Multiscale analysis of multi-channel signals

HÅKAN CARLQVIST

Doctoral Thesis
Stockholm, Sweden 2005
Akademisk avhandling som med tillstånd av Kungl Tekniska högskolan framlägges till offentlig granskning för avläggande av teknologie doktorsexamen i tillämpad matematik fredagen den 3 juni 2005 klockan 14.00 i sal D3, Kungl Tekniska högskolan, Lindstedsvägen 5, Stockholm.

© Håkan Carlqvist, June 2005

Tryck: Universitetsservice US AB
Abstract

This thesis consists of three papers.

I. Amplitude and phase relationship between alpha and beta oscillations in the human EEG (with V. V. Nikulin, J.- O. Strömberg and T. Brismar; To appear in Med. Biol. Eng. Comp.). We have studied the relation between two oscillatory patterns within EEG signals (oscillations with main frequency 10 Hz and 20 Hz), with wavelet-based methods. For better comparison, a variant of the continuous wavelet transform, was derived. As a conclusion, the two patterns were closely related and 70–90% of the activity in the 20 Hz pattern could be seen as a resonance phenomenon of the 10 Hz activity.

II. A local discriminant basis algorithm using wavelet packets for discrimination between classes of multidimensional signals (With R. Sundberg and J.-O. Strömberg). We have improved and extended the local discriminant basis algorithm for application on multidimensional signals appearing from multi-channels. The improvements includes principal-component analysis and cross-validation-leave-one out. The method is furthermore applied on two classes of EEG signals, one group of control subjects and one group of subjects with type I diabetes. There was a clear discrimination between the two groups. The discrimination follows known differences in the EEG between the two groups of subjects.

III. Improved classification of multidimensional signals using orthogonality properties of a time-frequency library (With J.-O. Strömberg). We further improve and refine the method in paper2 and apply it on 4 classes of EEG signals from subjects differing in age and/or sex, which are known factors of EEG alterations. As a method for deciding the best basis we derive an orthogonal-basis-pursuit-like algorithm which works statistically better (Tukey’s test for simultaneous confidence intervals) than the basis selection method in the original local discriminant basis algorithm. Other methods included were Fisher’s class separability, partial-least-squares and cross-validation-leave-one-subject out. The two groups of younger subjects were almost fully discriminated between each other and to the other groups, while the older subjects were harder to discriminate.
The Definition of Beauty is
That Definition is none -
Of Heaven, easing Analysis,
Since Heaven and He are one.

*Poem 988, Emily Dickinson*
Contents

1 Introduction 1
  1.1 Some background of Fourier series . . . . . . . . . . . . . . . . 1
  1.2 Windowed Fourier transform . . . . . . . . . . . . . . . . . . 2
  1.3 Continuous wavelets . . . . . . . . . . . . . . . . . . . . . . . 3
  1.4 Wavelet bases . . . . . . . . . . . . . . . . . . . . . . . . . . . 4
  1.5 A construction of compactly supported wavelets . . . . . . . . 5
  1.6 Wavelet packets . . . . . . . . . . . . . . . . . . . . . . . . . 6
  1.7 Best basis selection . . . . . . . . . . . . . . . . . . . . . . . 8
  1.8 Introduction to EEG . . . . . . . . . . . . . . . . . . . . . . . 11

2 Conclusions 13

Bibliography 15
Chapter 1

Introduction

The present thesis consists of three separate papers:

1. Amplitude and phase relationship between alpha and beta oscillations in the human EEG.

2. A local discriminant basis algorithm using wavelet packets for discrimination between classes of multidimensional signals.

3. Improved classification of multidimensional signals using orthogonality properties of a time-frequency library.

Since a common factor for all three papers within this thesis is wavelet-based methods for medical signal processing, the introduction mostly concerns about the background of wavelets. There is also a very brief introduction to the signal system in use—the EEG. Most of the historical notes originate in [8, 11, 16].

1.1 Some background of Fourier series

The concept of using sums of harmonically related sines and cosines, or complex exponentials, for describing periodic phenomena goes back to at least the Babylonians, who used this kind of ideas for predicting astronomical events. The modern history of the subject begins with L. Euler, who in 1748 examined the motion of a vibrating string. The main note of Euler was that the configuration of the vibrating string at some point in time is a linear combination of these normal modes, so is the configuration at any
subsequent time. Euler also showed that one could calculate the coefficients for the linear combination at later time from the coefficients at the earlier time. However, the previous property would not be particularly interesting unless it was true for a larger amount of functions. D. Bernoulli argued in 1753 that all physical motions of a string could be represented by linear combinations of normal modes. He did however not prove it and was heavily criticized by Lagrange and even Euler had strong doubts. Fourier did 50 years later study the heat diffusion and was in need of a powerful method for solving the heat equation. The tools he developed were related to the following equations

\[ f(t) = \sum c_n e^{int}, \]  

(1.1.1) 

and 

\[ c_n = \int f(t) e^{-int} \frac{dt}{2\pi}, \]  

(1.1.2) 

even though Fourier did not use complex exponentials, which was not used in Fourier series until well into the twentieth century. He also did extend (1.1.2) to the real line. Fourier did never state a convergence proof of any kind for his tools but showed an extended use of them in the solution of the heat equation. A first proof of his tools were instead done by Dirichlet in 1829, and the proof can be seen as a first attempt of a convergence theorem for Fourier series of piecewise continuous functions.

1.2 Windowed Fourier transform

Classically, the Fourier transform can be defined as 

\[ \hat{f}(\xi) = \int_{-\infty}^{\infty} f(t) e^{-i\xi t} dt. \]  

(1.2.1) 

The Fourier transform can be thought as the amount of sinusoidal oscillations \( e^{i\xi x} \) present in the function \( f \). However, the Fourier transform suffers from the matter of nonlocal representation. The behavior of a function in a very small open set will influence the global behavior or the Fourier Transform. Therefore, in signal analysis, one often encounters the so called short-time Fourier transform, or windowed Fourier transform, which was originally developed by Gabor in 1946 [9]. This consists of taking the inner product between the function \( f \in L^2(\mathbb{R}) \) and a window function, consisting of a real
1.3. CONTINUOUS WAVELETS

symmetric function \( g \in L^2(\mathbb{R}) \), \( g(t) = g(-t) \), which is translated at \( a \) and modulated by the frequency \( \xi \)

\[ g(a,\xi)(t) = e^{i\xi t}g(t-a), \]  
and the inner product can be calculated as

\[ Sf(u,\xi) = \langle f, g(a,\xi) \rangle = \int_{-\infty}^{\infty} f(t)g(a,\xi)(t)dt. \]  

For simplicity, \( g \) is chosen so \( ||g|| = ||g(a,\xi)|| = 1 \). With letting both the mean value of position \( \int t|g(t)|^2dx \) and the momentum \( \int \xi|\hat{g}(\xi)|^2d\xi \) equals zero \( \int |t||g(t)|^2dx < \infty \) and \( \int |\xi||\hat{g}(\xi)|^2d\xi < \infty \) then the state \( g(a,\xi)(x) \) will be centered around \( (a,\xi) \). \( g(a,\xi) \) defines a family of current states.

The Plancherel identity can be viewed as

\[ \int \int |\langle f, g(a,\xi) \rangle|^2d\xi da = 2\pi ||f||^2, \]  

where \( ||f||_2 = \langle f, f \rangle \) stands for the \( L^2 \)-norm of \( f \). This is a proof of energy conservation of the windowed Fourier transform. It is also easy to prove that the function \( f \) can be completely recovered from the projections \( \langle f, g(u,\xi) \rangle \) (see for example [12]).

\[ \frac{1}{2\pi} \int \int \langle f, g(a,\xi) \rangle g(a,\xi) d\xi da = f. \]  

1.3 Continuous wavelets

The idea of continuous wavelets arise from the windowed Fourier transform. When analyzing a signal with both high and low frequency components, the windowed Fourier transform turns non ideal due to at high frequencies the number of oscillations are too high, leading to numerical instability, and at low frequencies the number of oscillations are too low leading to bad resolution. In the early 1980s, Morlet developed a transform that could better handle such difficulties [10]. This transform is what we today name the continuous wavelet transform.

The wavelets here consist of a family of functions \( \{\psi_{a,b}\} \) where a mother wavelet \( \psi(t) \) is dilated and translated as

\[ \psi_{a,b}(t) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{t-b}{a}\right). \]  

Then the continuous wavelet transform is defined as

$$ W_f(a, b) = \int_{\mathbb{R}} \psi^*_a(t) f(t) dt. $$

(1.3.2)

The $L^2$ inversion formula then becomes

$$ f(t) = \frac{1}{\int_{\mathbb{R}} |\hat{\psi}(\xi)|^2 d\xi} \int_{\mathbb{R}} \int_{-\infty}^{\infty} W_f(a, b) \psi_a,b(t) \frac{dbda}{a^2}. $$

(1.3.3)

This equation is known as the Calderon’s resolution of the identity introduced in 1964 in the context of Banach space interpolation [1] and extensively used for providing “atomic decomposition” in Banach spaces of distributions. Morlet used it without knowledge of its existence.

### 1.4 Wavelet bases

The continuous wavelet transform has a high computational cost and therefore there exist other ways for decomposing a function into a wavelet basis. For discrete wavelet analysis, it is necessary to first define a multi resolution analysis. Let $V_n$ be an increasing sequences of subspaces $\{V_n\} \subseteq L^2(\mathbb{R})$ defined for $n \in \mathbb{Z}$ with

$$ \ldots V_{-1} \subset V_0 \subset V_1 \ldots, $$

(1.4.1)

together with a function $\phi \in L^2(\mathbb{R})$ such that

(i) $\bigcup_{n=-\infty}^{\infty} V_n$ is dense in $L^2(\mathbb{R})$, $\cap_{n=-\infty}^{\infty} V_n = \{0\}$

(ii) $f \in V_n$ if and only if $f(2^{-n} \cdot) \in V_0$

(iii) $\{\phi(x - k)\}_{k \in \mathbb{Z}}$ is an orthonormal basis of $V_0$

For each $j \in \mathbb{Z}$ then $\{2^{j/2} \phi(2^j t - k)\}_{k \in \mathbb{Z}}$ forms an orthonormal basis for $V_j$ and $\phi(t) = \sum_{n \in \mathbb{Z}} a_n \phi(2t - n)$. Then there is a function $\psi \in L^2(\mathbb{R})$ such that the set $\{2^{j/2} \psi(2^j t - k)\}_{j,k \in \mathbb{Z}}$ forms an orthonormal basis for the orthogonal complement $V_{j+1} \subset V_j$. Also $\psi(t) = \sqrt{\frac{1}{2}} \sum_{n \in \mathbb{Z}} b_n \phi(2t - n)$. The function $\psi$ is called the Wavelet. Taking Fourier transforms of the Scaling function as well as the wavelet function gives us

$$ \hat{\phi}(\xi) = m_0(\xi/2) \hat{\phi}(\xi/2) \quad \text{and} \quad \hat{\psi}(\xi) = m_1(\xi/2) \hat{\phi}(\xi/2), $$

(1.4.2)
where
\[ m_0(\xi) = \frac{1}{2} \sum_{n \in \mathbb{Z}} a_n e^{-i n \xi} \quad \text{and} \quad m_1(\xi) = \frac{1}{2} \sum_{n \in \mathbb{Z}} b_n e^{-i n \xi}. \quad (1.4.3) \]

The multipliers \( m_0 \) and \( m_1 \) is said to be the scaling filter (low-pass) and wavelet filter (high-pass) respectively. They are \( 2\pi \)-periodic and also satisfy
\[ |m_0(\xi)|^2 + |m_0(\xi + \pi)|^2 = 2. \quad (1.4.4) \]
\[ |m_1(\xi)|^2 + |m_1(\xi + \pi)|^2 = 2. \quad (1.4.5) \]
\[ m_1(\xi) m_1(\xi) + m_0(\xi + \frac{\pi}{2}) m_0(\xi + \frac{\pi}{2}) = 0. \quad (1.4.6) \]

1.5 A construction of compactly supported wavelets

The success of the wavelets lay in their ability to efficiently approximate particular classes of functions with few non-zero wavelet coefficients. For a particular function \( f \) the wavelet \( \psi \) must therefore be designed for producing a large number of wavelet coefficients \( \langle f, \psi_{j,n} \rangle \) that are close to zero. This depends mostly on the regularity of \( f \) but also of the number of vanishing moments and compact support of \( \psi \). A wavelet function \( \psi \) has \( p \) vanishing moments if
\[ \int_{-\infty}^{+\infty} t^k \psi(t) dt = 0 \quad \text{for} \quad 0 \leq k < p. \quad (1.5.1) \]

If \( f \) is regular and \( \psi \) has enough vanishing moments (if \( f \in \mathbb{C}^k \) and \( \psi \) has at least \( k - 1 \) vanishing moments) then the wavelet coefficients \( |\langle f, \psi_{j,n} \rangle| \) have size \( 2^{j-k+1} \) at fine scales \( 2^{-j} \). It can be shown (see for example [12]) that if \( \psi \) has \( p \) vanishing moments, \( \hat{\psi}(\xi) \) has \( p - 1 \) vanishing derivatives at \( \xi = 0 \) and \( \hat{h}(\xi) \) has \( p - 1 \) number of vanishing derivatives at \( \xi = \pi \).

The support size and the number of vanishing moments of a wavelet are a priori independent. However, the constraints imposed on orthogonal wavelets imply that if \( \psi \in V_0 \) has \( p \) vanishing moments then its support is at least of size \( 2p - 1 \) [6]. Wavelets with compact support can be constructed with finite impulse response conjugate mirror filters \( h \). We consider \( h \) to be real, which implies \( \hat{h} \) being a trigonometric polynomial:
\[ \hat{h}(\xi) = \sum_{n=0}^{N-1} h[n] e^{-i n \xi}. \quad (1.5.2) \]
Since \( \psi \) has \( p \) vanishing moments then, \( \hat{h} \) must have a zero of order \( p \) at \( \xi = \pi \). and can therefore be written as

\[
\hat{h}(\xi) = \sqrt{2} \left( \frac{1 + e^{-i\xi}}{2} \right)^p R(e^{-i\xi}).
\] (1.5.3)

Still the polynomial \( R(e^{-i\xi}) \) has to be designed to have minimum degree \( m \) and such that \( \hat{h} \) satisfies.

\[
|\hat{h}(\xi)|^2 + |\hat{h}(\xi + \pi)|^2 = 2.
\] (1.5.4)

As a result, \( h \) has \( N = m + p + 1 \) non-zero coefficients. In the paper by Daubechies [6], she proved that the minimum degree of \( R \) is \( m = p - 1 \). She furthermore constructed a wavelet that satisfied the minimal statement. The Daubechies recipe uses a real conjugate mirror filter \( h[n] \) and therefore \( |\hat{h}(\xi)|^2 \) is an even function and thus can be written as a polynomial in \( \cos \xi \) as

\[
|\hat{h}(\xi)|^2 = 2(\cos \frac{\xi}{2})^{2p} P(\sin^2 \frac{\xi}{2}),
\] (1.5.5)

and

\[
(1 - y)^p P(y) + y^p P(1 - y) = 1.
\] (1.5.6)

The Daubechies wavelets are optimal in the sense that they give a minimal support for a given number of vanishing moments.

1.6 Wavelet packets

Orthonormal wavelet bases have a frequency localization proportional to \( 2^j \) at resolution level \( j \). Starting from a compactly supported wavelet function \( \psi \), at resolution level \( j \) the measure of \( \text{supp}(\psi_{j,n}) \) is \( 2^{-j} \) times \( \text{supp}(\psi) \). Therefore, the wavelet bases have poor frequency localization for large \( j \). For some applications, especially in signal processing, orthonormal bases with better frequency localization is necessary. Wavelet packets, which are obtained from wavelets associated with MRA’s, were introduced by Coifman, Meyer and Wickerhauser [3]. They are generated by a decomposition of a space \( V_j \) into the lower resolution space \( V_{j+1} \) plus a detail space \( W_{j+1} \) as \( V_j = V_{j+1} \oplus W_{j+1} \) by dividing the orthogonal basis \( \{\phi_j(t - 2^j n)\}_{n \in \mathbb{Z}} \) of \( V_j \) into two new orthogonal bases

\[
\{\phi_{j+1}(t - 2^{j+1} n)\}_{n \in \mathbb{Z}} \text{ of } V_{j+1} \text{ and } \{\psi_{j+1}(t - 2^{j+1} n)\}_{n \in \mathbb{Z}} \text{ of } W_{j+1}.
\] (1.6.1)
1.6. WAVELET PACKETS

In the original article regarding wavelet packets [3] it was proved that if \( g[n] \) and \( h[n] \) defined a pair of conjugate mirror filters with

\[
g[n] = (-1)^{1-n} h[1-n],
\]

then the decompositions of \( \phi_{j+1} \) and \( \psi_{j+1} \) in the basis \( \{\phi_j(t-2^j n)\}_{n \in \mathbb{Z}} \) could be written as

\[
\phi_{j+1}(t) = \sum_{-\infty}^{\infty} h[n] \phi_j(t - 2^j n) \quad \text{and} \quad \psi_{j+1}(t) = \sum_{-\infty}^{\infty} g[n] \phi_j(t - 2^j n),
\]

and the family

\[
\{\phi_{j+1}(t - 2^{j+1} n), \psi_{j+1}(t - 2^{j+1} n)\},
\]

was an orthonormal basis for \( V_j \).

Instead of dividing only the approximation spaces \( V_j \) into detail spaces \( W_j \) we can divide the detail spaces further to derive new bases. Let the basis space \( W_{j}^p \) of basis \( B_{j}^p = \{\psi_{j}^p(t - 2^j n)\}_{n \in \mathbb{Z}} \) be the space at node \((j, p)\) within a binary wavelet packet tree. At the root of the tree, we start with the space \( W_{0,0} = V_0 \) consisting of basis functions \( \{\phi(t-n)\}_{n \in \mathbb{Z}} \) so \( \psi_0^0 = \phi_0 \). Recursively we can calculate the basis functions at all nodes of the wavelet packet tree. If the basis functions of \( W_{j}^p \) are already calculated, then the wavelet packet basis functions at the children nodes are

\[
\psi_{j+1}^{2p}(t) = \sum_{-\infty}^{\infty} h[n] \psi_{j}^{p}(t - 2^j n),
\]

and

\[
\psi_{j+1}^{2p+1}(t) = \sum_{-\infty}^{\infty} g[n] \psi_{j}^{p}(t - 2^j n),
\]

and that the orthonormal bases \( B_{j+1}^{2p} = \{\psi_{j+1}^{2p}(t - 2^{j+1} n)\}_{n \in \mathbb{Z}} \) and \( B_{j+1}^{2p+1} = \{\psi_{j+1}^{2p+1}(t - 2^{j+1} n)\}_{n \in \mathbb{Z}} \) are bases to the two orthogonal spaces \( W_{j+1}^{2p} \) and \( W_{j+1}^{2p+1} \) respectively such that

\[
W_j^p = W_{j+1}^{2p} \oplus W_{j+1}^{2p+1}.
\]
1.7 Best basis selection

When selecting a combination of basis coefficients from an over-complete library of bases that decomposes a signal in a desirable (defined by the user) way a few methods have been proposed. If we use the notation by [13], these methods originate from a library $\mathcal{D}$ of bases which is a collection of waveforms $(\psi_\gamma)_{\gamma \in \Gamma}$ and a decomposition of a signal (or function) $f$ into $\mathcal{D}$ as

$$f = \sum_{\gamma \in \Gamma} \alpha_\gamma \psi_\gamma. \quad (1.7.1)$$

Often it is more useful to approximate the decomposition of $f$ into $\mathcal{D}$ as

$$f = \sum_{i=1}^{m} \alpha_{\gamma_i} \psi_{\gamma_i} + R^{(m)}, \quad (1.7.2)$$

where $R^{(m)}$ is a residual. Finding the best decomposition of $f$ turns into an optimization problem in minimizing the residual $R^{(m)}$ for finding an (almost) complete basis or for minimizing the residual $R^{(m)}$ with just a very few basis functions.

For optimization issues it is better to arrange the library into a matrix $\Psi$ with the waveforms $\psi$ as columns. Let $p$ denote the number of bases within the library. If $f$ has length $n$ then the decomposition problem turns into

$$\Psi \alpha = f, \quad (1.7.3)$$

where $\Psi$ then is a $n \times p$ matrix and $\alpha$ is the vector coefficients in the decompositions problem. When the library furnishes a basis, then $\Psi$ is a nonsingular $n \times n$ matrix and the representation $\alpha = \Psi^{-1} f$. In the matter of finding an orthonormal basis then $\Psi^{-1} = \Psi^T$.

Methods of frames (MOF)

The Method of Frames method [7] picks out the solution of above to with finding

$$\min ||\alpha||_2 \text{ subject to } \Psi \alpha = f, \quad (1.7.4)$$

with respect to the $l^2$ norm. The solution $\alpha^\dagger$ to above is unique and can be written as

$$\alpha^\dagger = \Psi^\dagger f = \Psi^T (\Psi \Psi^T)^{-1} f. \quad (1.7.5)$$
1.7. BEST BASIS SELECTION

The two major disadvantages of the MOF algorithm is that first it is not sparsity-preserving. If the underlying object has a very sparest representation in terms of the library, then the coefficients found by MOF are likely to be very much less sparse. Second, MOF is intrinsically resolution-limited. No object can be reconstructed with features sharper than those allowed by the underlying operator $\Psi^\dagger\Psi$.

Matching pursuit

The Matching Pursuit algorithm [13] starts from an initial approximation $f^{(0)} = 0$ and a residual $R^{(0)} = f$, and iteratively builds up a sequence of sparse approximations, and at $k$ iterates, it identifies the library coefficient that best correlates to the residual then adds to the current approximation a scalar multiple of that coefficient, so that $f^{(k)} = f^{(k-1)} + \alpha_k \varphi_{\gamma_k}$, where $\alpha_k = \langle R^{(k-1)}, \varphi_{\gamma_k} \rangle$ and $R^{(k)} = f - f^{(k)}$. This algorithm is perfectly matched for orthogonal dictionaries, but for more general dictionaries, some examples have been constructed which badly foil the method. By forming the residual as $R^{(m)} = f - \sum_{i=1}^{m} \alpha_i \varphi_{\gamma_i}$, the residual will be orthogonal to all terms currently in the model. This improvement of the Basis Pursuit algorithm is called Orthogonal Matching Pursuit (OMP) [15]. The OMP can also be seen as the Matching Pursuit algorithm with a Gram-Schmidt orthogonalization of the directions of the projections. While the OMP works better for non-orthogonal basis libraries than MP, the computational cost is larger.

Basis pursuit

The basis pursuit algorithm is a more general method for an approximated decomposition that addresses the sparsity issue directly [2]. The method selects the representation of the signal whose coefficients have minimal $l^1$ norm.

$$\min ||\alpha||_1 \text{ subject to } \Phi \alpha = f.$$  \hspace{1cm} (1.7.6)

One can see this method as the Methods of frames with the difference that here the $l^1$ norm, instead of the $l^2$ norm, is used for minimization purposes. However, the exchange of norm has some major consequences. While the optimization problem in the Method of Frames is quadratic, it turns to a convex non-quadratic problem in the Basis Pursuit method and there is a strong connection to Linear programming [5]. The standard form of a LP-
problem of a variable $x \in \mathbb{R}^m$ can be written as
\[
\min C^T x \text{ subject to } A x = b, x \geq 0,
\] (1.7.7)
where $C^T x$ is the objective function, $A x = b$, a collection of equality constraints and $x \geq 0$ a set of bounds. If we put $c$ as a vector of ones, define two slack variables $u, v$ as
\[
\Phi = \Phi - \Phi = f. \tag{1.7.8}
\]
We obtain a standard form linear programming of size $L = 2P$ with
\[
A = (\Phi, -\Phi), \ x = \begin{pmatrix} u \\ v \end{pmatrix} \text{ and } b = f. \tag{1.7.9}
\]
Even though the Basis Pursuit algorithm is using an advanced LP-algorithm the method will have high computational cost due to the global cost function of all basis functions of the library.

Best orthogonal basis

The best orthogonal basis was derived as a fast algorithm for deciding which nodes to use in a library of orthogonal bases [4]. This method has its advantage when the library has a tree structure like in wavelet packets or local cosines. The basic idea is to define a cost function $C(f, B)$ for representing a function $f$ in a basis $B$. If we consider the library $D = \bigcup B^\lambda$, where $B^\lambda = \{\psi_m\}_{1 \leq m \leq N}$, the question is which basis $B^\lambda$ that approximates the function $f$ best? In [12] it is proved that for all concave functions $\phi$ and for two orthogonal bases $B^\alpha$ and $B^\gamma$, if
\[
\sum_{m=1}^N \phi(\frac{\langle f, \psi_m^\alpha \rangle}{\|f\|^2}) \leq \sum_{m=1}^N \phi(\frac{\langle f, \psi_m^\gamma \rangle}{\|f\|^2}),
\] (1.7.10)
then $B^\gamma$ is better than $B^\alpha$ in approximating the function $f$ in sense of lower cost.

However, practically there are no possibilities in testing for all concave functions. In [4] they did minimize the cost function with regard to the Shannon entropy $\phi(x) = -x \log x$, which is a concave function. They did then find the best basis regarding Shannon entropy. The solution is not unique regarding finding the very best basis, there may be other bases equally good.
When looking for the best basis in a tree structure the question is whether it is better to approximate a function $f$ (where $f$ is a signal of length $N$) with the basis of a node or with the bases of the leaves. If we start with the basis $B_j^p = \{ \psi_m \}_{0 \leq m < 2^{-j} N - 1}$ of $W_j^p$, then the cost function is calculated as

$$C(f, B_j^p) = \sum_{m=0}^{2^{-j} N - 1} \Phi \left( \frac{|\langle f, \psi_m \rangle|^2}{||f||^2} \right). \quad (1.7.11)$$

Since both $B_j^p$ and $B_{j+1}^{2p} \cup B_{j+1}^{2p+1}$ are orthogonal bases for $W_j^p$ we can calculate the cost function for approximating $f$ into these two bases. Because $\Phi$ is an additive function then we have

$$C(f, B_{j+1}^{2p} \cup B_{j+1}^{2p+1}) = C(f, B_{j+1}^{2p}) + C(f, B_{j+1}^{2p+1}). \quad (1.7.12)$$

The best basis $O_j^p$ of $W_j^p$ is then the basis that minimizes the cost among all the bases of $W_j^p$.

$$O_j^p = \begin{cases} O_{j+1}^{2p} \cup O_{j+1}^{2p+1} & \text{if } C(f, B_{j+1}^{2p}) + C(f, B_{j+1}^{2p+1}) < C(f, B_j^p) \\ B_j^p & \text{if } C(f, B_{j+1}^{2p}) + C(f, B_{j+1}^{2p+1}) \geq C(f, B_j^p) \end{cases} \quad (1.7.13)$$

The best basis $O_0^0$ for $W_0^0$ can be recursively calculated starting from the bottom of the tree at level $J$. The original algorithm uses Shannon entropy as a cost function.

$$C(f, B) = -\sum_{m=1}^N \frac{|\langle f, \psi_m \rangle|^2}{||f||^2} \log \frac{|\langle f, \psi_m \rangle|^2}{||f||^2}. \quad (1.7.14)$$

1.8 Introduction to EEG

Electrical brain signals of animals were measured for the first time by Richard Caton and the results were published in 1875. He also discovered that when he interrupted light falling on an animal’s eye, he detected variation in the electrical activity of the brain.

Dr. Hans Berger was early aware of the results by Caton and his goal was to measure human brain activity, an electroencephalograph (EEG). In the early 1920s Berger obtained his first results and in 1929 the first paper was published about EEGs of humans. In a publication from 1930 he showed
that the EEG (which he now named an encephalogram) consisted of both primary (which was named alpha) and secondary waves (which he named beta).

The electroencephalogram (EEG) has been in clinical use for more than 50 years. Even though it is not fully understood what an EEG is a measure of, clinical studies of changes during pathological conditions have been performed extensively. The waveforms recorded are thought to reflect the activity of the surface of the brain—the cortex. This activity is influenced by the electrical activity from the brain structures underneath the cortex.

The EEG is a measure of electrical potentials between electrodes on the scalp. Because the appearance of many electrodes, the EEG can be considered as a multi-channel- or a multidimensional signal. For an introduction to the normal findings of the EEG [14] is recommended. Based on Fourier methods, the EEG is split into several frequency bands. Today, the most studied are named the alpha band (8-13 Hz), beta band (13-30 Hz), gamma band (> 30 Hz), theta band (4-7 Hz) and delta band (< 4Hz). Many studies concern about changes within these bands during sensory stimuli events, and activity in these bands have been postulated to be linked to several brain functions.
Chapter 2

Conclusions

The main purpose of this thesis have been to improve wavelet based methods for analyzing multidimensional signals. Our choice of application have been the EEG, a multi-channel signal that can be seen as a multidimensional signal, but the methods might be extended to be adjusted for other multidimensional (or multi-channel) signals.

The first paper concerns about relationships between oscillatory patterns, differing in frequency content, within EEG signals. Due to the non-sinusoidal appearance of a neuronal signal, when performing harmonical frequency analysis of such a signal, resonance frequencies are seen at the harmonics of (an integer times) the base frequency of the signal. These harmonics can be mixed up with distinct patterns with base frequency of the harmonics. In the paper we show that the amplitude of the alpha (with base frequency 8-13 Hz) and beta (with base frequency 13-30 Hz) oscillations in spontaneous EEG at most are linearly dependent and that the alpha and beta oscillations are phase coupled. They should therefore be closely related. A mathematician would say the beta being a harmonic of the alpha oscillation. However, the methods used are not accurate enough to state a complete dependency, there still might be beta activity independent of alpha activity even though no proof of existence of such independency were found. These findings are very important when analyzing frequency bands other than the base frequency, because alterations in higher frequencies may be due to alterations in the base frequency. For a scientist with limited knowledge in mathematics this is not always complete clear.

The second and third paper concern about methods for classification of mul-
tidimensional signals. For some years there have existed methods for discriminating between classes of one-dimensional signals regarding differences within the time-frequency energy map of such classes of signals. We modify and improve this method to be used for multi-channel signals, like EEGs. As an example we show discrimination between two classes of EEG signals, one consisting of control subjects and one consisting of subjects with type I diabetes.

The third paper can be seen as an extension of paper two, where different modifications of the method derived in paper two are discussed. Four kinds of modifications were discussed and these modifications were done regarding

1. the choice of basis at component level,
2. the choice of coefficients at signal level
3. the choice of discriminant measure to use
4. the choice of classifiers.

The most successful method used 1. an orthogonal-basis-pursuit-like most discriminant orthogonal basis selection, 2. the most discriminant coefficients at each component independently, 3. Fisher’s class separability as discriminant measure, and 4. the partial least squares algorithm. 1. and 2. were statistically significant better \( p < 0.005 \), Tukey’s test for 95% simultaneous confidence intervals) than using other methods proposed. This method was further applied on discrimination analysis on 4 classes of EEG signals, from subjects differing in sex and/or age. Even though the classification into the four groups was not complete, the majority of the subjects were classified according to group membership.

To our knowledge, paper two and three are the first attempts to classify segments of spontaneous EEG regarding differences in the time-frequency map of classes of EEGs. Another area within EEG research where these methods would be excellent tools are within classification of EEG signals after different kinds of stimuli (event related signals).
Bibliography


Acknowledgments

A large number of people have contributed in their own ways to the making of this thesis. I am greatly indebted to them all. In particular I want to express my gratitude to:

- Jan-Olov-Strömberg for the never ending ideas and encouragement and for teaching me what is worth knowing about multiscale analysis.
- Rolf Sundberg for all enthusiasm and suggestions regarding improvements of my papers, and especially introducing me to the world of multivariate analysis.
- Tom Brismar and Vadim Nikulin for fruitful collaboration.
- Uno Nävert, the director of the NTM-network, for his support and for arranging all excellent network activities.
- Lars Villemoes for introducing me to the discrete world of wavelets.
- The PhD students at the department for making a social and scientific atmosphere.
- the people at the department of mathematics at KTH for creating a pleasant atmosphere to work in.
- All my friends for their endless support.
- My family for just being there for me.
- Malena for making it all worthwhile. Without you this thesis might not have been written.

Stockholm, April 2005

Håkan Carlqvist