Heim, Susanne

Working Paper

Wavelets for diffusion tensor imaging

Discussion Paper, No. 505

Provided in Cooperation with:
Collaborative Research Center (SFB) 386: Statistical Analysis of discrete structures - Applications in Biometrics and Econometrics, University of Munich (LMU)

Suggested Citation: Heim, Susanne (2007) : Wavelets for diffusion tensor imaging, Discussion Paper, No. 505, Ludwig-Maximilians-Universität München, Sonderforschungsbereich 386 - Statistische Analyse diskreter Strukturen, München, https://doi.org/10.5282/ubm/epub.1870

This Version is available at:
https://hdl.handle.net/10419/31074

Wavelets for Diffusion Tensor Imaging

– corrected version from February 14, 2007 –

Susanne Heim

Department of Statistics, Ludwig-Maximilians-University,

Ludwigstrasse 33/II, 80539 Munich, Germany

susanne.heim@stat.uni-muenchen.de

Abstract

In this paper, wavelet basis functions are investigated for their suitability for processing and analysing diffusion tensor imaging (DTI) data. First, wavelet theory is introduced and explained by means of 1d and 2d examples (Sections 1.1 – 1.3). General thresholding techniques, which serve as regularization concepts for wavelet based models, are presented in Section 1.4. Regularization of DTI data can be performed at two stages, either immediately after acquisition (Wirestam et al., 2006) or after tensor estimation. The latter stage of denoising is outlined in Section 6 together with the incorporation of the positive definiteness constraint using log-Cholesky parametrization. In Section 3, the procedure is examined in a simulation study and compared to standard processing and the space-varying coefficient model (SVCM) based on B-splines (Heim et al., 2007). In addition, a real data example is presented and discussed. Finally, an approach is proposed how a space-varying coefficient model could fairly be adapted to wavelet basis functions. The theoretical parts are based on books of Gençay et al. (2002, Chap. 1, 4-6), Härdle et al. (1998), Ogden (1997) and Jansen (2001) if not stated otherwise. For an introduction to diffusion tensor imaging refer to Heim et al. (2007, Chap. 2).

Key words: Wavelets; Varying coefficient model; Diffusion tensor; Brain imaging;

1 Introduction to wavelets

Before defining wavelets and proceeding with their applicability to DTI data, I would like to reason why the use of wavelets seems to be promising. Without loss of generality, space will be the location index below, comprising time as special 1d space.
Looking at Fig. 1, the partition of the space-frequency plane is displayed for different transformations. B-splines, as used in the previous chapter, are only localized in the space domain (Fig. 1 (a)). For example, zero-degree B-splines are piecewise constants covering one single interval. Hence, a 1d sequence of DTI signal intensities can fully be represented by as many zero-degree B-splines as the sequence contains observations. In contrast, a Fourier analysis using cosine and sine functions offers full resolution in the frequency domain only (Fig. 1 (b)). Additional coarse localization in space can be gained by windowing the original series and applying the Fourier analysis across each window (Fig. 1 (c)). This modification is also known as short-time Fourier transform or Gabor transform after its originator. However, spatial resolution is increased at the expense of frequency resolution and spatial localization is still lacking within the fixed windows. This phenomenon is due to Heisenbergs uncertainty principle, stating that a particle cannot exactly be described by its position and momentum at the same time. Wavelet transforms, by contrast, intelligently adapt to the space-frequency plane in that a narrow space-window is provided to examine high-frequency content, but a wide space-window is allowed when investigating low-frequency components (Fig. 1 (d)). Due to their resulting localization in time/space *and* in frequency the use of wavelets promises local adaptiveness. In other words, the wavelet transform can be expected to zoom in higher frequencies when distinctive spikes or sharp edges are met.

The switch to wavelets is furthermore motivated by recent reports about successful wavelet based filtering of MR images with best contrast rendering between different tissue types and best spatial adaptivity. A selection of articles on the application of wavelets to radiological imaging can be found in the special issue of the IEEE Journal on Transactions in Medical Medicine (March 2003). It was however in the early nineties when Weaver et al. (1991) initially proposed the removal of random noise from MR images using selective wavelet reconstruction which Donoho and Johnstone (1994) developed to a full statistical

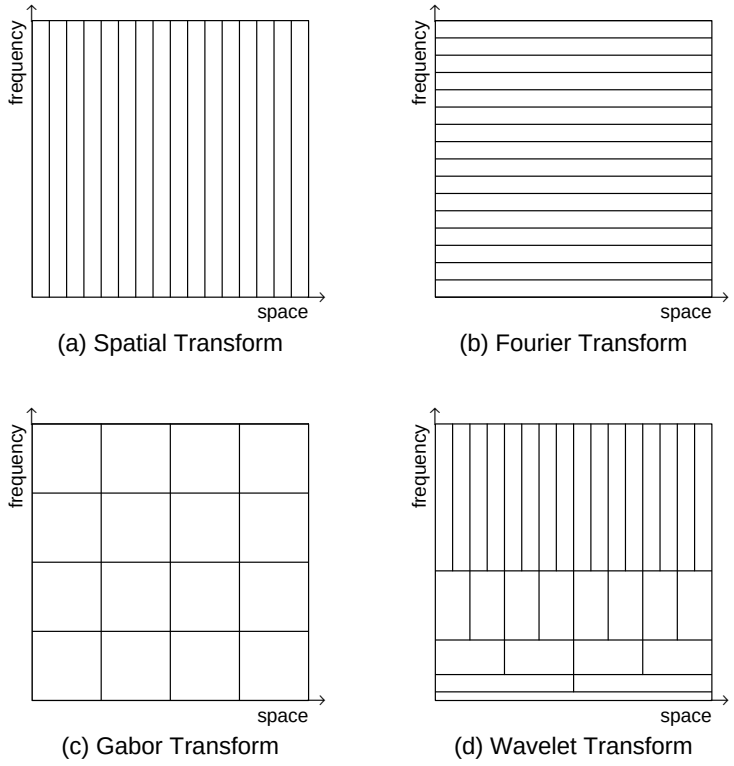


Figure 1: Partition of the space-frequency plane by different transforms.

technique. There is also a series of papers using wavelets in order to denoise fMRI data as well as to enhance the power of statistical analyses, see e. g. Van de Ville et al. (2006) and the references therein. Substantial contributions to the processing of MR modalities also stem from the Biomedical Imaging Group (École Polytechnique Fédérale, Lausanne). See the steadily updated homepage[1] for the most recent publications.

1.1 Univariate wavelet basis

With the aim in view to design space-varying coefficient surfaces using 3d wavelet basis functions, let me launch from the classical, more intuitive one-dimensional case. To give a first impression, Fig. 2 shows different basis functions of the Daubechies extremal phase wavelet class.

There is always a pair of so-called *father* and *mother* wavelet functions, $\phi(x)$ and $\psi(x)$,

[1] http://bigwww.epfl.ch/

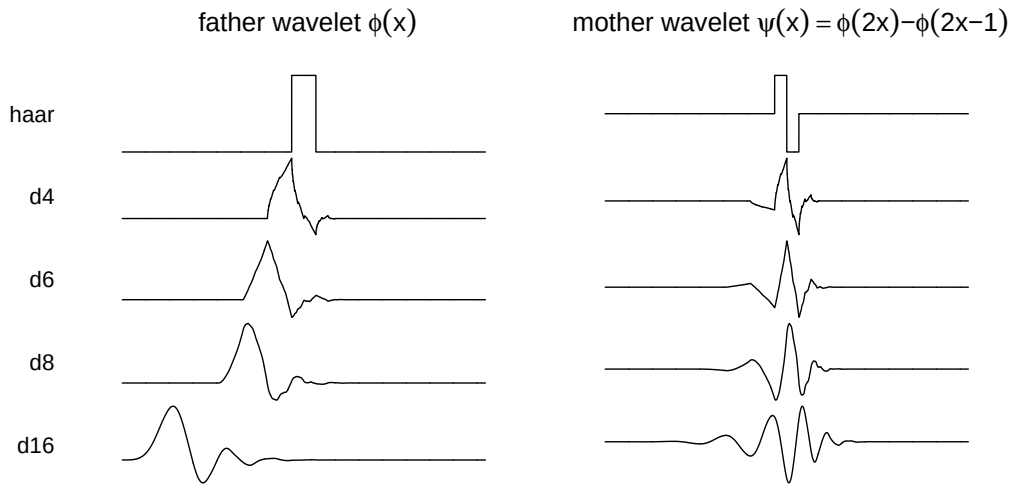


Figure 2: Principal representatives of the Daubechies extremal phase wavelet class.

respectively. The father wavelet is also called *scaling function*, while the mother wavelet is simply entitled *the* wavelet due to its prototype functionality for an entire set of wavelets. As most essential condition and kind of wedding ring, father and mother wavelet are orthogonal to each other. For the Haar family, this can immediately be seen from the graphic (top row) and, mathematically, by the corresponding definitions of the scaling function

$$\phi(x) \;=\; \begin{cases} 1, & x \in [0,1) \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

and the wavelet function

$$\psi(x) \;=\; \phi(2x) - \phi(2x-1) = \begin{cases} 1, & x \in [0,0.5) \\ -1, & x \in [0.5,1) \\ 0, & \text{otherwise}\,. \end{cases} \tag{2}$$

As furthermore obvious from Fig. 2, the representatives differ in (compact) support length and the degree of smoothness. Apart from the Haar family, asymmetry holds for the given wavelets and, in general, for most of the known families. Among the tremendous number, the main families are Daubechies extremal phase wavelets, Daubechies linear

phase wavelets (also known as least asymmetric wavelets or simply symlets) and coiflets. Some other classes involve B-splines such as the Battle-Lemarié and Chui-Wang cardinal spline families (both explained in Ogden (1997, pp. 23)), and the orthogonal fractional B-spline wavelets (Blu and Unser, 2003). Interestingly, the Haar wavelet cannot only be regarded as member of the Daubechies extremal phase but also of the Battle-Lemarié family. For an excellent description of how to construct wavelet bases see e. g. Härdle et al. (1998, Chap. 5-6).

Given its simplicity, the Haar wavelet will serve as the toy wavelet throughout this chapter. Most other wavelet classes do not even possess an explicit space-domain formula and are only defined implicitly.

There are two formal conditions for a function $\psi(x), x \in \mathbb{R}$, to be a wavelet. The *admissibility condition*

$$\int_{\mathbb{R}} \psi(x)dx = 0$$

ensures integration to zero and existence of the inverse. It is equivalent to a zero-valued zeroth moment. Further vanishing moments are reflected by the family attribute number, e. g. Daubechies-4 wavelets expose four vanishing moments. Seldom the father wavelet possesses vanishing moments, too, as it is the case for coiflets (Härdle et al., 1998). Note that some authors let the attribute encode for the length of the mother wavelet which is twice the number of vanishing moments. As evident from Fig. 2, the smoothness property is more pronounced for wavelets with larger support, as there is more time/space to die out to zero.

The second condition, namely the *unit energy condition*,

$$\int_{\mathbb{R}} \psi^2(x)dx = 1$$

constrains $\psi(x)$ to have normalized strength. The term *energy* refers to the "size" of a signal. Whenever a signal is imagined as a function of varying amplitude through time, the area under the curve seems to reason a good measurement of the signal strength. Since the possibly negative parts do not have less strength than a positive signal of the same size, energy is defined as the area under the squared signal. Hence the unit energy

condition forces $\psi(x)$, given its admissibilty, to take on nonzero values for some point x. As a consequence any wavelet function will oscillate or wave around the abscissa. The suffix 'let' comes in from discrete dilation and translation of the mother's shape, eventually resulting in baby wavelets

$$\psi_{j,k}(x) \quad = \quad \frac{1}{\sqrt{2^{-j}}}\psi\left(\frac{x-k2^{-j}}{2^{-j}}\right) = 2^{j/2}\psi(2^j x - k), \quad j \in \mathbb{Z}_0^+, k \in \mathbb{Z} \tag{3}$$

with support $[k2^{-j}, (k+1)2^{-j})$.

Wavelets are thus functions in two parameters. Parameter k accounts for dyadic translational displacement in the time domain whereas index j constitutes a dyadic change in frequency. The latter corresponds to compression or stretching of the wavelet on the time axis as illustrated by selected representatives in Figure 3. The amount of "localization" is therefore controlled by the dilation parameter j.

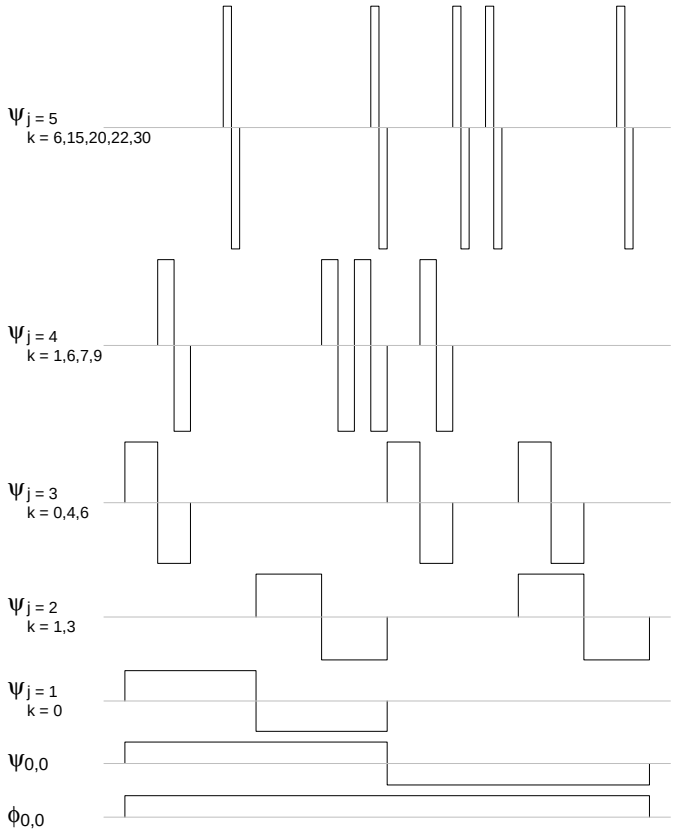


Figure 3: Selected members of the Haar family.

Application of these simple operations, namely dilation and translation, to the mother wavelet and multiplication with a normalization constant $2^{j/2}$ naturally leads to an orthonormal system

$$\left\{\{\phi_{0,k}\}, \{\psi_{j,k}\}, k \in \mathbb{Z}, j \in \mathbb{Z}_0^+\right\}$$

in $L^2(\mathbb{R})$, the space of square integrable functions[2] (Härdle et al., 1998). The system of wavelet functions allows to approximate any function $f \in L^2(\mathbb{R})$ aribitrarily well. For a pre-assigned precision, the system turns into an orthonormal basis for a subspace of $L^2(\mathbb{R})$. The shifted versions $\phi_{0,k}(x) = \phi(x-k)$ guarantee the coverage of $\mathbb{R}$. An outline of the theoretical underpinnings are postponed to the next section.

Orthogonality of the basis functions is achieved via critical sampling of the space-frequency plane at the nodes in Fig. 1 (d). This way the continuous and redundant partitioning becomes discrete and orthogonal with a minimum number of discrete points in space $k2^{-j}$ and discrete frequencies 2^{-j}. Wavelets obtained by Eq. (3) are therefore named discrete wavelets though being (piecewise) continuous functions.

1.2 Wavelet decomposition and reconstruction

In analogy to the Fourier or the zero-degree B-spline transform, any function $f \in L^2(\mathbb{R})$ can be approximated by a linear combination of wavelets

$$f(x) \quad = \quad \sum_{k=-\infty}^{\infty} \alpha_{0k}\phi_{0k}(x) + \sum_{j=0}^{\infty}\sum_{k=-\infty}^{\infty} \beta_{jk}\psi_{jk}(x),$$

with cofficients α_{0k} and β_{jk} (Härdle et al., 1998). This discrete decomposition directly results from the basis of discrete wavelets claimed in Section 1.1 which will roughly be reasoned in the following. In practice, only a discretized version of the function of interest is observed and this, in general, under some additional noise. Assume the moreover finite number of measurements to be equally spaced on the support region of dyadic length, say $[0, n = 2^J)$. This constitutes a square-summable[3] function f^J in the space V_J, defined as

$$V_J \quad = \quad \{f \in l^2(\mathbb{Z}) : f \text{ is piecewise constant on } [k2^{-J}, (k+1)2^{-J}), k \in \mathbb{Z}\}.$$

[2] A function f is square integrable or of finite energy if $\int_{\mathbb{R}} f^2(x)dx < \infty$.

[3] A sequence f is square-summable if $\sum_{x\in\mathbb{Z}} f^2(x) < \infty$.

Here, $l^2(\mathbb{Z})$ denotes the space of square-summable functions with integer domain. To gain a better understanding of how the concept of the discrete wavelet transform is derived subsequently, consider $J = 3$ and Figure 4.

An approximation to f^J can be generated by a member f^{J-1} of the subspace

$$V_{J-1} \quad = \quad \{f \in l^2(\mathbb{Z}) : f \text{ is piecewise constant on } [k2^{-J+1}, (k+1)2^{-J+1}), k \in \mathbb{Z}\}.$$

As the *level* $j, j = 0, \ldots, J$, decreases, the approximations f^j turn coarser with values on intervals of half the size as those for V_{j+1} (see gray-colored basis functions in Fig. 4). Eventually, the so-called *reference space* V_0 is reached, containing step functions on unit intervals. The overall sequence of spaces $(V_j)_{j=0,\ldots,J}$ fulfills

$$V_0 \subset V_1 \subset V_2 \subset \ldots$$

where V_0 can be build by the set of integer-shifted versions of the Haar scaling function $\{\phi_{0,k} = \phi(\cdot - k), k \in \mathbb{Z}\}$. As obvious from Fig. 4, an orthonormal basis for any other $V_j, j > 0$, is formally obtained by dilations of the father wavelet: $\{\phi_{j,k} = 2^{j/2}\phi(2^j \cdot - k), k \in$

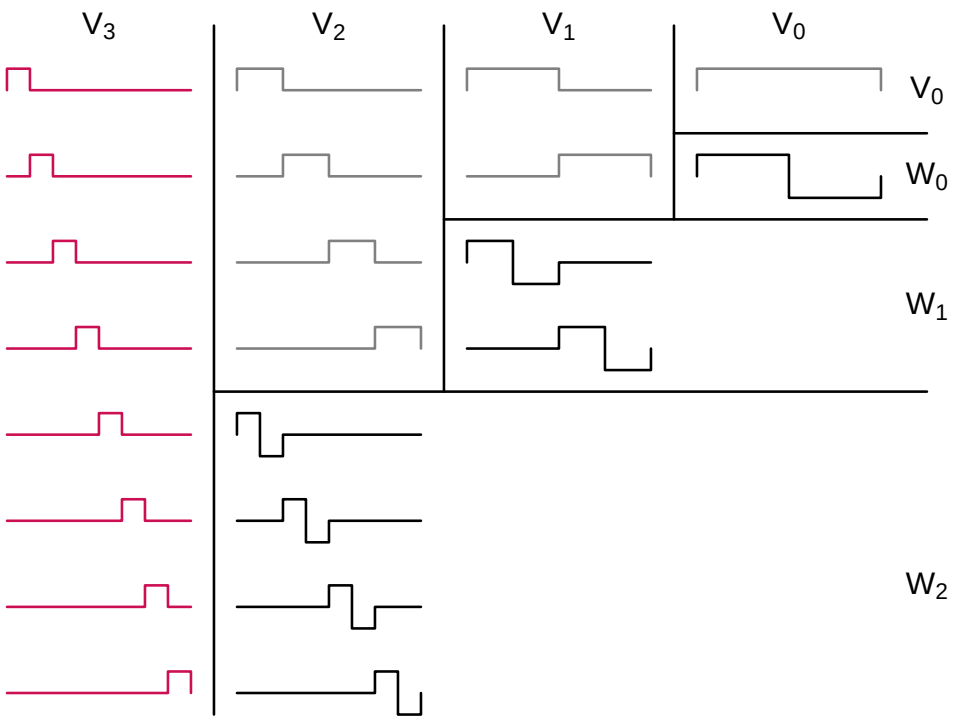


Figure 4: Repeated splitting of a vector space into two orthogonal subspaces build the basis for the discrete wavelet transform.

$\mathbb{Z}\}$. As a consequence, the central *two-scale relationship* can be established:

$$\phi_{j,k}(x) \quad = \quad \big(\phi_{j+1,2k}(x) + \phi_{j+1,2k+1}(x)\big)/\sqrt{2}.$$

This means that a function at a given level evolves by averaging at the next higher level (see also Figure 4).

Whenever a function f^j is expressed in terms of the next coarser approximation, the residuals represent the details of f^j that are lost due to the approximation. In this case, the details are constants of the same resolution as the function f^j itself and can thus be written as linear combination of the wavelets $\psi_{j-1,k}, k \in \mathbb{Z}$. As obvious from Figure 3, two wavelets with the same dilation index j but differing k cannot have overlapping support. Intra-level orthogonality is therefore guaranteed, so that the set of wavelets at a fixed level j defines a *detail space*:

$$W_j \quad = \quad \text{span}\{\psi_{j,k}, k \in \mathbb{Z}\}.$$

In the Haar example, it is easily ascertained that $\phi_{j,k}$ and $\psi_{j',k'}$ are orthogonal for $j \leq j'$ (compare black-colored functions in Figure 4). This implies that W_{j-1} is the orthogonal complement of V_{j-1} in V_j:

$$V_j \quad = \quad V_{j-1} \oplus W_{j-1}. \tag{4}$$

Similar to the two-scale equation, the relationship of Haar father (Eq. 1) and mother wavelet (Eq. 2) can be generalized:

$$\psi_{j,k}(x) \quad = \quad \big(\phi_{j+1,2k}(x) - \phi_{j+1,2k+1}(x)\big)/\sqrt{2}.$$

Hence, the residuals at a fixed level arise from differencing a function at the next higher level.

Now, mutual orthogonality across levels holds since one of the wavelets is constant in the shared support region (compare Figure 3). With this inter-space orthogonality between

W_j and $W_{j'}, j \neq j'$, recursive decomposition in a global approximative part and level specific details yields

$$V_j \;=\; V_{j-1} \oplus W_{j-1} = V_{j-2} \oplus W_{j-2} \oplus W_{j-1} = \ldots = V_0 \oplus \bigoplus_{l=0}^{j-1} W_l. \tag{5}$$

To be more general, replace V_0 with V_{j_0} in Eq. (5), where $j_0 < j$ denotes the primary decomposition level. It furthermore holds that $L^2(\mathbb{R}) = \overline{\bigcup_{j\in\mathbb{Z}} V_j} = \overline{V_0 \oplus \bigoplus_{j=0}^{\infty} W_j} = \overline{\bigoplus_{j\in\mathbb{Z}} W_j}$ (Härdle et al., 1998).

In conclusion, a finite and noisy signal of dyadic length, $x \in \{1, \ldots, n = 2^J\} \subset \mathbb{Z}$, has the following wavelet expansion:

$$\begin{aligned} y(x) &= f(x) + \varepsilon(x) \\ &= \sum_{k=0}^{2^{j_0}-1} \alpha_{j_0 k}\phi_{j_0 k}(x) + \sum_{j=j_0}^{J-1}\sum_{k=0}^{2^j-1} \beta_{jk}\psi_{jk}(x) + \varepsilon(x), \quad \varepsilon(x) \overset{\text{i.i.d.}}{\sim} N(0,\sigma^2). \end{aligned} \tag{6}$$

The choice of the initial dilation level j_0 determines the quality of the approximation (Ogden, 1997), but will depend on personal experience and the given data because of missing binding guidelines. If $j_0 > 0$, one speaks of *partial* decomposition and, if $j_0 = 0$, of *complete* decomposition. In general, the number of vanishing moments of the father wavelet determines the degree to which a polynomial is still exactly reproducible in V_0. Hence, V_0 is equivalent to the null space of the smoothing spline penalty (Hastie et al., 2001, Chap. 5.9).

Provided the recorded time series is scaled to the unit support interval, the complete decomposition involves as many wavelet coefficients as there are observations. Namely, there is the father plus 2^j contributions at each level $j = 0, \ldots, J-1$. Summation of the geometric sequence leads to $1 + \sum_{j=0}^{J-1} 2^j = 1 + (2^J - 1) = n$. Note that the coefficient $\alpha_{0,0}$ coincides with the sample mean in the Haar case. Since an increase in level entails a bisection of the wavelet's support of 2^{-j} on simultaneous doubling of the frequency (Figure 3), the *resolution* enhances accordingly. Attention should be paid to the notation. Some authors revert the j-axis so that father and mother do no more lie at level $j = 0$. Level j is then rather used in the meaning of decomposition depth. The definition in Eq. 3 changes accordingly. In addition, the technical term *scale* sometimes replaces *level* or two to the power of *level*.

(Inverse) discrete wavelet transform

The totality of the coefficients α_{0k} and β_{jk} is called *discrete wavelet transform* (DWT). It is determined by the projection of the function f onto the subspaces V_0 and $W_j, j = 0, \ldots, J-1$, and reflects the degree of congruity of the function with the wavelets ϕ_{0k} and ψ_{jk}, respectively. While the α_{0k}'s summarize the general form of the function, the β_{jk}'s represent local details. In contrast to the Fourier transform, that operates on a single resolution level, the wavelet transform works on multiple resolution levels (Figure 1). The multiresolution property constitutes the ability of the wavelet transform to capture nonstationary events and to map irregulary occurring features. Hence, the recursive decomposition of an arbitrary function in $L^2(\mathbb{R})$ in its smooth and rough parts is also termed *multiscale* or *multiresolution analysis*. This becomes also immediately plausible from the example given in Fig. 4. A formal definition can be found in e. g. Ogden (1997). Metaphorically speaking, the visual acuity reduces continuously while stepping forward in the hierachically organized decomposition. For instance, leaving a forest can be compared to going away from fine structures of tiny leaves over twigs and branches, passing by stems and whole trees to the entire wood till mere green color remains. The peculiarity of the wavelet transform arises from the concurrence of the fine and coarse patterns.
Giving an example, Figure 5 depicts the decomposition of an NMR signal of length $n = 1024$ (Wavelab[4]) on the left side. The vertical bars correspond by position to the location and by length to the quantity of the level specific wavelet contributions. The largest weights emerge from slowly changing features, whereas highly resolved features do not play a big role for the representation of the NMR signal. Solely, the exclusive hallmark of the sharp spike is correctly caught by a few high frequency wavelets that have small support. This ideally demonstrates the adaptivity of wavelet analysis.

To gain further insight into the properties of wavelet decomposition and to introduce the

[4]Wavelab is a free library of Matlab rountines for wavelet analysis, downloadable from http://www-stat.stanford.edu/∼wavelab

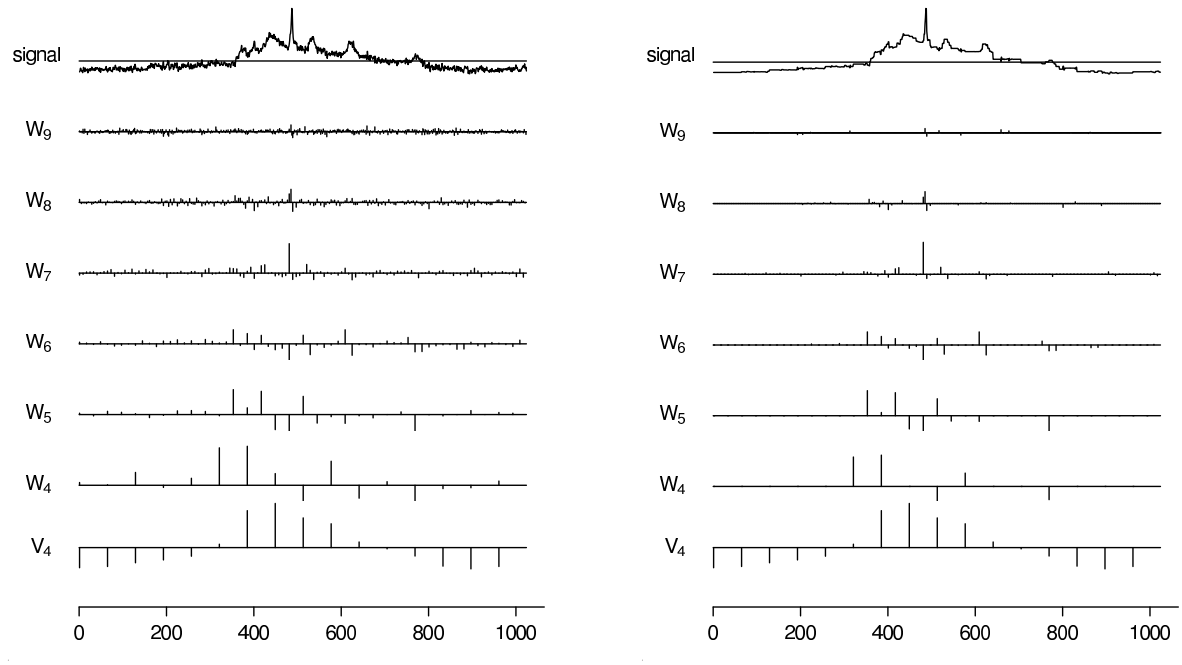


Figure 5: Decomposition (left) and denoised reconstruction (right) of an NMR signal using Haar wavelets.

inverse operation, Eq. (6) is reformulated in matrix notation setting $j_0 = 0$:

$$\begin{pmatrix} y(1) \\ y(2) \\ \vdots \\ y(n) \end{pmatrix} = \begin{pmatrix} \phi_{00}(1) & \psi_{00}(1) & \psi_{10}(1) & \psi_{11}(1) & \cdots & \psi_{J-1,\frac{n}{2}-1}(1) \\ \phi_{00}(2) & \psi_{00}(2) & \psi_{10}(2) & \psi_{11}(2) & \cdots & \psi_{J-1,\frac{n}{2}-1}(2) \\ \vdots & & & & & \\ \phi_{00}(n) & \psi_{00}(n) & \psi_{10}(n) & \psi_{11}(n) & \cdots & \psi_{J-1,\frac{n}{2}-1}(n) \end{pmatrix} \begin{pmatrix} \alpha_{00} \\ \beta_{00} \\ \beta_{10} \\ \beta_{11} \\ \vdots \\ \beta_{J-1,\frac{n}{2}-1} \end{pmatrix}$$

$$= \mathbf{W}\boldsymbol{\gamma}$$

Solving the corresponding system of normal equations leads to the ordinary least squares estimate of the wavelet transform:

$$\begin{aligned} \hat{\boldsymbol{\gamma}} &= (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{y} \\ &= \mathbf{W}'\mathbf{y}, \end{aligned}$$

the latter equality being due to the orthogonality of the basis functions in $\mathbf{W}$. In fact, this decomposition is loss-free, meaning that the information transported by $\mathbf{y}$ entirely

passes into the wavelet domain. Although concentrated in a much smaller fraction of the components than in the raw data, the wavelet transform maintains the sample variance, i. e. is *energy (variance) preserving*:

$$\text{Var}(\hat{\boldsymbol{\gamma}}) \quad = \quad \mathbf{W}'\text{Var}(\mathbf{y})\mathbf{W} = \sigma^2\mathbf{I} = \text{Var}(\mathbf{y}).$$

Even though, in the more general case, $\mathbf{y}$ might be correlated data with $\text{Var}(\mathbf{y}) = \boldsymbol{\Sigma}$, the DWT has the ability to dissolve the correlation structure and to produce a sparse covariance matrix $\text{Var}(\hat{\boldsymbol{\gamma}}) = \mathbf{W}'\boldsymbol{\Sigma}\mathbf{W}$. This means the wavelet coefficients exhibit a flat, almost white spectrum (Gençay et al., 2002, Chap. 5). Hence, the discrete wavelet transform is esteemed for its decorrelation or whitening property. The fMRI community recently detected this fundamental feature for their purpose of multiple test-correction: Significant clusters of activated voxels can be determined in the wavelet domain by the simple Bonferroni correction (Van de Ville et al., 2006).
Define the *inverse discrete wavelet transform* (IDWT) to be

$$\hat{\mathbf{y}} \quad = \quad \mathbf{W}\hat{\boldsymbol{\gamma}}.$$

Again supposing $j_0 = 0$, even perfect reconstruction of the input signal is accomplished:

$$\hat{\mathbf{y}} \quad = \quad \mathbf{W}\hat{\boldsymbol{\gamma}} = \mathbf{W}\mathbf{W}'\mathbf{y} = \mathbf{W}'\mathbf{W}\mathbf{y} = \mathbf{y}.$$

Here, 'perfect' is used in the sense of completely recoverable as no information is lost.

Preview on denoised reconstruction

Yet, the analyst's interest does usually not focus on the recorded measurements but on strategies to reveal the underlying function. As anticipated previously, the wavelet decomposition allows to determine basis functions of marginal meaning by their absolute weight. The left plot of Figure 5 suggests to discard, above all, coefficients at large levels where extremely localized features of high frequency are mapped. Similar to analysis of variance, coefficients are considered to be noise or, on the other hand, significantly important according to their explanatory strength. Applying an appropriate cut-off value and

performing the inverse wavelet transform yields a regularized version of the initial NMR signal. The result for the Haar basis is depicted on the right side of Figure 5. Particularly coefficients at large levels vanish, thereby suppressing rapid changes of the input signal. Though the essential form is clearly worked out, smoother wavelets are obviously required to meet the nature of the true NMR signal. On the other hand, the characteristic sharp peak is appropriately reproduced since wavelets concentrate the mass of oscillations on a small interval. Note the single distinct deflexion in W_7 that indicates the singularity of the original series. In total, a sparse and compressed wavelet representation of the input data is achieved. This mechanism also underlies the standardized image format JPEG2000[5] which allows the user to specify a compression rate. If no compression is desired, the loss-free reconstruction case is given with the same memory amount as required by the object itself. In the compression case, the use of special decoders leads to considerable memory reduction storing non-zero coefficients only. Multidimensional wavelets are presented in Section 1.3, while Section 1.4 deals with different *thresholding* and *shrinkage* approaches.

Practical matters

In view of the above, two more practical issues come to mind. First, the Haar basis clearly adapts to the unit interval, any finite sequence can be scaled to. But wavelets of other families cannot straightforwardly be forced to have support $\subseteq [0, 1)$. As a consequence, discontinuities at the interval boundaries can occur when a multiresolution analysis of the sequence is performed. These might be mitigated by symmetric or periodic expansion of the signal. Second, if a non-dyadic sampling rate is encountered, zero-padding or interpolating of the signal up to dyadic length has to be considered beside symmetric and periodic boundary handling (Ogden, 1997). Some approaches will introduce correlations among the coefficients. Rather than manipulating the data, other transforms for signals of unconstrained length seem advisable such as the maximum-overlap discrete wavelet transform (MODWT) (Gençay et al., 2002), biorthogonal wavelet transforms (Ogden, 1997), or lifting schemes (Jansen, 2001). The MODWT is also known as non-decimated, redundant, translation-invariant, or stationary wavelet transform.

[5]JPEG is based on cosine functions; JPEG2000 works with wavelets.

The fast wavelet transform algorithm

When it comes to implementation of the (inverse) discrete wavelet transform, Mallat's pyramidal algorithm is usually employed for its computational efficiency. The core aspect here is that the decomposition in smooth and rough parts can be interpreted as application of high- and low-pass filters. Indeed, the low-pass filter coefficients conform with the two-scale relationship of the scaling function:

$$l_k \;=\; <\phi, \phi_{1,k}> \overset{\text{Haar}}{=} \begin{cases} \frac{1}{\sqrt{2}}, & k = 0, 1 \\ 0, & \text{otherwise.} \end{cases}$$

By m-times application of filter operator L, represented by the sequence $\{l_k\}_{k\in\mathbb{Z}}$, to the scaling coefficients at a particular level, represented by $\alpha_{j,\cdot} = \{\alpha_{j,k}\}_{k\in\mathbb{Z}}$, scaling function coefficients at the m-times lower level are obtained:

$$\alpha_{j-m,\cdot} \;=\; L^m \alpha_{j,\cdot} = L^{m-1} \left\{ \sum_{i\in\mathbb{Z}} l_{i-2k} \alpha_{j,i} \right\}_{k\in\mathbb{Z}}.$$

The second equality sheds some light on the downsampling by two, inherent to each filtering step (compare Figure 6). This is hidden in the indices of $\{l_i\}$ that are all offset by two.

The high-pass filter operator H can be defined by the quadratur mirror relationship (see e. g. Gençay et al., 2002):

$$h_k \;=\; (-1)^k l_{1-k}, \quad k \in \mathbb{Z}.$$

Similar to the relation of father and mother wavelet, unique application of H to the scaling coefficients at level j leads to the wavelets coefficients at level $j-1$:

$$\beta_{j-1,\cdot} \;=\; H\alpha_{j,\cdot}.$$

Combining L and H yields

$$\beta_{j-m,\cdot} \;=\; HL^{m-1}\alpha_{j,\cdot}.$$

Iteratively subsampling every other coefficient in either filter steps gives the pyramidal or cascade scheme of the decomposition illustrated in Figure 6. The reconstruction follows the reverse principle. At a particular level, the upsampling step performs zero-padding and is necessary for the convolution with the filter sequence of double length.

The combination of low- and high-pass filters results in band-pass filters. Specifically, a band-pass filter over the frequency interval $[1/4, 1/2]$ results from serial application of an ideal filter pair L and H. For this purpose let L pass the frequencies lower than $1/2$ and H capture frequency dynamics greater than $1/4$. Therefore, one says that the pryamidal algorithm decomposes an input signal into sub-bands. In addtion, this viewpoint allows the following conclusion: The closer a wavelet filter approaches an ideal filter the less leakage of frequencies occurs, thus, the better the edge preservation property becomes (Gençay et al., 2002). Beyond this, the primary decomposition level is often chosen on a desired low-pass cut-off frequency.

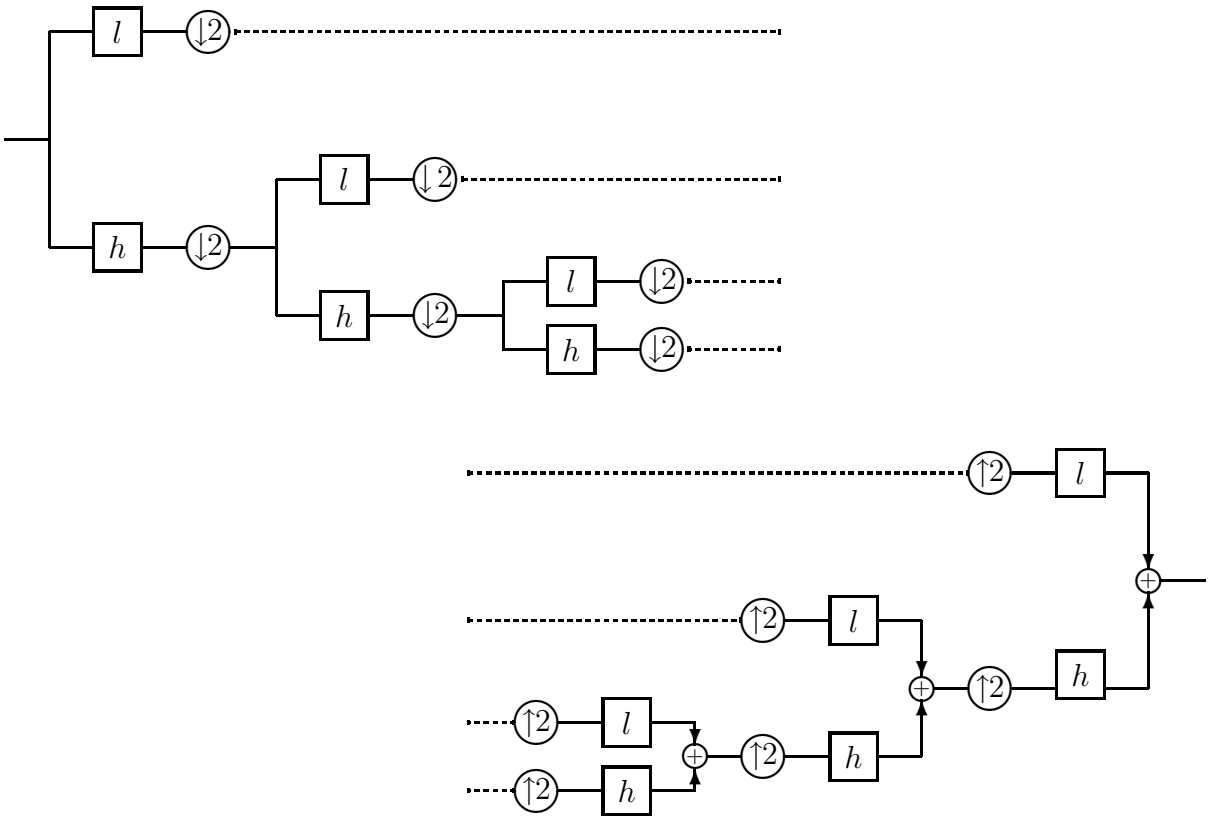


Figure 6: Decomposition (top) and synthesis (bottom) of a signal using the pyramidal algorithm. The branches refer to convolution with either low-pass ($\boxed{l}$) or high-pass filter coefficients ($\boxed{h}$). The icons $\downarrow 2$ and $\uparrow 2$ denote down- and upsampling by two.

1.3 Multidimensional wavelets decomposition - a 2d example

One-dimensional wavelets can easily be extended to the higher dimensional case via tensor products (see e. g. Ogden, 1997). Note that consecutive operations on columns, rows, layers etc of n-dimensional observational arrays generate *separable* basis functions. It is well known that these functions are likely to lack isotropy. In contrast, *non-separable* wavelets offer more freedom but at the expense of computational cost. Therefore, the discussion limits itself to separable wavelet functions. The decomposition idea is exemplified by means of 2d wavelets intended to analyze square-integrable 2d functions $f(x, y) \in L^2(\mathbb{R}^2)$.

Wavelet basis functions in $\mathbb{R}^2$

To begin with, define the 2d father: $\phi(x, y) = \phi(x)\phi(y)$. As before, translations and dilations of the father,

$$\{\phi_{j,k_1,k_2}(x, y) = 2^j\phi(2^j x - k_1, 2^j y - k_2), j, k_1, k_2 \in \mathbb{Z}\}$$

constitute a basis for $V_j^{x,y} = V_j^x \otimes V_j^y$. Deploying the multiresolution property and Eq. (4), the space $V_{j+1}^{x,y}$ of 2d block functions with resolution $2^{j+1} \times 2^{j+1}$ can be approximated by the successive space $V_j^{x,y}$ of twice as coarse squares and the complimentary residual space:

$$\begin{aligned} V_{j+1}^{x,y} &= V_{j+1}^x \otimes V_{j+1}^y \\ &= (V_j^x \oplus W_j^x) \otimes (V_j^y \oplus W_j^y) \\ &= (V_j^x \otimes V_j^y) \oplus \big((V_j^x \otimes W_j^y) \oplus (W_j^x \otimes V_j^y) \oplus (W_j^x \otimes W_j^y)\big) \\ &= V_j^{x,y} \oplus W_j^{x,y}. \end{aligned}$$

The rewriting reveals that the 2d space $W_j^{x,y}$ is made up of three orthogonal subspaces, each of which endowed with a tensor product basis of its components. So define the corresponding 2d mother wavelets

$$\begin{aligned} \psi^1(x, y) &= \phi(x)\psi(y) \\ \psi^2(x, y) &= \psi(x)\phi(y) \\ \psi^3(x, y) &= \psi(x)\psi(y) \end{aligned}$$

and build the wavelet basis for $W_j^{x,y}$ by means of translation and dilation:

$$\{\psi^d_{j,k_1,k_2}(x,y) = 2^j\psi^d(2^jx - k_1, 2^jy - k_2), d = 1,2,3, j, k_1, k_2 \in \mathbb{Z}\}.$$

Note that the basis functions are tensor products within the same scale. This distinguishes the *square* wavelet transform from the rectangular wavelet transform that involves tensor product across scales (Jansen, 2001). Figure 7 shows the four functions $\phi(x,y), \psi^1(x,y), \psi^2(x,y)$ and $\psi^3(x,y)$ associated with the 1d symlet with two vanishing moments. The clearly recognizable orientation in space tells us that ψ^1 emphazises edges in the horizontal direction, ψ^2 in the vertical and ψ^3 in the diagonal direction. These properties can be understood in view of the low- and high-pass filter analogons and are most intuitively accessed through the shapes of 2d Haar wavelets (Figure 8). Note also that the effective support of 2d tensor product wavelets shrinks by 2×2 when increasing the dilation parameter by one.

A finite and noisy 2d signal, e. g. an $(n_1 \times n_2)$-image Y, thus has the following wavelet

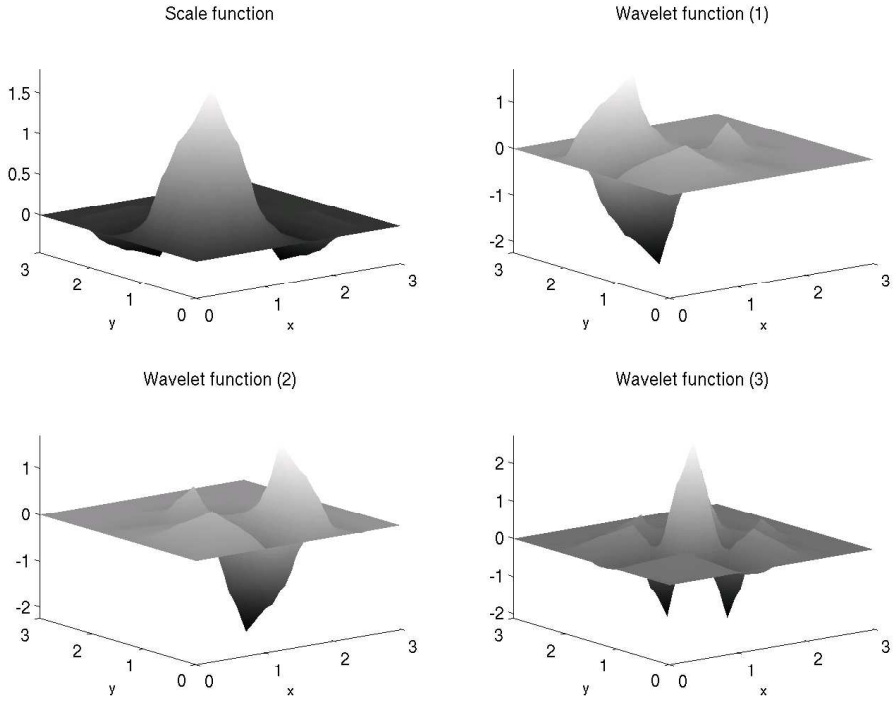


Figure 7: Scaling function and mother wavelet functions of 2d symlet with two vanishing moments.

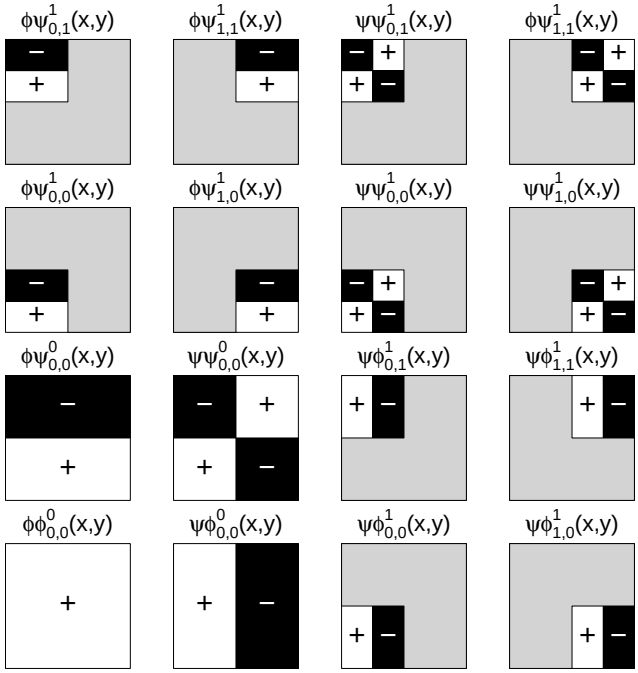


Figure 8: Exemplary basis functions of the 2d Haar family, projected onto the xy-plane. The exponent denotes the level j, while the subscripts correspond to shift parameters k_1 and k_2.

representation:

$$\begin{aligned} Y(x,y) &= f(x,y) + \varepsilon(x,y) \\ &= \sum_{k_1=0}^{2^{j_0}-1} \sum_{k_2=0}^{2^{j_0}-1} \alpha_{j_0,k_1,k_2} \phi_{j_0,k_1,k_2}(x,y) + \sum_{d=1}^{3} \sum_{j=j_0}^{J-1} \sum_{k_1=0}^{2^j-1} \sum_{k_2=0}^{2^j-1} \beta^d_{j,k_1,k_2} \psi^d_{j,k_1,k_2}(x,y) + \varepsilon(x,y), \end{aligned}$$

with $\varepsilon(x,y) \sim N(0,\sigma^2)$, $J = \log_2(\min(n_1,n_2))$.

2d decomposition scheme

In analogy to the 1d case, the pyramidal algorithm also adapts to the 2d wavelet decomposition. In a first step, a pair of low- and high-pass filters runs over the rows, and afterwards along the columns of the given image. This procedure is repeated on each purely low-pass filtered output, i. e. the coefficient matrix $(\alpha_{j,k_1,k_2})_{\substack{k_1=0,\dots 2^j-1 \\ k_2=0,\dots 2^j-1}}$. It is denoted by LL_j in Figure 9 for pointing out the filter link and filter order. Note that LL_J encodes for the input image itself. Convolution of the detail wavelets $\psi^1(x,y), \psi^2(x,y)$ and $\psi^3(x,y)$ with the approximating part LL_{j+1} results in horizontal, vertical and diagonal detail components LH_j,HL_j, and HH_j, respectively. Each iteration causes the filtered

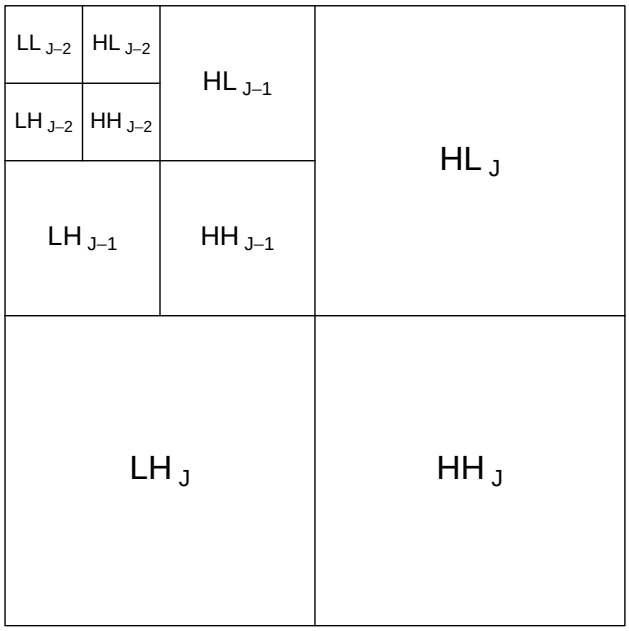


Figure 9: Nested two-dimensional signal decomposition on three scales.

output to be downsampled by 2 per direction which establishes the pyramidal structure and allows for the nested visualization. It also gives the algorithm its speed since in each step only 1/4 of the number of coefficients needs to be decomposed. The inverse process of reconstruction is governed by dyadic upsampling.

A black-and-white picture of the Siegestor in Munich serves as subject in Figure 10. Juxtaposed to the (256×256)-image matrix are two nested steps of the 2d DWT. The decomposition units exhibit their directional preferences in plain manner: brightness contrasts on columns (vertical features) are extracted separately from beams (horizontal features) and perspective alignments (diagonal features). It is worth emphazising that components including a high-pass filtering step have been adjusted for comparable visualization with the residual image. Background sprinkles are present, though hardly visible to the human eye due to poor reproduction quality. They can however be removed in an intermediate denoising, i. e. thresholding, step before performing the 2d IDWT.

By means of the Siegestor photography and the 2d Haar basis, I would like to elucidate the successive approximations associated with spaces $V_J \supset \ldots \supset V_0$. Figure 11 demonstrates how with decreasing level the grey colores contained in a scaling image concentrate on a

Figure 10: Original greyscaled image and 2d DWT based on Haar wavelets.

"grand mean" value. At the same time, the wavelet's blocky support reaches the size of the input image. Subsequent to the third decomposition level, the coefficients do no more reflect any visual features of the original image. Therefore, an appropriate DWT would

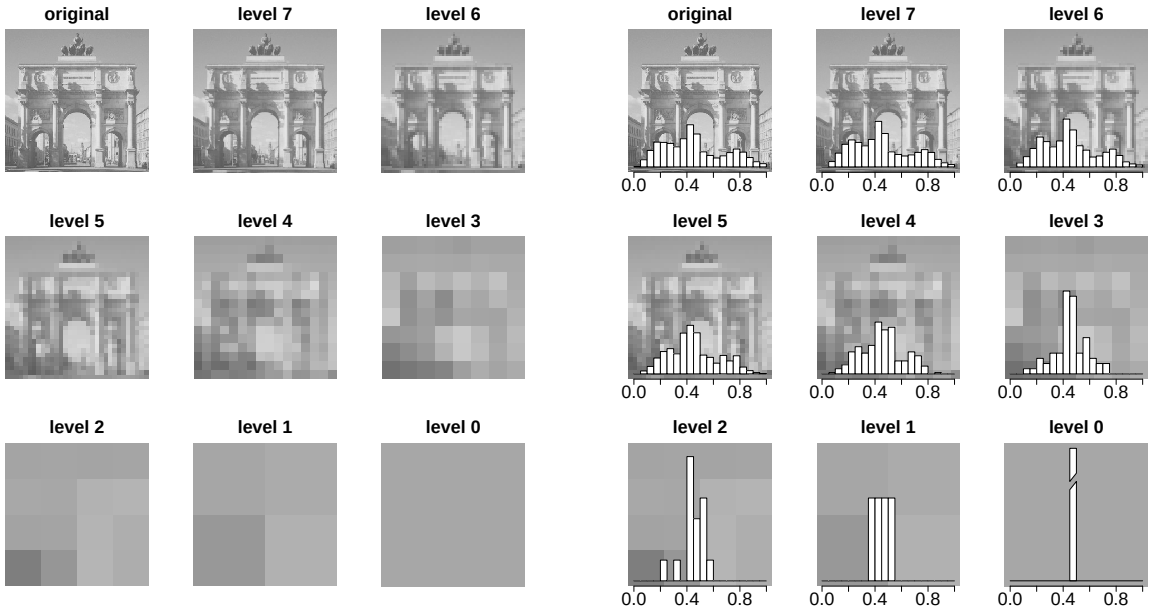


Figure 11: Purely low-pass filtered coefficient matrices of different levels (left) with overlaid normalized histograms (right).

stop at this specific level, just allowing to sort out features of reasonable frequency. In particular, variance reduces rapidly during the last four decomposition levels. Hence, a 'kink' in the curve of the squared sum of scaling coefficients could indicate an adequate primary decomposition level j_0.

Generalization to 3d

With the thorough introduction and examination of 1d and 2d wavelets as well as of the corresponding decomposition schemes, the extension to the 3d wavelet analysis is straightforward. In brief, the 3d scaling function $\phi(x, y, z)$ and seven wavelets in all cross-spatial horizontal, vertical and diagonal directions are designed as tensor product of 1d wavelets:

$$\begin{aligned}
\phi(x,y,z) &= \phi(x)\phi(y)\phi(z) \\
\psi^1(x,y,z) &= \phi(x)\phi(y)\psi(z) \\
\psi^2(x,y,z) &= \phi(x)\psi(y)\phi(z) \\
\psi^3(x,y,z) &= \psi(x)\phi(y)\phi(z) \\
\psi^4(x,y,z) &= \psi(x)\psi(y)\phi(z) \\
\psi^5(x,y,z) &= \psi(x)\phi(y)\psi(z) \\
\psi^6(x,y,z) &= \phi(x)\psi(y)\psi(z) \\
\psi^7(x,y,z) &= \psi(x)\psi(y)\psi(z).
\end{aligned}$$

All operations, such as dilation and translation, orthogonal subspace representation and the pyramidal filterbank algorithm, can be adapted to the above 3d initial wavelet functions. Finally, the wavelet analysis of square-summable function $f(x, y, z) \in l^2(\mathbb{R}^3)$ can be conducted.

1.4 Thresholding and shrinkage

Given the motivating preview on denoising (p. 12), this section serves to state more precisely what is meant by removing "small" coefficients. Let me derive noise reduction

procedures from the commonly used additive noise model. To ease readability, the location parameter is dropped for the moment. Suppose the observations $\mathbf{y}$, $\mathbf{y} \in \mathbb{R}^n$, relate to the true function f but are corrupted by Gaussian white noise, i. e. the observations are realizations of independent and identically distributed Gaussian random variables:

$$\mathbf{y} \quad = \quad f + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(0, \sigma^2 \mathbf{I}).$$

Due to the linearity of the wavelet transform, the additivity remains unchanged:

$$\begin{array}{ccccc} \underbrace{\mathbf{W}'\mathbf{y}} & = & \underbrace{\mathbf{W}'f} & + & \underbrace{\mathbf{W}'\boldsymbol{\varepsilon}} \\ \hat{\boldsymbol{\gamma}} & = & \boldsymbol{\gamma} & + & \boldsymbol{\eta}\,. \end{array}$$

Obviously, the wavelet coefficients of a noisy sample can be regarded as noise contaminated observations of hidden wavelet coefficients. Moreover, if $\mathbf{W}$ is orthogonal, the wavelet transform of the stationary and white noise will also be stationary and white:

$$\mathrm{Var}(\boldsymbol{\eta}) = \mathrm{Var}(\mathbf{W}'\boldsymbol{\varepsilon}) = \mathbf{W}'\mathrm{Var}(\boldsymbol{\varepsilon})\mathbf{W} = \sigma^2\mathbf{I} = \mathrm{Var}(\boldsymbol{\varepsilon}). \tag{7}$$

The wavelet transform creates a sparse signal representation in the sense that most empirical coefficients are zero or close to zero (compare Fig. 5). Only a few distinct coefficients reflect important signal singularities beyond the overall shape. In other words, the transform preserves the sum of squares (isometry of risks), but concentrated in a much smaller fraction of the components than in the raw data.

Thresholding rules

The most straightforward method for noise reduction suggests to let the informative coefficients survive while rigorously zeroing every "small" coefficient. This "keep-or-kill" strategy involves a well-chosen threshold λ that demarcates the cutoff between essential variables (or basis functions) and noise. The method is therefore also called *hard thresholding* and can formally be expressed as a function δ:

$$\hat{\gamma}_{\lambda,i} \quad = \quad \delta^H(\hat{\gamma}_i, \lambda) = \begin{cases} \hat{\gamma}_i, & |\hat{\gamma}_i| \geq \lambda \\ 0, & \text{otherwise.} \end{cases}$$

Figure 12 displays the corresponding function on the left hand side with $\lambda = 1$.

The second well-known denoising approach is *soft thresholding*:

$$\hat{\gamma}_{\lambda,i} \quad = \quad \delta^S(\hat{\gamma}_i, \lambda) = \text{sign}(\hat{\gamma}_i)\,(|\hat{\gamma}_i| - \lambda)_+ = \begin{cases} \hat{\gamma}_i + \lambda, & \hat{\gamma}_i < -\lambda \\ 0, & -\lambda \leq \hat{\gamma}_i < \lambda \\ \hat{\gamma}_i - \lambda, & \hat{\gamma}_i \geq \lambda\,. \end{cases}$$

The difference to hard thresholding is that, in addition to the truncation at λ, all other wavelet weights are shrunken by λ. This way soft thresholding becomes a continuous operation (right side of Fig. 12), offering considerable advantages. For instance, noise coefficients that have spuriously passed the hard threshold may appear as annoying 'blips'. Yet, soft thresholding shrinks these false structures and does therefore not suffer as much from high variability (Hastie et al., 2001). Soft thresholding can be termed a "high bias – low variance" solution to the shrinkage problem, whereas hard thresholding is a "low bias – high variance" solution.

The denoising mechanisms can be embedded into the general class of non-parametric regression models. Aiming at regularization, the objective to be minimized is formulated as trade-off between closeness to the data and sparsity/smoothness, governed by a tuning

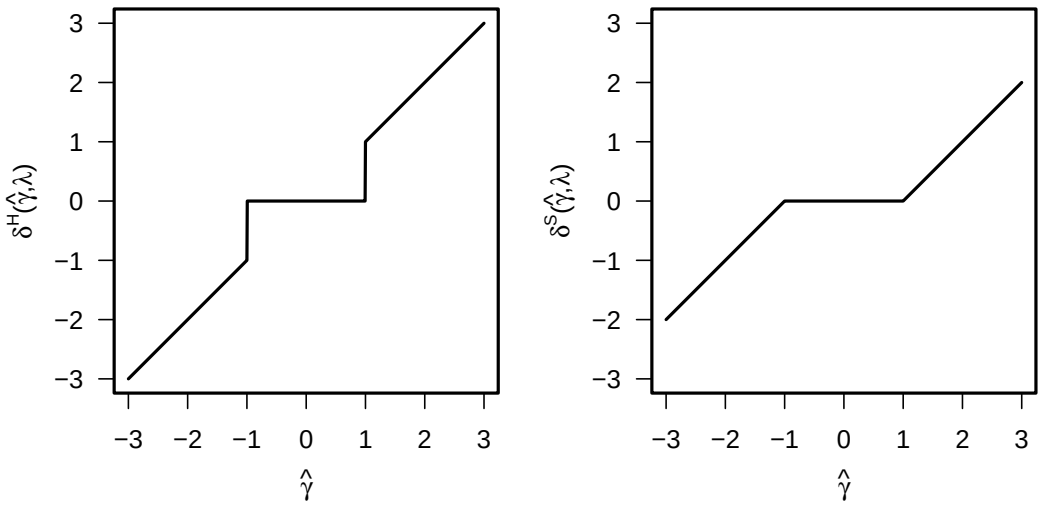


Figure 12: Classical hard (left) and soft (right) thresholding rules with $\lambda = 1$.

parameter:

$$\|\mathbf{y} - \mathbf{W}\boldsymbol{\gamma}_\lambda\|_2^2 + \lambda \mathrm{Pen}(\boldsymbol{\gamma}_\lambda) \quad \longrightarrow \quad \min_{\boldsymbol{\gamma}_\lambda} . \tag{8}$$

In the signal processing language, this balance between fidelity to the data and smoothness penalty is described as compromise between 'energy' and 'entropy'. Energy is an expression of signal size (compare p. 4) and related to variance. The entropy is a measure of the disorder of a system and is linked to the information content of that system. On the one hand, one could say that a high entropy state is information poor because there is so much disorder, and the disorder is essentially random. On the other hand, one could argue that a high entropy state is information rich because a truthful description of the exact state of randomness in all its detail would require lots of information. It is this latter sense which underlies Shannon's definition for discrete data: The information entropy is the log of the number of accessible states, and is dimensionless.Attention has to be paid to avoid possible confusion because this definition is a measure of the *lack* of order in a system. In Fig. 11, the entropy of the scaling image decreases with level until one single gray value remains. This monochromic state contains the least information according to Shannon's entropy. In contrast, a uniformly distributed histogram would have maximum entropy and maximum information.

To proceed with the topic of smoothing approaches, recapitulate that $\mathbf{y} = \mathbf{W}\hat{\boldsymbol{\gamma}}$ by perfect reconstruction. Then the orthogonality of the wavelet transform allows to replace the data term by the squared differences between untouched coefficients and thresholded counterparts:

$$\|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_\lambda\|_2^2 + \lambda \mathrm{Pen}(\boldsymbol{\gamma}_\lambda) \quad \longrightarrow \quad \min_{\boldsymbol{\gamma}_\lambda} .$$

The penalty term can take on different forms. For example, using the L_2 norm

$$\mathrm{Pen}(\boldsymbol{\gamma}_\lambda) \quad = \quad \|\boldsymbol{\gamma}_\lambda\|_2^2 = \sum_i \gamma_{\lambda,i}^2,$$

yields a scaled version of the ordinary least squares coefficients:

$$\hat{\boldsymbol{\gamma}}_\lambda \quad = \quad \frac{1}{1+\lambda}\,\hat{\boldsymbol{\gamma}}.$$

Unless the intercept, i. e. the scaling wavelet coefficients α_{0k}, is left out from penalization, this corresponds to *linear shrinkage* of the observed data:

$$\hat{\mathbf{y}} \;=\; \frac{1}{1+\lambda}\,\mathbf{y}.$$

The procedure is also known as ridge regression (Hastie et al., 2001, Chap. 3). However, the L_2 norm measures energy rather than sparsity or smoothness: selection does not take place. Hence consider the use of the L_1 norm

$$\mathrm{Pen}(\boldsymbol{\gamma}_\lambda) \;=\; \|\boldsymbol{\gamma}_\lambda\|_1^2 = \sum_i |\gamma_{\lambda,i}|,$$

which leads to the same framework as in Eq. (8), also familiar under the name of lasso regression (Hastie et al., 2001, Chap. 3). Because $\mathbf{W}$ is orthogonal, the lasso criterion results in the simple solution of soft thresholding. The method is highly esteemed for its adaptivity founded on least absolute shrinkage *and* selection from the parameter space. A wavelet fit constrained to the L_1 norm represents a *non-linear* smoothing procedure. Likewise, hard thresholding turns out to be a non-linear operation. In this case, the objective consists of the usual residual sum of squares and a penalty term that amounts to the number of non-zero coefficients:

$$\mathrm{Pen}(\boldsymbol{\gamma}_\lambda) \;=\; \sum_i 1_{[\gamma_{\lambda,i}\neq 0]},$$

where $1_{[\text{condition}]}$ is an indicator function.

Hard thresholding is said to perform better in spike detection, whereas soft thresholding is preferred if smoothness is paramount. For a discussion on risk, bias and variance of these two approaches refer to Jansen (2001, Chap. 5.5). More sophisticated rules are available such as n-degree garrote or Bayesian thresholding rules (Gençay et al., 2002; Jansen, 2001; Johnstone and Silverman, 2005). These schemes model the transition between rejected and shrunken coefficients in a smoother fashion but are computationally more intensive. Note that also ridge and lasso regression can be viewed as Bayesian estimates with different priors, either Gaussian or Laplace (e. g. Hastie et al., 2001). In both models, posterior modes are derived as Bayesian estimates instead of the more commonly used posterior

means. The equality of posterior mode and mean only holds in case of ridge regression. The lasso regularization constraint was promoted by Donoho and Johnstone (1995) as *SURE shrinkage* (Stein's Unbiased Risk Estimation). Donoho and Johnstone investigated optimality properties and contributed seminal work on the choice of the threshold value.

Thresholding values

Regardless of the selection of an available thresholding rule, the cutoff value λ needs to be chosen with care. If the noise proportion was known, a complete noise removal could be accomplished by hard thresholding with $\lambda = \sigma$ due to Eq. (7). A threshold less than the true error variance would cause undesired noise relicts while a too loose threshold leads to information loss. Yet, σ is unknown. A bunch of suitable procedures to determine λ is published, two of which will be presented here.

On the background of the additive Gaussian white noise model, a straightforward rationale is to remove all wavelet coefficients that are smaller than the expected maximum of an assumed i. i. d. normal noise sequence of given size. This approach yields the *universal threshold*:

$$\lambda^{UNIV} \quad = \quad \sigma\sqrt{2\log(n)}. \tag{9}$$

Its name reflects the global validity for all sufficiently smooth signals of length n (Jansen, 2001). It can even be stated that the universal threshold is adaptive to unknown smoothness. The interested reader may consult e. g. Härdle et al. (1998) for more detailed information on smoothness within the framework of Besov spaces.

The universal threshold operates as an oracle that reveals which coefficients are above σ but not their exact values. Donoho and Johnstone (1994) showed that, within a logarithmic factor, universal soft-thresholding performs as well as optimal wavelet coefficient selection which, in turn, is essentially not less powerful than any piecewise polynomial and spline method. Moreover, λ^{UNIV} possesses minimax property, i. e. the minimax threshold asymptotically coincides with the universal threshold. However, instead of balancing energy and entropy, smoothness comes prior to goodness of fit. This renders the results visually appealing, leading to the procedure's synonym *VisuShrink*.

The noise level in definition (9) has to be estimated from the data. Donoho and Johnstone (1994) use the median absolute deviation of the wavelet coefficients at the finest level $J = \log_2(n)$, divided by 0.6745. As motivated on p. 12, this level mirrors high frequency features and thus essentially noise. The use of a robust variance measure is intended to correct the bias due to some isolated signal occuring at that level. The various denoising techniques apply also to images and image volumes. Yet in the higher dimensional case, each level has multiple decomposition components, also called orientations (2d: three, 3d: seven). The question arises whether noise estimation should be based on purely high-pass filtered coefficients or, whether high- and low-pass filtered output at level J should be considered equally. Furthermore, n has to be adjusted accordingly.

If minimum risk optimality is paramount, soft-thresholding with the cutoff chosen according to Stein's Unbiased Risk Estimation (SURE) seems more suitable. Donoho and Johnstone (1995) call this procedure *SureShrink*. In contrast to the level invariant universal threshold, the SURE thresholds are independently derived from each resolution level $j = j_0, \ldots, J = \log_2(n)$ by minimizing the level specific risk $\mathrm{E}\|\hat{\boldsymbol{\gamma}}_{\lambda,j} - \boldsymbol{\gamma}_j\|^2$. Depending on j, the subvector $\boldsymbol{\gamma}_j = (\gamma_{j0}, \ldots \gamma_{jk}, \ldots, \gamma_{j(2^j-1)})'$ of $\boldsymbol{\gamma}$ comprises either the coefficients of the scaling functions or the ones of levelwise wavelets, here uniformly denoted by double-indexed γ_{jk}. In general, most of the wavelet literature formulates SURE shrinkage as a result in multivariate normal decision theory. Yet, I prefer to stay in the wavelet context though running the risk of complex notation.

After variance standardization in Eq. (7), the empirical OLS wavelet coefficients are multivariate normally distributed: $\boldsymbol{\gamma}_j^{OLS} \sim N(\boldsymbol{\gamma}_j, \mathbf{I})$. Stein proved that any (almost arbitrary) estimator for $\boldsymbol{\gamma}_j$ of the form $\hat{\boldsymbol{\gamma}}_{\lambda,j}(\boldsymbol{\gamma}_j^{OLS}) = \boldsymbol{\gamma}_j^{OLS} + g(\boldsymbol{\gamma}_j^{OLS})$ with $g : \mathbb{R}^{2^j} \longrightarrow \mathbb{R}^{2^j}$ weakly differentiable[6] leads to an unbiased estimate of the L_2 loss:

$$\mathrm{E}_{\boldsymbol{\gamma}_j}\|\hat{\boldsymbol{\gamma}}_{\lambda,j}\left(\boldsymbol{\gamma}_j^{OLS}\right) - \boldsymbol{\gamma}_j\|^2 \;\; = \;\; 2^j + \mathrm{E}_{\boldsymbol{\gamma}_j}\left(\|g(\boldsymbol{\gamma}_j^{OLS})\|^2 + 2 \bigtriangledown g(\boldsymbol{\gamma}_j^{OLS})\right),$$

where $\bigtriangledown g(\boldsymbol{\gamma}_j^{OLS}) = \sum_{k=0}^{2^j-1} \frac{\partial}{\partial \gamma_{jk}^{OLS}} g_k(\boldsymbol{\gamma}_j^{OLS})$. Employing soft thresholding allows the follow-

[6] Let g and h be functions defined on the real-line and integrable on every bounded interval. Then g is weakly differentiable with weak derivative h if $\int g(t)u'(t)dt = -\int h(t)u(t)dt$ is satisfied for all infinitely many times differentiable test functions u with compact support (Härdle et al., 1998). Generalization to the higher-dimensional case is straightforward.

ing specification:

$$g_k(\boldsymbol{\gamma}_j^{OLS}) = \delta_\lambda^S(\gamma_{jk}^{OLS}) - \gamma_{jk}^{OLS} = \begin{cases} \lambda, & \gamma_{jk}^{OLS} < -\lambda \\ -\gamma_{jk}^{OLS}, & -\lambda \leq \gamma_{jk}^{OLS} < \lambda \\ -\lambda, & \gamma_{jk}^{OLS} \geq \lambda\,, \end{cases}$$

so that $\|g(\boldsymbol{\gamma}_j^{OLS})\|^2 = \|\min(|\boldsymbol{\gamma}_j^{OLS}|, \lambda)\|^2$. Similarly, $\bigtriangledown g(\boldsymbol{\gamma}_j^{OLS}) = \sum_{k=0}^{2^j-1}(-1)1_{[-\lambda \leq \gamma_{jk}^{OLS} \leq \lambda]} = -\#\{k : |\gamma_{jk}^{OLS}| \leq \lambda\}$. To sum up, Stein's estimate of risk can be rewritten as

$$SURE(\lambda, \boldsymbol{\gamma}_j^{OLS}) \;=\; 2^j - 2\#\{k : |\gamma_{jk}^{OLS}| \leq \lambda\} + \|\min(|\boldsymbol{\gamma}_j^{OLS}|, \lambda)\|^2.$$

Recall that the expectation $\mathrm{E}_{\boldsymbol{\gamma}_j} SURE(\lambda, \boldsymbol{\gamma}_j^{OLS})$ is unbiased. The all-dominant and crucial threshold is given by

$$\lambda_j^{SURE} \;=\; \arg\min_{\lambda \geq 0} SURE(\lambda, \boldsymbol{\gamma}_j^{OLS}).$$

In fact, the SURE threshold takes on the value of one of the coefficients $|\gamma_{jk}^{OLS}|, k = 0, \ldots, 2^j - 1$. Moreover, one can show that SURE shrinkage is the same as lasso regression (Hastie et al., 2001).

Risk minimization in the sense of the L_2 norm can be regarded as seeking after the equilibrium between variance and (squared) bias. Without relying on a priori knowledge about the underlying smoothness, some noise artifacts will eventually sneak in at every scale. This, however, makes the SURE shrinkage automatically smoothness-adaptive (for theoretical results see Donoho and Johnstone, 1995). One could say, that the outcome of universal thresholding is visually attractive while results from SURE shrinkage are qualitatively convincing: The reconstructions jump where the true object jumps; at the same time, the fits are smooth where the true object is smooth.

In this category of minimum risk thresholds falls also the minimax estimator itself (named *RiskShrink* by Donoho and Johnstone (1994)) and the GCV strategy (Jansen, 2001). In general, SURE and related procedures are more conservative and are more convenient when small details of f lie near the noise range. A hybrid approach was conceived in order to switch from SURE to universal thresholding whenever the level specific signal-to-noise ratio is very small and the SURE result would be too noisy. Multiple alternative

threshold criteria are available such as the false discovery rate (FDR) for the multiple testing problem whether the n wavelet coefficients are significantly zero (Ogden, 1997, Chap. 8.2). The FDR based threshold yields similarly smooth results as the universal threshold (Jansen, 2001; Gençay et al., 2002).

1.5 Resolution increment

If wavelets act as smoothing tool, interpolation does not enter the equation. If however wavelets are to be supposed to act as proper basis functions within the SVCM framework, interpolation matters. Therefore, let me briefly comment on this issue. Since, in practice, wavelet analysis uses the repeated application of filter sequences (see Section 1.2, pp. 14), there exists no such possibility to enhance resolution as for B-splines. This means, we cannot simply evaluate the basis functions at the desired points and sum over their estimated weight. Yet, different alternatives are conceivable.

The naive approach would be to paste multiple zeros into the level's j_0 filtering output before reconstruction. This implies a modification of the upsampling step. A more sophisticated procedure involves lifting schemes that gradually increase the complexity of the overall transform (Jansen, 2001). This is achieved by repeated dual (prediction) and primal (update) lifting steps. Hence, lifting is not only an alternative to classical filter banks but also offers the wavelet transform to work on non-dyadic and even irregular grids. In addition, lifting allows the reconstruction to be limited to integers as e. g. desired in digital image processing.

Donoho (1992) discusses several interpolating scaling functions, for instance of the Deslauriers-Dubuc wavelet basis or the spline wavelet basis (see also recent work by the Biomedical imaging group, Lausanne[7]). It holds that the autocorrelation function of an orthogonal scaling function is an interpolating scaling function. Even more, the autocorrelation of a Daubechies scaling function is a Deslauriers-Dubuc scaling function. Image interpolation is then achieved within the multiresolution analysis (MRA) framework.

Other authors (e. g. Su and Ward, 2006) use bi-orthogonal wavelets and rely on the smooth reconstruction filter for improved image quality. The given image is considered

[7] http://bigwww.epfl.ch/publications/

to be the low-resolution image, from which the detail components can be predicted in a suitable way. Adding the horizontal, vertical and diagonal edge information to the low-resolution image yields a high-resolution image of twice the size as the original.
All these interpolation approaches allow image enlargement up to a fixed size. Yet in view of the application of interest, a fiber tracking algorithm needs to be able to calculate the desired tensor information at any arbitrary point. These considerations have to be kept in mind for future work. In this thesis, I will not explore wavelet interpolation schemes but rather use, on top of voxelwise tensor estimation and wavelet based smoothing, quadratic interpolation as implemented in the available fiber tracking tool (Gössl et al., 2002).

2 Denoising the diffusion tensor

The discrete wavelet decomposition was reported to successfully denoise diffusion weighted images in combination with a Wiener filter (Wirestam et al., 2006). The approach follows the minority of smoothing methods for DTI data which operate right on the complex signal intensities before the Fourier transformation turns the Gaussian noise into a Rician distribution, in particular at locally low SNR (Gudbjartsson and Patz, 1995). Attempts have been made to smooth the magnitude images, too, with the focus on edge-preservation and statistical robustness (Hahn et al., 2001). In this work, wavelet filtering is performed on processed data, namely on the array of diffusion tensor estimates. For this purpose, recall the estimation problem inherent to DTI (Heim et al., 2007, Chap. 2). Formally, it is a matter of repeated measurement design

$$\mathbf{y}(s) \;=\; \mathbf{X}\boldsymbol{\beta}(s) + \boldsymbol{\varepsilon}(s), \quad \boldsymbol{\varepsilon}(s) \sim N(\mathbf{0}, \sigma^2\mathbf{I}) \tag{10}$$

with $\boldsymbol{\beta}(s) = (\beta_1, \ldots, \beta_p)'(s)$ encoding for the $p = 6$ unknown elements that determine the symmetric covariance matrix at position $s, s = 1, \ldots, n$, in 3d space, i. e. the local diffusion tensor within the brain. Multiplication with the Moore-Penrose matrix inverse leads to the interpretation that the observed covariance entries $\mathbf{y}^*(s)$ are equal to the true

covariance entries $\boldsymbol{\beta}(s)$ corrupted by noise:

$$\underbrace{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}(s)}_{\mathbf{y}^*(s)} = \boldsymbol{\beta}(s) + \boldsymbol{\varepsilon}^*(s), \quad \boldsymbol{\varepsilon}^*(s) \sim N(\mathbf{0}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}). \tag{11}$$

More precisely, the pretended observations $\mathbf{y}^*(s)$ are nothing else than the standard voxel-by-voxel estimates of the tensor elements, i. e. ST1 results. The aim is to recover for each tensor element the underlying ideal function $\beta_j(\boldsymbol{s}), \boldsymbol{s} \in \mathbb{R}^3, j = 1, \ldots, 6$, using DWT as implemented in the fast and memory-saving algorithm of Mallat.

Incorporation of the positive definiteness constraint

Log-Cholesky parametrization (Pinheiro and Bates, 1996) is one possibility to meet the positive-definiteness constraint of the diffusion tensors. Note that the restriction had not yet been imposed in the B-spline based SVCM, nor in ST1. Of course, immediate consideration through the estimation process would have been more appropriate. Cholesky factorization states that if and only if $\mathbf{R}$ is upper triangular with non-zero diagonal elements, then $\mathbf{R}'\mathbf{R}$ will be positive definite. Thus as long as the diagonal entries are exclusively negative or exclusively positive, the off-diagonal elements of the Cholesky factor $\mathbf{R}$ are unique. Either way, the intended uniqueness of $\mathbf{R}$ represents once more a constrained estimation problem. Yet if the diagonal shall comply with positive values, the logarithm can serve as auxillary function to alter the parameter space unlimited on the real line. Hence, ensure positive definiteness of the diffusion tensor $\mathbf{D}$,

$$\mathbf{D} = \begin{pmatrix} \beta_1 & \beta_2 & \beta_4 \\ \beta_2 & \beta_3 & \beta_5 \\ \beta_4 & \beta_5 & \beta_6 \end{pmatrix} = \mathbf{R}'\mathbf{R} = \begin{pmatrix} r_1 & r_2 & r_4 \\ 0 & r_3 & r_5 \\ 0 & 0 & r_6 \end{pmatrix}' \begin{pmatrix} r_1 & r_2 & r_4 \\ 0 & r_3 & r_5 \\ 0 & 0 & r_6 \end{pmatrix}$$

by unique and simultaneously unconstrained log-Cholesky factorization:

$$\boldsymbol{\rho} = (\log(r_1), r_2, \log(r_3), r_4, r_5, \log(r_6))'.$$

Finally, reparameterize $\boldsymbol{\beta}$ as vector-valued function in $\boldsymbol{\rho}$ by definition of

$$\begin{aligned}\boldsymbol{\beta}(\boldsymbol{\rho}) &= \big(\exp(\rho_1)^2, \exp(\rho_1)\rho_2, \rho_2^2 + \exp(\rho_3)^2, \\ &\qquad \exp(\rho_1)\rho_4, \rho_2\rho_4 + \exp(\rho_3)\rho_5, \rho_4^2 + \rho_5^2 + \exp(\rho_6)^2\big)'.\end{aligned}$$

In summary, the initial estimation problem (11) with $\boldsymbol{\beta}(s)$ being constrained to build a positive deifinite diffusion tensor can be rendered into an unconstrained estimation problem using log-Cholesky parametrization:

$$\boldsymbol{\rho}_{obs}(s) = \boldsymbol{\rho}_{true}(s) + \boldsymbol{\varepsilon}^{\sharp}(s), \quad \boldsymbol{\varepsilon}^{\sharp}(s) \sim N(\mathbf{0}, \mathrm{diag}(\sigma^2_{\sharp,j})). \tag{12}$$

The error term is an approximation in that independence is postulated for the transformed covariance elements $\rho_j, j = 1, \ldots, p$. With the log-Cholesky parametrization no further care is then needed in keeping track of the error term: The 3d wavelet decomposition can promptly be applied to each 3d field $\{\boldsymbol{\rho}_{obs,j}(s) : s = 1, \ldots, n\}$, where $\sigma^2_{\sharp,j}$ is estimated from the high-frequency weights of the first decomposition step. Hence, project elements $\boldsymbol{\rho}_{obs,j}, j = 1, \ldots, p$, onto 3d tensor product wavelets:

$$\boldsymbol{\rho}_{obs,j} = \mathbf{W}\boldsymbol{\gamma}_j + \boldsymbol{\varepsilon}^{\sharp}_j.$$

The ordinary least squares estimate of the wavelet coefficient surfaces holds:

$$\boldsymbol{\gamma}_j^{OLS} = \underbrace{(\mathbf{W}'\mathbf{W})^{-1}}_{=\mathbf{I}} \mathbf{W}'\boldsymbol{\rho}_{obs,j}.$$

As explained in Section 1.4, the choice of a thresholding value and rule yields a sparse (and possibly shrunken) vector of wavelet contributions:

$$\hat{\boldsymbol{\gamma}}_{\lambda,j} = \delta_\lambda(\boldsymbol{\gamma}_j^{OLS}).$$

After having run the inverse discrete 3d wavelet transform, the regularized estimate

$$\hat{\boldsymbol{\rho}}_j = \mathbf{W}\hat{\boldsymbol{\gamma}}_{\lambda,j},$$

can be reparameterized to:

$$\hat{\boldsymbol{\beta}}_j \;=\; \boldsymbol{\beta}(\hat{\boldsymbol{\rho}}_j) = \begin{pmatrix} \exp(\hat{\rho}_1)^2 \\ \exp(\hat{\rho}_1)\hat{\rho}_2 \\ \hat{\rho}_2^2 + \exp(\hat{\rho}_3)^2 \\ \exp(\hat{\rho}_1)\hat{\rho}_4 \\ \hat{\rho}_2\hat{\rho}_4 + \exp(\hat{\rho}_3)\hat{\rho}_5 \\ \hat{\rho}_4^2 + \hat{\rho}_5^2 + \exp(\hat{\rho}_6)^2 \end{pmatrix}.$$

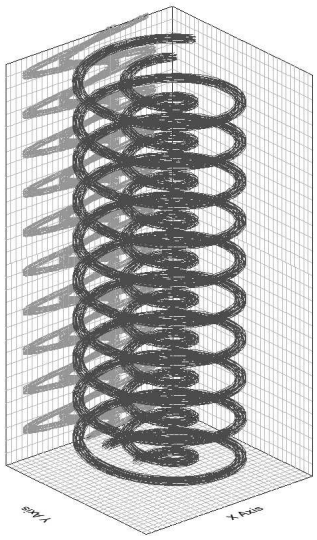
Note that the competitive log matrix transformation (Pinheiro and Bates, 1996) is more stable but does not lead to the desired independence of the parametrized elements either. Certainly, the assumption of independently distributed errors in Eq. (12) is strong but indispensable for looping over the six 3d fields of parametrized tensor elements. In contrary, spatial correlation could be considered by special wavelet transforms for colored noise. Also level dependent thresholding is said to be more suitable in such a case.
In conclusion, the bunch of processing steps described above yields an adaptive estimator for smooth 3d components of positive definite covariance matrices.

3 Simulation study

3.1 Simulation model

The simulation model as presented in Fig. 2 (Heim et al., 2007) is not apt for wavelet analysis. Of course, a mere resampling would satisfy the demand of dyadic dimensions. However, the miminum logarithm to the basis 2 of the 3d array size determines the maximum possible decomposition depth. For this reason, the spiral function is periodically extended along the z-axis prior to discrete sampling into $32 \times 32 \times 32$ voxels. As displayed in Fig. 13, the resulting imaging volume contains two concentrically nested spirals of five and twelve centimeters in radius, mimicking fiber bundles. Since each fiber-transit voxel is considered a fiber voxel, each transversal slice contains enough fiber structure (compare e. g. the projection on the xy-plane). The overall percentage amounts to 13.87% voxels. The degree of anisotropy in fiber as well as non-fiber voxels has been kept the same as

in Heim et al. (2007). Given the simulation model, $N = 100$ data sets were generated consisting of six diffusion weighted images corrupted by noise ($\sigma = 10$).



- 3d data grid of $\{32 \times 32 \times 32\} \subset \mathbb{R}^3$ voxels
- $2 \times 2 \times 4$ mm^3 voxel size
- six 3d varying coefficient surfaces
- spiral tensors are anisotropic;
  background tensors are isotropic
- simulated Gaussian error with $\sigma = 10$

Figure 13: Design of the simulation study and geometry of the underlying fiber bundles.

3.2 Study design

The parameter setup of the discrete wavelet analysis involves the wavelet family, the primary thresholding level j_0, the thresholding rule and the threshold itself. Boundary condition is not up for discussion as the employed R-package `waveslim` (Whitcher, 2005) handles border distortions only in a periodic fashion.

Since the basis function should mimic the underlying features of the given object, the choice of the wavelet family is crucial. Typical criteria are symmetry useful in avoiding dephasing in image processing, the number of vanishing moments important for compression purposes, and the regularity to get nice smoothness. Moreover, the length of the mother wavelet should not exceed the number of available observations. Apart from the Haar family, no wavelet system satisfies the demands of compact support (among others preferable for numerical stability) and symmetry at the same time. The latter condition is however indispensable in imaging if spatial directions are equally likely. Therefore, Daubechies *least asymmetric* wavelets with four vanishing moments (LA4) are investigated in addition to Haar (Haar) and Daubechies extremal phase wavelets (D2) with one and two vanishing moments, respectively. Also minimum bandwidth wavelets with four vanishing

moments (MB4) are considered. The MB family is closely related to Daubechies extremal phase family, i. e. the members coincide for two vanishing moments and are almost the reverse of each other for more than two vanishing moments ($\phi_{Daub}(x) = -\phi_{MB}(x)$). The important difference consists in superior frequency-domain properties and, thus, superior edge-preserving property of MB wavelets compared to Daubechies wavelets given the same number of vanishing moments (Gençay et al., 2002, Chap. 4.3.3).

The effect of the primary decomposition level was found to be neglibible according to a pilot study. Complete decomposition ($j_0 = 0$) is therefore performed.

Since hard thresholding can, in principle, be combined with the univeral or SURE threshold and since this seems more natural to a non-statistician (Donoho and Johnstone, 1994), it is taken into account. Thus, the comparison comprises four shrinkage methods: hard universal (hUNIV), soft universal (sUNIV), hard SURE (hSURE), and soft SURE (sSURE) thresholding. Every variant includes a noise estimate which, theoretically, can be level dependent and/or orientation dependent.

A preparatory study aims to investigate the impact of the noise estimation under the limitation to Haar wavelets. The estimation is either based on the highest resolution level (sigma1) or depends on each orientation (sigma2), each level (sigma3) or each level and each orientation (sigma4). As in Heim et al. (2007), the quality is assessed via the error measures averaged mean squared error (AMSE) of the tensor fit, depending on the voxel type 'spiral':

$$\mathrm{AMSE}_{\mathrm{sp}}^{(i)} = \frac{1}{p\, n_{\mathrm{sp}}} \sum_{s \in \mathrm{spiral}} \left\| \boldsymbol{\beta}(s) - \hat{\boldsymbol{\beta}}^{(i)}(s) \right\|^2 , \quad i = 1, \ldots, N ,$$

or 'background':

$$\mathrm{AMSE}_{\mathrm{bg}}^{(i)} = \frac{1}{p\, n_{\mathrm{bg}}} \sum_{s \in \mathrm{background}} \left\| \boldsymbol{\beta}(s) - \hat{\boldsymbol{\beta}}^{(i)}(s) \right\|^2 , \quad i = 1, \ldots, N ,$$

as well as by the voxelwise mean squared error (VMSE) defined as

$$\mathrm{VMSE}_{j}^{(s)} = \frac{1}{N} \sum_{i=1}^{N} \left(\boldsymbol{\beta}_j(s) - \hat{\boldsymbol{\beta}}_j^{(i)}(s) \right)^2 , \quad s = 1, \ldots, n .$$

A second study concentrates on the impact of the wavelet type. For this purpose, the DWT of the simulated data is conducted using wavelets from the Haar (Haar) , the Daubechies extremal phase (D2), the least asymmetric (LA4), and the minimum bandwidth (MB4) family. The DWTs are then thresholded according to the above rules hUNIV, sUNIV, hSURE, and sSURE in combination with the two different noise estimates sigma1 and sigma2.

3.3 Simulation results

Concerning the preparatory study, Figures 14 and 15 display the results by means of boxplots. Clearly recognizable is the familiar distribution of smaller errors in background than in spiral voxels. The discrepancy between these two tissue classes becomes more pronounced with soft thresholding (second and fourth row from top). Concerning the noise estimation, level dependent approaches fail as expected from the simulation design that does not incorporate correlated noise. This is also true for the SURE approach, although whatsoever less evident when combined with the hard rule (third row from top). In contrast, the exclusive specification with respect to orientation (horizontal, vertical, diagonal details) shows hardly an effect compared to the uniform approach sigma1. A slight improvement in terms of smaller log AMSE errors is visible for the SURE threshold applied either in a hard or soft way (Fig. 14, bottom rows).

Boxplots resulting from the second study design are not suited for visual inference because of apparent similarity. Tables 1 and 2 contain instead the principal boxplot statistics corresponding to the logarithmized average and voxelwise MSE, respectively.
At first glance, the wavelet type plays hardly a role, in particular with respect to the errors of D2, LA4 and MB4. The resembling behavior points at an oversized filter length given the scale of the features to be preserved. In view of this, especially the minimum bandwidth wavelet is not able to develop its superior edge-preserving property. The Haar wavelet has the shortest support among the examined families and can therefore still capture substantial structures at finer scales. Not surprisingly, the Haar wavelet performs best under almost all shrinkage conditions in terms of an overall error measurement (log AMSE). Regarding the local error (log VMSE), the D2 family outperforms the Haar

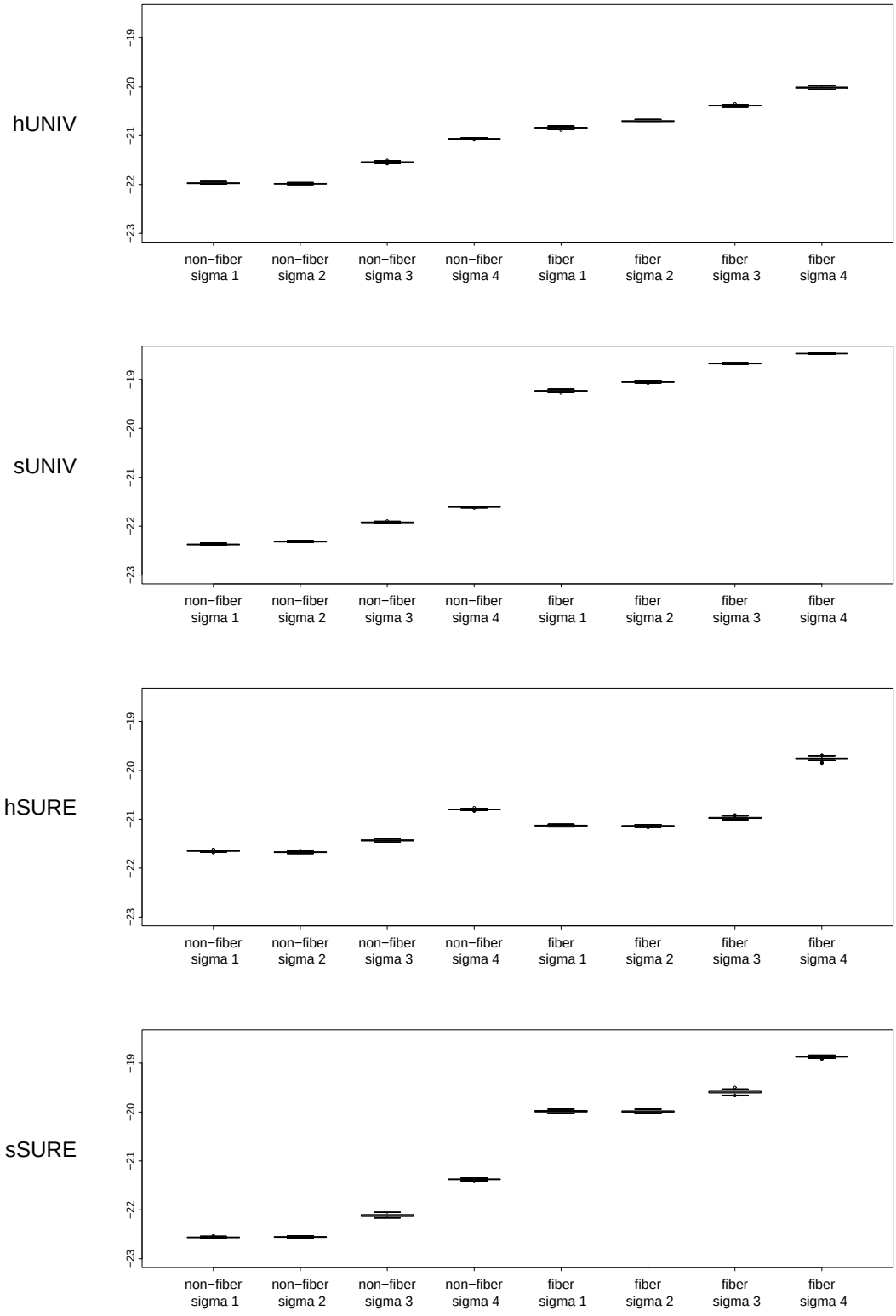


Figure 14: Boxplots of log AMSE for hUNIV, sUNIV, hSURE, and sSURE (top to bottom) with the noise estimated on the basis of the highest resolution level (sigma1), orientation dependent (sigma2), level dependent (sigma3), orientation and level dependent (sigma4) using Haar wavelets.

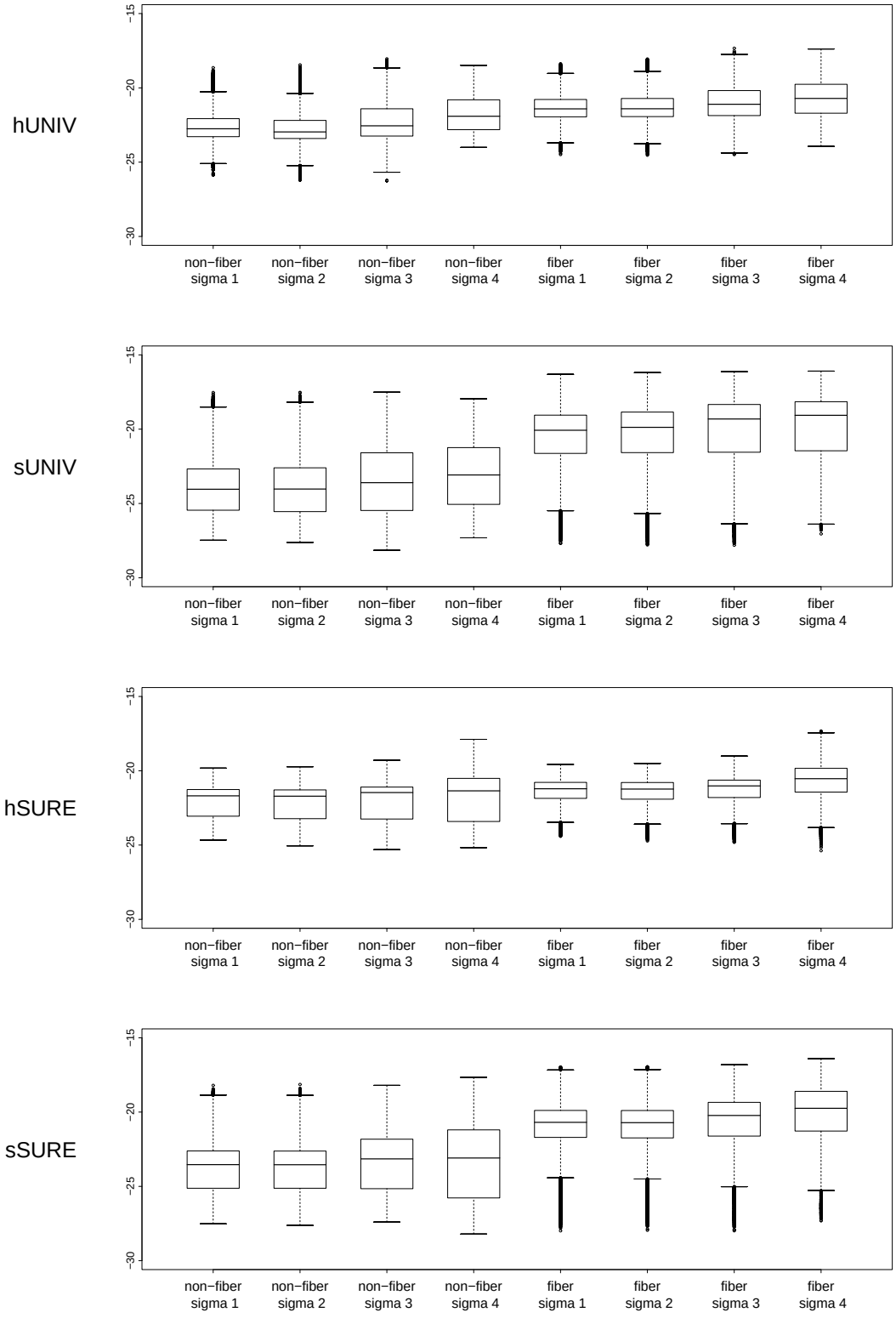


Figure 15: Boxplots of log VMSE for hUNIV, sUNIV, hSURE, and sSURE (top to bottom) with the noise estimated on the basis of the highest resolution level (sigma1), orientation dependent (sigma2), level dependent (sigma3), orientation and level dependent (sigma4) using Haar wavelets.

family in background voxels but not in spiral voxels.
The performance on these two tissue types is quite similar if hard thresholding is applied. In contrast, soft thresholding leads to distinctively larger errors in fiber associated regions. At the same time, the shrinkage entails a smooth and less erroneously reconstructed background. This holds for all parameter settings and both error measures.
Log AMSE errors obtained with the SURE threshold tend to be smaller than those of universal thresholding. If the focus is on log VMSE errors, universal hard thresholding seems to be preferrable over the SURE counterpart.
The choice of the noise estimate does, in general, not affect the MSE ranking of the wavelet types (compare black and gray prints). This result corroborates the previous finding obtained from analyses with the Haar wavelet.

3.4 Comparison to standard approaches

Coming back to the goal of improving regularized tensor estimation by modified DTI data processing, the comparison to previous methods is essential. Figure 16 therefore depicts the error boxplots of voxelwise regression without (ST1) and with subsequent Gaussian smoothing (ST2), sequential SVCM with linear B-splines, one knot per 1.28 voxels and global tuning parameter for first order difference penalties (DXglob-lin), as well as ST1 filtered once based on Haar wavelets combined with universal thresholding (hUNIV-Haar) and once based on Daubechies wavelets combined with SURE shrinkage (sSURE-D2). Noise estimation conforms to sigma1.
Clearly, both wavelet transforms convince with satisfactory error reduction. The improvement appears larger from an overall (top row) than from a local perspective (bottom row). In particular, errors in fiber voxels can be lowered using SURE shrinkage, outperforming ST2 and the B-splines based SVCM, but not ST1. The voxelwise least squares estimates show approximately the same error distribution regardless the underlying degree of anisotropy. This can be considered as proof of concept, since globally superimposed noise prolongates to the same amount into local estimates.

The VMSE errors, collapsed into distributional characteristics in Fig. 16, can alternatively be presented as maps to enable local assignment. For this purpose, log ratios of

		non-fiber				fiber			
		Haar	D2	LA4	MB4	Haar	D2	LA4	MB4
hUNIV	maximum	-21.99	-21.69	-21.65	-21.55	-20.89	-20.60	-20.61	-20.59
		-22.00	-21.62	-21.59	-21.49	-20.74	-20.37	-20.38	-20.34
	upper hinge	-21.98	-21.67	-21.64	-21.54	-20.85	-20.58	-20.59	-20.56
		-21.99	-21.62	-21.59	-21.47	-20.71	-20.35	-20.36	-20.31
	median	-21.97	-21.67	-21.64	-21.53	-20.84	-20.56	-20.58	-20.55
		-21.98	-21.61	-21.58	-21.47	-20.71	-20.34	-20.35	-20.30
	lower hinge	-21.96	-21.66	-21.63	-21.53	-20.83	-20.55	-20.57	-20.53
		-21.98	-21.61	-21.58	-21.46	-20.70	-20.33	-20.34	-20.29
	minimum	-21.94	-21.65	-21.62	-21.51	-20.80	-20.52	-20.54	-20.49
		-21.96	-21.59	-21.57	-21.45	-20.67	-20.31	-20.32	-20.27
		non-fiber				fiber			
		Haar	D2	LA4	MB4	Haar	D2	LA4	MB4
sUNIV	maximum	-22.39	-22.07	-22.00	-21.88	-19.28	-19.14	-19.20	-19.19
		-22.33	-21.96	-21.90	-21.80	-19.08	-18.86	-18.90	-18.87
	upper hinge	-22.38	-22.05	-21.99	-21.87	-19.24	-19.11	-19.17	-19.16
		-22.32	-21.95	-21.88	-21.79	-19.06	-18.84	-18.88	-18.86
	median	-22.37	-22.04	-21.98	-21.86	-19.23	-19.10	-19.16	-19.15
		-22.32	-21.94	-21.88	-21.78	-19.05	-18.84	-18.88	-18.85
	lower hinge	-22.36	-22.03	-21.97	-21.86	-19.22	-19.09	-19.15	-19.14
		-22.31	-21.94	-21.87	-21.78	-19.05	-18.83	-18.87	-18.85
	minimum	-22.35	-22.02	-21.96	-21.85	-19.20	-19.07	-19.12	-19.12
		-22.29	-21.93	-21.86	-21.76	-19.03	-18.82	-18.86	-18.83
		non-fiber				fiber			
		Haar	D2	LA4	MB4	Haar	D2	LA4	MB4
hSURE	maximum	-21.68	-21.59	-21.58	-21.53	-21.16	-21.02	-21.03	-21.02
		-21.70	-21.62	-21.61	-21.56	-21.17	-21.00	-21.01	-20.99
	upper hinge	-21.66	-21.57	-21.56	-21.51	-21.14	-20.99	-21.00	-20.99
		-21.68	-21.60	-21.59	-21.54	-21.14	-20.98	-20.99	-20.96
	median	-21.65	-21.56	-21.55	-21.51	-21.13	-20.99	-21.00	-20.98
		-21.67	-21.60	-21.58	-21.54	-21.14	-20.97	-20.98	-20.95
	lower hinge	-21.64	-21.56	-21.54	-21.50	-21.12	-20.98	-20.99	-20.97
		-21.67	-21.59	-21.58	-21.53	-21.13	-20.96	-20.97	-20.93
	minimum	-21.62	-21.53	-21.52	-21.48	-21.10	-20.95	-20.96	-20.91
		-21.64	-21.57	-21.56	-21.51	-21.11	-20.93	-20.94	-20.90
		non-fiber				fiber			
		Haar	D2	LA4	MB4	Haar	D2	LA4	MB4
sSURE	maximum	-22.58	-22.28	-22.22	-22.11	-20.04	-19.82	-19.87	-19.80
		-22.57	-22.27	-22.21	-22.09	-20.04	-19.78	-19.82	-19.74
	upper hinge	-22.57	-22.27	-22.21	-22.09	-20.00	-19.76	-19.79	-19.75
		-22.56	-22.25	-22.19	-22.08	-20.00	-19.75	-19.76	-19.72
	median	-22.56	-22.26	-22.20	-22.09	-19.99	-19.75	-19.78	-19.73
		-22.55	-22.25	-22.19	-22.07	-19.99	-19.73	-19.75	-19.70
	lower hinge	-22.56	-22.25	-22.20	-22.08	-19.97	-19.73	-19.76	-19.71
		-22.55	-22.24	-22.18	-22.06	-19.97	-19.72	-19.73	-19.68
	minimum	-22.53	-22.24	-22.19	-22.06	-19.94	-19.70	-19.73	-19.63
		-22.53	-22.23	-22.18	-22.04	-19.94	-19.68	-19.69	-19.65

Table 1: Boxplot statistics of log AMSE for different thresholding rules with either sigma1 (black) or sigma2 (gray) noise estimation using different wavelet families.

		non-fiber				fiber			
		Haar	D2	LA4	MB4	Haar	D2	LA4	MB4
hUNIV	maximum	-25.87	-26.97	-26.39	-25.97	-24.49	-24.42	-24.37	-24.22
		-26.20	-27.29	-26.76	-26.14	-24.53	-24.59	-24.39	-24.31
	upper hinge	-23.28	-23.23	-23.19	-23.14	-21.95	-21.86	-21.93	-22.00
		-23.41	-23.37	-23.32	-23.25	-21.93	-21.80	-21.86	-21.93
	median	-22.75	-22.57	-22.50	-22.33	-21.41	-20.93	-20.98	-21.00
		-22.97	-22.76	-22.67	-22.48	-21.41	-20.83	-20.86	-20.90
	lower hinge	-22.07	-21.46	-21.38	-21.20	-20.78	-20.31	-20.35	-20.36
		-22.19	-21.47	-21.39	-21.19	-20.72	-20.11	-20.15	-20.15
	minimum	-18.64	-18.73	-18.45	-18.26	-18.38	-17.90	-17.68	-17.50
		-18.45	-18.28	-18.40	-18.03	-18.06	-17.57	-17.37	-17.10

		non-fiber				fiber			
		Haar	D2	LA4	MB4	Haar	D2	LA4	MB4
sUNIV	maximum	-27.48	-28.49	-28.29	-27.56	-27.68	-27.61	-27.63	-27.38
		-27.62	-28.59	-28.45	-27.70	-27.79	-27.72	-27.79	-27.43
	upper hinge	-25.45	-25.64	-25.48	-25.35	-21.63	-21.57	-21.66	-21.71
		-25.55	-25.74	-25.56	-25.44	-21.58	-21.49	-21.58	-21.60
	median	-24.04	-24.03	-23.87	-23.65	-20.07	-20.06	-20.17	-20.23
		-24.03	-24.05	-23.91	-23.71	-19.88	-19.76	-19.86	-19.84
	lower hinge	-22.67	-22.24	-22.25	-22.06	-19.06	-18.83	-18.88	-18.91
		-22.60	-22.17	-22.18	-21.99	-18.85	-18.53	-18.56	-18.56
	minimum	-17.53	-17.74	-17.73	-17.27	-16.31	-16.39	-16.49	-16.23
		-17.53	-17.63	-17.80	-17.23	-16.19	-16.18	-16.31	-16.11

		non-fiber				fiber			
		Haar	D2	LA4	MB4	Haar	D2	LA4	MB4
hSURE	maximum	-24.66	-25.04	-24.53	-24.43	-24.41	-24.86	-24.61	-24.31
		-25.05	-25.14	-24.94	-24.72	-24.72	-25.30	-24.92	-24.57
	upper hinge	-23.06	-23.07	-23.07	-23.06	-21.86	-21.84	-21.89	-21.97
		-23.22	-23.21	-23.25	-23.26	-21.91	-21.93	-22.00	-22.11
	median	-21.69	-21.69	-21.69	-21.68	-21.21	-21.09	-21.09	-21.09
		-21.71	-21.76	-21.75	-21.75	-21.23	-21.11	-21.11	-21.10
	lower hinge	-21.26	-21.19	-21.15	-21.05	-20.78	-20.61	-20.63	-20.63
		-21.29	-21.23	-21.20	-21.10	-20.79	-20.60	-20.62	-20.62
	minimum	-19.82	-19.78	-19.59	-19.43	-19.57	-19.19	-18.99	-18.74
		-19.73	-19.54	-19.44	-19.18	-19.50	-18.91	-18.81	-18.55

		non-fiber				fiber			
		Haar	D2	LA4	MB4	Haar	D2	LA4	MB4
sSURE	maximum	-27.52	-28.42	-28.05	-27.84	-27.99	-27.84	-28.00	-27.64
		-27.63	-28.33	-28.14	-28.02	-27.95	-27.95	-27.75	-27.74
	upper hinge	-25.13	-25.35	-25.26	-25.14	-21.70	-21.61	-21.69	-21.76
		-25.13	-25.31	-25.25	-25.11	-21.74	-21.66	-21.74	-21.82
	median	-23.54	-23.55	-23.46	-23.30	-20.69	-20.53	-20.63	-20.66
		-23.54	-23.60	-23.51	-23.33	-20.71	-20.54	-20.63	-20.67
	lower hinge	-22.62	-22.20	-22.17	-22.01	-19.89	-19.58	-19.60	-19.57
		-22.62	-22.21	-22.19	-22.02	-19.89	-19.56	-19.57	-19.54
	minimum	-18.20	-18.30	-18.19	-17.74	-16.98	-16.92	-16.96	-16.68
		-18.14	-18.21	-18.08	-17.66	-16.95	-16.85	-16.93	-16.59

Table 2: Boxplot statistics of log VMSE for different thresholding rules with either sigma1 (black) or sigma2 (gray) noise estimation using different wavelet families.

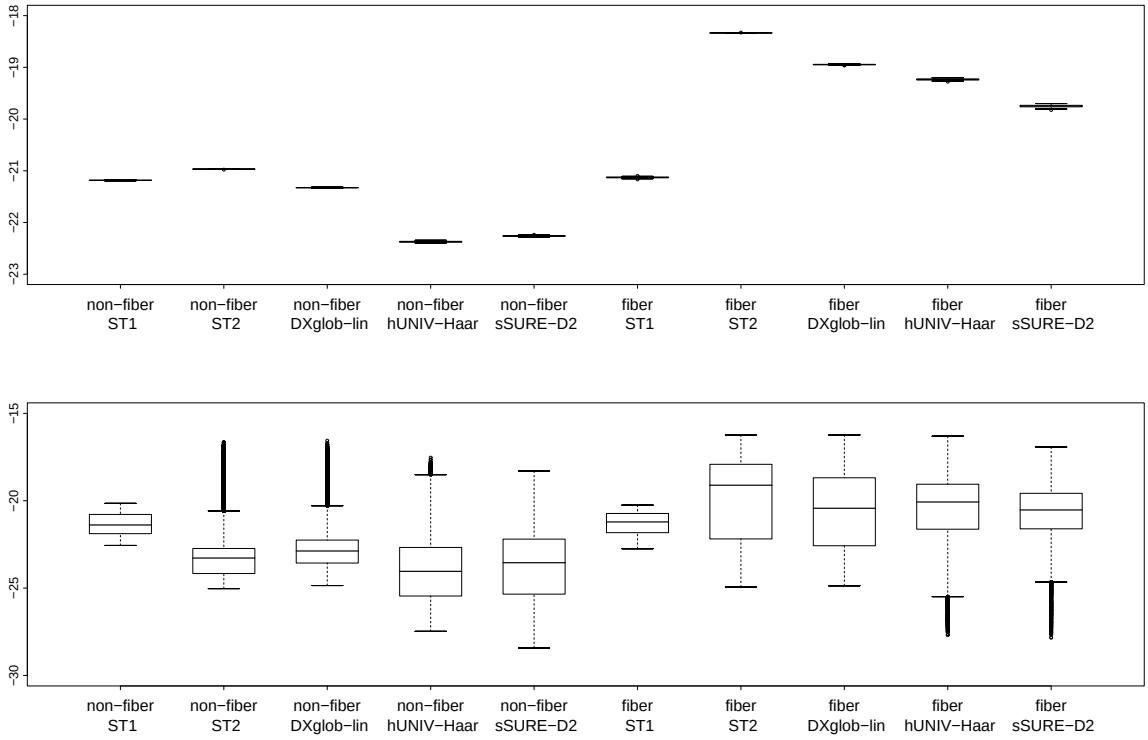


Figure 16: Boxplots of log AMSE (top) and log VMSE (bottom) for ST1, ST2, DXglob-lin, hUNIV-Haar and sSURE-D2.

$\text{VMSE}_{\text{method 1}}$ and $\text{VMSE}_{\text{method 2}}$ are build as in Section 4 in Heim et al. (2007). Where previously results from the B-spline based SVCM were contrasted with those obtained by ST2 (Heim et al., 2007, Fig. 5), wavelet filtered tensor estimates are now related to ST2 and to each other. Figure 17 contains the maps of $\log(\text{VMSE}_{\text{hUNIV-Haar}}/\text{VMSE}_{\text{sSURE-D2}})$ in the upper right, $\log(\text{VMSE}_{\text{hUNIV-Haar}}/\text{VMSE}_{\text{ST2}})$ in the lower left and $\log(\text{VMSE}_{\text{sSURE-D2}}/\text{VMSE}_{\text{ST2}})$ in the lower right corner. The interpretation is as follows: Rose colors indicate superiority of the numerator, while green colors hint to a better performance of denominator method. As obvious from the rose color dominance, both wavelet approaches outperform the standard voxelwise regression followed by Gaussian smoothing (bottom subfigures). In the interspiral space, however, ST2 shows higher accuracy. This becomes particularly evident when compared to sSURE-D2 (bottom right). The green pattern furthermore gives rise to the presumption of Gibb's artifacts caused by sSURE-D2. This phenomenon might be traceable in the estimates themselves (Fig. 18). In general, the wavelet filters seem to outperform the B-spline SVCMs with respect to background smoothness. This can be concluded without explicit comparison of the two methods (B-splines or wavelets), since

ST2 represents the benchmark in both figures 5 (Heim et al., 2007) and 17. However, the behavior at ridges is not as stable as in the B-spline case. For example, the wavelet approaches have difficulties to reproduce signals of low intensity (green spots in the fifth and sixth column of bottom maps).

As it can be judged from the upper right subfigure, hUNIV-Haar performs better at edges and yields smoother background in the first and second diagonal element, whereas sSURE-D2 primarily succeeds better in catching up the signal magnitude at ridges.

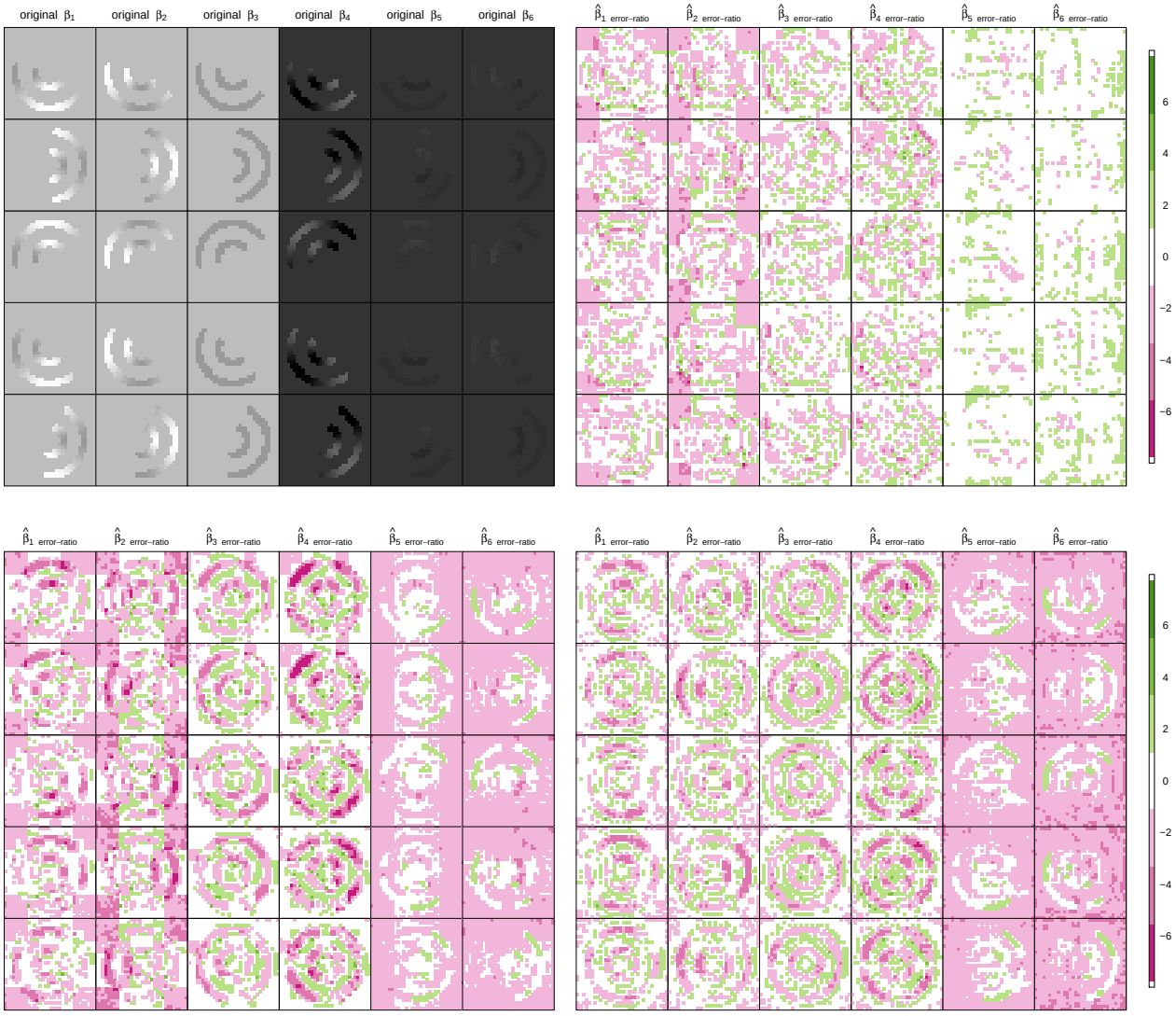


Figure 17: Template coefficient surfaces (top left) correspond to the six elements of the 3d tensor field. Log ratio of VMSE is given for both hUNIV-Haar (bottom left) and sSURE-D2 approach (bottom right) relative to ST2. Top right shows the log ratio of VMSE from hUNIV-Haar versus sSURE-D2. All subfigures show exemplary middle slices 14 to 18.

The white cross-shaped pattern hints to the presence of *blocking artifacts*, i. e. effects at different scales that are due to the squared support of the Haar wavelet. Blocking artifacts do not emerge with the D2 wavelet (bottom right).

Facing ambigously informative figures of the error ratios, a look onto the estimated diffusion tensors themselves seems worthwhile. For display feasibility, the focus is limited on the middle slices 14 to 18 and on the first diagonal tensor element for which error extremes as well as Gibbs and blocking distortions are both present (see Fig. 17). Figure 18 compares the original noise-free template with the standard approaches, the SVCM method and the wavelet filters applied to data from one exemplary simulation run.

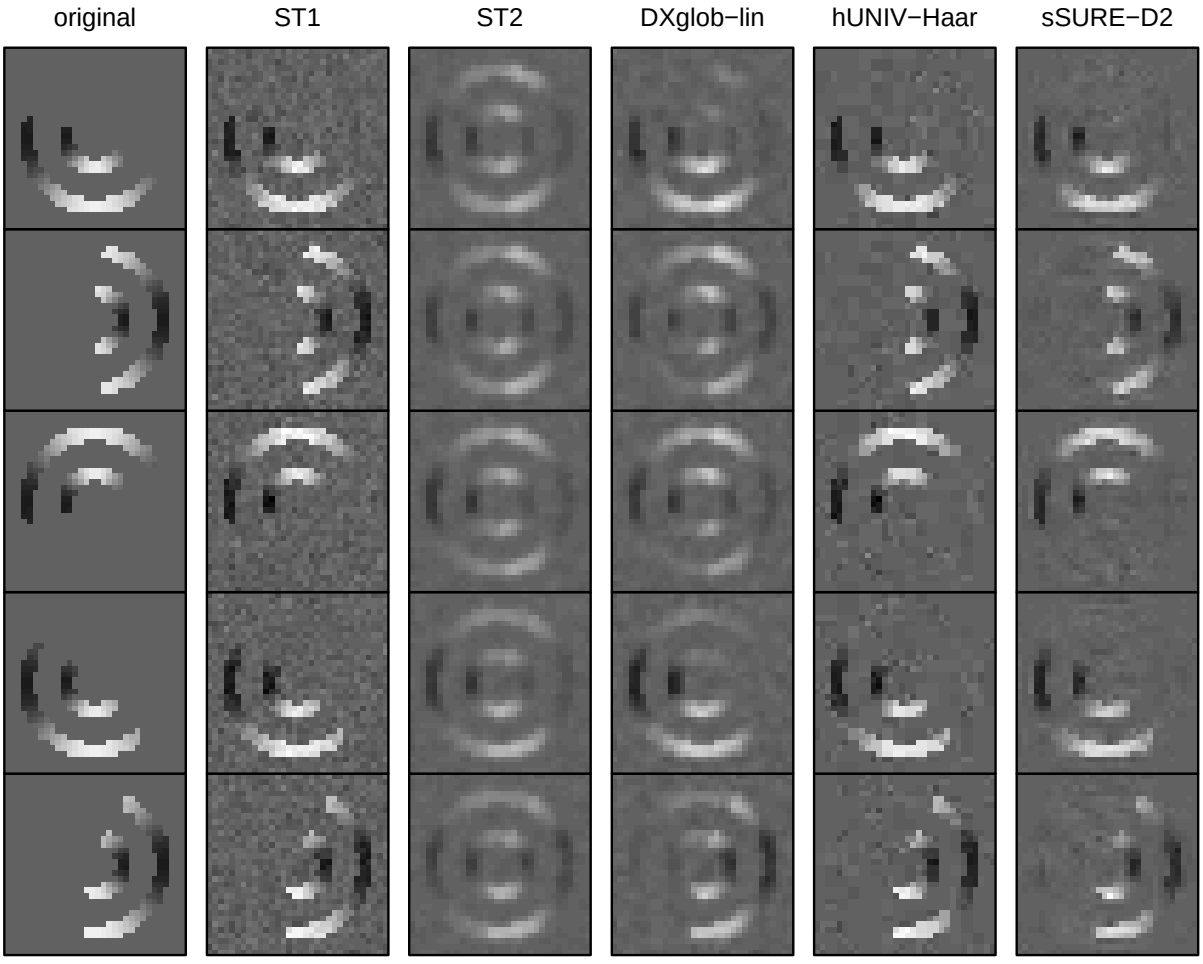


Figure 18: Surface excerpt (slices 14 to 18) showing the first diagonal diffusion tensor element as noise-free template (left) and estimated using voxelwise regression (ST1) with subsequent Gaussian smoothing (ST2) and space-varying coefficient approaches based on B-splines (DXglob-lin), Haar (hUNIV-Haar) or Daubechies wavelets (sSURE-D2).

Depending on the printing quality, all method specific drawbacks and advantages are more or less apparent at a glance:

ST1 Mere voxelwise regression completely lacks spatial smoothness though the signal magnitude seems to be caught satisfactorily in fiber voxels.

ST2 Consecutive application of a Gaussian kernel with FWHM equal to 0.75 voxels seriously oversmoothes, in particular in z-direction.

DXglob-lin Space-varying coefficient surfaces designed with penalized B-splines are less overregulated, contrasts are enhanced compared to ST2. Gibbs-like shadows are reduced to a minimum using linear basis functions.

hUNIV-Haar Features are correctly mapped regarding brightness and edge location. Smoothing is manifest but at the expense of blocking artifacts. For instance, partial failure of the spiral pathway happens at eight o'clock, top row. In the background, the Haar basis function is even fully recognizable (compare to Fig. 8). Gibbs phenomena are best visible as black voxels neighboring the bright curve segments.

sSURE-D2 For the first time, edge preservation appears satisfactorily balanced to smoothing. The interspiral space is slightly more turbulent for the D2 than for the Haar approach. Gibbs artifacts are still present, whereas blocking does hardly disturb the visual impression.

4 Real life example

In this section, an exemplary DTI data set of the human brain is processed using the wavelet filter that is rated best in the above simulations. For this purpose, the input data consists of six measurement repeats, each comprising 16 successive slices of originally 128×128 voxels. Although in-plane decomposition can in principle be determined by $j_0 = 7$, the possible maximum decomposition level in z-direction limits the thresholding to coefficients at the first four levels. This way, a realistic noise cutoff frequency is met at least within plane. Given the structural variety and the larger dimensions of real data, symmlets and minimum bandwidth wavelets serve for re-examination. The analysis is

therefore conducted with the wavelet families D2, LA4 as well as MB4 together with SURE shrinkage and level independent noise estimate (sigma1). After tensor estimation, the volume is resized to the same six slices of 90×75 voxels as used in Heim et al. (2007, Chap. 5).

A main difficulty, that did not occur with simulated samples, results from log-Cholesky parametrization of the voxel-by-voxel estimated tensor data (Eq. (11)). In few voxels, the Cholesky decomposition is not feasible straight away. Owing to the various sources of noise in MR imaging or to signal drop, the ST1 based covariance matrices are singular in about 0.5% of all 40500 voxels. For the locations of the affected voxels refer to Fig. 19. According to Skare et al. (2000b), negative eigenvalues are more likely with increasing noise and increasing FA. The latter statement holds for most of the affected voxels that seem to lie at the boundary to the ventricles where highly organized white matter is in direct neighborhood (Fig. 19). Hence, ensuring positive definiteness is particularly important with respect to fiber tracking. As remedy in such cases, small positive amounts are added on the diagonal elements prior to Cholesky decomposition.

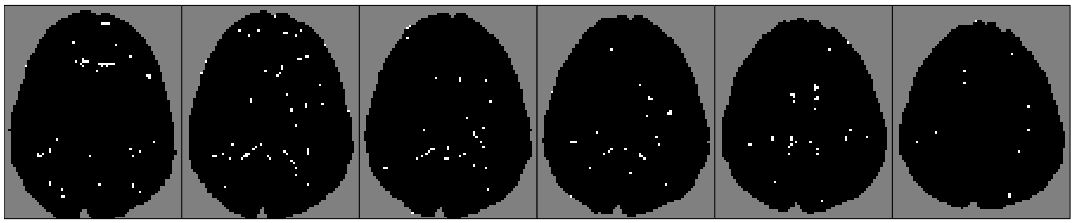


Figure 19: Mask of six middle human brain slices with problematic voxels superimposed in white. These indicate where the voxelwise estimates of the diffusion tensor are singular.

The problem of visualizing an estimated tensor field is equivalent to the problem of visualizing a vector field along with a scalar function on the same plot. Apart from showing the principal eigenvectors together with a color-code for the degree of local anisotropy, FA color-encoded ellipsoids are most intuitive and also somewhat more informative at first sight. For example, planar-shaped tensors may hint to an underlying fiber crossing or branching. Figure 20 was generated using the software module 'TensorViewer' developed by (Fillard and Toussaint, 2006).

Clearly, ST1 leads to the most speckled impression (Fig. 20 (a)). Interestlingly, the result is hardly discernible from the filtered version involving D4 wavelets (Fig. 20 (d)). The

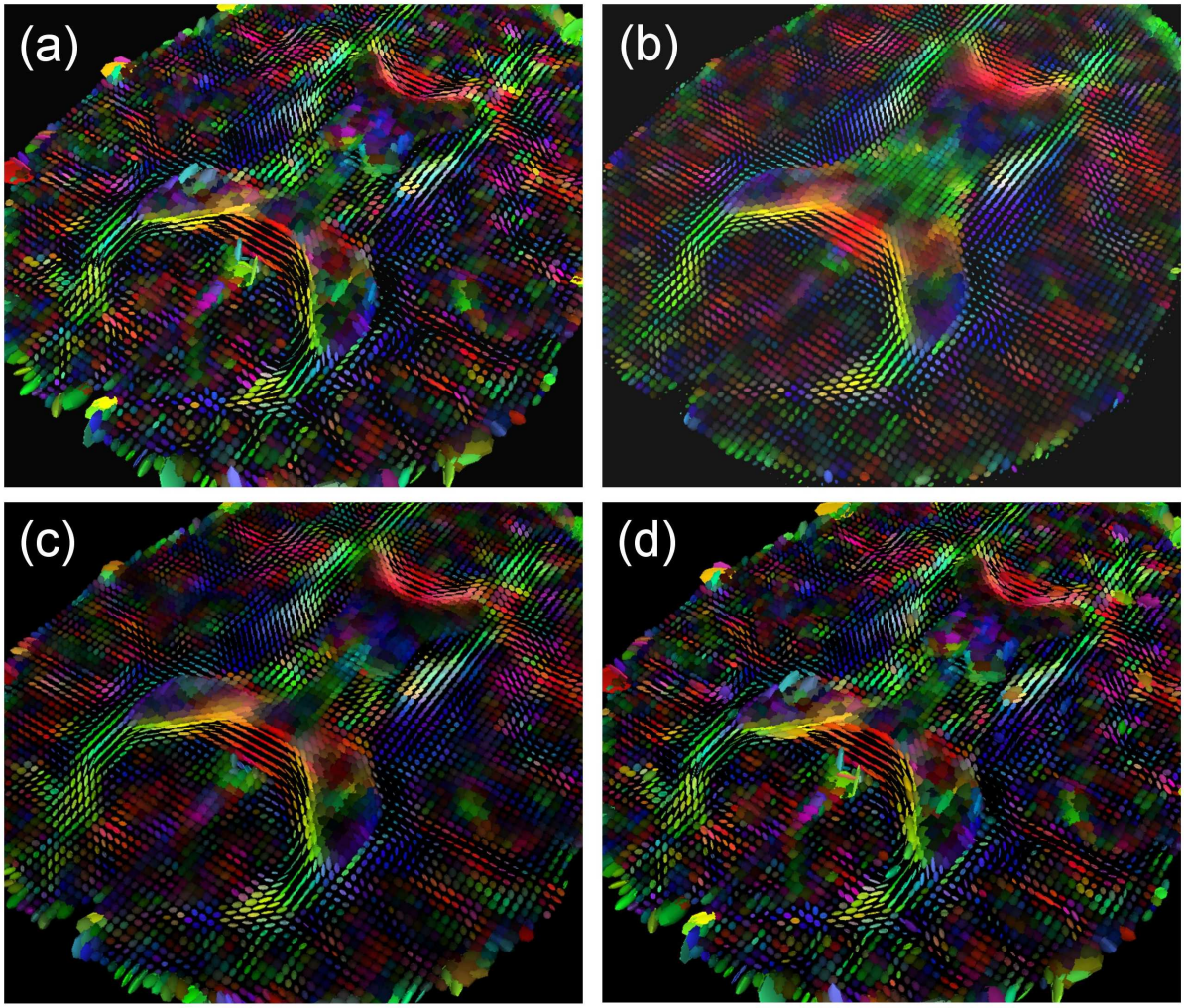


Figure 20: An exemplary slice of the 3d diffusion tensor field estimated from human DTI data using (a) ST1, (b) ST2, (c) DXglob-lin and (d) sSURE-D4, and displayed as FA color-encoded ellipsoids. The colorization is related to the principal diffusion direction and, thus, indicates left-right (red), anterior-posterior (green), or superior-inferior (blue) orientation of the associated fiber bundle. Intermediate directions are represented by the corresponding combination color.

disturbing effect is triggered by a few ellipsoids of bigger size than average, in particular present at the slice center and border. These tensors can be assigned to full or partial volumes of cerebro-spinal fluid (CSF), e. g. to the central ventricles where diffusion is free. Usually, a mask, built on the basis of the b_0 image, is applied to sort out CSF. In addition, these voxels are often excluded from fiber tractography. Since no masking was performed prior to estimation, the wavelet approach smears the ventricle compartement into the splenium of the corpus callosum, i. e. into the green-yellow-red-yellow-green c-shape. This result is contrary to what is anticipated from wavelet theory and lacks justification. Also, regularization is hardly given though presumably paramount for fiber

tractography. The results obtained by LA4 and MB4 wavelets do not differ significantly. In contrast, the tensor fields in Fig. 20 (b) and (c) convince by apparent smoothness. However, Gaussian kernel smoothing (Fig. 20 (b)) leads to overblurring in comparison to the SVCM with penalized B-splines. For example, the red-colored voxels in genu and splenium of the corpus callosum are less cigar-shaped though this region is highly anisotropic. Also in red, u-shaped region in grey matter contain more planar ellipsoids.

5 Discussion

Various simulation studies have been conducted accounting for different noise estimates, the use of different wavelet families and the application of different thresholding values and thresholding rules. As expected, scale dependent noise estimation proved to be inadequate to simulated data imposed with Gaussian white noise. Hard thresholding allows to reproduce structural aspects in almost full accuracy while uniform regions are still subjected to some error. Soft thresholding generally leads to smoother results, i. e. a correctly smooth background but also slightly blurred features. Depending on the application of interest, the soft shrinkage results are clearly to be favored with their moderate and adaptive regularization. The use of Haar wavelets should be avoided since undesired blocking effects are likely to emerge and the outcome equals a (higher-dimensional) step function. Least asymmetric separable 3d wavelets as well as minimum bandwidth wavelets showed neither beneficial nor detrimental over the use of Daubechies extremal phase family.

In contrast to former methods such as simple kernel smoothing or the B-spline based SVCM, the multiscale analysis with wavelets yields significantly less erroneous results due to the wavelets' built-in spatial-adaptivity. Space-frequency resolution particularly matters for the sensitivity to local curvature changes. B-splines (degree > 0), for example, cannot model aprupt changes due to their wide and overlapping support regions.
However, comparative error maps pointed to Gibbs-like phenomena accompanying wavelet filtering. Although these Gibbs artifacts are much more damped compared to the output from B-spline SVCMs, orthogonal wavelet analysis is not able to remedy them completely. Jansen (2001, p. 120) derives a theoretical explanation for the oscillation causes by looking at the optimal and experimental thresholding rules. According to (Jansen, 2001), hard

thresholding is more biased than soft thresholding and, thus, more likely to develop over- and undershoots near singularities. Yet, in the given simulation study, also the wavelet reconstruction using SURE shrinkage shows oscillatory behavior at fiber rims. Another reasoning considers the lack of translation invariance inherent to the DWT as origin of Gibbs effects (see Ogden (1997, p. 183) and the references therein). The use of the translation invariant wavelet transform (MODWT) may therefore be recommended.

Rather bizarre appears the fact that an SVCM involving Haar wavelets yields the best reconstruction in terms of MSE. The discrepancy between the resulting 3d step function and the intended smooth surface tackles the basic problem of the most relevant measure of error. Obviously, the "mean of squared errors (MSE) does not correspond to a human perception of generality. Since our visual system seems to work on a multiscale basis, a norm based on a multiresolution decomposition might be a better expression of visual quality" (Jansen, 2001, p. 49). In conclusion, the problem raises the question after the usefulness of theoretical optimality for experimental setups. Several proposals to encounter this "paradoxon" are made in Marron and Tsybakov (1995).

Finally, note that smaller errors of the tensor estimates do not necessarily imply smaller errors of the principal eigenvectors. Due to the non-linearity of the eigen decomposition, directional information of the main diffusivity could remain stable despite large AMSE and VMSE errors of the tensor entries. Recall that any tracking algorithm primarily exploit the eigenvectors. Hence, future studies should at least investigate the angular distribution between noise-free and reconstructed eigenvectors as well.

Concerning the human brain example, the SVCM on the basis of penalized linear B-splines yields the best visual result. Gaussian kernel smoothing leads to oversmoothing, as excepted, and voxelwise tensor estimation is obviously not apt for the ultimate goal of fiber tracking. Against the theoretical background and the results obtained in the simulation study, the wavelet approach fails in producing an appropriately regularized tensor field. Moreover, the edge-preserving property is challenged by unmasked CSF voxels. This disappointing result calls for further investigation of available wavelet transforms and thresholding mechanisms. Last but not least, the wavelet procedure needs to be adapted to the non-unit voxel size that is common in practice. Either the use of an anisotropic

basis or a scaling step seems advisable.

6 Proposal for a wavelet based SVCM

Ordinary least squares regression represents the standard approach to assess the diffusion tensor voxel-by-voxel. Subsequent wavelet filtering proved capable to preserve edges, while a low smoothing effect could be detected from the real life example (Fig. 4). More experience with this kind of basis functions is therefore needed, in particular in view of tractography. For the time being, let me sketch a possible solution to the original problem. The goal is to improve the B-spline based SVCM by substituting B-splines with wavelets. It has been substantially argued in Heim et al. (2007) that the covariance matrix elements are preferentially modelled non-parametrically as space-varying coefficient surfaces:

$$\boldsymbol{\beta}_j \;=\; \mathbf{B}\boldsymbol{\gamma}_j, \quad j = 1, \ldots, p.$$

The matrix $\mathbf{B}$ stands for B-splines and contains the basis functions evaluated at grid points $s, s = 1, \ldots, n$, as employed previously. A naive proposal would simply exchange $\mathbf{B}$ with the matrix $\mathbf{W}$ of wavelet basis functions. This is however hardly recommendable with Mallat's efficient pyramidal algorithm at hand (p. 14). Using the filterbank implementation, merely the wavelet coefficients are computed but not the wavelets themselves. Also, an analytical expression of the basis functions is often not available. Hence, the proposal must "eliminate" the matrix $\mathbf{W}$ as pointed out below.

A rearrangement of the r repeatedly measured $(n_1 \times n_2 \times n_3)$-dimensional diffusion weighted images into the vector $\mathbf{y}$,

$$\mathbf{y} = (\mathbf{y}_1{}', \ldots, \mathbf{y}_r{}')' = (y_{1,1}, \ldots, y_{n,1}, \ldots\ldots, y_{1,r}, \ldots, y_{n,r})', \quad n = n_1 n_2 n_3,$$

causes a swap of the Kronecker product in the joint model (Heim et al., 2007, Eq. (4)) for the n separate regression equations of the standard approach (ST1):

$$\mathbf{y} \;=\; \sum_{j=1}^{p} \big(\mathbf{X}(\cdot, j) \otimes \mathbf{I}_n\big)\boldsymbol{\beta}_j + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(0, \sigma^2 I_{rn}). \tag{13}$$

The jth vectorized surface $\boldsymbol{\beta}_j$, $\boldsymbol{\beta}_j = (\beta_j(1), \ldots, \beta_j(n))'$, is projected onto 3d tensor product wavelets:

$$\boldsymbol{\beta}_j \;=\; \mathbf{W}\boldsymbol{\gamma}_j, \quad j = 1, \ldots, p,$$

where $\mathbf{W}$ is $(n \times n)$-dimensional and encodes for the matrix of evaluated wavelet basis functions. In an image of dyadic dimensions, there are always as many wavelets as grid points. This allows to restate Eq. (13):

$$\mathbf{y} \;=\; (\mathbf{X} \otimes \mathbf{W})\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \tag{14}$$

with $(r \times p)$-dimensional regressor matrix $\mathbf{X}$ and $(pn \times 1)$-dimensional vector of wavelet coefficients $\boldsymbol{\gamma}$. For further reformulation, recall the basic properties of the Kronecker product (Dierckx, 1993, p. 170), where the vector operator $\text{vec}(\mathbf{X})$ returns the columnwise stacked matrix $\mathbf{X}$:

$$(\mathbf{A} \otimes \mathbf{B})' \;=\; \mathbf{A}' \otimes \mathbf{B}' \tag{15}$$

$$(\mathbf{A} \otimes \mathbf{B})^{-1} \;=\; \mathbf{A}^{-1} \otimes \mathbf{B}^{-1} \tag{16}$$

$$(\mathbf{AB}) \otimes (\mathbf{CD}) \;=\; (\mathbf{A} \otimes \mathbf{C})(\mathbf{B} \otimes \mathbf{D}) \tag{17}$$

$$\text{vec}(\mathbf{X}_{n\times m}\, \mathbf{B}_{m\times q}) \;=\; (\mathbf{B}' \otimes \mathbf{I}_n)\text{vec}(\mathbf{X}) \tag{18}$$

Subsequently, the OLS estimate of $\boldsymbol{\gamma}$ is derived from Eq. (14):

$$
\begin{aligned}
\hat{\boldsymbol{\gamma}}^{OLS} &= \big((\mathbf{X}\otimes\mathbf{W})'(\mathbf{X}\otimes\mathbf{W})\big)^{-1}(\mathbf{X}\otimes\mathbf{W})'\mathbf{y} \\
&\overset{(15)\text{–}(17)}{=} \big((\mathbf{X}'\mathbf{X})^{-1}\otimes(\underbrace{\mathbf{W}'\mathbf{W}}_{\mathbf{I}_n})^{-1}\big)(\mathbf{X}'\otimes\mathbf{W}')\mathbf{y} \\
&\overset{(17)}{=} \big((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\otimes\mathbf{W}'\big)\mathbf{y} \\
&\overset{(17)}{=} \big((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\otimes\mathbf{I}_n\big)\,(\mathbf{I}_r\otimes\mathbf{W}')\mathbf{y} \\
&= \left(\underbrace{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'}_{p\times r}\otimes\mathbf{I}_n\right)\begin{pmatrix}\mathbf{W}'\mathbf{y}_1\\ \vdots \\ \mathbf{W}'\mathbf{y}_r\end{pmatrix}_{nr\times 1} \qquad (19) \\
&\overset{(18)}{=} \operatorname{vec}\left(\underbrace{(\mathbf{W}'\mathbf{y}_1|\cdots|\mathbf{W}'\mathbf{y}_r)}_{n\times r}\,\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\right) \\
&= \begin{pmatrix}(\mathbf{W}'\mathbf{y}_1|\cdots|\mathbf{W}'\mathbf{y}_r)\ (\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1})\,(\cdot,1)\\ \vdots \\ (\mathbf{W}'\mathbf{y}_1|\cdots|\mathbf{W}'\mathbf{y}_r)\ (\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1})\,(\cdot,p)\end{pmatrix} = \begin{pmatrix}\hat{\boldsymbol{\gamma}}_1^{OLS}\\ \vdots \\ \hat{\boldsymbol{\gamma}}_p^{OLS}\end{pmatrix}
\end{aligned}
$$

Note that the vector in Eq. (19) contains the 3d wavelet transforms, $\mathbf{W}'\mathbf{y}_i$, $i = 1,\ldots,r$, of the raw images, thereby "eliminating" $\mathbf{W}$. If thresholding was directly applied to these wavelet coefficients, the proposal would work in analogy to the approach e. g. considered in Hahn et al. (2001), where smoothing comes prior to tensor estimation. In Eq. (19), the tensor model is however fitted in the wavelet domain. Hence, model (14) could alternatively be stated:

$$\mathbf{y}_{\mathrm{W}} = (\mathbf{X}\otimes\mathbf{I}_n)\boldsymbol{\gamma} + \boldsymbol{\varepsilon}_{\mathrm{W}},$$

with $\mathbf{y}_{\mathrm{W}} = (I_r\otimes\mathbf{W}')\mathbf{y}$ and $\mathrm{Var}(\boldsymbol{\varepsilon}_{\mathrm{W}}) = \mathrm{Var}\big((I_r\otimes\mathbf{W}')\boldsymbol{\varepsilon}\big) = \sigma^2 I_{rn}$.
The first approximations $\hat{\boldsymbol{\gamma}}_j^{OLS}$ to the amplitudes of the space-varying coefficient surfaces are independently thresholded using rule δ and cutoff value λ_j, $j = 1,\ldots,p$:

$$\hat{\boldsymbol{\gamma}}_{\lambda_j,j} = \delta_{\lambda_j}\left(\hat{\boldsymbol{\gamma}}_j^{OLS}\right).$$

Finally, smooth estimates

$$\hat{\boldsymbol{\beta}}_j \;=\; \mathbf{W}\hat{\boldsymbol{\gamma}}_{\lambda_j,j}$$

can be obtained for the varying coefficient surfaces and thus the 3d diffusion tensor elements. The thresholding mechanism certainly represents the crucial part of the proposed wavelet based SVCM since it determines the amount of shrinkage. For example, a surface specific universal threshold could be applied in combination with either hard or soft thresholding:

$$\lambda_j^{UNIV} = \sigma_j\sqrt{2\log(n)}.$$

According to Donoho and Johnstone (1994), the noise level σ_j is estimated on the basis of the contributions from the $n/8$ high frequency components. Assume that these elements of the wavelet transform are stored on positions $1,\ldots,n/8$ in vector $\mathbf{W}'\mathbf{y}_i$. Thus each of the first $n/8$ rows of $\left(\mathbf{W}'\mathbf{y}_1|\cdots|\mathbf{W}'\mathbf{y}_r\right)$ contains the weights associated with a specific high frequency wavelet in the r raw images. The final noise estimation would then involve the linear combinations of those r weights with the jth column of $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$ for the given totality of high frequency wavelets. See Section 1.4 for details on alternative methods.

In conclusion, the outlined SVCM permits to expect reasonable if not even better results than generated by the B-spline based SVCM. Yet, first endeavors point at concealed problems. A thorough implementation and investigation of performance is deferred to future research.

Acknowledgements

I am grateful to Ludwig Fahrmeir and Thomas Kneib (Ludwig-Maximilians-University, Munich, Germany) for helpful comments, to Peter Müller (University of Texas, Houston, USA) for his initial support and to Philipp Sämann (Max-Planck-Institute of Psychiatry, Munich, Germany) for advice on neuroscientifical applications. The human DTI data stems from my research period at the Max-Planck-Institute of Psychiatry (Munich, Germany). This study was kindly supported by the collaborative research center SFB386 (Statistical Analysis of Discrete Structures) as part of the German Science Foundation (D. F. G.).

References

Blu, T. and Unser, M. (2003). A complete family of scaling functions: The (α, τ)-fractional splines, *Proceedings of the 28th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'03)*, Vol. VI, Hong Kong SAR, People's Republic of China, pp. 421–424.

Dierckx, P. (1993). *Curve and surface fitting with splines*, Oxford: Monographs on numerical analysis, Oxford University Press.

Donoho, D. L. (1992). Interpolating wavelet transforms, *Technical report*, Department of Statistics, Stanford University, CA/USA.
URL: *http://stat.stanford.edu/∼donoho/Reports/1992/interpol.pdf*

Donoho, D. L. and Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage, *Biometrika* **81**: 425–455.

Donoho, D. L. and Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage, *Journal of the American Statistical Association* **90**: 1200–1224.

Fillard, P. and Toussaint, N. (2006). *Medical Image Navigation and Research Tool by INRIA (MedINRIA) Tutorial v1.0*, Institut National de Recherche en Informatique at en Automatique (INRIA) - Research Project ASCLEPIOS, Sophia Antipolis, France.
URL: *http://www-sop.inria.fr/asclepios/software/MedINRIA/*

Gençay, R., Selçuk, F. and Whitcher, B. (2002). *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics*, Academic Press London.

Gössl, C., Fahrmeir, L., Pütz, B., Auer, L. M. and Auer, D. P. (2002). Fiber tracking from dti using linear state space models: Detectability of the pyramidal tract, *NeuroImage* **16**: 378–388.

Gudbjartsson, H. and Patz, S. (1995). The rician distribution of noisy MRI data, *Magnetic Resonance in Medicine* **34**: 910–914.

Hahn, K., Prigarin, S. and Pütz, B. (2001). Edge preserving regularization and tracking for diffusion tensor imaging, *in* W. J. Niessen and M. A. Viergever (eds), *Proceedings*

of the 4th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'01), Utrecht, The Netherlands, p. 195.

Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998). *Wavelets, Approximation, and Statistical Applications*, Vol. 129 of *Lecture Notes in Statistics*, Springer-Verlag, New York.
URL: *http://www.xplore-stat.de/ebooks/scripts/wav/html/index.html*

Hastie, T., Tibshirani, R. and Friedman, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer Series in Statistics, Springer-Verlag, New York.

Heim, S., Fahrmeir, L., Eilers, P. H. C. and Marx, B. D. (2007). 3d space-varying coefficient models with application to diffusion tensor imaging, *Computational Statistics and Data Analysis* . Preversion: SBF386 discussion paper 455,.

Jansen, M. (2001). *Noise Reduction by Wavelet Thresholding*, Vol. 161 of *Lecture Notes in Statistics*, Springer-Verlag, New York.

Johnstone, I. M. and Silverman, B. W. (2005). Ebayesthresh: R programs for empirical bayes thresholding, *Journal of Statistical Software* **12**: 1–38.

Marron, J. S. and Tsybakov, A. B. (1995). Visual error criteria for qualitative smoothing, *Journal of the American Statistical Association* **90**: 499–507.

Ogden, R. T. (1997). *Essential Wavelets for Statistical Applications and Data Analysis*, Birkhäuser Boston.

Pinheiro, J. C. and Bates, D. M. (1996). Unconstrained parametrizations for variance-covariance matrices, *Statistics and Computing* **6**: 289–296.

Skare, S., Li, T., Nordell, B. and Ingvar, M. (2000b). Noise considerations in the determination of diffusion tensor anisotropy, *Magnetic Resonance Imaging* **18**: 659–669.

Su, W. and Ward, R. K. (2006). An edge-based image interpolation approach using symmetric biorthogonal wavelet transform, *Proceedings of the 8th IEEE International Workshop on Multimedia Signal Processing (MMSP'06)*, Victory, BC/Canada.

Van de Ville, D., Blu, T. and Unser, M. (2006). Surfing the brain—An overview of wavelet-based techniques for fMRI data analysis, *IEEE Engineering in Medicine and Biology Magazine* **25**: 65–78.

Weaver, J. B., Yansun, X., Healy, D. M. and Cromwell, L. D. (1991). Filtering noise from images with wavelet transforms, *Magnetic Resonance in Medicine* **21**: 288–295.

Whitcher, B. (2005). *waveslim: Basic wavelet routines for one-, two- and three-dimensional signal processing*. R package version 1.5.
URL: *http://www.image.ucar.edu/staff/whitcher/*
http://www.image.ucar.edu/staff/whitcher/book/

Wirestam, R., Bibic, A., Lätt, J., Brockstedt, S. and Ståhlberg, F. (2006). Denoising of complex MRI data by wavelet-domain filtering: Application to high-b-value diffusion-weighted imaging, *Magnetic Resonance in Medicine* **56**: 1114–1120.