國 立 交 通 大 學

網路工程研究所

碩 士 論 文

基於標準與特別演繹樣本訊號元素之稀疏訊號處理法

於腦電圖中觸發電位與獨立成份訊號之分析

Sparse Signal Representation of EEG and ICA Components

Based on Standard and Specially Trained Dictionaries

研 究 生： 邱俊瑋

指導教授： 邵家健　　教授

中 華 民 國 一 百 年 二 月

基於標準與特別演繹樣本訊號元素之稀疏訊號處理法

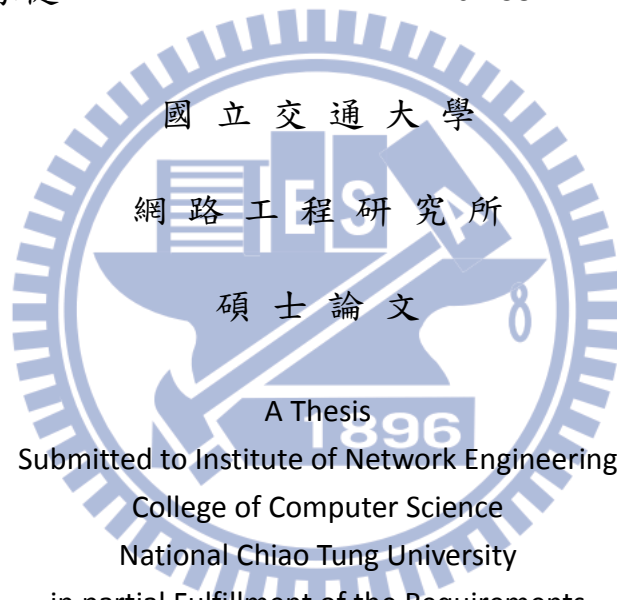於腦電圖中觸發電位與獨立成份訊號之分析

Sparse Signal Representation of EEG and ICA Components

Based on Standard and Specially Trained Dictionaries

研 究 生：邱俊瑋　　　　　　　　Student: Jun-wei Qiu

指導教授：邵家健　　　　　　　　Advisor: Dr. John Kar-kin Zao

國 立 交 通 大 學

網 路 工 程 研 究 所

碩 士 論 文

A Thesis
Submitted to Institute of Network Engineering
College of Computer Science
National Chiao Tung University
in partial Fulfillment of the Requirements
for the Degree of
Master
in

Computer Science

June 2010

Hsinchu, Taiwan, Republic of China

中 華 民 國 一 百 年 二 月

基於標準與特別演繹樣本訊號元素之稀疏訊號處理法

於腦電圖中觸發電位與獨立成份訊號之分析

學生：邱俊瑋　　　　　　　　指導教授：邵家健 博士

國立交通大學網路工程研究所碩士班

中文論文摘要

稀疏信號在選定某個恰當的訊號樣本庫之下，在使用其中的訊號元素拆解後有很大一部分的訊號能量會集中於某幾個少數的元素，此種訊號表達方式可以使用遠低於 Nyquist 訊號取樣限制的資料量來表達一個稀疏訊號，或擷取出其中占大部分能量的訊號成分。稀疏信號處理是應用數學與稀疏表達法的交集結果。也是未來實現遠距醫療與病患檢測上的一項重要技術，將會大大地降低儲存訊號所使用的資料量。

探討如何選用一個能僅用少量元素即可拆解一訊號大部分能量的訊號樣本庫在稀疏訊號處理上是非常重要的。在可使用時間、頻率、規模參數來控制樣本庫中訊號的波形的完備 Gabor 樣本庫已知為一種能有效率地拆解腦電圖儀（EEG）的訊號樣本庫。

本文將著重介紹一個基於 Matching Pursuit ，並配合改進後的 Grassmannian 參數取樣方式以及自然梯度設計而出的文化基因演算法。

Sparse Signal Representation of EEG and ICA Components Based on Standard and Specially Trained Dictionaries

Student: Jun-wei Qiu   Advisor: Dr. John Kar-kin Zao

Institute of Network Engineering

College of Computer Science

National Chiao Tung University

Abstract

A sparse signal has a large portion of its energy contained in a small number of coefficients, which can be represented with number of samples significantly beneath the Nyquist's criterion. Sparse signal processing is the application of the mathematics of sparse representations in signal processing. This is essential for ubiquitous medicine and healthcare due to the immense amount of data required for each individual patient monitoring.

A representation dictionary for the target signal space with only minimal number of supports is critical to the stability and consistency of the sparse signal processing schema when employing pursuit algorithms. The overcomplete Gabor dictionary, which can be parameterized into time, frequency, scale and phase, is already known to have efficient representation for electroencephalograms (EEGs).

This thesis includes a Memetic Algorithm based on Matching Pursuit for EEG signal decompositions, and an improved parameter sampling and optimization method for the Gabor dictionary based on natural gradient and Grassmannian concepts.

Acknowledgement

First appreciation is to my dearest members of the family; their selfless devotion undoubtedly provides the most sincere support during the period of my academic career far away from home, both physical and mental alike.

Secondly, in these days of my research, the instructors — John Kar-kin Zao and Peng-hua Wang — always encouraged me to overcome the difficulties which I originally thought to be insurmountable. This research topic is truly a challenge for my study in my college years, especially the part of digital signal processing and high dimensional information geometry, which are entirely whole brand new areas to me since I've never taken any related lectures during my college years. Their untiring and earnest advices have guided me through various obstacles. No achievement would have come out without their professional and genuine instructions.

There's a special thanks to Alvin Chou, his profound knowledge in clinical and biomedical engineering provides the most recent information and related technologies to the research team; his industrious and irreplaceable works have saved us a lot of time getting familiar with the state-of-the-art biomedical analysis tools during the research period.

Finally, I must thank my former roommate Ting-yu, Tse-wei and Sz-Hsien; the members of the softball team such as Yun-Hsiu, Shu-hua, Tz-Chaio and Zhe-wei. Those who stand with me and share my emotion just like my brothers should take the final prize. It's really a pleasure to have these valuable companions in such a fascinating and colorful journey.
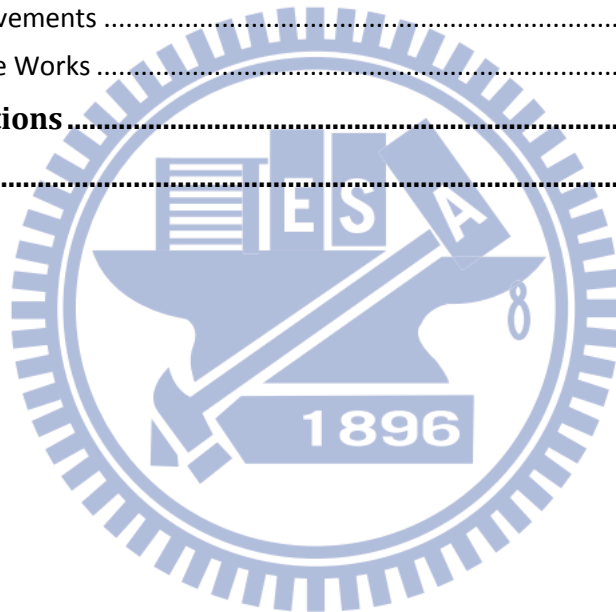
Jun'wei Qiu

Spring, 2011

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

1.1. Motivation

Electroencephalograms (EEGs) have gained its popularity in neurological research and clinical applications due to their mostly non-invasive recording procedures, simple and inexpensive operations and high time resolutions. Originally, the clinical application of EEG analysis is to distinguish epileptic seizures from normal brain activities such as syncope, sub-cortical movement disorders and migraine variants. Advanced signal processing techniques such as independent component analysis (ICA) [1,2] and event related spectral perturbation (ERSP) analysis [3,4] have also become increasingly popular in clinical brain research. In recent years, researchers began to view multichannel EEG recordings as correlated signals with quasi-sparse representations and analyze them with the powerful arsenal of digital signal processing techniques.

However, due to the fact that the recording of EEG for a meaningful and accurate analysis is often necessarily time consuming; the relatively high time resolution and the characteristic of concurrent multichannel recording will produce enormous amounts of data needed to be stored and transported over media. Therefore, interests in searching for sparse and compressible representations of EEG have rose considerably during recent decades.

1.2. Recent Difficulties and Solutions

Unfortunately, determining whether a signal can be optimally approximated and decomposed with a linear expansion over a redundant (over-complete) dictionary of atomic waveforms is an NP-complete problem, furthermore, finding such a solution is an NP-hard one [5]. However, an approximate method that has already been discovered long before is found to be capable enough to tackle this problem. It was the Matching Pursuit algorithm [6]

proposed by S. Mallat and Z. Zhang in 1993, which is a simple and economical way to produce a sub-optimal signal expansion by iteratively choosing the atom that best matches the signal structure and remove the atom ingredient from the signal until certain stopping condition is met. This algorithm has already been well studied and improved in the most recent decade, such as the randomized variant – the Stochastic Matching Pursuit (SMP) – pioneered by Piotr J. Durka in 2001.

Despite these efforts, sparse representation remains an impalpable objective. It's well-known that the sparse decomposition result produced by SMP is possible, however unlikely, to be inconsistent. The dominant atoms in the decomposed sequences for one signal can be significantly different between multiple instances of SMP operations.

1.3.   Objectives

The ultimate objective of this research is to identify the dictionary suitable for sparse representation of the EEG ERP signals and components, then design a Memetic Algorithm adopting the belief of Matching Pursuit, which is able to factorize a given EEG signal into a linear combination of only a small number of time-frequency atoms (Gabor functions, for instance) over a redundant dictionary, meanwhile preserving a large portion of signal energy.

1.4.   Research Approach

The sparsity of data assumed the ability of representing signals as linear combinations of only a small amount of atoms in a redundant, pre-specified dictionary. This requires the prior knowledge on the signal characteristic and a properly chosen dictionary.

In the aspect of prior knowledge to the signals, researchers performing brain activity analysis based on the equivalent electric dipole modeling [7,8] believed that the stimulus and response events may be well modeled as equivalent dipoles oscillating locally at certain

time period and frequency band. So far, on the aspect of proper selection of the dictionaries, Ron Rubinstein et al. indicated that there are two mainstream methods for searching for the proper redundant dictionary [9] currently.

1.4.1. Dictionary Chosen According to Mathematical Model of Signals

The build-up of dictionaries according to mathematical model is mainly functions with closed form mathematical expressions such as Gabor functions and wavelet atoms, which can be easily parameterized.

Further, about the searching method for the representation in the various types of standard dictionaries, while Durka's Stochastic MP algorithm focuses mainly on bias elimination and generates only sequences of atoms with integral multiplications of sampling steps over the time and frequency domain, a new method employing advanced optimization algorithms is proposed to improve the accuracy and efficiency of parameter searching in each MP iterative step by regarding the parameters as continuous variables and providing an even sparser parameterized representation.

1.4.2. Dictionary Evolved Based on Training Data

This type of dictionaries emerges, in Ron's words [9], from a given set of realizations of the data. This set serves as the examples of the signals to be modeled; with a given initial dictionary, the training algorithm will evolve the dictionary atoms iteratively to make the resulted dictionary performs better on the training set. In such case, an initialization method with minimal mutual coherence based on the Grassmannian frame for the dictionary is proposed to enhance the learning procedures.

1.5. Contributions

The major contribution for the standard dictionary approach is an efficient algorithm to

find a decomposition sequence in the scalable Gabor dictionary, which improves the Matching Pursuit by extending the search space into continuous real number domain and with one more degree of freedom with variable scaling factors. With the behaviors of the parameters of the Gabor standard dictionary are well understood, the signal can therefore be decomposed into merely a small number of parameter sets with minimal energy losses.

Another contribution is the successful application of the K-SVD on the EEG signal integrated with the concept of Grassmannian frames; the resulted dictionary contains several fingerprints of the event related potential of the EEG signal, which might be used further to identify different stimuli with the help of classification methods.

1.6.   Thesis Outline

Along the roadmap, in *Chapter 2*, the fundamental tools and prior knowledge will be introduced. Each succeeding chapters will focus on each milestone of major discoveries.

In *Chapter 3*, the identification of Gabor atoms as the most suitable type of waveform and an intuitive, however stable and consistent method based on Matching Pursuit to find the representation are described here, including some first-stage experiment results.

In *Chapter 4*, a better optimization method based on Natural Gradient and a sub-optimal uniform sampling method over Gabor parameter space using the Grassmannian concept will be introduced. The experiment results will be compared with their counterpart mentioned in *Chapter 3*. The *Chapter 3* and *Chapter 4* will cover the approach of realizing sparse representation via overcomplete standard waveform dictionaries.

Chapter 6 describes the improvements on the K-SVD algorithm via a new dictionary initialization method using the Grassmannian frame; the experiment result will be compared to the default initialization method using the normalized random Gaussian noise.

Chapter 7 includes the achievements and the future work of this research.

Chapter 2. Backgrounds

In recent years, a new signal processing paradigm called Compressive Sensing (CS) [10,11] promises to reduce the representation data size of a sparse signal up to two orders of magnitude while capturing their essential characteristics. The key to efficient compressive sensing and sparse signal analysis lies with successful identification of suitable signal dictionaries for the sparse representation of those signals.

A preliminary study by S. Aviyente in 2007 [12] suggested that the compressive sensing framework might be applicative for EEG signals. In 2009, Abdulghani et al [13] demonstrated the apparent sparsity of EEG signals with respect to several standard dictionaries including the Gabor atoms, Mexican Hat and Spline-based wavelets. However, the proper temporal-frequency and spatial-temporal dictionaries are yet to be discovered.

2.1.   Gabor Atoms and Dictionary

Real Gabor atoms, which are simply shifted and frequency modulated Gaussian functions, which can be expressed as follow:

$$g_\gamma = K(\gamma)e^{-\pi(t-t_0)^2/s^2}\cos(\omega_0(t-t_0)+\phi) \tag{1}$$

The set $\gamma = \{t_0, \omega_0, s, \phi\}$ in the parameter space Γ is the parameter set of a Gabor function, which contains the time center t_0, the frequency center ω_0, temporal scaling factor s and the sinusoidal phase ϕ of the atom. The normalization factor $K(\gamma)$ is the function ensuring the atom contains unit norm energy.

Figure 1 gives an example of Gabor atom (not energy normalized) waveforms in the time domain as shifted and modulated Gaussian functions.
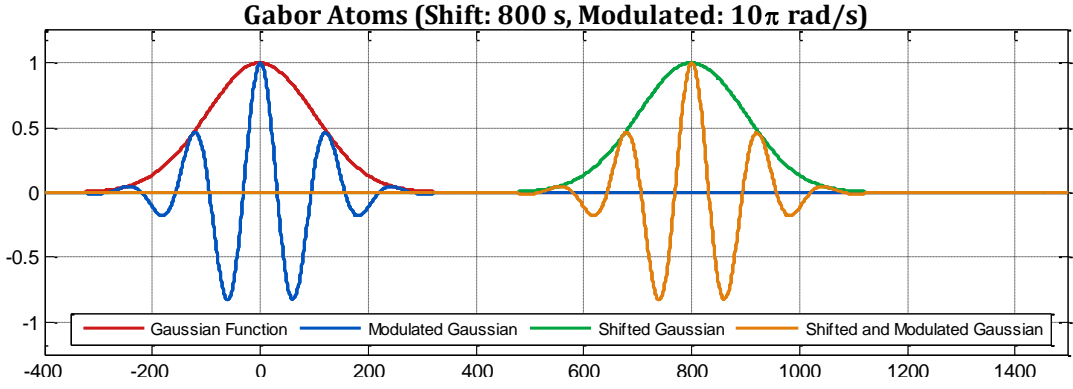
Figure 1 Gabor Atoms as Shifted and Modulated Gaussian Functions

The Gaussian envelope of a Gabor atoms leads to the optimal time-frequency localization for the Heisenberg's uncertainty principle in terms of energy concentration (the equality holds if g is a Gaussian function):

$$\sigma_t^2 \sigma_\omega^2 = \left(\int_{-\infty}^{\infty} t^2 |g(t)|^2 dt\right) \left(\int_{-\infty}^{\infty} \omega^2 |\hat{g}(\omega)|^2 d\omega\right) \geq \frac{\|g\|^4}{16\pi^2} \tag{2}$$

This property makes the Gabor atom suitable for performing local time-frequency brain activity observations and analyses; an explicit construction method for an overcomplete Grassmannian Gabor frame will be introduced later.

2.2. *Matching Pursuit*

2.2.1. *S. Mallat's Matching Pursuit (MP)*

Matching pursuit (MP) algorithm can provides a sub-optimal solution to the problem of acquiring an adaptive approximation of a signal in a redundant set of waveforms (dictionary).

The pursuit starts with a given dictionary with sufficiently abundant of unit energy atoms $\mathcal{D} \equiv \left\{g_{\gamma_n} : \|g_{\gamma_n}\| = 1\right\}_{n=1}^{N}$ of size N and a signal $f(t)$ to be analyzed; the main goal is to find a linear combination of atoms approximating the signal $f(t)$:

6

$$f(t) \approx \sum_n a_n g_{\gamma_n}(t) \tag{3}$$

In each iterative step, the algorithm chooses the atom having the largest inner product with the signal from $\mathcal{D}$. The signal is then subtracted by the atom waveform weighted by the inner product value. The procedure is repeated on the subsequent residual $Rf_n(t)$ after the subtractions until certain stopping condition is met, such as an acceptable energy extraction threshold or the number of iterations limited.

Algorithm   **MATCHING PURSUIT**

Input: Signal $f(t)$

Output: List of coefficients (a_n, g_{γ_n})

Initialization: $Rf_1 \leftarrow f(t)$

Repeat From $n = 1$:

$\qquad g_{\gamma_n} \leftarrow \underset{g_{\gamma_i} \in \mathcal{D}}{\mathrm{argmax}} \left| \langle Rf_n, g_{\gamma_i} \rangle \right|$

$\qquad a_n \leftarrow \langle Rf_n, g_{\gamma_n} \rangle$

$\qquad Rf_{n+1} \leftarrow Rf_n - a_n g_{\gamma_n}$

$\qquad n \leftarrow n + 1$

Until *Stopping Condition* met

Table 1   Mallat's Matching Pursuit Algorithm

The orthogonality of the residue Rf_{n+1} and the best matched atom g_{γ_n} selected in the n^{Th} iteration implies the signal energy conservation property:

$$\|f\|^2 = \sum_{n=1}^{N} \left| \langle Rf_n, g_{\gamma_n} \rangle \right|^2 + \|R_{N+1}f\|^2 = \sum_{n=1}^{N} |a_n|^2 + \|R_{N+1}f\|^2 \tag{4}$$

This implies that if the algorithm successfully extracts the atom with the maximum correlation with the signal in each iteration step, the residue will consequently be minimized in terms of signal energy.

2.2.2. Stochastic Matching Pursuit

In 2001, Piotr J. Durka has proposed the Stochastic Matching Pursuit algorithm (SMP) [14], which is a Monte Carlo approach aiming to find a statically unbiased linear expansion for quasi-sparse signals.

Algorithm **STOCHASTIC MATCHING PURSUIT**	

Input: Signal $f(t)$

Output: List of coefficients $\left(a_n, g_{\gamma_n}\right)$

$g_{\gamma_1} \leftarrow \underset{g_{\gamma_1} \in \mathcal{D}}{\operatorname{argmax}} \left|\langle f, g_{\gamma_1}\rangle\right|$

Take arbitrary percentages of atoms with large correlation with $f(t)$ to form $\mathcal{D}_\alpha$

$a_1 \leftarrow \langle f, g_{\gamma_1}\rangle$

$Rf_2 \leftarrow f - a_1 g_{\gamma_1}$

Repeat From $n = 2$:

    Generate $\mathcal{D}_e$ densely and exhaustively in the parameter center of atoms in $\mathcal{D}_\alpha$

    $g_{\gamma_n} \leftarrow \underset{g_{\gamma_i} \in \mathcal{D}_e}{\operatorname{argmax}} \left|\langle Rf_n, g_{\gamma_i}\rangle\right|$

    $a_n \leftarrow \langle Rf_n, g_{\gamma_n}\rangle$

    $Rf_{n+1} \leftarrow Rf_n - a_n g_{\gamma_n}$

    $n \leftarrow n + 1$

Until *Stopping Condition* met

Table 2   Durka's Stochastic Matching Pursuit Algorithm

First, a stochastic dictionary $\mathcal{D}$ is generated according to the required signal length N and chosen time-frequency resolutions in time, frequency and scale ($\Delta t, \Delta \omega$ and Δs). The parameter space is thereby divided into total $\pi N^2 / \Delta t \Delta \omega \Delta s$ cuboids of size $\Delta t \times \Delta \omega \times \Delta s$. In each cuboid, a Gabor atom is chosen by drawing its parameters from uniform distributions within the given ranges of continuous parameters.

In the first iteration, a percentage of prominent atoms with large correlations with the input signal are chosen to form an adaptive dictionary $\mathcal{D}_\alpha$. Afterwards in each matching

procedure, parameters of the atoms chosen from $\mathcal{D}_\alpha$ are further optimized by a dense and exhaustive search in the designated vicinity.

In [14], Durka stated that, instead of being driven by a quest to improve the speed or compression ratio, the goal of the stochastic method is attaining possible exact parameterization of certain signal structures freeing from bias and with a constant time-frequency resolution.

2.3.　Localized Optimization Methods

The localized optimization methods are employed on the parameter sets of the time-frequency atoms to maximize the correlation between the atom waveform and the input signal or residuals. Two of which that have been employed during the research period are listed below.

2.3.1.　Covariance Matrix Adaptation Evolution Strategy

The Covariance Matrix Adaptation Evolution Strategy (CMA-ES) is a stochastic numerical optimization method. As the tradition of the evolution strategies, each new generation of candidates are sampled according to a multivariate Gaussian distribution based on the pairwise dependencies between the variables in the parameter sets – the dependencies in the distribution are described by a covariance matrix. The covariance matrix adaptation (CMA) is a method to update the covariance matrix of this distribution.

2.3.2.　Natural Gradient

Generic optimization algorithm (black-box optimization) are designed for problem with unknown and sophisticate objective functions; one of the simplest and the most celebrious one is gradient based optimization, based on the fact that a minimum or a maximum of a function takes place at the point having zero partial derivatives in all dimensions, forming the

core concept of gradient descent/ascent method [15,16]. One of the most significant advantages of gradient descent over the other alternatives is the simplicity and efficiency; however, the most notable drawback is the vulnerability towards local minimum, especially when the objective function is mountainous, asperous or spiky.

In 1998, S. Amari et al introduced the natural Riemannian gradient or simply, natural gradient [17,18] in the field of optimization. Taking the curved space into consideration and changes the concept of distance in ordinary Euclidean space.

In 2008, Daan Wierstra and Tom Shaul presented the Natural Evolution Strategy (NES) [19,20], utilizing the canonical process of the gradient descent but replacing the standard gradient by empirical Natural Gradient derived using population based method to estimate Fisher information matrix and gradient of the objective function.

The next year in 2009, with Yi Sun's help in advance, the Natural Evolution Strategy benefited from a new computational method to attain actual Fisher information matrix instead of performing population based estimation on the objective function in the parameterized searching space. This improved version of the algorithm is called Efficient Natural Gradient Strategies [21].

In this research, the natural gradient based algorithms are employed as local optimization method to seek for the best atom parameter to decompose the signal.

2.4. Grassmannian Frame

This section includes the mathematical property of a frame, since frames with low mutual coherence are favored in the field of Compressive Sensing; the Grassmannian concept is employed to suppress the increment of mutual coherence when raising the redundancy of a dictionary.

2.4.1. Frame and Mutual Coherence

An N- element finite frame $X_d^N = \{x_1, x_2, \ldots, x_N\} \subseteq \mathbb{R}^d$ is a family of vectors that characterizes any signal from its inner product $\{\langle f, x_k \rangle\}$ and completely spans the d-dimensional Euclidean space, which necessitates $N \geq d$ and should hold the frame condition: There exists two constants A and B, $A \leq B$ for any function $f \in \mathbb{R}^d$ such that $A\|f\| \leq \sum_k |\langle f, x_k \rangle|^2 \leq B\|f\|$.

Of particular concern is the correlation value between the underlying atoms within frame, which is known as the mutual coherence of a frame. The definition is:

$$\mu(X_d^N) = \max_{k,l \in \Gamma, k \neq l} \frac{|x_k^T x_l|}{x_k^T x_k x_l^T x_l} \geq \sqrt{\frac{N-d}{d(N-1)}} \qquad (5$$

The maximum correlations are ideally zero for all orthonormal frames; for overcomplete dictionaries, those with small mutual coherence value are preferred in the case of sparse representation. The effect will be discussed later in 3.6.

2.4.2. Definition of Grassmannian Frames

A unit-energy (unit $\mathbf{L}_2$ norm) finite frame $U_d^N = \{u_1, u_2, \ldots, u_N\} \subseteq \mathbb{R}^d$, where $|u_n| = 1$, $\forall n = 1$ to N, is an (N, d)-Grassmannian frame [22] if satisfying the minimal mutual coherence criterion:

$$\mu(U_d^N) = \inf\{X_d^N\} \qquad (6$$

That is, the $N-$element Grassmannian frames are frames minimizing the mutual coherence in $\mathbb{R}^d$. In [22], T. Strohmer et al provided a method of constructing optimal Grassmannian frame with Gabor atoms by positioning the time and frequency center of the atoms on a triangular lattice over continuous time-frequency plane.

2.4.3. *Grassmannian Gabor Frame*

Recent attempts to represent a vector/function/operator in terms of a redundant spanning set of normalized elements gives rise to the concept of frames, which can be regarded as the generalization of orthonormal bases. The level of its over-completeness is measured by its redundancy ρ. Recently, about the acquisition of Grassmannian Gabor frame, T. Strohmer introduced two unitary operators of translation T_x and modulation M_ω in [22]:

$$\begin{cases} T_{t_0} f(t) = f(t - t_0) \\ M_{\omega_0}(t) = f(t) e^{j\omega_0 t} \end{cases} \tag{7}$$

A two dimensional lattice Λ is a discrete subgroup of time-frequency parameter space with compact quotient determined by a generator matrix L via $\Lambda = L\mathbb{Z}^2$. The volume of the lattice Λ is $\text{vol}(\Lambda) = \det(\Lambda)$. For the Gaussian window function $g \in \mathbf{L}^2(\mathbb{R})$ and a lattice Λ in the time frequency plane $\mathbb{R}^2$, the corresponding Gabor system can be defined as:

$$\mathcal{G}(g, \Lambda) = \{M_{\omega_0} T_{t_0} g, \text{ where } (t_0, \omega_0) \in \Lambda\} \tag{8}$$

The completeness of the system is therefore determined by the redundancy $\rho = 1/\text{vol}(\Lambda)$. A necessary condition for a Gabor system $\mathcal{G}(g, \Lambda)$ to be a frame is $\rho \geq 1$.

Since a Gaussian function in time domain remains Gaussian in frequency, also, both shift and rotational invariant with respect to every time frequency center (t_0, ω_0). The time frequency energy spread is an ellipse in the time frequency plane, the redundancy of the Gabor system can be viewed as the overlapping in time-frequency energy distribution.

Applying the classical sphere packing theory [23], the optimal lattice for a two dimensional space is the solution for $\text{argmax}_\Lambda\{(\text{volume of a sphere}) \times \rho\}$, and the generator matrix L of the solution lattice Λ is found in [22] to be a hexagonal lattice:

$$L = \begin{bmatrix} \dfrac{\sqrt{2}}{\sqrt[4]{3}\sqrt{\rho}} & \dfrac{1}{\sqrt[4]{3}\sqrt{2\rho}} \\ 0 & \dfrac{\sqrt[4]{3}}{\sqrt{2\rho}} \end{bmatrix} = \dfrac{\sqrt{2}}{\sqrt[4]{3}\sqrt{\rho}} \begin{bmatrix} 1 & \dfrac{1}{2} \\ 0 & \dfrac{\sqrt{3}}{2} \end{bmatrix} \tag{9}$$

Such result is also applied on OFDM design for time-frequency dispersive channel, the Fig. 4 (b) in [24] illustrate the resulted Gabor system energy concentration of the packing:
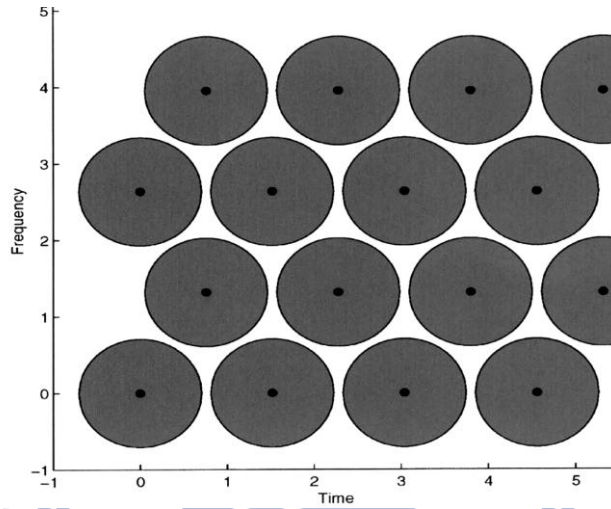
Figure 2 Supports of Energy Spreading of a Constant-Scale Gabor System Generated with a Hexagonal Lattice, Fig. 4 (b) of [24]

And the dilation effect of the scaling factor would make the form become:

$$L_s = \sqrt{\dfrac{2}{\sqrt{3}\rho}} \begin{bmatrix} s & \dfrac{s}{2} \\ 0 & \dfrac{\sqrt{3}}{2s} \end{bmatrix} \tag{10}$$

The overcomplete Grassmannian Gabor frame will be applied on both the advanced search of the optimal standard waveform representation and the dictionary training approach as well; from which the Matching Pursuit based algorithm selects the most prominent atom then performs further optimization in the first approach, and serves as the initial dictionary with configurable redundancy in the second.

2.5. Memetic Algorithms

The theory of "Universal Darwinism" was coined by Richard Dawkins in 1983 [25] to provide a unifying framework governing the evolution of any complex system. In particular, "Universal Darwinism" suggests that evolution is not exclusive to biological systems. The term "meme" was also introduced and defined by Dawkins in 1976 as "the basic unit of cultural transmission, or imitation", and in the English Oxford Dictionary as "an element of culture that may be considered to be passed on by non-genetic means".

Now, Memetic Algorithm (MA) [26] is one of the recent blooming areas of research in evolutionary computation, which is now widely used as a synergy of population-based approach with separate individual local improvement procedures. And finally, use evolutionary strategy such as Darwinian or Lamarckian filtering to select the proper chromosomes to reproduce the next generation.

The following code section shows a typical process of a canonical Memetic Algorithm:

Algorithm **CANONICAL MEMETIC ALGORITHM**

Initialize: Given starting point *p;*
 While *Stopping conditions* **Not** satisfied **Do**
     Evolve a new population Π using stochastic search operators.
     Evaluate all individuals in the population.
     Select the subset of individuals Ω_{il} that should take the learning procedure.
     For Each individual **In** Ω_{il} **Do**
         Perform learning with meme with probability of p_{il}
         Proceed with Lamarckian or Baldwinian learning.
     End For
 End While

Table 3   Canonical Memetic Algorithm

The concept of the Memetic Algorithm will be combined with the Grassmannian

dictionary and the natural gradient based search algorithm to assemble the ultimate

optimization method for the search of the Gabor parameter whose corresponding waveform

fit the given signal the most.

Chapter 3. CMA-ES Based Local Search

Based on the fundamental belief of Matching Pursuit, the goal of searching for the parameters of atom having the highest correlation value with the signal is quite obvious. In the first move, the quasi sparse characteristic of the EEG signals is justified by observing the distribution of Gabor atom centers from the result of time frequency analyses.

3.1.   Identifying the Quasi Sparse Property of EEG Signals,

Figure 3 is the normalized density map with top density re-scaled to 1.0 obtained by analyzing the time-frequency centers and scaling factors of the Gabor atoms after the decomposition of the 386-epoch, 145ms, 64-channel EEG signals recorded in the electro-magnetic isolation chamber of Brain Research Center, NCTU.
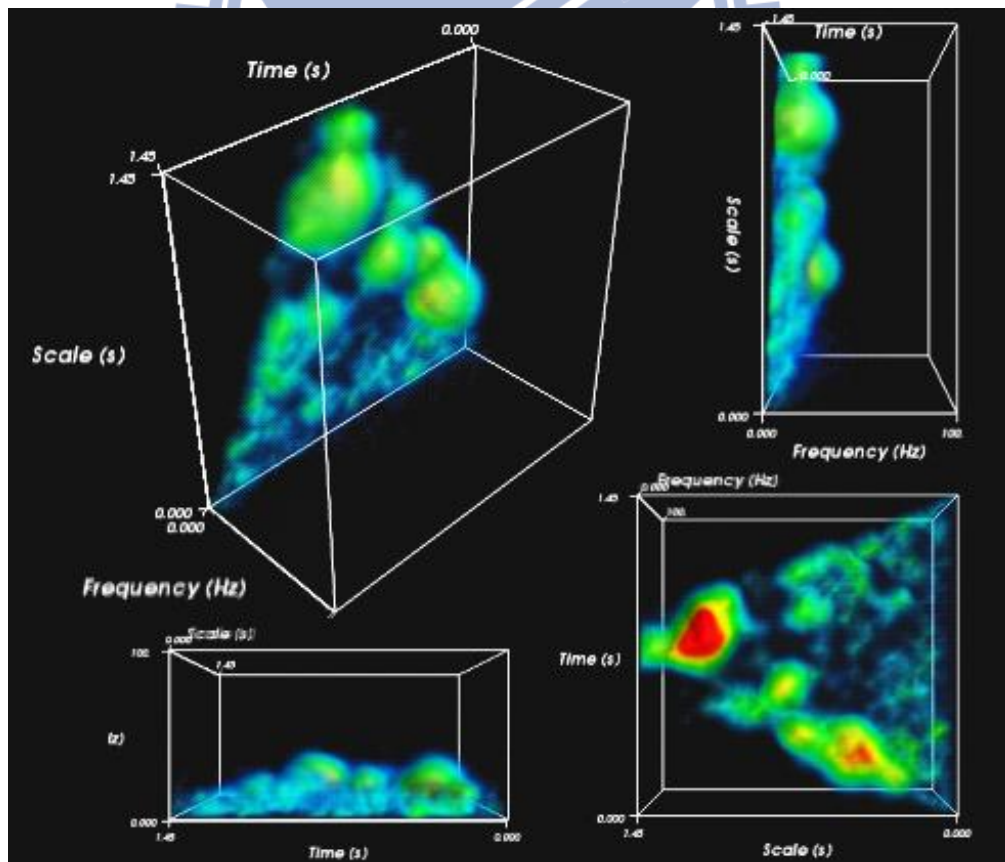
Figure 3   Relative Density Map of Time-Frequency Centre of Gabor Atoms

Each time-frequency-scale central point is weighted by the squared correlation value between the atom waveform and the signal according to the energy conservation property in Equation (4. To emphasis the concentration, a spherical kernel density function is used, the mathematical formula of the kernel is:

$$\begin{cases} k \cdot (1 - t^2)^2 & \text{if } t < 1 \\ 0 & \text{if } t \geq 1 \end{cases} \quad \text{where } t \text{ is the distance measure} \tag{11}$$

The concentration of time-frequency centers of Gabor dictionary atoms generated by decomposing the EEG signal with Durka's stochastic matching pursuit indicates the possibility of representing the signal with only few time-frequency localized atoms.

3.2. Sparse Representation over Standard Dictionaries

This section contains the description of four types of dictionaries – Gabor Chirp (Chirplet), Morlet wavelet, real Mexican Hat Wavelet and Cubic Cardinal B-Spline – used to obtain the sparse representation of the ERP and ICA signals.

3.2.1. Linear Gabor Chirp (Linear Chirplet) Dictionary

A linear Gabor chirp, or a chirplet by Steve Mann [27], is a piece of a chirp derived from windowing a chirp function where the window provides the time localization property similar to their Gabor siblings.

Most of the chirplet parameters are the same as those used in the Gabor atom, the additional parameter, the chirp rate, is introduced to represent the change of the oscillating frequency of the waveform over time.

Given the parameter set $\{t_0, \omega_0, s, c, \phi\}$ as time center, frequency center, scale, chirp rate and phase, the mathematical formula of a chirplet atom is:

$$g_{t_0, \omega_0, s, c, \phi}(t) = K \cdot \exp\left(-\pi\left(\frac{t - t_0}{s}\right)^2\right) \cdot \cos\left(\frac{c(t - t_0)^2}{2} + \omega_0(t - t_0) + \phi\right) \tag{12}$$

The normalizer K is a parameter set dependent function employed here and in the following sections to generate unit-norm atoms.

The following figure gives an example for a Gabor atom and a chirplet atom with the same time-frequency center, scale and phase.
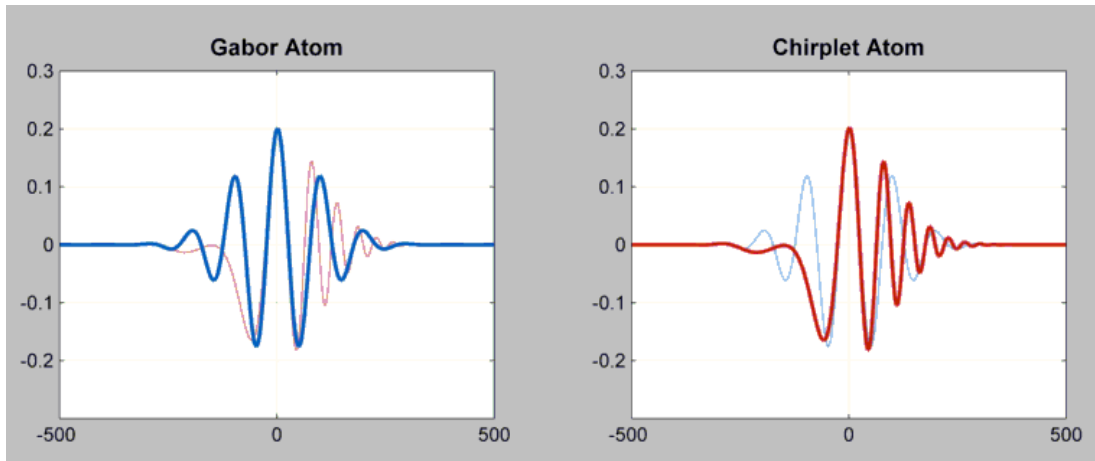
Figure 4 Gabor and Chirplet Atoms

Brown et al [28] have conducted some experiments using the chirplet atoms to decompose the ERP signals, the representation result modeled the signal only with a small number of chirplet logons and minimal energy lost. This indicates that the chirplet analysis might be suitable for perform sparse signal processing on the EEG signals.

3.2.2. Real Morlet and Mexican Hat Wavelet

Continuous wavelet analysis is another powerful instrument in the research field of joint time-frequency analysis [29,30]. As the ordinary constant-scale Gabor analysis (or short-time Fourier analysis) emphasized the localized time and frequency activities using time shifts and frequency modulations; the wavelet provides an alternative for frequency modulations by changing the scaling factors. According to the Heisenberg's uncertainty principle, a lower scaling factor will narrow the active period in time but expands the observable frequency band, and a larger one will do the opposite.

The Morlet wavelet family was introduced by Jean Morlet in 1984, which is a Gaussian

windowed sinusoid and actually embedded in the subspace of the Gabor atoms. Given the sampling rate f and the parameter set $\{t_0, s\}$ the parameterized Morlet atom can be represented as:

$$\psi_{t_0,s}(t) = K \cdot \exp\left(-\frac{1}{2}\left(\frac{t-t_0}{s}\right)^2\right) \cdot \cos\left(5 \cdot f \cdot \left(\frac{t-t_0}{s}\right)^2\right) \tag{13}$$

The Mexican Hat wavelet is proportional to the second order derivative of the Gaussian function. With the parameter set $\{t_0, s\}$, the shifted and scaled Mexican Hat wavelet atom can be formulated as:

$$\psi_{t_0,s}(t) = K \cdot \left(1 - \left(\frac{t-t_0}{s}\right)^2\right) \cdot \exp\left(-\frac{1}{2}\left(\frac{t-t_0}{s}\right)^2\right) \tag{14}$$

The following figure gives an example for the two types of wavelet atoms:
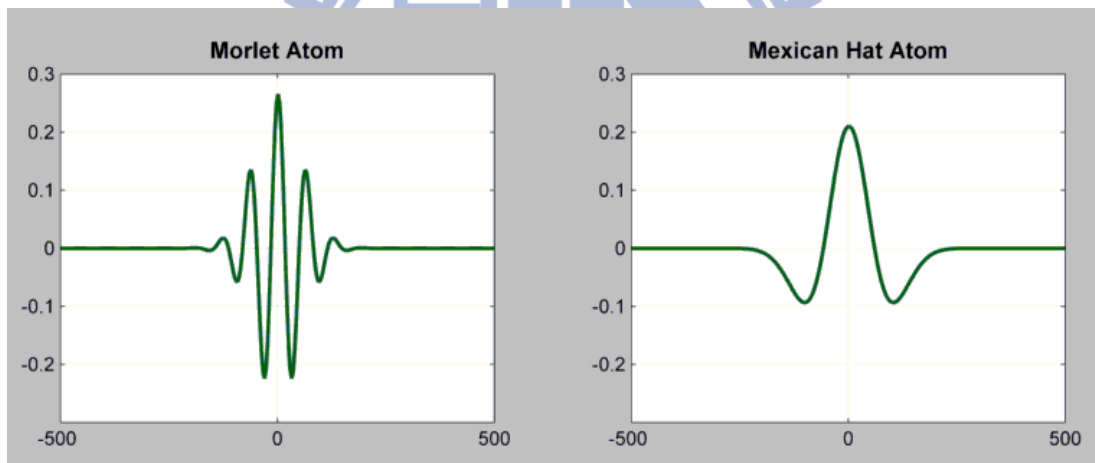
Figure 5 Morlet and Mexican Hat Atoms

Both the Morlet and Mexican Hat wavelet families are both widely used in biomedical signal analysis [31,32] and other areas like image compression.

3.2.3. Cubic Cardinal B-Spline

The cubic cardinal B-spline functions are continuous piecewise-polynomial compact waveforms introduced recently [33,34] for sparse representations and biomedical signal processing; these waveforms are frequently employed to absorb pulse-shaped features of

the signals. Given the parameter set of time shift and scaling factor $\{t_0, s\}$, the cubic cardinal

B-spline atom has the following mathematical form as in Equation (15 and shape in Figure 6:

$$B_{t_0,s}(t) = K \cdot \frac{1}{4!} \sum_{i=0}^{4} (-1)^i \binom{4}{i} \max\left[0, \left(\left(\frac{t-t_0}{4s}\right) - i\right)\right]^3 \tag{15}$$
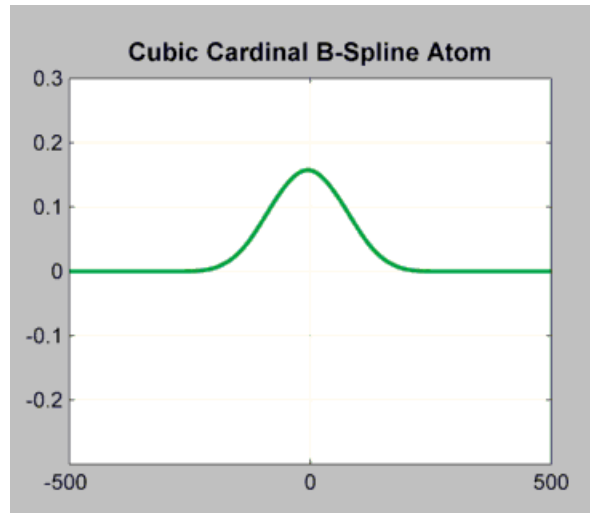
Figure 6 Cubic Cardinal B-Spline Atom

3.2.4. Mixed Dictionary

This design focuses on finding the sparse representations based on the combination of all the dictionaries mentioned above; the aim is to investigate whether more efficient representations may be produced under such configuration.

3.3. Signal Decomposition

3.3.1. Decomposition Problems

For the following sections, these variables are used for size/dimension measurements:

Variables	Meanings
N	Number of signal trials in the training template signal set
n	Number of samples in each signal trial
k	Number of atoms in the dictionary
L	Number of non-zeros in the decomposed coefficient matrix

Table 4　Size/Dimension Variables and Their Meanings

Their relation should be $L \ll n < k \ll N$.

The sparse signal decomposition (factorization) problem is formally stated here: Let $k > n, b \in \mathbb{R}, A \in \mathbb{R}^{n \times k}$ and $\text{Rank}(A) = n$, the noiseless single-trial signal decomposition problem P_0 is defined as:

$$P_0 \equiv \min_x \|x\|_0 \text{ subject to } Ax = b \tag{16}$$

The solution for the problem P_0 is definitely the sparse most representation. However, the case of the realistic world is not always in such an ideal situation; with the presence of noises, the bounded error version problem $P_{0,\epsilon}$ with the error measured in terms of the $\mathbf{L}_2$ norm is therefore defined for practice. The $P_{0,\epsilon}$ problem is suitable for modeling single-trial quasi-sparse standard waveform decomposition and recovery, which is presented in mathematical equation (17 and illustrated as Figure 7:

$$P_{0,\epsilon} \equiv \min_x \|x\|_0 \text{ subject to } \|Ax - b\|_2 < \epsilon \tag{17}$$
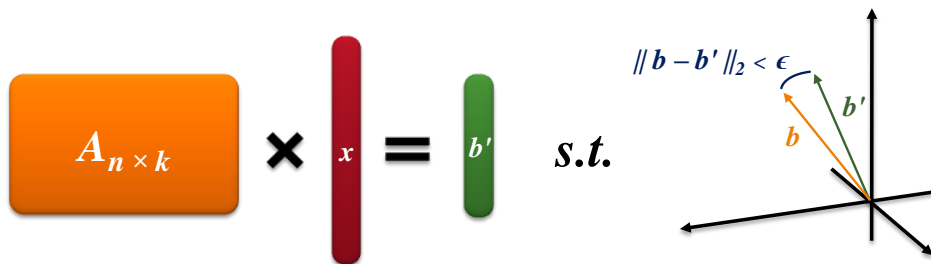
Figure 7　Illustration of Noisy Sparse Single-Trial Signal Decomposition Problem

3.3.2. Decomposition Algorithm

For the standard waveform dictionaries, the proposed method is basically based on the iterative procedures of Matching Pursuit.

An improved iteration procedure can be separated into 3 stages:

1. **Finding candidate atoms over loosely sampled parameter set**

   The candidates found in this stage are passed into the next stage as initial configuration of the optimization algorithm.

2. **Optimizing the candidates using CMA-ES**

   The CMA-ES algorithm will search the vicinity of the initial point acquired from the previous stage and iteratively converged to a certain parameter set. The most prominent set maximizing the absolute inner product value of the derived atom will be recognized as the best matched.

3. **Removing the best matched waveform from the signal space**

   This step is the same as the original Matching Pursuit.

To identify in which type of dictionaries can the signal be well and sparsely represented, the Matching Pursuit algorithm is applied on several types of standard dictionary including the generalized Gabor dictionary (consists of Fourier atoms, impulse functions and ordinary Gabor functions), Mexican Hat wavelet, Morlet wavelet and Cubic Cardinal B-Spline dictionaries for the comparison of the performance for each type of dictionary listed above.

3.4. Matching Pursuit Based on CMA-ES

Covariance matrix adaptation evolution strategy (CMA-ES) [35,36] is a stochastic numerical algorithm for solving non-linear non-convex optimization problems with search spaces of three to a hundred dimensions.

In essence, CMA-ES performs a principal components analysis in each mutation step so as to provide maximum likelihood estimation for subsequent parameter adaptation. Beside of computing the covariance matrix of its multivariate normal distribution of mutations, it also records the evolution paths of the mean of mutation distribution. The information is used to control the step size of mutations and determine the favorable directions for covariance adaptation. As a result, the algorithm is capable of handling problems with multimodal, discontinuous, even noisy fitness functions.

CMA-ES is a highly reliable and competitive algorithm for most local optimization and many global optimization problems. It is the de-factual benchmark for adaptive optimization.

3.5. Discovery of Inconsistency

Inconsistency described here means the decomposition sequences of the Gabor atom parameter sets may have significant differences in atom orders and values.

Given the same signal, the dominant atoms of one representation can be significantly different from those appear in another representation produced by a different run of the algorithm. Two causes of inconsistency have been identified:

1. **The presence of redundant atoms in a dictionary**

   For example, multiple atoms in a Gabor dictionary can produce similar or identical waveform; if the oscillating frequency of the sinusoid is too low, the behavior of the atom is just similar to a single envelope without oscillating. For

example, the two following unit-energy atoms A and B have almost identical waveform (both contain 290 sampling points at sampling rate of 200 Hz):

Atom	Length (s)	Normalizer	Shift (s)	Freq. (Hz)	Scale(s)	Phase (π)
A	1.45	36.486077	0.2491	0.0101	0.2533	0
B	1.45	0.538988	0.25	0.6897	0.26	0.5172

Table 5　Atom A and B with Significant Different Parameters

The mean square error is $\|A - B\|^2/290 = 2.7338 \times 10^{-7}$, and the following Figure 8 shows the waveform and frequency spectrum of the two atoms.

(a) Atom A　　　　　　　　(b) Atom B
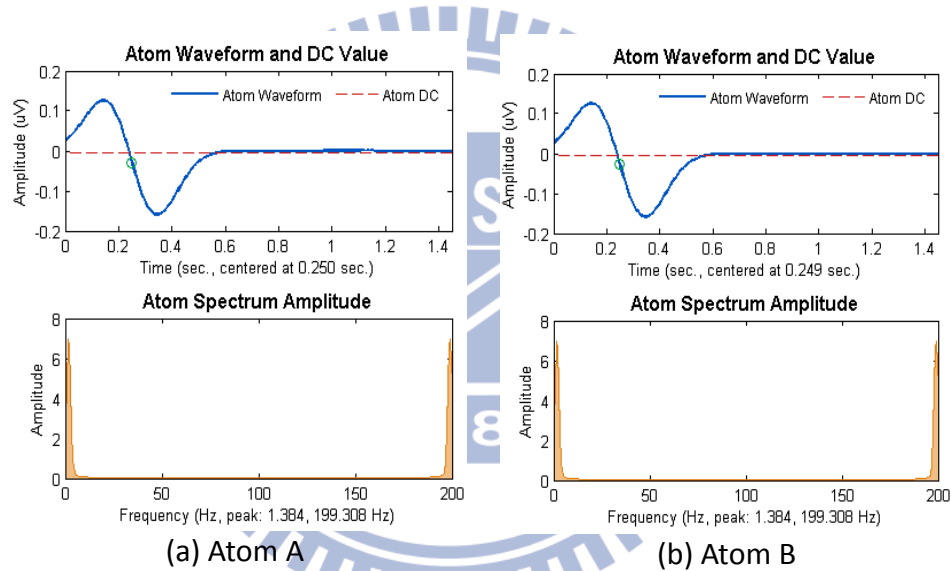
Figure 8　Example of Two Almost Identical Atoms with Different Parameters

2. The existence of degenerative cases in a sparse representation

For example, as a Gabor atom increases the standard deviation of its Gaussian envelope, it becomes more difficult to locate the time center of this atom.

In "Consistent Sparse Representations of EEG ERP and ICA Components Based on Chirplet and Wavelet Dictionaries" [37], several constraints on value ranges of the parameters are proposed to ensure the consistency during the decomposition.

3.6.   Consistency Conditions

For signal decomposition algorithms, there are several indexes for performance measurements – accuracy, efficiency and consistency [37]:

The accuracy of the decomposition is measured in the strength of the correlation between the found atom and the signal; the efficiency is ranked in terms of the number of atoms used to absorb a certain percentage of energy; the consistency is judged based on following two terms:

1. **The organization of the decomposed atoms**

    The ensemble of dominant atoms should form distinct clusters in their parameter spaces; these clusters of atoms accumulated through repeated random searches may be used to classify different ERP types.

2. **The reflections of ERP signal atoms in ICA components**

    The clusters of dominant atoms of ERP signals and their ICA components should manifest significant overlap with one another — these overlap atom clusters confirm the agreement between the sparse representations of the two EEG signal forms.

To ensure consistency, the decomposition should first have uniqueness (for exact reconstruction of noiseless signals), or stability (for signals with noise) through the sparsity of signals, which will be covered in 3.6.1; and the rule enforcing parameter consistency will be discussed in 3.6.2.

3.6.1.   Mathematical Uniqueness and Stability

In 2003, Donoho and Elad [38,39] identified the harsh conditions for the existence of a unique and stable representation in an over-complete dictionary. Elad et al. [40] then

proposed a method to find a unique dictionary that can produce stable representations for a given set of signals and devised the K-Singular Value Decomposition (K-SVD) algorithm [41] to design the dictionary consisting of non-parametric atoms with tailored waveforms only if the existence conditions of stable representations were satisfied at the first place; the part of the KSVD algorithm will be stated later in *Chapter 6*, here, the primary focus is the uniqueness and consistency condition on single-trial signal decomposition.

In [40] it describes that if the sought solution x for P_0 satisfies the sparsity condition,

$$\|x\|_0 \leq \frac{1}{2}\left(1 + \frac{1}{\mu(A)}\right)$$

(18

the pursuit algorithms will recover the original signal exactly. This suggest that for representation less than $\mathcal{O}(\sqrt{n})$ non-zero terms (see equation (5), Matching Pursuit algorithm can successfully reconstruct the signal.

3.6.2. Parameter Consistency

The Gabor parameters of different atoms span the search space of the CMA-ES. In order to exclude the redundant and degenerative cases in CMA-ES mutations, the following pruning rules of the search space are devised to prevent such undesired occasions:

Rule	Parameters	Pruning Rules (Tolerance: $\pm 5\%$)
1	**Time Shifts, t_0** (All Atoms)	Discard $t_0 \notin [T_{min},\ T_{max}]$
2	**Frequency Shifts, ω_0** (Gabor, Chirplets)	Discard $\omega_0 \notin [\omega_{min},\ \omega_{max}]$
3	**Scaling, s** (Gabor)	$\forall g_{t_0,\omega_0,s,\phi}(t)$ with $s > 100$, $g_{t_0,\omega_0,s,\phi}(t) \rightarrow f_{\omega_0}(t) = \cos(\omega_0 t)$
4	**Scaling & Chirp, (c, s)** (Gabor Chirp)	Discard $g_{K,t_0,\omega_0,s,c,\phi}(t)$ if $c \times s > \omega_{max} - \omega_{min}$
5	**Scaling, s** (All wavelets)	Discard $\psi_{t_0,s}(t)$ if $\omega_\psi/s \notin [\omega_{min}, \omega_{max}]$, where $\omega_\psi = $ central frequency of the wavelet mother function.

Table 6   Atom A and B with Significant Different Parameters

In rule 1 and 2, pruning of (t_0, ω_0) is meant to ensure that the chosen atoms always fall within the time and frequency spans of the target signal. Applying these rules will eliminate most of the redundant atoms.

In rule 3, replacing the Gabor atoms with $s > 100$ by Fourier atoms of the same frequency is meant to remove degenerated Gabor waveforms. The replacement will only be carried out if the inner product $\langle r_m(t), g_{\gamma_m}(t) \rangle$ of the Fourier atoms is large than those of the Gabor atoms.

In rule 4 and 5, discarding chirplet and wavelet atoms with large scale is meant to eliminate degenerated waveforms, whose spectra spread across the entire frequency band.

3.7.   Experiments

Here, the type of dictionary that is the best choice for decomposing the EEG signal will be identified, and a simple method will be introduced to achieve consistent results.

3.7.1.   EEG Signal Recording and Pre-processing

A 64-channel EEG signal recording session was conducted within an EM-insulated chamber in NCTU Brain Research Centre (BRC) in July 2009. During the recording session, the test subject was asked to listen to a randomized sequence of rising or falling tones and was instructed to press a button immediately after hearing the rising tone but do nothing otherwise. The entire session lasted 40 minutes and produced total 1600 auditory-motor ERP epochs.

The signals from all 62 scalp channels were sampled into single-precision (16 bit) data at 1 KHz sampling rate and passed through ideal 1 Hz – 75 Hz FFT filters. They were then down sampled to 200 samples per second and fed through the artefact removal procedures. Only 266 epochs of the no bottom pressing event and 108 epochs of the bottom pressing ones remained after the procedure. Each epoch contain 290 samples that cover a period from

40ms before to 1405ms after each Event On-Set (EOS) moment. Independent component analysis was then applied to all the epochs to produce their corresponding ICA components. Average ERP signals of the two events were generated by removing the DC components of individual epochs and then averaging their signal samples. Similar process was used to produce the average ICA components.

3.7.2. Experiment Configurations

The experiment goal in this early phase of the research period is to identify correctness and guarantee stability of the decomposition result, and select the most promising dictionary that can appropriately decompose the EEG signal with the smallest amount of atoms required for later use.

Let the signal elapses for $\mathcal{T}$ and has the bandwidth of $\mathcal{B}$, the resolution in time, angular frequency and scale is respectively $\Delta t, \Delta \omega$ and Δs (scaling factors s and Δs have the same unit as time shifts t_0 and Δt, both are time units). The following algorithm is implemented to decompose the signal in a trivial way (the restriction of scaling factor $\mathcal{S}_{\min}$ and $\mathcal{S}_{\max}$ is usually problem dependent, it's common for EEG signal having $\mathcal{S}_{\min} = 0.1s$ and $\mathcal{S}_{\max} = 1s$):

Algorithm **EXHAUSTIVE MATCHING PURSUIT (EMP)** (USING GABOR AS EXAMPLE)

Input: Signal $f(t)$

Output: A list of coefficients and waveform parameters (a_n, γ_n)

Initialization: $Rf_1 \leftarrow f(t)$;

While $n = 1$ to *Iteration Limit* and *Stopping Condition* not met

    $a_{\max} \leftarrow -\infty$; $\gamma_{\max} \leftarrow Null$

    For $t_0 = 0$ to $\mathcal{T}$ step Δt

        For $\omega_0 = 0$ to $\mathcal{B}$ step $\Delta \omega$

            For $s = \mathcal{S}_{\min}$ to $\mathcal{S}_{\max}$ step Δs

                $\gamma \leftarrow [t_0, \omega_0, s_0]; a \leftarrow \langle Rf_n, g_\gamma \rangle$

                If $|a| > a_{\max}$ **then**

                    $(a_{\max}, \gamma_{\max}) \leftarrow (a, \gamma)$ # the most correlated atom

                End If

            End For

        End For

    End For

    $\gamma'_{max} \leftarrow \textbf{\textit{CMAES}}(\gamma_{\max})$ # local optimization procedure with pruning rules

    $a'_{max} \leftarrow \langle Rf_n, g_{\gamma'_{max}} \rangle$ # use the updated correlation value

    $(a_n, \gamma_n) \leftarrow (a'_{\max}, \gamma'_{\max})$

    $Rf_{n+1} \leftarrow Rf_n - a_n g_{\gamma_n}$

End While

Table 7   Exhaustive Matching Pursuit Algorithm (Gabor example)

The objective function is simple, given the parameter set γ; the object function returns the value $\left| \langle Rf_n, g_\gamma \rangle \right|$. For the wavelets, simply remove the frequency scanning loop of ω_0.

3.7.3. Experiment Results

This experiment was suggested to use the event related potential of ICA component 20 and channel 17 as our input. (Signal elapse $\mathcal{T} = 1.45$, Bandwidth $\mathcal{B} = 75$Hz).
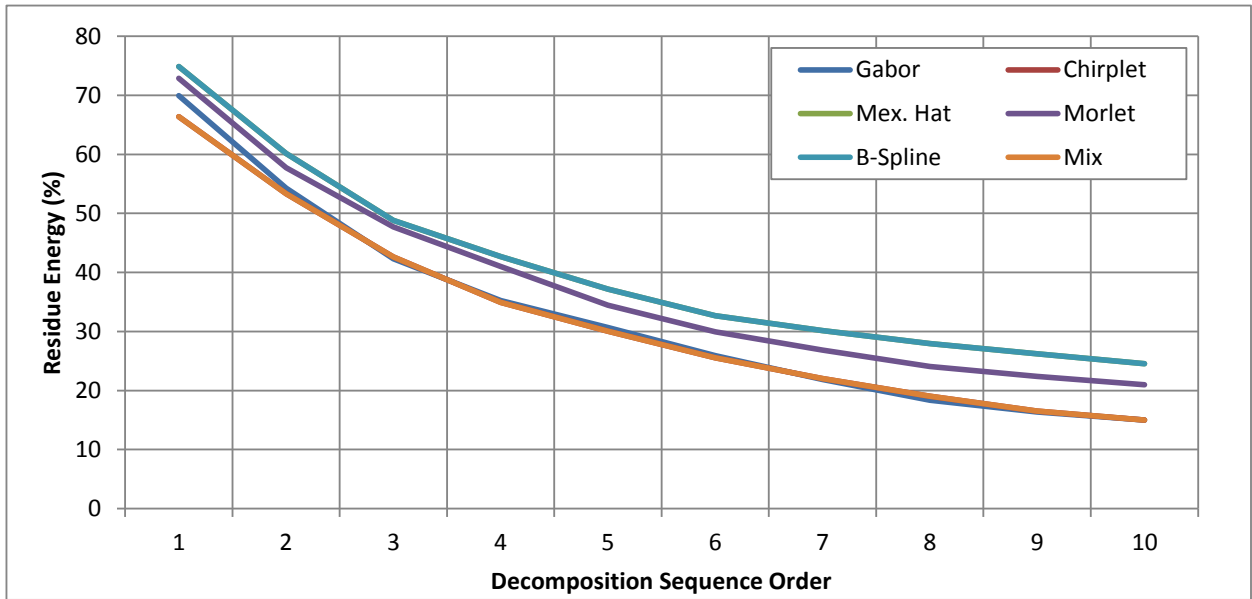
Figure 9 Comparison of Convergence Rate of Different Dictionaries over ICA Component #20
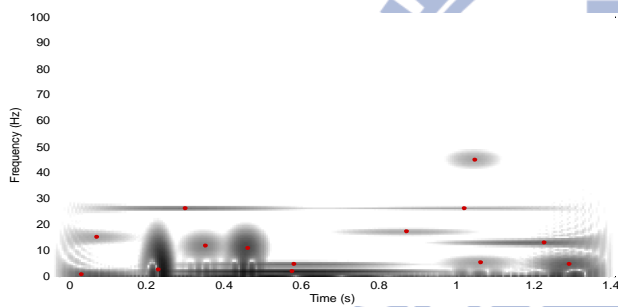
Figure 10   Gabor Energy Map (ICA #20)
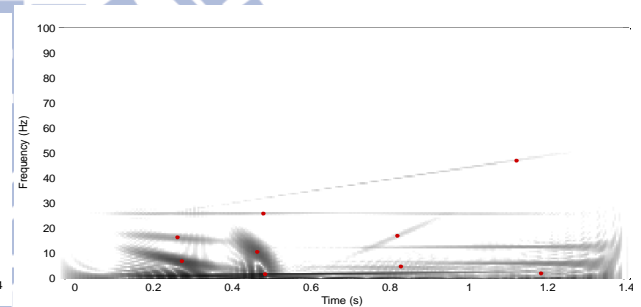
Figure 11   Chirplet Energy Map (ICA #20)

#	Time Shift	Freq. Shift	Scale	Phase (π)	Wgh. (%)
1	0.52	1.70	0.71	1.37	30.04
2	0.32	6.24	0.15	1.35	15.70
3	0.50	10.69	0.08	1.09	11.95
4	1.45	6.21	0.60	0.73	7.07
5	0.32	14.63	0.15	0.06	4.56
6	1.32	12.74	0.71	1.71	4.79
7	0.16	25.89	1.51	0.07	4.02
8	0.83	4.00	0.32	0.71	3.51
9	1.25	2.04	0.32	0.89	1.99
10	0.10	15.17	0.19	0.54	1.35

Table 8   Gabor Atom List (ICA #20)

#	Time Shift	Freq. Shift	Scale	Chirp Rate	Phase (π)	Wgh. (%)
1	0.51	3.25	0.59	-17.91	1.31	33.63
2	0.25	0.01	0.25	0.00	0.50	13.07
3	0.48	12.03	0.15	-19.02	0.60	10.69
4	1.98	5.11	1.31	-1.36	0.53	7.70
5	1.32	12.95	0.71	1.11	1.74	4.87
6	0.28	12.36	0.10	383.21	0.73	4.51
7	0.75	3.96	0.32	-8.19	0.07	3.48
8	0.15	17.37	0.14	183.46	1.90	3.03
9	0.98	25.86	1.94	0.03	1.00	2.50
10	1.31	0.00	0.12	-0.05	1.50	1.50

Table 9   Chirplet Atom List (ICA #20)
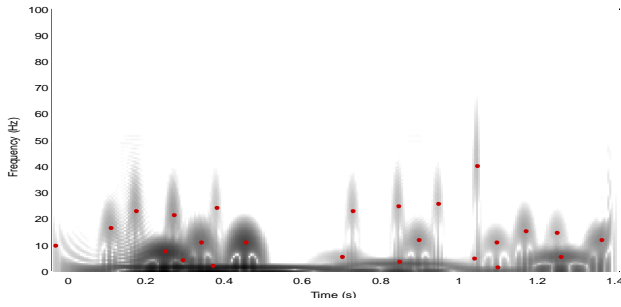
Figure 12 Mex. Hat Energy Map (ICA #20)

Figure 13 Morlet Energy Map (ICA #20)

#	Time Shift	Scale	Weight (%)
1	0.41	0.16	25.16
2	0.29	0.03	14.71
3	0.49	0.02	11.31
4	0.89	0.09	6.16
5	1.3	0.04	5.5
6	0.38	0.02	4.47
7	1.4	0.02	2.54
8	1.14	0.24	2.2
9	0.94	0.02	1.73
10	0.33	0.07	1.69

Table 10   Mex. Hat Atom List (ICA #20)

#	Time Shift	Scale	Weight (%)
1	0.41	0.46	27.12
2	0.38	0.15	15.18
3	0.46	0.06	10.01
4	1.39	0.14	6.68
5	0.29	0.07	6.54
6	0.88	0.27	4.51
7	1.33	0.06	3.12
8	1.34	0.83	2.75
9	0.19	0.03	1.68
10	0.31	0.04	1.42

Table 11   Morlet Atom List (ICA #20)
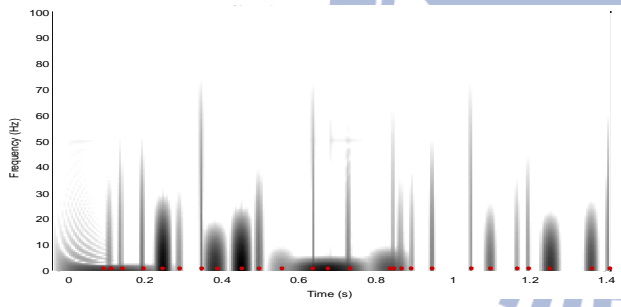
Figure 14   CB-Spline Energy Map (ICA #20)

Figure 15   Mixing Energy Map (ICA #20)

#	Time Shift	Scale	Weight (%)
1	0.41	0.16	25.16
2	0.29	0.03	14.71
3	0.49	0.02	11.31
4	0.89	0.09	6.16
5	1.30	0.04	5.50
6	0.38	0.02	4.47
7	1.40	0.02	2.54
8	1.14	0.24	2.20
9	0.94	0.02	1.73
10	0.33	0.07	1.69

Table 12   CB-Spline Atom List (ICA #20)

#	Type	Time Shift	Freq. Shift	Scale	Chirp Rate	Wgh. (%)
1	Chirplet	0.51	3.25	0.59	-17.91	33.63
2	Chirplet	0.34	0.50	0.30	-102.49	13.07
3	Gabor	1.90	31.04	1.19	45.60	10.69
4	Chirplet	0.09	0.69	0.71	42.30	7.70
5	Chirplet	1.11	12.48	1.79	-0.69	4.87
6	Chirplet	0.22	17.89	0.18	-97.71	4.51
7	Chirplet	0.70	4.42	0.50	-17.91	3.48
8	Chirplet	0.36	26.05	0.29	6.50	3.03
9	Chirplet	1.17	7.29	0.96	3.55	2.50
10	Chirplet	1.14	25.95	1.05	-0.64	1.50

Table 13   Mixing Atom List (ICA #20)
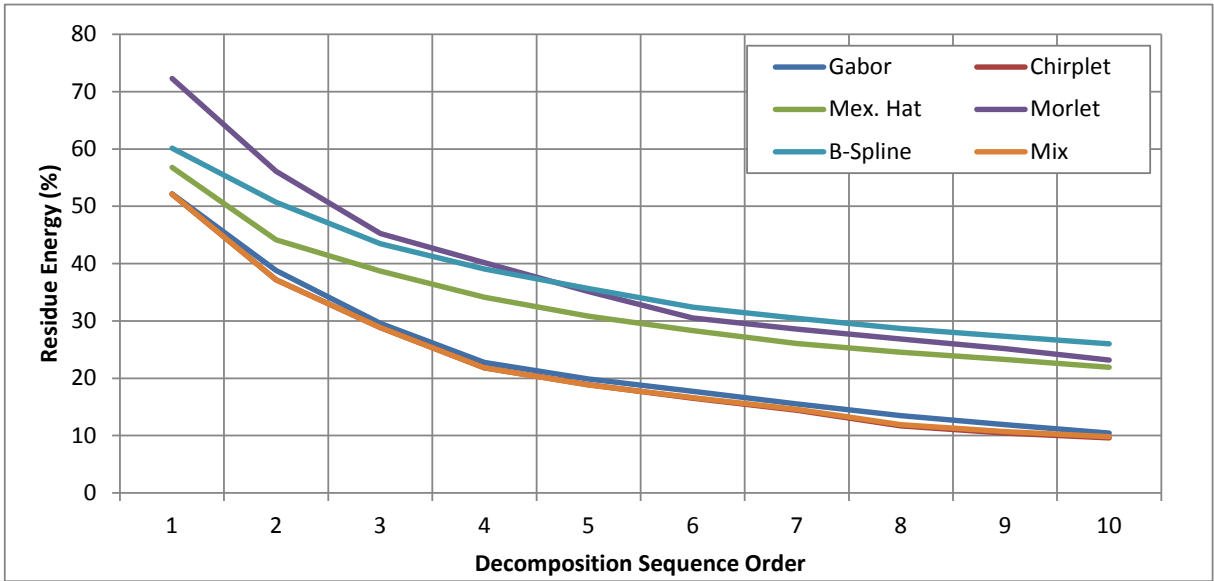
Figure 16 Comparison of Convergence Rate of Different Dictionaries over Event Related Potential of Channel #17
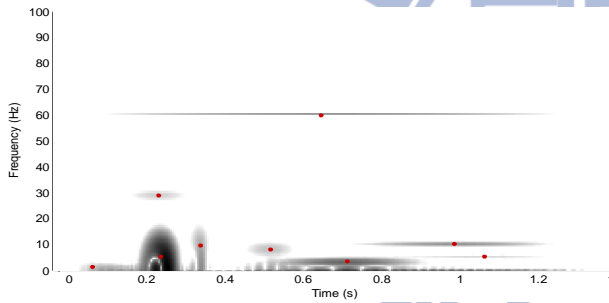
Figure 17 Gabor Energy Map (Ch. #17)
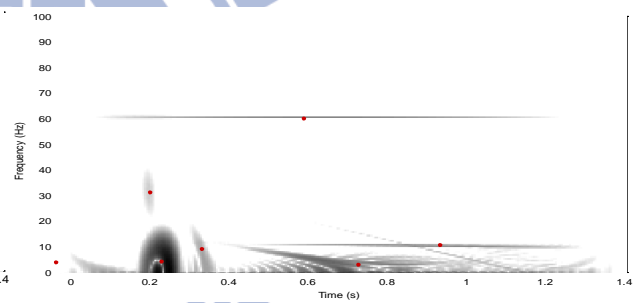
Figure 18 Chirplet Energy Map (Ch. #17)

#	Time Shift	Freq. Shift	Scale	Phase (π)	Wgh. (%)
1	0.27	0.02	0.06	0.50	47.85
2	0.75	3.43	0.30	1.98	13.36
3	0.45	59.97	20.55	0.24	9.20
4	0.88	10.76	0.99	1.06	6.83
5	0.38	7.95	0.05	1.78	2.86
6	0.06	2.14	0.27	1.60	2.19
7	1.02	6.48	0.14	0.07	2.19
8	0.27	28.94	0.19	0.76	2.05
9	1.36	1.68	0.16	1.45	1.56
10	0.60	7.31	0.28	0.65	1.49

Table 14 Gabor Atom List (Ch. #17)

#	Time Shift	Freq. Shift	Scale	Chirp Rate	Phase (π)	Wgh. (%)
1	0.27	4.16	0.06	31.25	1.31	47.87
2	0.77	2.83	0.44	-24.45	0.50	14.97
3	0.63	60.00	1.45	0.00	0.60	8.32
4	0.97	10.57	0.88	-1.66	0.53	7.00
5	0.37	9.00	0.07	-276.65	1.74	2.98
6	0.03	5.34	0.58	-46.8	0.73	2.36
7	0.24	30.6	0.04	-138.59	0.07	2.10
8	1.47	9.44	1.57	35.64	1.90	2.68
9	0.83	4.34	0.98	-14.36	1.00	1.26
10	0.11	54.17	0.12	-241.44	1.50	0.87

Table 15 Chirplet Atom List (Ch. #17)

32
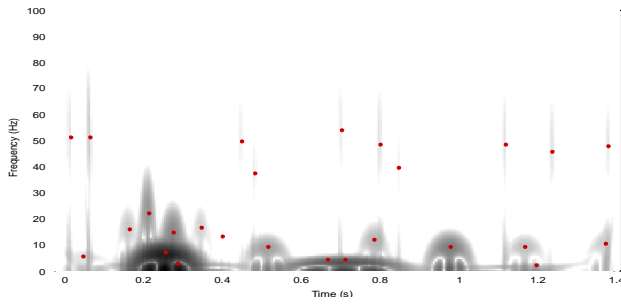
Figure 19   **Mex. Hat Energy Map (Ch. #17)**
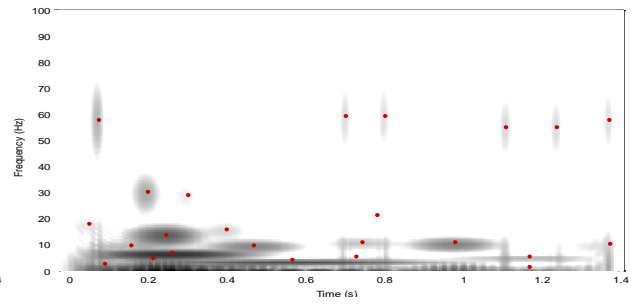
Figure 20   **Morlet Energy Map (Ch. #17)**

#	Time Shift	Scale	Weight (%)
1	0.30	0.04	43.20
2	0.75	0.07	12.67
3	1.02	0.03	5.46
4	0.32	0.02	4.53
5	0.25	0.01	3.33
6	0.56	0.03	2.50
7	0.33	0.10	2.23
8	0.83	0.02	1.53
9	1.21	0.03	1.25
10	1.24	0.15	1.36

Table 16 Gabor Atom List (Ch. #17)

#	Time Shift	Scale	Weight (%)
1	0.30	0.13	27.73
2	0.60	0.24	16.18
3	0.29	0.06	10.85
4	0.13	0.40	5.08
5	1.02	0.08	5.05
6	0.51	0.09	4.60
7	1.21	0.74	1.94
8	0.25	0.18	1.71
9	1.21	0.16	1.69
10	0.24	0.03	2.01

Table 17 Chirplet Atom List (Ch. #17)
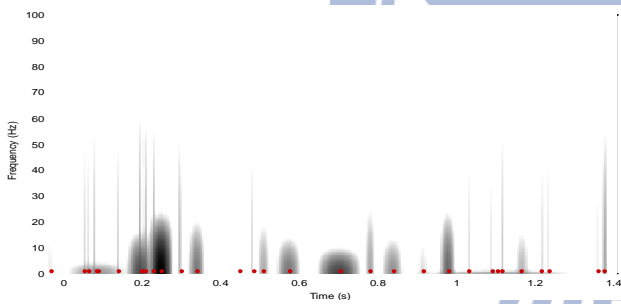
Figure 21   **CB-Spline Energy Map (ICA #20)**
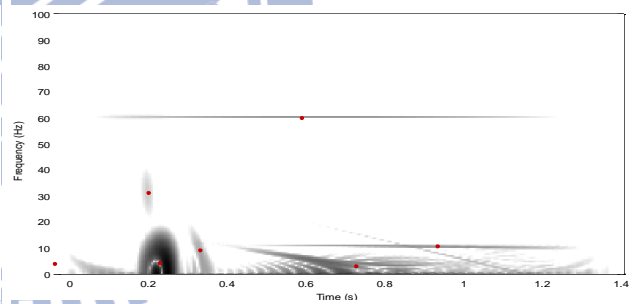
Figure 22   **Mixing Energy Map (ICA #20)**

#	Time Shift	Scale	Weight (%)
1	0.24	0.10	39.88
2	0.64	0.20	9.43
3	0.18	0.12	7.24
4	0.98	0.08	4.38
5	0.55	0.12	3.4
6	0.34	0.09	3.28
7	0.82	0.12	1.94
8	-0.04	0.34	1.78
9	0.68	0.93	1.36
10	1.40	0.03	1.32

Table 18 Gabor Atom List (Ch. #17)

#	Type	Time Shift	Freq. Shift	Scale	Chirp Rate	Wgh. (%)
1	Chirplet	0.27	4.16	0.06	31.25	47.87
2	Chirplet	0.77	2.83	0.44	-24.45	14.97
3	Gabor	0.63	60.00	1.45	N/A	8.32
4	Chirplet	0.97	10.57	0.88	-1.66	7.00
5	Chirplet	0.37	9.14	0.07	-280.45	2.99
6	Chirplet	0.00	3.93	0.48	-13.17	2.23
7	Chirplet	0.24	31.23	0.04	-129.88	2.06
8	Chirplet	1.49	10.11	1.57	35.60	2.63
9	Chirplet	0.77	4.92	0.57	-22.11	1.22
10	Chirplet	0.11	53.83	0.12	-242.38	0.88

Table 19 Chirplet Atom List (Ch. #17)

3.7.4. Result Analyses

From Figure 9 and Figure 16, the result is quite apparent that the Gabor atoms out-perform the other dictionaries during the decomposition process. In the following stages, Gabor atoms will be used as the primary decomposition dictionary in the research.

First, the decomposition sequence should be examined. The following Figure 23 shows the weights of the decomposed atoms and the decays of residue of ICA component #20 during the pursuit processes, generally, the amount of energy extracted by the exhaustive MP is higher than Durka's Stochastic MP. This phenomenon is expected since the CMA-ES optimization procedure breaks the parameter restrictions of integral resolution units, where atoms can be found to be generated by parameters in the field of real numbers $\mathbb{R}$.
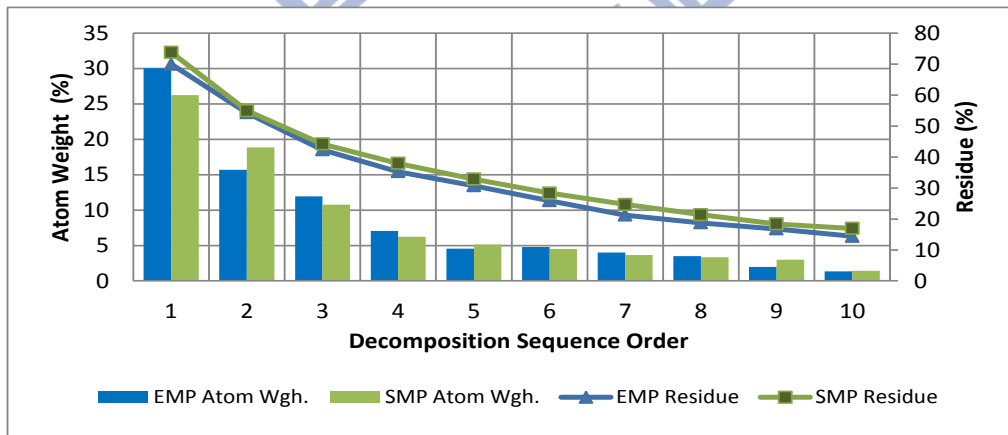
Figure 23 Residue and Atom Comparison between Exhaustive MP & Durka's SMP

#	Tme Shift	Freq. Shift	Scale	Phase (π)	Wgh. (%)	Residue (%)
1	0.52	1.70	0.71	1.37	30.04	69.96
2	0.32	6.24	0.15	1.35	15.70	54.26
3	0.50	10.69	0.08	1.09	11.95	42.31
4	1.45	6.21	0.60	0.79	7.07	35.24
5	0.32	14.63	0.15	0.06	4.56	30.68
6	1.32	12.74	0.71	1.70	4.79	25.89
7	0.16	25.89	1.51	0.07	4.02	21.25
8	0.83	4.00	0.32	0.71	3.51	18.74
9	1.25	2.04	0.32	0.89	1.99	16.75
10	0.10	15.17	0.19	0.54	1.35	14.40

Table 20 Gabor Atom List Generated by Exhaustive MP (ICA #20)

#	Time Shift	Freq. Shift	Scale	Phase (π)	Wgh. (%)	Residue (%)
1	0.62	2.07	0.53	1.70	26.22	73.78
2	0.27	2.76	0.05	0.57	18.85	54.93
3	0.50	11.03	0.08	1.14	10.75	44.18
4	1.33	4.83	0.16	1.32	6.23	37.95
5	1.27	13.10	0.43	0.41	5.12	32.83
6	0.62	4.83	0.49	0.64	4.50	28.33
7	0.34	26.21	0.5	1.65	3.64	24.69
8	0.07	0.69	0.23	1.73	3.34	21.35
9	0.39	11.72	0.10	0.30	2.97	18.38
10	1.10	5.52	0.19	0.47	1.43	16.95

Table 21 Gabor Atom List Generated by Durka's Stochastic MP (ICA #20)

Then, the correctness and precision of the decomposition results should be examined. Figure 24 and Figure 25 present the projection of the objective function on the axis of each parameter. For example, the "Time Shift Scanning" sub-graphs represent the changes of fitness values corresponding to the changes of time centers, while the other parameters are fixed during the scanning process.
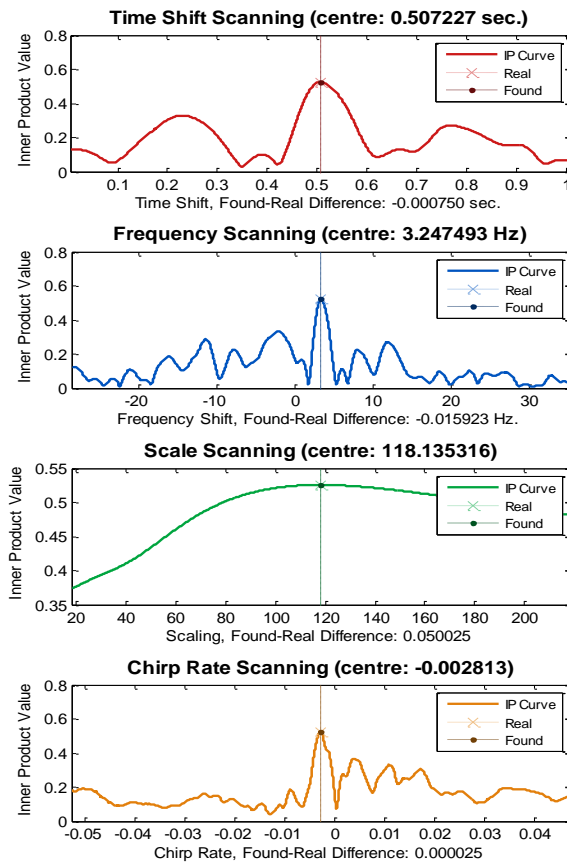
Figure 24 EMP on ICA #20
Chirplet Atom 1 in Table 9
(Desired Result)
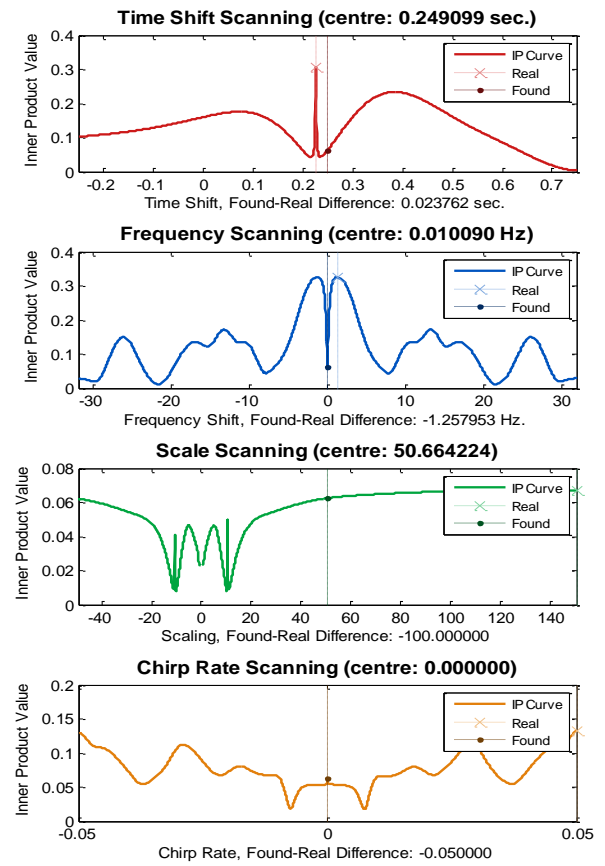
Figure 25 EMP on ICA #20
Chirplet Atom 2 in Table 9
(Typical Failure)

In Figure 24, it shows a perfectly optimized result of CMA-ES, each variable are closely lying beside their optimal values, which maximize the inner product of the atom waveform and the signal and consequentially, position themselves on the top of concaves with derivatives close to zero in their vicinities.

However Figure 25 presents another circumstances, which is a typical failure of insufficient sampling on large scaling, although the searched parameters are in close

proximity of the optimal value only with differences in only a small amount of samples (the resolutions are $\Delta t = 0.05$ s and $\Delta\omega = 0.689$ Hz. Therefore, the time difference in the number of Δt is approximately $0.024s \cong 0.48\,\Delta t$, and the frequency difference in the number of $\Delta\omega$ is $1.26\,Hz \cong 1.824\,\Delta\omega$). Notice that the real optimal value for scaling is quite large and locate on the edge of the objective curve. This is a classical degenerated case usually occurs when the scaling factor is too large in this case. The objective value can change drastically when modifying other parameters such as frequency and time centers due to the dilation and shrinkage of the resolution respectively. In the future design of the search algorithm, the dependencies of parameter should be considered in the design.

Still, the decompositions are precise enough to achieve its goal, but the heavily reliance on the exhaustive sampling in the 3-dimensional parameter space is the primary drawback. During the exhaustive scanning process, the number of samples in the EEG signal is 290, it requires $290 \times 145 \times 290 = 12,194,500$ iteration to acquire the densely sampled grid, and takes about 4 to 6 minutes for each Gabor atom in the table even on a 2.4GHz, Quad-core computer with 4GB RAM even written in C.

In the following section, we'll discover the mathematical properties of the Gabor atoms and used the derived result to decrease the cost of computation during the decompositions.

Chapter 4. Natural Gradient Search over Riemannian Space

The former experiment result of atoms clustering and concentration of signal energy has brought one positive response for the proclaimed quasi-sparse property of the EEG signal, but through an instinctive and lubberly method. In this chapter, the behavior of scaling (scaling factor) variables will be discovered to be non-Euclidean (non-linear); traditional optimization method ignored the impact of scaling factors on the time and frequency resolution, causing excessively redundant sampling and calculation at low resolution but probably insufficient at high. Therefore, in this case, the performance and precision of traditional approaches might be doubtfully satisfactory.

4.1.   Geometry of the Gabor Parameter Space

One of the major problems encountered when locating the candidates is to determine the way how algorithms take samples on the Gabor parameter space; instead of generating equidistance samples independently on the time, frequency and scale axes, the concept of Grassmannian frame is introduced to create dictionary with minimal mutual coherence.

The search domain for the real Gabor atom parameters space is practically three-dimensional; the phase ϕ is usually optimized in each step of matching procedure after the remaining three parameters are determined.

4.2.   Uniform Sampling on Gabor Parameter Space

The following figure is adopted from [42]. Case (a) is a simple example of failure that taking uniform samples with meridians and circles of latitude failed to achieve uniform sampling on a curved surface, whilst the case (b) is a more desirable case.

<center>(a)   (b)</center>
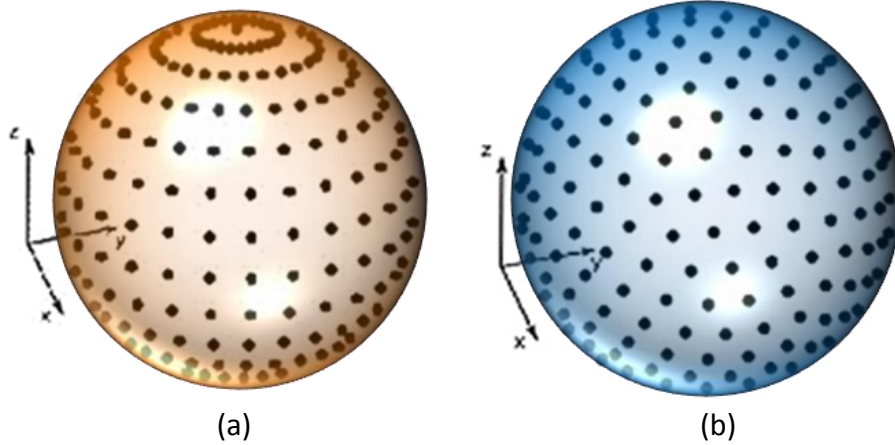
Figure 26 Uniform Sampling on Y and Z Axis (a) and a More Desirable Distribution of
Points Approximating Uniform Sampling on a Sphere (b)

The methods of taking samples in the Gabor parameter space can significantly affect the computational efficiency when performing optimizations or learning. In this section, the characteristic of the Gabor parameters will be discussed in details. Here, the complex Gabor atom (Equation (19) is used for calculation convenience during the analysis.

$$g_\gamma = K(\gamma)e^{-\pi(t-t_0)^2/s^2}e^{i\omega_0 t} \tag{19}$$

As mentioned before in 2.4 and explained in 3.6, dictionaries with small mutual coherence values gain favor in the case of sparse signal processing. Since the mutual coherence value of a dictionary is the maximum inner product value between the atoms within; it is intuitive to associate them with equiangular vector sets where any biasing will produce a greater, however undesired inner product value in Equation (5.

The sampling methods for time shift and frequency modulation are already described in [22]; followings are the derivations for how one should take samples on the Gabor parameters of scale factor s and chirp rate c.

Since the Gaussian integration and the behavior of the Heisenberg box remain unchanged regardless of the time and frequency center, the analyses will focus on the Gabor atoms centered at $(t_0, \omega_0) = (0, 0)$.

In the current construction method of the Grassmannian Gabor frame, the time and

<center>38</center>

frequency lattice Λ mentioned in 2.4 will be created first with a given redundancy ρ and a default value of scale s_0. To preserve the Grassmannian property, the value the rest of the parameters should be adjusted under the restriction of not to increase the mutual coherence of the dictionary set up by the given redundancy ρ.

4.2.1.   Verification of the Use of the Sphere Packing Concept

Here, the shape of the energy distribution of a Gabor atom on the time-frequency plane is going to be verified. The Wigner-Ville energy density of any waveform f can be written as:

$$P_V f(t, \omega) = \int f\left(t + \frac{\tau}{2}\right) \cdot f^*\left(t - \frac{\tau}{2}\right) e^{-j\omega\tau} \, d\tau \qquad (20$$

If the waveform $f = \exp[-t^2/\sigma^2]$ is a Gaussian function, then the Wigner-Ville energy distribution is:

$$P_V f(t_0, \omega_0) = \int \exp\left[-\frac{\left(t_0 + \frac{\tau}{2}\right)^2}{\sigma^2}\right] \cdot \exp\left[-\frac{\left(t_0 - \frac{\tau}{2}\right)^2}{\sigma^2}\right] e^{-j\omega_0\tau} \, d\tau$$

$$= \int \exp\left[-\frac{t_0^2 + \frac{\tau^2}{4}}{\sigma^2} - j\omega_0\tau\right] d\tau = 2\sigma\sqrt{\pi} \cdot \exp\left[-\frac{t_0^2}{\sigma^2} - \sigma^2\omega_0^2\right]$$

The shape of the contour of $P_V f$ at any given energy level c can be identified using:

$$P_V f(t_0, \omega_0) = c \quad \rightleftharpoons \quad 2\sigma\sqrt{\pi} \cdot \exp\left[-\frac{t_0^2}{\sigma^2} - \sigma^2\omega_0^2\right] = c$$

$$\Rightarrow \exp\left[-\frac{t_0^2}{\sigma^2} - \sigma^2\omega_0^2\right] = \frac{c}{2\sigma\sqrt{\pi}}$$

Re-arrange the equation, and finally there's a conic section with coordinate (t_0, ω_0):

$$\frac{t_0^2}{\sigma^2} + \sigma^2\omega_0^2 = \ln\left(2\sigma\sqrt{\pi}\right) - \ln c \qquad (21$$

Since both σ^2 and $1/\sigma^2$ are positive real numbers, the contour shape of the energy distribution $P_V f$ is an ellipse on the time-frequency plane. Therefore, using sphere packing to construct a Grassmannian frame is quite logical.

4.2.2. Relationship between Mutual Coherence μ and Redundancy ρ

For a Gaussian function with unit standard deviation $f(x) = e^{-x^2}$, the Gaussian integral, also known as the Euler-Poisson integral, can be formulated as:

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi} \tag{22}$$

The $\mathbf{L}^2$ energy of the unit σ and the dilated Gaussian function should be:

$$\|f(x)\|^2 \triangleq \int_{-\infty}^{\infty} f(x) \cdot f^*(x)dx = \int_{-\infty}^{\infty} e^{-2x^2} dx = \sqrt{\frac{\pi}{2}} \tag{23}$$

An energy-normalized Gaussian function can be written as

$$f_n(x) = \sqrt[4]{2/\pi} e^{-x^2} \tag{24}$$

Based on (22, if the standard deviation σ of the Gaussian function has become $\sigma = s$, the Gaussian integral will be amplified or diminished at with the ratio of s:

$$\int_{-\infty}^{\infty} f(x/s)dx = \int_{-\infty}^{\infty} e^{-(x/s)^2} dx = s\sqrt{\pi} \tag{25}$$

Using equation (25, the $\mathbf{L}^2$ energy of the unit σ and the dilated Gaussian function can therefore be calculated, by definition:

$$\|f(x/s)\|^2 \triangleq \int_{-\infty}^{\infty} f(x/s) \cdot f^*(x/s)dx = \int e^{-2(x/s)^2} dx = s\sqrt{\frac{\pi}{2}} \tag{26}$$

Notice that the Gabor atoms waveform with $s = 1$ have the Gaussian window with standard deviation of $\sqrt{\pi^{-1}}$.

$$\begin{cases} g(x) & = e^{-\pi(x/s)^2} & \Rightarrow \text{Gabor Atom} \\ \int_{-\infty}^{\infty} g\left(\frac{x}{s}\right) & = s & \Rightarrow \text{Gaussian Integral} \\ \|g\| = e^{-\pi(x/s)^2} & = \frac{s}{\sqrt{2}} & \Rightarrow \text{Norm} \end{cases} \tag{27}$$

Let $K = \sqrt[4]{2}/\sqrt{s}$ be the energy normalization factor, given the generating matrix L for the unit scaling Gabor atoms and redundancy ρ:

$$L = \begin{bmatrix} \ell_{1,1} & \ell_{1,2} \\ \ell_{2,1} & \ell_{2,2} \end{bmatrix} = \begin{bmatrix} \dfrac{\sqrt{2}}{\sqrt[4]{3}\sqrt{\rho}} & \dfrac{1}{\sqrt[4]{3}\sqrt{2\rho}} \\ 0 & \dfrac{\sqrt[4]{3}}{\sqrt{2\rho}} \end{bmatrix}$$

The mutual coherence of the resulted dictionary $\mathcal{D}$ will be determined by the adjacent atom having the Manhattan distance of 1.
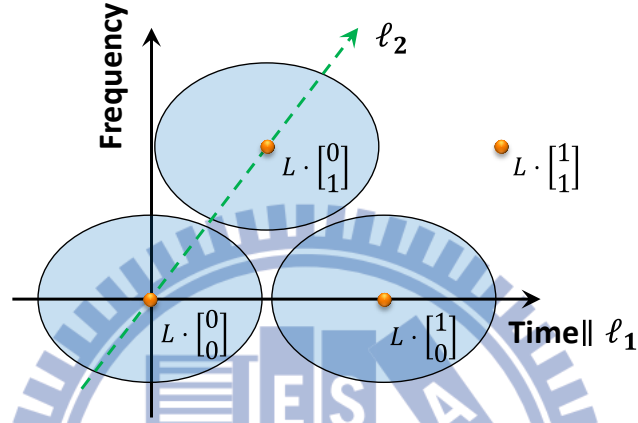
Figure 27   Energy Distribution of 3 Adjacent Gaussian Atoms Generated using Grassmannian Lattice Points

Since the lattice is proved to generate Grassmannian Gabor dictionary, all adjacent atom will certainly have the same inner product value as the mutual coherence of the dictionary. Therefore, in this case, the mutual coherence of the entire dictionary can be calculated using the inner product of any pairs of adjacent atoms.

Let g_{n1} be the atom centered at zero time and frequency, with a unit scaling factor. And, g_{n2} be the atom next to g_{n1}, different by a time shift of $\ell_{1,1}$.

The inner product of g_{n1} and g_{n2} is:

$$\langle g_{n1}, g_{n2} \rangle = \int_{-\infty}^{\infty} K e^{-\pi\left(\frac{x}{s}\right)^2} K e^{-\pi\left(\frac{x-\ell_{1,1}}{s}\right)^2} dx = K^2 \int_{-\infty}^{\infty} \exp\left[-\frac{\pi\left(2x^2 - 2\ell_{1,1}x + \ell_{1,1}^2\right)}{s^2}\right] dx$$

$$= K^2 \int_{-\infty}^{\infty} \exp\left[-\pi\left(2\left(x - \frac{\ell_{1,1}}{2}\right)^2 + \frac{\ell_{1,1}^2}{2}\right)/s^2\right] d$$

$$= \frac{K^2}{\sqrt{e^{\ell_{1,1}^2}}} \int_{-\infty}^{\infty} \exp\left[-2\pi\left(x - \frac{\ell_{1,1}}{2}\right)^2/s^2\right] dx = \frac{1}{e^{\pi/\sqrt{3}\rho}}$$

Therefore, the mutual coherence of the dictionary should be:

$$\mu(\mathcal{D}) = \frac{1}{e^{1/\sqrt{3}\rho}} \qquad (28$$

The relation between redundancy ρ and mutual coherence $\mu(\mathcal{D})$ can be shown as in Figure 28, notice that the value is independent of any scaling values.
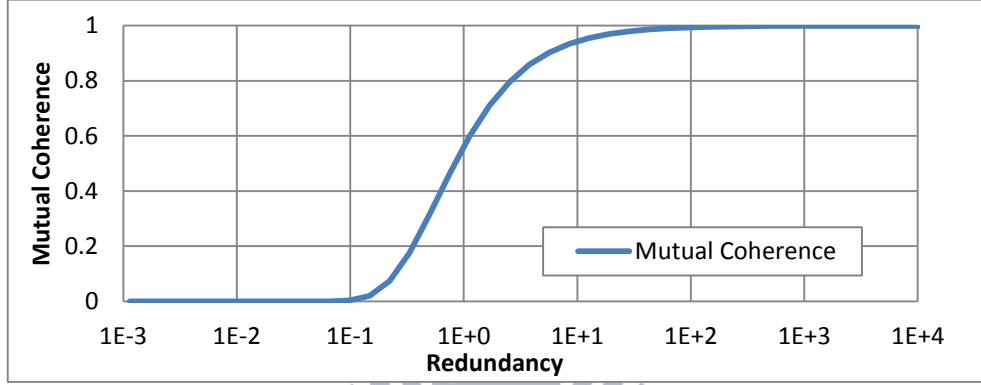
Figure 28   Relation between Dictionary Mutual Coherence and Redundancy

This proof can be applied on the seven other equivalent lattices of L, which are $\left(\text{let } \kappa = \sqrt{2}/\sqrt[4]{3}\sqrt{\rho}\right)$

$$L_1 = \kappa \begin{bmatrix} s & \dfrac{s}{2} \\ 0 & -\dfrac{\sqrt{3}}{2s} \end{bmatrix}, L_2 = \kappa \begin{bmatrix} s & -\dfrac{s}{2} \\ 0 & \dfrac{\sqrt{3}}{2s} \end{bmatrix}, L_3 = \kappa \begin{bmatrix} s & -\dfrac{s}{2} \\ 0 & -\dfrac{\sqrt{3}}{2s} \end{bmatrix},$$

$$L_4 = \kappa \begin{bmatrix} -s & \dfrac{s}{2} \\ 0 & \dfrac{\sqrt{3}}{2s} \end{bmatrix}, L_5 = \kappa \begin{bmatrix} -s & \dfrac{s}{2} \\ 0 & -\dfrac{\sqrt{3}}{2s} \end{bmatrix}, L_6 = \kappa \begin{bmatrix} -s & -\dfrac{s}{2} \\ 0 & \dfrac{\sqrt{3}}{2s} \end{bmatrix}, L_7 = \kappa \begin{bmatrix} -s & -\dfrac{s}{2} \\ 0 & -\dfrac{\sqrt{3}}{2s} \end{bmatrix}$$

4.2.3.   Behavior of Increment of the Scaling Factor (Conjecture)

Since the atoms employed in the pursuit algorithm should be normalized in terms of energy. The energy-normalized Gabor functions with $s = 1$ can be formulated as:

$$\begin{cases} g_n(x) = \sqrt[4]{2}e^{-\pi x^2} \\ g_{sn}(x) = \dfrac{\sqrt[4]{2}}{\sqrt{s}}e^{-\pi(x/s)^2} \end{cases} \qquad (29$$

Now, assumed that there are only two normalized atoms with different scaling

factors $s_1, s_2 \in \mathbb{R}$ in the dictionary $\mathcal{D}$, assume that $s_2 > s_1 > 0$:

$$\begin{cases} g_{s_1 n}(x) = \dfrac{\sqrt[4]{2}}{\sqrt{s_1}} e^{-\pi(x/s_1)^2} \\[3mm] g_{s_2 n}(x) = \dfrac{\sqrt[4]{2}}{\sqrt{s_2}} e^{-\pi(x/s_2)^2} \end{cases} \tag{30}$$

Their inner product $\langle g_{s_1 n}(x), g_{s_2 n}(x) \rangle$ is the mutual coherence value $\mu(\mathcal{D})$.

To observe the behavior of atom-wise correlation value when adjusting the scaling factor, try adding a new atom into $\mathcal{D}$ with a new scaling factor $s > s_2$ without affecting the mutual coherence $\mu(\mathcal{D})$.

To maintain the mutual coherence of the dictionary, the atom correlation value must follow that $\left| \langle g_{s_1 n}(x), g_{s_2 n}(x) \rangle \right| \geq \left| \langle g_{s_2 n}(x), g_{sn}(x) \rangle \right|$.

$$\left| \int_{-\infty}^{\infty} \frac{\sqrt[4]{2}}{\sqrt{s_1}} e^{-\pi(x/s_1)^2} \frac{\sqrt[4]{2}}{\sqrt{s_2}} e^{-\pi(x/s_2)^2} \, dx \right|$$

$$\geq \left| \int_{-\infty}^{\infty} \frac{\sqrt[4]{2}}{\sqrt{s_1}} e^{-\pi(x/s_2)^2} \frac{\sqrt[4]{2}}{\sqrt{s_2}} e^{-\pi(x/s)^2} \, dx \right| \tag{31}$$

$$\Rightarrow \left| \sqrt{s} \int_{-\infty}^{\infty} \exp\left[-\frac{x^2(s_1^2 + s_2^2)}{s_1^2 s_2^2} \right] dx \right| \geq \left| \sqrt{s_1} \int_{-\infty}^{\infty} \exp\left[-\frac{x^2(s_2^2 + s^2)}{s_2^2 s^2} \right] dx \right|$$

The positivity and strictly monotonically increasing characteristic of the exponential function in the real number domain implies:

$$\frac{s_1 s_2 \sqrt{s}}{\sqrt{s_1^2 + s_2^2}} \geq \frac{s \, s_2 \sqrt{s_1}}{\sqrt{s_2^2 + s^2}} \Rightarrow s_1(s_2^2 + s^2) \geq s(s_1^2 + s_2^2)$$

$$\Rightarrow s \geq \frac{s_2^2}{s_1} \quad (\text{or} \quad s s_1 \geq s_2^2 \text{ since } s_1, s_2, s > 0) \tag{32}$$

If the equality holds, the increasing sequence s_1, s_2, s is a geometric progression, one can easily discover that $s_2 = \sqrt{s_1 s}$. This suggests that the increment of the scaling factor while preserving a constant correlation value between adjacent atoms should be exponential. This process can be illustrated as in Figure 29; the expected result should be a vector that makes one of the given vectors lies on the equiangular surface defined by another given vector and the newly found one.
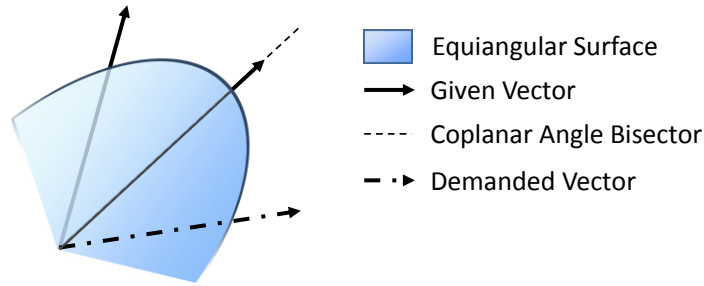
Figure 29   Searching for Additional Vectors without Lowering Mutual Coherence

This method will be applied later to calculate other behaviors of the Gabor parameters and the theoretical result will be realized in the devising of a method to construct a sub-optimal Grassmannian frame.

4.2.4.   Scaling Factor and Mutual Coherence (Conjecture)

From 4.2.3, the idea of treating the logarithm of the scaling factor s as linear is verified; the impact of the adjusting the scaling factor on the time-frequency energy distribution of an atom can be illustrated in Figure 30.
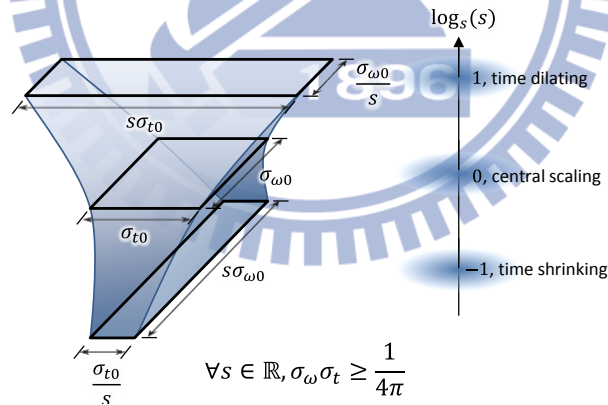
Figure 30   Dilation/Shrinking Effect of Scaling Factors

Now it's time to calculate the equivalence parametric distance between the time shift and scaling (dilation). Since the growth of scaling factor is geometric and the time shifting is additive, one can assume that the variation in s is ks, but the variation in x is $(x + \Delta x)$.

To maintain the mutual coherence amongst the existed atom, the following condition must be hold:

$$\left| \int_{-\infty}^{\infty} \underbrace{\frac{\sqrt[4]{2}}{\sqrt{s}} e^{-\pi\left(\frac{x}{s}\right)^2}}_{\text{original atom}} \cdot \underbrace{\frac{\sqrt[4]{2}}{\sqrt{ks}} e^{-\pi\left(\frac{x}{ks}\right)^2}}_{\text{dilated atom A}} dx \right| \geq \left| \int_{-\infty}^{\infty} \underbrace{\frac{\sqrt[4]{2}}{\sqrt{s}} e^{-\pi\left(\frac{x}{s}\right)^2}}_{\text{original atom}} \cdot \underbrace{\frac{\sqrt[4]{2}}{\sqrt{s}} e^{-\pi\left(\frac{x+\Delta x}{s}\right)^2}}_{\text{time shifted atom}} dx \right| \quad \text{(33}$$

$$\Rightarrow \left| \int_{-\infty}^{\infty} \frac{1}{s \cdot \sqrt{k}} e^{-\left(\frac{x}{s}\right)^2 - \left(\frac{x}{ks}\right)^2} dx \right| \geq \left| \int_{-\infty}^{\infty} \frac{1}{s} e^{-\left(\frac{x}{s}\right)^2 - \left(\frac{x+\Delta x}{s}\right)^2} dx \right|$$

Remove the common factor $1/s$ and since $\pi > 0$:

$$\frac{1}{\sqrt{k}} \left| \int_{-\infty}^{\infty} \exp\left[-\pi \left(\frac{k^2 x^2 + x^2}{k^2 s^2} \right) \right] dx \right| \geq \left| \int_{-\infty}^{\infty} \exp\left[-\pi \left(\frac{2x^2 + 2x\Delta x + \Delta x^2}{s^2} \right) \right] dx \right|$$

$$\Rightarrow \frac{1}{\sqrt{k}} \left| \int_{-\infty}^{\infty} \underbrace{\exp\left[-\left(\frac{1+k^2}{k^2} \right) x^2 \right]}_{\text{Dilated Gaussian}} dx \right|$$

(34

$$\geq \left| \int_{-\infty}^{\infty} \exp\left[-2\left(x^2 + x\Delta x + \frac{\Delta x^2}{4} \right) \right] \cdot \underbrace{\exp\left[-\frac{\Delta x^2}{2} \right]}_{\text{Constant}} dx \right|$$

Move the scalars outside and simplify the Gaussian integrals, therefore:

$$\begin{cases} \dfrac{1}{\sqrt{k}} \left| \displaystyle\int_{-\infty}^{\infty} \underbrace{\exp\left[-\left(\frac{1+k^2}{k^2} \right) x^2 \right]}_{\text{Dilated Gaussian}} dx \right| = \dfrac{1}{\sqrt{e^{\Delta x^2}}} \left| \displaystyle\int_{-\infty}^{\infty} \exp\left[-2\left(x + \frac{\Delta x}{2} \right)^2 \right] dx \right| = \sqrt{\dfrac{k\pi}{1+k^2}} \\[3em] \left| \displaystyle\int_{-\infty}^{\infty} \exp\left[-2\left(x^2 + x\Delta x + \frac{\Delta x^2}{4} \right) \right] \cdot \underbrace{\exp\left[-\frac{\Delta x^2}{2} \right]}_{\text{Constant}} dx \right| = \dfrac{1}{\sqrt{e^{\Delta x^2}}} \left| \displaystyle\int e^{-2x^2} dx \right| = \sqrt{\dfrac{\pi}{2e^{\Delta x^2}}} \end{cases}$$

Finally, according to the inequality (34:

$$\frac{1+k^2}{k} \geq 2e^{\Delta x^2} \rightleftharpoons k^2 - 2ke^{\Delta x^2} + 1 \geq 0 \quad \text{(35}$$

Using the solution form of quadratic equation $ax^2 + bx + c = 0$ of single variable x:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Since the polynomial is convex, the suitable range for k should be:

$$k \geq e^{\Delta x^2} + \sqrt{e^{2\Delta x^2} - 1} \quad \text{or} \quad k \leq e^{\Delta x^2} - \sqrt{e^{2\Delta x^2} - 1}$$

Notice that there are two possible ranges for k, and one can easily discover their product have the following property:

$$\underbrace{\left(e^{\Delta x^2} + \sqrt{e^{2\Delta x^2} - 1}\right)}_{\text{Expanding Factor}} \cdot \underbrace{\left(e^{\Delta x^2} - \sqrt{e^{2\Delta x^2} - 1}\right)}_{\text{Shrinking Factor}} = e^{2\Delta x^2} - e^{2\Delta x^2} + 1 = 1 \quad \text{(36}$$

4.2.5. *Verifications of Scalable Grassmannian Dictionary*

Statement I: The mutual coherence $\mu(\mathcal{D})$ of dictionary $\mathcal{D}(g, \Lambda_s)$ is determined by a unit shifted atoms with parameter center, i.e.

$$\text{If } \lambda_a = L \begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{ and } \lambda_b = L \begin{bmatrix} 0 \\ 1 \end{bmatrix} \text{ then } \mu(\mathcal{G}) = \left| \langle g_O, g_{\lambda_a} \rangle \right| = \left| \langle g_O, g_{\lambda_b} \rangle \right|$$

Proof: For any 2 atoms with parameter t_0, ω_0, s and $t_0 + \Delta t, \omega_0 + \Delta \omega, s$, their inner product value is

$$\int_{-\infty}^{\infty} \frac{\sqrt{2}}{s} \cdot \exp\left(-\frac{\pi(t - t_0)^2}{s^2} + j\omega_0 t - \frac{\pi(t - t_0 - \Delta t)^2}{s^2} - j(\omega_0 + \Delta \omega)t\right) dt$$

$$= \frac{\sqrt{2}}{s} \int_{-\infty}^{\infty} \exp\left(-\frac{\pi t^2}{s^2} - \frac{\pi(t - \Delta t)^2}{s^2} - j\Delta \omega t\right) dt$$

$$= \frac{\sqrt{2}}{s} \int_{-\infty}^{\infty} \exp\left(-\frac{\pi}{s^2}(2t^2 - 2t \cdot \Delta t + \Delta t^2) - j\Delta \omega t\right) dt$$

$$= \frac{\sqrt{2}}{s} \int_{-\infty}^{\infty} \exp\left(-\frac{2\pi}{s^2}\left(t - \frac{\Delta t}{2}\right)^2 - \frac{\pi}{2s^2}\Delta t^2 - j\Delta \omega t\right) dt$$

$$= \frac{\sqrt{2}}{s} \cdot \exp\left(-\frac{\pi \Delta t^2}{2s^2}\right) \int_{-\infty}^{\infty} exp\left(-\frac{2\pi}{s^2}\left(t - \frac{\Delta t}{2}\right)^2 - j\Delta \omega t\right) dt$$

$$= \exp\left(-\left(\frac{\pi \Delta t^2}{2s^2} + j\Delta t \Delta \omega + \frac{\Delta \omega^2 s^2}{8\pi}\right)\right)$$

The magnitude is:

$$\Re\left[\exp\left(-\left(\frac{\pi \Delta t^2}{2s^2} + j\Delta t \Delta \omega + \frac{\pi \Delta \omega^2 s^2}{2}\right)\right)\right] = \exp\left(-\left(\frac{\pi \Delta t^2}{2s^2} + \frac{\Delta \omega^2 s^2}{8\pi}\right)\right)$$

The result is a 2D Gaussian function centered at (0,0), since

$$
\begin{cases}
\lambda_a = \kappa \begin{bmatrix} s \\ 0 \end{bmatrix} = \begin{bmatrix} \dfrac{\sqrt{2}s}{\sqrt[4]{3}\sqrt{\rho}} \\ 0 \end{bmatrix} \\[4ex]
\lambda_a = \kappa \begin{bmatrix} \dfrac{s}{2} \\ \dfrac{\sqrt{3}}{2s} \end{bmatrix} = \begin{bmatrix} \dfrac{s}{\sqrt[4]{3}\sqrt{2\rho}} \\ \dfrac{\sqrt[4]{3}}{s\sqrt{2\rho}} \end{bmatrix}
\end{cases}
$$

And using $\omega = 2\pi f$ One can easily find out that

$$
\left|\langle g_O, g_{\lambda_a} \rangle\right| = \left|\langle g_O, g_{\lambda_b} \rangle\right| = \left|\langle g_{\lambda_a}, g_{\lambda_b} \rangle\right| = e^{-\pi/\sqrt{3}\rho} = \mu(\mathcal{D})
$$

And it can be applied on every equivalent lattice of L.

Statement II: For a Gabor dictionary $\mathcal{D}_{s_0}$ having scale $s = s_0$ with redundancy ρ and the mutual coherence $e^{-\pi/\sqrt{3}\rho}$, any atoms with the expanded scaling s_D

$$
s = s_D = r s_0 = s_0 \cdot \left(\exp\left(\frac{2}{\sqrt{3}\cdot\rho}\right) + \sqrt{\exp\left(\frac{4}{\sqrt{3}\cdot\rho}\right) - 1} \right)
$$

Will have the inner product with any atoms in $\mathcal{D}_{s_0}$ smaller or equal than $\mu(\mathcal{D}_{s_0})$, i.e.

$$
\forall g_k \in \mathcal{D}_{s_0}, \text{ and } t_0, \omega_0 \in \mathbb{R}, \qquad \left|\langle g_{(t_0,\omega_0,s_D)}, g_k \rangle\right| \leq \mu(\mathcal{D}_{s_0}) = e^{-1/\sqrt{3}\rho}
$$

Proof: For any two atoms with parameter $\gamma_1 = (t_0, \omega_0, 1)$ and $\gamma_2 = (t_0 + \Delta t, \omega_0 + \Delta\omega, s_D)$, their inner product $\langle g_{\gamma_1}, g_{\gamma_2} \rangle$ is

$$
\int_{-\infty}^{\infty} \frac{\sqrt{2}}{\sqrt{s_D}} \cdot \exp\left(-\frac{\pi(t-t_0)^2}{s_0^2} + j\omega_0 t - \frac{\pi(t-t_0-\Delta t)^2}{s_D^2} - j(\omega_0 + \Delta\omega)t \right) dt
$$

$$
= \frac{\sqrt{2}}{\sqrt{s_D}} \int_{-\infty}^{\infty} \exp\left(-\frac{\pi t^2}{s_0^2} - \frac{s_0^2}{s_D^2}\frac{\pi(t-\Delta t)^2}{s_0^2} - j\Delta\omega t \right) dt
$$

$$
= \frac{\sqrt{2}}{\sqrt{r}} \int_{-\infty}^{\infty} \exp\left(-\pi \left(\frac{r^2+1}{r^2}\right)\left(t^2 + \frac{2t\Delta t}{r^2+1} + \frac{\Delta t^2}{r^2+1} \right) - j\Delta\omega t \right) dt
$$

$$= \frac{\sqrt{2}}{\sqrt{r}} \int_{-\infty}^{\infty} \exp\left(-\pi\left(\frac{r^2+1}{r^2}\right)\left(\left(t^2+\frac{\Delta t}{r^2+1}\right)^2\right) + \frac{\Delta t^2}{(r^2+1)r^2} - \frac{\Delta t^2}{r^2} - j\Delta\omega t\right) dt$$

$$= \frac{\sqrt{2}}{\sqrt{r}} \int_{-\infty}^{\infty} \exp\left(-\pi\left(\frac{r^2+1}{r^2}\right)\left(t+\frac{\Delta t}{r^2+1}\right)^2 + \frac{\Delta t^2}{(r^2+1)r^2} - \frac{\Delta t^2}{r^2} - j\Delta\omega t\right) dt$$

$$= \frac{\sqrt{2}}{\sqrt{r}} \cdot \exp\left(-\frac{\pi\Delta t^2}{r^2}\right) \int_{-\infty}^{\infty} \exp\left(-\pi\left(\frac{r^2+1}{r^2}\right)\left(t+\frac{\Delta t}{r^2+1}\right)^2 - j\Delta\omega t\right) dt$$

$$= \sqrt{\frac{2r}{r^2+1}} \cdot \exp\left(-\frac{\pi\Delta t^2}{r^2}\right) \cdot \exp\left(-\frac{4\pi\Delta\omega\Delta t + j\Delta\omega^2 r^2}{4\pi(r^2+1)}\right)$$

The magnitude is

$$|\langle g_{\gamma_1}, g_{\gamma_2}\rangle| = \sqrt{\frac{2r}{r^2+1}} \cdot exp\left(-\frac{\pi\Delta t^2}{r^2}\right) \cdot exp\left(-\Im\left(\frac{4\pi\Delta\omega\Delta t + j\Delta\omega^2 r^2}{4\pi(r^2+1)}\right)\right)$$

$$= \sqrt{\frac{2r}{r^2+1}} \cdot exp\left(-\frac{\pi\Delta t^2}{r^2}\right) \cdot exp\left(-\frac{\Delta\omega^2 r^2}{4\pi(r^2+1)}\right)$$

Since $\Delta t, \Delta\omega \in \mathbb{R}$, the maximum occurs at $\Delta t = \Delta\omega = 0$, and the value is

$$\sqrt{\frac{2r}{r^2+1}} \cdot e^0 = \sqrt{\frac{\exp\left(\frac{2}{\sqrt{3}\cdot\rho}\right) + \sqrt{\exp\left(\frac{4}{\sqrt{3}\cdot\rho}\right)-1}}{\left(\exp\left(\frac{2}{\sqrt{3}\cdot\rho}\right) + \sqrt{\exp\left(\frac{4}{\sqrt{3}\cdot\rho}\right)-1}\right)^2 - 1}}$$

Let

$$q = \exp\left(\frac{2}{\sqrt{3}\cdot\rho}\right) > 0$$

$$\sqrt{\frac{\exp\left(\frac{2}{\sqrt{3}\cdot\rho}\right) + \sqrt{\exp\left(\frac{4}{\sqrt{3}\cdot\rho}\right)-1}}{\left(\exp\left(\frac{2}{\sqrt{3}\cdot\rho}\right) + \sqrt{\exp\left(\frac{4}{\sqrt{3}\cdot\rho}\right)-1}\right)^2 - 1}} = \sqrt{\frac{2q + 2\sqrt{q^2-1}}{\left(q+\sqrt{q^2-1}\right)^2 + 1}}$$

$$= \sqrt{\frac{2q + 2\sqrt{q^2-1}}{q^2 + 2q\sqrt{q^2-1} + q^2 - 1 + 1}} = \sqrt{\frac{2q + 2\sqrt{q^2-1}}{2q^2 + 2q\sqrt{q^2-1}}} = \sqrt{\frac{1}{q}} = \exp\left(\frac{-1}{\sqrt{3}\cdot\rho}\right) = \mu(\mathcal{D})$$

Conclusion: For the Gabor dictionary $\mathcal{D}_{s_0}$ having scale $s = s_0$, and the Gabor

dictionary $\mathcal{D}_{s_D}$ having scale s_D and the lattice generating matrix of L_{s_D}

$$s = s_D = r s_0 = \exp\left(\frac{2}{\sqrt{3} \cdot \rho}\right) + \sqrt{\exp\left(\frac{4}{\sqrt{3} \cdot \rho}\right) - 1}$$

$$L = \frac{\sqrt{2}}{\sqrt[4]{3}\sqrt{\rho}} \begin{bmatrix} s_D & \dfrac{s_D}{2} \\ 0 & \dfrac{\sqrt{3}}{2s_D} \end{bmatrix}$$

The merged dictionary has the mutual coherence $\mu\left(\mathcal{D}_{s_0} \cup \mathcal{D}_{s_D}\right) = \mu\left(\mathcal{D}_{s_0}\right)$

The proof can be easily extended, if $\mathcal{D}_{s_0}, \mathcal{D}_{s_1}, \mathcal{D}_{s_2} \dots$ are Gabor dictionaries generated using the hexagonal lattice with different scale $s_0, s_1, s_2 \dots$, and $s_1 = rs, s_2 = r^2 s, \dots, s_n = r^n s$, where the common ratio r being

$$r = \exp\left(\frac{2}{\sqrt{3} \cdot \rho}\right) + \sqrt{\exp\left(\frac{4}{\sqrt{3} \cdot \rho}\right) - 1}$$

Then the union dictionary $\mathcal{D}_{all}$

$$\mathcal{D}_{all} = \bigcup_i \mathcal{D}_{s_i}$$

will have the mutual coherence $\mu(\mathcal{D}_{all}) = \mu(\mathcal{D}_0) = \mu(\mathcal{D}_1) = \mu(\mathcal{D}_2) = \cdots$, which is a Grassmannian Gabor frame in $\mathbb{R}^2$ of the time frequency plane.

4.2.6. Behavior of the Increment of Chirp Rate Parameter (Conjecture)

Assume the dictionary $\mathcal{D}$ contains two normalized chirplet atoms $g_{c_1 n}$ and $g_{c_2 n}$ with chirp rate $c_1, c_2 \in \mathbb{R}$ share the same values in other parameters. Since the linear chirping term will not affect the energy of the atom, assume $c_2 > c_1 > 0$:

$$\begin{cases} g_{c_1 n}(x) = \sqrt[4]{2} \exp\left[\dfrac{jc_1 - 1}{2} x^2\right] \\ g_{c_2 n}(x) = \sqrt[4]{2} \exp\left[\dfrac{jc_2 - 1}{2} x^2\right] \end{cases} \tag{37}$$

Now find another chirplet atom g_{cn} with chirp rate $c > c_2$ and share the remaining parameters, such that $\left| \langle g_{c_1 n}(x), g_{c_2 n}(x) \rangle \right| \geq \left| \langle g_{c_2 n}(x), g_{cn}(x) \rangle \right|$. That is:

$$\left| \int_{-\infty}^{\infty} \sqrt[4]{2} \exp\left[\frac{jc_1 - 1}{2} x^2\right] \cdot \sqrt[4]{2} \exp\left[\frac{-jc_2 - 1}{2} x^2\right] dx \right|$$

$$\geq \left| \int_{-\infty}^{\infty} \sqrt[4]{2} \exp\left[\frac{jc_2 - 1}{2} x^2\right] \cdot \sqrt[4]{2} \exp\left[\frac{-jc - 1}{2} x^2\right] dx \right|$$

$$\Rightarrow \left| \int_{-\infty}^{\infty} \exp\left[\frac{j(c_1 - c_2) - 2}{2} x^2\right] dx \right| \geq \left| \int_{-\infty}^{\infty} \exp\left[\frac{j(c_2 - c) - 2}{2} x^2\right] dx \right|$$

$$\Rightarrow \left| \frac{1}{\sqrt{-2j(c_1 - c_2) + 4}} \right| \geq \left| \frac{1}{\sqrt{-2j(c_2 - c) + 4}} \right|$$

Finally,

$$c \geq 2c_2 - c_1 \implies \frac{c + c_1}{2} \geq c_2 \tag{38}$$

If the equality holds, the increasing sequence c_1, c_2, c is an arithmetic progression, one can easily find out that $c_2 = (c + c_1)/2$. This suggests that the increment of chirp rate while preserving a constant mutual coherence value of the dictionary should be arithmetic.

4.3. Summary

In this chapter, all required knowledge to construct a Grassmannian Gabor frame has been attained, which includes:

- *Verify the sphere packing of time-frequency centers on the TF plane*
- *Quantitate the exact relationship between the mutual coherence and the redundancy of a dictionary*
- *Identify the growth of the scaling factor, which is geometric*
- *Quantitate the exact increment ratio of the scaling factor without increasing the mutual coherence of a dictionary*
- *Identify the growth of the chirp rate parameter, which is linear (but unused)*

These facts and discoveries will be considered and applied during the design period of the Memetic search algorithm in the next chapter.

Chapter 5. Hybrid Memetic Natural Gradient Algorithm (HyMN-G)

This chapter describes the design details of the Hybrid Memetic Natural Gradient Algorithm for EEG signal decomposition. The main goal is to reduce the computation cost of searching in the Gabor parameter space by combining population based search algorithm and local optimization methods.

5.1.   Memetic Concept

The broad application of Memetic Algorithms in the field of modern computation is the result of compensations. Ordinary local search algorithms such as Gradient descent, suffered heavily from the stagnation in local minimum; whilst the population based methods, which are generally thought not to be trapped in the well of local minimum so easily if the distribution of the population is proper and the size of which is large enough, might take a serious amount of time to discover the optimal solution.

However, the advantages of those local search algorithms cannot be neglected. Their mathematical based theories, simplicity of implementation, and execution efficiency are the primary reason of their inveterate existence.

5.2.   Grassmannian Dictionary Initialization

For EEG and ICA components, the signals usually have their effective bandwidth lower than 100 Hz and the elapsing time of normal brain activities for an odd-ball experiment is usually shorter than 1 second.

The central scaling factor is set to 0.3 second in length, the reason is as followed: according to (1, the Gaussian window of the atom is $e^{-\pi(t-t_0)^2/s^2}$, therefore, the actual

standard deviation of that Gaussian window is $\sigma = 0.3\sqrt{\pi^{-1}} \cong 0.169 \ (s)$

According to the Three-Sigma rule, this window will have 68% of the sum of its support value in $\mathbf{L_1}$ concentrated in the range of $\pm\sigma$, which is equivalent to a time period of $0.169 \times 2 \cong 0.339 \ (s)$; 95% in the period of $0.677 \ (s)$ and 99.7% in $1.015 \ (s)$.

The creation of time and frequency parameter lattice is rather simple afterwards; the lattice points will pad the signal space uniformly after the scaling factor is decided.

Figure 31 is the scatter plot of the lattice points generated using the setting of $\rho = 0.5, s_0 = 0.3 \ s$ in the 3-dimensional space (this point of view does not have any physical meaning in signal processing; however, one can still observe the hexagonal lattice and the effect of different scaling factors on the distribution of points).

(a) 3D View

(b) Time-Scale Projection View

(c) Frequency-Scale Projection View
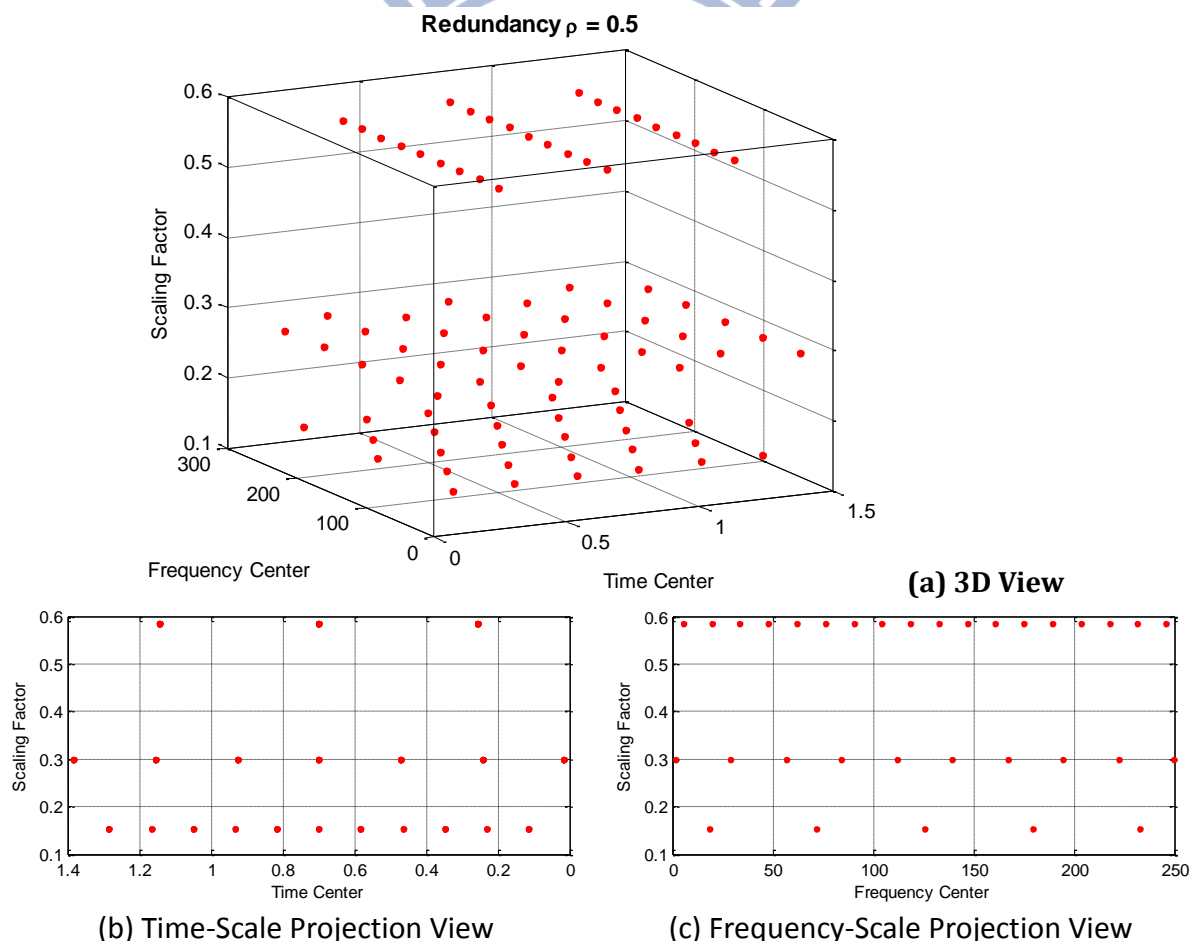
Figure 31   **Scatter Plot of the Gabor Parameter Centers ($\rho = 0.5, s_0 = 0.3 \ sec.$)**

Although the Gabor dictionary can well cover the effective signal space, however the artefacts, which might have relatively significant power over the EEG signals such as eye

movements, everlasting elapsing time and periodic such as civilian AC electricity source and ECG (heartbeats) or even unknown properties such as heat distortions, might significantly dropped the performance of the decomposition. The current solution is to use ICA to remove undesired component and re-project the component back to channel, even so, the small signal-to-noise ratio of raw EEG still poses insurmountable obstacle to accurate analysis.

5.3. Natural Gradient Optimization

Before the 3-dimensional time-frequency-scaling experiment, the application of efficient NES has brought prominent solution to the optimization problem. The following Figure 32 shows the convergence process of a Gabor atom with scaling factor $s = 0.3\,s$ and having its time frequency center values $(t_0, \omega_0) = (0.3s, 130\,\text{rad}/s)$.
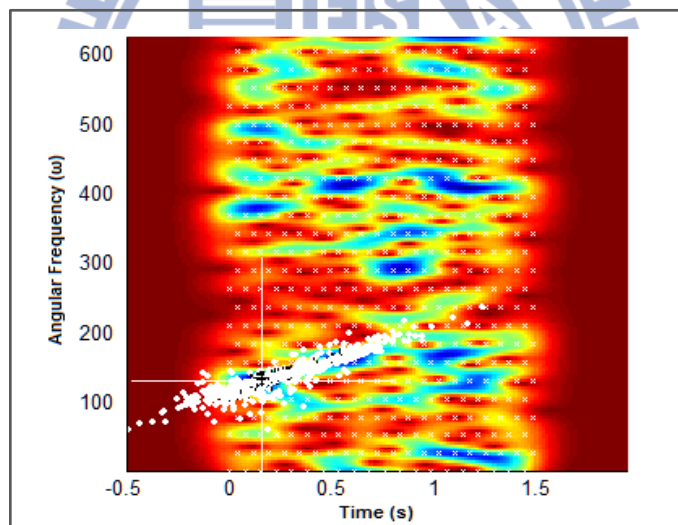
Figure 32   Natural Gradient Search over Signal Inner Product Space (ICA #20) Initialization Point = (0.3s, 130rad/s), Grid Redundancy ρ = 0.5

In the figure, colder colors represent lower objective function value (higher fitness). The big white '+' sign indicates the global minimum position, each white 'x' shows the position of the Grassmannian lattice point, and the white round points represent the distribution of population. In the 3-dimensional parameter space, the Memetic Pursuit algorithm is devised as:

Procedure **MEMETIC PURSUIT**

Input: Residue Waveform x,

    Generate uniform lattice $\Phi = \{\phi_1, \phi_2, \dots\}$ in Gabor parameter space

    For $j = 1$ to Memetic Iteration Limit **Do**

        Evaluate fitness $f(\phi_i) = -|\langle x, g_{\phi_i}\rangle|$ for all individual $\phi_i \in \Phi$

        Identify the subset $\Phi' = \{\phi'_1, \phi'_2, \dots\}$ of learning candidates

        Generate $\Phi^+ = \{\phi^+_1, \phi^+_2, \dots\}$ *by*

            Improving ϕ'_k with local Natural gradient optimization

        Set $\phi^* \leftarrow \text{argmax}_{\phi \in \Phi^+ \cup \Phi} f(\phi)$

        If stopping conditions met **Then Break**

        Select $\widehat{\Phi} = \{\hat{\phi}_1, \hat{\phi}_2, \dots\}$, $\widehat{\Phi} \subset (\Phi^+ \cup \Phi)$ of reproductive candidates

        Generate uniform proximity lattices $\Phi_1, \Phi_2, \dots$ centered at each $\hat{\phi}_i \in \widehat{\Phi}$

        Set $\Phi \leftarrow \cup \Phi_i$

    End For

End Procedure, Return decomposition sequence $\boldsymbol{\phi}^*$

<div align="center">Table 22   Memetic Pursuit Procedure</div>

The proximity lattices are generated using Manhattan distance; Figure 33 illustrates the shape of a proximity lattice with Manhattan distance of 2 (each white plate represents the surface of equal scaling, and the effect of different scaling on the distribution of time-frequency lattice points is ignored in the illustration).

$$s = \frac{a}{r^2} \qquad s = \frac{a}{r} \qquad s = a \qquad s = ar \qquad s = ar^2$$
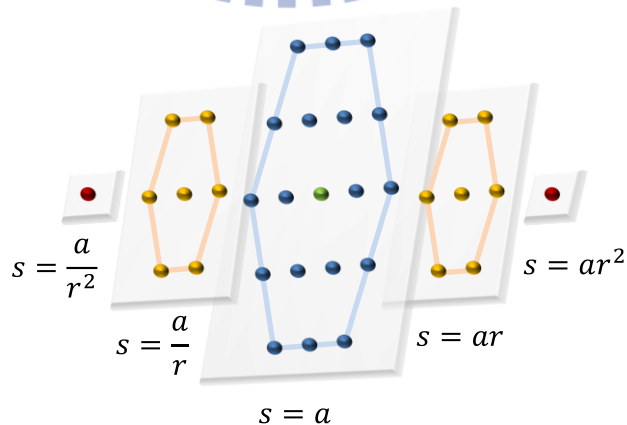
<div align="center">**Figure 33**   **Proximity Lattice of Manhattan Distance 2**</div>

The main functionality of the Natural Gradient optimization during the search of atom

parameters correlated the most with the signal is to perform appropriate local search; therefore, the search region will be restricted to cover only few number of grid points nearby, and let the population based optimization to handle the rest.

5.4.   Realisation: HyMN-G

Combining the method of lattice point generation, natural gradient optimization method and Memetic Algorithm, the new Hybrid Memetic Natural Gradient algorithm (HyMN-G) is developed to decompose the EEG signal in a more elegant way.

Algorithm   **HYBRID MEMETIC NATURAL GRADIENT**

Initialization: Atom sequence $L \leftarrow [\,]$

Input: Signal x, Energy extraction level $k\%$

  Calculate signal energy $\mathcal{E}_x = \|x\|^2$

  Set residue $r_0 \leftarrow x$

  For $i = 1$ to Atom Quantity Limit **Do**

    Perform Memetic Pursuit to find ϕ_i^*, $\phi_i^* = MemeticPursuit(r_{i-1})$
    Calculate the inner product $A_i \leftarrow \langle r_{i-1}, g_{\phi_i^*} \rangle$
    Calculate the residue of next iteration $r_i \leftarrow r_{i-1} - A \cdot g_{\phi_i^*}$
    Append the result to L, $L \leftarrow L \cup \{(A_i, \phi_i^*)\}$

    If $\|r_i\|^2 / \mathcal{E}_x \leq k\%$ **Then Break**

  End For

End Procedure, Return decomposition sequence L

Table 23 Hybrid Memetic Natural Gradient Algorithm

Similar to the exhaustive design of the previous work, the HyMN-G algorithm extends the belief of the greedy Matching Pursuit with the matching part of the atoms replaced by Memetic Pursuit procedure, but the computation cost can be reduced with a proper viewpoint on the behavior of the parameters.

5.5. Comparisons

5.5.1. Comparison with Previous Algorithms

The decomposition accuracy will be compared with the previous developed algorithm of stochastic matching pursuit and the Exhaustive Matching Pursuit.

First, the decomposition sequence will be verified, the decomposition of ICA component #20 is used to examine the correctness of the result. The five leading decomposed atom parameters in Table 24 found by HyMN-G seems to be very close to those found by Durka's stochastic matching pursuit and the exhaustive one.

#	Time Shift	Freq. Shift	Scale	Phase (π)	Wgh. (%)	Residue (%)
1	0.51	1.70	0.72	1.36	30.04	69.96
2	0.32	6.04	0.18	1.37	15.50	54.47
3	0.48	11.19	0.09	0.97	10.60	43.87
4	1.45	5.96	0.52	0.73	6.76	37.11
5	1.70	12.72	1.13	1.32	5.90	31.21
6	0.16	25.89	1.57	0.07	4.03	28.19
7	0.69	5.10	0.45	1.34	3.86	24.32
8	0.30	14.68	0.18	1.65	3.69	20.64
9	1.03	2.60	0.95	1.79	2.49	18.15
10	0.12	15.05	0.18	1.02	1.62	16.52
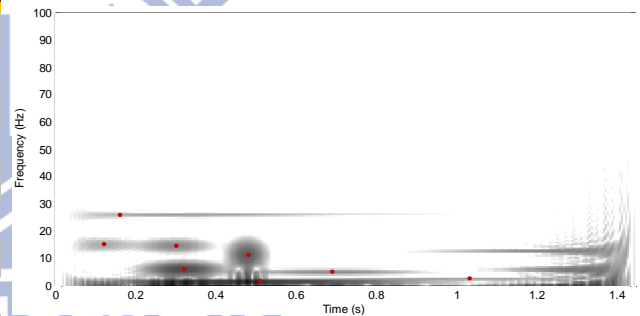
Table 24   Gabor Atom List Generated by HyMN-G (ICA #20)

Figure 34   Gabor Energy Map by HyMN-G (ICA #20)

#	Time Shift	Freq. Shift	Scale	Phase (π)	Wgh. (%)	Residue (%)
1	0.52	1.70	0.71	1.37	30.04	69.96
2	0.32	6.24	0.15	1.35	15.70	54.26
3	0.50	10.69	0.08	1.09	11.95	42.31
4	1.45	6.21	0.60	0.79	7.07	35.24
5	1.32	12.74	0.71	1.70	4.79	30.45
6	0.32	14.63	0.15	0.06	4.56	25.89
7	0.16	25.89	1.51	0.07	4.02	21.25
8	0.83	4.00	0.32	0.71	3.51	18.74
9	1.25	2.04	0.32	0.89	1.99	16.75
10	0.10	15.17	0.19	0.54	1.35	14.40

#	Time Shift	Freq. Shift	Scale	Phase (π)	Wgh. (%)	Residue (%)
1	0.62	2.07	0.53	1.70	26.22	73.78
2	0.27	2.76	0.05	0.57	18.85	54.93
3	0.50	11.03	0.08	1.14	10.75	44.18
4	1.33	4.83	0.16	1.32	6.23	37.95
5	1.27	13.10	0.43	0.41	5.12	32.83
6	0.62	4.83	0.49	0.64	4.50	28.33
7	0.34	26.21	0.5	1.65	3.64	24.69
8	0.07	0.69	0.23	1.73	3.34	21.35
9	0.39	11.72	0.10	0.30	2.97	18.38
10	1.10	5.52	0.19	0.47	1.43	16.95

Table 20 Gabor Atom List Generated by Exhaustive MP (ICA #20)

Table 21 Gabor Atom List Generated by Durka's Stochastic MP (ICA #20)

The percentages of energy extracted by the algorithms are plotted as in Figure 35:
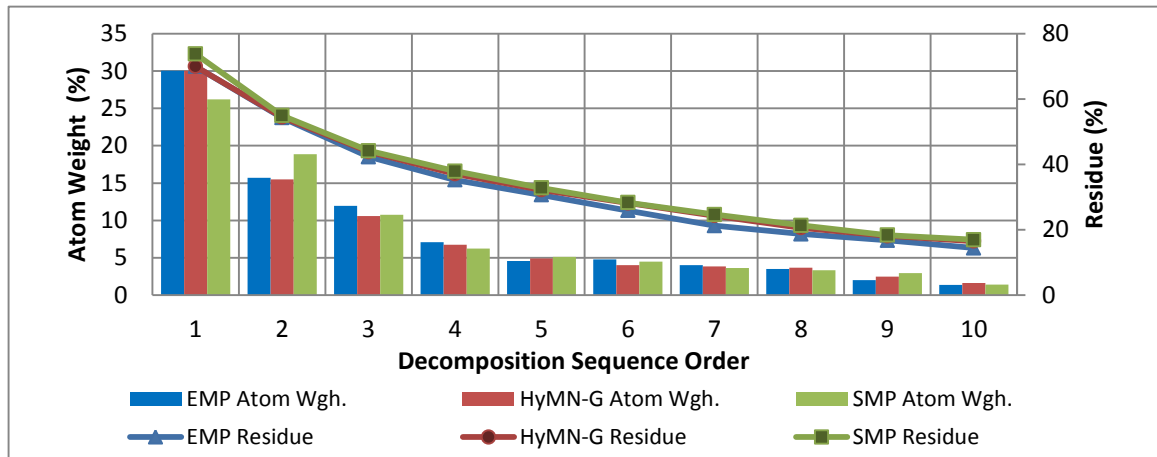
Figure 35   Convergences of Algorithms (Atom Weights & Residues Energies)

The result of exhaustive MP still generally being more precise than the others, but it comes with the cost of a time consuming process. On the other hand, the HyMN-G achieves almost similar goal with only several thousands of objective function evaluations.

The following table shows the computational cost of attaining one Gabor atom parameters (the computational cost of Natural Gradients has not been listed here since there is no comparable procedure in the other two algorithms):

Algorithms	Population	Objective Evaluations	Others Evaluations
Stochastic MP	Initial Population $= 100{,}000$	$100{,}000 + \underbrace{256}_{\text{Scan}} = 100{,}256$	None
Exhaustive MP	$\underbrace{290}_{\text{Scan } t} \cdot \underbrace{145}_{\text{Scan } \omega} \cdot \underbrace{290}_{\text{Scan } s}$ $= 12{,}194{,}500$	$12{,}194{,}500 +$ CMAES Pop. $\times$ Iteration $= 12{,}195{,}700$	Covariance Evaluation Covariance Modification
HyMN-G (Lamarckian)	Duration $\times$ Bandwidth $\times \sqrt{\rho} = 1.45 \cdot 200 \cdot \sqrt{1.2}$ $\cong 317$	$317 + \underbrace{\text{Candidates}}_{10\% \text{ of Pop.}}$ $\times \underbrace{\text{Iteration}}_{\text{Expected}}$ $\cong 5{,}500 \sim 15{,}000$	Fisher Information Matrix Gradient Estimation

Table 20 Comparable Computational Cost of Three Algorithms

5.5.2. Comparison with Cosine Dictionary

The following shows the decomposition result of component #20 generated using the HyMN-G algorithm and the discrete cosine transform.
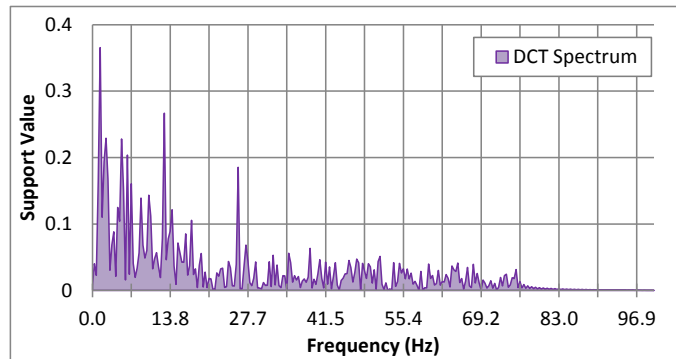
Figure 36   Discrete Cosine Spectrum
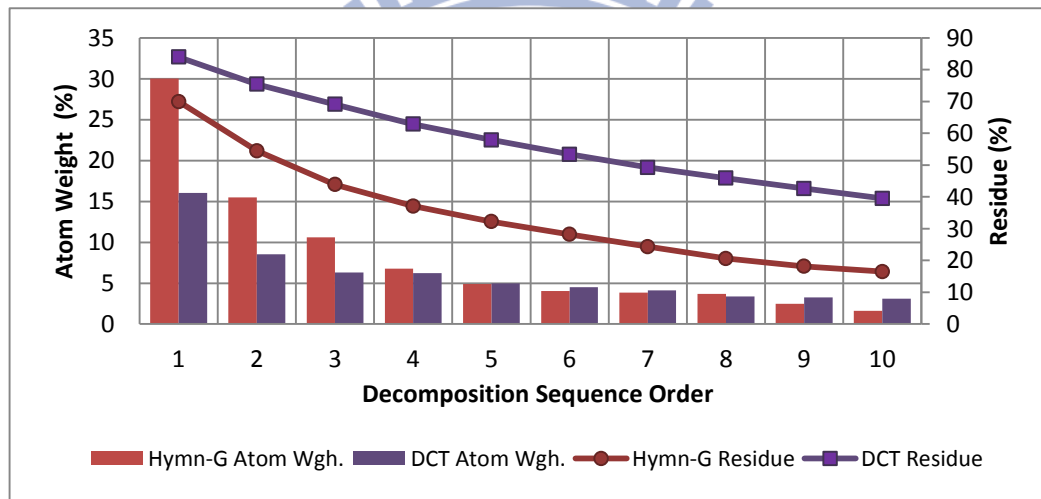
Figure 37   Decomposition of HyMN-G & Discrete Cosine Dictionary

Total 31 discrete cosine dictionary atoms should be used to extract the signal up to 85% of its energy, comparing to 11 when using HyMN-G algorithm. The scalable Gabor atoms have the advantages in number of atoms.

5.6.   Large Scale Analysis on Channel and ICA Signals

After verifying the correctness and convergence properties of HyMN-G, we'll use the algorithm to decompose a relatively large dataset instead of averaged ERP signal.

The dataset used in this experiment is the same one used in [43], which is an odd-ball

experiment recorded with 31 sensors in 512 Hz. During the recording, the subjects are asked to staring at a screen showing 4 blue boxes and 1 green box. After a certain period, a black circle will appear randomly in one of the boxes and disappear in 117 milliseconds; if the circle is right in the green box, the subject should press a button with his right thumb. The process is repeated with inter-stimulus interval ranging from 225 to 1000 milliseconds, total 524 epochs are extracted in the dataset.

In our experiment, the data is firstly filtered by a 0-to-40-Hz low pass filter, downsampled to 128 Hz and each contains 179 samples (duration = 1.4 seconds). After removing the artefacts, 262 are remained in the dataset and a channel is rejected due to the significant presence of muscle movement, and the ICA is performed afterwards. The size of the channel signal is $(\#\text{Channel} \times \#\text{Epoch} \times \text{Epoch Length}) = 31 \times 262 \times 179$, and the ICA component set is $(\#\text{Comp} \times \#\text{Epoch} \times \text{Epoch Length}) = 30 \times 262 \times 179$.

Finally, for all channels of raw EEG data and all ICA components, 10 epochs are selected and fed into the HyMN-G algorithm epoch by epoch as input; total 1393 atoms are generated for the channel data and 1864 for the ICA components.

Figure 38 and Figure 39 show the parameter centers in both scatter plots and bubble diagrams; with 3D perspective and the projection view on the normal plane of each axis. The radius of bubbles in the bubble charts are proportional to the weight of the atom and proportional to the square of the inner product value according to the conservation of energy described in (4.
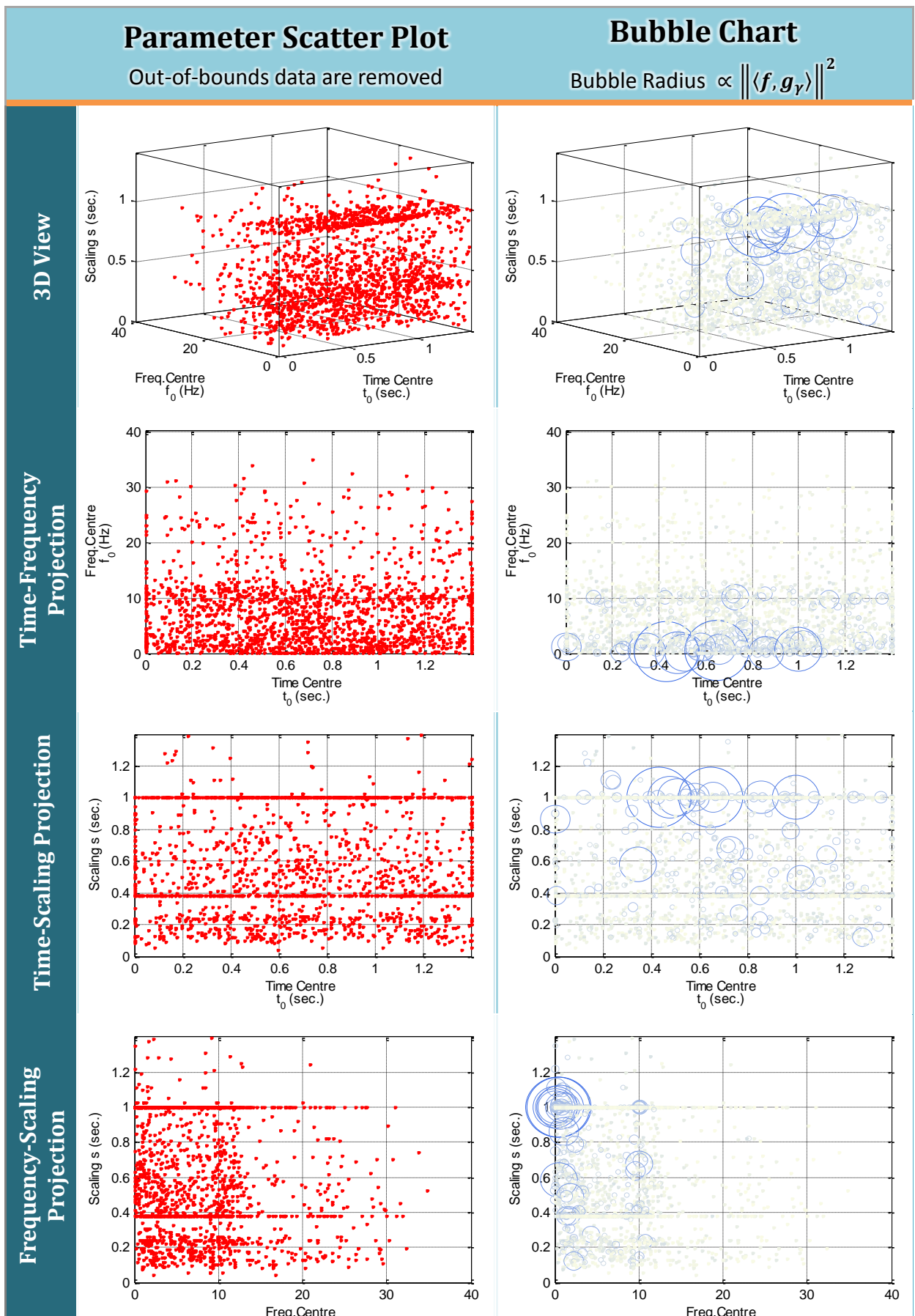
Figure 38   Large Scale Decomposition of 30 ICA Components (10 Epochs)
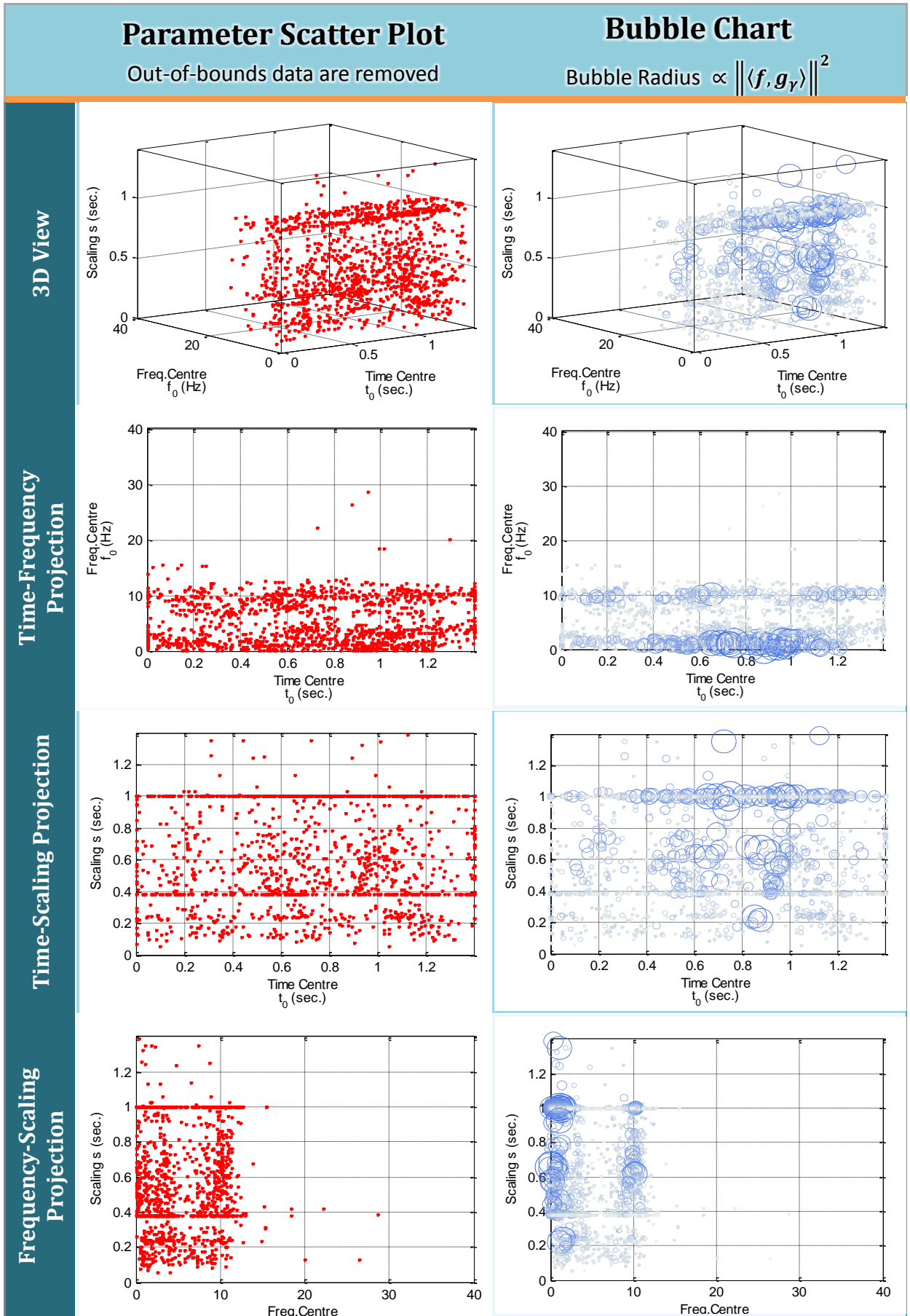
Figure 39   **Large Scale Decomposition of 31 Channels (10 Epochs)**

The atom centers derived from the ICA and channel decompositions seem to be concentrated on the same regions in the parameter space, and prove the consistency between the ICA decomposition result and the channel.

The most unexpected problem encountered during the experiment is the significant shifts in signal DC voltage; Figure 40 gives an example of such type of signals.
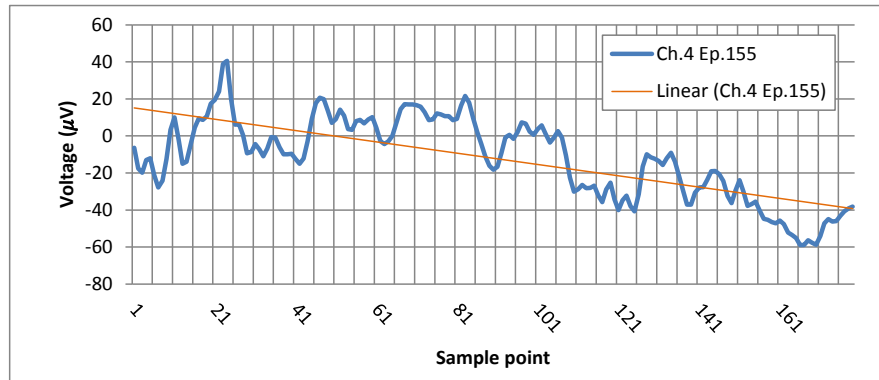
Figure 40 Channel #4, Epoch #155 and Linear Trendline

Since we've restricted the position of the time centers inside the signal duration, however in this case, the pursuit is not able to find a suitable set of parameters to fit the changes in DC voltage potentials since the Gaussian center will be out-of-bounds, even if the Gaussian center is at the boundary of the signal sample space, the standard deviation of the Gaussian window will become too large and eventually run out-of-bound, and causing a lot of atom centers fall on the border conditions.

Chapter 6. Dictionary Evolved Based on Training Signal Examples

6.1.   Training Method

In this approach, the result dictionary will not have a closed form expression. At the beginning, a set of training signals/models is provided and an initial dictionary are given as the initialization setting of the K-SVD algorithm [41], which is a generalized K-means clustering algorithm that iteratively alternates between sparse coding of the training set based on the current dictionary and an update of the dictionary atoms that makes the dictionary fit the training signals better; the update of the dictionary columns is combined with an update of the sparse representations while the sparsity level is configured as a parameter before executing the algorithm.

Several standard dictionaries and random Gaussian matrices are provided as the initial dictionaries of the learning procedures; the result will be compared in terms of convergence rate, final decomposition error and the performance under different sparse levels.

6.2.   Decomposition Problems

For the scheme of training a dictionary for signals of certain type; the training signal set, unlike the single-trial model, which serves as a template of the specific type of signals is given as a matrix, labeled as B. The problem now becomes 2 dimensional: (Equation (39 and Figure 41)

$$\forall 1 \leq i \leq N, \text{ find } x_i \text{ such that } \begin{cases} b_i = Ax_i \\ \quad \text{or} \\ \|Ax_i - b_i\|_2 < \epsilon \end{cases} \tag{39}$$

Figure 41   Illustration of Noisy Signal Factorization Problem

6.3.   Consistency Conditions

As mentioned in 3.6, similar to the single-trial model, Donoho and Elad [40] also proposed the condition for the approach of training dictionary.

The three following conditions identified by Donoho and Elad in [38-40] are considered when designing the size of the signal template set and the initial dictionary for training, notice that these following statements are sufficient conditions, so one can successfully compress and reconstruct the signal even if these conditions are not met:

1. **Boundary Support of the Coefficients** X

   The $\mathbf{L}_0$ norm of the coefficient vector should be smaller than half of the spark of the dictionary A, that is:

   $$\forall 1 \leq i \leq N \text{ and } b_i = Ax_i, \qquad \|x_i\|_0 = L < \frac{\sigma\{A\}}{2} \tag{40}$$

2. **Richness Relation between Dictionary** A **and Training Signal Set** B

   The signal template set B should be rich enough, which includes at least $L + 1$ signals for every possible combination of L atoms from A, so here, conservatively, assumed that B contains at least $(L + 1)\binom{k}{L}$ signal trials.

3. **Non-degeneracy of** B **when Representing in** A

64

This condition implies that the dictionary is utilized efficiently. Describe mathematically, for any group of $L + 1$ signals $B' \subset B$:

- If B' shares the same L atoms in A, then $\text{Rank}(B') = L$
- If B' shares different atoms in A, then $\text{Rank}(B') = L + 1$

Under the three conditions, the factorization relation $B = A\Pi X$ has the unique properties where:

- $A \in \mathbb{R}^{n \times k}, A = [a_1 \ a_2 \ a_3 \ \dots a_k]$ and $\forall i \leq 1 \leq k, \|a_i\|_2 = 1$
- $X \in \mathbb{R}^{k \times N}$, each column contains L non-zero terms
- $\Pi \in \{-1,0,1\}^{k \times k}$, which is a signed permutation matrix

For analysis convenience, the atoms in the dictionary are frequently sorted according to their dominant frequencies in our result visualization.

6.4. Experiment

The experiment use the previously described 64-Channel dataset recorded in the radio silent room, BRC, but the sampling rate has been lowered to 100 Hz and therefore the number of samples in each epoch is only 145 now, the primary goal is to successfully use the K-SVD algorithm to detect common "signal templates" or "fingerprints" in the ERP signal of Channel #17.

### 6.4.1. K-SVD Initialization and Sparsity Configurations

After the Exhaustive Matching Pursuit experiment, there's a discovery about that most of the ERP (both raw channel EEG and ICA components) can be deduced to contain only 10%

of energy (or $20 \log_{10} 0.1 = -20 \, dB$) with about 10 to 15 Gabor atoms. Therefore, the sparsity setting, which is the restriction of the number of non-zero terms in the coefficient matrix, of the K-SVD algorithm is 15, which is also close to the square root of the number of samples in the signal referred in 3.6.1, since $\sqrt{145} \cong 12$.

The redundancy ρ of the Grassmannian frame is set to 1.2, which generates the dictionary of size 145×445 and has a theoretical mutual coherence of 0.6181.

Finally, the training signal is generated by averaging of 100 over total 266 specimens with round robin selection method due to the shortage of epochs in the dataset, and it usually require 60 to 100 epoch to generate a valid ERP signal. The size of the generated training signal matrix is 145×16492.

The result is plotted in both the time domain representation and spectrum and finally, ordered according to the significant frequency value, shown in Figure 1Figure 42:
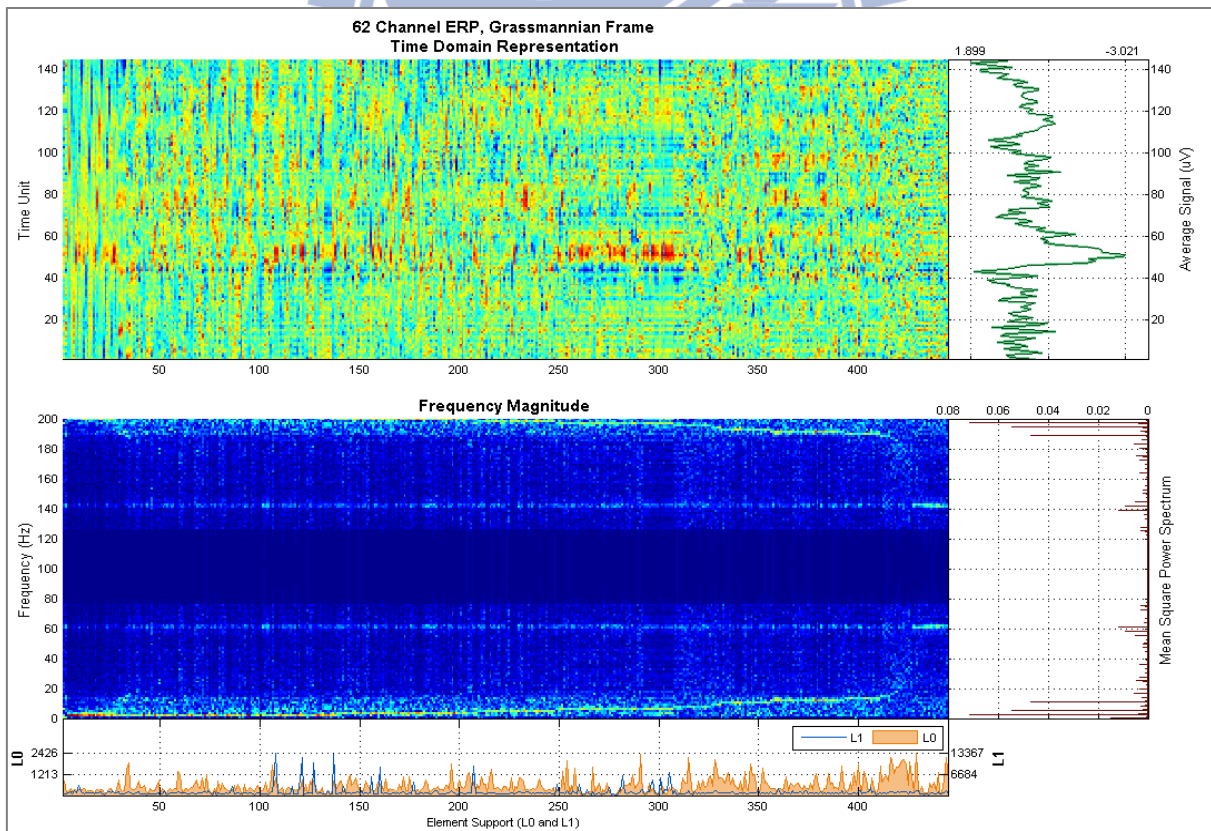
Figure 42   K-SVD Decomposition of Channel 17 ERP

In the bottom of the Figure 42, there's the summation of the L_0 and L_1 support of

each template waveforms in the dictionary, one can observed that the decomposed coefficient in terms of L_1 support is concentrated in only 14 signal templates. Furthermore, the 4 templates are found to be the top 4 selected atoms that are used to start a Matching Pursuit for the channel #17 ERP signal with the entire dictionary.
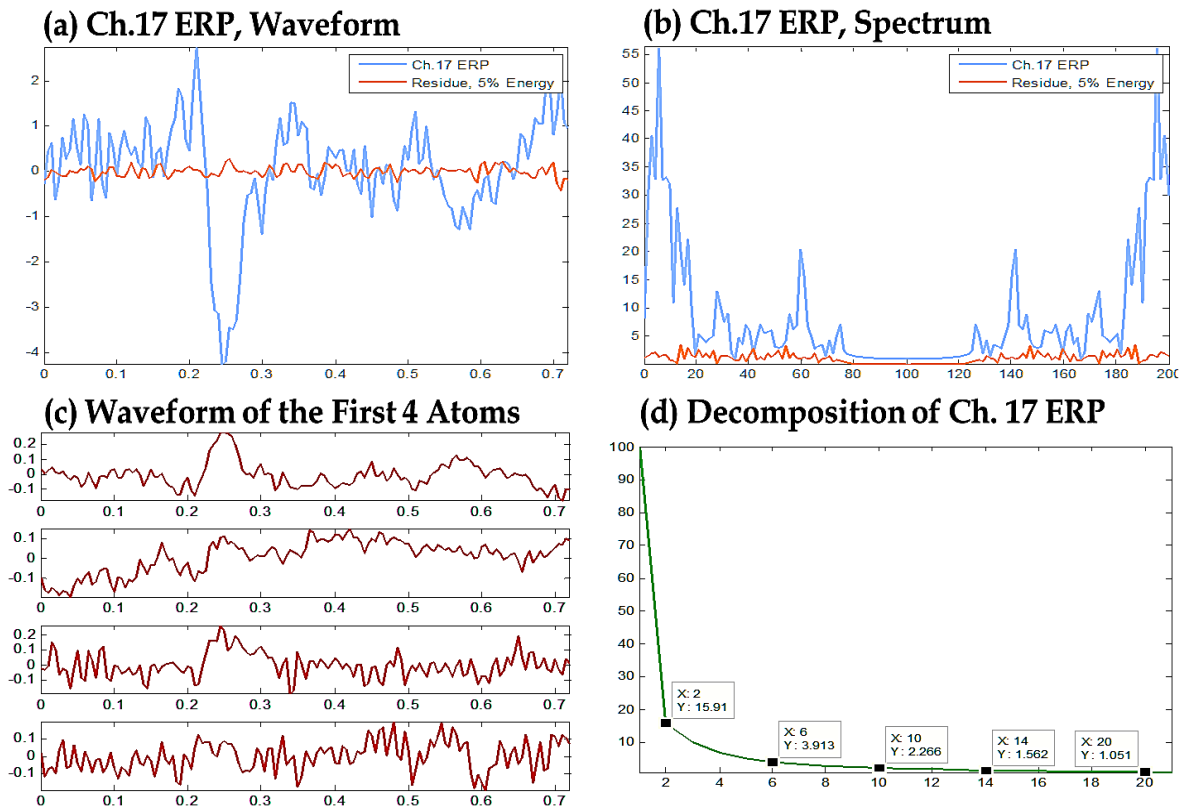
Figure 43   Decomposition of Channel 17 ERP Using the Trained Dictionary

Figure 43 shows the result of executing Matching Pursuit over channel #17 ERP signal using the trained dictionary, the four subfigures are detailed as follow:

a) Represents the signal and the residue carrying only 5% of the signal energy in temporal domain.

b) Shows the spectrum of the same waveforms in a).

c) This is the first four atoms that are used during the pursuit process, are identified to be the same four atoms with the highest sum over L_1 support.

d) The red curve shows the decay of the residue energy in top 20 iterations.

The residue also has the kurtosis of 3.053 and skewness of 0.0037.

The 1st, 3rd and the 4th atoms in the part (d) of Figure 43 are identified as the "fingerprint" of the ERP oscillating in different harmonic frequencies, with these four atoms, about 95% of the signal energy can be extracted, and therefore, the trained dictionary can be concluded as Ron Rubinstein said in [9], perform well as the "realization" of the signal of the event related potential of channel #17 EEG signals.

Chapter 7. Conclusion

7.1. Achievements

In this research, an efficient method has been presented and verified being able to search for the atom parameters having the highest correlation value with the signal in the continuous scalable Gabor parameter space. Also, the scalable Gabor parameter space is well perceived by now.

The major achievements are listed below, for the standard dictionary approach:

1. **Identify the generalized Gabor dictionary as the best suited standard dictionary for sparse ERP/ICA representation**

   For various types of dictionaries composed of standard waveforms, observation shows that the number of atoms used for representing the signals over the Gabor dictionary is the smallest among all types of waveforms that have been tested during the experiments.

2. **Formulate pruning rules of illogical parameter value for consistent sparse representation over the generalized Gabor parameter space**

   Since most of the recent decomposition algorithms focused mainly on sparsity efficiency and reconstruction accuracy; on the other hand, physical meaning of the atoms and the consistency of the decomposition are usually ignored, which should be carefully considered during biomedical signal analysis. To avoid ambiguous and illogical representations, several rules are brought up to regulate the decomposition procedures and eliminate the unreasonable results.
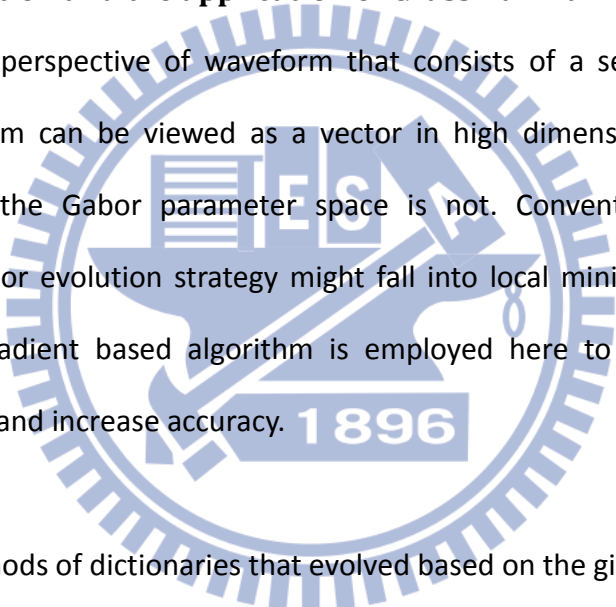
3. **Discover the behavior of non-linear parameters of Gabor atoms, and**

4. **Devise a construction method of Grassmannian lattice for Gabor atoms**

    With the concept of searching for the dictionary with low mutual coherence, the discovery of the behavior will be a key component to construct a redundant but lowly correlated frame.

5. **Design and implement a Memetic search algorithms for consistent sparse representation utilizing the Natural Gradient based optimization and the application of Grassmannian frame**

    From the perspective of waveform that consists of a sequence of samples, a Gabor atom can be viewed as a vector in high dimensional Euclidean space; however, the Gabor parameter space is not. Conventional gradient search algorithm or evolution strategy might fall into local minimum in this case. The natural gradient based algorithm is employed here to avoid pitfalls of local minimum and increase accuracy.

And, for the methods of dictionaries that evolved based on the given signal examples:

6. **Improve the K-SVD method by using Grassmannian frame as initial dictionary, and successfully locate the fingerprints of the ERPs in the decomposition result**

    The concept of the Grassmannian Gabor frame is applied on the initialization of the dictionary in this case, i.e., the development of the dictionary starts from a state that closes to a uniform sampling on the Gabor parameter space.

7.2.   Future Works

Nonetheless, this preliminary effort merely marks the beginning of our continuous search for better sparse representations of EEG signals and suitable dictionaries for these representations.

The primary concern is about the biomedical meaning of the decomposition result, most of the pursuit algorithms are not designed to be applied on a specific type of signals. Due to the sophisticated propagation of EEG signal from their actual brain location to the scalp, through skin, hair and finally to the sensors, the meaningful ingredient might become insignificant and difficult to extract. Always searching for the best fit waveform and justifying the precision of the decomposition of the signals in terms of energy is not always beneficial.

A temporarily possible solution might be the cross reference between the channel and ICA components, with the application of ICA components, one might be able to eliminate the undesired signal components and extract the *de facto* EEG events well represent the brain activities.

Establishment of the relationship between the result and Compressive Sensing might become a logical next step; during the analyses process of EEG and especially the ERP data, with both the standard dictionary approach and the K-SVD experiments, atoms are found to be clustered into several region and share similar parameter values, further investigation might be carried out to classify the effect, if an jittering and scaling resistive decomposition method can be found, a large number of the single-trial decomposition results can be compressed in advance to save even more data storage resources.

For the dictionary trained by the K-SVD algorithm, an intuitive thought is to discover the proper measurement matrix for the compressive sensing framework described in [12], but still, since the experiment requires an astronomical amount of data to form a proper training signal matrix.

Mathematical Notations

Spaces

$\mathbb{R}^N$	Real signal of size N
$\mathbb{C}^N$	Complex signal of size N
$\mathbf{L}^2(\mathbb{R})$	Finite energy continuous functions: $\int |f(t)|^2 dt < \infty$
$\mathbf{l}^2(\mathbb{Z})$	Finite energy discrete functions: $\sum_{n=-\infty}^{\infty} |f[n]|^2 < \infty$
$\mathbf{H}$	Hilbert Space

Operators and Operations

z^*	Complex conjugate of $z \in \mathbb{C}$
$\|f\|$	$\mathbf{L}^2$ Norm
$|f|_p$	$\mathbf{L}^p$ Norm
$\langle f, g \rangle$	Inner Product (Mathematically varied in different spaces)
A^T	Transpose of Matrix
$\inf\{S\}$	Infimum of set S
$\sigma\{A\}$	Spark of matrix A
T_{t_0}	Shift t_0 in time
M_{ω_0}	Modulate ω_0 in frequency

Transforms and Representations

$\hat{f}(\omega)$	Fourier transform
$\hat{f}[n]$	Discrete Fourier transform
$Sf[t, \omega]$	Short-time Fourier transform
$P_S f(t, \omega)$	Spectrogram
$P_V f(t, \omega)$	Wigner-Ville distribution

References

[1] T.P. Jung, S. Makeig, M.J. McKeown, A.J. Bell, T.W. Lee, and T.J. Sejnowski. "Imaging Brain Dynamics Using Independent Component Analysis", *Proc. IEEE*, 89(7):1107-22, 2001.

[2] S. Makeig, M. Westerfield, T.P. Jung, S. Enghoff, J. Townsend, E. Courchesne, T.J. Sejnowski. "Dynamic Brain Sources of Visual Evoked Responses". *Science*, 295:690-4, Jan. 25, 2002

[3] S. Makeig. "Auditory Event-Related Dynamics of EEG Spectrum and Effects of Exposure to Tones". *Electroencephalography and Clinical Neurophysiology*, 86:283-293, 1993.

[4] S. Makeig, S. Debener, J. Onton, A. Delorme. "Mining Event-Related Brain Dynamics". *Trends in Cognitive Science (TICS)*, Jan. 2004.

[5] G. Davis, S. Mallat, and M. Avellaneda. "Adaptive Greedy Approximations". *J. Constr; Approx.*, 13:57-98, 1997.

[6] S. Mallat, Z. Zhang, "Matching Pursuits with Time-Frequency Dictionaries". *IEEE Trans. Signal Process*. 41:3397–3415, Dec. 1993.

[7] Mingui Sun, Baumann, S.B., Xiaopu Yan, Shie Qian, Sonmez, M., Sclabassi, R.J. "Localization of dipole sources with time-frequency pre-processing of the EEG", *Engineering in Medicine and Biology Society, 1995., IEEE 17th Annual Conference*, 10.1109/IEMBS.1995.579502

[8] David Papo, Abdel Douiri, Florence Bouchet, Jean-Claude Bourzeix "Time--Frequency Intracranial SourceLocalization of Feedback-Related EEGActivity in Hypothesis Testing", *Oxford Journals: Life Sciences & Medicine, Cerebral Cortex, Volume17, Issue6, Pp. 1314-1322*. August 2, 2006

[9] Ron Rubinstein, Alfred M. Bruckstein. "Dictionaries for Sparse Representation Modeling". *Proceedings of the IEEE.,* (98), 0918-9219, June 2010

[10] Emmanuel J. Candès, "Compressive sensing", *Proc. International Congress of Mathematicians*, Madrid, Spain, 2006.

[11] Dror Baron, Michael B. Wakin, Marco F. Duarte, Shriram Sarvotham and Richard G. Baraniuk, "Distributed Compressed Sensing", *IEEE Trans. Information Theory*, November 2005.

[12] S. Aviyente, "Compressed Sensing Framework for EEG Compression," *Proc. IEEE Statistical Signal Processing (SSP)*, Madison, August 2007.

[13] Amir M. Abdulghani, Alexander J. Casson and Esther Rodriguez-Villegas, "Quantifying the Performance of Compressive Sensing on EEG Signals", submitted *to IEEE Trans. Biomedical Engineering*, July 2009.

[14] P.J. Durka, D. Ircha, and K.J. Blinowska, "Stochastic Time-Frequency Dictionaries for Matching Pursuit," *IEEE Trans. Signal Process*., (49)3:507–510, Mar. 2001

[15] Mordecai Avriel (2003). *Nonlinear Programming: Analysis and Methods*. Dover Publishing. ISBN 0-486-43227-0.

[16] Jan A. Snyman (2005). *Practical Mathematical Optimization: An Introduction to Basic Optimization Theory and Classical and New Gradient-Based Algorithms*. Springer Publishing. ISBN 0-387-24348-8

[17] S. Amari, S.C. Douglas, "Why Natural Gradient?", *Proc. IEEE,* 1998.

[18] S. Amari. "Natural Gradient Works Efficiently in Learning." *Neural Computation*, 10(2):251–276, 1998

[19] S. Yi, D. Wierstra, T. Schaul, J. Schmidhuber, "Stochastic Search using the Natural Gradient", *Proc. 26th Int'l Conf. Machine Learning*, Montreal, Canada, 2009.

[20] S. Yi, D. Wierstra, T. Schaul, J. Schmidhuber, "Efficient Natural Evolution Algorithm", *Proc. GECCO*, July 8-12, 2009.

[21] Y Sun, D Wierstra, T Schaul . "Efficient natural evolution strategies". *Proceedings of the 11th Annual conference on Genetic and evolutionary computation, 2009*, doi:10.1145/1569901.1569976.

[22] Thomas Strohmer, Robert W. Heath, Jr., "Grassmannian frames with applications to coding and communication "*Applied and Computational Harmonic Analysis, Vol. 14, Issue 3, P. 257-275,* May 2003

[23] A. R. Calderbank, R. H. Hardin, E. M. Rains, P. W. Shor, and N. J. A. Sloane. "A group-theoretic framework for the construction of packings in Grassmannian spaces". J. Algebraic Combin., 9(2):129–140, 1999.

[24] Thomas Strohmer, Scott Beaver, "Optimal OFDM design for time-frequency dispersive channels", *IEEE transactions on communications ISSN 0090-6778,* vol. 51, no7, pp. 1111-1122, 2003.

[25] Dawkins R."The Selfish Gene". *Oxford University Press* 1989.

[26] Moscato, P. "On Evolution, Search, Optimization, Genetic Algorithms and Martial Arts: Towards Memetic Algorithms". *Caltech Concurrent Computation Program* (report 826), 1989.

[27] S. Mann and S. Haykin, "The Chirplet transform: A generalization of Gabor's logon transform*", Proc. Vision Interface 1991*, 205–212 (3–7 June 1991).

[28] Brown, M.L., Williams, W.J., Hero, A.O., III. "Non-orthogonal Gabor representation of biological signals", *Acoustics, Speech, and Signal Processing, 1994*. 10.1109/ICASSP.1994.389742. 1994

[29] C. Tallon-Baudry, O. Bertrand, C. Delpuech, and J. Pernier. "Oscillatory γ-Band (30-70 Hz) Activity Induced by a Visual Search Task in Humans". *J. Neuroscience*, 17(2):772–734, Jan. 1997.

[30] M . Latka, Z. Was, A. Kozik, and B. J. West. "Wavelet Analysis of Epileptic Spikes". *J. Phys. Rev.* E67(5) #052902, Jan. 2003.

[31] C. Tallon-Baudry, O. Bertrand, C. Delpuech, and J. Pernier. "Oscillatory γ-Band (30-70 Hz) Activity Induced by a Visual Search Task in Humans". *J. Neuroscience*, 17(2):772–734, Jan. 1997.

[32] M . Latka, Z. Was, A. Kozik, and B. J. West. "Wavelet Analysis of Epileptic Spikes". *J. Phys. Rev.* E67(5) #052902, Jan. 2003.

[33] M. Andrle, L. Rebollo-Neira. "Cardinal B-spline dictionaries on compact interval" *Applied Comput. Harm. Analysis*, 18(3):336-346, 2005.

[34] L. Rebollo-Neiraa, Z. Xu. "Sparse signal representation by adaptive non-uniform B-spline dictionaries on a compact interval*". EURASIP* 90(7):2308-2313, Jul. 2010.

[35] N. Hansen, A. Ostermeier. "Completely Derandomized Self Adapta-tion in Evolution Strategies", *Evolutionary Computation*, 9(2):159–195, 2001.

[36] N. Hansen, S.D. Müller, P. Koumoutsakos. "Reducing Time Complexity of De-randomized Evolution Strategy with Covariance Matrix Adaptation (CMA-ES)". *Evolutionary Computation*, 11(1):1–18, 2003.

[37] J.W. Qiu, John Kar-kin Zao, Peng-hua Wang, Alvin Choi, "Consistent Sparse Representations of EEG ERP and ICA Components Based on Chirplet and Wavelet Dictionaries", 32nd Annual International IEEE EMBS Conference, Sep, 2010.

[38] D.L. Donoho, M. Elad, "Optimally Sparse Representation in General (Non-orthogonal) Dictionaries via l1 Minimization", *PNAS* 100(5): 2197–2202, 2003.

[39] D.L. Donoho, M. Elad, V. Temlyakov. "Stable Recovery of Sparse Over-complete Representations in the Presence of Noise". *IEEE Trans. Information Theory*, Feb. 2004.

[40] M. Aharon, M. Elad, A.M. Bruckstein. "On the uniqueness of overcomplete dictionaries and a practical way to retrieve them". *Linear Alg. And Appl.*, 416:48-67, 2006.

[41] M. Aharon, M. Elad, and A.M. Bruckstein, "K-SVD: An Algorithm for Designing of Overcomplete Dictionaries for Sparse Representation". *IEEE Trans. Signal Processing*, 54(11):4311-4322, Nov. 2006.

[42] Albert Solernou and Juan Fernandez-Recio, "Protein docking by Rotation-Based Uniform Sampling (RotBUS) with fast computing of intermolecular contact distance and residue desolvation", *BMC Bioinformatics*, 11:352doi:10.1186/1471-2105-11-352, Jun. 2010.

[43] Scott Makeig, Marissa Westerfield, Tzyy-Ping Jung, James Covington, Jeanne Townsend, Terrence J. Sejnowski, and Eric Courchesne. "Functionally Independent Components of the Late Positive Event-Related Potential during Visual Spatial Attention", *The Journal of Neuroscience*, April 1, 1999, 19(7):2665-2680