

Wavelets in NMR-imaging.

V. 2.5

# Chapter 1

## Introduction

The topic of the report is wavelets in NMR-imaging. Nuclear Magnetic Resonance is the phenomenon of absorption and emission of radio frequency pulses by nuclei under specific conditions. Under even more specific conditions we can acquire an image of a slice of an object containing nuclei, in magnetic resonance. This is called Magnetic Resonance Tomography (MRI).

Today, we regard MRI as one of the most frequently used methods as for medical practice as for scientific research. To be more precise we should regard functional MRI as the leading direction in the research, which uses the MRI methods. Functional MRI is the observation of brain activity, while one is performing specific tasks. The problem, often appearing in MRI research is the low signal to noise ratio (which is about one or two in many fMRI experiments). Still, noise suppression is a delicate and difficult task, because the separation of noise from the original data has to be done so that it preserves the actual image features, acquiring diagnostically relevant image. This leads to construction of versatile methods, applicable to various circumstances, rather than fitting for specific conditions.

On the other hand, there are many types of transform which can represent our data in such a way, that filtering (and noise suppression as a part of it) becomes easier and more productive.

In this case, the theory of wavelets gives us a tool, applicable to many practical tasks. Mainly, wavelets are used in the field of analysis of highly inhomogeneous signals and functions, which have to be transformed in a way, capable of representing not only the distribution of the energy in the frequency domain, but also to localize the smallest features of the signal.

This leads us to the idea of using wavelets in our experiments. But first, let's look at the theory of wavelets itself.

## Chapter 2

# Wavelets and wavelet transform

### 2.1 Introduction

Wavelet analysis is a specific type of linear transform of the signal and the representation of the data, carried by the signal. The basis of the transformation has a number of interesting properties, which allow us to concentrate on either these or that features of the process observed.

Wavelets are the functions, shaped as a small wave, well localized on the independent variable axis, having zero mean value and forming a basis in  $L^2$ . Basis functions are acquired from special mother functions, determining their shape and properties, what we'll see now on several historical examples.

### 2.2 A short history of wavelets

Let's look now at the origins of the wavelet transform, that are to the Fourier transform. It's well known that any  $2\pi$  periodic function  $f(x)$  can be represented by its Fourier series in the form of:

$$f(x) = a_0 + \sum_1^{\infty} (a_k \cos(kx) + b_k \sin(kx))$$
$$k = 1, 2, 3, \dots$$

where

$$a_0 = \frac{1}{2\pi} \int_0^{2\pi} f(x) dx$$

and

$$a_k = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos(kx) dx$$

$$b_k = \frac{1}{\pi} \int_0^{2\pi} f(x) \sin(kx) dx$$

However, it appeared, that not all  $2\pi$  periodic functions can be represented in such a way. In 1897 Du Bois-Reymond constructed a  $2\pi$  periodic function, which Fourier series diverged at a certain point.

This has lead to a question: "Is there a different orthonormal system of functions, defined on  $[0, 1]$ , such that for any function, continuous on  $[0, 1]$  its series converges uniformly on  $[0, 1]$ ?"

This question led to the creation of Haar system, using which we can write a series

$$\langle f, h_0 \rangle h_0(x) + \langle f, h_1 \rangle h_1(x) + \dots + \langle f, h_n \rangle h_n(x) + \dots$$

which converges to  $f(x)$  uniformly on  $[0, 1]$ .

Here  $\langle a, b \rangle$  stands for scalar product of  $a$  and  $b$ , and  $h_n$  stand for the elements of the Haar system which are acquired from the basic function

$$h(x) = \begin{cases} 1, & x \in [0, 1/2) \\ -1, & x \in [1/2, 1) \end{cases}$$

For  $n \geq 1$  we take  $n = 2^j + k$  where  $j \geq 0$  and  $0 \leq k \leq 2^j$  and thus define  $h_n$  as follows:

$$h_n = 2^{\frac{j}{2}} h(2^j x - k)$$

Finally, to complete the  $h_n$  set and to acquire an orthonormal basis for  $L^2[0, 1]$  we define  $h_0(x) = 1$ .

The approximation of  $f(x)$  by the Haar series is simply the approximation of a continuous function by step functions with values equal to the mean values of  $f(x)$  on the appropriate intervals.

There is a serious disadvantage in the Haar system, suppose that  $f(x)$  is not simply continuous, but also has a continuous derivative. Then the approximation by step functions is completely inappropriate. To avoid it, Faber and Schauder suggested to replace  $h_n(x)$  functions by their primitives, thus defining the Schauder basis. This basis may also be defined as one constructed from the mother function  $\Delta(x)$ , as the Haar basis was. Here we have:

$$\Delta(x) = \begin{cases} 2x, & x \in [0, 1/2] \\ 2(1-x), & x \in [1/2, 1] \end{cases}$$

and as in the previous case we define

$$\Delta_n(x) = \Delta(2^j x - k)$$

for  $n = 2^j + k$  and with  $j \geq 0$  and  $0 \leq k \leq 2^j$ . Then after adding  $\Delta_0(x) = x$  and a constant 1 we define a Schauder basis. And the function  $f(x)$  is then written as

$$f(x) = a + bx + \sum_1^{\infty} \alpha_n \Delta_n(x)$$

If we derive this expression term by term, we can see then that it turns into an expression for  $f'(x)$  in the Haar basis, thus solving the disadvantage of the Haar basis mentioned.

## 2.3 Into the wavelet theory

We have seen now that it is possible to construct a basis from a function, following a simple rule of shifts and scaling. We can also notice that the properties of the basis constructed and the decomposition of the function, obtained in that basis are mostly defined by the properties of the mother function.

Here we follow the works of Grossmann and Morlet and define the "mother wavelet"  $\psi(x)$  function as an infinitely differentiable, rapidly decreasing function, whose Fourier transform  $\hat{\psi}(\xi)$  satisfies the following condition:

$$\int_0^{\infty} |\hat{\psi}(t\xi)| \frac{dt}{t} = 1$$

for almost all  $\xi \in \mathbf{R}^n$ . And if  $\psi \in \mathbf{L}^1(\mathbf{R}^n)$  then this condition turns into:

$$\int_{\mathbf{R}^n} \psi(x) dx = 0$$

From the mother function  $\psi(x)$  we acquire the corresponding wavelet basis by translations of the  $\psi(x)$  function:

$$\psi_k(x) = \psi(x - k)$$

and by scaling of the same function:

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

It was shown that, represented like this, wavelets form an orthonormal basis in  $L^2(\mathbf{R}^n)$ , thus any function  $f(x)$  of the same class can be represented as a set of projections onto this basis. The elements of the decomposition, called wavelet coefficients, are then computed as scalar products:

$$C_{j,k} = \langle f, \psi_{j,k} \rangle = 2^{\frac{j}{2}} \int f(x) \overline{\psi(2^j x - k)} dx$$

The set of  $C_{j,k}$  is then called the analysis of the function and the reconstruction of  $f(x)$  from the coefficients is called the synthesis and is computed as:

$$f(x) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} C_{j,k} \psi_{j,k}(x)$$

## 2.4 Multiresolution analysis

Many wavelet families can be described not only by the wavelet function  $\psi$  but also by the scaling function  $\varphi$ , bound with the  $\psi$  function by equation:

$$\psi(x) = \sqrt{2} \sum_k g_k \varphi(2x - k)$$

where  $g_k$  are constant and are found out of a linear system rising from the condition of orthogonality of the scaling function and from conditions of vanishing moments of the wavelet function.

Wavelet families containing scaling and wavelet functions are said to produce multiresolution analysis, which means that basis is now formed not only by  $\psi_{j,k}$  functions, but also by  $\varphi_{j,k}$  functions, defined from the  $\varphi$  functions in the same manner  $\psi_{j,k}$  were defined from  $\psi$ :

$$\varphi_{j,k}(x) = 2^{\frac{j}{2}} \varphi(2^j x - k)$$

Then any function can be decomposed at a pre defined resolution level  $j_n$  to a set of different wavelet coefficients:

$$f(x) = \sum_k A_{j_n,k} \varphi_{j_n,k} + \sum_{j \geq j_n,k} C_{j,k} \psi_{j,k}$$

where  $A_{j,k}$  are the approximation and  $C_{j,k}$  the detail coefficients, taken from:

$$A_{j,k} = 2^{\frac{j}{2}} \int f(x) \varphi(2^j x - k) dx$$

$$C_{j,k} = 2^{\frac{j}{2}} \int f(x) \psi(2^j x - k) dx$$

If we regard the pre mentioned Haar basis, then the approximation coefficients will give us the averaged values of the function on the dyadic intervals and the detail coefficient will represent the fluctuation of the function from the mean value on the same intervals.

## Chapter 3

# Wavelet applications

### 3.1 2D analysis

As was shown any signal can be represented as a number of detail coefficient sets and a set of approximation coefficients. In MRI we deal not just with one dimensional signals, but, most common, with a 2D set of data, corresponding either to the image itself or to the  $K$  space representation of the distribution of spin magnetisation in the sample observed. Thus we need to apply the multiresolution analysis idea to our matrix data.

The solution proposed is getting four sets of data on each step, instead of just two. These four are a set of approximation coefficients, used for next steps, and three sets of detail coefficients, called vertical, horizontal and diagonal detail coefficients. First two of these are acquired by applying wavelet transform to the raw data using two types of transform (with  $\psi$  and with  $\varphi$  function) in different directions. Diagonal elements come after applying the  $\psi$ -function transform twice (same as the approximation coefficients do with the  $\varphi$  function).

### 3.2 Denoising

#### 3.2.1 Soft and hard thresholding

Proposed by Ruttimann, the method is based on the assumption that the signal is represented by number of strong coefficients and noise is uniformly spread through wavelet space.

The first step of processing is then the decomposition of data into several sets of wavelet coefficients, using the 2D transform algorithm. After that a threshold is computed in two steps. At first we answer a question if significant signal power lies in any of the acquired detail coefficients. The approximation ones are left unprocessed, because they contain only the dc or extremely low-pass information. Then on the second step, the sets which do not contain signal information are discarded and are replaced by a reduced coefficients. The significant sets are

then subject to the thresholding procedure. The cutoff level of the coefficients  $\lambda$  is taken as follows:

$$\lambda = \frac{1}{n_{pix}} \sum_{n \in \mathbf{IC}} \Sigma^2(n)$$

Here  $\mathbf{IC}$  stand for intracarnial region of the image and  $n_{pix}$  is the number of pixels falling into that region.  $\Sigma^2(n)$  is the sample variance at each point.

After  $\lambda$  is computed we can proceed with the denoising itself. The two types of thresholding offered, are soft and hard suppression.

Hard thresholding is described as:

$$C_{j,k} = \begin{cases} 0, & |C_{j,k}| < \lambda \\ C_{j,k}, & |C_{j,k}| > \lambda \end{cases}$$

where  $\lambda$  is the pre-computed threshold. Soft method deals with the same  $\lambda$ , but is slightly different in the coefficient behavior:

$$C_{j,k} = \begin{cases} C_{j,k} - \lambda, & C_{j,k} > \lambda \\ 0, & |C_{j,k}| \leq \lambda \\ C_{j,k} + \lambda, & C_{j,k} < -\lambda \end{cases}$$

After passing the threshold step, the coefficients are subject to the inverse wavelet transform to acquire the source denoised data.

This approach differs from the common one of applying a low pass filter to the raw data, because in that case we usually have the information about the shape of activated brain regions. For if the activation is focal then only small amounts of smoothing are best and a corresponding filter is picked then, but if the region of activation is diffuse, then smoothing would be appropriate and the other filter is used. With wavelets we do not make assumptions on the activation properties ad thus we can successfully analyze not only the two previous cases, but also cases where both kinds of activation are present.

### 3.2.2 Generalized likelihood

A different, more complex and more flexible algorithm is proposed by Pižurica and Williams. The method is based on preliminary detection of wavelet coefficients, representing features of interest for the following estimation of the probability density function, responsible for separation of noise from the signal. Such detection is possible under the approach that regions of interest propagate well across the scales of wavelet decomposition.

The first step of the algorithm is decomposing a noisy image to a number of wavelet coefficients  $w_k$ , then the estimation of a noise-free signal is given by:

$$y_k = \frac{r\xi_k\eta_k}{1 + r\xi_k\eta_k} w_k$$

where  $r$ ,  $\xi$  and  $\eta$  are the relations between probability density functions, calculated along the regions of interest.



The features of interested are estimated using the assumption about propagation of signal along wavelet scales. To use it we should take a cross product of detail coefficient processed and of the one of the coarser scale and then compare it to a heuristic tunable parameter. This operation results in a binary mask  $x_k$ , which shows us regions of interest ( $x_k = 1$ ) and the rest of the image ( $x_k = 0$ ). The next step is computing the parameters  $r$ ,  $\xi$  and  $\eta$ ;  $r$  here stands for probability ratio and equals to:

$$r = \frac{\sum_{k=1}^N x_k |_{x_k=1}}{N - \sum_{k=1}^N x_k |_{x_k=1}}$$

$\xi_k$  is a ratio between probability densities of magnitude distribution and  $\eta_k$  is the ration between the probability densities of energy distribution.

### 3.3 Non fourier encoding of MRI data

#### 3.3.1 Fourier encoding

Here we should look at the methods, used to acquire an image, using NMR. The basic principle of the phenomenon is the proportionality of the radio frequency absorbed/emitted by nuclei to the magnetic field applied. Thus if we apply a magnetic field uniformly changing along a direction (what MRI-people call a gradient), and look at the frequency of NMR, we'll see a distribution of spin density along the axis. That is called a frequency encoding.

But if we want to acquire an image, we need to encode spin density in two dimensions. It's easy to show that applying two gradients in different directions doesn't help, so what is called a phase encoding is used. With phase encoding we apply a gradient to a system, changing its value during pulse sequence repetition? thus acquiring an analogue of free induction decay not in a single experiment, but in step-by-step mode. The number of points in the direction of phase encoding. is then equal to the number of gradient value changes. The combination of phase and frequency encoding is the most MRI experiments.

#### 3.3.2 Wavelet encoding

We can show that, if a RF pulse is applied during the gradient, then only the spins with frequency, corresponding to the frequency of the signal are subject to NMR. This means that the signal, acquired is the product of the frequency distribution in the sample and a spectrum of the pulse. This is usually used to select as a slice, while the distribution in a slice is phase and frequency encoded, but it can also be suited to a implement a different encoding method.

The idea is to present pulses, which are will give us the multiplication by the basic wavelet functions  $\psi_{j,k}$  or  $\varphi_{j,k}$ . Then, having spin distribution  $\rho(\nu)$ , after the pulse sequence we will acquire approximation and detail coefficients:

$$A_{j,k} = \langle \rho, \varphi_{j,k} \rangle$$

and

$$C_{j,k} = \langle \rho, \psi_{j,k} \rangle$$

What is shifts ( $k$ ) and scaling ( $j$ ) in terms of frequency, will result in pulse shape ( $j$ ) and central frequency ( $k$ ) in terms of time. Thus for a decomposition at 3 levels and for example 16 points we need to apply the pulses, corresponding to the next functions. Twice -  $\varphi_{3,k}$ , twice -  $\psi_{3,k}$ , 4 times -  $\psi_{2,k}$ , 8 times -  $\psi_{1,k}$ . Every time the pulse is applied, central frequency is changed to supply the change of  $k$ . The frequency shift is taken so, that each for  $j$  full field of view is covered.

After the acquisition we apply the inverse wavelet transform to the coefficients and get the distribution of spin density along the direction of the gradient (which is applied during the pulse). The second direction is frequency encoded.

Up to this place there is no difference in the acquisition time between phase and wavelet encoding, The advantages appear when we try to decrease the experiment time by acquiring less information from the object (retaining resolution by replacing the missed data with zeros). With phase encoding such procedure will result in the loss of contrast and appearance of fake image components, because during each step of phase encoding information from entire object is acquired. Wavelet encoding, on the other hand, has low spatial resolution on low scales and better resolution no high scales. This gives us the opportunity to choose, which parts of image would lack detail after the number of acquired points is reduced.

Finally, with the same number of points, phase encoding image would lack a general contrast, but wavelet image would only be affected in the chosen area, which means that if we choose the area, where no features are present, then no important information is lost in the experiment, still, thanks to the low resolution components, the area is visible and large elements are represented. The wavelet encoding is used mainly in dynamical MRI, where it helps observing the dynamics only of the chosen areas.