Representation of Some Class of Time Frequency Distributions Towards Their Determination Through Remote Sensing Techniques

GEORGIOS KLADIS

MSc by Research in Pattern Analysis and Neural Networks

ASTON UNIVERSITY

September 2006

ASTON UNIVERSITY

Representation of Some Class of Time Frequency Distributions Towards Their Determination Through Remote Sensing Techniques

GEORGIOS KLADIS

MSc by Research in Pattern Analysis and Neural Networks, 2006

Thesis Summary

This project is concerned with reconstruction problems arising in the context of radar signalling. The problem arising in remote sensing is a classical inverse problem. The aim is to reconstruct the density of the target tracked from the echo, which bears two features distance and velocity, with respect to the source (radar). Also adaptive and recursive procedures are described for signal representation. The representation is build up non-linearly through functions selected from a redundant family of waveforms in a forward "greedy" kind of manner or backward.
Inspired by the work of H.Narparst [5] and L.Rebollo-Neira [14], the project will include experiments concerning the robustness of such methods in the case of prior knowledge of both the target environment and optimal support of the redundant family of waveforms (dictionary) used so as to represent the density of targets bearing these common features.

Keywords: Radar, remote sensing, echo, frames, wavelets, matching pursuit strategies, best orthogonal base, method of frames, prior knowledge, dictionary, optimal support.

Acknowledgements

I must express my deepest appreciation to the precious help and guidance of my supervisor Dr Laura Rebollo Neira for completing this thesis. In addition, I would like to thank the examiners Dr Juan P. Neirotti and Dr Pier Luigi Dragotti for their comments and help.
My greatest debt belongs to my beloved family and friends that supported and tolerated me throughout this year of studies. Guys you made it an unforgettable year.
A due appreciation and dedication is to Prof Grigorios Zeibekakis and Dr Anastasios Tzerachoglou for his support to fulfil my dream of searching beyond the expected.

Epimenides (6th c.BC) Hymn to Zeus 8-9
Quoted by Paul of Tarsus
at Acts 17:28 and Titus 1:12

They fashioned a tomb for thee [O Zeus], O holy and high one-
The Cretans, always liars, evil beasts, idle bellies!
But thou art not dead: thou livest and abidest forever,
For in thee we live and move and have our being.

Contents

Chapter 1

Radars

1.1 Introduction

It is widely known if a pulse or continuous wave is propagated through a device to the atmosphere then, if a receiver is used to 'listen', the echo of the signal that is sent will be collected. This occurs since materials have a special property to reflect and absorb portion of the frequency wave that is propagated to them. These special devices that perform such a task are called RADARS and are used for remote sensing of objects. Remote sensing is the science and art of obtaining information about an object, area, or phenomenon through the analysis of data acquired by a device that is not in contact with the object, area, or phenomenon under investigation.

1.2 Radar modelling

In general terms a radar comprises of a transmitter, a receiver, a multiplexer (in order to switch between the two different functions send a signal-receive its echo) and an imaging device. So a problem one should address is what could be the input of such a device, how long should we 'listen' in order to acquire the echo of the signal propagated, what is the object of radar interest and how can one recognize or reconstruct the target from the echo that is received? In essence, the echo that is received is a modulated signal which is the process of varying a *carrier signal* in order to use that signal to convey information. On the other hand the inverse process is called demodulation. In essence when a signal is sent, which is reflected by an unknown object, then due to the properties of atmospheric air, the signal received is

corrupted with noise or generally modulated [1][2]. Consequently by demodulating it, certain features and properties, such as the delay can be calculated. So by measuring these properties the position and the speed of the target observed can be maintained. In this thesis we will concentrate in ideal circumstances where the signal that is propagated is reflected by the target in total and that the echo received is not corrupted with noise. Moreover we will concentrate on a density of many reflecting objects or a dense-target environment, inspired by the work of Naparst [3], rather than just a point target. *A dense-target environment is a group of targets that might be closely spaced with its own range and velocity from the source.*

In recent years, wavelet analysis has been very successfully applied to many problems in signal and image processing [4]. In the case of observing a point target one can send just one signal and investigate the properties of the reflected signal (echo) in order to pinpoint the location and velocity of the target. However in the case that a point target is considered, which is not rigid, the question that immediately arises is *whether one could determine the reflectivity density by transmitting not just one signal but a family of signals.* The previous was answered positively since the pioneering work of H.Naparst [3][5].Assuming that the reflectivity density remains completely invariant during observation time and the same density produces all echoes received, Narpast has shown that it can be obtained by transmitting a set of signals and measuring the corresponding echoes. In other words by transmitting a set of signals and receiving the corresponding echoes one can form a set of equations, one for each transmitted signal, that will allow inversion of the problem, such as the reconstruction of the density. So he suggested a method to determine the spreading function from the echoes produced by a set of orthogonal waveforms.

Traditional methods for signal representation involve the use of orthogonal bases in order to perform a M-term approximation. That approximation where one fixes a basis and searches for a combination of M-terms of the basis of N-terms ($M \subseteq N$) is a nonlinear problem. The nonlinearity is due to the fact that this subset of M-terms depends on the algorithm that is used to select it. Consequently the problem of finding the best combination for such an approximation has been subject of much mathematical work, even nowadays. The approximation has a basic ingredient which is a basis called a *dictionary* and is formed by a set of functions referred as *atoms*. Some methods of approximations that fall in the above category are adaptive pursuit [4][6],[7],[8],[9],[10] and adaptive basis selection[11].

Moreover, through a particular experiment of representation of a target environment of two parameters, Naparst showed how deviations in those might cause the method to fail when the set of signals is chosen priori to what the target was likely to be and left some room for further analysis.

The aim of that project will be, through these nonlinear methods, to use *prior knowledge* both of the density of objects being tracked and the orthogonal basis used in order to examine robustness of their representation phase. Based on these, a series of experiments was designed hoping to draw some theoretical conclusions.

Most of the methods used in order to obtain an approximation of the density of targets fall in the category of adaptive pursuit. Some numerical examples of these approaches are going to be presented, without comparing them with regards to convergence rate, since we are in lead for a different motivation. Our goals would be to use prior knowledge of both the density of targets being tracked or the combination of waveforms being sent in order to approximate that input target.

This thesis is organized as follows: In chapter 2 the radar problem is firstly addressed and we will analyze how to obtain the reconstruction of the spreading function. Chapter 3 refers to some preliminaries concerning dictionaries and atoms. In addition, in chapter 4 adaptive methods are explained, such as the methods used in order to send a finite number of elementary signals (chosen from a dictionary) that will approximate the target, which will be used extensively in the experiments. In chapter 5 the analysis with some simple examples on Orthogonal Matching Pursuit for the representation of functions of two variables is described. Thereafter prior knowledge is considered for the target being tracked and the subset of waveforms sent, chapter 6 and chapter 7, respectively. Moreover a subset of waveforms is proposed in order to approximate the density of targets bearing assumptions considered. Thereafter, in chapter 8 it is shown with simple numerical experiments the capabilities of such subsets to approximate a target environment based on prior knowledge. Lastly Chapter 9 illustrates through some simple examples the advantage of combining different dictionaries (mixed dictionaries) for the representation phase. Partial conclusions and remarks are given in each chapter and some general ones are drawn in the last one.

Chapter 2

Radar setting

2.1 Basic Radar Setting

To explain the basic radar setting, suppose that the object under consideration can be described as a point moving with a constant velocity u towards or away the source. The distance between the object observed and the source is R. So the task of a radar technique is to image or locate an object by analyzing electro-magnetic waves that are reflected from these objects. In general, the range and the velocity of the target observed is determined by transmitting an electro-magnetic pulse $y(t)$ and receiving and analyzing the echo $f(t)$ produced by the object. Adapting the notation and terminology in [12][13], for objects moving at constant velocity then the *wide-band model* for the received echo $f(t)$ is a delayed and scaled signal of y (signal transmitted) given below:

$$f(t) = ay[s(t-\tau)] \qquad (2.1)$$

where, s is obtained from the speed of propagation of the transmitted pulse c (speed of light) and the object velocity u as

$$s = \frac{c-u}{c+u} \qquad (2.2)$$

and the delay τ is determined by the distance R between the object and the source as

$$\tau = \frac{2R}{c-u} \qquad (2.3)$$

Hence from s and τ the range R between the target and the source and the velocity u of the target can be computed as $u = \frac{1-s}{1+s}c$ and $R = \frac{cs\tau}{1+s}$. Moreover the factor a $(0 \le a \le 1)$ is a constant representing the *reflectivity* of the object where in the case of an energy preserving model it should be considered as $\sqrt{|s|}$. Since a perfectly reflecting object is assumed, in the experiments following the reflectivity factor will be considered equal to one (energy is

conserved). The reader may refer to the particular work ([12][13]) for a thorough discussion of the derivation of the wide-band model, the time-scaling factor s and the time delay τ. Due to the fact that all objects travel with a speed which is much smaller than the speed of light the above equations reduces to $s \approx 1$ and $\tau \approx 2R/c$.

In addition to that, if the outgoing signal bandwidth is narrow and centred around a high frequency f_c there is the *narrow-band model*. This is of the following form

$$f(t) = a\exp(-2\pi\imath\phi t)g(t-\tau) \qquad (2.4)$$

where, $\phi = 2\frac{u}{c}f_c$ (Doppler frequency shift) , $\tau = \frac{2R}{c}$ (delay), ($c >> u$ so $c-u \approx c$) and $g(t)$ is the transmitted waveforms. The above is widely used in classical radar. More details on how to reduce from the wide-band model to the narrow-band one, can be obtained in references [12][13]. Consequently the echo *f(t)* is a delayed and modulated version of the transmitted signal for the narrow-band approximation.

The Doppler coordinates (ϕ, τ) can uniquely determine the distance and the velocity of the target due to their one to one correspondence. However, assuming a target that can be represented by a point scatter which is determined by a single velocity and range is not always realistic. For instance the object may not be rigid and its different components might have different velocities, such as a cloud. Moreover though the object is rigid, it may be changing orientation with respect to the radar or there may be many reflecting objects densely situated with its own ranges and velocities. So, a more realistic assumption is in the presence of a density function describing the scatter spreading effects upon the transmitted signal and the total echo is modelled as the superposition of the single echoes. In general, the total echo of a reflecting continuum with varying reflectivity as described in Doppler coordinates can be obtained by formula (2.5) as described in [3][12] [13] [14].

$$f(t) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p(\tau,\phi)\exp(-2\pi\imath\phi t)g(t-\tau)d\tau d\phi \qquad (2.5)$$

The function $p(\tau,\phi)$ is called the spreading function, ϕ is the Doppler frequency shift which is equal to $2\frac{u}{c}f_c$, $\tau = \frac{2R}{c}$ the delay and $g(t)$ the transmitted waveform. Note that the total echo *f(t)* is a delayed and modulated version of the transmitted signal and that the factor α is equal to one due to the fact that there is a perfectly reflecting target environment. So the task is to reconstruct the density $p(\tau,\phi)$ from the received echo. The problem can be dealt in an inverse manner provided the echo and the waveform that is transmitted are known initially. However this problem has an infinite number of functions $p(\tau,\phi)$, which are able to reproduce the particular echo $f(t)$. This derives from the fact that the reconstruction of $p(\tau,\phi)$ is

analogous to the reconstruction of a signal from its *Window Fourier Transform*. It was clearly stated in [15] that the most general solution of the above (equation (2.5)) is of the form

$$p(\tau,\phi)=\frac{w_g f(\tau,\phi)}{\|g\|^2}+p^{\perp}(\tau,\phi) \qquad (2.6)$$

where $w_g f(\tau,\phi)$ is the Windowed Fourier Transform, $p^{\perp}(\tau,\phi)$ is any function in the orthogonal complement to the range of the Windowed Fourier Transform with respect to g and $\|g\|^2=\int_{-\infty}^{+\infty}|g|^2\,dx$. The Windowed Fourier Transform can be defined as:

$$w_g f(\tau,\phi)=\int_{-\infty}^{+\infty} f(t)\exp(2\pi\iota\phi t)\overline{g(t-\tau)}dt \qquad (2.7)$$

where $\overline{g(t)}$ is the complex conjugate of $g(t)$. The signal $f(t)$ can be reconstructed by the values of $w_g f(\tau,\phi)$ through the following equation

$$f(t)=\frac{1}{\|g\|^2}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} w_g f(\tau,\phi)\exp(-2\pi\iota\phi t)g(t-\tau)dt \quad (2.8)$$

Due to the *Window Fourier Transform* failure of invertibility the most general solution to equation (2.5) is of the form of equation (2.6). Consequently for a fixed signal g this solution yields to be an optimal. However, when the signal transmitted is an arbitrary one, then there is no unique solution for the representation by sending only one of these signals. In this case, this leads to whether one can determine the reflectivity density by propagating a family of signals instead of just one. The previous was answered by Naparst, where he stated that the spreading function can be determined uniquely by sending a particular set of orthogonal functions based on the echoes. Hence by a family of signals transmitted, one will have a set of equations for the corresponding echoes to obtain the representation of the reflectivity density. In other words, a variety of outgoing signals is needed to be used that constitute a 'frame' and yield an orthogonal projection to the signal, which is going to be analyzed in a next section. So the aim is to reconstruct a function of two variables by means of signals of one variable that are reflected from the dense-target environment being observed. However let us first examine what could be a density of targets in practice.

2.2 Distributed targets (target environment)

To be more realistic to the whole problem let us first consider, what a dense-target environment might be? Some examples for possible distributed targets in an environment are given below.

- The blood in the heart or arteries and echo cardiology.

- Fluid flow in a pipe or a wind tunnel where the length of the experiment will determine whether a turbulent environment exists or not.
- A rain cloud.
- The dust debris in the air around airports maybe imaged to determine the presence of dangerous wind currents.
- The surface of a large moving object such as the hull of ship underwater, using sonar.
- Imaging ground clutter, when a missile flies over ground it may realize the ground as a distributed moving target (Synthetic Aperture Radar-SAR)

2.3 Reconstruction of the spreading function

In this section a brief description of the reconstruction of the spreading function $p(\tau,\phi)$, as stated in equation (2.5), is given. Let us assume that for each τ the Fourier transformation of $p(\tau,\phi)$ as a function of ϕ exists and that g_m are the waveforms transmitted. g_m are well-localised, linearly dependent functions that constitute a frame under certain criteria (refer to Appendix-frame definition). These, will be used in order to produce the echoes f_m of the distribution. The index m is meant to show the echoes received to the corresponding waveforms transmitted. So according to the narrow band-model (2.5) the following holds, as detailed in reference [14]:

$$f_m(t) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p(\tau,\phi)\exp(-2\pi i\phi t)g_m(t-\tau)d\tau d\phi,\ m\in N \qquad (2.3.1)$$

The Fourier transformation of $p(\tau,\phi)$ is:

$$P(\tau,t) = \int_{-\infty}^{+\infty} p(\tau,\phi)\exp(-2\pi i\phi t)d\phi \qquad (2.3.2)$$

Consequently the echoes will be

$$f_m(t) = \int_{-\infty}^{+\infty} P(\tau,t)g_m(t-\tau)d\tau,\quad m\in N \qquad (2.3.3)$$

or

$$f_m(t) = \int_{-\infty}^{+\infty} P(t-\mu,t)g_m(\mu)d\mu \equiv \left\langle \overline{P(t-\mu,t)}, g_m(\mu)\right\rangle_{L^2},\quad m\in N \qquad (2.3.4)$$

(The inner product in $L^2(R)$ for $f,g\in L^2(R)$ is $\langle g,f\rangle_{L^2} = \int_{-\infty}^{+\infty}\overline{g(t)}f(t)dt$). By multiplying both sides of the above equation with $\overline{g^m(t-\tau)}$ and adding up the equations, where the set of g^m used, constitute the corresponding reciprocal frame $\{g^m\in L^2(R); m\in N\}$, and are computed as $g^m = \hat{G}^{-1}g_m$, $m\in N$ and $\hat{G}^{-1}$ is a *bounded inverse of a frame condition* $\hat{G}$. The construction of the reciprocal frame is defined in appendix A. Consequently it is obtained as:

$$\sum_{m\in N}\overline{g^m(t-\tau)}f_m(t)=\sum_{m\in N}\overline{g^m(t-\tau)}\int_{-\infty}^{+\infty}P(\tau',t)g_m(t-\tau')d\tau'$$
$$=\sum_{m\in N}\overline{g^m(t-\tau)}\left\langle\overline{P(t-\mu,t)},g_m(\mu)\right\rangle_{L^2}= \qquad (2.3.5)$$
$$=\sum_{m\in N}\overline{g_m(t-\tau)}\left\langle\overline{P(t-\mu,t)},g^m(\mu)\right\rangle_{L^2}=P(\tau,t)$$

Equation (2.3.5) holds providing equality (a.7 in Appendix) exists. Multiplying both sides by the exponential term $\exp(2\pi i\phi t)$ and integrating over t, the spreading function can be obtained by

$$\int_{-\infty}^{+\infty}\sum_{m\in N}\overline{g^m(t-\tau)}f_m(t)\exp(2\pi i\phi t)dt=\int_{-\infty}^{+\infty}P(\tau,t)\exp(2\pi i\phi t)dt=p(\tau,\phi) \qquad (2.3.6)$$

which does not depend on any chosen frame although the dependency of $p(\tau,\phi)$ on the transmitted waveforms. Consequently by transmitting any frame of waveforms one can reach a unique solution of the problem. However in practice it may not be possible to transmit a set of signals, ample enough to approximate the spreading function, but just a few signals that yield an orthogonal projection to it. So by having a finite set of signals, that are orthogonal to the spreading function, a good approximation of the left side of (2.3.6) is yield. This was clearly stated in [14](section 4). Hence, if $P(\tau,t)$ (which is the Fourier transform of $p(\tau,\phi)$) can be obtained as a product of a vector of transform coefficients times the finite set of signals that is transmitted then a unique approximation of the spreading function is obtained. The previous is of the following form

$$P(\tau,t)=\sum_{m=1}^{K}c_m(t)g_m(t-\tau) \qquad (2.3.7)$$

where, $c_m(t)$ are the transform coefficients that depict the dependency of them on the values of t. The transform coefficients in equation (2.3.7) are calculated with the help of certain non-linear adaptive strategies apart from the traditional ones used in the Fourier kingdom. The subset of waveforms transmitted, is a combination of M-terms selected from a redundant family of functions N ($M\subseteq N$) that constitutes an orthogonal bases. This subset is selected through strategies, which are going to be analyzed briefly in a next chapter, with respect to the projection of the signal to the basis. In order to obtain the approximation of the target environment one should minimize the distance between the approximation and the actual target by computing the coefficients $c_m(t)$. These are calculated by the inner product $c_m(t)=<g_m(t-\tau),P(\tau,t)>$. This family of waveforms is chosen priori to characteristics of the environment and each function of g_m, selected, is well-localized which comes from a family of orthonormal basis $B=\{g_m\}_{m\in N}$. Consequently this approximation clearly depends

on the properties of the target environment relative to that basis. *However what could happen if the target environment wished to be represented is not known for certain? Would this subset of waveforms chosen to be transmitted prove adequate to approximate it?* But Let us first introduce a vocabulary notation, over dictionaries and splines, before the description of these strategies and their application when using prior knowledge over the target environment.

Chapter 3

Spline Dictionaries

3.1 Dictionaries and atoms

There are several methods proposed for obtaining signal representations apart from the traditional ones [2] [16]. These range from general approaches, like the Method Of Frames[14][17] and the Matching Pursuit [4] [7], more efficient schemes like the method of Best Orthogonal Basis [18] which will be described briefly thereafter. Most of these methods mentioned above are used in, generally, signal and image processing such as speech recognition, range-Doppler radars, compression, face recognition and identification, image compression, signal restoration and recovery etc.

Like mentioned earlier the task in the radar problem is to represent a function of one or two variables by using elementary functions out of a redundant family of functions of one variable. Their selection is performed through the above iterative approaches to best "match" this function observed. However before the analysis of the above existing methods let us establish a vocabulary notation. Adopting the terminology earlier introduced [4], emphasis on terms such as *overcomplete representations*, *dictionaries* and *atoms* is going to be made. By the former let us assume that f is a discrete signal of length n, with $(f_t : 0 \le t < n)$, which can also be viewed as a vector, or matrix in the 2D case, in R^n. We are interested in the reconstruction of this signal using superpositions of elementary waveforms as discussed earlier. Traditional methods [4][7][14][19] of analysis and reconstruction involve the use of orthogonal bases(Fourier basis), various discrete cosine transform bases and orthogonal wavelet ones. So given a list of n waveforms one wishes to represent f as a combination of these waveforms. All this list of waveforms, which are viewed as vectors, are linearly independent and consequently the representation of f is unique.

Concerning the latter, adapting the terminology introduced by Mallat and Zhang [7], let H be a Hilbert space, Γ a set of indices ($\Gamma \subseteq \mathrm{H}$) and $D=(a_n : n \in \Gamma)$ a family of functions in H, where each of them are normalized to unity ($\|a_n\|=1$). Since in practice we deal with a finite number of these functions it is assumed that Γ is a finite set of N indices where n=1,2,…,N. The set of D is a collection of parameterized waveforms called a **dictionary**. Each of the functions $\{a_n\}_{\in\Gamma}$ constituting D are discrete-time, linear independent fixed elements that are normalized to unity, which are called **atoms**. These are created by simple translations of a basis function with respect to a sequence of knots and are defined by equation (3.2.2). Concerning a dictionary, n parameter is interpreted as of indexing frequency (Fourier dictionary), time/scale jointly, (time-scale dictionary-Haar), or time/frequency jointly (time-frequency dictionary-Gabor) [Coifman et al.[18]]. We denote S a linear span of the dictionary functions a_n. Dictionaries are *complete* or *overcomplete*. In the first case S=H for $\lim_{N\to\infty}$. However, a dictionary is in general redundant which implies that in the finite case the dimension of S is less than N and that in the case of $\lim_{N\to\infty}$ a complete dictionary is actually overcomplete. A complete dictionary is not necessarily a basis because its elements need not be linearly independent. Also one could have continuum dictionaries containing infinite atoms, and undercomplete dictionaries for special purposes, containing fewer than n atoms.

According to the atoms that D consists of, the dictionary can be described as *orthogonal* or *non-orthogonal*. By the former an orthonormal set of atoms forms an orthogonal dictionary, whereas non-orthogonal dictionaries can be constructed from all sorts of waveforms.

Numerous interesting dictionaries have been proposed over the last few years. However this thesis is focused on *Cardinal B-spline* dictionaries as already analyzed in [20]. These are going to be used thoroughly in order to obtain the representation phase with the help of adaptive pursuit approaches. Before we proceed with Cardinal B-splines let us first introduce some relevant definitions on splines.

3.2 Splines

A spline is a piecewise polynomial satisfying continuity conditions between the *pieces*. Since polynomials are continuous in all derivatives, it must be the joints between the *pieces* in a spline where continuity is a question. These points are called *knots*. The support of a spline function f denotes that it is zero outside a compact interval ($\sup port(f)=closure\{x: f(x)\neq 0\}$). In the particular section some basic definitions are stated concerning splines on a compact interval as detailed in [20][21].

Definition 1:

Given a compact interval [c,d], a partition [c,d] is a finite set of points

$$\Delta = \{x_i\}_{i=0}^{N}, N \in \mathbf{N} \text{ such that } c = x_0 < x_1 < ... < x_{N+1} = d$$

We also define N subintervals I_i, $i = 0,...,N$ as $I_i = [x_i, x_{i+1})$, $i = 0,...,N-1$ and $I_N = [x_N, x_{N+1})$.

Definition 2:

If Π_m the space of the polynomials of order (or degree) $\leq m \in \mathbf{N_0}=\mathbf{N} \cup \{0\}$. For m a positive integer then

$$S_m(\Delta) = \{f \in C^{m-2}[c,d]; f\big|_{I_i} \in \Pi_{m-1}, i = 0,...,N$$

where $f\big|_{I_i}$ denotes the restriction of function f on the interval I_i. $S_m(\Delta)$ is the *space* of polynomial splines of order m with simple, non decreasing sequence of control points $x_1,...,x_N$, known as knots. If the knots are equidistantly distributed in the interval $[c,d]$ then this is referred as a *uniform* spline, otherwise it is *non-uniform.*

It has been shown in [21] (theorem4.4) that $S_m(\Delta)$ is a linear space of dimension $m+N$. Furthermore if Δ and Δ' are two partitions of the interval $[c,d]$ such that $\Delta \subseteq \Delta'$ then $S_m(\Delta) \subseteq S_m(\Delta'), m \in \mathbf{N}$. In order to construct a basis for $S_m(\Delta)$ one should introduce the *extended partition.*

Definition 3:

For Δ being a partition of $[c,d]$ let us consider

$$y_1 \leq y_2 \leq ... \leq y_{2m+N} \text{ such that}$$

$y_1 \leq y_2 \leq ... \leq y_m \leq c$, $d \leq y_{m+N+1} \leq ... \leq y_{2m+N}$ and $y_{m+1} < ... \leq y_{m+N} = \overbrace{x_1,..x_1}^{m_1},...,\overbrace{x_N,..x_N}^{m_N}$.

We define $\tilde{\Delta} = \{y_i\}_{i=1}^{2m+N}$ an *extended partition* with simple inner knots that are associated with $S_m(\Delta)$.These points in the extended partition are uniquely determined, however the first and the last one can be chosen arbitrarily subject to $y_1 \leq y_2 \leq ... \leq y_m \leq c$, $d \leq y_{m+N+1} \leq ... \leq y_{2m+N}$. Thus with each fixed extended partition there is a unique B-spline basis denoted as

$\{B_i\}_{i=1}^{m+N}$ (details on the construction of a basis can be found in [21] (theorem-4.9)). The basis functions B_i are equal to zero if $x \notin [y_i, y_{i+m}]$ otherwise $B_i(x) > 0$ for any $x \in [c,d]$. Moreover $\sum_{i=1}^{m+N} B_i(x) = 1$ for all $x \in [c,d]$. If the knots are equally spaced then the splines are called *cardinal*, which is going to be explained further.

3.3 Cardinal B-splines

Like already mentioned in the previous section we are going to concentrate on dictionaries of Cardinal B-splines [20]. It is widely known that splines have been used with success in wavelet theory and applications of signal processing. Cardinal spline spaces on a bounded interval are finite dimensional linear spaces. All B-splines with equally spaced knots are called Cardinal B-Splines. The set of knots $..., y_i, y_{i+1}, ...$ is equally spaced provided that $y_{i+1} - y_i = b$ for all i. All Cardinal B-Splines of order m can be obtained from one basic B-Spline $B(x)$ (3.2.1) ,by translation and scaling, with respect to a simple knot sequence $0,1,2,..,m$.as detailed in [21]

$$B(x) = \frac{1}{m!} \sum_{i=0}^{m} (-1)^i \binom{m}{i} (x-i)_+^{m-1} \qquad (3.2.1)$$

where $(x-i)_+^{m-1}$ is equal to $(x-i)^{m-1}$ if $x - i > 0$ and 0 otherwise.

Consequently for the equally spaced sequence of points $y_i, ..., y_{i+m}$ we will have the following relation

$$B_i(x) = \frac{1}{b} B(\frac{x - y_i}{b}) \qquad (3.2.2)$$

where b is the distance between two adjacent knots. A dictionary is formed by single translations of the prototype function (3.2.1) introduced earlier. For example in [fig.1] we observe a Cardinal B-splines dictionary. It comprises of 47 translated atoms based on a prototype function. The order of splines is four, the number of discrete point in the interval is 2049 and the length of support equal to two. Notice the atoms that are partially contained in the interval. These are called *boundary* functions.

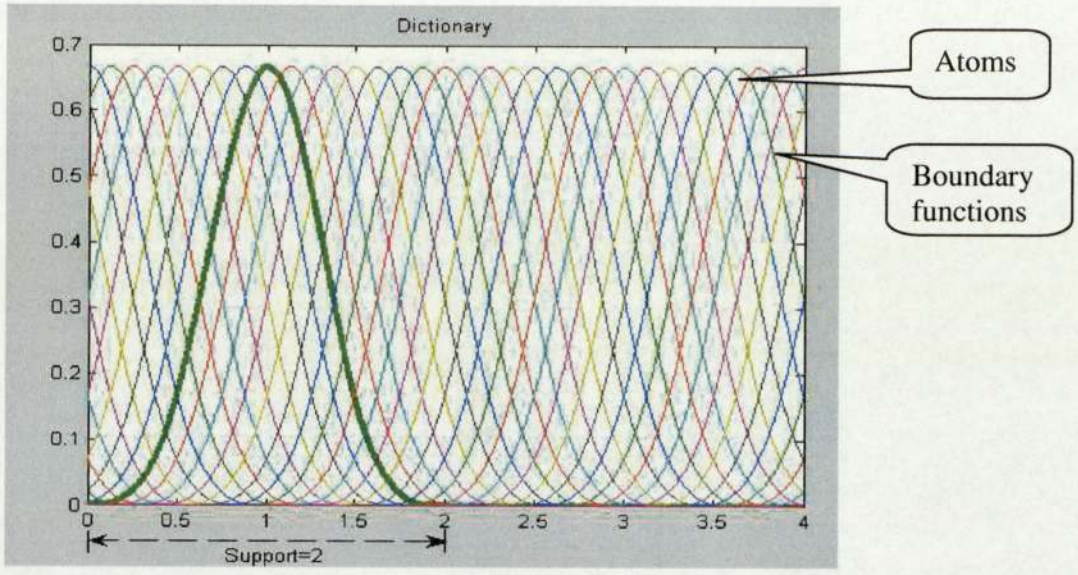


Figure 1: Dictionary of Cardinal B-Splines of length of support of two and order four.

What is essential in signal representations and reconstructions is fixing the space that one is going to work with. In the experiments performed later the above was managed through increasing the compact support, which is going to be analyzed thoroughly in another chapter. Generally there are two different ways to increase the subspace's dimension by maintaining a fixed order of the splines. The first one is by increasing the number of knots, say decreasing the distance b between two adjustment knots. And the other way is by increasing the length of the support of each basis function of the cardinal spline space, in other ways the width mb. In this thesis the latter is considered in order to fix the basis function that is going to be used in the later experiments.

Chapter 4

Adaptive techniques for signal representation

4.1 Introduction

It was stated in chapter two, that in practice it may not be possible to transmit a number of signals, ample enough to approximate the spreading function (2.3.6). It was stated in reference [14] that one can obtain a good approximation by sending just a finite number of signals that yield an orthogonal projection to the target environment. So if $P(\tau,t)$ can be obtained from (2.3.7), one can have a unique approximation for just a finite number of signals.

In this section the problem of signal representation, often referred as atomic decomposition, is discussed. Moreover, some techniques for adaptive signal representation that are going to be used for the experiments and a numerical example on these algorithms are presented.

4.2 The real problem

Let us assume that a function f in a Hilbert space is given. Also a basis of N-terms (dictionary) is also available. So, if the inner product <.,.> and its norm $\| f \| = < f, f >^{1/2}$ exists, then the *task is to approximate it by a M-term ($M \subseteq N$) linear superposition* of the following form

$$f^M = \sum_{n=1}^{M} c_n a_n \qquad (4.2.1)$$

where $\{a_n\}_{n=1}^{M}$ are fixed elements of a basis (*dictionary*) and are called *atoms,* which are well-localised. In order to obtain the approximation f^M of f one should minimize the distance

$\| f - f^M \|$ by computing the coefficients c_n. These are calculated depending on features of these atoms. In essence, if $V_k = span\{a_1, a_2, ..., a_n\}$ and the spanning set is an orthonormal basis for V_k then $c_n =< a_n, f >$. On the other hand, if the spanning set is a basis for the subspace but not necessarily orthogonal then the coefficients are calculated by the inner product $c_n =< \tilde{a}_n, f >$. where the dual sequence $\{\tilde{a}_n^M\}_{n=1}^M$ is biorthogonal to $\{a_n\}_{n=1}^M$. The biorthogonal atoms are computed through a recursive procedure illustrated in [22]. The superscript means that $span\{a_n\}_{n=1}^M \equiv span\{\tilde{a}_n^M\}_{n=1}^M$. Hence, in order to include one term in (4.2.1) the elements of the dual sequence need to be modified for the coefficients of the new approximation to minimize the above distance.

The selection of the adequate atoms to be used for the approximation is obtained by using non-linear greedy algorithms for sparse approximation, such as the ones that are going to be analyzed further in this chapter. In addition the computational complexity depends on the dictionaries used which impose drastically different computational burdens-not to mention the nominal cost of storing and applying. The previous is needed to be considered in future work by means of fast adaptive processing. However we are going to focus, generally, on the effectiveness and robustness of these methods in the representation phase.

4.3 Existing Decomposition Methods

Like already mentioned above there are several approaches to obtaining an optimum selection of atoms that will yield the minimum error between the approximation and the signal *f* wished to be represented. These will be briefly discussed in this section and are going to be used extensively in the experiments following in another chapter.

4.3.1 Matching pursuit methods

The literature review for these methods is vast and these are implied in a forward or a backward manner.

In general terms each of the methods requires two inputs to be initialized and produce one output. Their inputs are a family of functions and the signal that is going to be approximated. The dictionary includes translations of a single function, like Cardinal B-splines or a combination of different forms of atoms. Then these two inputs activate the algorithm of reconstruction by making successive approximations to the initial signal. In other words, the arbitrary signal is decomposed to an expansion of elementary waveforms chosen from a

dictionary. In order to approximate the initial signal, atoms are chosen from the dictionary, which minimize the approximation error. The algorithms stop as soon a criterion is reached, which is going to be described thoroughly in another section. Its block diagram is depicted below.

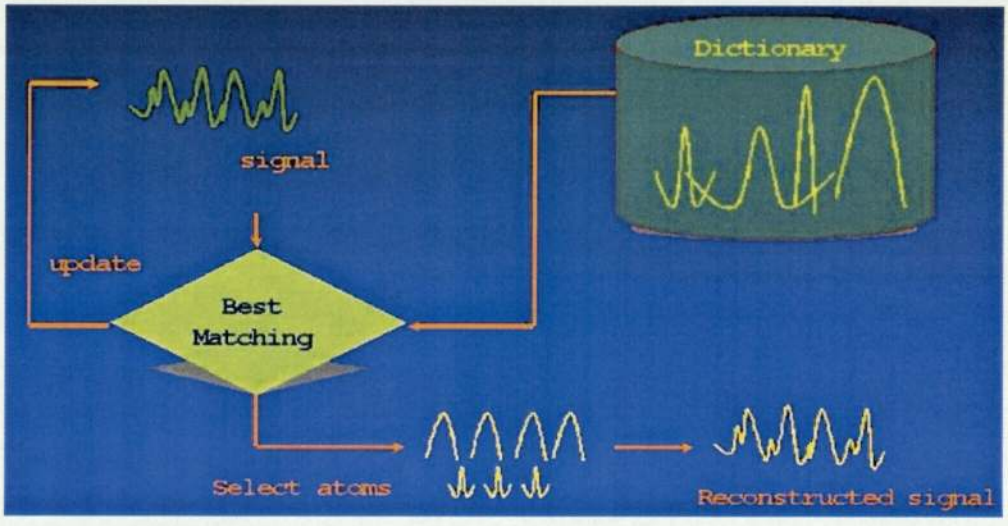


Figure 2: Block diagram of Best Matching methods.

A general approach to approximate the decomposition was introduced by Mallat and Zhang[7] and addresses the sparsity issue directly.

Let H be a Hilbert space, Γ a set of indices and $D=\{a_n;n\in\Gamma\}$ a family of functions in H each of them normalized to unity. Since in practice one deals with a finite number of such functions lets assume that Γ is a finite set of N indices (n=1,2..,N) .Let us also assume that there exists a linear span of the dictionary D denoted as S. Now given a signal $f\in$ H the aim is to represent it as a linear combination of atoms selected from D.

The **Matching pursuit** approach (MP) proposes to make selection by successive approximations of f. Vectors are selected through a dictionary D one by one, while the signal approximation is optimized at each step. So if the dictionary comprises from functions a_n the approach begins by projecting f on each vector and computes the residue R. If R_k is the kth order residue and l_k the index n for which the corresponding dictionary atom a_{l_k} yields a maximal value of $|\langle a_n, R_k\rangle|$, where <.,.> indicates the inner product in H, then the algorithm will evolve as follows. Starting with an initial approximation of $f_1=0$ and residue $R_1=f$ the kth order residue will be:

$$R_k=\langle a_n, R_k\rangle a_n + R_{k+1} \qquad (4.3.2.1)$$

where the term R_{k+1} is the residue vector after approximating f in the direction of a_n. Clearly the residue vector is orthogonal to a_n (meaning that $\langle R_{k+1}, a_n\rangle=0$ for n=1,2,…,k+1), so

$$\|R_k\|^2 = |\langle a_n, R_k \rangle|^2 + \|R_{k+1}\|^2 \qquad (4.3.2.2)$$

Hence in order to minimize $\|R_{k+1}\|^2$, a_{l_k} must be chosen such that $|\langle R_k, a_n \rangle|$ is maximum. So from the equation of the kth order residue it follows that at iteration k the MP approach results in the representation of the form:

$$f = f_k + R_{k+1} \qquad (4.3.2.3)$$

with

$f_k = \sum_{n=1}^{k} \langle a_{l_n}, R_n \rangle a_{l_n}$. An intrinsic feature of this algorithm -and the following ones- is that when stopped after a few steps k it yields an approximation f_k of the signal f using only a few atoms (elementary signals from a dictionary) with an error equal to R_{k+1}. Consequently if k tends to infinity then the error converges to zero.

However this technique does not yield at each iteration the linear span of the selected atoms that approximates the signal in its minimum sense. In other words, while the MP approach chooses an atom in order to minimize the residue error then by subtracting the projection of the residual over that atom the algorithm produces new components. These components are not in general orthogonal since atoms chosen from the dictionary may not necessarily be orthogonal to the subspace created by these atoms. The previous is avoided by projecting the residues on an orthogonal family $\hat{P}_{V_k}$ derived from V_k subspace consisting of atoms $\{a_{l_n}; n = 1,..., k\}$. $\hat{P}_{V_k}$ is assumed to be an orthogonal projector on V_k. This refinement of the algorithm that improves the convergence rates since it can converge in only a finite number of steps is called **Orthogonal Matching Pursuit** (OMP) [6]. Let g be an arbitrary function in V_k, $\hat{P}_{V_k}$ is self-adjoin and idempotent, which means that $\langle \hat{P}_{V_k} f, g \rangle = \langle f, \hat{P}_{V_k} g \rangle$, for $k \geq 1$ and all $f,g \in H$ and $\hat{P}_{V_k} \hat{P}_{V_k} = \hat{P}_{V_k}$, respectively. The closest function to f that can be written as a linear expansion of k-atoms from the dictionary is $\hat{P}_{V_k} f$. If the distance is calculated and we write $g \equiv g - \hat{P}_{V_k} f + \hat{P}_{V_k} f$ where $(f - \hat{P}_{V_k} f) \in V_k^{\perp}$ (the orthogonal complement of V_k in H) then:

$$\|f - g\|^2 = \|f - g - \hat{P}_{V_k} f + \hat{P}_{V_k} f\|^2 = \|f - \hat{P}_{V_k} f\|^2 + \|\hat{P}_{V_k} f - g\|^2$$

So the distance is minimized for $g \equiv \hat{P}_{V_k} f$. On the other hand, if the subset of atoms chosen $\{a_{l_n}; n = 1,..., k\}$ is orthogonal then f_k given in (4.3.2.3) is not the orthogonal projection of f on V_k and the approximation obtained is not the closest obtained in the particular subspace.

However at each iteration OMP keeps selecting the dictionary atoms as prescribed by the MP approach which is not an optimal one concerning the residue. Following the above Rebollo et al. provided a new method called **Optimized Orthogonal Matching Pursuit** (OOMP) that guaranties minimization of the residue. Adapting the notation in [8] at iteration k+1 the OOMP approach prescribes an approximation that is both the orthogonal projection of f onto the subspace generated by these atoms and minimizes the residue.

If f_k is the orthogonal projection of f in subspace V_k which yields the best possible approximation then through the iteration of the algorithm the signal will decompose in

$$f = \sum_{n=1}^{k} c_n^{(k)} a_n + \tilde{R}_k \qquad (4.3.2.4)$$

In order to minimize the residue, we should have $\sum_{n=1}^{k} c_n^{(k)} a_n = \hat{P}_{V_k} f$. Coefficients c are modified through iterations by means of an adaptive biorthogonalisation technique [8] [22]. Lets say that out of dictionary (D), OOMP chooses an atom a_{l_1}, where $V_1 = a_{l_1}$ and $V_{k+1} = V_k \oplus a_{l_{k+1}}$ ($\oplus$ denoting the projection of $a_{l_{k+1}}$ to subspace V_k). If W_{k+1} the orthogonal compliment of V_k in V_{k+1} then the orthogonal projection onto V_{k+1} is $\hat{P}_{V_{k+1}} = \hat{P}_{V_k} + \hat{P}_{W_{k+1}}$ and the orthogonal projection of $a_{l_{k+1}}$ onto W_{k+1} is Ψ_{k+1} such that

$$\Psi_{k+1} = \hat{P}_{W_{k+1}} a_{l_{k+1}} = \hat{P}_{V_{k+1}} a_{l_{k+1}} - \hat{P}_{V_k} a_{l_{k+1}} = a_{l_{k+1}} - \hat{P}_{V_k} a_{l_{k+1}} \qquad (4.3.2.5)$$

W_{k+1} is spanned by a single function Ψ_{k+1} of zero norm and the normalized to unity function is

$$\tilde{\Psi}_{k+1} = \frac{\Psi_{k+1}}{\|\Psi_{k+1}\|} \qquad (4.3.2.6)$$

The orthogonal projection onto W_{k+1} is $\hat{P}_{W_{k+1}} f = \tilde{\Psi}_{k+1} \langle \tilde{\Psi}_{k+1}, f \rangle$. In order to obtain the coefficients c, an adaptive biorthogonalisation technique is used. Like mentioned earlier those have the following properties:

- $\beta_{l_n}^{(k+1)} = \beta_{l_n}^{(k)} - \frac{\tilde{\Psi}_{k+1}}{\|\Psi_{k+1}\|} \langle a_{l_{k+1}}, \beta_{l_n}^{(k)} \rangle$ *and* $\beta_{l_{k+1}}^{(k+1)} = \frac{\tilde{\Psi}_{k+1}}{\|\Psi_{k+1}\|}$
- also they are biorthogonal to the functions $\{a_{l_n}; n = 1,..., k+1\}$ with $\langle \beta_{l_n}^{(k+1)}, a_{l_m} \rangle = \delta_{l_n, l_m}$ where $n = 1,..., k+1$ and $m = 1,..., k+1$

- the representation will be $\hat{P}_{V_{k+1}} f = \sum_{n}^{k+1} \left\langle \beta_{l_n}^{(k+1)}, f \right\rangle a_{l_n}$ (the proof is detailed in [8])

The coefficients β are used, seen above, in order to recalculate $f = \sum_{n=1}^{k} c_n^{(k)} a_n + \tilde{R}_k$ and we will have the following recursive equations

$$c_{l_n}^{(k+1)} = c_{l_n}^{(k)} - \left\langle \beta_{l_n}^{(k)}, a_{l_{k+1}} \right\rangle \left\langle \frac{\tilde{\Psi}_{k+1}}{\|\Psi_{k+1}\|}, f \right\rangle \text{ and } c_{l_{k+1}}^{(k+1)} = \left\langle \frac{\tilde{\Psi}_{k+1}}{\|\Psi_{k+1}\|}, f \right\rangle \qquad (4.3.2.7)$$

with $c_1^{(1)} = \left\langle a_{l_1}, f \right\rangle$.

The selection is made by choosing the atom $a_{l_{k+1}}$ that at (k+1)th iteration minimizes the norm of the residue and is the one that yields a maximal value for the functional below (proof in [8]):

$$e_n = \frac{b_n}{d_n} = \frac{\left|\left\langle a_n, \tilde{R}_k \right\rangle\right|^2}{1 - \left\langle a_n, \hat{P}_{V_k} a_n \right\rangle} = \frac{\left|\left\langle \Psi_n, f \right\rangle\right|^2}{\|\Psi_n\|^2} \qquad (4.3.2.8)$$

the algorithm is stopped when $\left\|\tilde{R}_k\right\|^2 \le \delta$, where δ is the desired precision.

Last but not least there is a backward approach of the matching pursuit called **Backward-Optimised Orthogonal Pursuit** (Boomp) [10] which provides a criterion to choose the atom to be deleted in order to leave an approximation minimizing the norm of the residual error. In its implication one should assume that the atomic decomposition is given, from a forward method (OMP or OOMP), and wishes to eliminate some coefficients. So let us assume that after the selection of k atoms the forward approach (OOMP) provides a representation of

$$f^k = \hat{P}_{V_k} f = \sum_{n=1}^{k} c_n^{(k)} a_{l_n} \qquad (4.3.2.9)$$

where V_k denotes the subspace $\{a_{l_n}; n = 1,..., k\}$ chosen from D and $\hat{P}_{V_k}$ is the orthogonal projector of f on V_k (signal $f \in H$). The coefficients $c_n^{(k)}$ are the inner product of $\left\langle \beta_n^{(k)}, f \right\rangle$, where $\{\beta_n^{(k)}(n = 1,..., k)\}$ are the biorthogonal functions and are computed recursively as detailed in [10].

In addition let us assume that one wishes to eliminate an atom indexed with j out of the previous spanning set V_k obtained. The new subspace will be of the form $V_{k/a_j} = span\{a_{l_1},..., a_{l_{j-1}}, a_{l_{j+1}},..., a_{l_k}\}$ and the resulting approximation the following

$$f^{k/j} = \hat{P}_{V_{k/a_j}} f = \sum_{n=1}^{k} a_n \left\langle \beta_n^{(k/j)}, f \right\rangle = \sum_{n=1}^{k} c_n^{(k/j)} a_{l_n} \quad (4.3.2.10)$$

The coefficients $c_n^{(k/j)}$ to be eliminated are not selected randomly. However, there is a certain criterion to be fulfilled for their exclusion. In other words the one to be removed is the one that minimizes the norm of the residual error $\Delta = f^k - f^{k/j}$. It has been detailed in [10] that the norm of the residual error is minimized for a minimum value of the quantity $\frac{\left|c_j^{(k)}\right|^2}{\left\|\beta_j^{(k)}\right\|^2}$. Consequently the remaining subspace will be V_{k/a_j} and the new approximation of the form (4.3.2.10). Lastly the new coefficients are recursively modified by a backward adaptive biorthogonalisation technique, described in the same work. These are of the following form for n=1,2,…,j-1,j+1,…k :

$$c_n^{(k/j)} = c_n^{(k)} - \frac{\left\langle \beta_n^{(k)}, \beta_j^{(k)} \right\rangle}{\left\|\beta_j^{(k)}\right\|^2} c_j^{(k)} \text{ and } \beta_n^{(k/j)} = \beta_n^{(k)} - \frac{\beta_j^{(k)} \left\langle \beta_j^{(k)}, \beta_n^{(k)} \right\rangle}{\left\|\beta_j^{(k)}\right\|^2} \quad (4.3.2.11)$$

So by fixing the above equations and the criterion in order to omit an atom the algorithm is iterated as soon as the tolerance error initially put is reached.

Finally there is the **Swapping-based** refinement of orthogonal matching pursuit strategies [9] which implies a backward step to eliminate one atom from the atomic decomposition, according the Boomp criterion, and a forward step to replace such an atom by another one according the OMP or OOMP criterion.

4.4 Stopping criteria

Since all the algorithms used for the reconstruction are iterative procedures, there a two criteria for deciding when an algorithm should be halted iterating.

- After a fixed number of cycles
- Reach a desired precision

By the first one the method can be instructed, from the beginning, to stop after a fixed number of iterations. In other words, fix the size of the subset of atoms that is going to be selected in order to obtain the approximation of the signal. Concerning the later there is also the possibility, while initialization of the method, to order the representation phase to stop as soon as the residue error returned declines to a particular level ε. The residue error is an L_2 - norm or Euclidean norm and is the following:

$$Error = \frac{1}{N}\sqrt{\sum_{i=0}^{N}(f_i - f_i^{'})^2} \quad (4.4.1)$$

where f (f_i are elements of vector f) is the target environment wished to be represented and $f^{'}$ ($f_i^{'}$ are elements of vector f') is the approximation obtained by the decomposition. Both of them are vectors of N rows.

4.5 Numerical example on BOOMP algorithm

In order to understand further how the above algorithms perform let us illustrate it in a simple example of approximation. For the purpose of the experiment the *Backward Optimized Orthogonal Matching Pursuit* is going to be used.

Let us suppose that we wish to reconstruct a particular signal f, depicted in [fig.4]. Like discussed earlier, in order to apply the backward approach it needs to be considered that a reconstruction is already obtained by using a forward method, such as OMP or OOMP. Let us assume that in the decomposition the Optimized Orthogonal Matching Pursuit is used and only thirteen atoms of the dictionary D are chosen to approximate the initial target. The dictionary was created by the translation of Cardinal B-Splines with support functions of length two and order of four. This comprises of 47 functions depicted in [fig.3]. So let us assume that three particular coefficients are eliminated out of the decomposition in order to obtain the approximation of f with BOOMP. The reconstruction method (BOOMP) is initialized by applying both the signal f and the thirteen atoms that the forward approach used to reconstruct the signal with error 0.207149 (Table 1).

Like stated previously these three atoms are not eliminated at random, however the backward approach eliminates them with respect to a particular prescription mentioned earlier. [fig.6] depicts the decomposition obtained by using firstly the forward method and below [fig.8] the approximation from BOOMP with error 0.295606. The thirteen atoms that OOMP selects to reconstruct the signal and the ten ones that BOOMP selects are depicted in [fig.5-7].

Now, continuing further and trying to use these ten atoms to reconstruct the signal f with OOMP an approximation is obtained with an error of 0.324695.From the above it is worth saying that not only is there shrinking of coefficients or further compression of the approximation, but also the residue error is minimized, like stated previously. [fig.9] shows the block diagram of the whole procedure. Table 1 coincides with results of the particular experiment. The first column from the left depicts the method used for the decomposition. The middle one the size of the subset of atoms chosen for the approximation of the initial signal and the third one corresponds to the residue error returned.

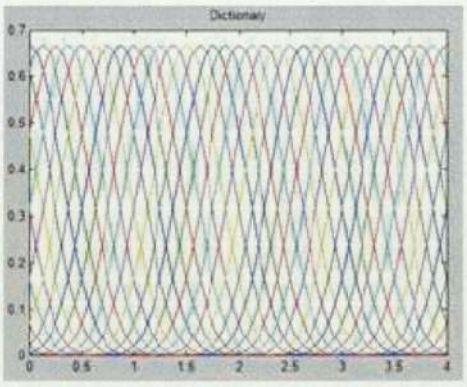


Figure 3: Dictionary of Cardinal B-Splines of support two and order four.

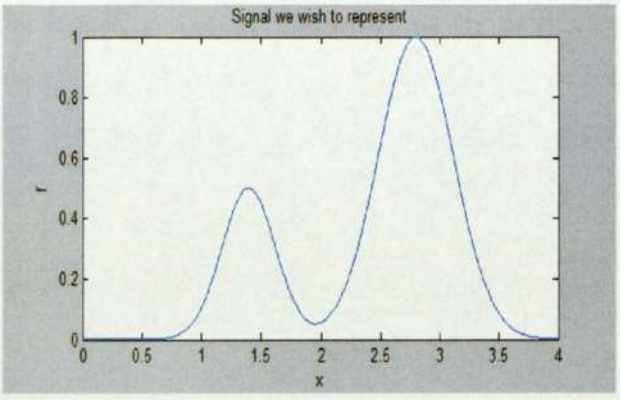


Figure 4:Signal f that we wish to represent.

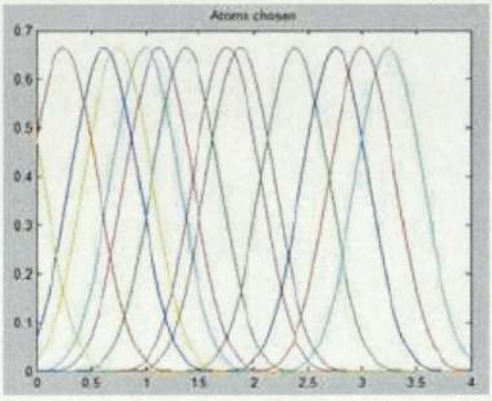


Figure 5:The subset of atoms that OOMP selected after the decomposition.

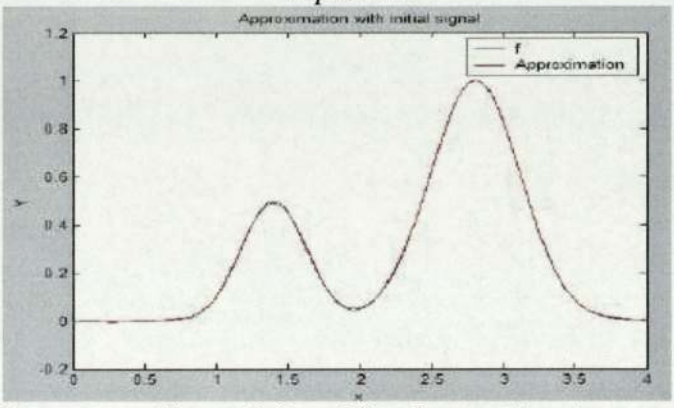


Figure 6:Signal f and approximation obtained by OOMP.

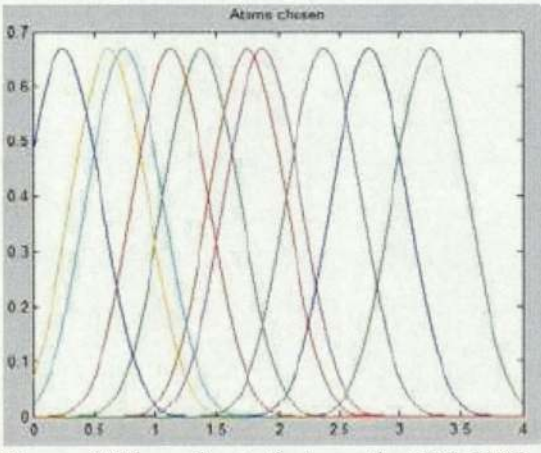


Figure 7:The subset of atoms that BOOMP selected after the decomposition.

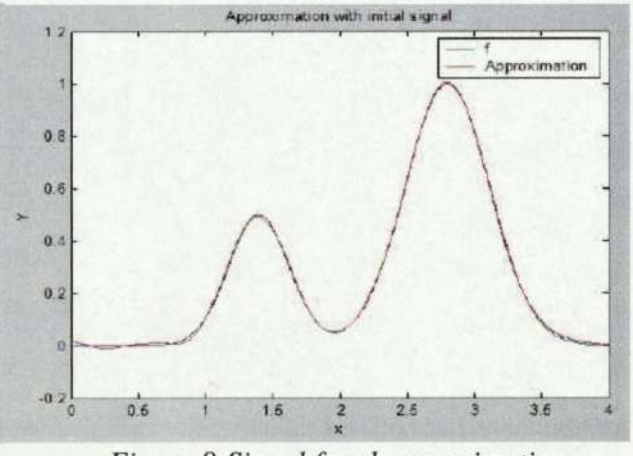


Figure 8:Signal f and approximation obtained by BOOMP.

Strategies	Number of Atoms Chosen	Norm of the Error
OOMP	13	0.207149
BOOMP	10	0.295606
OOMP'	10	0.324695

Table 1: Concentrated table of the size of the subset (number of atoms) that the method selects and the norm of the error returned, respectively.

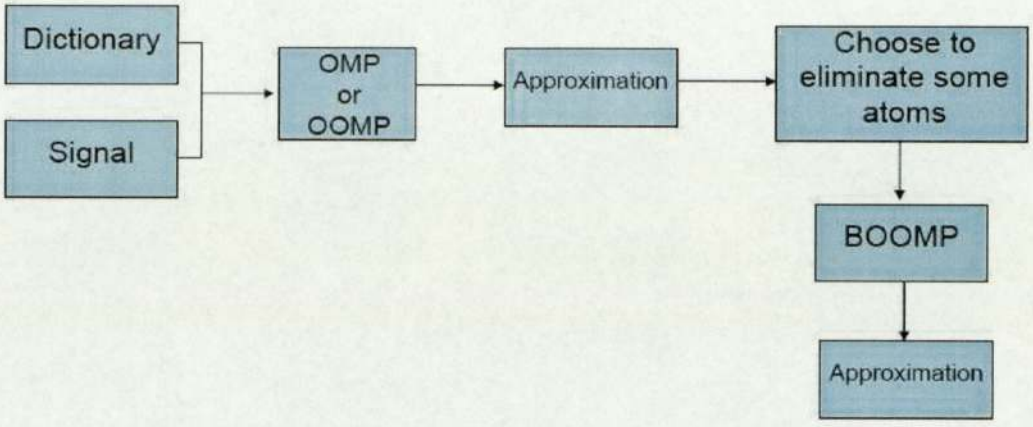


Figure 9: Block diagram of the whole reconstruction phase.

ANALYSIS

Chapter 5

Simulations of the representation using Orthogonal Matching Pursuit method.

5.1 Introduction

The literature for the above methods of reconstruction is certainly vast. One could carry on analyzing further interesting methods which apply with great success to the representation of signals. However the aim of that project is not to compare each of them, which is already performed in previous interesting works. In that chapter some typical features concerning the performance of OMP introduced in the section (4.3.2) is going to be illustrated. The calculations are going to be carried out for functions of two variables. The aim is to stress out the ways one can obtain an approximation of a function of two variables using OMP and remarks concerning the computational time required. Moreover it is shown that when the target environment is known in total and the basis functions, used, are chosen priori to features of the target environment, then it is possible to represent it by just using a subset of these functions of one variable (selected by a dictionary). This has shown to yield a sparse approximation with a compromise of error.

5.2 Implementation in a target environment of two variables

From the previous chapter, it is an easy task to represent a function of one variable with the above methods. However the aim is to represent a function of two variables with a linear combination of elementary functions of one variable (these are selected from a dictionary by

means of the previous approaches already described). In order to perform such a task it was essential to manipulate each method's routings used. There are two ways of performing the approximation by using OMP for two dimensional target environments, which can be easily applied to all the methods described earlier. The later is analyzed in this section.

The first one is by performing *slices* in the environment and iterating, for these, the representation phase and the second one is done by *introducing the target environment as an input in total*. Numerical calculations are performed with OMP. Let us assume that the target environment is known initially and consists of seven peaks that are fairly separated, these have the form below and are depicted in [fig.11].

$$f(x,y) = \sum_{i=1}^{K} a_i e^{\left[-\left(\frac{\left(x-mean_x^i\right)^2+\left(y-mean_y^i\right)^2}{\left(b_i\sigma_i\right)^2}\right)\right]} \qquad (5.2.1)$$

where x, y are matrices (N by M), K=7, $mean_x^i$ and $mean_y^i$ denote the location, $b_i\sigma_i$ is the width and a_i the amplitude of each peak, respectively.

In all, the norm of the error between the approximation obtained and the function wished to be represented is the following:

$$Error = \frac{1}{NM}\sqrt{\sum_{j=1}^{M}\sum_{i=1}^{N}(f_{ij} - f_{ij}^{'})^2} \qquad (5.2.2)$$

where f is the input target environment and $f^{'}$ is the approximation obtained. These are matrices with N rows and M columns. In general terms $f^{'}$ is a product of a matrix $D(:,1:k)$ (subset of the dictionary D), k the indexes of atoms that each method chooses throughout the whole dictionary D to approximate f, and a matrix of transform coefficients c that are calculated, like stated in an earlier chapter. The value of the error is subjective since what we are interested in, is to have a sparse representation that is bearable in proportion to the input target. For example producing the number of peaks in total or their localization individually is a good approximation.

5.2.1 Slicing the environment

Let us assume that the target environment $\boldsymbol{f}$ is a matrix $\boldsymbol{f} = [f_{ij}]$ (f_{ij} are elements of $\boldsymbol{f}$ matrix where i=1,…,N and j=1,…,M) with N, M number of rows and columns, respectively. The elements of each column represent a different signal represented in a 2D space. Each column of matrix $\boldsymbol{f}$, that is a vector of a set of values, will be referred as a *slice* of the target environment. Consequently if the number of *slices* is M then from the representation of each of

these, for instance through OMP, one can form the approximation f ' of the whole matrix f. The block diagram of the representation phase of f is depicted below.

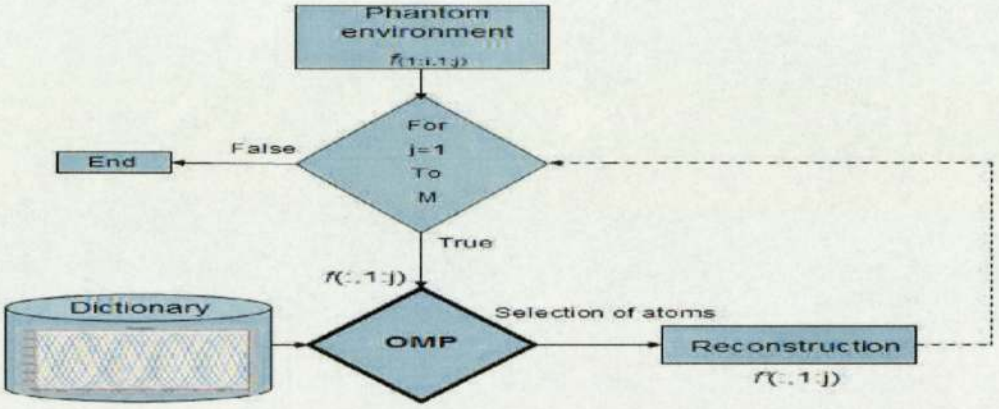


Figure 10: Block diagram of the reconstruction phase with slices.

So at the M^{th} iteration every vector of the input matrix f is reconstructed with the selection of a subset of, let us say, thirteen functions out of the dictionary. Hence [fig.12] shows an approximation of the whole matrix obtained with the total mean error of 1.1258e-006. It can be also mentioned that the slices for the particular environment are done in parallel with X-axis and M=41 for the particular case. Moreover the dictionary used for the approximation of each of the slices, is one of 47 functions with order of B-Spline functions=4 and length of the support of two.

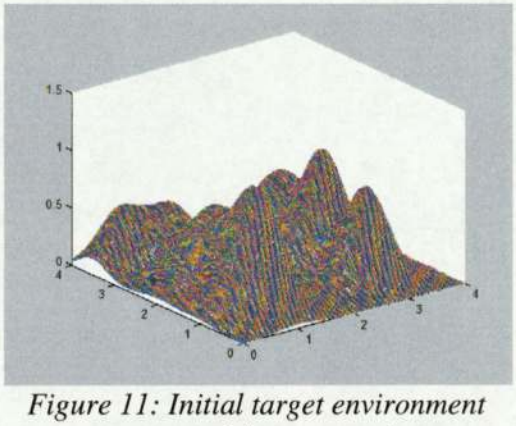


Figure 11: Initial target environment we wish to approximate.

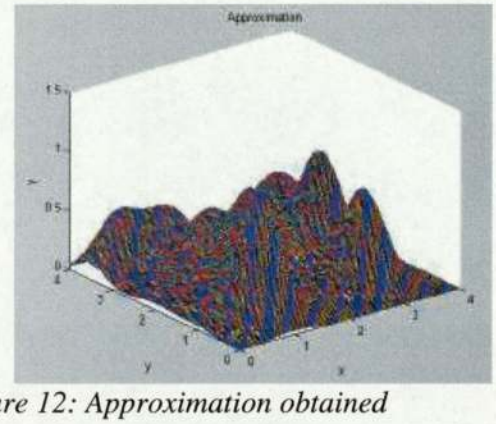


Figure 12: Approximation obtained

Obviously this target environment is more complex than that used in the previous example with BOOMP and the results are clearly more impressive. The approximation is really close to the initial input target environment. This can be seen in [fig.13]. In the particular figure the actual mean error for each slice and its approximation, respectively is observed. Like mentioned earlier in each iteration, OMP is using a different subset of thirteen atoms in order to approximate each vector of the matrix f, however its size is the same.

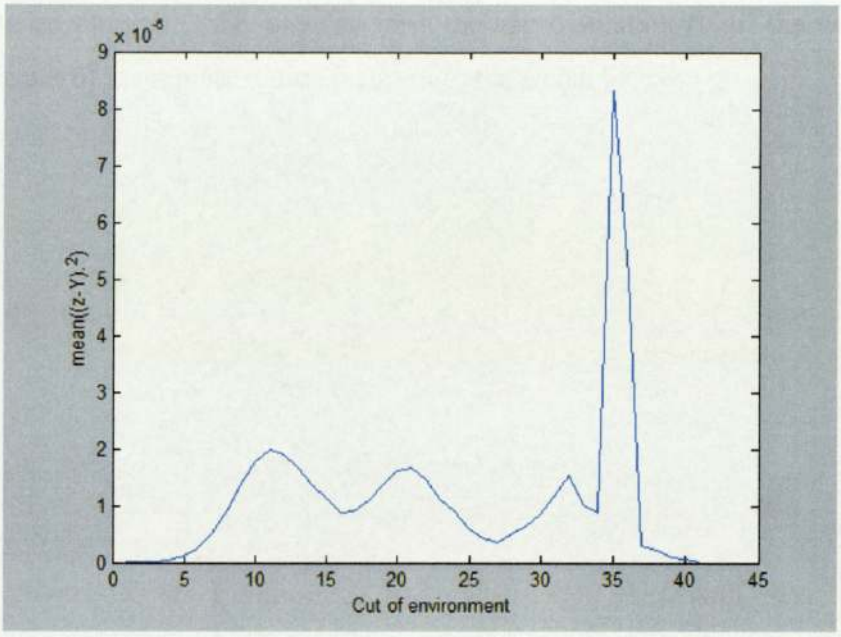


Figure 13: Plot of the actual mean error of the slice used and approximation obtained.

Furthermore the large deviation in the error plot above is due to the large complexity of the particular slice, such as that it consists of three peaks that are fairly separated. Concerning the computational time, in order to produce such an approximation, this is extremely huge, especially when a large M (number of columns in $\boldsymbol{f}$) is considered in the target environment $\boldsymbol{f}$. Moreover, in a classical radar problem it is known, that the subset of elementary signals propagated must be fixed, which is not happening for the particular method of representation. In other words although there is a fixed number of atoms (thirteen) the same subset is not used to approximate the whole environment but only every vector of $\boldsymbol{f}$.

Consequently one could consider using a *random* or *optimal slice* in order to reproduce the whole environment. In essence by using OMP for the particular slice one could select the thirteen atoms and still use this particular subset to approximate the environment in total, which is going to be analyzed further below.

5.2.2 Using the Best Slice (BS)

From the previous simulation it was considered that although the approximation is extremely good it would certainly take a huge number of computational steps, depending on the dimensions of $\boldsymbol{f}$. Consequently it would take a long time to be obtained. Moreover it wouldn't provide us with a solution due to the fact that the selected atoms of each subset used in order to approximate every vector of $\boldsymbol{f}$ vary. In other words the algorithm selects a different subset of atoms to approximate each corresponding slice. There was also reference to a way for further minimization of computational costs by just using a certain slice from the environment. In other words, using the selected subset of functions, that OMP chooses from the dictionary,

for a particular slice in order to approximate the rest of the environment. In this section simulations of the approximation phase are illustrated by using different slices from the environment and select the one in terms of optimality.

Let us assume that the same target environment as the one in the previous section is used in the reconstruction phase. Firstly the same dictionary, as previously, and each of the vectors f_j of the matrix $\boldsymbol{f}$ is introduced to OMP. Then OMP is iterated in order to approximate every slice by using only a subset of thirteen atoms out of the dictionary. As soon as the subset of atoms for every different slice is acquired, respectively, the atoms are stored and OMP is initialized for a second time. This time we are going to consider only as a dictionary the particular subset obtained, thirteen atoms for every slice, respectively, and we are iterating OMP for the rest of the environment. Lastly, in order to obtain the total approximation a new matrix is formed that each column will be created from the formula below.

$$f_j^{(k)} = \sum_{n=1}^{k} \left\langle \beta_n^{(k)} \mid f_j^{(k)} \right\rangle a_{l_n} \qquad (5.2.2.1)$$

where a_{l_n} the subset of the thirteen atoms selected for slice j of the initial target environment and $\beta_n^{(k)}$ the biorthogonal functions calculated as prescribed in section (4.3.2).Further below [fig.14] coincides with the mean error obtained by using the subset of atoms, deriving from the reconstruction of each of the slices, respectively, to approximate the rest of the environment.

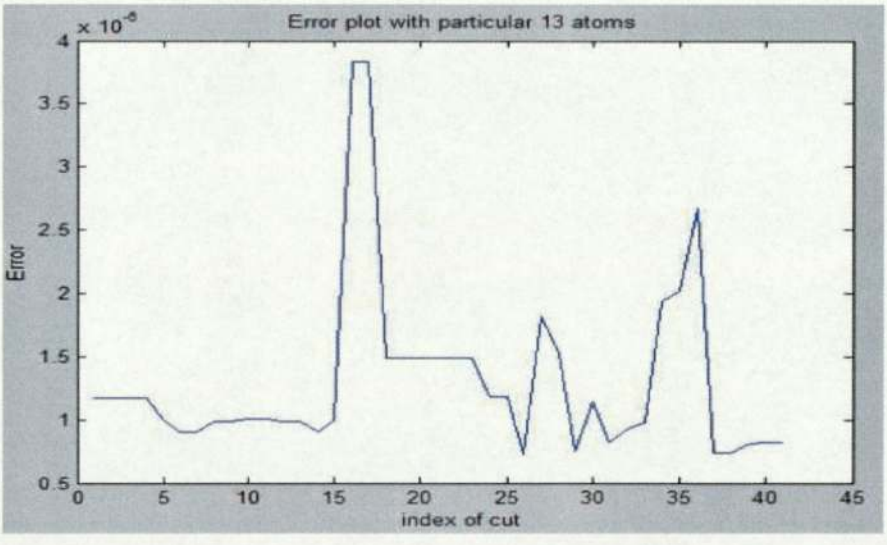


Figure 14: Plot of the mean error of the target environment and approximation obtained.

The approximation obtained is still good and of an order of 10^{-6} (as depicted above) if the algorithm is halted after a finite number of steps. Obviously the computational steps can still be reduced further for the particular environment, if we consider in the representation the slice that produces the minimum error, which is the one with index 26, and an error of 7.348e-7 or just a random slice. The slice and the dictionary in order to produce the approximation of the

whole target environment are depicted below in [fig.15] and [fig.16]. Lastly the approximation obtained is depicted in [fig.17].However in real life it is not possible to know from the start which is the optimal slice to use especially if the target environment in not known initially. In essence, one could have another environment of greater complexity and chose a slice, randomly, in order to represent it that would provide a poor approximation.

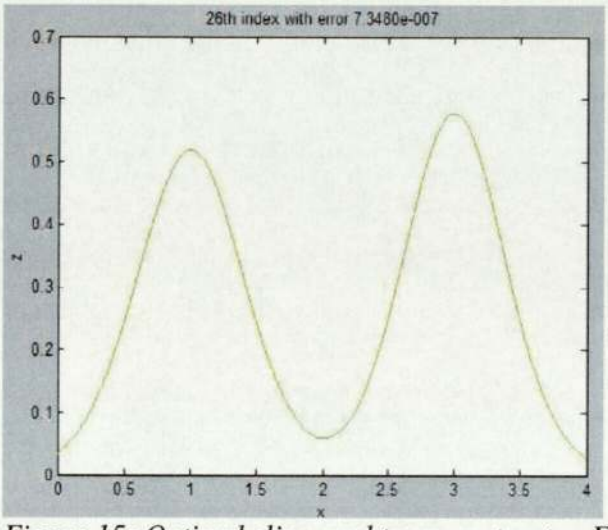


Figure 15: Optimal slice used to reconstruct the whole environment..

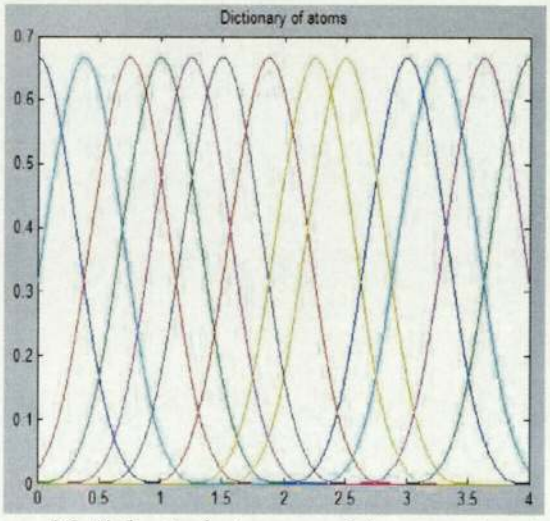


Figure 16: Subset of atoms used to reconstruct the whole environment.

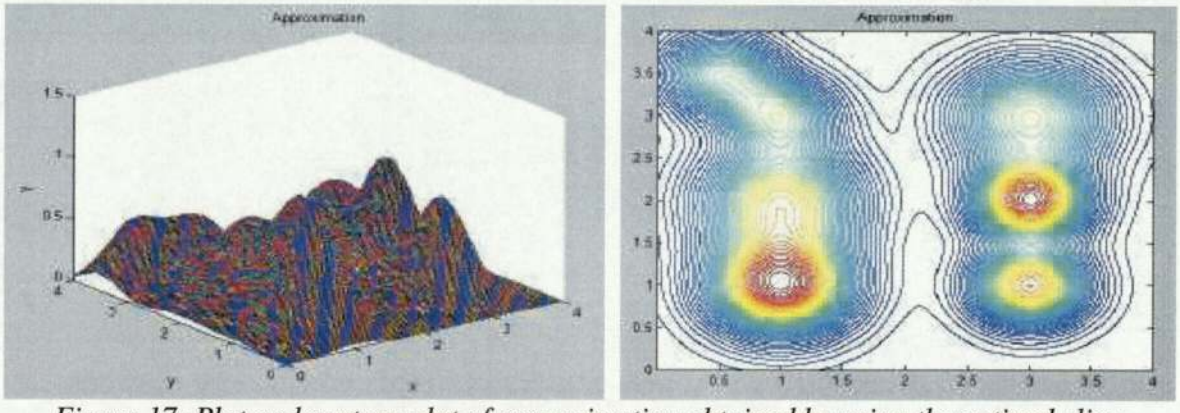


Figure 17: Plot and contour plot of approximation obtained by using the optimal slice.

5.2.3 Simultaneous Orthogonal matching pursuit

In the previous section it was discussed earlier how one can iterate OMP for a matrix and obtain a reconstruction of the target environment. However good the approximation might be, the computational steps required are extremely great in number. In this section it is going to be described how one can introduce the whole matrix in OMP and perform the representation phase simultaneously, an action that will minimize the computational steps, as well.

In a previous chapter the prescription used by OMP to select the most adequate subset of atoms from the dictionary in order to represent a vector (signal) was described. Like already stated in its first steps it calculates the inner product of $\left|\langle a_n, R_k \rangle\right|$ and selects the atom a_{l_n} that

yield a maximum value to that inner product. Adapting the notation in [23], Tropp et al proposed a method to represent a 2D signal simultaneously. This involved mostly the way it selects the atom maximizing the above inner product. So instead of calculating the previous inner product $a_{l_n} = \max_K \sum_{i=1}^{N} \left|\langle f_K, a_i \rangle\right|$ is calculated, where f_K is each column in the signal matrix and a_i the atom of the dictionary involved in the decomposition. The intuition behind maximizing that sum is that we wish to find an atom that can contribute a lot of energy to every column of the signal matrix. Thereafter the greedy selection is performed as prescribed in OMP. By that means the whole input matrix is introduced to the method and this save a lot of computational steps. Obviously one would expect the approximation to be worse in comparison to the previous two ways of representation; however both the results are extremely close.

Below [fig.18] the simulation of the reconstruction of the same target environment used previously is observed. This was obtained by using, exactly, the same dictionary (D13) as used in the previous simulation. The point wise error of the approximation(Y) and the initial signal (z) is depicted in [fig.19].In [fig.21] the error of the approximation obtained with respect to the number of atoms selected for the representation for SOMP and BS is depicted. Notice that the more the atoms the better the approximation. The previous can only imply-a *greedy* algorithm. Lastly [fig.20] depicts the thirteen atoms used in both cases for the approximations, the values on the bottom coincide with the actual index of each atom in the dictionary.

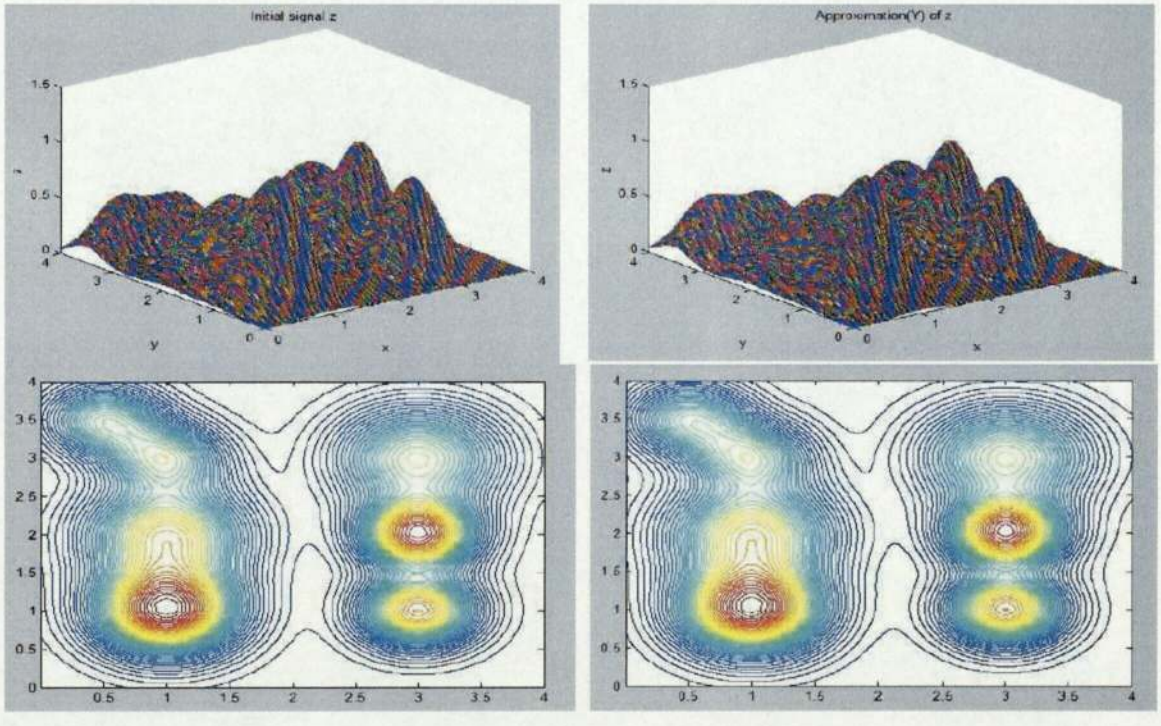


Figure 18: Plots and contour plots of the target environment we wish to reconstruct and approximation obtained by using SOMP.

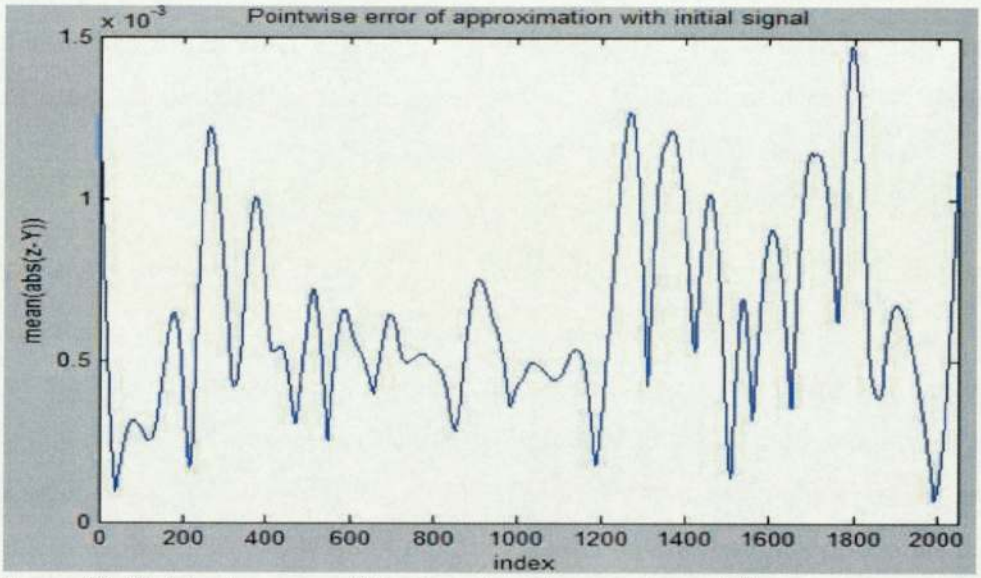


Figure 19: Point wise error of the target environment we wish to reconstruct and approximation obtained by using SOMP.

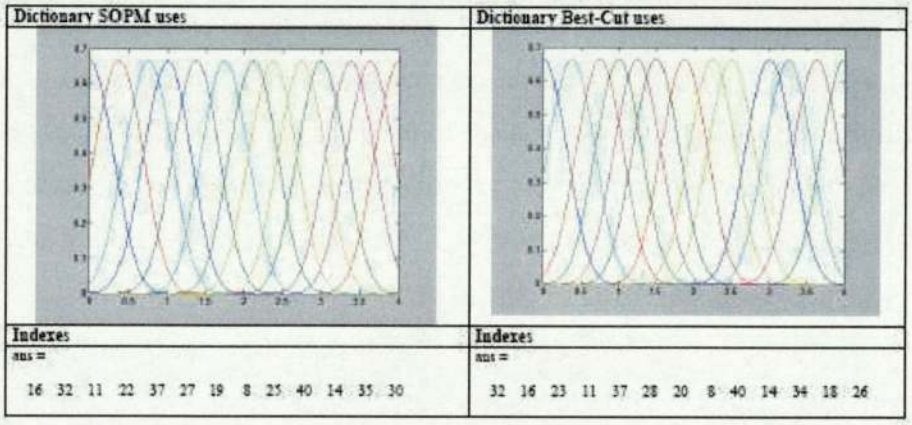


Figure 20: Subset of atoms and their indexes used for the reconstruction.

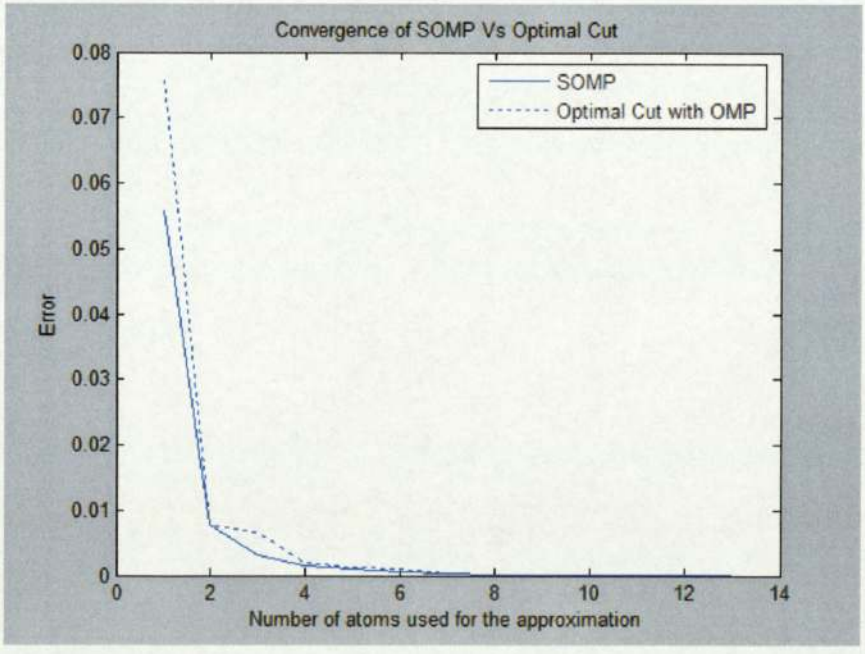


Figure 21: Error plot of the convergence of SOMP and Optimal Slice.

5.3 Remarks

In this chapter ways OMP can be implemented for adaptive approximation of a target environment of two variables were illustrated. Moreover it is shown that when the target environment is known in total and the basis functions, used, are chosen priori to features of the target environment, then it is possible to represent it by just using a subset of these functions. This has shown to yield a sparse approximation with a compromise of error.

The means of representing the target environment was performed with a random or an optimal slice. The difference of each approach is with respect to the computational time and the subset of atoms that it chooses to solve the approximation problem. Consequently the first method was proved inadequate due to the fact that it chooses a different subset for every slice to approximate the target environment. The second approach with the best using slice and SOMP, although it used a certain subset of atoms to approximate the whole environment bear differences with respect to the convergence and computational time. Hence, in all relevant experiments SOMP approach is going to be used.

Chapter 6

Prior Knowledge

Prior knowledge is defined as the knowledge that stems from previous experience. It is sometimes referred to as a *schema*. Schema comes from the Greek word "*σχήμα*", which means shape or more generally plan. While a scheme refers to a loosely described plan, a schema usually refers to specific, well documented, and consistent plans or preconception. In the particular problem addressed there are two different cases of attributes that might be assumed a priori. The first one is assuming partial or total information about the target and the other assume features of the dictionary used to represent a target environment. In this chapter some calculations with OMP are to be shown, a target environment is considered priori and is known in total. On the other hand there will be little information about the dictionary used for the decomposition.

In a previous chapter it was mentioned how important it is for signal processing and reconstruction, to work in a specific space. The non-linear methods described earlier bear robustness to the dictionary used for the representation. In essence, apart from the number of atoms used, the more these atoms of the dictionary resemble the signal being represented the better the representation. The previous is of obvious character since there could be a possibility that the input signal might belong to the space that is being used. So the possibility of spanning a fixed space by B-Spline dictionaries, each of which consists of functions of different support, arise the question as to how to choose in an effective manner the dictionary of B-Splines of optimal support for representing a given signal.

Like stated earlier there are two different ways to increase the subspace's dimension by having a fixed order of the splines. The first one is by increasing the number of knots, say decreasing the distance b between two adjustment knots. And the other way is by increasing the length of the support of each basis function of the cardinal spline space, in other ways the width mb .The later motivated us to search a dictionary of B-Splines of optimal support with

respect to the signal wished to be represented. Let us illustrate the above through a simple example of approximation.

6.1 Increase compact support

Let us assume that the input signal to be approximated is f which is depicted in [fig.22].The target environment consists of two peaks that are fairly separated of the form

$$f(x)=\sum_{i=1}^{N} a_i e^{\left[-\left(\frac{x-mean_i}{b_i\sigma_i}\right)^2\right]} \text{,with N=2} \quad (6.1.1)$$

where the localization factor of each peak is controlled by $mean_i, b_i\sigma_i$ controls the width of each function, N is the complexity of the environment (number of peaks) and a_i the amplitude of each peak.

The particular environment is to be reconstructed with the help of one of the above forward or backward methods described. By means of simplicity let us suppose that the forward approach OMP is used and an approximation is to be obtained with tolerance level of 0,01. In essence, the method is instructed to stop iterating as soon as it reaches a norm of the error equal to the above desired precision. Such a signal has an acceptable approximation in a subspace (D33). D33 is a dictionary created with support of functions of length 0.5, spanned by a B-Spline basis of order four (the length of the support of the dictionary is the order of B-Splines times the distance between knots). Changing the length of the support of the B-spline functions introduces redundancy to a dictionary if it consists of only the basis functions. It has been shown in [20] that changing the support of the functions might result to an identical approximation with less elementary functions used. It should also be noted that the functions used to approximate the target environment are chosen priori to features of it. The graph indicating the approximation of the environment coincides with the one of [fig.27] and the atoms selected to be used for the representation are depicted in [fig.24].Out of all the 35 functions [fig.23] the forward approach has selected only twenty elementary signals (M=20) with a norm of the error of 0.0058.

Now let us suppose that we are trying to find the coarsest possible approximation following the same experiment as above. Let us try and increase the compact support of the dictionary. In the new space the new dictionary (D13) [fig.25] would have length of the support of two for the same order of B-Splines and the same distance between knots. In this case the approximation [fig.27] is obtained by means of only M=13 [fig.26] atoms chosen through the same forward method with error of 0.0066. In that case too, OMP is initialized to halt iterations with a desired precision of 0,01. If an approximation is to be obtained in the

former space (D33) with a subset of thirteen atoms this will be a worse one (norm of error=0.050) in comparison to the latter one. So a better approximation is obtained with the space of length of the support two.

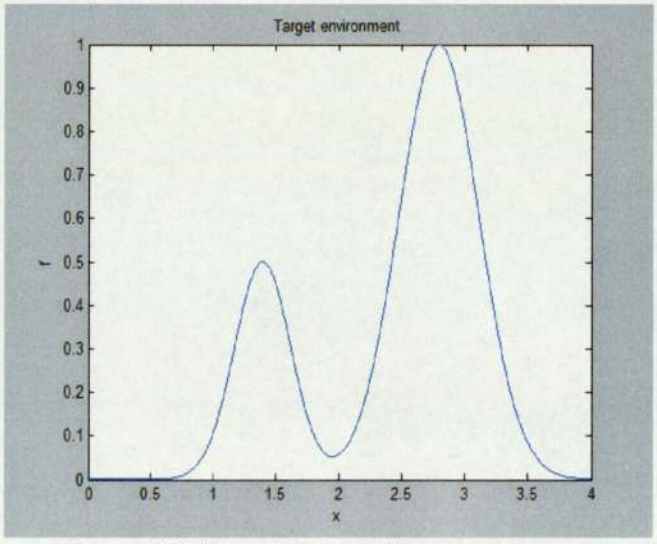


Figure 22: Signal f we wish to reconstruct.

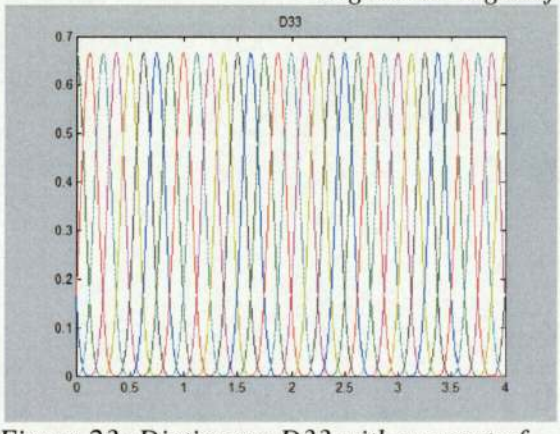


Figure 23: Dictionary D33 with support of functions of length 0.5.

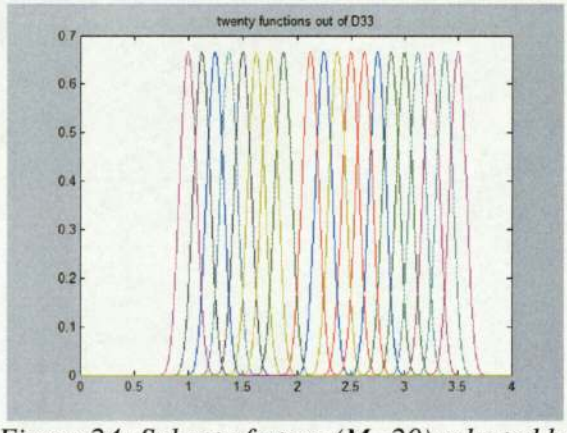


Figure 24: Subset of atoms(M=20) selected by OMP in order to reconstruct f with a tolerance value of 0.01.

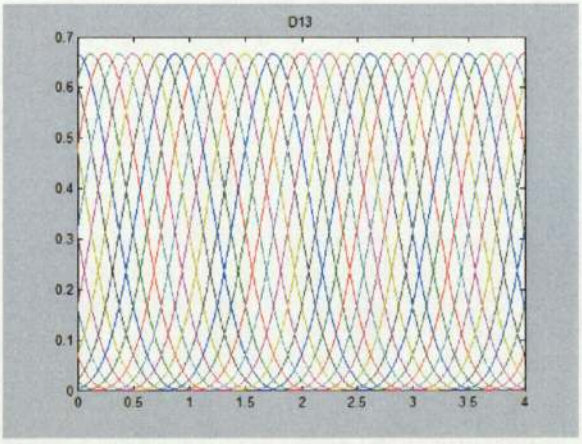


Figure 25: Dictionary D33 with support of functions of length 2.

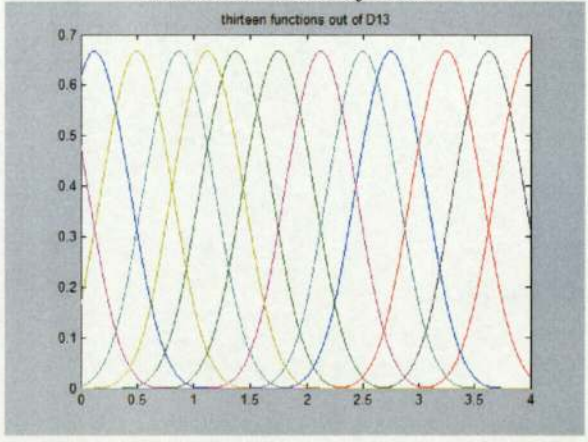


Figure 26: Subset of atoms(M=13) selected by OMP in order to reconstruct f with a tolerance value of 0.01.

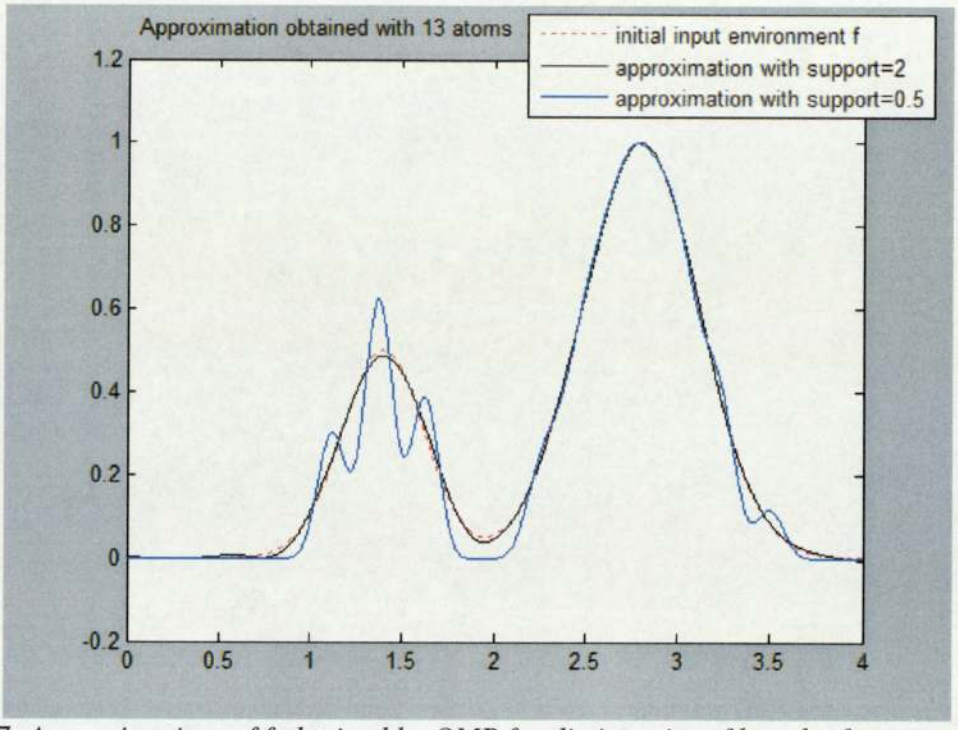


Figure 27: Approximations of f obtained by OMP for dictionaries of length of support 2 and 0.5 by selecting a subset of 13 atoms, respectively.

From the previous it is concluded that by choosing to work with a different space, consequently a different dictionary, the approximation obtained will be either a better or worse one. Still the previous does not provide any solution to the question posed about optimal support of the B-Spline basis. Yet, there is little to say about the support when there is no information about the target environment either. However if the form of the target environment is known initially and we look for certain subspaces , that are included in a particular space, for the one that provides the best approximation with the least selected atoms a hypothesis over the above can be introduced.

In essence, an experiment was performed where, for each iteration of each different subspace that is used for the representation of the input, we chose the dictionary that, both, produced the best approximation and used the least selected atoms from that. The previous is going to be described below.

6.2 Experiment for finding the Optimal Support

Aim

Find the optimal length of the support of B-Spline functions to be used to represent the input environment or find the optimal dictionary for the approximation.

Input environment

The input environment being considered for the problem consists of seven peaks fairly separated, is depicted in [fig.11] and has the form of (5.2.1).

Method

Due to computational complexity and speed SOMP is used.

Procedure and results

A basis of B-spline functions of distance between knots of 2^{-7} (see table 3) is selected and we increase their distance to 2^{-1} at each stage. Then for each nested subspace of the basis functions we increase the length of the support of each B-spline function used in the particular subspace by maintaining the same distance between knots. It has been mentioned in [20] that by increasing the length of the support of the B-spline basis functions we obtain a new subspace that is of larger dimension than the one created of basis functions. In other words the number of functions included in the new space will be increased. The number of functions in each space is calculated by the following manner

$$Number\ of\ functions = m\frac{b'}{b} + N2 \text{ and } N2 = \frac{\text{int}}{b} - 1$$

where b the distance between two adjacent knots, b' the new translation parameter of the B-Spline basis functions such that b'/b is an integer, m the order of the B-spline functions and int the distance of the interval considered.

The intuition behind this is to create the approximation of signal (5.2.1) by using dictionary functions of much larger compact support and chose the length of the functions that will produce the smallest error (4.4.1) and the sparsest representation. In other words we are introducing values for the distance between knots (b) for values between 2^{-7} to 2^{-1}. For each of the values of (b) the length of the support of B-Splines is increased from order_of_Splines*(2^{-7}) to order_of_Splines*(2^{-1}). Moreover the tolerance parameter (desired precision) of the approximation is fixed to 0.01 and the method is initialized. Lastly the norm of the error of the approximation obtained and the number of atoms that are responsible for to such an approximation, respectively is observed. For reasons of computational costs we are using order of B-Splines up to 10.

In the first step of the algorithm, the order of splines that is going to be used is chosen by initializing the method with only the basis functions, that is for b=b'. Let us assume that b=b'=2^{-3} and that the decomposition involves 20 atoms out of the dictionary. From the table [2]

the order of B-Splines that obtain the best possible solution to the problem of approximation with respect to the norm of the error are B-Splines of order ten. However the ones of order four are chosen due to computational complexity. The reconstruction of the target environment is depicted in [fig.28]. As soon as the order of splines is fixed, the method is applied again by creating each time a new dictionary with functions of greater dimension in the particular subspace. Notice that by increasing the length of the support of the functions the dictionary obtained consists of more functions to select from. Then the results are sorted with respect to the least selected atoms. The dictionaries that produce a decomposition with a large number of atoms are eliminated, let us say more than twenty. Lastly out of these, the one that is producing the approximation with the smallest error is selected. It is seen from table [3] that the optimal support of functions, if such a target environment is assumed, is a dictionary of length of the support of two which consists of 47 functions out of which the forward approach selects only thirteen atoms to approximate the input environment with an error equal to 0.0065954.The dictionary and the atoms that the method selects are depicted respectively in [fig.29].The approximation is seen in [fig.31]. The point wise error of the approximation obtained and the initial target environment can be seen in [fig30]. Notice that although a more complex environment is considered the results of the method are remarkable. *Since this is a positive answer to the question of optimal support, the particular dictionary obtained is to be used in the experiments following*. For simplicity we are going to refer to the particular dictionary as D13.

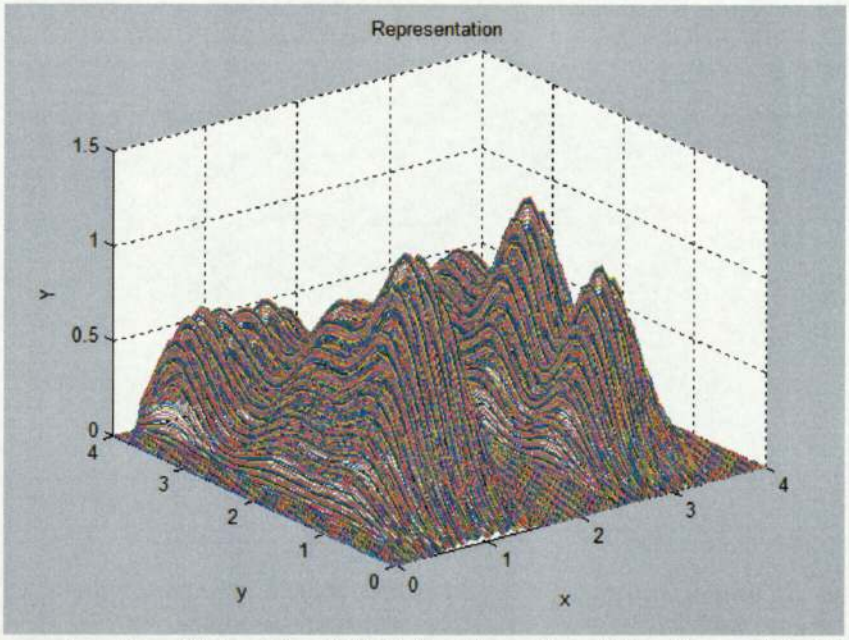


Figure 28: Approximation obtained by SOMP, for the order of B-splines=4, b=b'=2^{-3} *and interval of four. The dictionary consists of 67 functions and SOMP chooses 20 atoms to represent the target environment with a norm of error 0,6385.*

Order of Splines	Interval	b'	b	Functions in Dictionary	Atoms chose n	Norm of the error
1	4	0,0625	0,0625	64	20	1,1045
2	4	0,0625	0,0625	65	20	1,0199
3	4	0,0625	0,0625	66	20	0,82455
4	**4**	**0,0625**	**0,0625**	**67**	**20**	**0,63846**
10	4	0,0625	0,0625	73	20	0,46836

Table 2: Table with results, obtained by SOMP, for the purpose of fixing the order of B-splines. The order of splines that we select with respect to the norm of the error is depicted in blue.

Order of Splines	Interval	b'	b	Functions in Dictionary	Tolerance	Atoms chosen	Norm of the error
4	4	0,0078125	0,0078125	515	0,01	507	0,0095939
4	4	0,015625	0,0078125	519	0,01	321	0,0085132
4	4	0,03125	0,0078125	527	0,01	167	0,0095143
4	4	0,0625	0,0078125	543	0,01	69	0,0096121
4	4	0,125	0,0078125	575	0,01	36	0,0099177
4	4	0,25	0,0078125	639	0,01	21	0,0073105
4	4	0,5	0,0078125	767	0,01	13	0,0066071
4	4	0,015625	0,015625	259	0,01	253	0,0088182
4	4	0,03125	0,015625	263	0,01	149	0,0094907
4	4	0,0625	0,015625	271	0,01	74	0,0043036
4	4	0,125	0,015625	287	0,01	39	0,0093278
4	4	0,25	0,015625	319	0,01	22	0,0077819
4	4	0,5	0,015625	383	0,01	13	0,0067339
4	4	0,03125	0,03125	131	0,01	128	0,0052431
4	4	0,0625	0,03125	135	0,01	71	0,0093189
4	4	0,125	0,03125	143	0,01	39	0,0096773
4	4	0,25	0,03125	159	0,01	21	0,00917
4	4	0,5	0,03125	191	0,01	13	0,0067069
4	4	0,0625	0,0625	67	0,01	64	0,0096727
4	4	0,125	0,0625	71	0,01	32	0,0089694
4	4	0,25	0,0625	79	0,01	23	0,0089734
4	4	0,5	0,0625	95	0,01	13	0,0066075
4	4	0,125	0,125	35	0,01	32	0,0090451
4	4	0,25	0,125	39	0,01	25	0,0086736
4	**4**	**0,5**	**0,125**	**47**	**0,01**	**13**	**0,0065954**
4	4	0,25	0,25	19	0,01	17	0,0029402
4	4	0,5	0,25	23	0,01	10	0,0094998
4	4	0,5	0,5	11	0,01	11	0,016987

Table 3: Table with results, obtained by SOMP, for the purpose of deciding the dictionary of Optimal Support with respect to the norm of the error of the approximation obtained. The dictionary of Optimal Support with a desired precision of 0.01 consists of 47 functions, with length of two and SOMP chooses 13 atoms to represent the target environment with a norm of error 0,0065954.

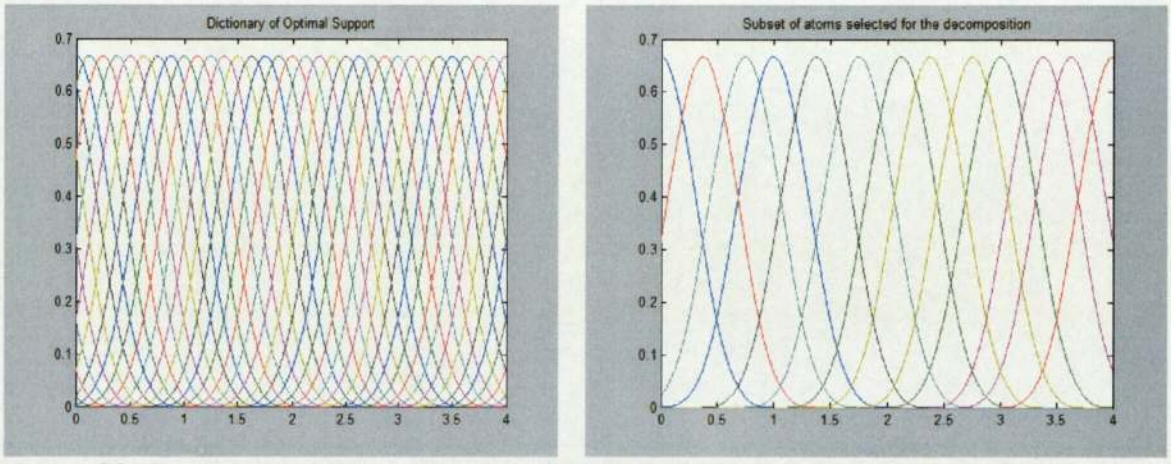


Figure 29: Dictionary of Optimal Support (left) and subset of atoms selected to represent the target environment (right).

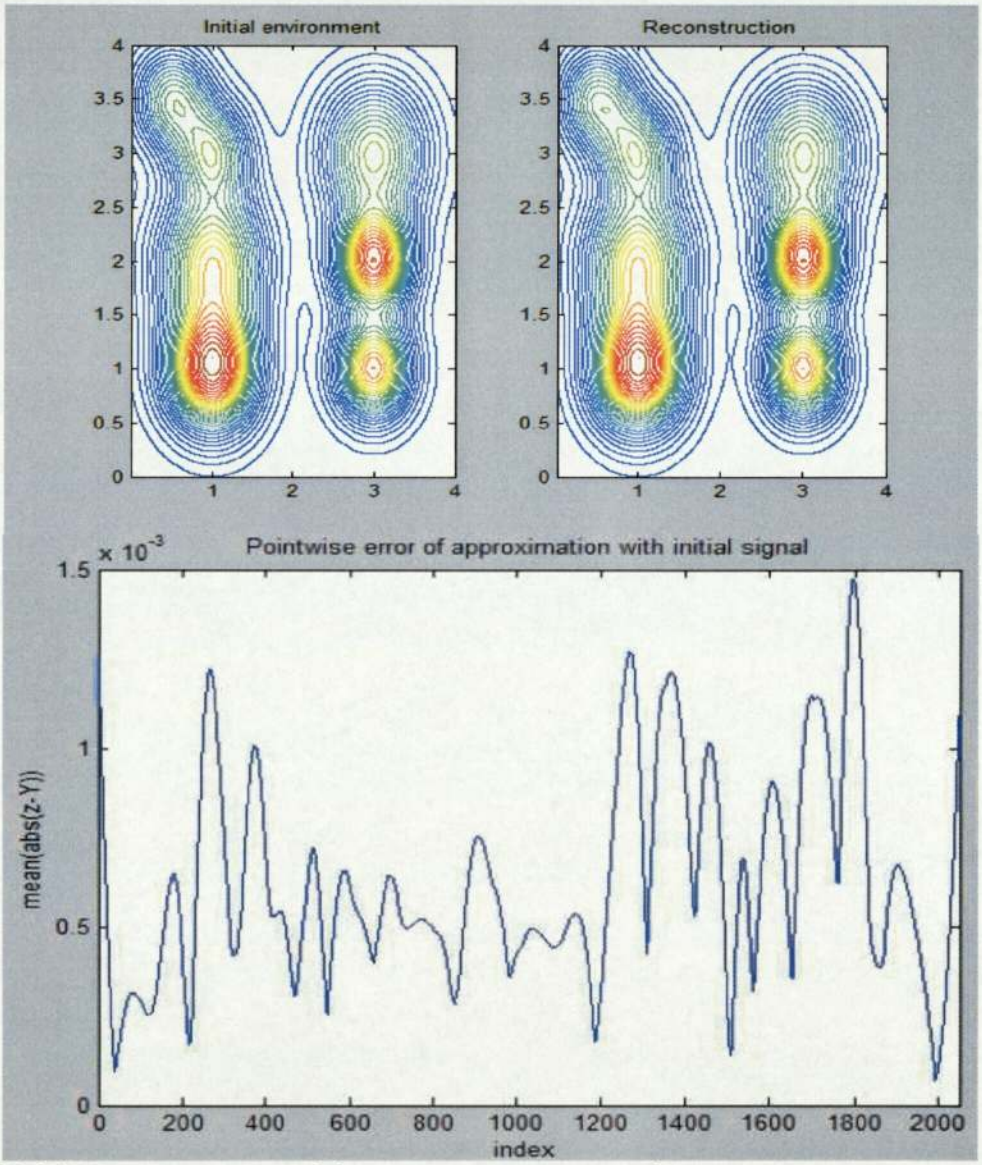


Figure 30-31: Point wise error of the target environment and its approximation returned by using SOMP.(Bottom)
Contour plots of approximation (right) obtained by SOMP and initial target environment (left) we wish to represent with the dictionary of Optimal Support.

6.3 Remarks

In this chapter examples were illustrated on increasing the compact support of the dictionary used in the approximation phase. Prior knowledge of the target environment was considered and a particular dictionary of optimal support was selected through experiments. Firstly the order of B-Spline functions was fixed with respect to the norm of the error of the approximation. Thereafter, an adequate dictionary that produced an approximation with the smallest norm of the error and the least selected atoms was chosen. It was witnessed that increase on the length of the support of the functions in the space considered, resulted to a dictionary with more functions. Consequently the basis functions were fixed and the particular dictionary is going to be used in the following experiments. However, there is still need for a particular subset out of the dictionary of 47 functions when the target environment is not known initially. Since the above environment can be well approximated with a subset of only thirteen atoms we are going to perform experiments when we consider *prior knowledge of the target environment*. These will involve its approximation with a combination of only thirteen elementary signals that are going to be used for the rest of the experiments. For simplicity this dictionary chosen is going to be denoted as D13.

CHAPTER 7

Fixing a particular subset of atoms for the representation of a target environment

7.1 Introduction

In the previous chapters SOMP was used for the representation phase of a target environment of two variables. Moreover prior knowledge was considered over the target environment and the space that we are going to work with was fixed. In other words attributes of the basis functions of the dictionary were fixed, such as the order of B-Spline functions, their support and the number of atoms that are adequate enough to provide us with a good approximation. A good one means a compromise between the norm of the error of the approximation and a sparse one. However what is still needed is a particular subset for the representation when the target environment is partially known. The previous was managed through some simple experiments that are going to be illustrated in this chapter.

In all experiments prior knowledge is considered, of the basis functions (dictionary) used for the decomposition and certain features of the target environment, wished to be represented. So an environment of seven peaks fairly separated is taken into account, however the location of the peaks is not known, in the first one, or the width of each peak, or both of them in the latter one. Moreover in others there is no information about both the amplitude and the width or the location, or all of the three attributes together. Consequently by iterating SOMP one atom is fixed, at a time, of the subset that will determine the initial subspace to influence the decomposition, until we obtain the whole subset. However let us illustrate the above through the particular experiments.

Input environment

An environment of the form (5.2.1) is assumed. It is an environment with a complexity of seven peaks (K=7) that are fairly separated. In each of the experiments following it is considered, that there is partial information about the target environment. In other words, in each case, random values are chosen for each parameter, individually, or the combination of them. In all calculations the same dictionary D13 (comprising of 47 atoms, as chosen in the previous chapter) is used and we obtain approximations with a subset of 13 atoms. Moreover **SOMP** is used to obtain the approximation. Lastly in order to measure the error of the approximation the Euclidean distance (5.2.2) is used.

The value of the error is subjective since what we are interested in is to have an approximation that is bearable in proportion to the input target. For example producing the number of peaks in total or their localization individually is a good approximation. In other words we are searching for a compromise between the norm of the error of approximation and the number of elementary signals that participate in the linear combination (a sparse approximation).

7.2 Selecting a subset of atoms for a target environment where the location of each peak is unknown.

Let us assume that an input environment (5.2.1) consisting of seven peaks (K=7) is considered. For the particular experiment all parameters are known apart from the location $(mean_x^i, mean_y^i)$ of each of the seven peaks. The aim of the experiment is to try to fix the atoms selected for the representation phase based on the frequency they appear. Since thirteen atoms are considered for the decomposition we are going to have twelve iterations. In essence, in each iteration we are going to fix one atom each time until we have the whole most probable subset of thirteen.

In each iteration, of the twelve ones, the procedure is going to be the following. All the peaks are shifted randomly by just choosing values between the interval [0.5 , 3.5]. Such an interval is chosen in order to have the whole environment totally contained in the space being considered. Moreover there is a case of overlapping of the peaks with each other. In essence two different vectors of seven values are created, each, in the interval considered. These values represent the location of each peak in xx' and yy' respectively. For each different set of values for the location of the peaks throughout the space SOMP is initialized. In order to represent each target environment a dictionary with support of B-Spline functions of two, distance between knots 2^{-3} and order of splines of four is used. The dictionary consists of 47 functions.

Furthermore SOMP is instructed to use only a subset of thirteen atoms out of the dictionary in order to represent the input environment. 10000 iterations were performed, or 10000 different target environments were produced for SOMP to approximate.

In essence as soon as the first 10000 iterations finish we observe a vector returned (Di) that depicts the indices of atoms, such as their position in the initial dictionary before the decomposition, and has a length of 1x47.Out of that vector, the first thirteen elements (Dii) are the ones that contribute a lot of energy in the decomposition. Then in the next iterations the above are repeated and we calculate the times that each index of the initial form of the dictionary participates in the decomposition, for example, number 20 appeared as the first element of that vector (Dii) etc. Thereafter as soon as the 10000^{th} iteration is reached, the probability of each of the 47 atoms that appeared in each index of the particular new vector (Dii) is computed. From the new vector Prob (with dimension 47x13) that is produced, from every 10000 iterations, the frequency that the indices appear is observed. There are two possible occasions that might occur in order to select an atom. The first one is that it appears with the highest frequency and the other one that of the highest probability (if all indices in the vector Prob are different). Consequently we choose the particular atom of highest frequency or probability as an input in the next 10000 iterations in order to determine the initial subspace that will influence SOMP.

Finally in the twelfth iteration of the 10000 ones the most probable subset of atoms for the particular experiment is constructed. In [fig.32] we depict the actual norm of the error returned for the subset of thirteen atoms that we chose to influence SOMP. Although the deterministic nature of the algorithm, we notice the approximation is bearable with the particular subset selected under the assumptions taken. The norm of the error, for the particular subset, has values that are concentrated around an error of order 10^{-2}. The atoms that determine the subset selected out of the dictionary [fig.33] are depicted in [fig.34]. The subset of atoms selected through the above procedure is denoted as S_L.

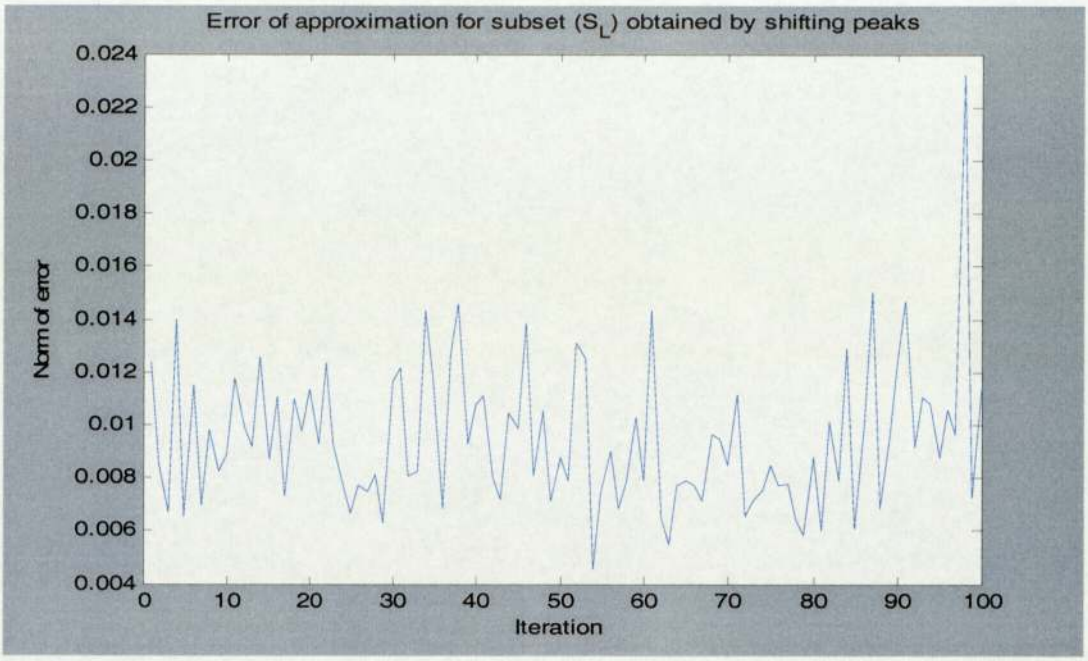
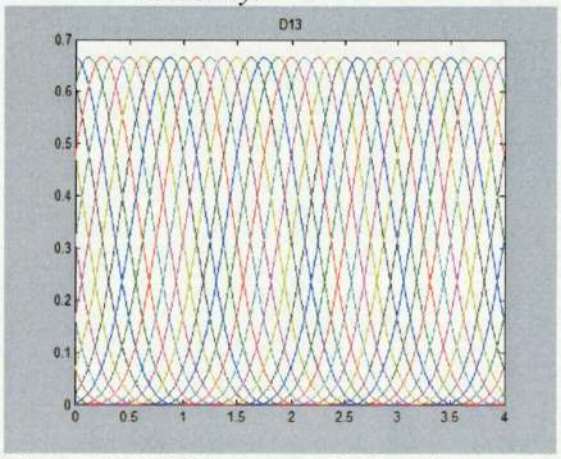


Figure 32: Error plot of the norm of the error for the particular subset when we shift the peaks randomly.



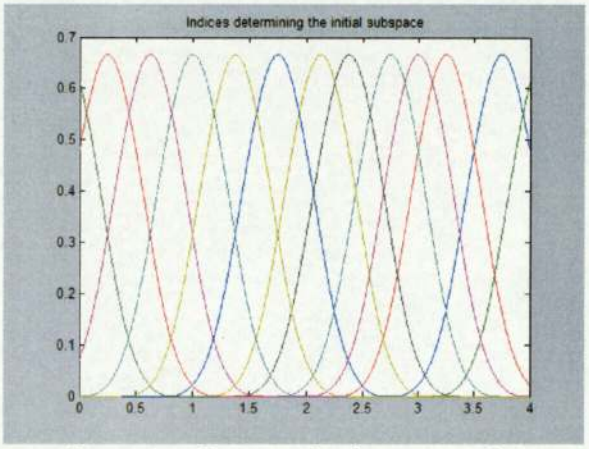


Figure 33-34: Dictionary (left) and atoms chosen to determine the initial subspace (right)

7.3 Selecting a subset of atoms for a target environment where the width of each peak is unknown.

For the particular experiment the same procedure as above is followed. However for that case it is assumed that only the localization and amplitude of each peak is known but the width is not known priori. Let us assume an input environment of seven peaks is given that each has a particular location of $(mean_x^i, mean_y^i)$ and amplitude a_i and are of the form of (5.2.1).

In other words the parameter $(b_i\sigma_i)^2$ that controls the width of each of the seven peaks is unknown. Moreover the interval used to pick random values for each peak respectively is [0.2 0.4].Consequently the width of each peak is from one to two (the width of each peak can

be calculated if we solve 5.2.1 with respect to x or y when all parameters are known, f is approaching zero and x or y equals to zero).The aim of that experiment is to try to fix the atoms selected for the representation phase based on the frequency they appear. In that experiment too a subset of thirteen atoms is considered to influence SOMP.

After the whole procedure the most probable subset of atoms for the particular experiment is depicted in [fig.37].In [fig.35] the actual norm of the error returned for the subset of thirteen atoms that we chose to influence SOMP is depicted. We notice that for the particular subset selected, the approximation is bearable. Moreover we observe in the error plot that the norm of the error, for the particular subset, has values that are bounded in an interval of an error of order 10^{-3}. The subset of atoms selected through the above procedure is denoted as S_V.

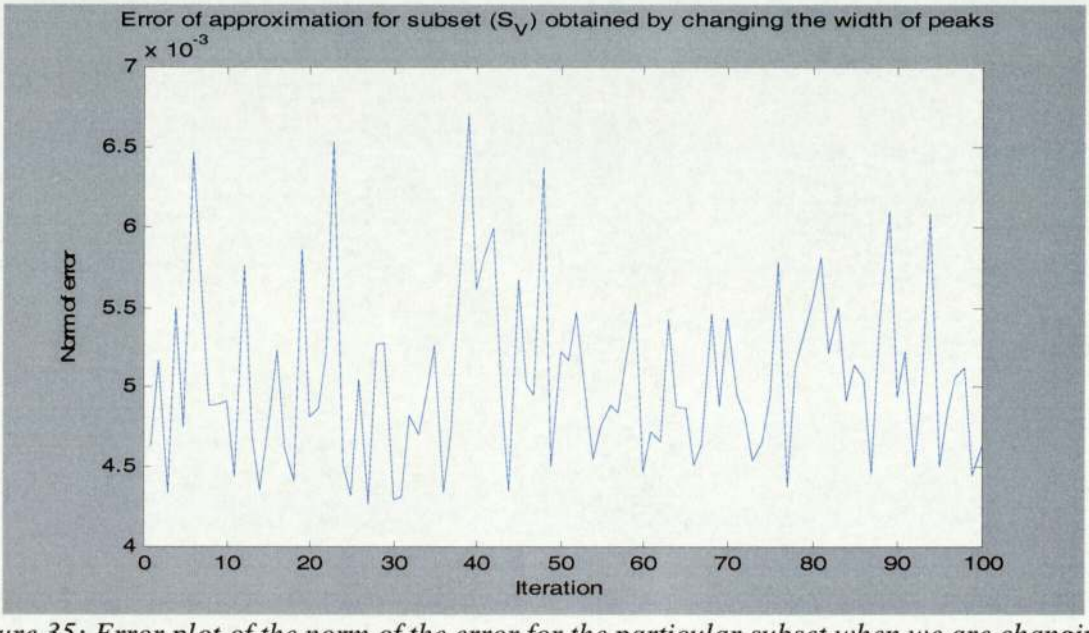


Figure 35: Error plot of the norm of the error for the particular subset when we are changing the width of the peaks randomly.

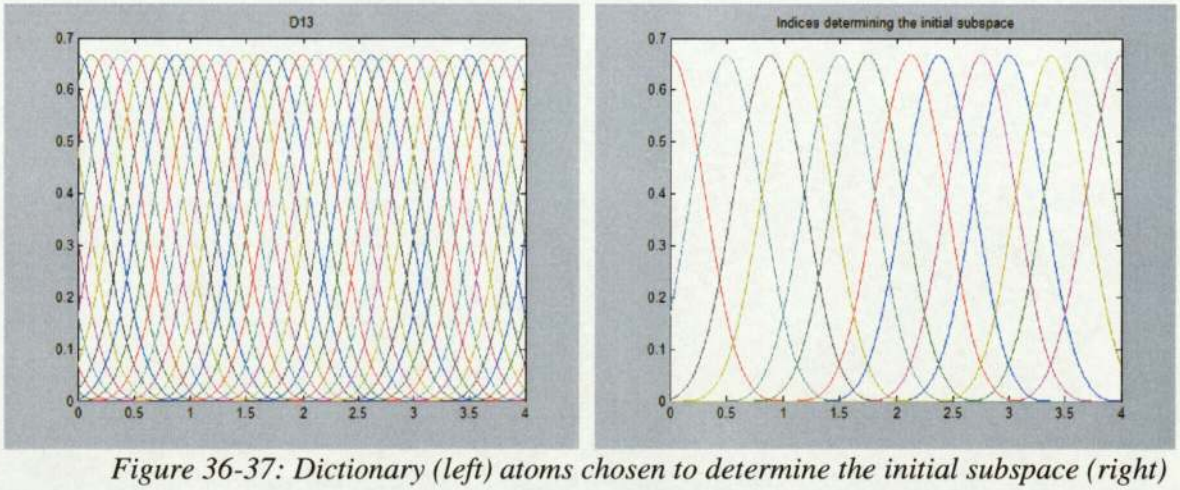


Figure 36-37: Dictionary (left) atoms chosen to determine the initial subspace (right)

7.4 Selecting a subset of atoms for a target environment where the location and width of each peak are unknown.

In the previous experiments a particular subset of atoms was chosen for each different case of prior knowledge. However which subset can one choose that is adequate enough when one combines the above? For the particular experiment the same procedure as above is going to be followed, however for that case it is assumed that we do not know the localization and the width of each peak respectively. Let us assume that an input environment of seven peaks (5.2.1) is considered, that each has a particular amplitude a_i. In other words, it is assumed that there is no information about both the location $(mean_x^i, mean_y^i)$ and the parameter $(b_i \sigma_i)^2$ of each of the seven peaks. Moreover the interval used to pick random values for the parameter that controls the width of each peak respectively is [0.2 0.4] and for the location [0,5 3,5].The aim of that experiment is to try to fix the atoms selected for the representation phase based on the frequency they appear. In that experiment a subset of thirteen atoms to influence SOMP is also considered.

After the whole training phase the most probable subset of atoms obtained are depicted in [fig.40].In [fig.38] we have the actual norm of the error returned for the subset of thirteen atoms that was chosen to determine the initial subspace in SOMP. For that case too the norm of the error of the approximation is deviated around the value of 10^{-2}, which is a bearable error for the approximation. The subset of atoms selected through the above procedure is denoted as S_{L+V}.

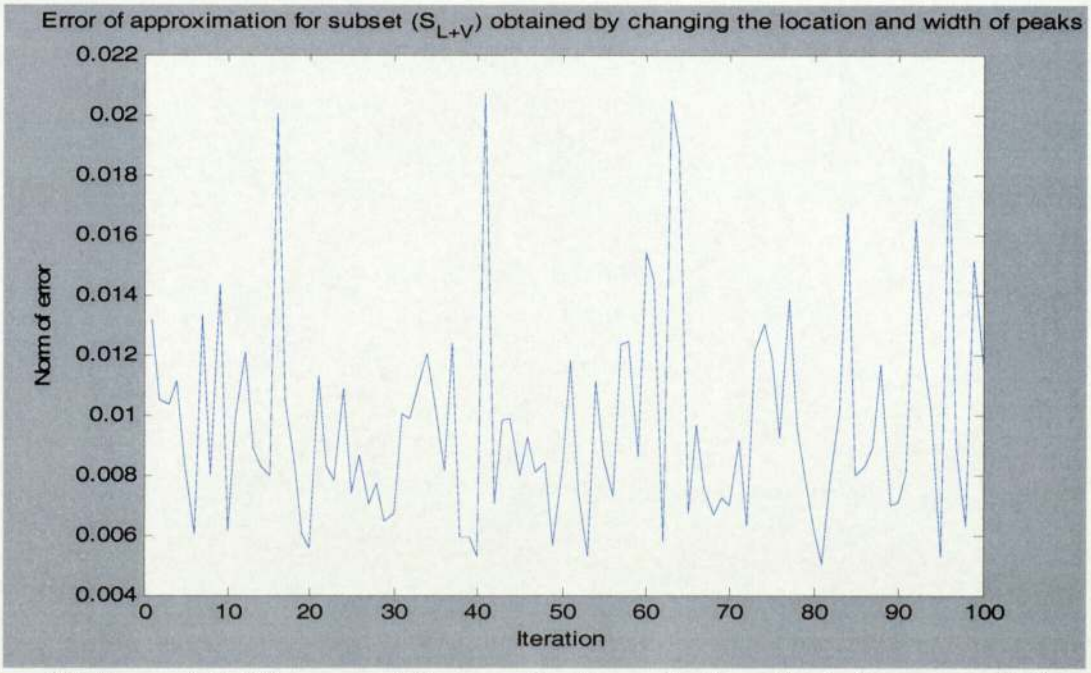


Figure 38: Error plot of the norm of the error for the particular subset when we are both shifting and changing the width of the peaks randomly.

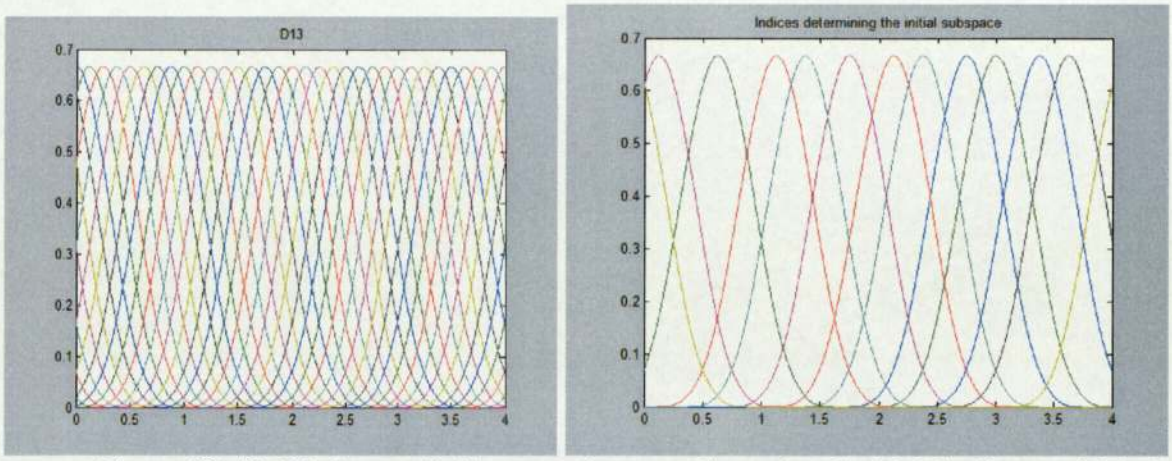


Figure 39-40: Dictionary (left) atoms chosen to determine the initial subspace (right)

7.5 Selecting a subset of atoms for a target environment where the amplitude and width of each peak are unknown.

In the particular experiment it is assumed that there is knowledge about the location $(mean_x^i, mean_y^i)$ of each peak, but no information about its amplitude a_i and the parameter $(b_i \sigma_i)^2$ .In other words it is considered that each peak is "sharpened" randomly. By sharpening peaks we mean that for large values of the amplitude factor we are going to have small values of the width. In essence these two parameters are not proportional. For the particular experiment the same procedure as previously is followed. Furthermore the same dictionary of support of B-spline functions of two is used. Let us assume that an input environment of seven peaks of the form (5.2.1) is considered, which we only know where the peaks are localized. In other words the location $(mean_x^i, mean_y^i)$ of each of the seven peaks are known priori. Moreover the interval considered to peak random values for sharpening the peaks will produce peaks that will vary between interval [0,07 0,4] for the $(b_i \sigma_i)^2$, of each peak, and [0,4 1,6] for the amplitude, respectively. The aim of that experiment is to try to fix the atoms selected for the representation phase based on the frequency they appear. In that experiment too a subset of thirteen atoms to influence SOMP is used.

After the whole procedure the most probable subset of atoms for the particular experiment is depicted in [fig.43].In [fig.41] we have the actual norm of the error returned for the subset of the thirteen atoms that was chosen to influence SOMP. We observe from the error plot that the norm of the error of the approximation is concentrated around the value of 0.04. The subset of atoms selected through the above procedure is denoted as S_{V+A}.

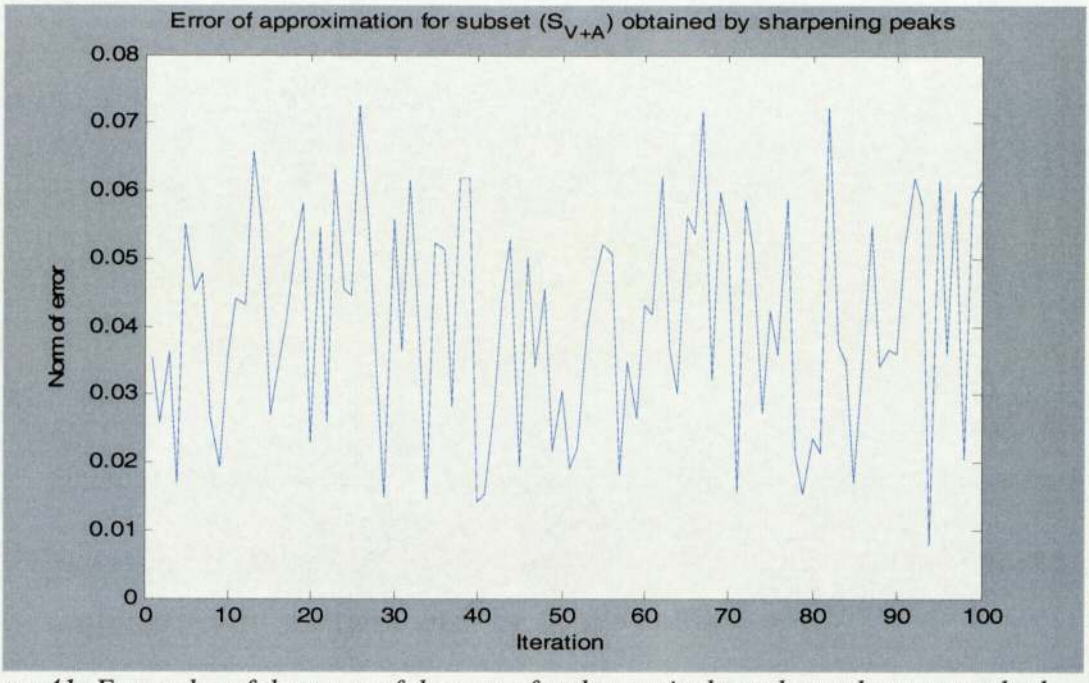


Figure 41: Error plot of the norm of the error for the particular subset when we are both changing the amplitude and the width of the peaks (sharpening peaks) randomly.

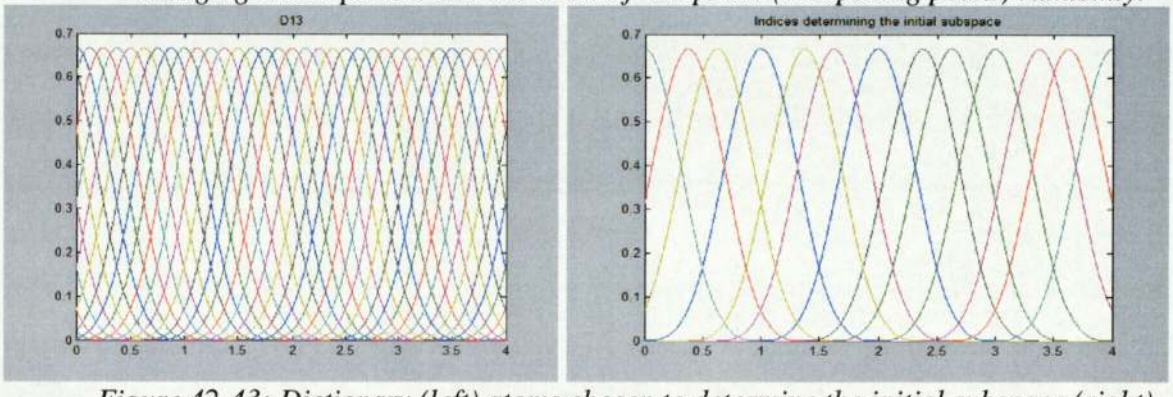


Figure 42-43: Dictionary (left) atoms chosen to determine the initial subspace (right)

7.6 Selecting a subset of atoms for a target environment where the amplitude, width and the location of each peak are unknown.

In the previous experiments we observed the subset obtained if different combinations of prior knowledge of the target are considered. In this experiment it is assumed that the only information known about the environment is the number of peaks. The procedure for that experiment is the same as above. In other words, the atoms selected for the representation phase are fixed based on the frequency they appear. Furthermore the same dictionary of support of B-spline functions of two is used. Let us assume that an input environment of seven peaks of the form (5.2.1) is used.

Random values are picked for every parameter, such as location, width and amplitude. Consequently the input environment that is going to be produced will have parameters that vary between intervals [0,07 0,4] for $(b_i, \sigma_i)^2$, [0,5 3,5] for the location, and [0,4 1,6] the

amplitude for each peak, respectively. In that experiment too we are going to have a subset of thirteen atoms to influence SOMP.

After the whole procedure the most probable subset of atoms for the particular experiment is depicted in [fig.46].In [fig.44] we have the actual norm of the error returned for the subset of thirteen atoms that we chose to influence SOMP. For that case too, the norm of the error of the approximation is bounded on an interval that is bearable for the representation phase. The subset of atoms selected through the above procedure is denoted as S_{V+L+A}.

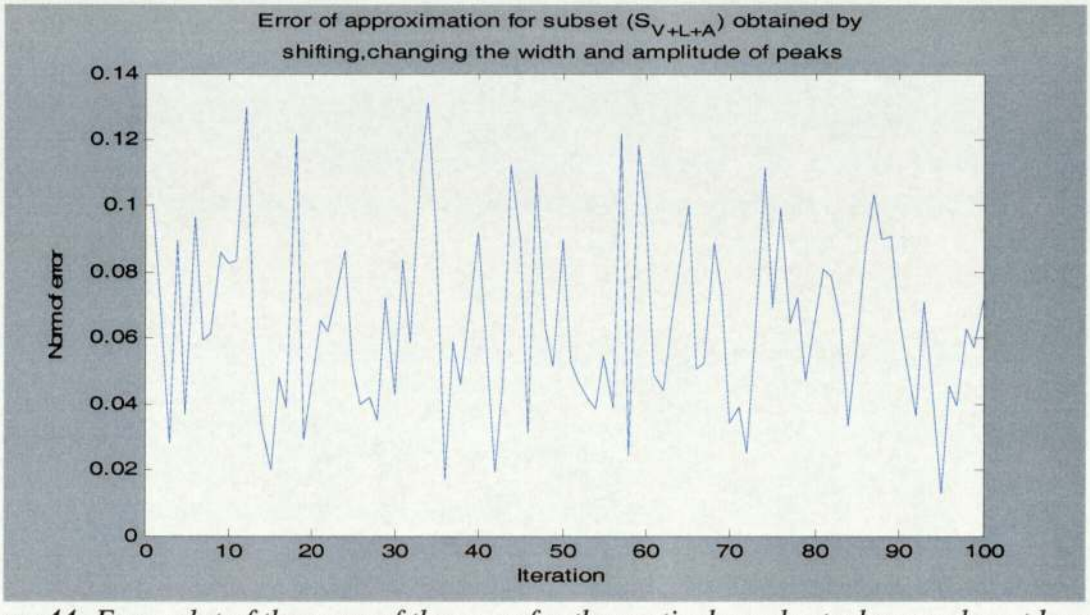


Figure 44: Error plot of the norm of the error for the particular subset when we do not know any features of the target environment apart from the number of peaks.(changing randomly the width, the location and the amplitude of the peaks)

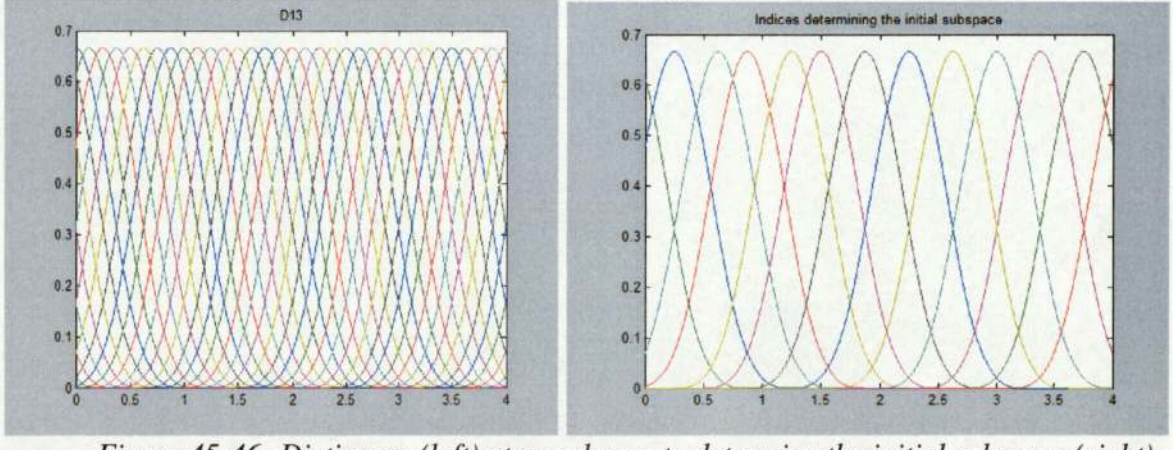


Figure 45-46: Dictionary (left) atoms chosen to determine the initial subspace (right)

7.7 Conclusions and remarks

In this chapter it was assumed that the space we are going to work with is known, such as features of the dictionary and the size of the subspace of atoms involved in the decomposition. Moreover partial information was considered about the parameters that control the target environment that we wish to represent. Based on this prior knowledge we used a

dataset of 10000 different values, that is 10000 different target environments for every experiment separately, and a particular subspace was obtained that will influence SOMP. This subset of atoms was chosen according to the frequency or probability that each index appeared in the decomposition phase. Consequently five different sets of atoms were obtained, for each case of prior knowledge, respectively, that produced an approximation of the input environment with a norm of the error that is of order 10^{-2}. The previous means, that the error is bearable for the representation phase. For instance, if we have information for the width and amplitude of each peak then the localization of the total environment is preserved correctly. Furthermore one should notice that in the subset of atoms chosen there are certain cases where some atoms are the same. Something which is very logical since the algorithm is initialized to choose the atoms to represent the target environment with the same attributes. In essence notice the first three different cases where we do not have information about the location or width or both of them for each peak respectively. In these three cases we observe that there are five common indexes 30,32,25,27 and 22.

Apart from the above it is worth noting that there in no repetition of the appearance of the indices in the subset that were obtained. The previous is due to the means the algorithm chooses the atoms to take part in the decomposition phase. In other words each column of the residual is orthogonal to the atoms indexed in the subset obtained, therefore no atom is chosen twice.

Concerning the value of the norm of the error for the particular subsets that was obtained we seek for a compromise between the approximation error, measured with Euclidean distance, and the number of elementary signals that participate in the linear combination. Consequently for that error we can identify a good approximation involving only few elementary signals out of the dictionary, in other words a sparse approximation. Hence we assumed from the previous chapter that with a subset of thirteen atoms out of a dictionary of a certain support of B-Spline function one could obtain a good approximation for such an environment wished to be represented. On the other hand due to the greediness of SOMP one could increase the atoms participating in the decomposition phase which would lead to a smaller order of error. In other words a better approximation. There is also a case that the function of two variables, we wish to represent, could be outside the space that we work with. This implies that even if the whole dictionary was used as a subset of atoms to participate in the decomposition phase then the approximation obtained would still be bad. However this does not happen in the above experiments due to the intervals chosen for each of the random values.

Concluding there is still need to produce evidence for the effectiveness of the approximation phase. The above is illustrated in the next chapter by iterating SOMP for a

different dataset that is a random test set, for the same experiments and the subset of atoms obtained in this chapter.

CHAPTER 8

Prior knowledge of target

8.1 Introduction

In the previous chapters it was shown how one can obtain a good approximation of the target environment with SOMP by seeking for a compromise between the approximation error, measured with Euclidean distance, and the number of elementary signals that participate in the linear combination. Moreover through certain procedures a subset was obtained, for particular assumptions of prior knowledge of the target environment. That compromise between the approximation error and the combination of each particular set involved implied a sparse approximation.

The aim of this chapter is to visualize the robustness of the particular subsets, obtained for different cases of prior knowledge for a function of two variables. In other words we are going to perform experiments for the representation of the target environment that we might have no information about the location, width, amplitude and their combinations, respectively, of each of the peaks. Moreover we are going to test the subset of atoms in the decomposition, obtained by the training phase (previous chapter) with a random test set of values that is different from the training set that was already used. Last but not least the approximations obtained with the subset that influenced SOMP are compared with one obtained when one does not use such a subset to influence SOMP. It should also be noted that the approximations obtained without influencing SOMP are obtained by the selection of a subset of thirteen atoms out of the dictionary used. Moreover some examples of the representation are illustrated. These are included for the purpose of not only comparing the case when one influences each reconstruction algorithm or not, but also both the approximations obtained and the initial target wished to be represented. Lastly we are hoping to draw some general conclusions and remarks. For simplicity subsets are going to be denoted as $\mathbf{S_L}$ (fig.34) for the one obtained from not knowing where the peaks are localized, $\mathbf{S_V}$ (fig.37) for having no information about the width

of peaks, $\mathbf{S_{V+L}}$ (fig.40) for both unknown attributes (location and width of peaks), $\mathbf{S_{V+A}}$ (fig.43) for sharpening the peaks (unknown width and amplitude of each peak), $\mathbf{S_{V+L+A}}$ (fig.46) knowing only the number of the peaks and **D13** (fig.29) the dictionary that these subsets were drawn.

***Assumptions*:**

In the following experiments it is assumed that there exists an environment of the same form (5.2.1) as previously. Moreover, based on chapter six, the optimal dictionary that can approximate such a target environment is one of B-Spline functions of length of two. Furthermore for each different case of prior knowledge about the environment a subset of thirteen atoms is going to be used for the decomposition with SOMP, which was obtained from the training phase (chapter six). In all experiments we are going to compare the approximation obtained with the particular subset and the one obtained with not having an initial subset determining the subspace by recourse of the norm of the error (5.2.2). Finally, where applicable, we are going to illustrate through numerical examples the approximation obtained by SOMP, of the target environment we wish to represent.

Like stated previously the value of the error is subjective since what we are interested in is to have an approximation that is bearable in proportion to the input target. For example producing the number of peaks in total or their localization individually is a good approximation. Considering the test data, random values are drawn from intervals. These will be different from the ones used in the training phase.

8.2 Experiments

The experiments that are following are aiming on testing the subsets obtained from the training previously and have to do with:

- **Shifting peaks randomly (unknown parameters** $mean_x^i$ and $mean_y^i$ **)**
- **Changing the width of the peaks randomly (unknown parameter** $b_i\sigma_i$ **)**
- **Changing the variance and shifting peaks randomly (unknown parameters** $mean_x^i$, $mean_y^i$ and $b_i\sigma_i$ **)**
- **Sharpening peaks (unknown parameters** a_i and $b_i\sigma_i$ **)**
- **Changing the location, width and amplitude of the peaks randomly (unknown parameters** $mean_x^i, mean_y^i$, a_i and $b_i\sigma_i$ **)**

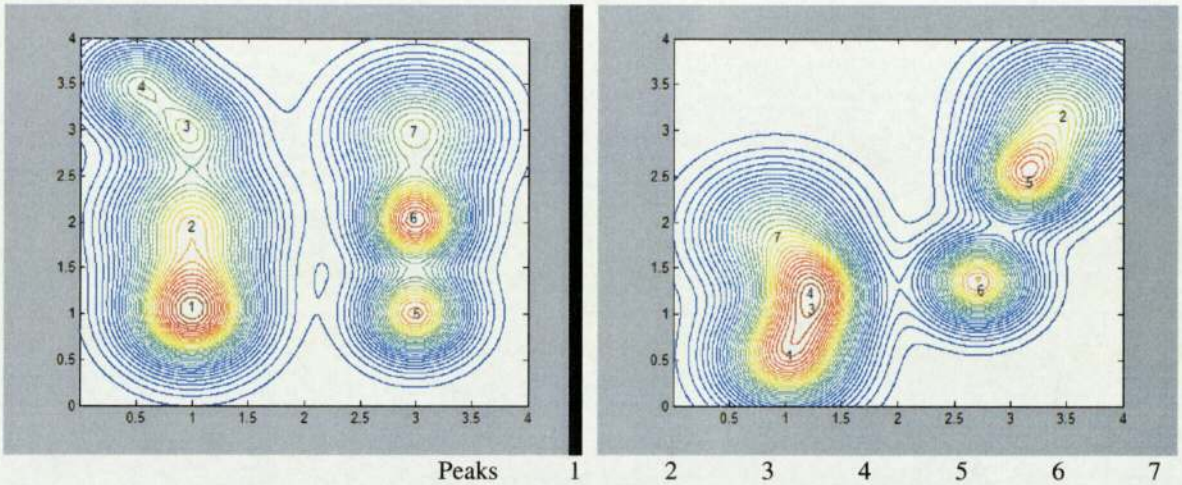
8.2.1 Shifting peaks randomly

Aim:

In that experiment there is only preconception about the rest of the environment apart from its localization, $mean_x^i$ and $mean_y^i$. The aim is to try to obtain an approximation with SOMP by using the subset ($\mathbf{S_L}$) that we obtained from the training phase when there is no information about the location of the peaks.

Procedure:

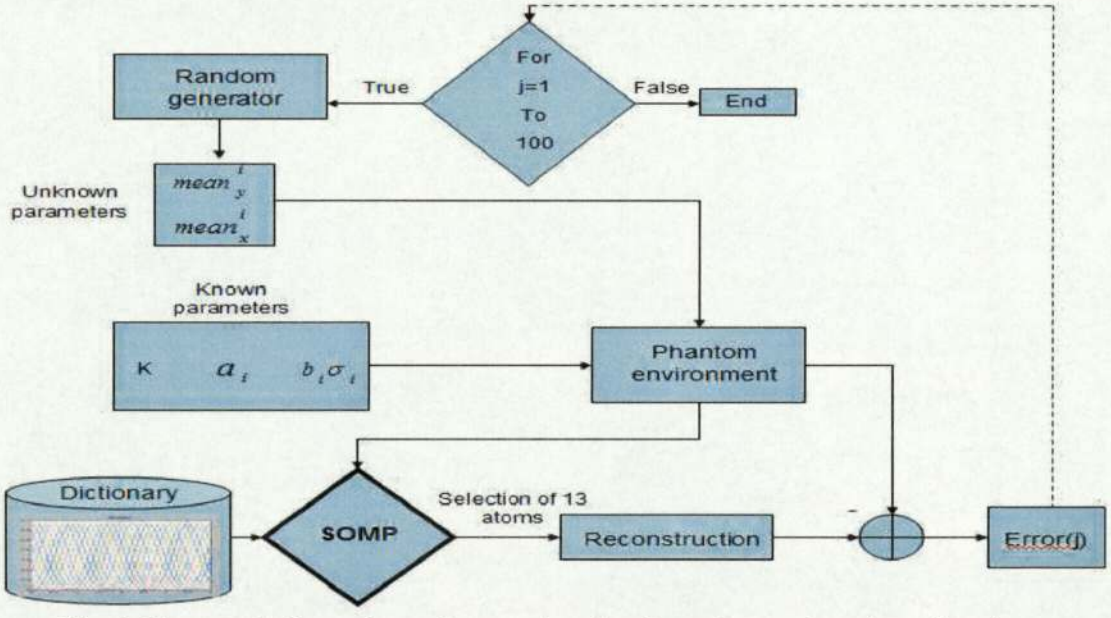
It is assumed the parameters that control the width ($b_i\sigma_i$) and the amplitude (a_i) for each of the seven peaks, respectively in 5.2.1 are known initially. However the location of these peaks into the space ($mean_x^i$ and $mean_y^i$) is unknown. All the peaks are shifted by just choosing random values between the interval [0.5 , 3.5]. In essence, two different vectors of seven values are created, each, in the interval considered. These values represent respectively the location of each peak in xx' and yy' respectively. For this experiment subset (S_L) is used to determine the initial subspace in SOMP in order to represent the target environment at each iteration. Moreover SOMP is iterated for the same environment as input and we try to approximate it without initializing a subset to influence SOMP, but the whole dictionary (D13).For both, 200 vectors of seven values for the location of each peak are initialized, as prescribed above and we continue, at each iteration, with the approximation of the target function by using SOMP. Lastly after each iteration we observe the norm of the error (5.2.2) of the input target and the representation in both cases is observed. Below it is depicted a simple example where there are seven random pairs of values for the location of each peak chosen. Lastly there is also possibility of overlapping of the peaks with each other, due to the localization of them respectively.



Peaks	1	2	3	4	5	6	7
Meanx = 1.0189	3.4392	1.3143	1.2570	3.1272	2.7119	0.9096	
Meany = 0.5353	3.1817	1.0974	1.3962	2.4843	1.3532	1.9077	

Parameters in Target environment:

Known parameters (i=1,…,7)	Unknown parameters (i=1,…,7)
$b_i\sigma_i$	$mean_x^i$
a_i	
K	$mean_y^i$



Block diagram 1: Procedure of approximation for unknown location of peaks

Results:

[fig.48] illustrates the norm of the error of the two different approximations with subset S_L to influence SOMP and without an initial subset to determine the initial sunspace in the decomposition. The mean error of the approximation by using the subset S_L is 0.0097 were as with D13 it is 0.011. In the following plots [fig.47] an example of representations is depicted, for an error value of 0.0089 and 0.013 respectively in both cases.

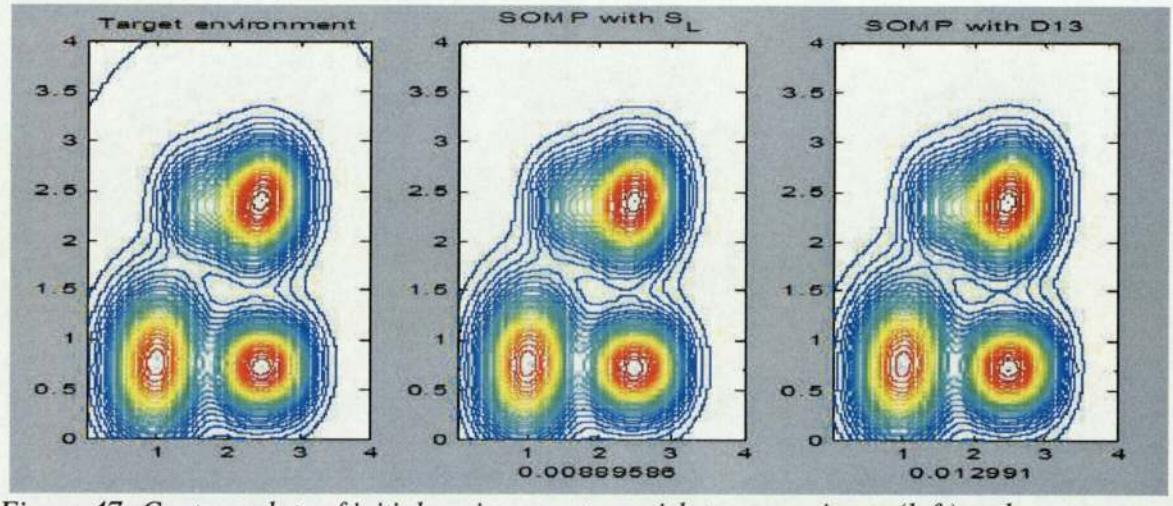


Figure 47: Contour plots of initial environment we wish to approximate (left) and approximation obtained for subset S_L (centre) and thirteen atoms of D13(right) with norm of error 0.009 and 0.013, respectively.

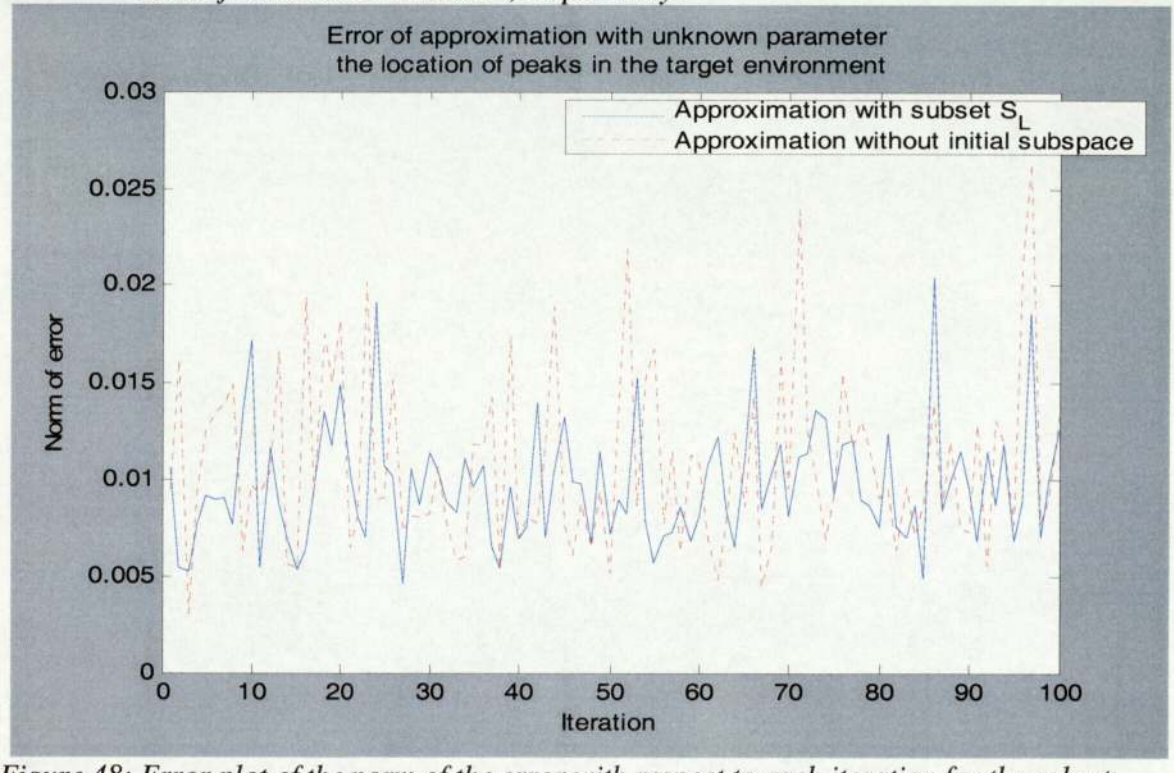


Figure 48: Error plot of the norm of the error with respect to each iteration for the subset obtained from the training phase (blue) and without an initial subset to influence SOMP(red).

Conclusion

From the plots above it results that although both experiments have error of approximation very close to each other, the approximation of the target environment is good. However the representation obtained with the subset S_L is better than the one obtained with no initial subset in SOMP to determine its subspace. The mean error for the former is 0.0097 and for the latter 0.011.Consequently by using such a linear combination of atoms, obtained by the training phase, one can approximate the target well, instead of letting SOMP to decide which

subset to use in the decomposition for each iteration, respectively. Concerning the prior knowledge, there is a good compromise of error using the thirteen elementary signals, which is a good approximation for the location of each of the peaks. Furthermore one should notice that the deviation in the error plot is smaller for the subset S_L.

The effectiveness of the reconstruction formula is of high order, that's due to the dictionary used which has atoms of compact support (4*0.5=2) very close to each peak in total. However one could have a poorer reconstruction if another dictionary or a target environment with narrower peaks was used. The previous is analyzed further in the next experiments.

We understand in that simple experiment we could have certain values for the localization of the peaks where there could be overlapping with other ones or even the phenomenon of vanishing from the environment. In the later case this means that the peak is totally overlapped by another. So, under that circumstances an environment will exist that comprises of six rather than seven peaks, in other words the complexity of it will diminish and, consequently, the approximation would be with a smaller error, instead. The number depicted in the contour plots above is the actual norm of the error for the particular representation.

8.2.2 Change the parameter that controls the width of the peaks

Aim:

The aim of that experiment is to change the parameter that controls the width ($b_i\sigma_i$, i=1,...,7) of each peak, arbitrarily, of the target environment (5.2.1) and observe how well it will be reconstructed from the method that we are using (SOMP) with the subset $\mathbf{S_V}$ that was obtained from the training phase previously. In other words it is considered that everything is known about the input environment we wish to represent apart from the width of each peak.

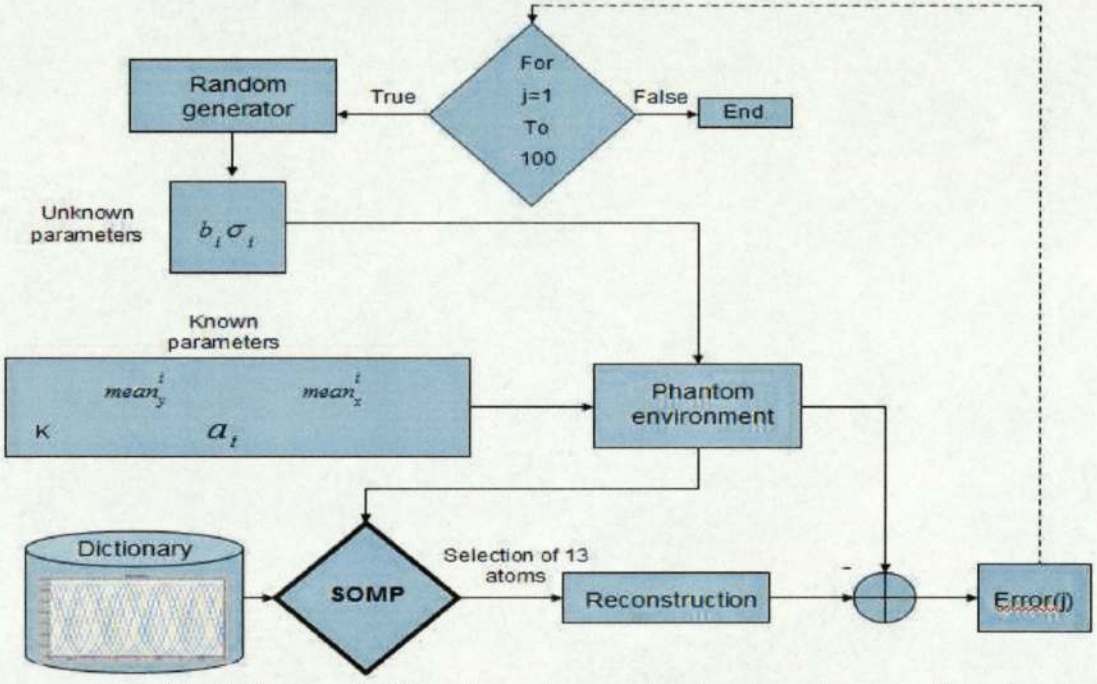
Procedure:

It is assumed that the width of all peaks is changed randomly (the dilation factor is unknown). The environment consists of seven peaks fairly separated. The location ($mean_x^i$ and $mean_y^i$) of all the peaks and their amplitude (a_i), respectively are known, apart from $b_i\sigma_i$ of each one. In other words the unknown factor is the dilation for each peak of the environment. A random set of seven values between the interval [0.2, 0.4] is picked for $(b_i\sigma_i)^2$ of each peak, respectively, and an input target environment (5.2.1) is constructed at each case. That means that we are going to have peaks in the target environment with width from one to two. For

every new set of values ($(b_i\sigma_i)^2$), SOMP is initialized to reconstruct the target. The algorithm is iterated for 100 different sets of values. For this experiment a subset S_V is used that was obtained from the training phase previously. Lastly, after each iteration, the error (5.1.2) of the input target and the approximation obtained is observed. There is also a chance of overlapping between the peaks. In that case too we also approximate the same input environment with SOMP when we do not determine any initial subspace.

Parameters in Target environment:

Known parameters (i=1,…,7)	Unknown parameters (i=1,…,7)
$mean_x^i$ $mean_y^i$ a_i K	$b_i\sigma_i$



Block diagram 2: Procedure of approximation for unknown width of peaks

Results

In [fig.49] below we observe an example of the initial target environment (left) and the approximation (centre) obtained with subset S_V and the one without an initial subspace (right).The norm of the error is respectively 0.006 and 0.01. The norm of the error for the two different cases is depicted in [fig.50].The mean value of the error is 0.005 for the subset and 0.006 without a subset to influence SOMP.

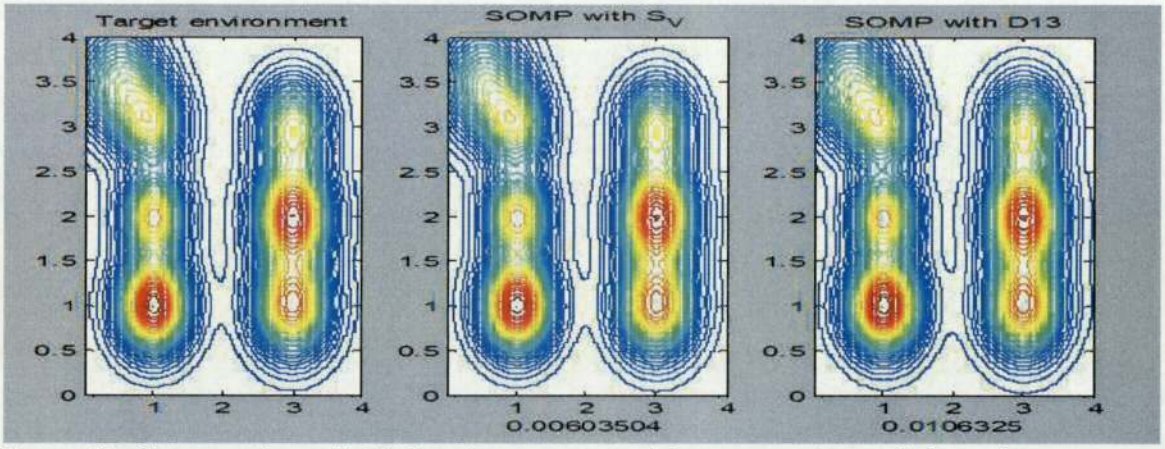


Figure 49: Contour plots of initial environment we wish to approximate (left) and approximation obtained for subset S_V (centre) and thirteen atoms of D13(right) with norm of error 0.006 and 0.01, respectively.

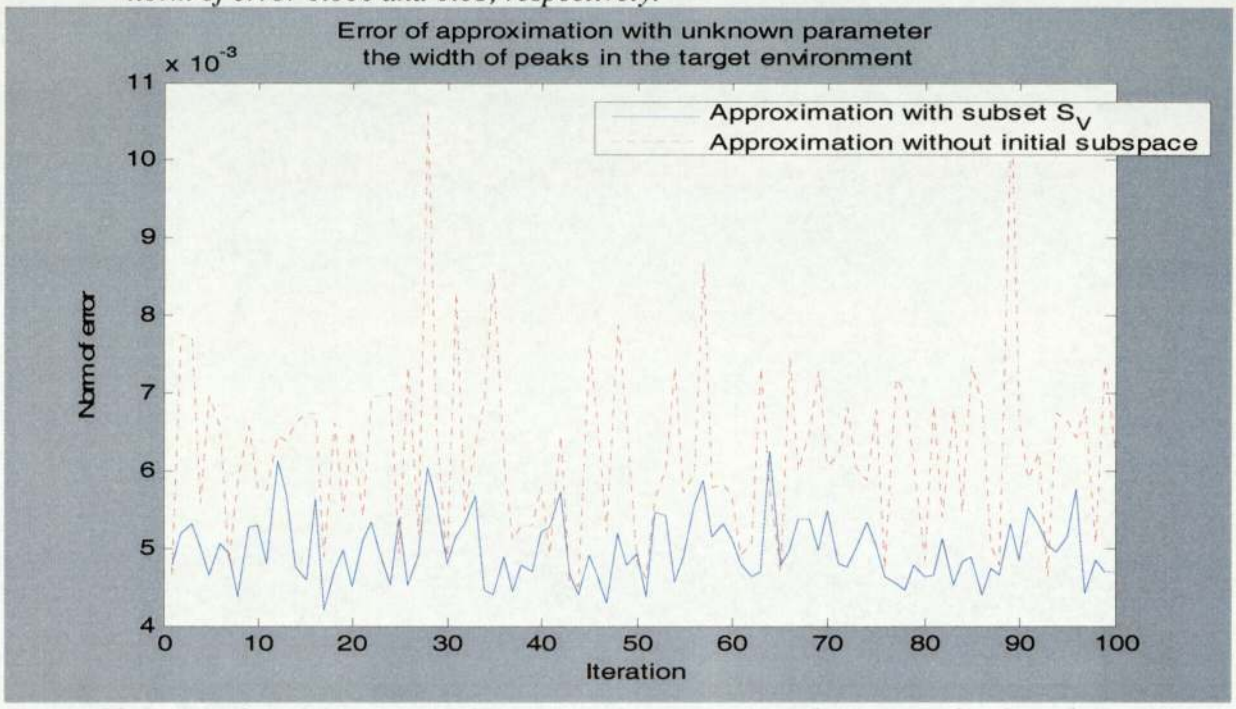


Figure 50: Error plot of the norm of the error with respect to each iteration for the subset obtained from the training phase (blue) and without an initial subset to influence SOMP(red).

Conclusion

In that case too the approximation obtained for both of them is still very good. On the other hand the one obtained by the subset S_V is by far better in proportion to the case where SOMP is not influenced with an initial subspace. In essence the mean norm of the error is 0.006 and 0.01, respectively. Consequently for the particular experiment one could have a particular combination of elementary signals to send in order to represent such a target environment that was assumed.

Apart from the above, one could think that the interval considered to pick these random values for the input environment is too small. We could also try a different interval with peaks

of much narrower base. However by doing so the error might be even greater (worse approximation) due to the fact that we are outside space. The previous means that even if the whole dictionary is used in the decomposition phase, that is 47 atoms, the approximation obtained would still be worse. Hence, in order to approximate such functions one could choose a dictionary of smaller support of B-Spline functions. However that's not the case in the particular experiment and that's the main reason that a target environment that has peaks with width from one to two was used. Even if for an extreme case of having the smallest width for all the peaks the dictionary used can still obtain a good approximation with a small subset of atoms participating in the decomposition.

Like said previously the deviation of the norm of the error for S_V is still smaller than the other one. However let us change the experiment and consider even less prior knowledge for the representation phase.

8.3 Less prior knowledge for the target environment

It was illustrated in the previous experiments how one can obtain a good approximation, by using the particular combination of elementary signals, of the environment by changing only one parameter in the formula of the target environment (5.2.1). However what could happen if there is even fewer information about the target? Such as both the translation and dilation factor or the amplitude and width factors are totally unknown.

Aim:

In the experiments following it is considered even less prior knowledge of the target environment (5.2.1). In the first one, there is known information only about the amplitude of each peak, individually. For the second one, the location and for the last one that we only know the number of peaks. Like stated previously for each of the different experiments we are going to test the subsets of atoms acquired during the training phase. The aim is to illustrate how well the target is going to be approximated, if some partial information over it is assumed.

8.3.1 Randomly change the width and the localization of all peaks

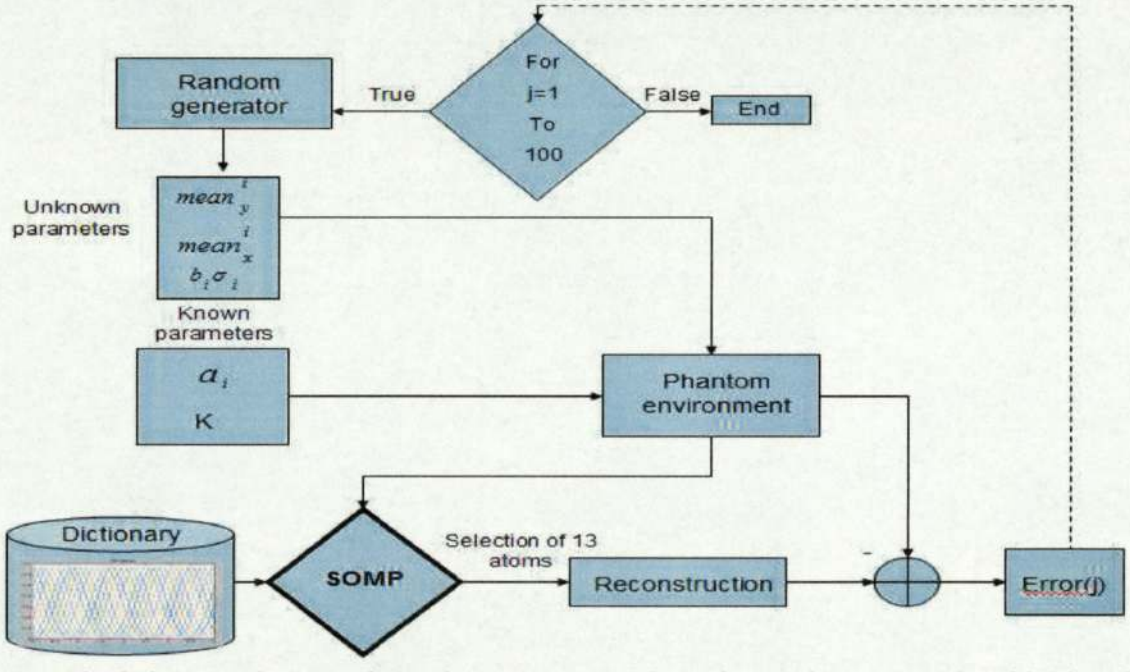
Procedure:

For this experiment it is assumed that there is prior information of the target environment about the amplitude a_i of each of the peaks. Values for the parameter that controls the width between interval [0.2 0.4] and for the location between [0.5 3.5] are going to

be randomly picked. That's due to the same reasons already stated previously. So at each iteration of different values for the location and the width of each peak a target environment (5.2.1) is formed and SOMP is initialized to represent the target. Moreover for the decomposition phase the subset ($\mathbf{S_{V+L}}$) we acquired previously is used. Furthermore SOMP is iterated separately without a subset to determine its initial subspace. There is also a possibility of overlapping while shifting the environment. Lastly after each iteration the error (5.2.2) of the input target and the approximation is observed.

Parameters in Target environment:

Known parameters (i=1,…,7)	Unknown parameters (i=1,…,7)
a_i	$b_i\sigma_i$
K	$mean_x^i$
	$mean_y^i$



Block diagram 3: Procedure of approximation for unknown location and width of peaks

Results:

In [fig.51] below we observe an example of the initial target environment (left) and the approximation when SOMP is influenced with S_{V+L} (centre) and the one obtained when SOMP uses whichever subset of thirteen atoms from D13. Thereafter [fig.52] depicts the error returned from SOMP with respect to each iteration for both cases. The mean error for the former is nearly equal to the latter 0.0096 and 0.0095, respectively.

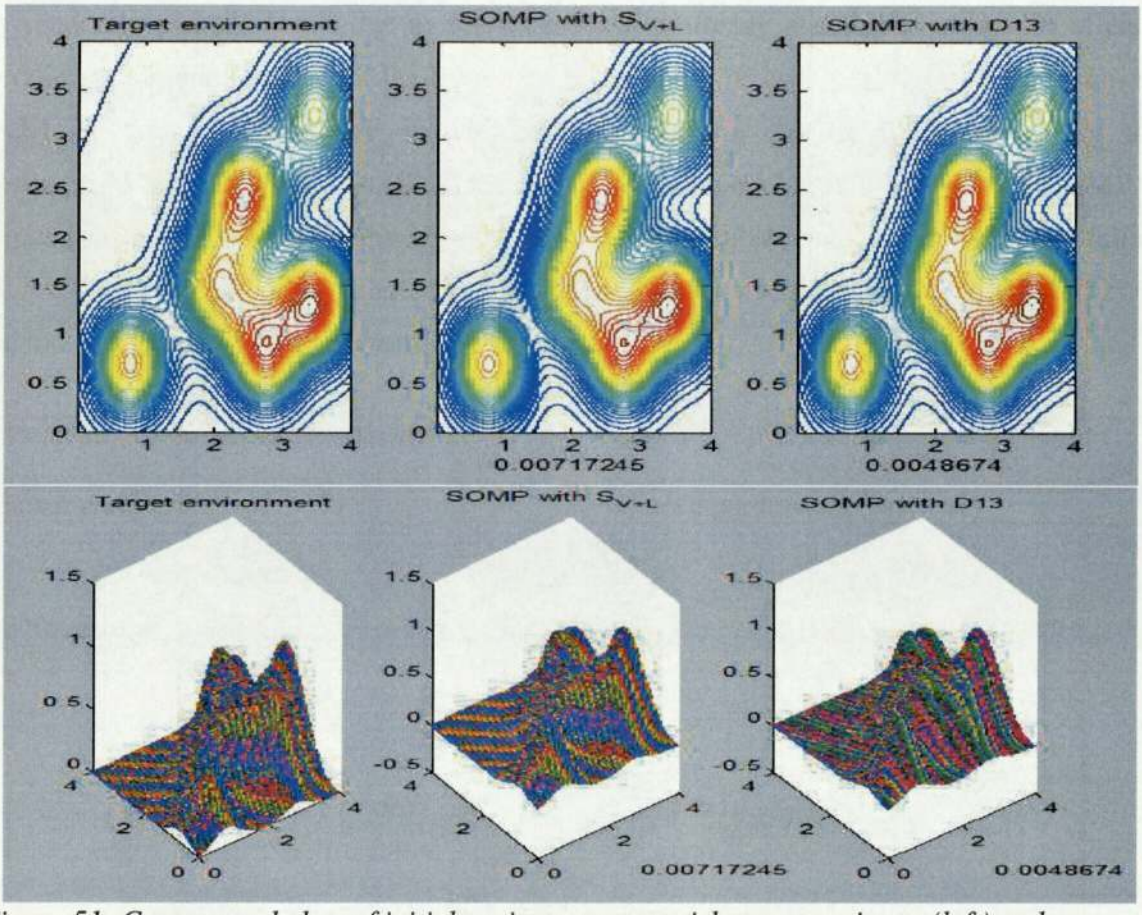


Figure 51: Contour and plots of initial environment we wish to approximate (left) and approximation obtained for subset S_{V+L} (centre) and thirteen atoms of D13(right) with norm of error 0.007 and 0.005, respectively.

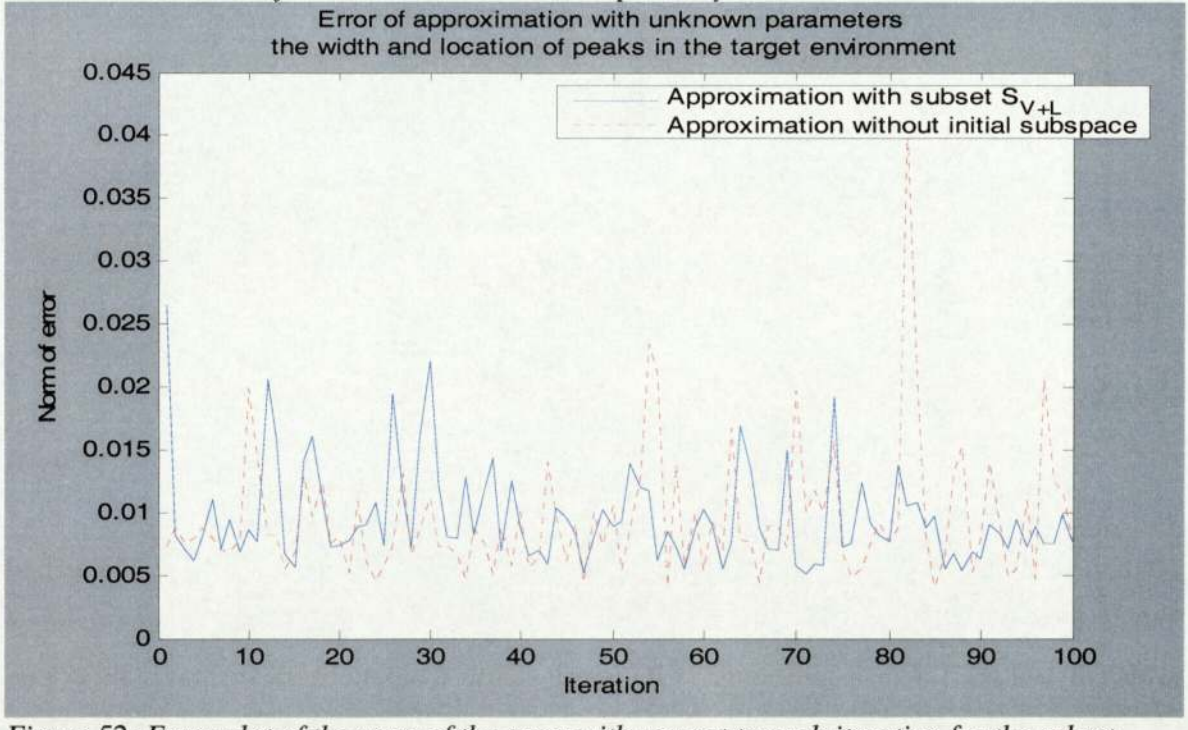


Figure 52: Error plot of the norm of the error with respect to each iteration for the subset obtained from the training phase (blue) and without an initial subset to influence SOMP(red).

Conclusions

Even if the experiment became more complicated and less prior knowledge for the target environment that we wish to approximate was considered there is a good compromise of error and spare approximation for the two different approximations. The mean norm of the error for the subset S_{V+L} and one without an initial subspace are nearly the same. Apparently one could use for such an environment assumed the subset obtained in chapter six for the decomposition.

Concerning the order of error that is of order 10^{-2} it is big due to the fact that, if we consider the previous cases of prior knowledge, the unknown factor of the location of the peaks overwhelms the one of the width in the approximation phase.

Furthermore the same stated previously stands for that experiment too as far as the width of the peaks is concerned. The energy of the whole target environment is going to be preserved to a certain loss, in proportion to an error, and the peaks of the environment are well localized with good approximations of their width, respectively.

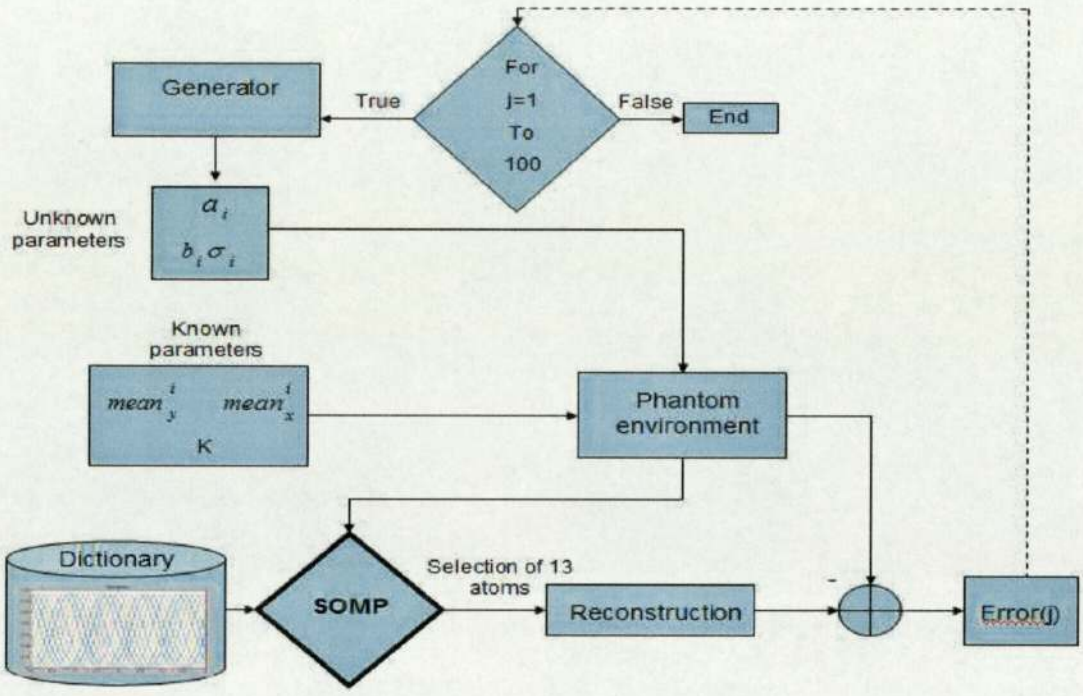
8.3.2 Sharpening peaks

Procedure:

Like stated earlier for this experiment it is assumed that the only information that we have about the environment (5.2.1) are its complexity (number of peaks) and localization. By sharpening a peak we mean that we are going to increase the amplitude factor (a_i) and decrease the width of the each peak. In the particular experiment there is prior information over the location of all of the peaks. However there is no information for $(b_i\sigma_i)^2$ and amplitude (a_i).So we are going to increase a_i from 0.4 to 1.6 and simultaneously decrease $(b_i\sigma_i)^2$ from 0.4 to 0.07, for each peak. For these values obtained the input environment that we wish to approximate with SOMP is constructed at each time. For each case we observe the error returned from the approximation obtained and the input environment. Moreover subset ($\mathbf{S_{V+A}}$) of thirteen atoms is used to determine the initial subspace for SOMP to approximate the input target. Lastly SOMP is iterated without an initial subset in order to compare both procedures.

Parameters in Target environment:

Known parameters (i=1,…,7)	Unknown parameters (i=1,…,7)
$mean_x^i$	$b_i\sigma_i$
$mean_y^i$	a_i
K	



Block diagram 4: Procedure of approximation for unknown amplitude and width of peaks

Results:

[fig.53] depicts an example of the approximation obtained with both cases, an initial subset to determine SOMP (centre) and with thirteen atoms of D13 (right).Also the error plot for both cases is shown in [fig.54].The mean error for the former case is 0.038 and for the latter 0.042.

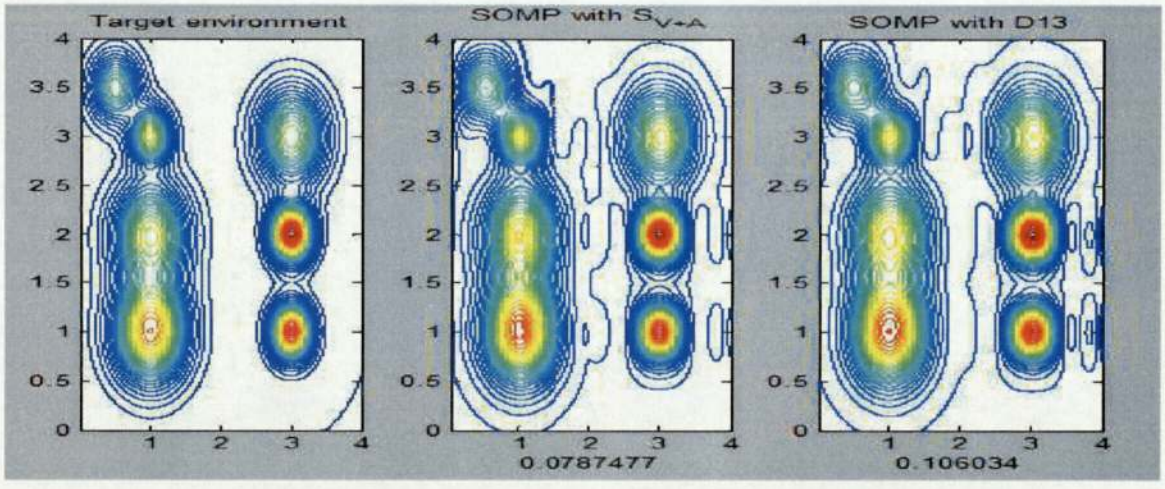


Figure 53: Contour plots of initial environment we wish to approximate (left) and approximation obtained for subset S_{V+A} (centre) and thirteen atoms of D13(right) with norm of error 0.08 and 0.11, respectively.

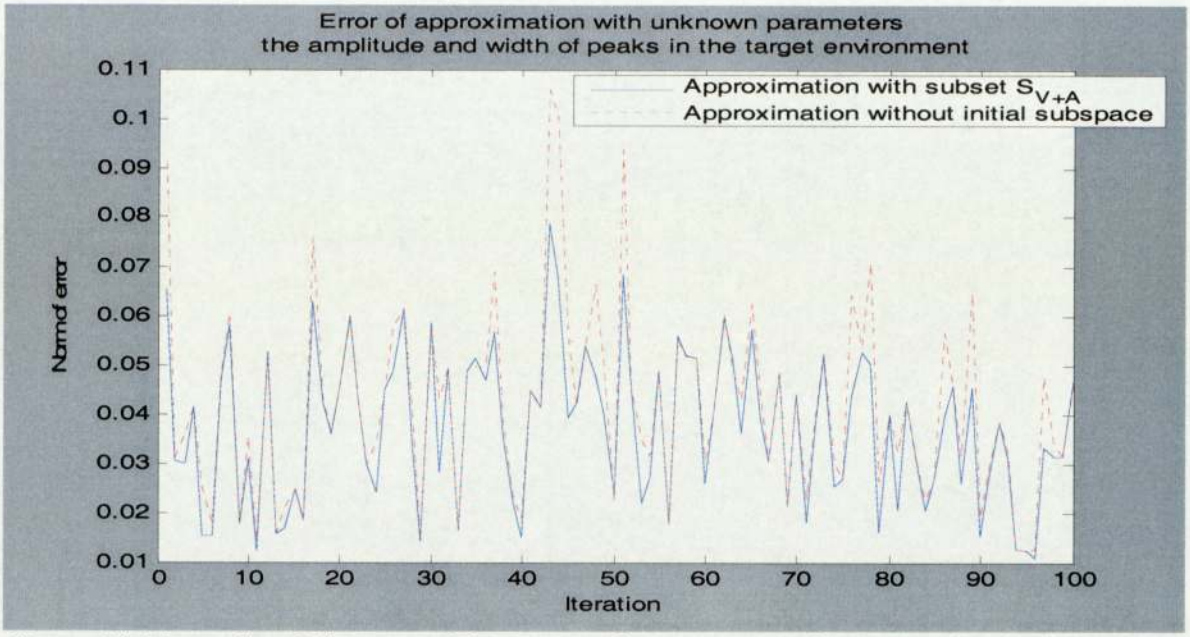


Figure 54: Error plot of the norm of the error with respect to each iteration for the subset obtained from the training phase (blue) and without an initial subset to influence SOMP(red).

Conclusions

In the particular experiment a good representation of error of order 10^{-2} to 10^{-1} was obtained for both cases. However it is obvious from the error plot that although the representations are extremely close to one another, the one obtained by influencing SOMP with the subset S_{V+A} is better. Again the compromise of error with a sparse approximation is of that order due to the fact that the unknown factor of amplitude overwhelms the one of the width of the peaks. Furthermore the energy and the width of the peaks are preserved correctly with respect to an error.

In that case too, a possible solution for an even better approximation would be to use another dictionary or a combination of them with a narrower support of B-Spline functions (width of each function). However let us change the experiment and consider the worst scenario of prior knowledge. In essence let us assume that the only information known about the target environment that we wish to approximate is the number of peaks.

8.4 Knowing only information about the number of peaks.

Aim:

In the previous experiments we witnessed how well SOMP can approximate a target environment when the subsets of atoms obtained as stated in the previous chapter are used. We realize that one can use a particular combination of atoms to approximate a certain target environment that was assumed.

error plot for both cases is shown in [fig.56].The mean error for the former case is 0.06 and for the latter 0.053.

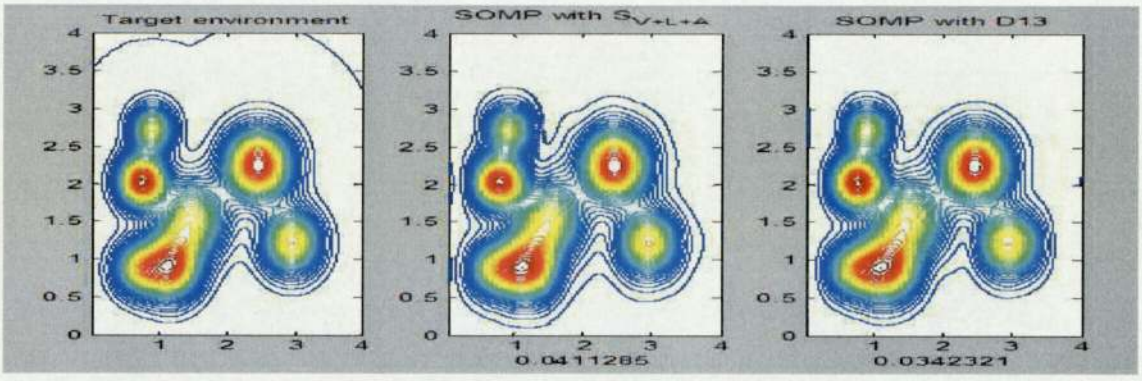


Figure 55: Contour plots of initial environment we wish to approximate (left) and approximation obtained for subset S_{V+L+A} (centre) and thirteen atoms of D13(right) with norm of error 0.04 and 0.034, respectively.

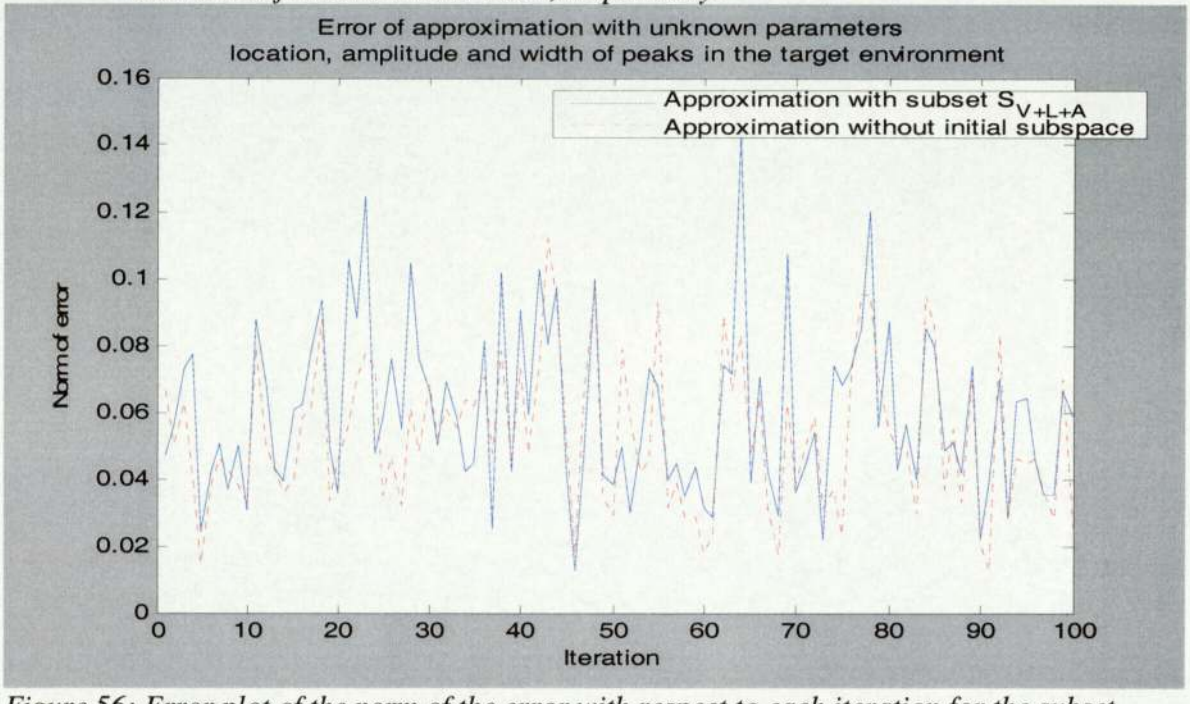


Figure 56: Error plot of the norm of the error with respect to each iteration for the subset obtained from the training phase (blue) and without an initial subset to influence SOMP(red).

Conclusion

Again the approximation obtained for the target environment is still bearable with a compromise of the norm of the error for both cases. The factor that controls the amplitude of the peaks determined the outcome of the order of the error and due to this the representation without an initial subspace had better results which was only 0.07 in comparison to the subset S_{V+L+A}.

The most important fact is that although we have an overwhelmingly complicated environment with such less prior information, the representation obtained with the particular

subset was still very good, for instance this can be depicted in [fig.55].Which means that one could use a particular combination of atoms in SOMP instead of iterating SOMP without determining an initial subspace.
Apparently for this experiment too we could obtain a smaller error in the approximation by just involving another dictionary (with narrower support) or a combination of them.

8.5 General conclusions and remarks

In this chapter the effectiveness of SOMP was investigated when using a particular subset of atoms obtained by the latter chapter from considering prior knowledge of attributes of the target environment we wish to represent and different combinations of them. The most significant conclusion that one can draw from the experiments above is one of a *greedy* nature. In other words the more prior knowledge considered about attributes of the target environment we wish to approximate, the better the approximation obtained. In addition all the subsets used from the training phase proved to yield a good approximation in comparison to the target wished to be approximated. Moreover the norm of error obtained from the representation when an initial subspace of SOMP is determined with a particular subset, was very close to the one that the technique is not influenced by a particular subset.

Furthermore in all experiments the location and width of peaks was preserved correctly with the particular subsets obtained by the training in the previous chapter. In addition the energy of the signal is preserved nearly in total by recourse of an error. Even in the most extreme case that we only have knowledge of the complexity of the environment (number of peaks) and its shape the approximation was obtained with a rather small error.

Concerning the approximation when unknown parameters such as the amplitude and width of peaks is concerned, we discussed that one could obtain an even better approximation with three different ways. The one first was to increase the number of atoms involved in the decomposition. Last but not least, a possible solution would be to apply a different dictionary with a narrower support of B-Spline functions when cases of peaks with smaller widths are involved. Also there is the case of even considering combining dictionaries of different support of B-Spline functions. This is illustrated by a simple example in the next chapter.

In general terms the most important conclusion of all is that the subset of atoms used for the decomposition, obtain an approximation that is very close to the one without involving an initial subspace to influence SOMP with cases that is even better than that one. Lastly the approximation obtained in both cases is good based on the assumptions made for the target environment. Consequently one can use a particular combination of elementary signals, based on prior knowledge of the target environment, which can be propagated in order to visualize a

target environment with a compromise between the norm of the error of the representation and a sparse approximation.

CHAPTER 9

Using a mixed set of elementary functions of different length of the support.

9.1 Introduction

In the previous chapter the subsets, acquired from the training phase previously, were tested in the method of reconstruction for the case of prior knowledge in the input target environment. Moreover there were stated reasons for the failure of SOMP and possible solutions. In this chapter a possible solution is provided through simulations to the case where SOMP might fail.

Particularly in the case of prior knowledge we faced problems with the representation phase at the point where a target environment had too narrow functions in comparison with the dictionary used. In essence the method would fail due to the fact that the functions wished to be reconstructed were not included in the space used for that purpose. So, as a possible solution to the above it was suggested that the approximation can become far better if a dictionary was used in the decomposition that included functions of B-Splines with narrower support. On the other hand due to the morphology of the input target environment, that would increase the number of atoms that SOMP might select in order to approximate it in total and, consequently, increase of computational burdens. Moreover if the same subsets of atoms are to be used as previously, that might produce an approximation that is far worse than the ones previously obtained. This is due to the fact that we chose the particular dictionary of Optimal Support to approximate the input environment in total and not to approximate individually every peak within.

Consequently one could consider combining dictionaries of different support. This will not only produce a better approximation, but also select a small subset in size of atoms for that purpose. The previous is going to be investigated in this chapter. Despite the above, a combination of dictionaries of completely different form, like mixing a Mexican wavelet

dictionary with one of Cardinal B-Splines can be used. Their choice totally depends in the complexity of the target environment wished to be represented.

On the other hand this is beyond the scope of this chapter. The new dictionary that is going to be used will be built of functions that are coming from the same B-Spline basis function. These have different compact support. We realize that the combinations one could use are infinite ones. Unfortunately due to time limitation we have not been able to implement a thorough investigation as to which could be the most probable subset one can use when combining dictionaries. We leave this line of research as one of our proposals of future work. Finally the aim of this chapter is just to illustrate with some numerical calculations the advantage of using a combination of dictionaries of different support at the representation phase.

9.2 Multimodal case

Let us consider a more complicated case of the reconstruction of a peak that gradually becomes separated in twenty five identical others, a multimodal case, of the same attributes. Obviously it's a far more complicated environment than the one previously used and it is almost certain that SOMP will fail. It is assumed that the target environment is known and that comprises of one peak that is localized in x=2 and y=2. This peak is going to be gradually separated in 25 identical peaks with particular localization and the same width. A value is randomly picked between the interval [0.01, 0.4] for the width of the peak. Further on, random values between the interval [0.5, 3.5] are chosen for the location of each peak which are shifted, respectively, in certain directions of the space. The form of the target environment is the (5.2.1). The way we create the environment can be seen in [fig.57] from left to right.

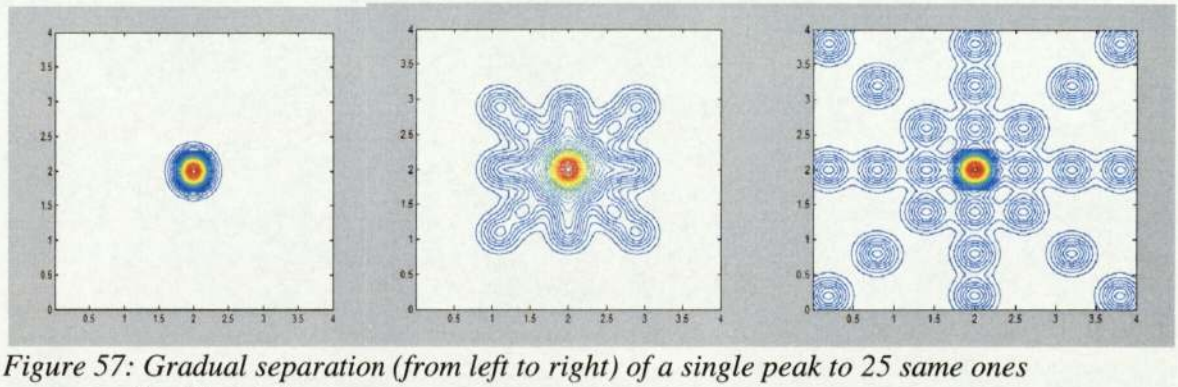


Figure 57: Gradual separation (from left to right) of a single peak to 25 same ones (multimodal).

Then SOMP is initialized for each new environment and the number of atoms to be selected is fixed to 13. In essence SOMP is going to use a subset of thirteen atoms out of the dictionary in order to reconstruct each input environment. Lastly, the norm of the error returned is observed, as described earlier in formula (5.2.2), from the approximation and the

input signal. As in the previous experiments, a dictionary with length of support of B-Splines functions of two is used.

In the [fig.58] below we observe two of the approximations obtained. On the left we observe the input target and its approximation before it becomes separated. On the right side we have the input target and its approximation when all the peaks are totally separated. The error plot of the error returned from SOMP with respect to each iteration is depicted in [fig.59].

As stated earlier although the localization of each peak is preserved correctly the approximation of the width of each function is not so good. Notice in the figures below the "noise" that is created that creates a problem in distinguishing all the peaks that are totally separated at that point. It is obvious that the method failed due to the reasons already stated, however what might happen if we try to combine different dictionaries in order to approximate the same target environment?

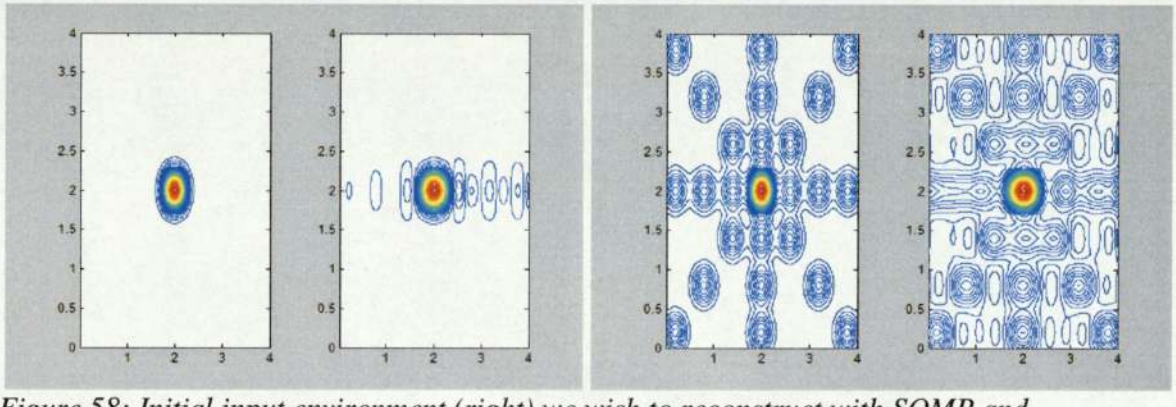


Figure 58: Initial input environment (right) we wish to reconstruct with SOMP and approximation obtained(left) for two different iterations.

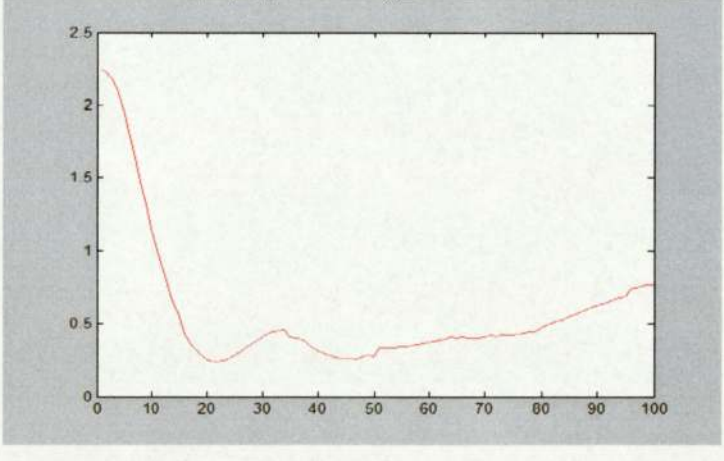


Figure 59: Norm of the error of the approximation obtained with respect to each iteration.

9.3 Combined dictionaries

Aim:

It was stated in the previous experiment how SOMP might fail reconstructing the whole target environment. Moreover in previous experiments we discussed about a possible solution of obtaining a better approximation of the input target instead of increasing the number of atoms used. In essence a better approximation might be obtained by combining dictionaries together.

For the purpose of this experiment three different dictionaries and their combinations are going to be used. The functions in these dictionaries will have the same distance between knots and the same order of B-Splines but they will have different support. In other words we are going to have B-Splines of order four, distance between knots of 2^{-3} and width of 0.5 (D33), one (D23) and two (D13), respectively. Dictionary D33 consists of 35 functions, D23 of 39 and D13 of 47.We should also mention that it is, in fact, a case of increasing the compact support of the basis functions. Moreover for the case of simplicity, their combinations are denoted as D3323 for the mixing of D33 and D23 dictionary, D3313, D2313, and D332313 etc. The attributes of each dictionary can be observed in table [4]. For each different dictionary the previous experiment is repeated and we are going to decide out of which of the dictionaries and their combinations which produces the optimum approximation. For the decomposition a subset of only thirteen atoms is used.

Dictionary	Order of B-Spline functions	Width of functions	Distance between knots	Atoms	Size of subset used for the decomposition	Mean norm of error of approximation
D33	4	0.5	0.125	35	13	0.4176
D23	4	1	0.125	39	13	0.0670
D13	4	2	0.125	47	13	0.5761
D3323	4	0.5 and 1		71	13	0.0668
D3313	4	0.5 and 2		82	13	0.2943
D2313	4	1 and 2		86	13	0.1981
D332313	4	0.5, 1 and 2		121	13	0.2063

Table 4: Concentrated table of attributes and results obtained with each dictionary.

From the error plot [fig.60] it is obvious that if the dictionary D2313 or D332313 is chosen the approximation will be a far better one than using D13.Thats occurring due to the fact that the method is choosing a mixed subset of atoms out of the dictionary. This is depicted in [fig.66-67].

However one might say that there is a better support of functions in order to represent that particular target environment which is D23 with support of B-Splines functions of one.

Without doubt the particular dictionary can approximate the target environment better than the other three dictionaries that we have already considered. Nevertheless if a combination of D23 with D33 is considered there are certain intervals that SOMP will select a subset that includes not only functions of support of one but also of 0.5.This is depicted in [fig.65]. Under these circumstances the approximation is still better than just selecting a subset of functions only from D23 dictionary.

It follows that with the combination of dictionaries a better approximation might be obtained. [fig.61-67] depicts analytically examples of the approximation obtained with SOMP (left) and the signal we wish to approximate (centre). Moreover on the (right) it is shown the subset used each time for the representation.

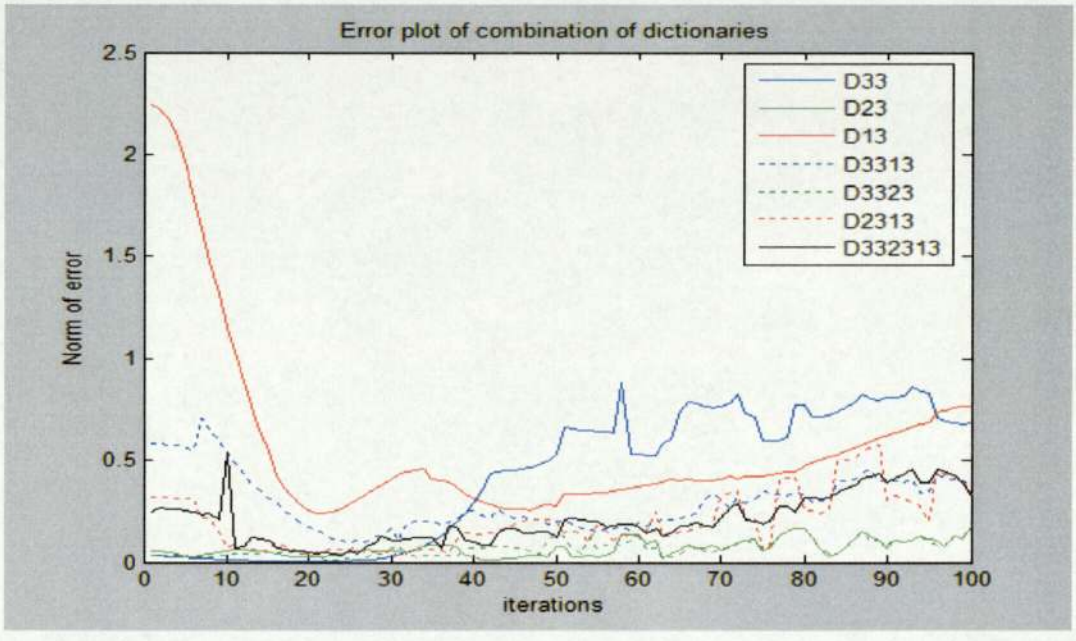


Figure 60: Norm of the error of the approximation obtained with respect to each iteration for every dictionary respectively. The different combinations of each of the three dictionaries are depicted with the dashed line.

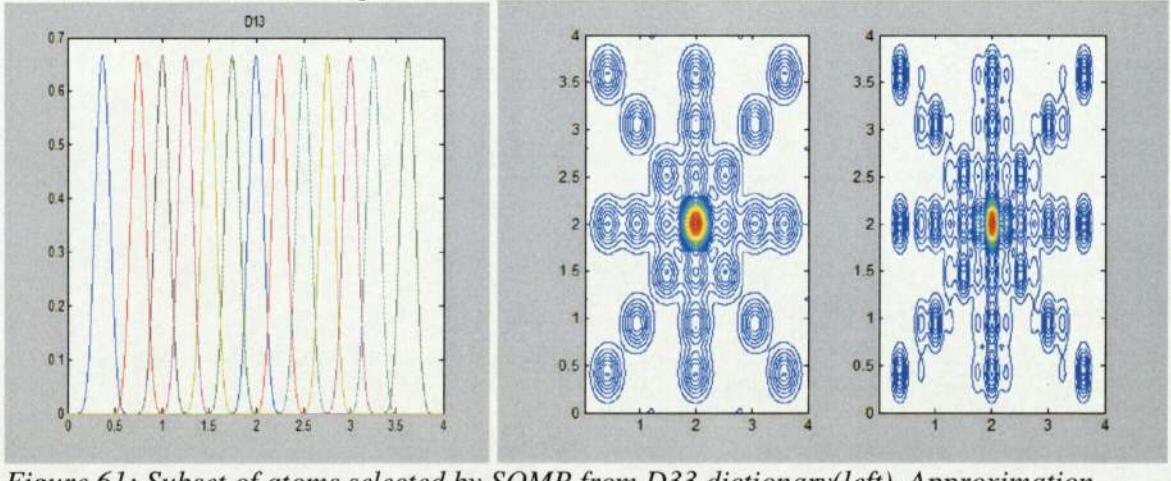


Figure 61: Subset of atoms selected by SOMP from D33 dictionary(left). Approximation (right) obtained and initial environment (centre).

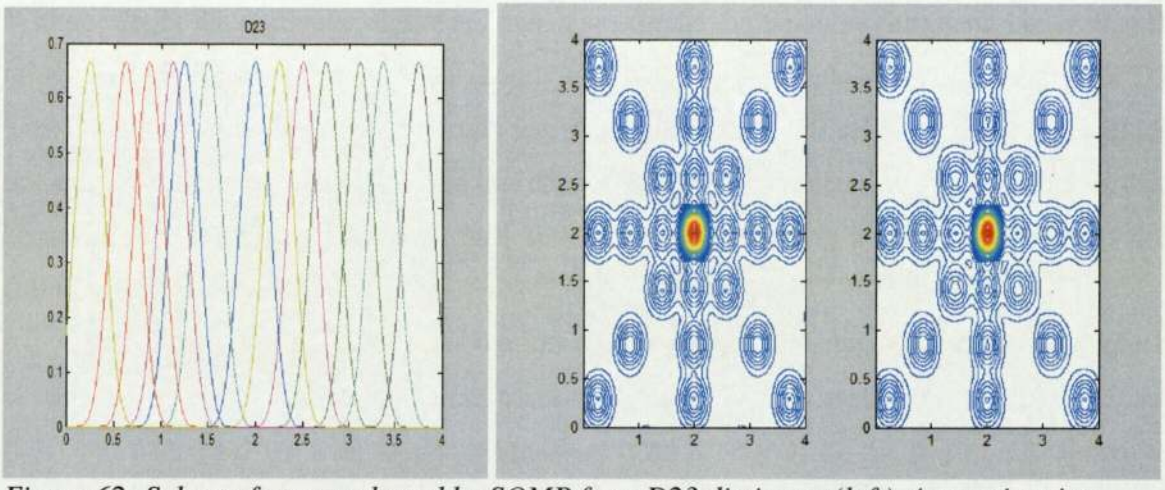


Figure 62: Subset of atoms selected by SOMP from D23 dictionary(left). Approximation (right) obtained and initial environment (centre).

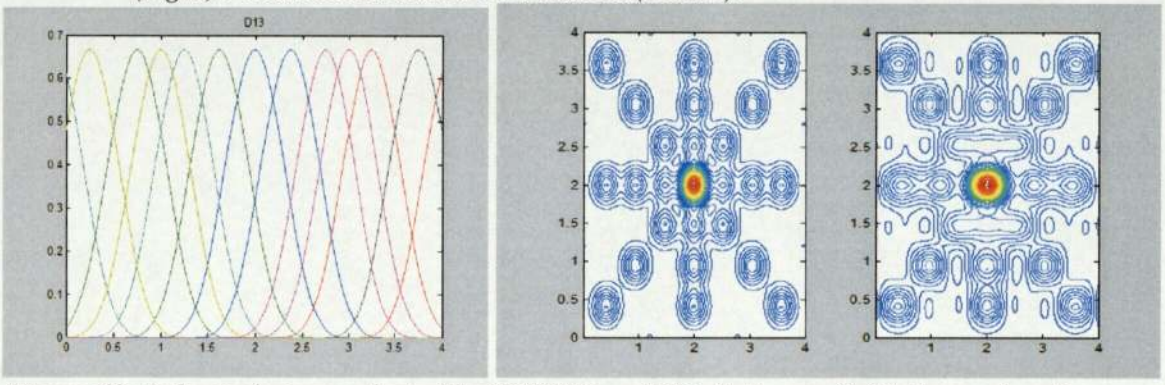


Figure 63: Subset of atoms selected by SOMP from D13 dictionary(left). Approximation (right) obtained and initial environment (centre).

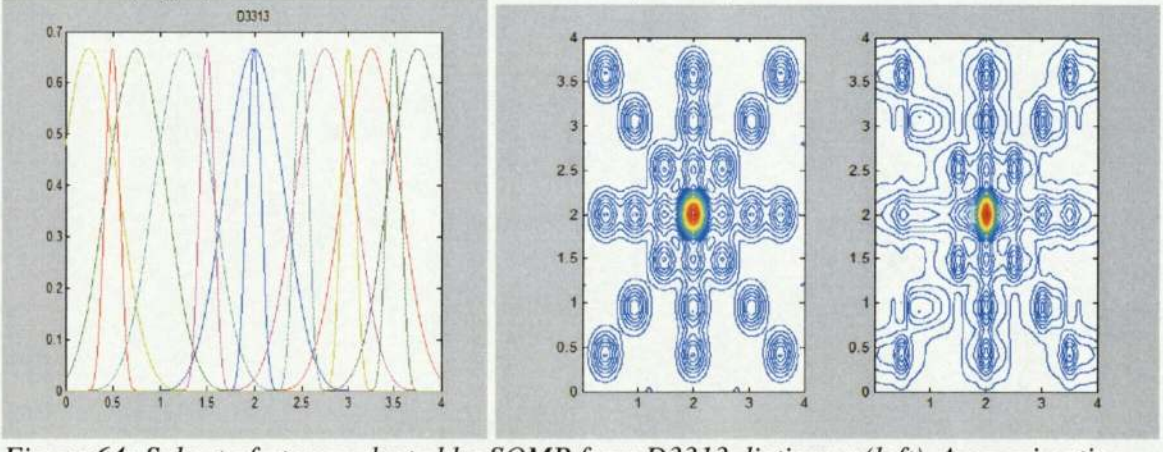


Figure 64: Subset of atoms selected by SOMP from D3313 dictionary(left). Approximation (right) obtained and initial environment (centre).

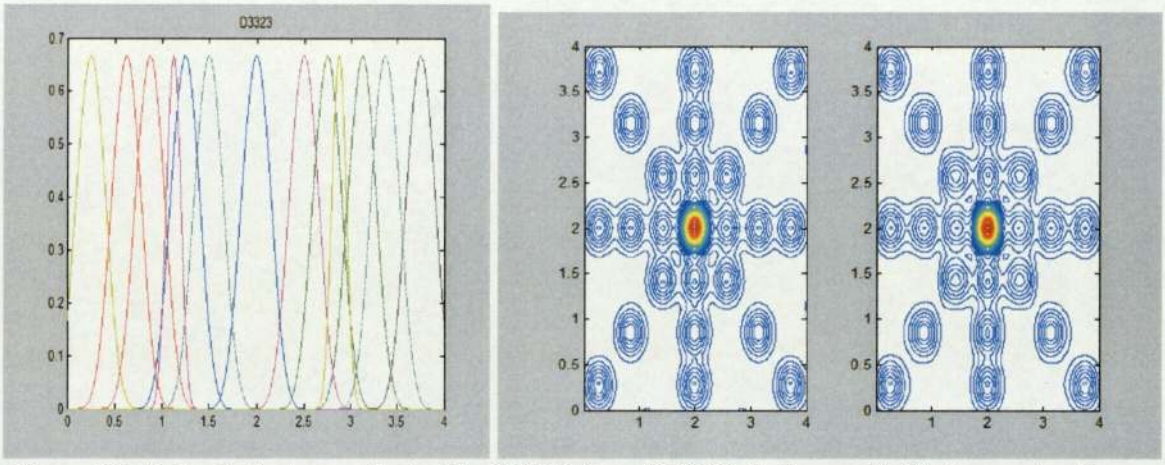


Figure 65: Subset of atoms selected by SOMP from D3323 dictionary(left). Approximation (right) obtained and initial environment (centre).

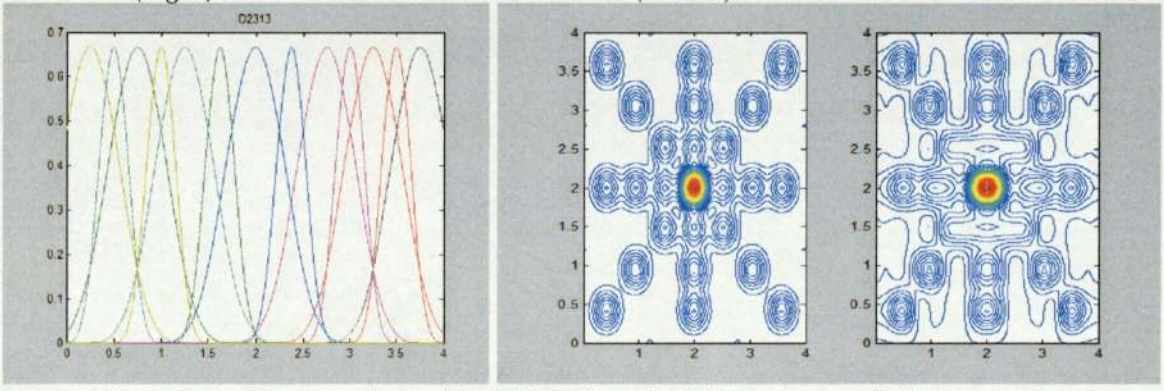


Figure 66: Subset of atoms selected by SOMP from D2313 dictionary(left). Approximation (right) obtained and initial environment (centre).

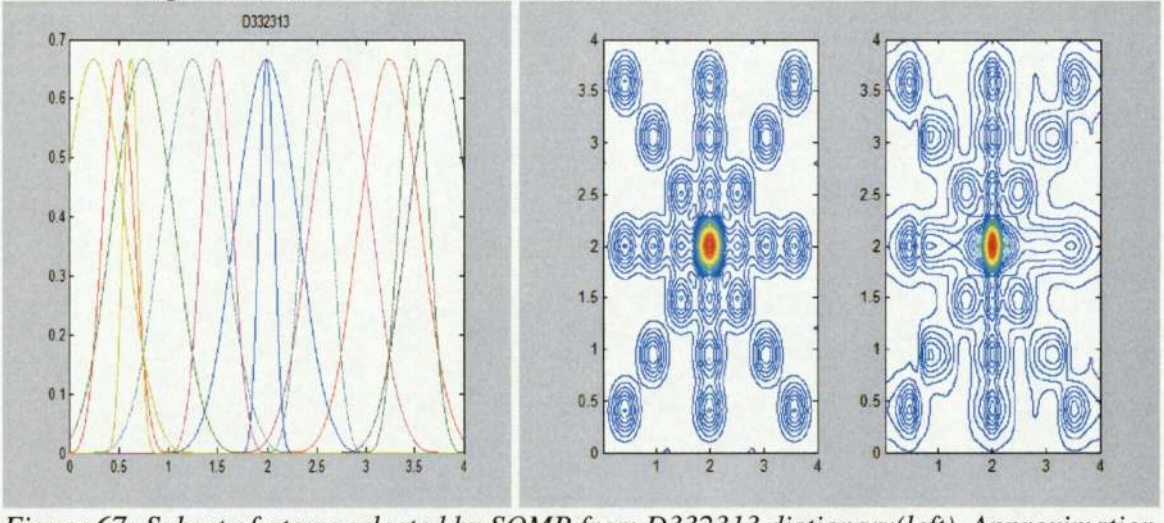


Figure 67: Subset of atoms selected by SOMP from D332313 dictionary(left). Approximation (right) obtained and initial environment (centre).

9.4 Remarks and conclusions

In this chapter it was illustrated a particular case of combining B-Spline functions of different support in order to approximate better a target environment. We considered an environment of one narrow peak that is gradually separated and we tried to approximate every single step of its separation until every peak was totally separated. Firstly we tried the

dictionary that was assumed, previously, to be the one of optimum support for the representation phase and then others of narrower support and their combinations, respectively were considered.

The aim of that chapter was just to illustrate potential proof that one could obtain a sparse approximation with a mixed dictionary. However due to time limitations we weren't capable to perform a thorough investigation in order to find a particular subset of atoms to send. It is clear that there is still much work to be done on this subject. In essence finding an adequate way of selecting atoms from mixed dictionaries would certainly be one of significant importance for the particular purpose.

Chapter 10

Conclusions

10.1 General remarks

The aim of this thesis was divided in two main goals. The first one was to illustrate by some simple examples of how to reconstruct a function of one or two variables by means of elementary functions of one variable selected from a dictionary. The methods used for the above purpose fall into the group of adaptive matching pursuit. Such techniques operate by iteratively selecting a combination of elementary signals chosen from a dictionary of functions, which is created by the translation of a mother basis function in order to participate in a decomposition problem. This selection is performed nonlinearly due to the nature of these algorithms. In most of the experiments performed the Orthogonal Matching Pursuit was used extensively due to purposes of calculation speed and nominal cost of storage. Then it was illustrated ways one can represent a target environment when it is considered as a form of a matrix (function of two variables). The previous was shown in three different ways. By the former we had successive slices of the environment and approximated each of these until the whole environment was represented. By the second one, the best slice was used with respect to the norm of the error of the approximation for the particular one, in order to approximate the rest of the environment. Lastly by the latter the whole matrix was introduced as an input to the algorithm and we obtained the combination of elementary signals (subset of atoms) to take part in the decomposition simultaneously. The last technique is called Simultaneously Orthogonal Matching Pursuit and was used finally for the second goal of the thesis. In addition a simple numerical example was performed to illustrate the advantages one can get from combining recursively a backward technique (BOOMP) with a forward one (OMP or OOMP), which might result in a more sparse approximation with respect to a smaller error. Due to time

limitation we have not been able to illustrate this through a thorough investigation of their combination, a thing that is considered as one of our proposals of future work.

The latter objective of the thesis was twofold. In the first part an adequate dictionary was to be found for representing a completely known target environment. The latter objective was to assume partial information, priori, of the target environment and try to represent it with an *optimal* dictionary (previously obtained). Our experiments were particularly focused on using elementary functions of Cardinal B-Splines. In a few words, firstly the space that we are going to work with in the representation phase was fixed, by assuming a perfectly known target environment. Those were attributes of the dictionary obtained such as the order of B-Splines, the distance between knots, the length of the B-Spline functions and the size of the subset that is capable to represent such an environment of the complexity, that was assumed, with respect to a compromise between the norm of the error of the approximation and a sparse representation (least selected atoms). The latter was managed by increasing the support of the B-Spline functions which resulted in dictionaries consisting of even more atoms. In addition, as soon as the space was obtained, we assumed for the same target environment different combinations of prior knowledge, that is knowing certain attributes of the target wished to be approximated priori, and through a certain procedure of training we obtained particular combinations of atoms to solve each approximation problem respectively. Thereafter the particular subsets obtained were tested through a particular random test set of values for the unknown parameters of the target environment in order to show whether or not one can have a good compromise of error and a sparse. We found the results obtained most interesting and it was concluded that the more prior knowledge one assumed about the signal's properties the better the approximation. Furthermore solutions were provided for a better approximation. These were, increasing the atoms involved in the decomposition phase or choosing a combination of dictionaries of B-Spline functions to work with. Unfortunately due to time limitations evidence was not produced of using a particular subset of such a combination, however we illustrated the advantages of a possible solution to the above through a simple example of approximation. The previous task should be left as one of our main proposals of future work.

Concluding what was illustrated through the course of the particular experiments was a rather obvious remark. In essence the first one is obtaining a better approximation when one assumed, a priori, more features of the signal wished to be represented. The other one would be to increase the size of the subset used for the decomposition in order to have a better compromise between the norm of the error of the approximation and a sparse representation. That's due to the means the adaptive techniques select the combination of elementary signals for the decomposition which is of greedy nature. Lastly, to use a more complicated form of a

dictionary involving mixed functions in order to have an even better approximation. On the other hand evidence was produced of a good compromise between error and a sparse representation if one assumed prior knowledge both for the signals sent to approximate a dense group of targets and attributes of the target environment known beforehand. In essence, the subsets obtained from the training phase proved to obtain a good approximation if such a shape of a target is assumed. Consequently one can determine an initial subset to influence SOMP in order to approximate these target environments, without letting the approach, used in the decomposition, to choose another one.

10.2 Future work proposals

Based on the example illustrated with BOOMP it came up by combining strategies a remarkable attribute of these techniques. That is we can obtain further compression and better approximation of the target environment. So what is considered as a future work proposal is to try to combine forward and backward strategies in order to design new techniques for adaptive signal representation.

Moreover looking for further criteria in choosing atoms from the dictionary or the combination of them would be another one. It is even considered applying an architecture of artificial neural networks that are used extensively for signal processing. These nonlinear procedures are said to yield good representations and reconstructions with even fewer computational steps, than the traditional ones, when applied. Some of the most popular ones are MLP and RBF Networks.

Furthermore another it would worth investigating in obtaining a particular combination of elementary signals from mixing dictionaries, in order to approximate a signal. For instance, consider using a mixed dictionary of Mexican wavelets with Cardinal B-Splines.

However, in the above analysis a rather deterministic model was considered. In other words it is known that in a classical radar problem the signal is corrupted with noise. Something due to time limitations wasn't considered in our experiments. Moreover there is also the parameter, in the radar equation, that represents the reflectivity of the density of the targets which was considered as one (a perfectly reflecting object). If one considers that partial energy of the signals emitted from the source is absorbed by the target environment, instead of being reflected in total, can we still yield a good approximation?

Lastly from the literature review it was noticed that these adaptive methods can also be used for image representation, restoration and reconstruction. It might be interesting and there is great potential to use the particular approaches in order to detect edges of an image, compression or even obtaining, in that case too, a possible subset out of the dictionary that will come up with bearable approximations of an image.

Appendix A

A.1 Frames

Let $\{e_m; m \in N\}$ be the standard orthonormal basis in $l^2(N)$, ($\langle e_n, e_m \rangle_{l^2} = \delta_{n,m}$), where $\langle .,. \rangle_{l^2}$ denotes the inner product in $l^2(N)$. Every vector $c \in l^2(N)$ can be expressed as

$$c = \sum_{m \in N} e_m \langle e_m, c \rangle_{l^2} \qquad \text{(a.1)}$$

Frame definition:

A family of functions $\{g_m \in L^2(R); m \in N\}$ is called a *frame* if for every $f \in L^2(R)$ there exists a pair of constants $0 < A \leq B < \infty$ such that

$$A\langle f, f \rangle_{L^2} \leq \sum_{m \in N} \left| \langle g_m, f \rangle_{L^2} \right|^2 \leq B\langle f, f \rangle_{L^2} \qquad \text{(a.2)}$$

A, B are called *frame bounds* and the above is the *frame condition*.

Derivation of the reciprocal functions:

Adapting the notation in [14], if we built up a vector b such that $b = \sum_{m \in N} e_m \langle g_m, f \rangle_{L^2}$ then $\|b\|_{l^2}^2 = \langle b, b \rangle_{l^2} = \sum_{m \in N} \left| \langle g_m, f \rangle_{L^2} \right|^2$ where the frame condition implies that $b \in l^2(N)$ whenever $f \in L^2(R)$. Thus we can define for a mapping of $L^2(R)$ to $l^2(N)$ an operator $\hat{T}$ called the *frame operator* which can be written in the following form

$$\hat{T} = \sum_{m \in N} e_m \langle g_m, . \rangle_{L^2} \qquad \text{(a.3)}$$

Consequently for f we will have $\hat{T}f = \sum_{m\in N} e_m \langle g_m, f\rangle_{L^2}$ and the adjoin operator of $l^2(N)$ to $L^2(R)$ will be $\hat{T}^+ f = \sum_{m\in N} g_m \langle e_m, f\rangle_{l^2}$ .So expressing the frame condition in terms of an operator $\hat{G} = \hat{T}^+\hat{T} = \sum_{m\in N} g_m \langle g_m, .\rangle_{L^2} : L^2(R) \to L^2(R)$ we will have $A\hat{I}_{L^2} \le \hat{G} \le B\hat{I}_{L^2}$, where $\hat{I}_{L^2}$ is the unity operator in $L^2(R)$. From the previous, one can obtain the bounded $\hat{G}^{-1}$ inverse in the following manner

$$B^{-1}\hat{I}_{L^2} \le \hat{G}^{-1} \le A^{-1}\hat{I}_{L^2} \qquad \text{(a.4)}$$

$g^m \in L^2(R); m \in N$ is the *reciprocal frame* which is computed by $g^m = \hat{G}^{-1} g_m$ and provides a decomposition of the unity operator in the following manner

$$\hat{G}^{-1}\hat{G} = \hat{G}\hat{G}^{-1} = \hat{I}_{L^2} \text{ where } \hat{I}_{L^2} = \sum_{m\in N} g^m \langle g_m, .\rangle_{L^2} = \sum_{m\in N} g_m \langle g^m, .\rangle_{L^2} \qquad \text{(a.5)}$$

For the case where A=B=1 and $\|g_m\| = 1$, $\hat{G}$ is the identity that is $g^m = g_m$ and the functions g_m are orthonormal. Following the above, the identity $f = \hat{I}_{L^2} f$ expands to

$$f = \hat{I}_{L^2} f = \sum_{m\in N} g^m \langle g_m, f\rangle_{L^2} = \sum_{m\in N} g_m \langle g^m, f\rangle_{L^2} \qquad \text{(a.6)}$$

and its complex conjugate is of the form below

$$\overline{f} = \hat{I}_{L^2} \overline{f} = \sum_{m\in N} \overline{g^m} \langle f, g_m\rangle_{L^2} = \sum_{m\in N} \overline{g_m} \langle f, g^m\rangle_{L^2} \qquad \text{(a.7)}$$

The above implies that f can be recovered either by $\langle g_m, f\rangle_{L^2}$ or $\langle g^m, f\rangle_{L^2}$. If there exists many different set of the reciprocal functions which give rise to the decomposition of the unity operator in terms of the same frame. These are going to be used in order to reconstruct the spreading function $p(\tau, \phi)$, by building a decomposition of the unity operator from the echoes produced.

List of Figures

List of Tables

List of Block Diagrams

Bibliography

[1] C. E. Cook, M. Bernfeld, Radar signals, New York, Academic press, (1967)

[2] S. Haykin, "Communication systems",4th edition, John Wiley & Sons, Inc. 2001

[3] H. Naparst, "Dense Target signal processing", IEEE Trans. Information theory 37,2, 317-327, (1991)

[4] S. Mallat, "A wavelet tour of Signal Processing", Academic Press, Second edition (1998)

[5] H. Naparst, "Radar signal choice and Processing for Dense Target Environment".(PhD Thesis, University of California, Berkeley,1988)

[6] Y.C. Pati, R. Rezaiifar, and P.S. Krishnaprasad, "Orthogonal matching pursuits: recursive function approximation with applications to wavelet decomposition", Proceedings of the 27th Asilomar Conference in Signals, Systems and computers 1993.

[7] S. Mallat and Z. Zhang, "Matching Pursuit in time-frequency dictionary", IEEE Transactions on Signal Processing, Vol 41, 3397-3415 (1993)

[8] L. Rebollo-Neira, and D. Lowe, "Optimised Orthogonal Matching Pursuit Approach", IEEE Signal processing letters, Vol 9,4, 137-140 (2002)

[9] L. Rebollo-Neira, Miroslav Andrle, "A swapping-based refinement of orthogonal matching pursuit strategies", Signal processing, vol.86, pp.480-495 2006

[10] L. Rebollo-Neira,M. Andrle,E.Sagianos, "Backward-optimized orthogonal matching pursuit approach", IEEE Signal process. Lett. 11 (9) (2004) 705-708

[11] R.R. Coifman, M. V. Wicherhauser. "Entropy-based algorithms for best-basis selection", IEEE Transactions on information theory, Vol 38, pp 713-718 (1992)

[12] G.Kaiser, A Friendly Guide to Wavelets (Birkhauser, Berlin, 1994)

[13] G. Kaiser, "Physical Wavelets and Radar , a Variational Approach to Remote Sensing", IEEE Antennas and Propagation Magazine, 38,1, 15-25 (1996)

[14] L. Rebollo-Neira, J. Fernandez-Rubio, "A frame theory based approach for determining the time-frequency distribution characterising a dense group of reflecting objects", vol. 41, no. 8, pp.5325-5341, 2000

[15] L. Rebollo-Neira, J. Fernandez Rubio, On The Inverse Windowed Fourier Transform. *IEEE Trans. Information Theory* , Vol(45, 7), 2668-2671, (1999).

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

[17] I. Daubechies, Time-frequency localization operators: a geometric phase space approach, IEEE Transactions on Information Theory, 34 (1988), pp. 605-612.

[18] R.R. Coifman, M. V. Wicherhauser. "Entropy-based algorithms for best-basis selection", IEEE Transactions on information theory, Vol 38, pp 713-718 (1992)

[19] I. Daubechies, Ten lectures on wavelets (Society for industrial and applied mathematics, Philadelphia, 1992)

[20] L. Rebollo-Neira, M. Andrle, "Cardinal B-splines dictionaries on a compact interval". Appl. Comput. Harmon. Anal. 18 (3) (2005) 336-346

[21] L. L. Schumaker, Spline Functions: Basic theory, Wiley, New-York,1981

[22] L. Rebollo-Neira, "Recursive biorthogonalization approach and orthogonal projectors", math ph/0209026, 2002.

[23] J. A. Tropp, A. C. Gilbert, M. J. Strauss, "Algorithms for simultaneous sparse approximation, Part I: Greedy pursuit", *Signal Processing*, special issue 86 April (2006) 572-588.