

---

# Multiscale Approximation

S. Dahlke, K. Koch, M. Werner<sup>1</sup>, P. Maaß, D. Lorenz, S. Schiffler<sup>2</sup>, G. Teschke<sup>3</sup>, H. Thiele<sup>4</sup>, and S. Müller, A. Stämpfli<sup>5</sup>

<sup>1</sup> Fachbereich 12 Mathematik und Informatik, Philipps-Universität Marburg  
Hans-Meerwein-Straße, Lahnberge, D-35032 Marburg  
{dahlke,koch,werner}@mathematik.uni-marburg.de

<sup>2</sup> Zentrum für Technomathematik, Fachbereich 3, Universität Bremen  
Postfach 330440, D-28334 Bremen  
{pmaass,dlorenz,schiffi}@math.uni-bremen.de

<sup>3</sup> Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB)  
Takustraße 7, D-14195 Berlin-Dahlem  
teschke@zib.de

<sup>4</sup> Bruker Daltonics GmbH  
Fahrenheitstraße 4, D-28359 Bremen  
Herbert.Thiele@bdal.de

<sup>5</sup> F. Hoffmann-La Roche AG  
Grenzacherstraße 124, CH-4070 Basel  
{stephan.mueller,andreas.staempfli}@roche.com

## 1 Introduction

During the last two decades wavelet methods have developed into powerful tools for a wide range of applications in signal and image processing. The success of wavelet methods is based on their potential for resolving local properties and to analyze non-stationary structures. This is achieved by multi-scale decompositions, e.g., a signal or image is mapped to a phase space parametrized by a time/space- and a scale/size/resolution parameter. In this respect, wavelet methods offer an alternative to classical Fourier- or Gabor-transforms which create a phase space consisting of a time/space- frequency parametrization. Hence, wavelet methods are advantageous whenever local, non-stationary structures on different scales have to be analyzed.

The diversity of wavelet methods, however, requires a detailed mathematical analysis of the underlying physical or technical problem in order to take full advantage of wavelet methods. This first of all requires to choose an appropriate wavelet. The construction of wavelets with special properties is still a central problem in the field of wavelet and multi-scale analysis. We will review a recently developed construction principle for multi-variate multi-wavelets in the next section. As a second step one needs to develop tools for analyzing the result of the wavelet decomposition. Recently, non-linear wavelet meth-

ods (shrinkage techniques) have been developed for solving ill-posed operator equations and inverse problems in imaging. The success of wavelet methods in this field is a consequence of the following facts:

- Weighted sequence norms of wavelet expansion coefficients are equivalent in a certain range (depending on the regularity of the wavelets) to Sobolev and Besov norms.
- For a wide class of operators their representation in the wavelet basis is nearly diagonal.
- The vanishing moments of wavelets remove smooth parts of a function and give rise to very efficient compression strategies.

This will be demonstrated in a section on applications in signal and image processing, where we highlight the potential of wavelet methods for non-linear image decomposition and deconvolution tasks. The numerical results include evaluations of real life data from MALDI/TOF mass spectroscopy from Bruker Daltonics GmbH, Bremen, and HofmannLaRoche AG, Basel.

## 2 Multiwavelets

In this section we present a construction principle for multivariate multi-wavelets, which remedies some fundamental drawbacks of classical wavelet constructions.

The general setting can be described as follows. Let  $M$  be an integer  $d \times d$  scaling matrix which is *expanding*, i.e., all its eigenvalues have modulus larger than one. If for a finite set  $\mathcal{I}$  the system

$$\psi_{i,j,\beta}(x) := m^{j/2} \psi_i(M^j x - \beta), \quad i \in \mathcal{I}, j \in \mathbb{Z}, \beta \in \mathbb{Z}^d, \quad (1)$$

where  $m = |\det M|$ , is a basis of  $L_2(\mathbb{R}^d)$ , then  $\{\psi_{i,j,\beta}\}_{i \in \mathcal{I}, j \in \mathbb{Z}, \beta \in \mathbb{Z}^d}$  is called a *wavelet basis*. Within this classical setting there are still some serious bottlenecks. It has turned out that some desirable properties cannot be achieved at the same time. For instance, it would be optimal to construct an *orthonormal* basis that is also *interpolating*, since orthonormality gives rise to very efficient decomposition and reconstruction algorithms, and the interpolation property yields a Shannon-like sampling theorem. However, it can be checked that for sufficiently smooth wavelets such a construction is impossible [41].

To overcome this problem, a more general approach that provides more flexibility is needed. One way is to consider *multiwavelets*, i.e., a collection of function *vectors*  $\Psi^{(n)} := \left( \psi_0^{(n)}, \dots, \psi_{r-1}^{(n)} \right)^\top \in L_2(\mathbb{R}^d)^r$ ,  $0 < n < m$ , for which

$$\left\{ \psi_0^{(n)}(M^j \cdot -\beta), \dots, \psi_{r-1}^{(n)}(M^j \cdot -\beta) \mid j \in \mathbb{Z}, \beta \in \mathbb{Z}^d, 0 < n < m \right\}$$

forms a (Riesz) basis of  $L_2(\mathbb{R}^d)$ . Compared to the classical scalar setting, this notion of wavelets is much more general.

This section is organized as follows. In Subsection 2.1, we briefly recall the basic properties of scaling functions and wavelets as far as they are needed for our purposes. Then, in Subsection 2.2, we describe a new construction method that enables us to construct orthogonal families of multiwavelets with sufficiently high smoothness which are in addition also interpolating. This is a very surprising result which clearly demonstrates the usefulness of the multiwavelet approach, since nothing similar can be done in the classical wavelet setting. Finally, in Subsection 2.3, our approach is further extended by presenting construction principles for biorthogonal pairs of symmetric compactly supported interpolating scaling vectors with nice approximation and smoothness properties. The results in this subsection have been published in a series of papers [27, 28, 29, 30], we refer to these papers for further details.

## 2.1 General Setting

### Refinable Function Vectors

Let  $\Phi := (\phi_0, \dots, \phi_{r-1})^\top$ ,  $r > 0$ , be a vector of  $L_2(\mathbb{R}^d)$ -functions which satisfies a *matrix refinement equation*

$$\Phi(x) = \sum_{\beta \in \mathbb{Z}^d} A_\beta \Phi(Mx - \beta), \quad A_\beta \in \mathbb{R}^{r \times r}, \quad (2)$$

with the *mask*  $A := (A_\beta)_{\beta \in \mathbb{Z}^d}$ , then  $\Phi$  is called  $(A, M)$ -*refinable*. We shall always assume that the mask has only a finite number of non-vanishing entries,  $A \in \ell_0(\mathbb{Z}^d)^{r \times r}$ , and these entries are denoted by

$$A_\beta = \begin{pmatrix} a_\beta^{(0,0)} & \dots & a_\beta^{(0,r-1)} \\ \vdots & \ddots & \vdots \\ a_\beta^{(r-1,0)} & \dots & a_\beta^{(r-1,r-1)} \end{pmatrix}. \quad (3)$$

Applying the Fourier transform component-wise to (2) yields

$$\widehat{\Phi}(\omega) = \frac{1}{m} \mathbf{A}(e^{-iM^{-\top}\omega}) \widehat{\Phi}(M^{-\top}\omega), \quad \omega \in \mathbb{R}^d, \quad (4)$$

where  $e^{-i\omega}$  is a shorthand notation for  $(e^{-i\omega_1}, \dots, e^{-i\omega_d})^\top$ . The *symbol*  $\mathbf{A}(z)$  is the matrix valued Laurent series with entries

$$a_{i,j}(z) := \sum_{\beta \in \mathbb{Z}^d} a_\beta^{(i,j)} z^\beta, \quad z \in \mathbb{T}^d,$$

and  $\mathbb{T}^d := \{z \in \mathbb{C}^d : |z_i| = 1, i = 1, \dots, d\}$  denotes the  $d$ -dimensional torus. All elements of  $\mathbb{T}^d$  have the form  $z = e^{-i\omega}$ ,  $\omega \in \mathbb{R}^d$ , thus we have  $z^\beta = e^{-i\langle \omega, \beta \rangle}$ ,

and for  $\xi \in \mathbb{R}^d$  we use the notation  $z_\xi := e^{-i(\omega+2\pi\xi)}$ . In addition, we define  $z^M := e^{-iM^\top\omega}$  such that  $(z^M)^\beta = z^{M\beta}$  and  $z_\xi^M := e^{-iM^\top(\omega+2\pi\xi)}$ .

One central aim is the construction of families of *interpolating*  $m$ -scaling vectors  $\Phi$  with compact support, i.e., all components of  $\Phi$  are at least continuous and satisfy

$$\phi_n(M^{-1}\beta) = \delta_{\rho_n,\beta} \quad \text{for all } \beta \in \mathbb{Z}^d, 0 \leq n < m, \quad (5)$$

where  $R := \{\rho_0, \dots, \rho_{m-1}\}$  denotes a complete set of representatives of  $\mathbb{Z}^d/M\mathbb{Z}^d$ . Note that the interpolation condition (and the length of the scaling vector) is determined by the determinant of the scaling matrix. One advantage of interpolating scaling vectors is that they give rise to a Shannon-like sampling theorem as follows. For a compactly supported function vector  $\Phi \in L_2(\mathbb{R}^d)^m$ , let us define the shift-invariant space

$$S(\Phi) := \left\{ \sum_{\beta \in \mathbb{Z}^d} u_\beta \Phi(\cdot - \beta) \mid u \in \ell(\mathbb{Z}^d)^{1 \times m} \right\}.$$

A direct computation shows that, if  $\Phi$  is a compactly supported interpolating  $m$ -scaling vector, then for all  $f \in S(\Phi)$  the representation

$$f(x) = \sum_{\beta \in \mathbb{Z}^d} \sum_{i=0}^{m-1} f(\beta + M^{-1}\rho_i) \phi_i(x - \beta) \quad (6)$$

holds. The interpolation requirement is quite strong and implies the following necessary condition on the mask.

**Lemma 2.1** *Let  $\rho_k \in M\mathbb{Z}^d$ , then the mask of an interpolating  $m$ -scaling vector has to satisfy*

$$a_{M\alpha + \rho_j - M^{-1}\rho_k}^{(i,k)} = \delta_{0,\alpha} \delta_{i,j} \quad \text{for all } \alpha \in \mathbb{Z}^d, 0 \leq i, j < m.$$

For simplicity of notation, we shall assume  $\rho_0 = 0 \in \mathbb{Z}^d$  without loss of generality. Then the above lemma implies that the symbol of an interpolating  $m$ -scaling vector has to have the form

$$\mathbf{A}(z) = \begin{pmatrix} z^{\rho_0} & a^{(0,1)}(z) & \dots & a^{(0,m-1)}(z) \\ \vdots & \vdots & \ddots & \vdots \\ z^{\rho_{m-1}} & a^{(m-1,1)}(z) & \dots & a^{(m-1,m-1)}(z) \end{pmatrix}. \quad (7)$$

For the case  $m = 2$  we can choose  $R = \{0, \rho\}$  and obtain

$$\mathbf{A}(z) = \begin{pmatrix} 1 & a^{(0)}(z) \\ z^\rho & a^{(1)}(z) \end{pmatrix}. \quad (8)$$

## Multiwavelets

Next we want to briefly explain how a multiwavelet basis can be constructed, provided that a suitable (interpolating) refinable function vector is given. As in the classical setting, a multiwavelet basis can be constructed by means of a *multiresolution analysis* which is a sequence  $(V_j)_{j \in \mathbb{Z}}$  of closed subspaces of  $L_2(\mathbb{R}^d)$  which satisfies:

- (MRA1)  $V_j \subset V_{j+1}$  for each  $j \in \mathbb{Z}$ ,
- (MRA2)  $g(x) \in V_j$  if and only if  $g(Mx) \in V_{j+1}$  for each  $j \in \mathbb{Z}$ ,
- (MRA3)  $\bigcap_{j \in \mathbb{Z}} V_j = \{0\}$ ,
- (MRA4)  $\bigcup_{j \in \mathbb{Z}} V_j$  is dense in  $L_2(\mathbb{R}^d)$ , and
- (MRA5) there exists a vector  $\Phi \in L_2(\mathbb{R}^d)^r$ , called the *generator*, such that

$$V_0 = \overline{\text{span}\{\phi_i(x - \beta) \mid \beta \in \mathbb{Z}^d, 0 \leq i < r\}}.$$

Let  $W_0$  denote an algebraic complement of  $V_0$  in  $V_1$  and define  $W_j := \{g(M^j \cdot) \mid g \in W_0\}$ . Then, one immediately obtains that  $V_{j+1} = V_j \oplus W_j$  and consequently, due to (MRA3) and (MRA4),  $L_2(\mathbb{R}^d) = \bigoplus_{j \in \mathbb{Z}^d} W_j$ . If one finds function vectors  $\Psi^{(n)} \in L_2(\mathbb{R}^d)^r$ ,  $0 < n < m$ , such that the integer translates of the components of all  $\Psi^{(n)}$  are a basis of  $W_0$ , then, by dilation, one obtains a multiwavelet basis of  $L_2(\mathbb{R}^d)$ . Since  $W_0 \subset V_1$ , each  $\Psi^{(n)}$  can be represented as

$$\Psi^{(n)}(x) = \sum_{\beta \in \mathbb{Z}^d} B_\beta^{(n)} \Phi(Mx - \beta) \quad (9)$$

for some  $B^{(n)} \in \ell(\mathbb{Z}^d)^{r \times r}$ . By applying the Fourier transform component-wise to (9) one obtains

$$\widehat{\Psi}^{(n)}(\omega) = \frac{1}{m} \mathbf{B}^{(n)}(e^{-iM^{-\top} \omega}) \widehat{\Phi}(M^{-\top} \omega), \quad \omega \in \mathbb{R}^d,$$

with the symbol

$$\mathbf{B}^{(n)}(z) := \sum_{\beta \in \mathbb{Z}^d} B_\beta^{(n)} z^\beta, \quad z \in \mathbb{T}^d.$$

Therefore, the task of finding a stable multiwavelet basis can be reduced to constructing the symbols  $\mathbf{B}^{(n)}(z)$ .

Consequently, to obtain some multiwavelets, we first have to find a way to construct a suitable MRA. Under mild conditions, (MRA1) and (MRA2) imply that the function vector  $\Phi$  in (MRA5) satisfies a refinement equation of the form (2). Therefore refinable function vectors are the natural candidates for generators. Fortunately, it can be shown that any compactly supported interpolating scaling vector indeed generates an MRA, therefore the whole construction problem is reduced to finding suitable interpolating scaling vectors. For further information, the reader is referred to [14, 32, 40].

### Approximation Order

In order to obtain efficient numerical algorithms, the power of the MRA to approximate (sufficiently smooth) functions is essential. For a compactly supported function vector  $\Phi \in L_2(\mathbb{R}^d)^r$  and  $h > 0$ , let

$$S_h(\Phi) := \left\{ f \left( \frac{\cdot}{h} \right) \mid f \in S(\Phi) \cap L_2(\mathbb{R}^d) \right\}$$

denote the space of all  $h$ -dilates of  $S(\Phi) \cap L_2(\mathbb{R}^d)$ .  $\Phi$  (or  $S(\Phi)$ ) is said to provide *approximation order*  $k > 0$  if the Jackson-type inequality

$$\inf_{g \in S_h(\Phi)} \|f - g\|_{L_2} = \mathcal{O}(h^k), \quad \text{as } h \rightarrow 0,$$

holds for all  $f$  contained in the Sobolev space  $H^k(\mathbb{R}^d)$ . The approximation properties of a scaling vector are closely related to its ability to reproduce polynomials. A function vector  $\Phi : \mathbb{R}^d \rightarrow \mathbb{C}^r$  with compact support is said to provide *accuracy order*  $k + 1$ , if  $\pi_k^d \subset S(\Phi)$ , where  $\pi_k^d$  denotes the space of all polynomials of total degree less or equal than  $k$  in  $\mathbb{R}^d$ . It was shown by Jia, see [23], that if a compactly supported scaling vector  $\Phi$  has linear independent integer translates, then the order of accuracy is equivalent to the approximation order provided by  $\Phi$ .

A mask  $A \in \ell_0(\mathbb{Z}^d)^{r \times r}$  of an  $r$ -scaling vector with respect to a scaling matrix  $M$  satisfies the *sum rules of order*  $k$ , if there exists a set of vectors  $\{y_\mu \in \mathbb{R}^r \mid \mu \in \mathbb{Z}_+^d, |\mu| < k\}$  with  $y_0 \neq 0$  such that for some uniquely determined numbers  $w(\mu, \nu)$

$$\sum_{0 \leq \nu \leq \mu} (-1)^{|\nu|} \left( \sum_{\beta \in \mathbb{Z}^d} \frac{(M^{-1}\rho + \beta)^\nu}{\nu!} A_{\rho+M\beta}^\top \right) y_{\mu-\nu} = \sum_{|\nu|=|\mu|} w(\mu, \nu) y_\nu \quad (10)$$

holds for all  $\mu \in \mathbb{Z}_+^d$  with  $|\mu| < k$  and all  $\rho \in R$ . It was proven in [8, 24] that if the mask of a compactly supported scaling vector  $\Phi$  satisfies the sum rules of order  $k$ , then  $\Phi$  provides accuracy of order  $k$ .

### 2.2 Multivariate Orthonormal Interpolating Scaling Vectors

In this subsection, we want to derive a construction method to obtain a multiwavelet basis  $\Psi^{(n)}$  which is orthogonal and interpolating. As already outlined above, the whole construction can be reduced to the task of finding a suitable refinable scaling vector. In particular, we will focus on scaling matrices with  $|\det(M)| = 2$ . Then, in the interpolating setting, we obtain  $r = m = 2$ , and since the number of multiwavelets is determined by  $m$ , cf. Subsection 2.1, this approach enables us to construct a basis of  $L_2(\mathbb{R}^d)$  generated by one single mother multiwavelet consisting of two functions only. Therefore the final goal in this section is to construct a multiwavelet basis that satisfies

$$\langle \psi_i, m^{j/2} \psi_{i'}(M^j \cdot -\beta) \rangle = c \delta_{i,i'} \delta_{0,j} \delta_{0,\beta}, \quad i, i' = 0, 1, j \in \mathbb{Z}, \beta \in \mathbb{Z}^d, \quad (11)$$

with a constant  $c > 0$ , as well as  $\Psi(M^{-1}\beta) = \begin{pmatrix} \delta_{0,\beta} \\ \delta_{\rho,\beta} \end{pmatrix}$  for all  $\beta \in \mathbb{Z}^d$ , and  $R = \{0, \rho\}$ .

### Main Ingredients

As already explained, the starting point is an interpolating scaling vector whose integer translates of all component functions are mutually orthogonal, i.e.,

$$\langle \phi_i, \phi_j(\cdot - \beta) \rangle = \hat{c} \delta_{i,j} \delta_{0,\beta}, \quad i, j = 0, 1, \beta \in \mathbb{Z}^d, \quad (12)$$

with a constant  $\hat{c} > 0$ . By using Fourier transform, it can be checked that the symbol  $\mathbf{A}(z)$  of an orthonormal scaling vector has to satisfy

$$\sum_{\tilde{\rho} \in \tilde{R}} \mathbf{A}(z_{M^{-\tau}\tilde{\rho}}) \overline{\mathbf{A}(z_{M^{-\tau}\tilde{\rho}})}^\top = m^2 \mathbf{I}_r, \quad (13)$$

where  $\tilde{R}$  denotes a complete set of representatives of  $\mathbb{Z}^d/M^\top\mathbb{Z}^d$ . For the special case of an interpolating 2-scaling vector with compact support we obtain the following simplified conditions.

**Theorem 2.2** *Let  $\mathbf{A}(z)$  be the symbol of an interpolating 2-scaling vector with mask  $A \in \ell_0(\mathbb{Z}^d)^{2 \times 2}$ .  $\mathbf{A}(z)$  satisfies (13) if and only if the symbol entries  $a^{(0)}(z)$  and  $a^{(1)}(z)$  in (8) satisfy*

$$\left| a^{(0)}(z) \right|^2 + \left| a^{(0)}(z_{M^{-\tau}\tilde{\rho}}) \right|^2 = 2 \quad (14)$$

and

$$a^{(1)}(z) = \pm z^\alpha \overline{a^{(0)}(z_{M^{-\tau}\tilde{\rho}})} \quad (15)$$

for some  $\alpha \in [\rho]$ , where  $[\rho]$  denotes the coset of  $\rho$ , and with  $\tilde{R} = \{0, \tilde{\rho}\}$ .

So any construction of orthonormal and interpolating scaling functions has to start with the necessary conditions stated in Theorem 2.2. In order to obtain a useful result, also the approximation order, i.e., the sum rules have to be taken into account. For an orthonormal interpolating scaling vector with  $m = 2$  we obtain the following simplification.

**Theorem 2.3** *If we choose  $a^{(1)}(z) = z^\rho \sum_{\beta \in \mathbb{Z}^d} (-1)^{\mathbf{1}_{[\rho]}(\beta)} a_\beta z^{-\beta}$ ,  $a^{(0)}(z) = \sum_{\beta \in \mathbb{Z}^d} a_\beta z^\beta$  in (15), then for an orthonormal interpolating 2-scaling vector the sum rules of order  $k$  are reduced to*

$$\begin{aligned} (M^{-2}\rho)^\mu &= \sum_{\beta \in \mathbb{Z}^d} a_\beta (-M^{-1}\beta)^\mu, \\ (M^{-2}\rho)^\mu &= \sum_{\beta \in \mathbb{Z}^d} a_\beta (M^{-1}\beta)^\mu (-1)^{\mathbf{1}_{[\rho]}(\beta)} \end{aligned}$$

for all  $\mu \in \mathbb{Z}_+^d$ ,  $|\mu| < k$ , with  $R = \{0, \rho\}$ .

### Explicit Construction

Based on the results in the previous subsection, we suggest the following construction principle:

1. Choose a scaling matrix  $M$  with  $|\det(M)| = 2$  and the nontrivial representative  $\rho$  of  $\mathbb{Z}^d/M\mathbb{Z}^d$  such that  $R = \{0, \rho\}$ .
2. Start with the first symbol entry  $a^{(0)}(z) = \sum_{\beta \in \Lambda} a_\beta z^\beta$  by choosing the support  $\Lambda \subset \mathbb{Z}^d$  of  $(a_\beta)_{\beta \in \Lambda}$ . Centering the coefficients around  $a_0$  provides the best results, therefore we suggest the choice of  $\Lambda = [-n, n]^d \cap \mathbb{Z}^d$ .
3. According to Theorem 2.2 the second symbol entry  $a^{(1)}(z)$  has to have the form  $a^{(1)}(z) = \pm z^\alpha \overline{a^{(0)}(z_{M^{-\tau}\bar{\rho}})}$  with  $\alpha \in [\rho]$ . Based on our observations we suggest to choose  $\alpha = \rho$  and a positive sign, since this seems to provide the highest regularity and the smallest support.
4. Apply the orthogonality condition (14) to the coefficient sequence  $(a_\beta)_{\beta \in \Lambda}$ . This will consume about one half of the degrees of freedom.
5. Finally, apply the sum rules of Theorem 2.3 up to the highest possible order to the coefficient sequence  $(a_\beta)_{\beta \in \Lambda}$ .

Starting with an index set  $\Lambda = \{-n, \dots, n\}^2$  we obtain a sequence of scaling vectors denoted by  $\Phi_n$  with increasing accuracy order and regularity. For the special case of the *quincunx matrix*  $M_q$  defined by

$$M_q := \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix},$$

let us denote the resulting scaling vector by  $\Phi_n^q$ . In Table 1 the properties of the constructed examples are shown. Note that for  $n \geq 2$  all our solutions have

$n$	accuracy order	$\mathfrak{s}(\Phi_n^q)$
0	1	0.238
1	1	0.743
2	2	1.355
3	3	1.699
4	3	1.819
5	4	2.002

**Table 1.** Accuracy order and critical Sobolev exponent  $\mathfrak{s}$  of the  $\Phi_n^q$

critical Sobolev exponents strictly larger than one. Therefore, by the Sobolev embedding theorem, see [1], all these scaling vectors are at least continuous.

For  $n = 5$  we obtain an example that is continuously differentiable. The corresponding functions are shown in Figure 1. The reader should note that these scaling vectors are very well localized.



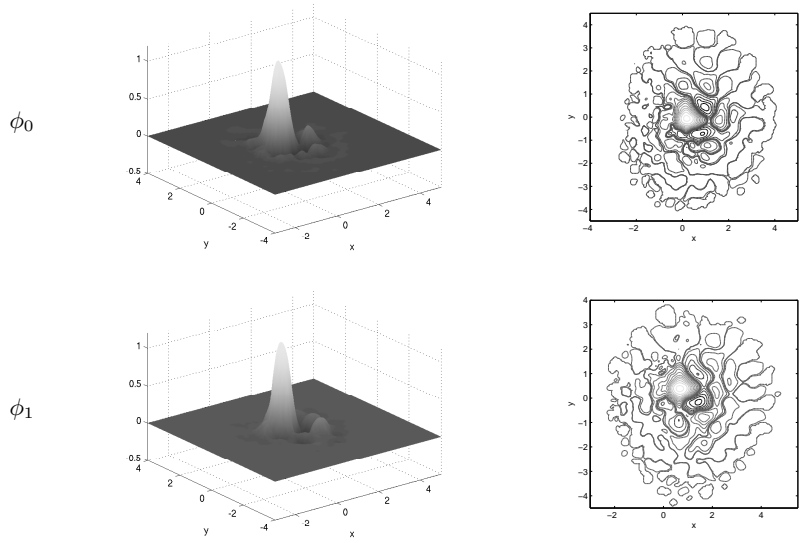


Fig. 1. Component functions of  $\Phi_5^g$

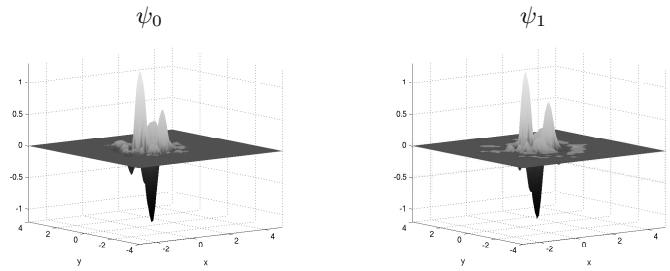


Fig. 2. Multiwavelets corresponding to  $\Phi_5^g$

**Multiwavelets**

Once a suitable scaling vector is found, the construction of an associated multiwavelet basis is easy.

**Theorem 2.4** *Let  $\mathbf{A}(z)$  be the symbol of a compactly supported orthonormal interpolating 2-scaling vector  $\Phi$ . Furthermore, let the function vector  $\Psi$  be defined by (9), where  $\mathbf{B}(z)$  is given by*

$$\mathbf{B}(z) = \begin{pmatrix} 1 & -a^{(0)}(z) \\ z^\rho & -a^{(1)}(z) \end{pmatrix} \tag{16}$$

with  $a^{(0)}(z)$  and  $a^{(1)}(z)$  as in (8). Then,  $\sqrt{2}\Psi$  gives rise to an orthonormal multiwavelet basis, and  $\Psi$  is also interpolating.

One example of such an interpolating multiwavelet corresponding to our scaling vector  $\Phi_5^g$  is shown in Figure 2.

### 2.3 Multivariate Symmetric Interpolating Scaling Vectors

In this subsection, our aim is to extend the approach obtained in the previous subsection by incorporating an additional property, namely symmetry. Since symmetry is hard to achieve in the orthonormal setting, we focus on the construction of a pair of *biorthogonal* scaling vectors  $(\Phi, \tilde{\Phi})$ , i.e.,  $\langle \phi_i, \tilde{\phi}_j(\cdot - \beta) \rangle = c \cdot \delta_{i,j} \delta_{0,\beta}$ ,  $0 \leq i, j < r$ , holds for all  $\beta \in \mathbb{Z}^d$  and a constant  $c > 0$ , the interpolation property being solely satisfied by the primal scaling vectors. A necessary condition for  $\Phi$  and  $\tilde{\Phi}$  to be biorthogonal is that their symbols  $\mathbf{A}(z)$  and  $\tilde{\mathbf{A}}(z)$  satisfy

$$\sum_{\tilde{\rho} \in \tilde{R}} \mathbf{A}(z_{M^{-\tau} \tilde{\rho}}) \overline{\tilde{\mathbf{A}}(z_{M^{-\tau} \tilde{\rho}})}^\top = m^2 \mathbf{I}_m. \quad (17)$$

The concept of biorthogonality provides more flexibility compared to orthonormality, and likely scaling vectors can be obtained providing reasonable approximation power together with much smaller supports than for the case of orthonormal vectors.

#### Main Ingredients

The following notion of symmetry was introduced in [20], see also [19].

A finite set  $\mathcal{G} \subset \{U \in \mathbb{Z}^{d \times d} \mid |\det U| = 1\}$  is called a *symmetry group with respect to  $M$*  if  $\mathcal{G}$  forms a group under matrix multiplication and for all  $U \in \mathcal{G}$  we have  $MUM^{-1} \in \mathcal{G}$ . Since  $\mathcal{G}$  is finite,  $U \in \mathcal{G}$  implies  $M^{-1}UM \in \mathcal{G}$  as well. A scaling vector  $\Phi$  is called  *$\mathcal{G}$ -symmetric*, if for  $0 \leq i < m$  and  $c_i = M^{-1}\rho_i$  holds

$$\phi_i(U(x - c_i) + c_i) = \phi_i(x).$$

Since all elements of  $\mathcal{G}$  are integer matrices, the notion of symmetry can be used for sequences as well. In particular it holds

**Proposition 2.5** *Let  $\mathcal{G}$  be a symmetry group with respect to  $M$  and let  $\Phi$  be a  $\mathcal{G}$ -symmetric interpolating  $m$ -scaling vector with mask  $A \in \ell_0(\mathbb{Z}^d)^{r \times r}$ . If  $U[\rho] = [\rho]$  holds for all  $\rho \in R$  and for all  $U \in \mathcal{G}$ , then the mask entries  $(a_\beta^{(i,j)})_{\beta \in \mathbb{Z}^d}$  are  $\mathcal{G}$ -symmetric with centers  $\rho_i - M^{-1}\rho_j =: c(i, j)$ , meaning that*

$$a_\beta^{(i,j)} = a_{U(\beta - c(i,j)) + c(i,j)}^{(i,j)},$$

for all  $\beta \in \mathbb{Z}^d$  and all  $U \in \mathcal{G}$ ,  $0 \leq i, j < m$ .

Proposition 2.5 shows that a symmetric interpolating scaling vector is completely determined by a small part of its mask.

Furthermore, we are able to give the following decomposition of the supports of our masks.

**Proposition 2.6** *Let  $a^{(i,j)} \in \ell_0(\mathbb{Z}^d)$ ,  $0 \leq i, j < m$ , be  $\mathcal{G}$ -symmetric with centers  $c(i, j) := \rho_i - M^{-1}\rho_j$ . Then there exist finite sets  $\Omega_j \subset \mathbb{Z}^d$  such that  $\Omega_j + M^{-1}\rho_j$  is  $\mathcal{G}$ -symmetric (i.e.,  $U(\Omega_j + M^{-1}\rho_j) \subset \Omega_j + M^{-1}\rho_j$  for all  $U \in \mathcal{G}$ ), and  $\text{supp}(a^{(i,j)}) \subset \Omega_j + \rho_i$ . Furthermore, there exist sets  $\Lambda_j \subset \Omega_j$  such that we have the disjoint decomposition*

$$\Omega_j + \rho_i = \bigcup_{\beta \in \Lambda_j + \rho_i} \bigcup_{U \in \mathcal{G}_{\beta - c(i,j)}} \{U(\beta - c(i,j)) + c(i,j)\}. \quad (18)$$

Similar to Theorem 2.2, using that  $\Phi$  is intended to be interpolating, the biorthogonality condition can be considerably simplified.

**Proposition 2.7** *Let  $(\Phi, \tilde{\Phi})$  be a pair of dual  $m$ -scaling vectors with masks  $(A_\beta), (\tilde{A}_\beta) \in \ell_0(\mathbb{Z}^d)^{m \times m}$ . If  $\Phi$  is interpolating, then the biorthogonality condition (17) holds if and only if*

$$\tilde{a}_{\rho_i - M\alpha}^{(j,0)} + \sum_{n=1}^{m-1} \sum_{\beta \in \mathbb{Z}^d} a_\beta^{(i,n)} \tilde{a}_{\beta - M\alpha}^{(j,n)} = m \cdot \delta_{0,\alpha} \delta_{i,j}, \quad 0 \leq i, j < m, \quad (19)$$

holds for all  $\alpha \in \mathbb{Z}^d$ .

Thus, given the mask of a primal interpolating scaling vector, the biorthogonality condition leads to simple linear conditions on the dual mask.

Similar to the orthonormal case, before we can incorporate the sum rules (10) into our construction, the vectors  $y_\mu$  have to be determined, see [28] for details on how this can be realized.

## Explicit Construction

In this section, we give an explicit construction method for the masks of symmetric interpolating scaling vectors on  $\mathbb{R}^d$  with compact support, as well as for the masks of the dual scaling vectors which are also symmetric and compactly supported. We start with the primal side:

1. Choose the scaling matrix  $M$  and a complete set of representatives  $R = \{0, \rho_1, \dots, \rho_{m-1}\}$  of  $\mathbb{Z}^d/M\mathbb{Z}^d$ . Choose an appropriate symmetry group  $\mathcal{G}$ .
2. To determine the support of the mask  $A \in \ell_0(\mathbb{Z}^d)^{m \times m}$ , choose the sets  $\Omega_j$  in Proposition 2.6 for  $1 \leq j < m$  and compute some minimal generating sets  $\Lambda_j \subset \Omega_j$ . Thus, we start with  $m \cdot \sum_{j=1}^{m-1} |\Lambda_j|$  degrees of freedom.
3. Apply a proper sum rule order  $k$  (i.e. as high as possible) taking into account the symmetry conditions in Proposition 2.5.
4. Find the best solution.

If the sets  $\Lambda_j$  are not too large, we have to deal with a moderate number of linear equations only. Hence, the system in step 3 can be solved analytically. In general, this system is underdetermined and thus, as step 4 of our scheme, we can use these remaining degrees of freedom to maximize the regularity of the corresponding scaling vector  $\Phi$ .

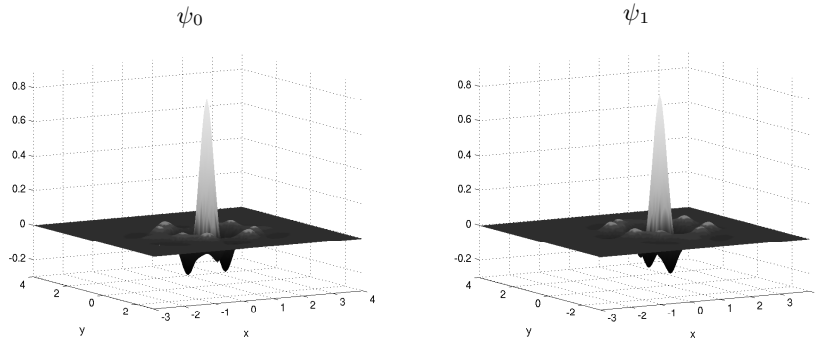
Given the mask of a symmetric interpolating scaling vector, the mask of a dual scaling vector can be obtained as follows:

1. For  $0 \leq i < m$  choose the symmetry center  $c_i$  of  $\tilde{\phi}_i$ . Due to the biorthogonality of  $\Phi$  and  $\tilde{\Phi}$ , the choice  $c_i = M^{-1}\rho_i$  suggests itself.
2. Determine the support of  $\tilde{A} \in \ell_0(\mathbb{Z}^d)^{m \times m}$  by choosing  $\tilde{\Omega}_j$ ,  $0 \leq j < m$ , and compute some minimal generating sets  $\tilde{\Lambda}_j \subset \tilde{\Omega}_j$  corresponding to Proposition 2.6. Thus, we have  $m \cdot \sum_{j=0}^{m-1} |\tilde{\Lambda}_j|$  degrees of freedom.
3. Apply the biorthogonality condition (19) to the coefficient sequence  $(\tilde{A}_\beta)_{\beta \in \mathbb{Z}^d}$  with respect to the symmetry conditions on the dual mask.
4. Choose a proper sum rule order  $\tilde{k}$  and compute the vectors  $\tilde{y}_\mu$ ,  $|\mu| < \tilde{k}$ .
5. Apply the sum rules of order  $\tilde{k}$  to the coefficient sequence  $\tilde{A}$  with respect to the symmetry conditions on the dual mask.
6. Proceed analogously to step 4 for the primal vectors.

In [28] it has been shown that for a pair of compactly supported biorthogonal  $r$ -scaling vectors biorthogonal multiwavelets  $\Psi^{(n)}$  and  $\tilde{\Psi}^{(n)}$ ,  $1 \leq n < m$ , can be obtained by solving appropriate matrix extension problems. For instance, for the quincunx matrix  $M_q$  and the symmetry group

$$\mathcal{G} := \left\{ \pm \mathbf{I}, \pm \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \pm \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \right\},$$

we obtain the primal multiwavelets shown in Figure 3.



**Fig. 3.** Primal multiwavelets corresponding to  $M_q$  with six vanishing moments and critical Sobolev index  $\mathfrak{s} = 3.664$

### 3 Multiscale Approximation in Image Processing

In this section, we discuss some aspects of multiscale approximation in the context of signal and image processing. Essentially, we are focusing on problems where it is reasonable to assume that the solution has a sparse expansion with respect to a wavelet basis. The starting point is always a wavelet-based variational formulation of the underlying signal or image restoration problem, which incorporates a Besov-penalty for ensuring a sparse approximation. The main ingredient to solve the regularized variational problem is therefore the norm equivalence between Besov norms and sequence norms for the orthonormal wavelet decomposition:

$$\|f\|_{H^s(\mathbb{R})} < \infty \Leftrightarrow \sum_k |\langle f, \varphi_{J,k} \rangle|^2 + \sum_{j \geq J} \sum_k 2^{2js} |\langle f, \psi_{j,k} \rangle|^2 < \infty \quad (3.1)$$

$$\begin{aligned} \|f\|_{B_p^s(L_p(\mathbb{R}))} < \infty \Leftrightarrow & 2^{J(1/2-1/p)} \sum_k |\langle f, \varphi_{J,k} \rangle|^p \\ & + \sum_{j \geq J} \sum_k 2^{pjs} 2^{j(\frac{p}{2}-1)} |\langle f, \psi_{j,k} \rangle|^p < \infty \quad (3.2) \end{aligned}$$

see [9]. By using tensor products, an analogous result can also be derived for the multivariate case, see the appendix in [13]. A first result of this type was presented in [15], where the definition of an appropriate surrogate functional led to an iterated soft shrinkage procedure. The shrinkage is due to the  $\ell_p$ -penalization term in the variational formulation and leads to sparse signal representation. The importance of sparse representations for various tasks in image processing such as compression, denoising, deblurring and texture analysis has been highlighted in various papers, which also led to substantial generalizations for solving this type of variational problems in image processing, [2, 3, 37]. In this section emphasis is placed on the special case of simultaneously denoising, decomposing and deblurring as well as some particular deconvolution tasks for peak-like objects. We also discuss the potential of these methods for analyzing real life data from 1D and 2D applications in mass spectroscopy.

#### 3.1 Simultaneous Decomposition, Deblurring and Denoising of Images by Means of Wavelets

We follow approaches presented by Vese–Osher and Osher–Solé–Vese, see [34, 35] and discuss a special class of variational functionals that induce a decomposition of images into oscillating and cartoon components and possibly an appropriate ‘noise’ component; the cartoon part is, ideally, piecewise smooth with possibly abrupt edges and contours; the texture part on the other hand ‘fills’ in the smooth regions in the cartoon with, typically, oscillating features. Osher, Solé and Vese propose to model the cartoon component by the space  $BV$ ; this induces a penalty term that allows edges and contours in the

reconstructed cartoon images, leading however to a numerically intensive PDE based scheme. Our hope is to provide a computationally thriftier algorithm by using a wavelet-based scheme that solves not the same but a very similar variational problem, in which the  $BV$ -constraint is replaced by a  $B_1^1(L_1)$ -term. This relies on the fact that elementary methods based on wavelet shrinkage solve similar extremal problems where  $BV(\Omega)$  is replaced by the Besov space  $B_1^1(L_1(\Omega))$ . Since  $BV(\Omega)$  can not be simply described in terms of wavelet coefficients, it is not clear that  $BV(\Omega)$  minimizers can be obtained in this way. Yet, it is shown in [11], exploiting  $B_1^1(L_1(\Omega)) \subset BV(\Omega) \subset B_1^1(L_1(\Omega))$ -weak, that methods using Haar systems provide near  $BV(\Omega)$  minimizers. So far there exists no similar result for general (in particular smoother) wavelet systems. We shall nevertheless use wavelets that have more smoothness/vanishing moments than Haar wavelets, because we expect them to be better suited to the modeling of the smooth parts in the cartoon image. Though we may not obtain provable ‘near-best- $BV$ -minimizers’, we hope to nevertheless be ‘not far off’. This approach allows us, moreover, to incorporate bounded linear blur operators into the problem so that the minimization leads to a simultaneous decomposition, deblurring and denoising.

### Wavelet-Based Variational Formulation and Iterative Strategy for Image Decomposition

The basic idea of the variational formulation of the decomposition problem goes back to the famous total variation framework of Rudin, Osher and Fatemi [38] and was improved in a series of papers, see e.g. [33, 34, 35, 36], finally amounting to the following minimization problem

$$\inf_{u,v} G_2(u, v) , \quad \text{where} \quad (3.3)$$

$$G_2(u, v) = \int_{\Omega} |\nabla u| + \lambda \|f - (u + v)\|_{L_2(\Omega)}^2 + \mu \|v\|_{H^{-1}(\Omega)}^2 ,$$

where  $u$  stands for the cartoon part and  $v$  for the oscillatory part of a given image  $f$ . In general, one drawback is that the minimization of (3.3) leads to numerically intensive schemes. Instead of solving problem (3.3) by means of finite difference schemes, we propose a wavelet-based treatment by replacing  $BV(\Omega)$  by the Besov space  $B_1^1(L_1(\Omega))$ . Incorporating, moreover, a bounded linear operator  $K$ , we end up with the following variational problem:

$$\inf_{u,v} \mathcal{F}_f(v, u) , \quad \text{where}$$

$$\mathcal{F}_f(v, u) = \|f - K(u + v)\|_{L_2(\Omega)}^2 + \gamma \|v\|_{H^{-1}(\Omega)}^2 + 2\alpha |u|_{B_1^1(L_1(\Omega))} . \quad (3.4)$$

At first, we may observe the following

**Lemma 3.1** *If the null-space  $\mathcal{N}(K)$  of the operator  $K$  is trivial, then the variational problem (3.4) has a unique minimizer.*

In order to solve problem (3.4) by means of wavelets we have to switch to the sequence space formulation. When  $K$  is the identity operator the problem simplifies to

$$\inf_{u,v} \left\{ \sum_{\lambda \in J} \left( |f_\lambda - (u_\lambda + v_\lambda)|^2 + \gamma 2^{-2|\lambda|} |v_\lambda|^2 + 2\alpha |u_\lambda| \right) \right\}, \quad (3.5)$$

where  $J = \{\lambda = (i, j, k) : k \in J_j, j \in \mathbb{Z}, i = 1, 2, 3\}$  is the index set used in our separable setting. The minimization of (3.5) is straightforward, since it decouples into easy one-dimensional minimizations. This results in an explicit shrinkage scheme, presented also in [16, 17]:

**Proposition 3.2** *Let  $f$  be a given function. The functional (3.5) is minimized by the parameterized class of functions  $\tilde{v}_{\gamma,\alpha}$  and  $\tilde{u}_{\gamma,\alpha}$  given by the following nonlinear filtering of the wavelet series of  $f$ :*

$$\tilde{v}_{\gamma,\alpha} = \sum_{\lambda \in J_{j_0}} (1 + \gamma 2^{-2|\lambda|})^{-1} [f_\lambda - S_{\alpha(2^{2|\lambda|} + \gamma)/\gamma}(f_\lambda)] \psi_\lambda$$

and

$$\tilde{u}_{\gamma,\alpha} = f_{(j_0)} + \sum_{\lambda \in J_{j_0}} S_{\alpha(2^{2|\lambda|} + \gamma)/\gamma}(f_\lambda) \psi_\lambda,$$

where  $S_t$  denotes the soft-shrinkage operator,  $J_{j_0}$  all indices  $\lambda$  for scales larger than  $j_0$  and  $f_{(j_0)}$  is the approximation at the coarsest scale  $j_0$ .

In the case where  $K$  is not the identity operator the minimization process results in a coupled system of nonlinear equations for the wavelet coefficients  $u_\lambda$  and  $v_\lambda$ , which is not as straightforward to solve. To overcome this problem we proceed as follows. We first solve the quadratic problem for  $v$ , and then construct an iteration scheme for  $u$ . To this end, we introduce the differential operator  $T := (-\Delta)^{1/2}$ . Setting  $v = Th$ , problem (3.4) reads as

$$\inf_{(u,h)} \mathcal{F}_f(h, u), \quad \text{with} \quad (3.6)$$

$$\mathcal{F}_f(h, u) = \|f - K(u + Th)\|_{L_2(\Omega)}^2 + \gamma \|h\|_{L_2(\Omega)}^2 + 2\alpha |u|_{B_1^1(L_1(\Omega))}.$$

Minimizing (3.6) with respect to  $h$  results in

$$\tilde{h}_\gamma(f, u) = (T^* K^* K T + \gamma)^{-1} T^* K^* (f - K u)$$

or equivalently

$$\tilde{v}_\gamma(f, u) = T(T^* K^* K T + \gamma)^{-1} T^* K^* (f - K u).$$

Inserting this explicit expression for  $\tilde{h}_\gamma(f, u)$  in (3.6) and defining

$$f_\gamma := T_\gamma f, \quad T_\gamma^2 := I - K T (T^* K^* K T + \gamma)^{-1} T^* K^*, \quad (3.7)$$

we obtain

$$\mathcal{F}_f(\tilde{h}_\gamma(f, u), u) = \|f_\gamma - T_\gamma K u\|_{L_2(\Omega)}^2 + 2\alpha|u|_{B_1^1(L_1(\Omega))} . \quad (3.8)$$

Thus, the remaining task is to solve

$$\inf_u \mathcal{F}_f(\tilde{h}_\gamma(f, u), u) , \quad \text{where} \quad (3.9)$$

$$\mathcal{F}_f(\tilde{h}_\gamma(f, u), u) = \|f_\gamma - T_\gamma K u\|_{L_2(\Omega)}^2 + 2\alpha|u|_{B_1^1(L_1(\Omega))} .$$

**Proposition 3.3** *Suppose that  $K$  is a linear bounded operator modeling the blur, with  $K$  maps  $L_2(\Omega)$  to  $L_2(\Omega)$  and  $\|K^*K\| < 1$ . Moreover, assume  $T_\gamma$  is defined as in (3.7) and the functional  $\mathcal{F}_f^{sur}(\tilde{h}, u; a)$  is defined by*

$$\mathcal{F}_f^{sur}(\tilde{h}_\gamma(f, u), u; a) = \mathcal{F}_f(\tilde{h}_\gamma(f, u), u) + \|u - a\|_{L_2(\Omega)}^2 - \|T_\gamma K(u - a)\|_{L_2(\Omega)}^2 .$$

*Then, for arbitrarily chosen  $a \in L_2(\Omega)$ , the functional  $\mathcal{F}_f^{sur}(\tilde{h}_\gamma(f, u), u; a)$  has a unique minimizer in  $L_2(\Omega)$ . The minimizing element is given by*

$$\tilde{u}_{\gamma, \alpha} = \mathbf{S}_\alpha(a + K^*T_\gamma^2(f - Ka)) ,$$

where the operator  $\mathbf{S}_\alpha$  is defined component-wise by

$$\mathbf{S}_\alpha(x) = \sum_\lambda S_\alpha(x_\lambda)\psi_\lambda .$$

The proof follows from [17]. One can now define an iterative algorithm by repeated minimization of  $\mathcal{F}_f^{sur}$ :

$$u^0 \text{ arbitrary ; } u^n = \arg \min_u \left( \mathcal{F}_f^{sur}(\tilde{h}_\gamma(f, u), u; u^{n-1}) \right) \quad n = 1, 2, \dots \quad (3.10)$$

The convergence results shown in [15, 16, 18] can be applied directly:

**Theorem 3.4** *Suppose that  $K$  is a linear bounded operator, with  $\|K^*K\| < 1$ , and that  $T_\gamma$  is defined as in (3.7). Then the sequence of iterates*

$$u_{\gamma, \alpha}^n = \mathbf{S}_\alpha(u_{\gamma, \alpha}^{n-1} + K^*T_\gamma^2(f - K u_{\gamma, \alpha}^{n-1})) , \quad n = 1, 2, \dots ,$$

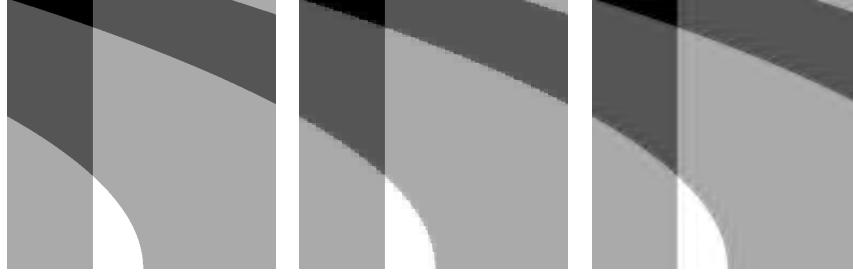
*with arbitrarily chosen  $u^0 \in L_2(\Omega)$ , converges in norm to a minimizer  $\tilde{u}_{\gamma, \alpha}$  of the functional*

$$\mathcal{F}_f(\tilde{h}_\gamma(f, u), u) = \|T_\gamma(f - Ku)\|_{L_2(\Omega)}^2 + 2\alpha|u|_{B_1^1(L_1(\Omega))} .$$

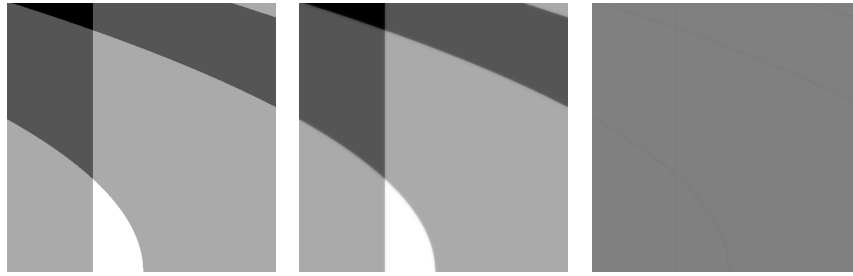
*If  $\mathcal{N}(T_\gamma K) = \{0\}$ , then the minimizer  $\tilde{u}_{\gamma, \alpha}$  is unique, and every sequence of iterates converges to  $\tilde{u}_{\gamma, \alpha}$  in norm.*

Combining the result of Theorem 3.4 and the representation for  $\tilde{v}$  we summarize how the image can finally be decomposed in cartoon and oscillating components.





**Fig. 4.** An initial geometric image  $f$  (left), and two versions of  $f$  (the middle decomposed with the Haar wavelet basis and the right with the Db3 basis) where the soft-shrinkage operator with shrinkage parameter  $\alpha = 0.5$  was applied.



**Fig. 5.** From left to right: initial geometric image  $f$ ,  $\tilde{u}$ ,  $\tilde{v} + 150$ , computed with Db3 in the translation invariant setting,  $\alpha = 0.5$ ,  $\gamma = 0.01$ .

**Corollary 3.5** *Assume that  $K$  is a linear bounded operator modeling the blur, with  $\|K^*K\| < 1$ . Moreover, if  $T_\gamma$  is defined as in (3.7) and if  $\tilde{u}_{\gamma,\alpha}$  is the minimizing element of (3.9), obtained as a limit of  $u_{\gamma,\alpha}^n$  (see Theorem 3.4), then the variational problem*

$$\inf_{(u,h)} \mathcal{F}_f(h, u), \text{ with } \mathcal{F}_f(h, u) = \|f - K(u + Th)\|_{L_2(\Omega)}^2 + \gamma \|h\|_{L_2(\Omega)}^2 + 2\alpha |u|_{B_1^1(L_1(\Omega))}$$

is minimized by the class

$$(\tilde{u}_{\gamma,\alpha}, (T^*K^*KT + \gamma)^{-1}T^*K^*(f - K\tilde{u}_{\gamma,\alpha})) .$$

where  $\tilde{u}_{\gamma,\alpha}$  is the unique limit of the sequence

$$u_{\gamma,\alpha}^n = \mathbf{S}_\alpha(u_{\gamma,\alpha}^{n-1} + K^*T_\gamma^2(f - Ku_{\gamma,\alpha}^{n-1})) , \quad n = 1, 2, \dots .$$

**Numerical Experiments - Additional Redundancy and Adaptivity**

The non-linear filtering rule of Proposition 3.2 gives explicit descriptions of  $\tilde{v}$  and  $\tilde{u}$  that are computed by fast discrete wavelet schemes. However, non-redundant filtering very often creates artifacts in terms of undesirable oscillations, which manifest themselves as ringing and edge blurring, see Figure 4.



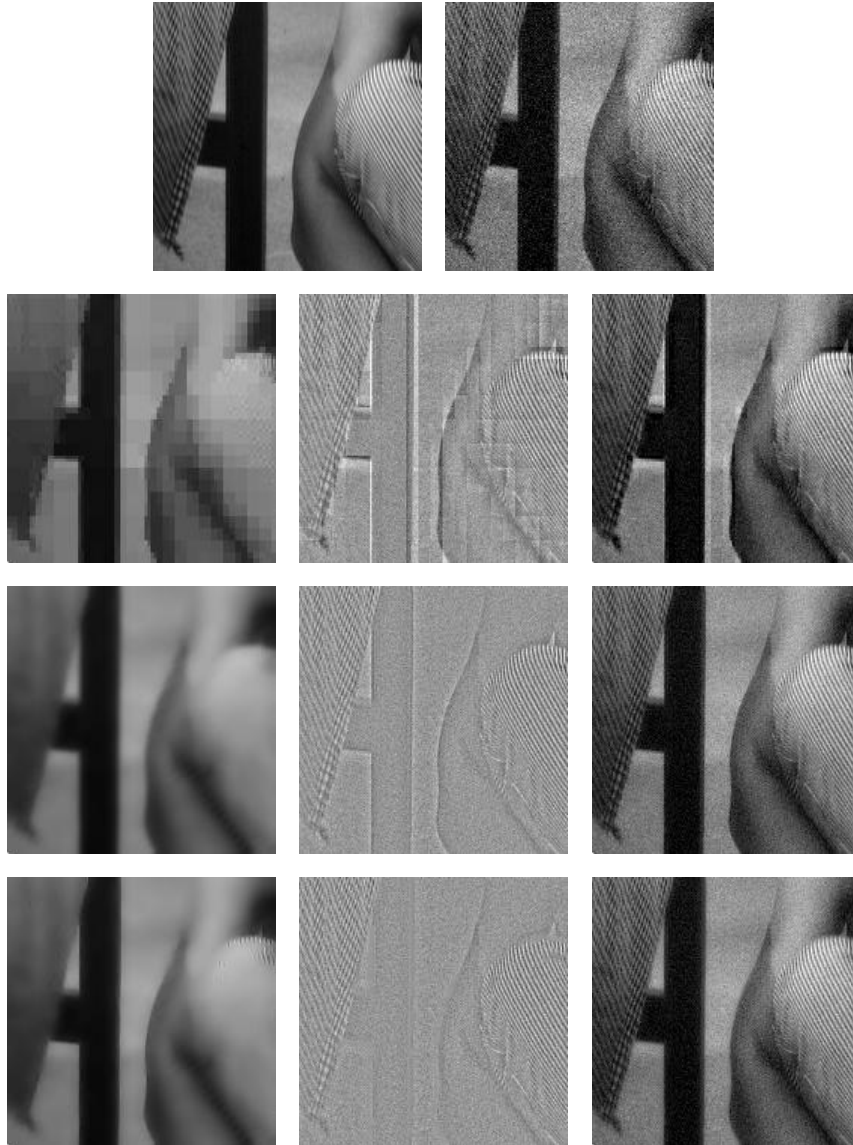
**Fig. 6.** Left: noisy segment of a woman image, middle and right: first two scales of  $\mathbf{S}(f)$  inducing the weight function  $w$ .

Poor directional selectivity of traditional tensor product wavelet bases likewise cause artifacts. Therefore we apply various refinements on the basic algorithm that address this problem. In particular, we shall use redundant translation invariant schemes, see [12], complex wavelets, see e.g. [25, 39], and additional edge dependent penalty weights introduced in [17]. Here we limit ourselves to presenting the numerical results for the particular problem of simultaneously decomposing, deblurring and denoising a given image.

We start with the case where  $K$  is the identity operator. In order to show how the nonlinear (redundant) wavelet scheme acts on piecewise constant functions we decompose a geometric image (representing cartoon components only) with sharp contours, see Figure 5. We observe that  $\tilde{u}$  represents the cartoon part very well. The texture component  $\tilde{v}$  (plus a constant for illustration purposes) contains only some very weak contour structures.

Next, we demonstrate the performance of the Haar shrinkage algorithm successively incorporating redundancy (by cycle spinning) and local penalty weights. The local penalty weights are computed the following way: first, we apply the shrinkage operator  $\mathbf{S}$  to  $f$  with a level dependent threshold  $\alpha'$ . Second, for those  $\lambda$  according to the non-zero values of  $S_{\alpha'}(f_\lambda)$  we put an extra weight  $w_\lambda \gg 1$  in the  $H^{-1}$  penalty. The coefficients  $S_\xi(f_\lambda)$  for the first two scales of a segment of the image ‘Barbara’ are visualized in Figure 6. In Figure 7, we present our numerical results. The upper row shows the original and the noisy image. The next row visualizes the results for non-redundant Haar shrinkage (Method A). The third row shows the same but incorporating cycle spinning (Method B), and the last row shows the incorporation of cycle spinning and local penalty weights. Each extension of the shrinkage method improves the results. This is also being confirmed by comparing the signal-to-noise-ratios (which is here defined as follows:  $SNR(f, g) = 10 \log_{10}(\|f\|^2 / \|f - g\|^2)$ ), see Table 2.

In order to compare the performance with the Vese–Osher TV model and with the Vese–Solé–Osher  $H^{-1}$  model we apply our scheme to a woman image (the same that was used in [34, 35]), see Figure 8. We obtain very similar



**Fig. 7.** Top: initial and noisy image, 2nd row: non-redundant Haar shrinkage (Method A), 3rd row: translation invariant Haar shrinkage (Method B), bottom: translation invariant Haar shrinkage with edge enhancement (Method C); 2nd-4th row from left to right:  $\tilde{u}$ ,  $\tilde{v} + 150$  and  $\tilde{u} + \tilde{v}$ ,  $\alpha = 0.5$ ,  $\gamma = 0.0001$ , computed with Haar wavelets and critical scale  $j_e = -3$ .

results as obtained with the TV model proposed in [35]. Compared with the

Haar Shrinkage	SNR( $f, f_\epsilon$ )	SNR( $f, u + v$ )	SNR( $f, u$ )
Method A	20,7203	18,3319	16,0680
Method B	20,7203	21,6672	16,5886
Method C	20,7203	23,8334	17,5070

**Table 2.** Signal-to-noise ratios of the several decomposition methods (Haar shrinkage, translation invariant Haar shrinkage, translation invariant Haar shrinkage with edge enhancement).

Data basis	"Barbara" image (512x512 pixel)
Hardware Architecture	PC
Operating System	linux
OS Distribution	redhat7.3
Model	PC, AMD Athlon-XP
Memory Size (MB)	1024
Processor Speed (MHz)	1333
Number of CPUs	1
Computational cost	(average over 10 runs)
PDE scheme in Fortran (compiler f77)	56,67 sec
wavelet shrinkage Method A (Matlab)	4,20 sec
wavelet shrinkage Method B (Matlab)	24,78 sec
wavelet shrinkage Method C (Matlab)	26,56 sec

**Table 3.** Comparison of computational cost of the PDE- and the wavelet-based methods.

results obtained with the  $H^{-1}$  model proposed in [34] we observe that our reconstruction of the texture component contains much less cartoon information. In terms of computational cost we have observed that even in the case of applying cycle spinning and edge enhancement our proposed wavelet shrinkage scheme is less time consuming than the Vese-Solé-Osher  $H^{-1}$  restoration scheme, see table 3, even when the wavelet method is implemented in Matlab, which is slower than the compiled version for the Vese-Solé-Osher scheme.

We end this section with an experiment where  $K$  is not the identity operator. In our particular case  $K$  is a convolution operator with Gaussian kernel. The implementation is simply done in Fourier space. The upper row in Figure 9 shows the original  $f$  and the blurred image  $Kf$ . The lower row visualizes the results: the cartoon component  $\tilde{u}$ , the texture component  $\tilde{v}$ , and the sum of both  $\tilde{u} + \tilde{v}$ . One may clearly see that the deblurred image  $\tilde{u} + \tilde{v}$  contains (after a small number of iterations) more small scale details than  $Kf$ . This definitely shows the capabilities of the proposed iterative deblurring scheme (3.10).



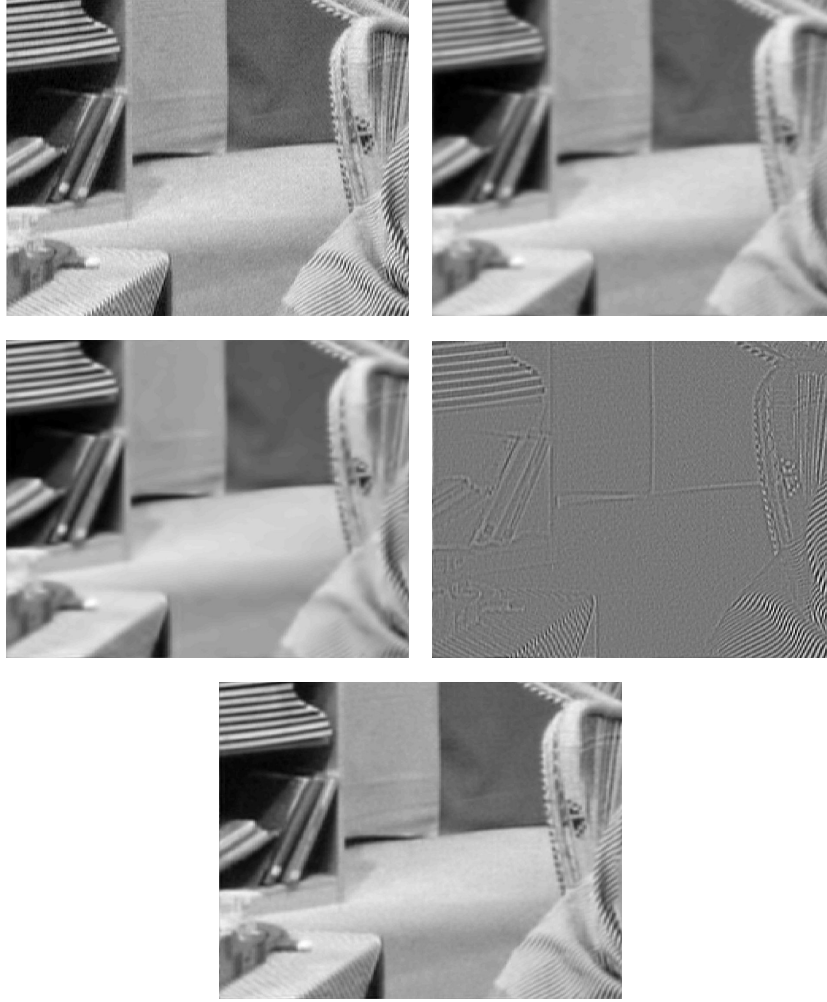
**Fig. 8.** Top from left to right: initial woman image  $f$ ,  $\tilde{u}$  and  $\tilde{v} + 150$ , computed with Db10 (Method C),  $\alpha = 0.5$ ,  $\gamma = 0.002$ ; bottom from left to right:  $u$  and  $v$  obtained by the Vese–Osher TV model and the  $v$  component obtained by the Vese–Solé–Osher  $H^{-1}$  model.

### 3.2 Deconvolution of $\delta$ -Sequences

This section was inspired by discussions with signal and imaging experts in the field of preprocessing 1D and 2D mass spectroscopy data in proteomics. Both applications are mathematically modeled by a convolution operator, hence the data are some blurred and noisy signals or images. Accordingly, the classical approach for solving this inverse ill-posed problem would consist of a regularized deconvolution applied to the given data.

However, the sought-after signals or images are mathematically modeled by finite sums of delta peaks, which are not captured by the classical theory. Moreover, numerical experiments in both fields indicate, that a somewhat ‘practical approach’ yields better results. This ‘practical approach’ proceeds by computing a wavelet-shrinkage on an appropriate wavelet decomposition followed by simply plotting the positions and amplitudes of the remaining coefficients. It has been shown, [26], that this approach is indeed equivalent to a regularized deconvolution scheme in Besov scales. Besov scales are needed in order to obtain an appropriate mathematical model for the reconstruction of such sequences of delta peaks.

The aim of the present section is to summarize the mathematical justification of this approach as given in [26] and to present reconstruction results



**Fig. 9.** Top from left to right: initial image  $f$ , blurred image  $Kf$ ; middle from left to right: deblurred  $\tilde{u}$ , deblurred  $\tilde{v} + 150$ ; bottom: deblurred  $\tilde{u} + \tilde{v}$ , computed with Db3 using the iterative approach,  $\alpha = 0.2$ ,  $\gamma = 0.001$ .

for deconvolving 1D MALDI/TOF-data (Bruker Daltonics GmbH) and 2D LCMS spectra (HofmannLARoche AG).

This approach will require to measure the defect  $\|Af - g^\delta\|_{B_p^s(L_p)}$  in an appropriate Besov space. Hence the resulting regularization method extends the recently proposed sparsity schemes for solving inverse problems [3, 15, 37], which treated  $L_2$  defects, i.e.  $\|Af - g^\delta\|_{L_2}$ , in combination with Besov sparsity constraints.

### A Mathematical Formulation of the Practical Approach

As a first observation, let us note, that in both applications, the sought-after function can be modeled by a finite set of delta peaks. Hence, the sought-after function has a sparse structure and, in addition, the mass spectroscopy data is equispaced by multiples of the unitary atomic mass.

Therefore, let us now state the mathematical formulation which is the basis of the above mentioned practical approach:

We consider some noisy data  $g^\delta$ , a function with finite support. The ‘practical approach’ is a two-step procedure, which starts by a shrinkage operation on an appropriate wavelet decomposition followed by plotting the amplitudes and positions of the remaining coefficients.

Let us formalize this procedure. Applying a shrinkage operation  $S_\lambda$  to  $g^\delta$  starts by computing a wavelet decomposition with a bi-orthogonal wavelet basis  $\varphi, \psi, \tilde{\varphi}, \tilde{\psi}$ ,

$$g^\delta = \sum_{k \in \mathbb{Z}} c_k^J 2^{J/2} \varphi(2^J \cdot -k) + \sum_{j \geq J, k \in \mathbb{Z}} d_k^j 2^{j/2} \psi(2^j \cdot -k)$$

where the coefficients are given by  $c_k = \langle g^\delta, 2^{J/2} \tilde{\varphi}(2^J \cdot -k) \rangle$  and  $d_k^j = \langle g^\delta, 2^{j/2} \tilde{\psi}(2^j \cdot -k) \rangle$ , see, e. g., [32]. It follows a shrinkage of the coefficients, i. e. deleting all coefficients  $|d_k^j| \leq \lambda$ . This yields a function  $S_\lambda g^\delta$  which is the finite sum of wavelet and scaling functions. We choose a ‘finest scale’  $j_0 > J$  (for convenience of notation we set  $j_0 = 0$ ) and delete all coefficients on scales finer than  $j_0$ . This amounts to a projection of  $S_\lambda g^\delta$  and hence,

$$P_0 S_\lambda g^\delta = \sum_{k \in \mathbb{Z}} c_k^J 2^{J/2} \varphi(2^J \cdot -k) + \sum_{0 \geq j \geq J, k \in \mathbb{Z}} d_k^j 2^{j/2} \psi(2^j \cdot -k) = \sum_{k \in \mathbb{Z}} c_k \varphi(\cdot - k)$$

can be represented as a finite sum of scaling functions on this scale. The reconstruction by plotting the position of the coefficient sequence  $\{c_k\}$  is equivalent to a reconstruction

$$R_\alpha P_0 S_\lambda g^\delta = \sum_{k \in \mathbb{Z}} c_k \delta(\cdot - k) ,$$

which is a deconvolution of  $P_0 S_\lambda g^\delta$  with the scaling function  $\varphi$ . This should give good results, whenever  $\varphi$  is a good approximation to the kernel of the true convolution operator.

This ‘practical approach’ also shares some ingredients with the compressive sampling techniques, [6, 7], however, they proceed in a different direction by analyzing achievable levels of resolution as well as deriving sampling theorems.

Finally we want to mention some of the prominent papers in the vast literature, which analyze specific properties of deconvolution problems. There are at least two fairly recent papers, which start from a precise mathematical model for specific applications. In [22] cumulative spectra in Hilbert scales are

treated, and [21] analyzes a deconvolution problem in astronomy in combination with an efficient CG solver. Both papers use models in  $L_2$  spaces with source conditions in Hilbert scales.

The publications [4, 5] give an overview on inverse problems in astronomy, in particular they address the relevant convolution problems in this field. We are well aware of the fact, that this is an incomplete list of even the most basic results. However, to our best knowledge, deconvolution problems in Besov spaces have not yet been addressed in the accessible literature.

### Basic Ingredients

Despite the rather basic mathematical model (convolution operator) of the underlying application a precise definition of all ingredients of the related inverse problem (function spaces, convolution kernels, source conditions) requires some care. The most frequently used models use  $L_2$ -function spaces, mainly for convenience and in order to apply standard regularization theory. However, the nature of the specific convolution problems under consideration is characterized by

- a sparse structure of the solution and
- a model which needs to capture spectral lines or point like objects, i. e. a chain of delta peaks.

Neither of these two requirements is captured by the standard theory.

In this section we will first introduce the convolution operators under consideration. They will be rather straightforward and classical. We then introduce the appropriate function spaces, which leads to Besov spaces and sparsity constraints.

The natural models for the applications described in the introduction is given by an operator  $A : X \rightarrow Y$  which is an integral transform with a convolution kernel:

$$Af = \varphi * f = \int \varphi(\cdot - y)f(y) dy,$$

where  $\varphi$  approximates the point spread function of the measurement device. We will first address the case of a 1D B-spline kernel  $\varphi$ , i.e.

$$\varphi_k(x) = \underbrace{\varphi * \dots * \varphi}_{k \text{ times}}(x), \text{ where } \varphi(x) = \frac{1}{2r}\chi_{[-r,r]}(x) \quad (3.11)$$

These convolution kernels define standard operators  $A_k$  by

$$A_k f = \varphi_k * f. \quad (3.12)$$

They are continuous smoothing operators of the same order in Sobolev as well as in Besov scales



$$A_k : H^s \rightarrow H^{s+k}$$

$$A_k : B_p^s(L_p) \rightarrow B_p^{s+k}(L_p) .$$

The extension to higher dimensions by tensor products is straightforward, the case of general kernels  $k$  is addressed in Theorem 3.9.

The appropriate mathematical model is a sum of delta peaks

$$f(x) = \sum_{k=1}^N f_k \delta(x - k). \quad (3.13)$$

Suitable model spaces for point-like objects are defined via Besov norms: delta peaks in  $\mathbb{R}$  are elements of any Besov space  $B_p^s(L_p(\mathbb{R}^d))$  satisfying  $(s+1)p < 1$ . The most important cases are  $p = 2$ , i.e. the classical Sobolev case of negative order ( $s < -1/2$ ), and  $p = 1$  which requires  $s < 0$ :

$$\delta \in B_1^{-\epsilon}(L_1(\mathbb{R})) \quad \text{or} \quad \delta \in H^{-1/2-\epsilon}(\mathbb{R}) \quad \text{for any } \epsilon > 0 .$$

### Model Problems and Regularization Techniques

As we have described in the previous section, there are various meaningful choices for different model spaces as well as for defining convolution operators. Discussing the most general choice would involve a jungle of indices, which would obscure the main objective of the present section: to show the importance of Besov regularization schemes for solving inverse convolution problems.

We will therefore concentrate in the following on analyzing one model problem in detail and address the general cases in some remarks.

Following the discussion on the different modeling alternatives in the previous section we will now select some illustrative choices for the convolution kernel, the model space for source conditions, the noise model and a particular solution  $f^+$ . We will focus on the one-dimensional case ( $d = 1$ ) in the following.

As convolution kernel we always choose a B-spline of order two in this section, the general case of an approximate kernel is discussed below.

The corresponding convolution operator  $A_2$ , defined by (3.12) is smoothing of order two in Sobolev- as well as in Besov scales:

$$A_2 : H^s(\mathbb{R}) \rightarrow H^{s+2}(\mathbb{R}) \quad \text{or} \quad A_2 : B_1^s(L_1(\mathbb{R})) \rightarrow B_1^{s+2}(L_1(\mathbb{R})) .$$

As usual, the unavoidable data error will require to choose some weaker norms in the image space.

**Problem 1** *This model is the appropriate physical model for the above mentioned real life deconvolution problems in mass spectroscopy as well as in astronomy. It has no direct analogon in the classical regularization theory.*

The problem of reconstructing sequences of Dirac peaks  $f^+ = \sum f_k \delta(\cdot - k)$  requires a model space  $X = B_1^{-\epsilon}(L_1(\mathbb{R}))$  with an arbitrary small but fixed  $\epsilon$ . We use a white noise model, hence  $g^\delta = g^+ + \delta dW$ .

We exploit the smoothing properties of  $A_2$ , i.e.  $Af^+ \in B_1^{2-\epsilon}(L_1(\mathbb{R}))$ . No additional source condition on the smoothness of  $f^+$  is required.

In all cases,  $A_2$  is smoothing images about two orders, i.e.  $f \in H^s(\mathbb{R})$  implies  $A_2 f \in H^{2+s}(\mathbb{R})$ .

## Convergence Analysis

Our primary objective concerns a mathematical analysis of the ‘practical approach’ as explained in the introduction. To this end we will first analyze Problem 1 and compare the convergence results with other settings.

The approximation properties of wavelet shrinkage operators are well studied by now. We will use the results of [10, Theorem 4], which state the following.

**Theorem 3.6** *Let  $s, \sigma, p, q, \alpha$  denote real numbers s.t.  $f \in B_q^s(L_q)$ ,  $0 < \alpha \leq 1$ ,  $q = (1 - \alpha)p$ ,  $s = \sigma p/q + (p/q - 1)/2$ ,  $\sigma - 1/p > -\alpha/2$ . Let further  $f^\delta = f + \delta dW$  and denote with  $S_\lambda$  the hard shrinkage operator. Define the threshold  $\lambda$  and the cut-off scale  $J$  via*

$$\lambda \simeq \sqrt{2|\log \delta|} \delta, \quad 2^J = -\frac{1}{2\delta^2 \log \delta} \quad (3.14)$$

and the projection  $P_J$  via

$$P_J f = \sum_{j \leq J, k \in \mathbb{Z}} \langle f, \tilde{\psi}_{j,k} \rangle \psi_{j,k}. \quad (3.15)$$

Then the estimator

$$f_\lambda^\delta = P_J S_\lambda f^\delta = \sum_{j \leq J, k \in \mathbb{Z}, |\langle f^\delta, \tilde{\psi}_{j,k} \rangle| > \lambda} \langle f^\delta, \tilde{\psi}_{j,k} \rangle \psi_{j,k} \quad (3.16)$$

fulfills

$$E(\|f_\lambda^\delta - f\|_{B_p^s(L_p)}^p) \lesssim (\sqrt{|\log \delta|} \delta)^{\alpha p}. \quad (3.17)$$

Here and in the following ‘ $a \simeq b$ ’ means that both quantities can be uniformly bounded by constant multiples of each other. Likewise, ‘ $\lesssim$ ’ indicates inequalities up to constant factors. We will need this result only for measuring the approximation error on the scale  $B_1^s(L_1)$  of Besov spaces.

## Analysis of the Model Problem

This is the basic deconvolution problem for sparse, peak-like structures, e.g. mass spectroscopy data or certain astrophysical images.

The convolution operator  $A_2$  maps a delta sequence  $f^+ = \sum f_k \delta(\cdot - k)$  to a sum of hat functions  $g^+ = A_2 f^+ = \sum f_k \varphi_2(\cdot - k)$  with  $\varphi_2$  defined by (3.11). For the rest of this section we denote  $\varphi = \varphi_2$ . Hence, the exact inverse deconvolution operator is well defined on such finite sums  $g = \sum c_k \varphi(\cdot - k)$  and yields a sequence of delta peaks:  $A_2^{-1} g = \sum c_k \delta(\cdot - k)$ .

We now exploit the denoising properties of wavelet shrinkage methods on the data side for given noisy data  $g^\delta = g^+ + \delta dW$ . The general result for the present situation is given by the following Corollary, see [26].

**Corollary 3.7** *Let  $g^+ = \sum_{k \in \mathbb{Z}} f_k \varphi(\cdot - k)$  denote a finite sum of second order B-splines, i.e.  $\{f_k\}$  is a finite set of non-zero indices. Then for every  $\varepsilon > 0$*

$$g^+ \in B_q^{1/q+1-\varepsilon}(L_q). \quad (3.18)$$

Assume  $g^\delta = g^+ + \delta dW$  and let  $\lambda$  and  $J$  be chosen according to (3.14).

For  $0 < \varepsilon < 3/2$  and for every  $3/2 > \tau \geq \varepsilon$  and  $\alpha = 1 - \frac{3-2\tau}{3-2\varepsilon}$  we obtain the convergence rate

$$E(\|P_J S_\lambda g^\delta - g^+\|_{B_1^{2-\tau}(L_1)}) \lesssim (\sqrt{|\log \delta|} |\delta|)^\alpha. \quad (3.19)$$

This is an approximation result on the data side, which needs to be transferred to an estimate on the reconstruction side, see again [26].

**Theorem 3.8** *Let  $f^+ = \sum f_k \delta(\cdot - k)$ ,  $g^+ = A_2 f^+$  and  $g^\delta = g^+ + \delta dW$ . The ‘practical approach’ as described in the introduction produces a regularized deconvolution of  $g^\delta$  as*

$$f_\lambda^\delta = A_2^{-1} P_0 S_\lambda g^\delta. \quad (3.20)$$

If  $\lambda$  is chosen according to (3.14) the following convergence rate holds for every  $3/2 \geq \tau > \varepsilon > 0$ :

$$E(\|f_\lambda^\delta - f^+\|_{B_1^{-\tau}(L_1)}) \lesssim (\sqrt{|\log \delta|} |\delta|)^{1 - \frac{3-2\tau}{3-2\varepsilon}}. \quad (3.21)$$

We want to remark, that this theorem justifies the ‘practical approach’ of just plotting the remaining wavelet decomposition as described in the reconstruction. Hence, this approach, which is a pure shrinkage technique, can indeed be interpreted as a regularization method, which converges—arbitrarily slow—to the delta sequence of the exact solution. The convergence rate  $\left(1 - \frac{3-2\tau}{3-2\varepsilon}\right)$  can be made better by choosing  $\tau$  larger and  $\varepsilon$  smaller, but this weakens the norm in which we measure the convergence. Moreover, a generalization to infinite sequences of delta peaks  $f^+ = \sum f_k \delta(\cdot - k)$  is obvious as long as  $\sum_{k \in \mathbb{Z}} |f_k| < \infty$ .

*Remark 1.* Implicitly, the practical approach makes use of the equivalent description of shrinkage methods via a variational approach. Minimizing

$$\|Af - g^\delta\|_{L_2} + \alpha\|f\|_{L_0}$$

for an operator which can be diagonalized by a wavelet basis also leads to a hard shrinkage approach. Hence, the practical approach can be regarded as some type of Tikhonov regularization in  $L_p$  spaces. However, we use a different noise model and measure the reconstruction error in a Besov space.

### Approximate Kernels

The previous sections have analyzed regularization methods for reconstruction sequences of delta peaks from convolution data. These results were based on the assumption, that the convolution kernel equals a B-spline. Obviously, they immediately extend to other wavelet kernels, i.e. we obtain the same convergence results for any convolution operator with kernel function  $\varphi$  whenever  $\varphi$  can be extended to a bi-orthogonal wavelet bases with a norm equivalence as stated in (3.1) and (3.2).

However, this is still very restrictive. In this section we address the case of a general kernel  $k$ , see again [26].

**Theorem 3.9** *Let the assumptions of Theorem 3.8 be satisfied. Assume that the kernel  $k$  is approximated by the scaling function  $\varphi$ :  $\|k - \varphi\|_{B_p^\tau(L_p)} < \epsilon$ , and let  $Af = k * f$ . The ‘practical approach’ with kernel  $\varphi$  applied to the noisy data  $g^\delta$  yields an approximation*

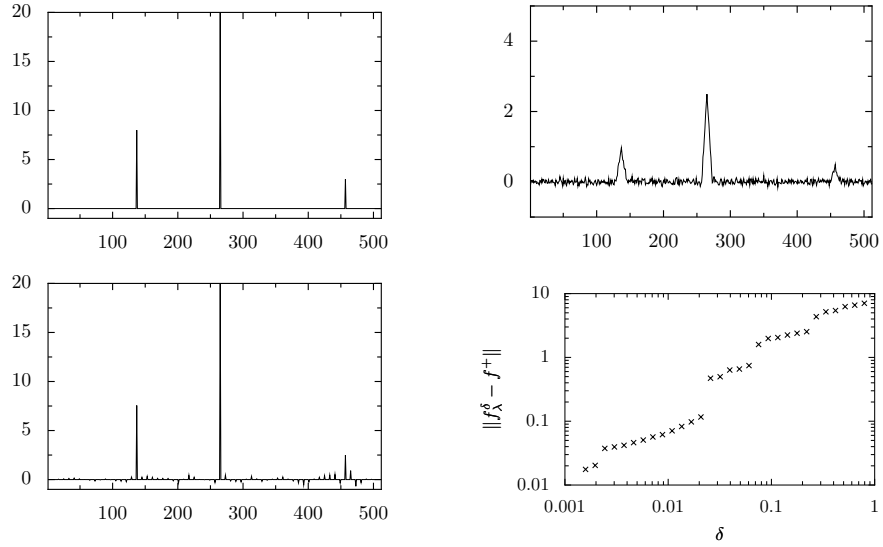
$$E(\|f - A^{-1}P_0S_\lambda g^\delta\|_{B_1^{-\tau}(L_1)}) \lesssim \|f\|_{B_1^{-\epsilon}(L_1)} + \left(\delta\sqrt{|\log \delta|}\right)^{\frac{2(\tau-\epsilon)}{3-2\epsilon}}.$$

### Numerical Simulations

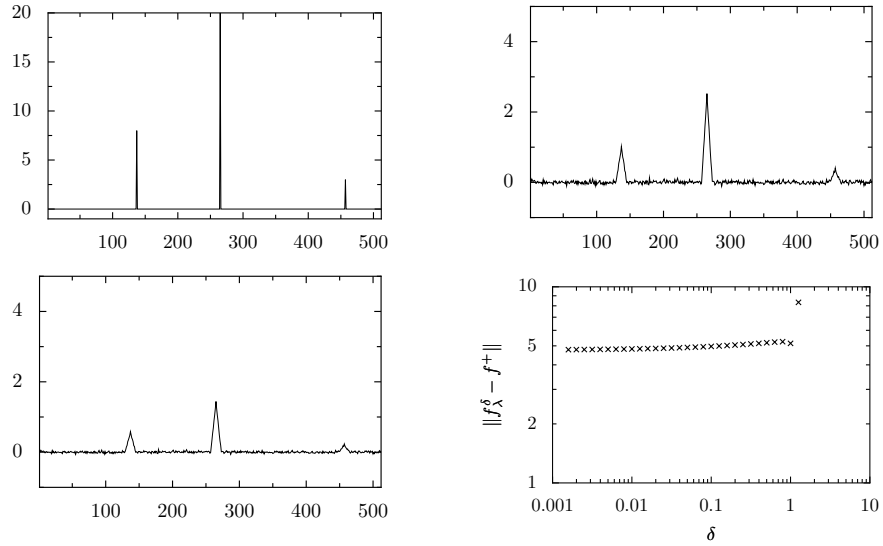
For the numerical simulations we use an artificial example as well as a real world example from mass spectrography data. We start with an example where the convolution kernel coincides with a B-spline scaling function. We will discuss a kernel, which is only roughly approximated by the B-spline scaling function in the following example.

First we are going to illustrate the practical approach as described in the introduction for our model problem with simula data. We choose discrete data sets of 512 data-points. The solution  $f^+$  is given by three delta-peaks of different height and the convolution kernel is a hat function. In our case we used a bi-orthogonal wavelet base of the class `bior2.x` and hence the reconstructing scaling function is a hat function. The noisy data  $g^\delta$  was generated by adding white noise of variance  $\delta$ . We chose the shrinkage parameter  $\lambda$  according to (3.14).

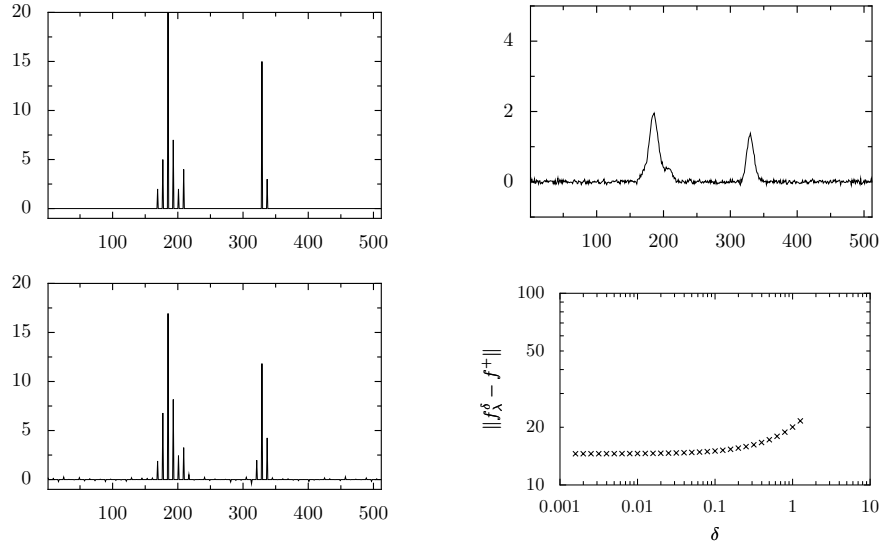
Figure 10 shows the true solution, the data, the reconstruction and an illustration of the convergence rate for  $\delta \rightarrow 0$ . Note that the convergence for  $\delta \rightarrow 0$  shows very different behavior in different regions: slow convergence



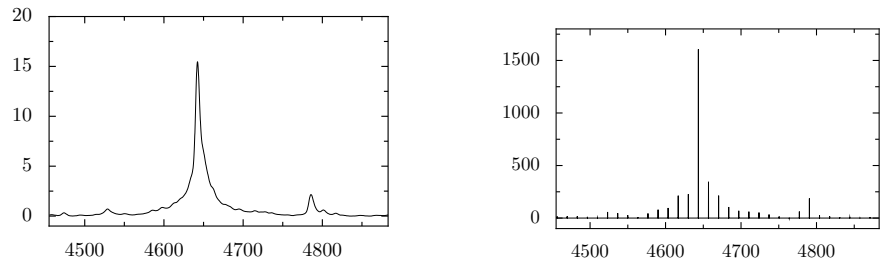
**Fig. 10.** Illustration of the performance of the practical approach. Upper left: The true solution  $f^+$ , upper right: the noisy data  $g^\delta$  with approximately 12% relative error, lower left: the reconstruction by the practical approach, lower right: log-log plot of the reconstruction error measured in  $B_1^{-\tau}(L_1)$  against the noise level.



**Fig. 11.** Illustration of the performance of Tikhonov regularization for the reconstruction of delta peaks. Upper left: The true solution  $f^+$ , upper right: the noisy data  $g^\delta$  with approximately 12% relative error, lower left: the reconstruction by Tikhonov regularization, lower right: log-log plot of the reconstruction error measured in  $H^{-1/2}$  against the noise level.

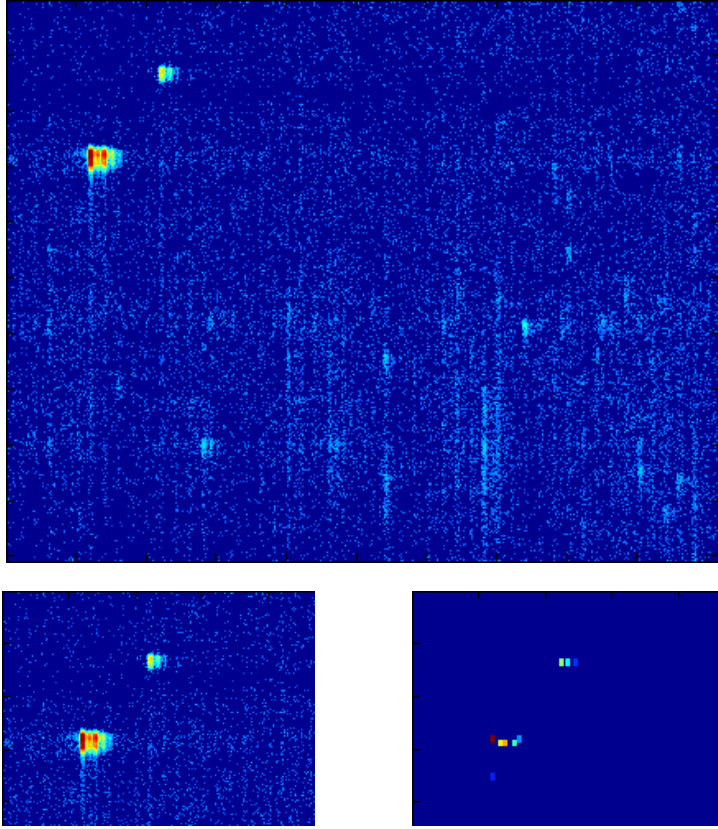


**Fig. 12.** Illustration of the performance of the practical approach for the reconstruction from overlapping peaks where the reconstruction scaling function does not fit to the convolution kernel. Upper left: The true solution  $f^+$ , upper right: the noisy data  $g^\delta$  with approximately 12% relative error, lower left: the reconstruction by the practical approach, lower right: log-log plot of the reconstruction error measured in  $B_1^{-\tau}(L_1)$  against the noise level.



**Fig. 13.** Deconvolution of real world data. Left: A section a spectrogram, right its deconvolution by the practical approach.

interrupted by jumps. The slow convergence is the behavior which is expected asymptotically (since we used  $\tau = .1$ , formula (3.21) predicts a convergence rate of  $1/15$  which is close to the results). The jumps have a simple explanation: As observable in Figure 10 the reconstruction does not only show the delta peaks but also a number of smaller peaks, which are a result of the data errors which escape the shrinkage step. When noise level  $\delta$  and shrinkage parameter  $\lambda$  tend to zero it happens that more and more of these false peaks



**Fig. 14.** Deconvolution of real world data. Top: 2D spectrum from drug metabolism, bottom left: part of the spectrum, bottom right: deconvolved spectrum part

are wiped out and a jump in the reconstruction error occurs every time this happens.

In a second experiment we compare the ‘practical approach’ with standard Tikhonov regularization. We used the same data as for the first experiment and minimized the functional  $J_\gamma(f) = \|Af - g^\delta\|_{H^{-1/2}}^2 - \alpha\|f\|_{H^{-1/2}}^2$ . To illustrate the convergence rate for  $\delta \rightarrow 0$  we used the optimal regularization parameter  $\gamma \approx \delta^{1/2}$ . The results shown in Figure 11 show the expected behavior: oversmoothing of the regularized reconstruction and a very low convergence rate.

In the last experiment we used as convolution kernel a B-spline of order four but the same reconstruction scaling function (hat function) as in the other experiment: a B-spline of order two. The data consists of overlapping peaks. According to Theorem 3.9 the reconstruction by the practical approach leads to very good results. As Figure 12 shows, the position of the major peaks is reconstructed perfectly. The height of the major peaks is slightly wrong and

there are some small sidepeaks in the reconstruction which are due to the fact the kernel and reconstruction scaling function do not fit together.

We now present reconstruction results for 1D and 2D mass spectroscopy data. Figure 13 shows the deconvolution of real world data from a MALDI/SELDI-TOF mass spectrometer provided by an AutoFlex II by Broker Daltonics [31]. We have chosen a section of with a large peak consisting of different isotopes and two small peaks. We deconvolved the data by the practical approach with the `bior2.8` bi-orthogonal wavelet base and threshold and finest scale chosen by hand. The deconvolution of 2D LCMS mass spectroscopy data follows the same outline. Figure 14 shows the practical approach applied to some real world data provided by Roche. The used data are from drug discovery metabolite identification and show metabolites of a drug produced by rat liver microsomes. The used spectrometer was a Sciex API 365 Triple Quadrupole operated in positive ion electrospray mode. Applying the practical approach is to calculate the wavelet decomposition with the `bior6.8` bi-orthogonal wavelet, cut off 90% of the detail coefficients, choose a finest scale and plot the positions of the major coefficients.

## References

1. R. A. Adams. *Sobolev Spaces*. Academic Press, New York, 1975.
2. K. Bredies and D. A. Lorenz. Iterated hard shrinkage for minimization problems with sparsity constraints. To appear in *SIAM J. Sci. Comput.*, 2007.
3. K. Bredies, D. A. Lorenz, and P. Maaß. A generalized conditional gradient method and its connection to an iterative shrinkage method. To appear in *Comput. Optim. Appl.*, 2005.
4. J. C. Brown. Inverse problems in astrophysical spectrometry. *Inverse Probl.*, 11:783–794, 1995.
5. J. C. Brown. Overview of topical issue on inverse problems in astronomy. *Inverse Probl.*, 11:635–638, 1995.
6. E. J. Candès. Compressive sampling. Sanz-Solé, Marta (ed.) et al., Proceedings of the international congress of mathematicians (ICM), Madrid, Spain, August 22–30, 2006. Volume III: Invited lectures. Zürich: European Mathematical Society (EMS), 2006.
7. E. J. Candès, J. K. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Commun. Pure Appl. Math.*, 59:1207–1223, 2005.
8. C. Cabrelli, C. Heil, and U. Molter. Accuracy of lattice translates of several multidimensional refinable functions. *J. Approximation Theory*, 95:5–52, 1998.
9. A. Cohen. *Numerical analysis of wavelet methods*. Studies in Mathematics and its Applications. 32. Amsterdam: North-Holland. xviii, 336 p., 2003.
10. A. Cohen, R. DeVore, G. Kerkycharian, and D. Picard. Maximal spaces with given rate of convergence for thresholding algorithms. *Appl. Comput. Harmon. Anal.*, 11(2):167–191, 2001.
11. A. Cohen, R. DeVore, P. Petrushev, and H. Xu. Nonlinear approximation and the space  $BV(\mathbb{R}^2)$ . *Am. J. Math.*, 121:587–628, 1999.



12. R. Coifman and D. Donoho. Translation-invariant de-noising. *Wavelets and Statistics*, A. Antoniadis and G. Oppenheim, eds., Springer-Verlag, 21:125–150, 1995.
13. S. Dahlke, E. Novak, and W. Sickel. Optimal approximation of elliptic problems by linear and nonlinear mappings II. *J. Complexity*, 22:549–603, 2006.
14. I. Daubechies. *Ten lectures on wavelets*. CBMS-NSF Regional Conference Series in Applied Mathematics. 61. Philadelphia, PA: SIAM, Society for Industrial and Applied Mathematics. xix, 357 p., 1992.
15. I. Daubechies, M. Defrise, and C. DeMol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Commun. Pure Appl. Math.*, 57(11):1413–1457, 2004.
16. I. Daubechies and G. Teschke. Wavelet-based image decomposition by variational functionals. *Wavelet Applications in Industrial Processing; Frederic Truchetet; Ed.*, 5266:94–105, 2004.
17. I. Daubechies and G. Teschke. Variational image restoration by means of wavelets: simultaneous decomposition, deblurring and denoising. *Appl. Comput. Harmon. Anal.*, 19(1):1–16, 2005.
18. I. Daubechies, G. Teschke, and L. Vese. Iteratively solving linear inverse problems with general convex constraints. *Inverse Probl. Imaging*, 1(1):29–46, 2007.
19. B. Han. Symmetry property and construction of wavelets with a general dilation matrix. *Linear Algebra Appl.*, 353(1–3):207–225, 2002.
20. B. Han. Symmetric multivariate orthogonal refinable functions. *Appl. Comput. Harmon. Anal.*, 17(3):277–292, 2004.
21. M. Hanke and J. G. Nagy. Restoration of atmospherically blurred images by symmetric indefinite conjugate gradient techniques. *Inverse Probl.*, 12(2):157–173, 1996.
22. M. Hegland and R. S. Anderssen. Resolution enhancement of spectra using differentiation. *Inverse Probl.*, 21(3):915–934, 2005.
23. R. Q. Jia. Shift-invariant spaces and linear operator equations. *Isr. J. Math.*, 103:259–288, 1998.
24. Q. T. Jiang. Multivariate matrix refinable functions with arbitrary matrix dilation. *Trans. Am. Math. Soc.*, 351:2407–2438, 1999.
25. N. Kinsbury. *Image processing with complex wavelets*. Phil. Trans. R. Soc. Lond., 1999.
26. E. Klann, M. Kuhn, D. A. Lorenz, P. Maaß, and H. Thiele. Shrinkage versus deconvolution. Submitted for publication, 2007.
27. K. Koch. Interpolating scaling vectors. *Int. J. Wavelets Multiresolut. Inf. Process.*, 3(3):389–416, 2005.
28. K. Koch. *Interpolating Scaling Vectors and Multiwavelets in  $\mathbb{R}^d$* . PhD thesis, Philipps-University of Marburg, 2006.
29. K. Koch. Multivariate symmetric interpolating scaling vectors with duals. Preprint Nr. 145, SPP 1114, to appear in *J. Fourier Anal. Appl.*, 2006.
30. K. Koch. Nonseparable orthonormal interpolating scaling vectors. *Appl. Comput. Harmon. Anal.*, 22(2):198–216, 2007.
31. M. Lindemann, M. Diaz, P. Maaß, F.-M. Schleif, J. Decker, T. Elssner, M. Kuhn, and H. Thiele. Wavelet based feature extraction in the analysis of clinical proteomics mass spectra. In *54th Conference of the American Society for Mass Spectrometry*, 2006.
32. S. Mallat. Multiresolution approximation and wavelet orthonormal bases for  $L^2(\mathbb{R})$ . *Trans. Am. Math. Soc.*, 315:69–88, 1989.

33. Y. Meyer. Oscillating patterns in image processing and nonlinear evolution equations. *University Lecture Series Volume, 22*, 2002.
34. S. Osher, A. Sole, and L. Vese. Image decomposition and restoration using total variation minimization and the  $H^{-1}$  norm. *SIAM J. Multiscale Model. Simul.*, 1(3):349–370, 2003.
35. S. Osher and L. Vese. Modeling textures with total variation minimization and oscillating patterns in image processing. *J. Sci. Comput.*, 19(1-3):553–572, 2003.
36. S. Osher and L. Vese. Image denoising and decomposition with total variation minimization and oscillatory functions. *J. Math. Imaging Vis.*, 20(1-2):7–18, 2004.
37. R. Ramlau and G. Teschke. Tikhonov replacement functionals for iteratively solving nonlinear operator equations. *Inverse Probl.*, 21:1571–1592, 2005.
38. L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variations based noise removal algorithms. *Physica D*, 60:259–268, 1992.
39. I. W. Selesnick. Hilbert transform pairs of wavelet bases. *IEEE Signal Processing Letters*, 8(6):170–173, 2001.
40. P. Wojtaszczyk. *A Mathematical Introduction to Wavelets*. Cambridge University Press, 1997.
41. X.-G. Xia and Z. Zhang. On sampling theorem, wavelets and wavelet transforms. *IEEE Trans. Signal Process.*, 41:3524–3535, 1993.