Wavelet decompositions of non-refinable shift invariant spaces

S. Dekel and D. Leviatan

Abstract: The motivation for this work is a recently constructed family of generators of shift-invariant spaces with certain optimal approximation properties, but which are not refinable in the classical sense. We try to see whether, once the classical refinability requirement is removed, it is still possible to construct meaningful wavelet decompositions of dilates of the shift invariant space that are well suited for applications.

AMS Subject classifications: 42C40, 65T60, 41A25, 41A44

Key words: wavelets, non-stationary wavelets, shift-invariant spaces, approximation order.

Authors affiliation:

School of Mathematical sciences
Tel-Aviv University
Tel-Aviv 69978, Israel

E-mail:

S. Dekel: shai.dekel@turboimage.com
D. Leviatan: leviatan@math.tau.ac.il

1 Introduction

In classical refinable wavelet theory ([Ch], [Da], [M]) one begins with a **finitely generated shift invariant** (FSI) space $S(\Phi) := \overline{span}\{f(\cdot - k) \mid f \in \Phi, k \in \mathbb{Z}^d\}$, where Φ is a finite set and the closure is taken in some Banach space X. Typically, $S(\Phi)$ is selected to have **approximation order** $m \in \mathbb{N}$. This means that for any $h > 0$ and $f \in X$

$$E\left(f, S(\Phi)^h\right)_X := \inf_{g \in S(\Phi)^h} \|f - g\|_X \le C h^m |f|_X, \tag{1.1}$$

where

$$S(\Phi)^h := \overline{span}\left\{f\left(h^{-1} \cdot - k\right) \mid f \in \Phi, k \in \mathbb{Z}^d\right\},$$

and $|\cdot|_X$ is a semi-norm, measuring the smoothness of the elements of X.

To allow the construction of wavelets associated with $S(\Phi)$, one assumes that the shift invariant space is **two-scale refinable**, namely

$$S(\Phi) \subset S(\Phi)^{1/2}. \tag{1.2}$$

One then selects a complementary set of generators, so called **wavelets**, Ψ so that

$$S(\Phi)^{1/2} = S(\Phi) + S(\Psi). \tag{1.3}$$

It is easy to see that (1.3) can be dilated to any given scale $J \in \mathbb{Z}$ that is,

$$S(\Phi)^{2^{-J}} = S(\Phi)^{2^{J+1}} + S(\Psi)^{2^{-J+1}}.$$

Assume $f_\Phi^J \in S(\Phi)^{2^{-J}}$ so that $f_\Phi^J = f_\Phi^{J-1} + f_\Psi^{J-1}$, where $f_\Phi^{J-1} \in S(\Phi)^{2^{-J+1}}$, $f_\Psi^{J-1} \in S(\Psi)^{2^{-J+1}}$. Then, f_Φ^{J-1} plays the role of a low resolution approximation to f_Φ^J, while f_Ψ^{J-1} is the difference between the two, the detail. Typically, if f_Φ^J is a sufficiently smooth function or J is sufficiently large, then $f_\Phi^{J-1} \approx f_\Phi^J$ and $f_\Psi^{J-1} \approx 0$. Under certain conditions (1.3) leads to a wavelet decomposition

$$S(\Phi)^{2^{-J}} = S(\Psi)^{2^{-J+1}} + S(\Psi)^{2^{-J+2}} + S(\Psi)^{2^{-J+3}} + \cdots, \tag{1.4}$$

i.e., any $f_\Phi^J \in S(\Phi)^{2^{-J}}$ possesses a decomposition

$$f_\Phi^J = f_\Psi^{J-1} + f_\Psi^{J-2} + f_\Psi^{J-3} + \cdots. \tag{1.5}$$

In applications, FSI spaces are used as follows. Let f be some signal that one wishes to approximate. Using property (1.1), one chooses a fine enough scale $J \in \mathbb{Z}$ and computes an approximation

$$f \approx f_\Phi^J \in S(\Phi)^{2^{-J}} . \tag{1.6}$$

In some applications there is no need to further decompose the approximation f_Φ^J into the wavelet sum (1.5). Typical examples are curve and surface (linear) approximations in CAGD or re-sampling in image processing. However, the wavelet decomposition (1.4) is effective in applications that require a compact representation of the signal such as compression, denoising, segmentation, etc.

Let $S(\Phi_0)$ be a non-refinable FSI space. Namely, $S(\Phi_0) \not\subset S(\Phi_0)^{1/2}$. There are many examples of non-refinable FSI spaces that perform well in approximations of type (1.6). In fact, there is an interesting recent construction [BTU] of shift invariant spaces that are "optimal" in some approximation theoretical sense and are not two-scale refinable. Nevertheless, we would still like to decompose the space $S(\Phi_0)^{2^{-J}}$ into a sum of difference (wavelet) spaces in the sense of (1.4) (see [CSW] for a different approach). Since our FSI space is not refinable we need to replace $S(\Phi_0)$ by a different space $S(\Phi_1)$ to play the role of a low resolution space and a (wavelet) space $S(\Psi_1)$ to serve as a difference space in a decomposition similar to (1.3), namely,

$$S(\Phi_0)^{1/2} = S(\Phi_1) + S(\Psi_1) .$$

In this work we show that such meaningful decomposition techniques exist. They allow us, to further decompose $S(\Phi_1)^{1/2} = S(\Phi_2) + S(\Psi_2)$ and so on and to obtain a non-stationary wavelet decomposition similar to (1.4), i.e.,

$$S(\Phi_0)^{2^{-J}} = S(\Psi_1)^{2^{-J+1}} + S(\Psi_2)^{2^{-J+2}} + S(\Psi_3)^{2^{-J+3}} + \cdots$$

Thus, the (non-stationary) sequence $\{\Phi_j\}$ is a means to obtain the non-stationary wavelet sequence $\{\Psi_j\}$. The sequence $\{\Phi_j\}$ is also used to determine the (linear) approximation properties of the wavelets. It is interesting to note that our techniques enable us to recover the stationary choice $\Phi_j = \Phi_0$, $\Psi_j = \Psi$, whenever $S(\Phi_0)$ is two-scale refinable and $S(\Phi_0)^{1/2} = S(\Phi_0) + S(\Psi)$.

Another interesting question addressed in this work is the following. Let $S(\Phi_0)$ be an "optimal" non-refinable FSI space under some approximation theoretical gauge. Obviously, if $S(\Phi_0)$ has an "optimal" approximation property, no constructed $S(\Phi_1) \subset S(\Phi_0)^{1/2}$ can inherit this exact property. One then asks how close are the approximation properties of $S(\Phi_1)$ to those of $S(\Phi_0)$? Another question is the following. In what way (if any) are wavelets that decompose dilations of "optimal" non-refinable FSI spaces better than known existing wavelets?

In Section 2 we present the basic theory on the structure of shift invariant spaces which serves as framework throughout the work. We also present some new "regularity" results that are required for the wavelet constructions in Section 3.1. In section 3 we construct non-stationary wavelet decompositions of shift invariant spaces which are not required to be two-scale refinable. There are two such constructions. The Superfunction wavelet construction described in Section 3.1 is inspired by the superfunction theory of [BDR1], [BDR2], [BDR3]. In Section 3.2 we introduce Cascade wavelets. Their construction exploits properties of the Cascade operator (see for example [Da]). In Section 4 we first present results on approximation from shift invariant spaces. We then proceed to justify the constructions of Section 3, by showing that our non-stationary sequence $\{\Phi_j\}$ inherits the approximation properties of the decomposed non-refinable shift invariant space. Consequently, the non-stationary wavelet sequence $\{\Psi_j\}$ span "detail spaces" and are therefore suitable for signal processing applications.

2 Shift invariant spaces

Shift invariant spaces are a special case of invariant subspaces in Banach spaces. Here we use the framework of [BDR2] and present results that are required for the constructions in Section 3.

Definition 2.1 For any $k \in \mathbb{Z}^d$ we denote the linear shift operator S_k by $S_k(f) := f(\cdot - k)$.

Definition 2.2 Let V be a closed subspace of $L_p(\mathbb{R}^d)$, $1 \le p \le \infty$. We say that V is a **shift invariant** (SI) space if it is invariant under the operators $\{S_k \mid k \in \mathbb{Z}^d\}$. We say that a set Φ **generates** V if $V = S(\Phi) := \overline{span}\{f(\cdot - k) \mid f \in \Phi, k \in \mathbb{Z}^d\}$. We say that V is a **finite shift invariant** (FSI) space, if there exists a finite generating set Φ, $|\Phi| = n$, such that $V = S(\Phi)$. In such a case we say that V is of **length** $\le n$. We denote $len(V) := \min\{|\Phi| \mid V = S(\Phi)\}$. An SI space V is called a **principal shift invariant** (PSI) space if $len(V) = 1$.

To approximate functions with arbitrary precision one uses dilates of shift invariant spaces. For a given subspace V and $h \in \mathbb{R}_+$ we denote by V^h the dilated space

$$V^h := \{f(\cdot / h) \mid f \in V\}.$$

We note that is if $S(j)$ is a PSI space, then for $j \ge 0$, $S(j)^{2^{-i}}$ is a FSI space of length 2^{dj}.

We now restrict our discussion to $L_2(\mathbb{R}^d)$. It is well known that Fourier techniques appear naturally in the analysis of SI spaces. The following is simple characterization of SI spaces in the Fourier domain.

Lemma 2.3 [BDR2] Let $S(\Phi)$ be an FSI subspace of $L_2(\mathbb{R}^d)$ and let $f \in L_2(\mathbb{R}^d)$. Then the following are equivalent:

1. $f \in S(\Phi)$.

2. There exist $\mathbb{T}^d$ – periodic functions $\{t_f\}$ such that $\hat{f} = \sum_{f\in\Phi} t_f \hat{f}$.

We see that we can regard the generators of an FSI space as vectors spanning a finite dimensional vector space, with periodic functions playing the role of coefficients in the representations. Thus, we turn to Fourier based techniques. For each $f \in L_2(\mathbb{R}^d)$ we denote

$$\hat{f}_{|w} := \left(\hat{f}(w+2\pi k)\right)_{k\in\mathbb{Z}^d}, \qquad w\in\mathbb{T}^d .$$

The **bracket operator** $[\ ]: L_2(\mathbb{R}^d)\times L_2(\mathbb{R}^d)\to L_1(\mathbb{T}^d)$ is defined by

$$\left[\hat{f},\hat{g}\right](w) := \left\langle \hat{f}_{|w},\hat{g}_{|w}\right\rangle_{l_2(\mathbb{Z}^d)}, \quad w\in\mathbb{T}^d .$$

It is easy to see that the Fourier expansion of $\left[\hat{f},\hat{g}\right]$ is

$$\left[\hat{f},\hat{g}\right](w) \sim \sum_{k\in\mathbb{Z}^d} \left\langle f, g(\cdot+k)\right\rangle_{L_2(\mathbb{R}^d)} e^{ikw}, \tag{2.1}$$

Observe that if f, g are compactly supported, then the bracket $\left[\hat{f},\hat{g}\right]$ is a trigonometric polynomial and so we have an equality in (2.1). For $f \in L_2(\mathbb{R}^d)$ the function $[f, f]\in L_1(\mathbb{T}^d)$ is called the **auto-correlation** of f . Auto-correlations play a major role in our analysis. They are used in the definitions of stability constants, error kernels and “fine” error estimation constants. Our analysis requires the following simple result on the convergence of auto-correlations.

Lemma 2.4 Assume that $r_j \underset{L_2(\mathbb{R}^d)}{\to} f$ such that $supp(f), supp(r_j) \subseteq \Omega$ where Ω is a bounded domain. Then for any $m \ge 0$ we have the convergence

$$\left[\hat{r}_j^{(m)}, \hat{r}_j^{(m)}\right] \underset{C(\mathbb{T}^d)}{\to} \left[\hat{f}^{(m)},\hat{f}^{(m)}\right]. \tag{2.2}$$

Proof It is easy to see that we also have $(\cdot)^m r_j \underset{L_2(\mathbb{R}^d)}{\to} (\cdot)^m f$ for any $m \ge 0$. By virtue of (2.1) we have that $\left[\hat{f}^{(m)},\hat{f}^{(m)}\right], \left[\hat{r}_j^{(m)}, \hat{r}_j^{(m)}\right]$ are trigonometric polynomials of uniformly bounded degree. Therefore, the convergence of the Fourier coefficients

$$\left(\left[\hat{r}_j^{(m)},\hat{r}_j^{(m)}\right]\right)_k = \left\langle(\cdot)^m r_j, (\cdot+k)^m r_j(\cdot+k)\right\rangle \underset{j\to\infty}{\to} \left\langle(\cdot)^m f, (\cdot+k)^m f(\cdot+k)\right\rangle = \left(\left[\hat{f}^{(m)},\hat{f}^{(m)}\right]\right)_k ,$$

implies the convergence (2.2).

♦

We now proceed to present “regularity” results for shift invariant spaces in $L_2(\mathbb{R}^d)$. The motivation for working with regular shift invariant spaces comes from applications where it is required to have a stable representation or approximation of signals. Stability implies that small changes in the input function do not change much the representation and small changes in the representation change the reconstructed function only a little. We begin with definitions and notions from [BDR2].

Let $S(\Phi)$ be an SI space**.** The **range** function associated with $S(\Phi)$ is

$$J_S(w) := span\left\{\hat{f}_{|w} \;\middle|\; f \in \Phi\right\} \tag{2.3}$$

The **spectrum** of $S(\Phi)$ is defined by

$$s\,S(\Phi) := \left\{ w \in \mathbb{T}^d \;\middle|\; \dim J_S(w) > 0 \right\}, \tag{2.4}$$

or equivalently

$$s\,S(\Phi) := \left\{ w \in \mathbb{T}^d \;\middle|\; \left[\hat{f},\hat{f}\right](w) \neq 0, \text{ for some } f \in \Phi \right\}.$$

It can be shown ([BDR2]) that the range and spectrum of an SI space are invariants of the space. In particular they do not depend on the generating set. If $\dim J_S(w) \equiv const$ a.e. we say that S is **regular**. Observe that regularity implies a full spectrum. In the other direction, a full spectrum implies regularity only in the PSI case. We say that Φ is a **basis** for S if for each $f \in S(\Phi)$ there are periodic functions t_f where $\hat{f} = \sum_{f \in \Phi} t_f \hat{f}$ and t_f are uniquely determined. Observe that if $\mathbb{T}^d \setminus s\,S(\Phi)$ is of positive measure then $S(\Phi)$ does not have a basis. The set Φ is called a **stable generating set** or a **stable basis** (for its span) if there exist constants $0 < A \le B < \infty$ such that for every $c = \{c_{f,k}\}_{f \in \Phi, k \in \mathbb{Z}^d} \in l_2(\Phi \times \mathbb{Z}^d)$

$$A\|c\|^2_{l_2(\Phi\times\mathbb{Z}^d)} \le \left\| \sum_{f\in\Phi, k\in\mathbb{Z}^d} c_{f,k} f(\cdot - k) \right\|^2_{L_2(\mathbb{R}^d)} \le B\|c\|^2_{l_2(\Phi\times\mathbb{Z}^d)}. \tag{2.5}$$

It can be shown that a stable basis is indeed a basis. Since stable bases are necessary for applications, the next result leads towards the construction of regular spaces.

Theorem 2.5 [BDR2] Let $S(\Phi)$ be an FSI space. Then $S(\Phi)$ is regular if and only if it contains a stable generating set. Furthermore, an FSI space is regular if and only if it is the orthogonal sum of $len(S(\Phi))$ regular PSI spaces.

We recall the connection between the definition of stability (2.5) and the notion of the range function (2.3) for the simple case of PSI spaces (see [RS] Theorem 2.3.6 for the general case of FSI spaces).

Theorem 2.6 [Me] A function $f \in L_2(\mathbb{R}^d)$ is stable iff there exist $0 < A \le B < \infty$ such that $A \le \left[\hat{f},\hat{f}\right] \le B$, a.e.

Assume that we have constructed a non regular FSI subspace $S(\Phi_m)$ of a regular FSI space $S(\Phi_n)$ so that

$$len\left(S(\Phi_m)\right) = m < n = len\left(S(\Phi_n)\right).$$

We can certainly define $S(\Psi)$ as the orthogonal complement of $S(\Phi_m)$ in $S(\Phi_n)$ such that

$$S(\Phi_m) \oplus S(\Psi) = S(\Phi_n).$$

But the decomposition will have two undesirable features. First, there is no choice of generators $\widetilde{\Phi_m}, \widetilde{\Psi}$ so that $S\left(\widetilde{\Phi_m}\right) = S(\Phi_m)$, $S\left(\widetilde{\Psi}\right) = S(\Psi)$ and $\left\{\widetilde{\Phi_m}, \widetilde{\Psi}\right\}$ is stable. Secondly, the decomposition may be somewhat redundant, namely, $len\left(S(\Psi)\right) > n-m$. We will show that this can be fixed by constructing $S(\Phi_m')$ such that $S(\Phi_m) \subseteq S(\Phi_m') \subset S(\Phi_n)$, $len\left(S(\Phi_m')\right) = m$ and $S(\Phi_m')$ is regular. In doing so we ensure that the orthogonal complement is also regular and of length $n-m$. Hence, such a correction can produce a stable and efficient decomposition of $S(\Phi_n)$.

Lemma 2.7 Let $S(\Phi)$ be a regular FSI space and let $r \in S(\Phi)$. Then there exists $j \in S(\Phi)$, such that $S(r) \subseteq S(j)$ and $S(j)$ is a regular PSI subspace of $S(\Phi)$.

Proof If $S(r)$ is regular, we are done. Otherwise, by Corollary 3.31 in [BDR2], we may assume the decomposition $S(\Phi) = \bigoplus_{i=1}^{n} S(f_i)$ so that each $S(f_i)$ is a (regular) PSI subspace and the shifts of f_i are an orthonormal basis for $S(f_i)$. Therefore there exists a unique representation $\hat{r} = \sum_{i=1}^{n} t_i \hat{f}_i$ with t_i periodic functions. Since $\left[\hat{f}_j, \hat{f}_k\right](w) = d_{j,k}$ for $1 \le j,k \le n$ we have that $[\hat{r}, \hat{r}] = \sum_{i=1}^{n} |t_i|^2$ and so $s\,S(r) = \bigcup_{i=1}^{n} \mathrm{supp}(t_i)$. Define $j \in S(\Phi)$ by

$$\hat{j} = t_1' \hat{f}_1 + \sum_{i=2}^{n} t_i \hat{f}_i, \qquad t_1'(w) = \begin{cases} 1 & w \in \mathbb{T}^d \setminus s\,S(r), \\ t_1(w) & else. \end{cases}$$

Then $[\hat{j}, \hat{j}] = |t_1'|^2 + \sum_{i=2}^{n} |t_i|^2$ and we can conclude the following. The space $S(j)$ is regular since

$$\begin{aligned} s\,S(j) &= \mathrm{supp}\left([\hat{j}, \hat{j}]\right) = \mathrm{supp}(t_1') \cup \bigcup_{i=2}^{n} \mathrm{supp}(t_i) \\ &= \left(\mathbb{T}^d \setminus s\,S(r)\right) \cup \mathrm{supp}(t_1) \cup \bigcup_{i=2}^{n} \mathrm{supp}(t_i) \end{aligned}$$

$$= \left(\mathbb{T}^d \setminus s\,S(r)\right) \cup s\,S(r)$$
$$= \mathbb{T}^d .$$

Finally, $\hat{r} = c_{sS(r)} \hat{j}$ implies that $S(r) \subseteq S(j)$.

♦

Lemma 2.8 Let V, U be FSI spaces where $V \subseteq U$. Then $len(V) \le len(U)$.

Proof This is a direct consequence of the fact that the shift and orthogonal projection into an SI space commute. This implies that if $\Phi = \{f_i\}$ generate U, then $\{P_V f_i\}$ generate V.

♦

Theorem 2.9 Let U be a regular FSI. Then for any FSI subspace $S(\Phi_m) \subseteq U$ of length m there exists a regular subspace $S\left(\Phi_m{}'\right)$ of length m such that $S(\Phi_m) \subseteq S\left(\Phi_m{}'\right) \subseteq U$.

Proof The proof is essentially a Gram-Schmidt type construction, where we construct the "correction" $S\left(\Phi_m{}'\right)$ as an orthogonal sum of regular PSI spaces. We use induction on the length $|\Phi_m| = m$. The case $m = 1$ follows by virtue of Lemma 2.7. Assume the claim is true for $k < m$. Denote $\Phi_{m-1} = \{f_1, \ldots, f_{m-1}\}$, where $\Phi_m = \{f_1, \ldots, f_m\}$. Then by the induction hypothesis there exists a regular FSI subspace $S\left(\Phi_{m-1}{}'\right)$ such that

$$S(\Phi_{m-1}) \subseteq S\left(\Phi_{m-1}{}'\right) \subset U ,$$

and $len\left(S\left(\Phi_{m-1}{}'\right)\right) = \left|\Phi_{m-1}{}'\right| = m-1$. By [BDR2] the orthogonal complement in U of $S\left(\Phi_{m-1}{}'\right)$, denoted by W_{m-1}, is a regular FSI space. Let $S(y_m) := P_{W_{m-1}} S(f_m)$. Observe that $S(y_m)$ is not trivial since this would imply $S(\Phi_m) \subseteq S\left(\Phi_{m-1}{}'\right)$ which by Lemma 2.8 contradicts $len(S(\Phi_m)) = m$. Using again Lemma 2.7, we can find a regular PSI space $S\left(f_m{}'\right)$ such that

$$S(y_m) \subseteq S\left(f_m{}'\right) \subseteq W_{m-1} .$$

Since by Theorem 2.5 the orthogonal sum of two regular FSI spaces is regular, we have that $S\left(\Phi_m{}'\right)$, $\Phi_m{}' = \Phi_{m-1}{}' \cup f_m{}'$ is a regular FSI subspace of U. To conclude, observe that $S\left(\Phi_m{}'\right)$ also possesses the required properties of minimal length, $len\left(S\left(\Phi_m{}'\right)\right) = \left|\Phi_m{}'\right| = m$ and that $S(\Phi_m) \subseteq S\left(\Phi_m{}'\right)$.

♦

Next we discuss the special structure of the orthogonal projection into SI spaces.

Lemma 2.10 [BDR2] Let Φ be a basis for an FSI space $S(\Phi)$ and let $f \in L_2(\mathbb{R}^d)$. Then the orthogonal projection $P_{S(\Phi)} f$ is given by

$$\widehat{P_{S(\Phi)} f} = \sum_{f \in \Phi} \frac{\det G_{\hat{f}}\left(\hat{f}\right)}{\det G\left(\hat{\Phi}\right)} \hat{f} \quad , \tag{2.6}$$

where $G\left(\hat{\Phi}\right) := \left(\left[\hat{f}, \hat{y}\right]\right)_{f, y \in \Phi}$ and $G_{\hat{f}}\left(\hat{f}\right)$ is obtained from $G\left(\hat{\Phi}\right)$ by replacing the f-th row with $\left(\left[\hat{f}, \hat{y}\right]\right)_{y \in \Phi}$.

In the PSI case the formula for the orthogonal projection (2.6) leads to the definition of the natural dual. For any $f \in L_2(\mathbb{R}^d)$, the **natural dual** $\tilde{f}$ is defined by its Fourier transform

$$\hat{\tilde{f}} := \frac{\hat{f}}{\left[\hat{f}, \hat{f}\right]}, \tag{2.7}$$

where we interpret $0/0 = 0$.

Equation (2.6) implies that in the PSI case $\widehat{P_{S(f)} f} = \left[\hat{f}, \hat{\tilde{f}}\right] \hat{f}$. Transforming this back to the "time domain" we obtain the well known quasi-interpolation representation for the orthogonal projection, namely,

$$P_{S(f)} f = \sum_{k \in \mathbb{Z}^d} \left\langle f, \tilde{f}(\cdot - k) \right\rangle f(\cdot - k). \tag{2.8}$$

An FSI space V is called **local** if there exist a finite set of compactly supported functions, Φ, such that $V = S(\Phi)$. In applications compactly supported generators are frequently used to minimize the time and space complexities of the algorithms. An example is Daubechies' [Da] construction of compactly supported orthonormal wavelets. Observe that a local FSI is always regular ([BDR2]). We require the following result on the special case of orthogonal projections of local SI spaces into local SI spaces.

Theorem 2.11 Let V, U be local FSI spaces. Then the orthogonal projection of V into U is a local FSI subspace. In particular it is a regular FSI space.

Proof Let $U = S(\Phi)$, $V = S(\Psi)$ be so that Φ, Ψ are compactly supported generating sets for U, V respectively. Using the commutativity of the orthogonal projection into an SI space and the shift operator, we have that $P_U V = P_U S(\Psi) = S(P_U \Psi)$. Thus, it suffices to prove that for each $y \in \Psi$, there exists a compactly supported function $y' \in U$, such that $S(y') = S(P_U y)$. By virtue of (2.6) we have

$$\widehat{P_U y} = \sum_{f \in \Phi} \frac{\det G_{\hat{f}}(\hat{y})}{\det G\left(\hat{\Phi}\right)} \hat{f}. \tag{2.9}$$

Since the set Φ is composed of compactly supported functions, it follows from (2.1) that the elements of the Gramian $G\left(\hat{\Phi}\right)$ are trigonometric polynomials. Thus, $\det G\left(\hat{\Phi}\right)$ is also a trigonometric polynomial so that $\det G\left(\hat{\Phi}\right) \neq 0$ a.e. on $\mathbb{T}^d$. Let $y' \in S\left(P_U y\right)$ be defined by its Fourier transform, $\widehat{y'} := \det G\left(\hat{\Phi}\right)\widehat{P_U y}$. Then the constructed generator y' has the required compact support property. Indeed, from (2.9) we have the representation $\widehat{y'} = \sum_{f\in\Phi} \det G_{\hat{f}}\left(\hat{y}\right)\hat{f}$ where each $\det G_{\hat{f}}\left(\hat{y}\right)$ is a trigonometric polynomial. This means that y' is a finite sum of compactly supported functions hence it is compactly supported. To conclude we observe that since $\det G\left(\hat{\Phi}\right) \neq 0$ a.e., we have that $\widehat{P_U y} = \left(\det G(\Phi)\right)^{-1}\widehat{y'}$, thus $S(y') = S\left(P_U y\right)$.

♦

The following theorem is the main result of this section. It provides meaningful decompositions of FSI spaces with good approximation properties to an orthogonal sum of two FSI subspace. Naturally, there are many ways to represent FSI spaces as a sum of two FSI subspaces. But our construction is such that the first subspace inherits the good approximation properties of the decomposed space, so that the second subspace is a difference (wavelet) space. The key to the construction is the use of an auxiliary reference space. The underlying principal which justifies this approach is "superfunction theory" [BDR1] and is elaborated upon in Section 4.

Theorem 2.12 Let U_0 be a (local) regular FSI space of length $l_{U_0} \geq 2$. Let V be a (local) FSI space of length $1 \leq l_V < l_{U_0}$. Then U_0 can be decomposed $U_0 = U_1 \oplus W_1$ such that:

1. U_1 is a (local) regular FSI space of length $l_{U_1} = l_V$.

2. W_1 is a (local) regular FSI space of length $l_{W_1} = l_{U_0} - l_V$.

3. $W_1 \perp V$.

Proof

1. Let $\tilde{U}_1 = P_{U_0} V$. Note that $\tilde{U}_1$ is an FSI subspace of U_0 with $len\left(\tilde{U}_1\right) \leq \min\left(l_{U_0}, l_V\right) = l_V$. Without loss of generality, $\tilde{U}_1$ is regular, otherwise, by virtue of Theorem 2.9, we can replace it by a regular subspace of U_0, containing $\tilde{U}_1$ and of the same length, which we will continue to call $\tilde{U}_1$. Observe that in the "local" case, Theorem 2.11 implies that $\tilde{U}_1$ is local.
2. Since $\tilde{U}_1$ is (local) regular, by (Theorem 3.38) Theorem 3.13 in [BDR2] its orthogonal complement in U_0 denoted by $\tilde{W}_1$ is (local) regular and of length $l_{\tilde{W}_1} \geq l_{U_0} - l_V$. Let $\tilde{W}_1 = S\left(y_1, \ldots, y_{l_{\tilde{W}_1}}\right)$ where $S\left(y_1, \ldots, y_i\right)$ is (local) regular for $1 \leq i \leq l_{\tilde{W}_1^\perp}$. By Theorem 2.5 it is always possible to find a generating set with that property. Define $W_1 := S\left(y_1, \ldots, y_{l_{W_1}}\right)$ where $l_{W_1} = l_{U_0} - l_V$. Then clearly $W_1 \perp V$.

3. We conclude the construction by setting U_1 to be the orthogonal complement of W_1 in U_0. By (Theorem 3.38) Theorem 3.13 in [BDR2], U_1 is a (local) regular subspace of U_0 of length $l_{U_1} = l_{U_0} - l_{W_1} = l_V$.

♦

Example 2.13

1. Let f, y be any known pair of univariate semi-orthogonal scaling function and wavelet, e.g., B-splines and B-wavelets [Ch]. Define $U_0 = S(f)^{1/2}$ and $V = S(f)$. Then, since $S(f) \subset S(f)^{1/2}$, the above construction recovers the (refinable) decomposition

$$S(f) \oplus S(y) = S(f)^{1/2}. \tag{2.10}$$

2. Let $S(r_0)$ be a univariate regular PSI space that is not refinable. Assume that r_0 provides L_2 approximation order m. Select $U_0 = S(r_0)^{1/2}$, $V = S(r_0)$. Then the above construction finds a decomposition

$$S(r_1) \oplus S(y_1) = S(r_0)^{1/2}, \quad S(y_1) \perp S(r_0),$$

which in some sense mimics the refinable decomposition (2.10). Furthermore, we show in Section 4.2 that r_1 inherits the approximation order m from r_0 while the wavelet y_1 has m vanishing moments.

3 Non-stationary wavelets

Our first results are simple modifications of the classical "symbol approach" to wavelet construction for the non-refinable setting. Assume $r \in S(j)^{1/2}$ where $j \in L_2(\mathbb{R}^d)$ is stable. Define the **symbol**

$$P(w) := 2^{-d} \sum_{k \in \mathbb{Z}^d} p_k e^{-ikw}, \qquad \text{where } r = \sum_{k \in \mathbb{Z}^d} p_k j(2 \cdot -k). \tag{3.1}$$

To justify the pointwise validity of (3.1) and resolve technical difficulties concerning convergence, we require that these symbols be taken from the Wiener algebra. Namely, $f \in L_2(\mathbb{T}^d)$ is in the **Wiener Algebra** ($f \in \mathbb{W}$) if its Fourier coefficients are in $l_1(\mathbb{Z}^d)$.

The following partitioning of the lattice $\mathbb{Z}^d$, known to be useful in the analysis of refinable functions, is also useful in our non-refinable setting

$$\mathbb{Z}^d = \bigcup_{e \in E_d} \left(e + 2\mathbb{Z}^d\right), \quad E_d := \{0,1\}^d. \tag{3.2}$$

We begin with a "stability" lemma (see [Ch] Theorem 5.16 for the univariate case).

Lemma 3.1 Let $r \in S(j)^{1/2}$ have a symbol $P \in \mathbb{W}$ such that

$$\sum_{e \in E_d} \left| P(w + p\, e) \right|^2 > 0, \qquad \forall w \in \mathbb{T}^d,$$

and assume that $j \in L_2(\mathbb{R}^d)$ is stable. Then r is a stable generator for $S(r)$.

Proof The proof for the univariate case can be found in [Ch] Theorem 5.16. To obtain the proof for the multivariate case one uses the lattice (3.2).

♦

We observe that the following result, which is well known for the refinable case $r = j$, is still valid for the more general case.

Theorem 3.2 Let $j \in L_2(\mathbb{R})$ be a basis for $S(j)$ and let $r, y \in S(j)^{1/2}$. Assume $P, Q \in \mathbb{W}$ where P, Q are the symbols of r, y respectively. A necessary and sufficient condition for $\{r, y\}$ to be a basis for $S(j)^{1/2}$ is

$$\Delta_{P,Q}(w) := P(w)Q(w+p) - P(w+p)Q(w) \neq 0, \qquad \forall w \in \mathbb{T}. \tag{3.3}$$

Furthermore, if j is stable, then r and y are stable bases of $S(r)$ and $S(y)$, respectively.

Proof The proof basically follows the method of [Ch] Theorem 5.16 with the observation that refinability ($r = j$) is not required.

♦

Next we discuss the special case of a decomposition $S(j)^{1/2} = S(r) + S(y)$, with the additional orthogonality constraint $S(r) \perp S(y)$.

Definition 3.3 Let $j \in L_2(\mathbb{R})$ and $r, y \in S(j)^{1/2}$. In case $S(r) \oplus S(y) = S(j)^{1/2}$, we call the decomposition **semi-orthogonal** and r, y a **semi-orthogonal** pair.

Note that the term semi-orthogonality comes from the fact that $S(r) \perp S(y)$, but the shifts of r, respectively y, are not necessarily orthogonal to each other. Assume r has a two-scale symbol $P \in \mathbb{W}$ so that

$$\hat{r}(w) = P\left(\frac{w}{2}\right)\hat{j}\left(\frac{w}{2}\right).$$

Recall that the natural dual $\tilde{r}$ (see (2.7)) can be used to compute the orthogonal projection into $S(r)$. For the dual we also have the following dual two-scale relation

$$\hat{\tilde{r}} = \frac{\hat{r}}{[\hat{r}, \hat{r}]} = \frac{1}{[\hat{r}, \hat{r}]} P(2^{-1}\cdot)\hat{j}(2^{-1}\cdot) = \frac{[\hat{j}, \hat{j}](2^{-1}\cdot)}{[\hat{r}, \hat{r}]} P(2^{-1}\cdot)\hat{\tilde{j}}(2^{-1}\cdot).$$

Hence $\hat{\hat{r}} = G^*\left(2^{-1}\cdot\right)\hat{j}\left(2^{-1}\cdot\right)$ where

$$G^* := \frac{[\hat{j},\hat{j}]}{[\hat{r},\hat{r}](2\cdot)}P\,. \qquad (3.4)$$

Denoting

$$G := \overline{G^*}\,, \qquad (3.5)$$

it is easy to see that we have the duality relation

$$P(w)G(w) + P(w+p)G(w+p) \equiv 1\,. \qquad (3.6)$$

Equipped with the notion of the **dual symbol**, we now characterize the univariate semi-orthogonal (wavelet) complement of a given generator in a space of type $S(j)^{1/2}$.

Theorem 3.4 Let $r \in S(j)^{1/2}$ with a two-scale symbol $P \in \mathbb{W}$, where j and r are stable. Assume further that $G \in \mathbb{W}$, where G is defined by (3.5). Then, $y \in S(j)^{1/2}$ is a stable semi-orthogonal complement such that $S(j)^{1/2} = S(r) \oplus S(y)$ with a two-scale symbol $Q \in \mathbb{W}$ if and only if

$$Q(w) = e^{iw}G(w+p)K(2w)\,, \qquad (3.7)$$

where $K \in \mathbb{W}$ does not vanish on $\mathbb{T}$.

Proof The proof is similar to [Ch] Theorem 5.19.

♦

Using the above we can always complement any generator by a semi-orthogonal counterpart. In particular, in the case of local spaces, this gives us a method to construct a (minimal) compactly supported generator, as done in [Ch], by a proper selection of the periodic function K. Namely, assume j, r are stable and compactly supported and that the symbol P of r is a trigonometric polynomial. By (3.4), the choice $K = [\hat{r},\hat{r}]$ in (3.7) leads to the following two-scale symbol

$$Q(w) = -e^{-iw}[\hat{j},\hat{j}](w+p)\overline{P(w+p)}\,. \qquad (3.8)$$

It is easy to see that for compactly supported j, r, the above symbol produces a complementary compactly supported wavelet.

We conclude this section with the following observation. Let j be stable and two-scale refinable such that $S(j)^{1/2} = S(j) + S(y)$ is a decomposition where P, Q are the corresponding symbols of j, y. In image coding applications perfect reconstruction subband filters banks derived from the symbols P, Q are used in discrete settings (see Section 7.3.2 in [M]). In many applications, one is not required to understand wavelet theory but simply to implement an efficient discrete filtering process. Furthermore, computational steps, that seem necessary according to sampling theory, are ordinarily neglected (see the discussion in [M] pp. 257-258), but still good coding results are obtained. How can one explain this phenomenon? A plausible explanation can be given using the results of this section. As is well known in the signal processing community, the "perfect reconstruction decomposition condition" (3.3) is a property of the symbols P, Q and does not depend on the generator φ. Assume that condition (3.3) holds for the two-scale symbols P, Q and replace the generator φ by some other stable generator ρ_0 which need not be refinable. Then, by Theorem 3.2, the functions $\rho_1, \psi_1 \in S(\rho_0)^{1/2}$ that have P, Q as their two-scale symbols are a basis for $S(\rho_0)^{1/2}$. This means that (3.3) is a universal property of the two-scale symbols P, Q and the subband filters derived from them, regardless of the underlying functions. Furthermore, we will see in Section 4.3 that if in addition, the symbols P, Q have certain approximation properties, then the corresponding basis $\{\rho_1, \psi_1\}$ provides a decomposition which is meaningful in the context of wavelet theory, whenever $S(\rho_0)$ has good approximation properties.

3.1 Non-stationary Superfunction wavelets

In this section we present the construction of non-stationary wavelets inspired by the superfunction techniques of [BDR1]. In our case the projection is done from a stationary reference space, but the superfunction and wavelet spaces are non-stationary. The abstract decomposition of Theorem 2.12 already tells us that, given a reasonable FSI space U, we can decompose it into $U = U_1 \oplus W_1$ using a reference space V, with $len(V) < len(U)$, such that $W_1 \perp V$ and U_1, V are of the same length. The heuristics of the superfunction decompositions presented in this section is justified in Section 4.2 where the approximation properties of the decomposition subspaces are discussed in detail.

Theorem 3.5 Let $U_0 \subset L_2(\mathbb{R}^d)$ be a (local) regular FSI space. Let V be a (local) FSI space with $len(V) = len(U_0)$. Then there exists a sequence of subspaces U_j, W_j, $j \ge 1$ such that

1. U_j and W_j are (local) regular FSI spaces with $len(U_j) = len(U_0)$, $len(W_j) = (2^d - 1)\,len(U_0)$.
2. $U_j \oplus W_j = U_{j-1}^{1/2}$.
3. $W_j \perp V$.

Proof Since dilation by 2^{-j}, $j \ge 1$, preserves the property of (localness) regularity, $U_0^{1/2}$ is a (local) regular FSI of length $2^d len(U_0)$. By Theorem 2.12, $U_0^{1/2}$ can be decomposed into $U_0^{1/2} = U_1 \oplus W_1$ where $len(U_1) = len(V) = len(U_0)$, $W_1 \perp V$ and such that U_1, W_1 are (local) regular. We now continue and decompose $U_1^{1/2}$ in the same manner. By repeated decomposition we obtain an half-multiresolution with the required properties.

♦

Corollary 3.6 Let $U_0 \subset L_2(\mathbb{R}^d)$ be a (local) regular FSI space. Let V be a (local) FSI space with $len(V) = len(U_0)$. Then for any scale $J \in \mathbb{Z}$ we have the following formal wavelet decomposition

$$U_0^{2^{-J}} = \mathop{\oplus}_{j=-\infty}^{J-1} W_{J-j}^{2^{-i}}, \tag{3.9}$$

where $W_j := S(\Psi_j) \perp V$, are non-stationary (local) regular wavelet spaces.

Clearly, the fact that we construct only half-multiresolutions is not a real restriction. By dilating the construction to any given (fine) scale, it can be used to approximate any function in $L_2(\mathbb{R}^d)$ at any required level of accuracy. Also, since we have ensured that each wavelet space W_j is regular, by [BDR2] Corollary 3.31, one may select for each $j \ge 1$ an orthonormal wavelet basis for W_j. From the orthogonality $W_j \perp W_k$ for $j \ne k$, any selection of orthonormal bases Ψ_j for W_j (with the appropriate normalization) provides an orthonormal basis for $U_0^{2^{-J}}$, $J \in \mathbb{Z}$.

Next we discuss actual constructions that realize the decomposition of Theorem 3.5. There are two strategies we can employ. First, we can follow the method of Theorem 2.12 by constructing the superfunction spaces U_j using projection and then complementing them by the wavelet spaces W_j. The second approach is to construct the wavelet space first using methods mostly applied for wavelet constructions over (multivariate) non-uniform grids (see [LM], [LMQ]). Let $j, f \in L_2(\mathbb{R})$ such that $\mathrm{supp}(j) \subseteq [0, m_j]$, $\mathrm{supp}(f) \subseteq [0, m_f]$ with $m_j, m_f \in \mathbb{N}$. We wish to find compactly supported generators r, y so that $S(j)^{1/2} = S(r) \oplus S(y)$ and $S(y) \perp S(f)$. We begin with the construction of the wavelet y. Assume $\mathrm{supp}(y) \subseteq [0, y]$, $y \in \mathbb{N}$. Since $y \in S(j)^{1/2}$, we need to compute $2y - m_j + 1$ unknowns $\{q_k\}_{k=0}^{2y-m_j}$ where

$$y = \sum_{k=0}^{2y-m_j} q_k j(2 \cdot - k).$$

The assumption that $\mathrm{supp}(f) \subseteq [0, m_f]$ implies the following $y + m_f - 1$ constraints

$$\langle y, f(\cdot - j) \rangle = 0, \qquad j = 1 - m_f, \ldots, y - 1.$$

In order to have a non-trivial solution, the number of constraints must be strictly smaller than the number of unknowns. Thus,

$$\underbrace{y + m_f - 1}_{\text{number of orthgonality constraints}} + 1 \le \underbrace{2y - m_j + 1}_{\text{number of unknowns}}.$$

The smallest possible value $y = m_j + m_f - 1$ leads to the following definition for y (up to a multiplicative constant)

$$y(x) = \det \begin{pmatrix} \langle f_{1-m_f}, j_0 \rangle & \cdots & \langle f_{1-m_f}, j_{2m_f+m_j-2} \rangle \\ \langle f_{2-m_f}, j_0 \rangle & \cdots & \langle f_{2-m_f}, j_{2m_f+m_j-2} \rangle \\ \vdots & & \vdots \\ \langle f_{m_f+m_j-2}, j_0 \rangle & \cdots & \langle f_{m_f+m_j-2}, j_{2m_f+m_j-2} \rangle \\ j_0(x) & \cdots & j_{2m_f+m_j-2}(x) \end{pmatrix},$$

where we have denoted $f_k := f(\cdot - k), j_k := j(2\cdot - k)$. We see that $q_k = (-1)^{m_j - k} d_k$ where the minor d_k is defined by the Gram matrix

$$d_k := \det Gram \begin{pmatrix} f_{1-m_f} & & \cdots & & f_{m_f+m_j-2} \\ j_0 & \cdots & j_{k-1} \quad j_{k+1} & \cdots & j_{2m_f+m_j-2} \end{pmatrix}. \tag{3.10}$$

Thus, we obtain the following result.

Theorem 3.7 Let $f, j \in L_2(\mathbb{R})$ where, with $\operatorname{supp}(j) \subseteq [0, m_j]$, $\operatorname{supp}(f) \subseteq [0, m_f]$, $2 \le m_j, m_f \in \mathbb{N}$. Assume that the sequence $\{d_k\}_{k=0}^{m_j+2m_f-2}$ defined by (3.10) is not identically zero. Then for $y := \sum_{k=0}^{m_j+2m_f-2} q_k j(2\cdot - k)$, $q_k = (-1)^{m_j-k} d_k$ we have that $S(y) \perp S(f)$ and $|\operatorname{supp}(y)| \le m_j + m_f - 1$.

Example 3.8

1. Let $j = f = N_m$, where N_m is the univariate B-spline of order m. Then (see [LM]) the B-splines fulfill the conditions of Theorem 3.7. Since $|\operatorname{supp}(N_m)| = m$, we recover the result of Chui that the support of the B-wavelet (minimally supported semi-orthogonal wavelet) is of size $2m-1$.
2. Let $j = f = OM_4$ where $OM_4 := N_4 + N_4''/42$. This generator, constructed in [BTU], has "optimal" approximation properties, but is not two-scale refinable (see Example 4.6). Then, $y_1 \in S(OM_4)^{1/2}$ defined by $y_1 = \sum_{k=0}^{10} q_k OM_4(2\cdot - k)$ with $\{q_k\}$ given (up to a multiplicative constant) by the table below, is stable and fulfills the orthogonality condition $S(y_1) \perp S(OM_4)$.

k	q_k
0,10	-0.000347466
1,9	0.011939448
2,8	-0.099178639
3,7	0.374225526
4,6	-0.786638869
5	1.000000000

Even before the analysis of approximation properties is presented, it is easy to see that y_1 has all the required properties of a wavelet:

- The coefficients $\{q_k\}$ oscillate in sign.
- The coefficients $\{q_k\}$ as “high pass” filters have four vanishing moments.
- The function y_1 has four vanishing moments.

In fact, with the right normalizations, the fifth (non vanishing) moment of $\{q_k\}$ or y_1 is closer to zero than the corresponding one of the cubic B-spline wavelet with the same support size.

♦

Still, according to our theory, the wavelet y_1 constructed in Example 3.8 is only the first wavelet in a series of non-stationary wavelets that must be constructed if one wishes to decompose spaces of the type $S(OM_4)^{2^{-J}}$. The next wavelets in the sequence $y_2, y_3, \dots$ still have four vanishing moments and as we shall see, their fifth moment remains closer to zero than the fifth moment of the cubic wavelet. In such examples, the price paid for removing the refinability property is that the support of the constructed wavelets might grow.

Once the wavelet y is constructed, one may construct a complementary “superfunction” as follows. Assume $|\mathrm{supp}(y)| \le m_j + m_f - 1$ such that $S(y) \perp S(f)$. Now we assume the conditions of Theorem 3.7 again, this time allowing y to play the role of the reference generator. This leads to the construction of a generator $r \in S(j)^{1/2}$ with $S(r) \oplus S(y) = S(j)^{1/2}$ and

$$|\mathrm{supp}(r)| \le m_j + m_y - 1 \le m_j + (m_j + m_f - 1) - 1 = 2m_j + m_f - 2 .$$

Observe that $S(y) \perp S(f)$ implies $P_{S(j)^{1/2}} S(f) \subseteq S(r)$. Since by Theorem 2.11 $P_{S(j)^{1/2}} S(f)$ is a local PSI space, Corollary 2.6 in [BDR2] implies that $S(r) = P_{S(j)^{1/2}} S(f)$.

3.2 Non-stationary Cascade wavelets

It is well known that the cascade operator can be used to obtain a refinable function corresponding to a subdivision scheme, or equivalently, a solution of a two-scale functional equation. Given a mask $P = \{p_k\}_{k \in Z^d}$, we define the **cascade operator** $\mathbb{C}$ by

$$\mathbb{C} f := \sum_{k \in \mathbb{Z}^d} p_k f(2 \cdot - k).$$

Starting with an initial function $r_0 \in L_p(\mathbb{R}^d)$ one iterates $r_{j+1} = \mathbb{C} r_j$.

For our construction we require the general results of [R2] on the cascade operator. We have an initial generator r_0, possibly not refinable, but with good approximation properties. We would like to decompose the space $S(r_0)^{2^{-J}}$, corresponding to a certain scale J, into a sum of meaningful wavelet subspaces. By carefully choosing an appropriate cascade operator and applying it to r_0, we obtain a sequence of generators $r_j = \mathbb{C}^j r_0$ such that:

1. The sequence $\{r_j\}$ converges in some (or all) p-metrics to a refinable function f which is a “fixed point” of the operator $\mathbb{C}$.
2. The spaces $\{S(r_j)\}$ satisfy a nesting property, i.e., $S(r_j) \subset S(r_{j-1})^{1/2}$.

Such a cascade sequence can be used to construct a “wavelet type” decomposition of the space $S(r_0)^{2^{-J}}$ in the following way. First we construct for each level $j \ge 1$ a complement FSI space $S(\Psi_j)$ of length $2^d - 1$ so that $S(r_j) \oplus S(\Psi_j) = S(r_{j-1})^{1/2}$. Once such a non-stationary sequence of spaces is found we can (formally) decompose

$$S(r_0)^{2^{-J}} = \bigoplus_{j=1}^{\infty} S(\Psi_j)^{2^{-J+j}}.$$

The orthogonality $S(\Psi_j) \perp S(\Psi_k)$ for $j \ne k$ simplifies the construction of a stable basis for $S(r_0)^{2^{-J}}$. Indeed, we will construct wavelet generators Ψ_j that are a stable basis for $S(\Psi_j)$ with stability constants A_j, B_j which are uniformly bounded from below and above, i.e., $0 < A \le A_j \le B_j \le B$. Then, from the orthogonality $S(\Psi_j) \perp S(\Psi_k)$, we can immediately derive that their union is a stable basis for $S(r_0)^{2^{-J}}$, with stability constants bounded from below and above, respectively, by A, B.

The following is a simple form of Theorem 3.2.8 in [R2].

Theorem 3.9 [R2] Let $f \in W_p^m(\mathbb{R}^d)$ be a two-scale refinable and stable generator for $S(f)$. Denote by $\mathbb{C} := \mathbb{C}(f)$ the corresponding cascade operator. Let g be a bounded stable compactly supported function for which $\hat{f} - \hat{g} = O(|\cdot|^n)$ near the origin. If the shifts of g provide approximation order $\ge m$, then the cascade algorithm converges at the rate

$$\left\| \mathbb{C}^j g - f \right\|_{L_p(\mathbb{R}^d)} \le A_g 2^{-\min\{m,n\} j}.$$

We see that by a careful selection of the underlying refinable function f we not only ensure convergence of the cascade process, but we can also estimate the convergence rate. For example, a typical application of Theorem 3.9 in our setting for the univariate case is as follows. Let $r_0 = (I + D) N_m$ be a stable generator where N_m is the B-spline of order m and D is some homogeneous differential operator of degree $n \le m-2$. Select the cascade operator $\mathbb{C}(N_m)$. Then, near the origin we have $\left| \left(\widehat{N_m} - \widehat{r_0} \right)(w) \right| \le C|w|$. As we shall see, r_0 provides the same approximation order as N_m and therefore the conditions of Theorem 3.9 are satisfied.

In contrast to the convergence acceleration sought in [R2] using a smart choice of initial seed, in our settings there are cases where slow convergence is preferable. As we shall see in Section 4.3, this is the case whenever the initial function r_0 has better properties then the limit function f. In such a case the first few levels of the cascade process have properties that are "close" to the properties of the initial function. This is useful in applications, since in practice only the first levels of the cascade are used.

Definition 3.10 Let r_0 be an initial function for the cascade process $\mathbb{C}$ defined by a refinable f. Let $r_j = \mathbb{C}^j r_0$ and assume $\lim_{j\to\infty} \left\| r_j - f \right\|_{L_2(\mathbb{R}^d)} \to 0$. We call any sequence $\{\Psi_j\}$ such that $\{r_{j+1}, \Psi_{j+1}\}$ is a basis for $S(r_j)^{1/2}$ a **Cascade Wavelet** sequence.

For the rest of the section we assume that the masks of the cascade operators are finitely supported, hence also the corresponding refinable function. We now show that the cascade process interpolates the stability of the endpoints r_0, f.

Theorem 3.11 Let $r_0 \in L_2(\mathbb{R}^d)$ be a stable compactly supported initial function and let $\mathbb{C}$ be a cascade operator associated with a stable $f \in L_2(\mathbb{R}^d)$. If $\lim_{j\to\infty} \left\| r_j - f \right\|_{L_2(\mathbb{R}^d)} = 0$ where $r_j := \mathbb{C}^j r_0$, then there exist uniform stability constants $0 < \tilde{A} \le \tilde{B} < \infty$ such that $\tilde{A} \le \left[\hat{r}_j, \hat{r}_j \right] \le \tilde{B}$ for all $j \ge 0$.

Proof For $j \ge 0$**,** let A_j, B_j be min/max values of $\left[\hat{r}_j, \hat{r}_j \right]$. Since the Cascade mask is finitely supported, by Lemma 2.4 we have the convergence $A_j \to A$, $B_j \to B$ where A, B are the min/max values of $\left[\hat{f}, \hat{f} \right]$. Thus, we need only prove that each $A_j > 0$.

To this end, let $P(w) = 2^{-d} \sum_{k\in\mathbb{Z}^d} p_k e^{-iwk}$ be the trigonometric polynomial corresponding to the finite mask of the cascade operator $\mathbb{C}$. Since f is stable, we have

$$\sum_{e\in E_d} \left| P(w + pe) \right|^2 > 0, \quad \forall w \in \mathbb{T}^d, \tag{3.11}$$

where we have used the lattice decomposition (3.2). Indeed, otherwise $P(w_0 + pe) = 0$, $\forall e \in E_d$, for some $w_0 \in \mathbb{T}^d$. Then, by the refinability of f

$$\left[\hat{f}, \hat{f} \right](2w_0) = \sum_{e\in E_d} \left| P(w_0 + pe) \right|^2 \left[\hat{f}, \hat{f} \right](w_0 + pe) = 0.$$

Since f is compactly supported, $\left[\hat{f}, \hat{f} \right]$ is a trigonometric polynomial and by Theorem 2.6, this contradicts the stability of f. We can now apply Lemma 3.1 inductively to obtain that each $A_j > 0$.

♦

An immediate consequence of the bounds obtained in Lemma 3.1 and Theorem 3.11 is the following.

Corollary 3.12 Assume that $d=1$ and let r_0 and f be as in Theorem 3.11. Assume further that $\bigcap_{j=-\infty}^{\infty} S(f)^{2^{-j}}=\{0\}$ and let $\{y_j\}$ be a univariate Cascade wavelet sequence such that $S(r_{j+1})\oplus S(y_{j+1})=S(r_j)^{1/2}$ for all $j\ge 0$. If the two-scale symbols of the wavelets satisfy

1. $Q_j\in\mathbb{W}$ with $\|Q_j\|_{C(\mathbb{T})}\le B'<\infty$,

and

2. $|Q_j(w)|^2+|Q_j(w+p)|^2\ge A'>0$, $\forall w\in\mathbb{T}$,

then for any $J\in\mathbb{Z}$ the dilated non-stationary wavelet set $\left\{2^{(J-j)/2}y_j\left(2^{J-j}\cdot-k\right)\right\}_{j\ge1,k\in\mathbb{Z}}$ is a stable basis for $S(r_0)^{2^{-J}}$.

Next we use the general tools presented at the beginning of this section to construct, for a univariate cascade sequence $\{r_j\}_{j=1}^{\infty}$, a sequence of semi-orthogonal wavelets $\{y_j\}_{j=1}^{\infty}$ for which the conditions of Corollary 3.13 hold.

Assume r_0 and f are as in Theorem 3.11. Following (3.4) and (3.5) we define for $j\ge1$

$$G_j=\frac{\left[\hat r_{j-1},\hat r_{j-1}\right]}{\left[\hat r_j,\hat r_j\right](2\cdot)}\overline{P}. \tag{3.12}$$

Since $\left[\hat r_j,\hat r_j\right]>0$ is a trigonometric polynomial for $j\ge0$, by Wiener's lemma [K], we have that $G_j\in\mathbb{W}$ for each $j\ge1$. By Theorem 3.4, any wavelet y_j such that $S(r_{j+1})\oplus S(y_{j+1})=S(r_j)^{1/2}$ has a symbol Q_j of the form

$$Q_j(w)=e^{iw}G_j(w+p)K_j(2w), \tag{3.13}$$

where $K_j\in\mathbb{W}$ never vanishes. Recall that in this local setting we can use (3.8) to choose $\{K_j\}$ so that $\{Q_j\}$ are trigonometric polynomials and thus construct $\{y_j\}$ with compact support. For each $j\ge1$ we select $\hat y_j=Q_j(\cdot/2)\,\hat r_{j-1}(\cdot/2)$ where

$$Q_j(w):=e^{iw}\left[\hat r_{j-1},\hat r_{j-1}\right](w)\,\overline{P}(w).$$

This is equivalent to the selection $K_j=\left[\hat r_j,\hat r_j\right]^{-1}$ in (3.13). We already know that y_j is a semi-orthogonal complement to r_j so that $S(r_j)\oplus S(y_j)=S(r_{j-1})^{1/2}$. Also, observe that since the auto-correlation $\left[\hat r_{j-1},\hat r_{j-1}\right]$ and P are trigonometric polynomials, so is Q_j. Thus, the $\{y_j\}$'s have compact support. Furthermore, we can uniformly bound their support due to the convergence $r_j \to f$ and the fact that we are using a finitely supported cascade mask. It remains to show that the conditions specified in Corollary 3.13 on the wavelet symbols are met. To this end, by Theorem 3.11 there exist $0 < \tilde{A} \le \tilde{B} < \infty$ such that for each $j \ge 0$ we have

$$\tilde{A} \le \left[\hat{r}_j, \hat{r}_j \right] \le \tilde{B}. \tag{3.14}$$

Hence

$$\left\| Q_j(w) \right\|_\infty \le \left\| \left[\hat{r}_{j-1}, \hat{r}_{j-1} \right] \right\|_\infty \left\| P \right\|_\infty \le \tilde{B} \left\| P \right\|_\infty =: B < \infty. \tag{3.15}$$

Also, (3.14) together with (3.11) imply

$$\begin{aligned} \left| Q_j(w) \right|^2 + \left| Q_j(w+p) \right|^2 &= \left(\left[\hat{r}_{j-1}, \hat{r}_{j-1} \right](w) \right)^2 \left| P(w) \right|^2 + \left(\left[\hat{r}_{j-1}, \hat{r}_{j-1} \right](w+p) \right)^2 \left| P(w+p) \right|^2 \\ &\ge \tilde{A}^2 \left(\left| P(w) \right|^2 + \left| P(w+p) \right|^2 \right) \ge A > 0. \end{aligned} \tag{3.16}$$

By virtue of Corollary 3.13 we can conclude that $\left\{ 2^{(J-j)/2} y_j \left(2^{J-j} \cdot - k \right) \right\}_{j \ge 1, k \in \mathbb{Z}}$ is a stable basis for $S\left(r_0 \right)^{2^{-J}}$.

4 Approximation properties

We recall that in classical refinable setting, it is a standard practice to construct wavelets from a given multiresolution analysis of "scaling" function(s). Any reasonable wavelet construction ensures that the (linear) approximation properties of wavelets are derived directly from the approximation properties of the "scaling" function(s). Let us briefly review this point. Throughout this chapter we use the standard notation for the error of approximation

$$E(f, V)_X := \inf_{g \in V} \left\| f - g \right\|_X,$$

where $V \subseteq X$ is a closed subspace of a Banach space X.

First recall that a closed subspace $V \subset L_p\left(\mathbb{R}^d \right)$ is said to provide L_p **approximation order** m if for any function f in the Sobolev space $W_p^m\left(\mathbb{R}^d \right)$

$$E\left(f, V^h \right)_p \le C(V, f) h^m. \tag{4.1}$$

Most results on approximation from shift invariant spaces use the Sobolev semi-norm of the approximated function for the constant in (4.1), namely, a Jackson-type estimate,

$$E\left(f, V^h \right)_p \le C_V h^m \left| f \right|_{W_p^m}. \tag{4.2}$$

If $V = S(\Phi)$ is an FSI space we write C_Φ for C_V. In wavelet theory it is a common practice to ensure that the so called “scaling” functions provide approximation order. Also recall that a generator j of a PSI space $S(j)$ satisfies the **Strang-Fix (SF) conditions of order** m if

$$\hat{j}(0) \neq 0 \text{ and } D^a \hat{j}(2pk) = 0 \text{ for all } k \in \mathbb{Z}^d \setminus 0 \text{ and } |a| < m. \tag{4.3}$$

It is well known that, under certain mild restrictions, if j satisfies the SF conditions of order m then the polynomials of degree $m-1$ can be represented using a superposition of the integer shifts of j, and $S(j)$ provides approximation order m.

On the other hand, wavelets should have the complimentary feature of m vanishing moments. That is, y is a “wavelet” if for all polynomials of degree $m-1$, $p \in \Pi_m$

$$\int_{\mathbb{R}^d} py = 0.$$

The connection between approximation order of the “scaling” functions and the wavelets is simple. Assume $j, r, \Psi \in L_2(\mathbb{R}^d)$, where $\Psi = \{y\}$ and $S(j) = S(r) \oplus S(\Psi)$. It can be shown that if j, r provide approximation order m then all $y \in \Psi$ have m vanishing moments. In such a case the space $S(\Psi)$ will be orthogonal to all polynomials of degree $m-1$.

In this section we show that the nested sequence of non-stationary (“scaling” function) spaces we have constructed using the Superfunction or Cascade methods, beginning with some given non-refinable shift invariant space, inherits the approximation properties of the initial space. Also, the nested spaces share uniform approximation properties. Specifically, we provide simultaneous estimates using uniform constants for the approximation of functions from these spaces. Consequently, our non-stationary wavelet spaces will have the desired vanishing moments property. This is what makes them suitable for signal processing applications.

We now state a Strang-Fix type result that will become useful in Section 4.3. It is quite basic, but handles the case of approximation from a sequence of PSI spaces. First we need the following definitions.

Let $\mathbb{E}_m(\mathbb{R}^d)$ denote the space of bounded measurable functions that decay faster than an inverse of a polynomial of degree $m+d$, i.e.,

$$\mathbb{E}_m(\mathbb{R}^d) := \left\{ f \;\middle|\; |f(x)| \le C(1+|x|)^{-(m+d+e)}, \text{ for some } e > 0 \right\}.$$

Definition 4.1 Let $f \in \mathbb{E}_m(\mathbb{R}^d)$. We say that f satisfies the **Poisson summation condition of order** m if the Poisson Summation Formula holds for all $(\cdot)^n f(x_0 - \cdot)$, $|n| < m$, $x_0 \in \mathbb{R}^d$. Recall that the Poisson summation formula for $g \in L_1(\mathbb{R}^d)$ is

$$\sum_{k \in \mathbb{Z}^d} g(x-k) = \sum_{k \in \mathbb{Z}^d} \hat{g}(2pk) e^{2pikx}.$$

The above requirement holds for example if f is compactly supported, continuous and of bounded variation.

Theorem 4.2 Let $\{r_j\}_{j\geq 1}$ be a sequence of measurable univariate functions and $m \geq 1$. Assume the following conditions hold for each $j \geq 1$.

1. (uniformly bounded support) $supp(r_j) \subseteq [-L, L]$.
2. (uniform bound) $\|r_j\|_\infty \leq M$.
3. (Poisson Summation) The Poisson summation condition of order m holds for r_j.
4. (Strang-Fix) $\hat{r}_j(0) = 1,\ \hat{r}_j^{(l)}(2pk) = 0$, $l = 0, \ldots, m-1,\ k \neq 0$.

Then, there exist constants $\tilde{C}_1, \tilde{C}_2$ which depend on L, M, m (but do not depend on p) such that:

(i) For any $f \in W_p^m(\mathbb{R})$

$$E\left(f, S(r_j)^h\right)_p \leq \tilde{C}_1 h^m |f|_{W_p^m(\mathbb{R})}, \qquad j \geq 1. \tag{4.4}$$

(ii) For any $f \in L_p(\mathbb{R})$

$$E\left(f, S(r_j)^h\right)_p \leq \tilde{C}_2 w_m(f, h)_p, \qquad j \geq 1. \tag{4.5}$$

Proof The proof essentially follows the approach of [DL] Chapter 13 Section 7, with the observation that the constants can be estimated using values of the derivatives of the Fourier transform at the origin. Conditions 1 and 2 ensure that this can be achieved. Namely, there are constants $C_1, \ldots, C_{m-1}$ such that

$$\left|\hat{r}_j^{(n)}(0)\right| \leq C_n,\ 1 \leq n \leq m-1, \quad j \geq 1.$$

♦

4.1 L_2 approximation from shift invariant spaces

For the case of $p = 2$, two tools allow the analysis to be both elegant and powerful, the Hilbert space geometry and the Plancharel-Parseval equality. The latter allows us to carry out the analysis in the frequency domain. An excellent survey of L_2 approximation from shift invariant spaces is [JP]. Henceforth we denote $H^m(\mathbb{R}^d) := W_2^m(\mathbb{R}^d)$.

Definition 4.3 [BDR1] For $f \in L_2(\mathbb{R}^d)$, define the error kernel $\Lambda_f \in L_\infty\left([-p, p]^d\right)$ by

$$\Lambda_f := \left(1 - \frac{|\hat{f}|^2}{[\hat{f}, \hat{f}]}\right)^{\frac{1}{2}}, \tag{4.6}$$

where 0/0 is interpreted as 0.

Applying Fourier methods one can use the error kernel (4.6) to obtain L_2 estimates. The following theorem characterizes the approximation order of an SI space, by the existence of a **superfunction**. The superfunction is required to have an error kernel (4.6) with fast decay to zero about the origin.

Theorem 4.4 [BDR3] Let V be an SI space. Then V provides approximation order $m \geq 1$ such that

$$E\left(f, V^h\right)_2 \leq C_V h^m \|f\|_{H^m}.$$

if and only if there exists $f \in V$ for which $|\cdot|^{-m} \Lambda_f \in L_\infty(B)$, for some neighborhood B of the origin.

As proved in [BU1] the kernel (4.6) can also be used to produce very accurate error estimates.

Theorem 4.5 [BU1] Assume that $f \in \mathbb{E}_m(\mathbb{R})$ is stable with $\hat{f}(0) = 1$ and provides L_2 approximation order m. Then for any function $f \in H^{m+1}(\mathbb{R})$

$$E\left(f, S(f)^h\right)_2 = C_f^- h^m |f|_{H^m(\mathbb{R})} + O\left(h^{m+1}\right), \qquad C_f^- = \frac{1}{m!}\sqrt{\sum_{k \neq 0}\left|\hat{f}^{(m)}(2p k)\right|^2}. \tag{4.7}$$

One of the results in [U] is that the leading constants of type C_f^- in (4.7) are much smaller for the B-Spline generators than for the Daubechies orthonormal scaling functions [Da]. Since the wavelets inherit in some sense this constant from the scaling functions, it might explain the empirical evidence in image coding that spline wavelets outperform the Daubechies wavelets with the same number of vanishing moments.

Example 4.6 OM_m, O-Moms (Optimal Maximum Order and Minimal Support)

The generator OM_m ([BTU], [TBU]) minimize for a given support size (and approximation order) m, the constant C_f^- in (4.7). For each order $m \geq 1$, OM_m can be defined as the outcome of applying a differential operator $I + D_m$ to the B-spline N_m, where D_m is homogeneous of degree $\leq m-1$. It is easy to see that for any differential operator of the type $I + D$, the resulting $(I + D)N_m$ is piecewise polynomial with degree $m-1$ and support size m. Also, since the SF conditions remain valid, OM_m provides approximation order m. The O-Moms functions are continuous for the even orders. For example,

$$OM_4 = N_4 + \frac{1}{42}N_4^{(2)}, \qquad OM_6 = N_6 + \frac{1}{33}N_6^{(2)} + \frac{1}{7920}N_6^{(4)}.$$

The (normalized) gains in sampling density brought by using O-Moms instead of the b-Splines are

$$\left(\frac{C_{N_4}^-}{C_{OM_4}^-}\right)^{1/4} \approx 1.463, \qquad \left(\frac{C_{N_6}^-}{C_{OM_6}^-}\right)^{1/6} \approx 1.951.$$

♦

We augment the L_2 – superfunction theory with a more careful treatment of constants. We combine the finer error estimates of [BU1] related to optimal constants with the superfunction theory of [BDR1]. We show that the superfunction provides asymptotically exactly the same approximation as the “full” space, with the same (sharp) leading constant. First we require the following

Lemma 4.7 [BDR1] Let V be an SI space. Then for any $f, g \in L_2\left(\mathbb{R}^d\right)$

$$E\left(f, V\right)_2 \le E\left(f, S\left(P_V g\right)\right)_2 \le E\left(f, V\right)_2 + 2E\left(f, S\left(g\right)\right)_2 .$$

Theorem 4.8 Let V be an FSI space which provides approximation order $m \ge 1$, such that for any function $f \in H^r\left(\mathbb{R}^d\right)$, $r \ge m$,

$$E\left(f, V^h\right)_2 \le C_V^- h^m \left|f\right|_{H^m} + O\left(h^r\right). \tag{4.8}$$

Then there exists a superfunction $f \in V$ such that for any $f \in H^r\left(\mathbb{R}^d\right)$, $r \ge m$ one has

$$E\left(f, S\left(f\right)^h\right)_2 \le C_V^- h^m \left|f\right|_{H^m} + O\left(h^r\right).$$

Proof Let $f \in H^r\left(\mathbb{R}^d\right)$. We use a dilated version of (4.8)

$$E\left(f\left(h\cdot\right), V\right)_2 = h^{-d/2} E\left(f, V^h\right)_2 \le h^{-d/2}\left(C_V^- h^m \left|f\right|_{H^m} + C\left(V, r; f\right) h^r\right).$$

Select $f = P_V g^*$, where g^* is the multivariate sinc-function

$$g^* := \prod_{i=1}^{d} \frac{\sin p x_i}{p x_i}, \qquad \widehat{g^*} = c_{[-p,p]^d} .$$

It is well known (see for example [JP]) that

$$E\left(f, S\left(g^*\right)^h\right)_2 \le h^r \left|f\right|_{H^r}. \tag{4.9}$$

By virtue of Lemma 4.7, (4.8) and (4.9) we obtain

$$\begin{aligned} E\left(f, S\left(f\right)^h\right)_2 &= h^{d/2} E\left(f\left(h\cdot\right), S\left(f\right)\right) \\ &\le h^{d/2}\left[E\left(f\left(h\cdot\right), V\right) + 2E\left(f\left(h\cdot\right), S\left(g^*\right)\right)\right] \\ &\le C_V^- h^m \left|f\right|_{H^r} + C\left(V, r; f\right) h^r + 2h^r \left|f\right|_{H^r} \end{aligned}$$

$$\le C_V^- h^m |f|_{H^r} + O(h^r).$$

♦

Next we present a similar result for local shift invariant spaces. We require the following "superfunction" result for the local case.

Theorem 4.9 [BDR2] Let V be a local FSI space. Let g be any compactly supported function (not necessarily in V). Then there exists a compactly supported function $f \in V$, such that for every $f \in L_2(\mathbb{R}^d)$

$$E(f, S(f))_2 \le E(f, V)_2 + 2E(f, S(g))_2. \tag{4.10}$$

Theorem 4.10 If in addition to the assumptions of Theorem 4.8 we further assume that V is local, then for each $r > m$ there exists a compactly supported function $f_r \in V$ such that for every $f \in H^r(\mathbb{R}^d)$ one has

$$E\left(f, S(f_r)^h\right)_2 \le C_V^- h^m |f|_{H^m} + O(h^r).$$

Proof The method of proof is very similar to Theorem 4.8, only this time we apply Theorem 4.9 with the selection $g = N_r$, where N_r is the tensor-product B-spline of order r

♦

4.2 Approximation properties of the non-stationary Superfunction wavelets

We now go back to the superfunction decompositions of Section 3.1 and verify that the non-stationary half-multiresolution inherits the approximation properties of the initial space and the reference space. First, we need the following result.

Theorem 4.11 Let $r_0, f \in L_2(\mathbb{R}^d)$ have approximation order m and assume $S(r_0)^{1/2} = S(r_1) \oplus S(\Psi)$ where $S(\Psi) \perp S(f)$. Then r_1 has approximation order m. Furthermore:

1. If for all functions $f \in H^m(\mathbb{R}^d)$ and $h > 0$ the following two estimates hold

$$E\left(f, S(r_0)^h\right)_2 \le C_{r_0} h^m |f|_{H^m(\mathbb{R}^d)}, \qquad E\left(f, S(f)^h\right)_2 \le C_f h^m |f|_{H^m(\mathbb{R}^d)}, \tag{4.11}$$

then for all functions $f \in H^m(\mathbb{R}^d)$ and $h > 0$,

$$E\left(f, S(r_1)^h\right)_2 \le C_{r_1} h^m |f|_{H^m(\mathbb{R}^d)}, \qquad C_{r_1} \le C_{r_0} 2^{-m} + 2C_f.$$

2. If for all functions $f \in H^r(\mathbb{R}^d)$, $r > m$ and $h > 0$ the following two estimates hold

$$E\left(f, S(r_0)^h\right)_2 \le C_{r_0}^- h^m |f|_{H^m(\mathbb{R}^d)} + O(h^r), \quad E\left(f, S(f)^h\right)_2 \le C_f^- h^m |f|_{H^m(\mathbb{R}^d)} + O(h^r), \tag{4.12}$$

then for all functions $f \in H^r\left(\mathbb{R}^d\right)$ and $h > 0$,

$$E\left(f, S\left(r_1\right)^h\right)_2 \le C_{r_1}^- h^m \left|f\right|_{H^m\left(\mathbb{R}^d\right)} + O\left(h^r\right), \qquad C_{r_1}^- \le C_{r_0}^- 2^{-m} + 2C_f^- .$$

Proof

1. Let $f \in H^m\left(\mathbb{R}^d\right)$ and $h > 0$. Since r_0, f have approximation order m, we can obtain a dilated version of (4.11) for both generators

$$E\left(f(h\cdot), S\left(r_0\right)^{1/2}\right)_2 \le h^{-d/2} C_{r_0} \left(h/2\right)^m \left|f\right|_{H^m\left(\mathbb{R}^d\right)}, \qquad E\left(f(h\cdot), S(f)\right)_2 \le h^{-d/2} C_f h^m \left|f\right|_{H^m\left(\mathbb{R}^d\right)}.$$

Since $S(\Psi) \perp S(f)$, it follows that $P_{S(r_0)^{1/2}} S(f) \subseteq S(r_1)$. Since the orthogonal projection onto an SI space and the shift commute we get

$$P_{S(r_0)^{1/2}} S(f) = S\left(P_{S(r_0)^{1/2}} f\right).$$

We now apply Lemma 4.7 to derive

$$\begin{aligned} E\left(f, S\left(r_1\right)^h\right)_2 &= h^{d/2} E\left(f(h\cdot), S\left(r_1\right)\right) \\ &\le h^{d/2} E\left(f(h\cdot), S\left(P_{S(r_0)^{1/2}} f\right)\right) \\ &\le h^{d/2}\left[E\left(f(h\cdot), S\left(r_0\right)^{1/2}\right) + 2E\left(f(h\cdot), S(f)\right)\right] \\ &\le \left(C_{r_0} 2^{-m} + 2C_f\right) h^m \left|f\right|_{H^m}. \end{aligned}$$

2. Let $f \in H^r\left(\mathbb{R}^d\right)$. Then the same arguments yield

$$\begin{aligned} E\left(f, S\left(r_1\right)^h\right)_2 &\le h^{d/2}\left[E\left(f(h\cdot), S\left(r_0\right)^{1/2}\right) + 2E\left(f(h\cdot), S(f)\right)\right] \\ &\le \left(C_{r_0}^- 2^{-m} + 2C_f^-\right) h^m \left|f\right|_{H^m} + \left(C\left(r_0, f\right) 2^{-r} + 2C(f, f)\right) h^r. \end{aligned}$$

♦

We are now ready to justify the superfunction construction of Theorem 3.5.

Corollary 4.12 Let $f, r_0 \in L_2\left(\mathbb{R}^d\right)$ have approximation order m and let $\left\{r_j\right\}_{j \ge 1}$ be so that for all $j \ge 1$:

(i) $S\left(r_j\right) \oplus S\left(\Psi_j\right) = S\left(r_{j-1}\right)^{1/2}$,

(ii) $S\left(\Psi_j\right) \perp S(f)$.

1. If f, r_0 satisfy (4.11), then we have the uniform estimate for any $f \in H^m\left(\mathbb{R}^d\right)$

$$E\left(f,S\left(r_j\right)^h\right)_2 \le \frac{2^{m+1}}{2^m-1}\max\left(C_{r_0},C_f\right)h^m\,|f|_{H^m\left(\mathbb{R}^d\right)},\quad j\ge 0\,. \tag{4.13}$$

2. If f, r_0 satisfy (4.12), then we have the uniform estimate for any $f \in H^r\left(\mathbb{R}^d\right)$, $r > m$,

$$E\left(f,S\left(r_j\right)^h\right)_2 \le \frac{2^{m+1}}{2^m-1}\max\left(C^-_{r_0},C^-_f\right)h^m\,|f|_{H^m\left(\mathbb{R}^d\right)} + O\left(h^r\right),\ j\ge 0\,. \tag{4.14}$$

Proof The proof is by induction. We only prove (4.13), the proof for (4.14) being similar. The estimate (4.13) is certainly true for the initial function r_0. Assume that r_{j-1} has approximation power m. By Theorem 4.11 we see that the generator r_j, constructed using the projection method of Theorem 3.5, inherits the approximation power m with a constant $C_{r_j} \le 2^{-m}C_{r_{j-1}} + 2C_f$. The relation leads to the uniform bound

$$\begin{aligned}
C_{r_j} &\le 2^{-jm}C_{r_0} + \left(\sum_{n=0}^{j-1} 2^{1-nm}\right)C_f \\
&\le \left(2^{-jm} + \sum_{n=0}^{j-1} 2^{1-nm}\right)\max\left(C_{r_0},C_f\right) \\
&\le \frac{2^{m+1}}{2^m-1}\max\left(C_{r_0},C_f\right).
\end{aligned}$$

♦

Example 4.13 Select f, r_0 in Corollary 4.12 to be $OM_4 := N_4 + N_4''/42$ (see Example 4.6). Then for any $f \in H^r(\mathbb{R})$, $r > 4$

$$E\left(f,S\left(r_j\right)^h\right)_2 \le \frac{2^5}{2^4-1}C^-_{OM_4}h^4\,|f|_{H^4(\mathbb{R})} + O\left(h^r\right),\qquad j\ge 0\,.$$

Therefore for all $j \ge 0$,

$$\left(\frac{C^-_{N_4}}{C^-_{OM_4}}\right)^{1/4} \approx 1.463 \Rightarrow \left(\frac{C^-_{N_4}}{C^-_{r_j}}\right)^{\frac{1}{4}} \ge 1.21\,.$$

Assume $\{y_j\}_{j\ge 1}$ are any non-stationary (compactly supported) wavelets, complementing the half-multiresolution generated by $\{r_j\}_{j\ge 0}$ where OM_4 generates both the initial space and the reference space. Then, these wavelets have a sharp constant smaller then the B-wavelets of [Ch] with a gain of about 20%. This result is not very surprising. We have shown (Example 3.8) that we can choose the first wavelet y_1 such that $|\mathrm{supp}(y_1)| = 7$, which is exactly the support size of the cubic B-wavelet. But as explained in Section 3.1, for any such non-stationary wavelet sequence, the support of the wavelets in general grows.

♦

4.3 Approximation properties of the non-stationary Cascade wavelets

The first results of this section verify that the application of a cascade operator with good properties to a given function with good approximation properties yields a function that inherits these properties. These results are connected to the known so called "zero conditions on the mask symbol" (see Section 3.2 in [JP]). The main difference with previous work is that we use "zero conditions" on the cascade mask when applied to non-refinable functions.

Lemma 4.14 Assume $r_0 \in \mathbb{E}_m\left(\mathbb{R}^d\right)$ satisfies the SF conditions of order m and let $P \in \Pi_N$ be a trigonometric polynomial defined by

$$P(w) = R(w)\prod_{r=1}^{d}\left(\frac{1+e^{-iw_r}}{2}\right)^{m_r} = 2^{-d}\sum_{|k|\le N} p_k e^{-ikw}\,, \tag{4.15}$$

with $m_r \ge m$, $r = 1,\dots,d$ and $R(0)\neq 0$. Then r_1 defined by

$$r_1 = \sum_{|k|\le N} p_k r_0(2\cdot - k), \tag{4.16}$$

is in $\mathbb{E}_m\left(\mathbb{R}^d\right)$ and satisfies the SF conditions of order m.

Proof The Fourier equivalent of (4.16), is the two-scale relation

$$\hat{r}_1(w) = P\left(\frac{w}{2}\right)\hat{r}_0\left(\frac{w}{2}\right). \tag{4.17}$$

Since $P \in \Pi_N$, the sum in (4.16) is finite so that $r_1 \in E_m\left(\mathbb{R}^d\right)$ and $\hat{r}_0, \hat{r}_1 \in C^m\left(\mathbb{R}^d\right)$. Since r_0 satisfies SF, it follows that $\hat{r}_1(0) = R(0)\,\hat{r}_0(0) \neq 0$. It is quite easy to show that r_1 satisfies the other SF conditions (4.3). This is done using the two-scale relation (4.17) and the multivariate form of Leibniz' rule.

♦

It is known [R1] that any univariate generator f that provides approximation order m is a convolution of a B-spline of order m and a tempered distribution. Thus, the smallest support possible for a given approximation order m is m. Next we see that the B-spline cascade operator can help preserve this optimal feature.

Corollary 4.15 Assume that $r_0 \in L_\infty(\mathbb{R})$ satisfies the SF and Poisson summation conditions of order m and has (minimal) support size m. Then there exists $r_1 \in S(r_0)^{1/2}$ that provides approximation order m and has (minimal) support size m.

Proof Observe that by Theorem 4.2 r_0 provides approximation order m. We may assume that $supp(r_0) \subseteq [0,m]$ (we can always shift the construction below to this interval and then back). Select P_{N_m}, the (minimally supported) two-scale symbol of the B-Spline of order m, defined by

$$P_{N_m}(w)=\left(\frac{1+e^{-iw}}{2}\right)^m=\frac{1}{2}\sum_{k=0}^{m}p_k e^{-ikw}\ ,\ \ p_k=2^{-m+1}\begin{pmatrix} m \\ k \end{pmatrix}. \tag{4.18}$$

Clearly, for P_{N_m} condition (4.15) holds. Thus by Lemma 4.14, r_1 defined by

$$\hat{r}_1(w)=P_{N_m}\left(\frac{w}{2}\right)\hat{r}_0\left(\frac{w}{2}\right),$$

is in $L_\infty(\mathbb{R})$, has compact support and satisfies the SF and Poisson summation conditions of order m. Using Theorem 4.2 this implies that r_1 provides approximation order m. Also, since $p_k=0$ for all $k\neq 0,\dots,m$, r_1 has the required (minimal) support property.

♦

Thus, we see that a good cascade operator is actually an algorithm to extract a superfunction r_1 from the FSI space $S(r_0)^{1/2}$. We need to verify that the cascade process preserves approximation properties in a uniform sense. The next result overcomes this technical point.

Corollary 4.16 Let r_0 be a univariate function with compact support that satisfies the SF and Poisson summation conditions of order m. Let P be a finite mask of type (4.15) associated with a cascade operator $\mathbb{C}$ and a refinable function $f\in L_\infty(\mathbb{R})$ and let $r_j:=\mathbb{C}^j r_0$ be so that,

1. $supp(r_j)\subseteq[-L,L]$ for all $0\le j<\infty$,
2. $\|r_j\|_\infty\le M$ for all $0\le j<\infty$.

Then the following hold,

1. There exists a constant $\tilde{C}_1$ such that for any $f\in W_p^m(\mathbb{R})$, $1\le p\le\infty$

$$E\left(f,S(r_j)^h\right)_p\le\tilde{C}_1 h^m\left\|f^{(m)}\right\|_{L_p(\mathbb{R})}.$$

2. There exists a constant $\tilde{C}_2$ such that for any for any $f\in L_p(\mathbb{R})$, $1\le p\le\infty$

$$E\left(f,S(r_j)^h\right)_p\le\tilde{C}_2 w_m(f,h)_p.$$

Proof It is easy to see that under our assumptions, conditions 1-3 of Theorem 4.2 hold. Also by Lemma 4.14 it follows that the SF conditions of order m hold for all functions in the sequence and so condition 4 of Theorem 4.2 is also fulfilled. We now apply Theorem 4.2 to obtain the required estimates.

♦

Remark It is interesting to observe that for the last result we did not require that the cascade sequence converges to a refinable function, just that it remained bounded in some box.

For a finer analysis of the inheritance of approximation properties through the cascade process we wish to inspect the sharp constants (4.7). Assume $r \in S(j)^{1/2}$ where j, r are univariate functions such that j satisfies the SF conditions of order m. Assume further that $\hat{r} = P(2^{-1}\cdot)\hat{j}(2^{-1}\cdot)$ where P is of type (4.15) and that we have the normalization $\hat{j}(0) = P(0) = 1$. For each $0 \neq n \in \mathbb{Z}$,

$$\hat{r}^{(m)}(2p\,n) = \left(P\left(\frac{w}{2}\right)j\left(\frac{w}{2}\right)\right)^{(m)}\Bigg|_{w=2p\,n} = 2^{-m}\sum_{k=0}^{m}\binom{m}{k}P^{(k)}(p\,n)\hat{j}^{(m-k)}(p\,n).$$

There are two cases:

1. If $n \equiv 0(\mathrm{mod}\,2)$, then

$$\hat{r}^{(m)}(2p\,n) = 2^{-m}P(0)\hat{j}^{(m)}(p\,n) = 2^{-m}\hat{j}^{(m)}(p\,n),$$

2. If on the other hand $n \equiv 1(\mathrm{mod}\,2)$, then

$$\hat{r}^{(m)}(2p\,n) = 2^{-m}P^{(m)}(p)\hat{j}(p\,n).$$

By (4.7),

$$\begin{aligned}
\left(C_r^-\right)^2 &= \frac{1}{(m!)^2}\sum_{n\neq 0}\left|\hat{r}^{(m)}(2p\,n)\right|^2 \\
&= \frac{1}{(m!)^2}\left[\sum_{k\neq 0}\left|\hat{r}^{(m)}(2p\,2k)\right|^2 + \sum_{k}\left|\hat{r}^{(m)}(2p\,(2k+1))\right|^2\right] \\
&= \frac{1}{(m!)^2}\left[\sum_{k\neq 0}\left|2^{-m}\hat{j}^{(m)}(2p\,k)\right|^2 + \sum_{k}\left|2^{-m}P^{(m)}(p)\hat{j}(p\,(2k+1))\right|^2\right] \\
&= \frac{1}{(m!)^2\,2^{2m}}\left[(m!)^2\left(C_j^-\right)^2 + \left|P^{(m)}(p)\right|^2[\hat{j},\hat{j}](p)\right].
\end{aligned} \tag{4.19}$$

If we may take $r = j$, then $j \in S(j)^{1/2}$ and thus j is refinable. In such a case, we obtain from (4.19) a formula for the constant C_j^-

$$\left(C_j^-\right)^2 = \frac{\left|P^{(m)}(p)\right|^2[\hat{j},\hat{j}](p)}{(m!)^2\left(2^{2m}-1\right)}. \tag{4.20}$$

Formula (4.20) for the refinable case is exactly equation (26) in [BU3].

Lemma 4.17 Let $m \geq 0$ and assume that $r_j \underset{L_2(\mathbb{R}^d)}{\longrightarrow} f$ where $f, r_j \in L_2(\mathbb{R}^d)$ are such that $supp(f), supp(r_j) \subseteq \Omega$ where Ω is some bounded domain. Then we have

$$C^-_{r_j} = \frac{1}{m!}\sum_{k\neq 0}\left|\hat{r}_j^{(m)}(2\pi k)\right|^2 \to \frac{1}{m!}\sum_{k\neq 0}\left|\hat{f}^{(m)}(2\pi k)\right|^2 = C^-_f . \qquad (4.21)$$

Proof By Lemma 2.4, for any $w \in \mathbb{T}^d$,

$$\sum_{k\in\mathbb{Z}^d}\left|\hat{r}_j^{(m)}(w+2\pi k)\right|^2 \to \sum_{k\in\mathbb{Z}^d}\left|\hat{f}^{(m)}(w+2\pi k)\right|^2 . \qquad (4.22)$$

In particular we have (4.22) for $w=0$. It is easy to see that $\hat{r}_j^{(m)}(0)\to\hat{f}^{(m)}(0)$. Hence (4.21) follows.

♦

Lemma 4.18 Let $r_j \underset{L_2(\mathbb{R}^d)}{\to} f$ where f is stable, $\text{supp}(f), \text{supp}(r_j) \subseteq \Omega$ and Ω is some bounded domain in $\mathbb{R}^d$. Then there exists $J>0$ such that for $j \ge J$, $\left\|\Lambda_{r_j} - \Lambda_f\right\|_{C(\mathbb{R}^d)} \to 0$, where we recall that for any $f \in L_2(\mathbb{R}^d)$, Λ_f is the error kernel (4.6).

Proof From the continuity of the Fourier transform, is easy to see that $\left\|\hat{r}_j - \hat{f}\right\|_{C(\mathbb{R}^d)} \to 0$. By Lemma 2.4 we also have $\left\|\left[\hat{r}_j, \hat{r}_j\right] - \left[\hat{f}, \hat{f}\right]\right\|_{C(\mathbb{T}^d)} \to 0$. Since f is stable and compactly supported, its auto-correlation does not have any zeros in $\mathbb{T}^d$. Thus, there exists $J>0$ for which $\left[\hat{r}_j, \hat{r}_j\right]$, $j \ge J$, are uniformly bounded from below. This implies the uniform convergence of the error kernels for $j \ge J$.

♦

An important application of the discussion so far is the following result.

Theorem 4.19 Let $\{r_j\}_{j\ge 0}$ be defined by

$$\hat{r}_{j+1}(w) = P_{N_m}\left(\frac{w}{2}\right)\hat{r}_j\left(\frac{w}{2}\right), \quad j \ge 0 ,$$

where

1. P_{N_m} is the B-spline two-scale symbol (4.18),
2. $r_0 \in L_\infty(\mathbb{R})$, satisfies SF and Poisson summation conditions of order m,
3. r_0 has (minimal) support size m,
4. $\left|\left(\hat{r}_0 - \hat{N}_m\right)(w)\right| = O(|w|)$ near the origin.

Then

1. For any $1 \le p \le \infty$, the sequence r_j converges to the B-spline N_m in every p norm.
2. Each r_j has (minimal) support size m.
3. There exists a constant $\tilde{C}$ such that for any $f \in W_p^m(\mathbb{R})$, $1 \le p \le \infty$,

$$E\left(f, S\left(r_j\right)^h\right)_p \le \tilde{C} h^m \left| f \right|_{W_p^m(\mathbb{R})}, \qquad j \ge 0 .$$

4. The sharp constants $C_{r_j}^-$ converge to $C_{N_m}^-$. Also there exist a constant C^- such that for any function $f \in H^r(\mathbb{R})$

$$E\left(f, S\left(r_j\right)^h\right)_2 \le C^- h^m \left| f \right|_{H^m(\mathbb{R})} + O\left(h^r\right), \; j \ge 0 . \tag{4.23}$$

Proof

1. We use the cascade result Theorem 3.9.
2. See the proof of Corollary 4.15.
3. For each $j \ge 1$ we use Corollary 4.15. We then apply Theorem 4.2.
4. The convergence $C_{r_j}^- \to C_f^-$ follows from Lemma 4.17. The estimate (4.23) is obtained using Lemma 4.18 and some techniques from [BU3].

♦

Example 4.20 Let $r_0 = OM_4$, $f = N_4$ and let $\{r_j\}$ be the sequence constructed in Theorem 4.19. Then one can compute using (4.19),

$$\left(\frac{C_{N_4}^-}{C_{OM_4}^-}\right)^{\frac{1}{4}} \approx 1.463 \Rightarrow \left(\frac{C_{N_4}^-}{C_{r_1}^-}\right)^{\frac{1}{4}} \approx 1.07 .$$

This means that the first generator r_1 constructed by the cascade process is not as good as the initial optimal $r_0 := OM_4$, but still better than the B-spline. The corresponding minimally supported semi-orthonormal wavelets $\{y_j\}$ can be constructed using the methods of Section 3.2 so that for any $J \ge 0$, $\left\{2^{(J-j)/2} y_j\left(2^{J-j} \cdot - k\right)\right\}_{j \ge 1, k \in \mathbb{Z}}$ is a stable basis for $S\left(r_0\right)^{2^{-J}}$. Therefore any approximation obtained from dilations of the PSI space $S\left(r_J\right)$ has a representation in the form of a non-stationary wavelet sum.

In applications such as signal processing, one usually approximates a function and then decomposes the approximation to a sum of a coarse approximation and a few wavelet subspaces. Thus, at least in theory, the non-stationary wavelets derived from a B-spline cascade multiresolution initialized by OM_4, outperform spline-wavelets [Ch], [Da] on these first decomposition levels. Observe that this increase in approximation performance is achieved for exactly the same computational effort. This is due to the fact that the generators $\{r_j\}$ have (minimal) support size 4 and thus the non-stationary Cascade wavelets $\{y_j\}$ have support size 7, which is exactly the support size of the cubic B-wavelet ([Ch]).

♦

References

[BDR1] C. de Boor, R. DeVore and A. Ron, "Approximation from shift invariant subspaces of $L_2\left(\mathbb{R}^d\right)$", Trans. Amer. Math. Soc., 341, pp.787-806, 1994.

[BDR2] C. de Boor, R. DeVore and A. Ron, "The structure of finitely generated shift-invariant spaces in

$L_2\left(\mathbb{R}^d\right)$", J. of Functional Analysis 119(1), pp. 37-78, 1994.

[BDR3] C. de Boor, R. DeVore and A. Ron, "Approximation orders of FSI spaces in $L_2\left(\mathbb{R}^d\right)$", CA 14, pp. 631-652, 1998.

[BTU] T. Blu P. Thevenaz and M. Unser, "Minimum support interpolators with optimum approximation properties", Proc. IEEE-ICIP, 1998.

[BU1] T. Blu and M. Unser, "Approximation error for quasi-interpolators and (multi-) wavelet expansions", App. Comp. Harmonic Anal., vol. 6, no. 2, pp. 219-251, 1999.

[BU2] T. Blu and M. Unser, " Quantitative Fourier Analysis of approximation techniques, part I-Interpolators and projectors", IEEE Trans. Signal Processing, vol. 47, no. 10, pp. 2783-2795, 1999.

[BU3] T. Blu and M. Unser, " Quantitative Fourier Analysis of approximation techniques, part II-Wavelets", IEEE Trans. Signal Processing, vol. 47, no. 10, pp. 2796-2806, 1999.

[Ch] C. K. Chui, "An introduction to wavelets", Academic Press, 1992.

[CSW] C. K. Chui, J. Stockler and J. D. Ward, "Analytic wavelets generated by radial functions", Advances in comp. Math., 5, pp. 95-123, 1996.

[Da] I. Daubechies, "Ten lectures on wavelets", CBMS-NSF Reg. Conf. Series in Applied Math., Vol. 61, SIAM, 1992.

[De] R. DeVore, "Non-linear approximation", Acta Numerica, pp. 51-150, 1998.

[DL] R. DeVore and G. Lorentz, Constructive Approximation, Springer-Verlag, 1991.

[JL] R. Q. Jia and J. Lei, "Approximation by multilinear translates of functions having global support", JAT 72, pp. 2-23, 1993.

[JP] K. Jetter and G. Plonka, "A survey on L_2 – aproximation order from shift invariant spaces", Multivariate approximation and applications (ed. A. Pinkus, D. Leviatan, N. Dyn, and D. Levin), Cambridge Univ. Press, 2001.

[K] Y. Katznelson, "An introduction to Harmonic Analysis", Dover, 1976.

[LM] T. Lyche and K. Morken, "Spline wavelets of minimal support", Numer. Methods of Approx. Theory, Vol 9, pp. 177-194, 1992.

[LMK] T. Lyche, K. Morken and E. Quak, "Theory and algorithms for non-uniform spline wavelets", Multivariate approximation and applications (ed. A. Pinkus, D. Leviatan, N. Dyn, and D. Levin), Cambridge Univ. Press, 2001.

[M] S. Mallat, "A wavelet tour of signal processing", Academic Press, 1998.

[Me] Y. Meyer, "Wavelets and operators", Cambridge, 1992 (French edition: Hermann, 1990).

[R1] A. Ron, "Factorization theorems for univariate splines on regular grids", Israel J. Math. 70, pp. 48-68, 1990.

[R2] A. Ron, "Wavelets and their associated operators", in Approximation Theory IX (ed. C. K. Chui and L. L. Schumaker), Vanderbilt University Press, Nashville, 1999.

[RS] A. Ron and Z. Shen, " Frames and stable bases for shift-invariant subspaces of $L_2\left(\mathbb{R}^d\right)$", Canad. Math. J., 1995.

[TBU] P. Thevenaz, T. Blu and M. Unser, "Image interpolation and resampling", preprint, 1999.

[U] M. Unser, "Approximation power of biorthogonal wavelet expansions", Trans. Signal Proc. 44 (3), pp. 519-527, 1996.