wavelets.* Wavelet Analysis and its Applications, v.1, Academic Press. San Diego, California, 1992.

[7] Cohen, A., DeVore, R., Petrushev, P., and Xu, H. *Nonlinear Approximation and the Space* $BV(\mathbb{R}^2)$. American Journal of Mathematics, v.121, pp.587-628, 1999.

[8] Daubechies, Ingrid. *Ten Lectures on Wavelets.* CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Sciences. Philadelphia, Pennsylvania, 1992.

[9] DeVore, R. A. and Lucier, B. J. *Wavelets.* Acta Numerica, Cambridge University Press, v.1, pp.1-56, 1992.

[10] Evans, L. C. and Gariepy, R. F. *Measure theory and fine properties of functions.* Studies in Advanced Mathematics. CRC Press, Boca Raton, Florida, 1992.

[11] Giusti, Enrico. *Minimal Surfaces and Functions of Bounded Variation.* Monographs in Mathematics, v.80, Birkhäuser Boston, 1984.

[12] Mallat, Stéphane. *A Wavelet Tour of Signal Processing.* Academic Press, 1998.

[13] Meyer, Yves. *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations.* The Fifteenth Dean Jacqueline B. Lewis Memorial Lectures. University Lecture Series, v.22, 2001.

[14] Meyer, Yves. *Variational methods in Image Processing.* Notes from the Conference in Honor of Haim Brezis. Paris, June 2004.

[15] Meyer, Yves. *Wavelets and Operators.* Cambridge Studies in Advanced Mathematics, v.37, Paris, 1992.

[16] Osher, S., Vese, L. *Modeling Textures with Total Variation Minimization and Oscillating Patterns in Image Processing.* Journal of Scientific Computing, v.19, Issues 1-3, pp.553-572, December, 2003.

[17] Osher, S., Rudin, L., Fatemi, E. *Nonlinear Total Variation Based Noise Removal Algorithms.* Physica D, v.60, pp.259-268, 1992.

[18] Rudin, W. *Real and Complex Analysis.* Third Edition. McGraw-Hill Series in Higher Mathematics, 1987.

[19] "Wavelet." Wikipedia. May 10, 2006. ⟨http://www.wikipedia.org⟩.

[20] Ziemer, William P. *Weakly Differentiable Functions: Sobolev Spaces and Functions of Bounded Variation.* Graduate Texts in Mathematics. Springer-Verlag, New York, 1989.