Technical Research Report

Quotient Signal Decomposition and Order Estimation

by D. Napoletani, C.A. Berenstein and P.S. Krishnaprasad

TR 2002-47

Quotient Signal Decomposition and Order Estimation

D. Napoletani, C.A. Berenstein and P. S. Krishnaprasad.

Institute for Systems Research
University of Maryland
College Park, MD 20742
(dnapolet, carlos, krishna @glue.umd.edu)

current address for the first author:
School of Computational Sciences
George Mason University
Fairfax, VA 22030
(dnapolet@gmu.edu)

Abstract

In this paper we propose a method for blind signal decomposition that does not require the independence or stationarity of the sources. This method, that we consider a simple instance of non-linear projection pursuit, is based on the possibility of recovering the areas in the time-frequency where the original signals are isolated or almost isolated with the use of suitable quotients of linear combinations of the spectrograms of the mixtures.

We then threshold such quotients according to the value of their imaginary part to prove that the method is theoretically sound under mild assumptions

on the mixing matrix and the sources. We study one basic algorithm based on this method.

The algorithm has the important feature of estimating the number of sources with two measurements, it then requires $n-2$ additional measurements to provide a reconstruction of n sources. Experimental results show that the method works even when several shifted version of the same source are mixed.

1 Introduction

Independent Component Analysis can recover signals that are linearly mixed with an unknown mixing matrix. All algorithms are essentially based on some local learning rule (see [L] and references therein, but also [QKS]). This procedure is effective, but it suffers from the need to assume that sources are independent and stationary. A different approach is taken in [CC], where sources are assumed to be independent and non-stationary and only time-delayed correlations of the observations are used to recover the mixing matrix. None of the previous methods can handle the case of mixtures of sources and their echoes, since clearly a source and its shifted versions are not independent.

In this paper we suggest an algorithm that requires a different set of assumptions on the sources. This algorithm allows to estimate the number of sources given at least two mixtures and we show that, if there are additional observations so that the total number of mixtures is equal to the number of sources, a full reconstruction algorithm is possible.

More specifically let $x_1 = a_1 s_1 + ... + a_n s_n, \;\; x_2 = b_1 s_1 + ... + b_n s_n$ be the two mixtures of n real-valued discrete sources s_i, $i = 1, ..., n$ with $a_i, b_i \in \mathbb{R}$.

Compute the spectrograms of x_1 and x_2, say X_1 and X_2, where by spectrograms we mean the complex-valued matrices of windowed discrete Fourier transforms.

Clearly, if we denote the spectrograms of s_i by S_i, we have:

$$X_1 = a_1 S_1 + ... + a_n S_n, \;\; X_2 = b_1 S_1 + ... + b_n S_n.$$

Let R be a non-singular 2×2 real-valued matrix that we call *exploratory matrix*, and consider the quotient

$$Q_R(t,w) = \frac{R(1,1)X_1(t,w) + R(1,2)X_2(t,w)}{R(2,1)X_1(t,w) + R(2,2)X_2(t,w)}$$

where t is the time coordinate and w the frequency one.

We use $Q_R(t,w)$ to find regions in the time-frequency plane where sources are isolated or almost isolated. This in turn reduces the search for the unmixing matrix (up to left multiplication by a diagonal matrix) to the solution of an underdetermined system of linear equations.

The algorithm does not require the sources to be independent or stationary, but rather it relies on geometrical separation conditions on the spectrograms of the sources. In essence, two related data sets of dimension two (X_1 and X_2) are projected onto a one dimensional space through the non-linear function $Z = \frac{R(1,1)X_1+R(1,2)X_2}{R(2,1)X_1+R(2,2)X_2}$, therefore we can view the underlining method as a type of non-linear projection pursuit in which the choice of the exploratory matrix determines the specific non-linear projection of interest (see [H] for an extensive treatment of projection pursuit).

The second section of this paper introduces the basic idea and we introduce the definitions and tools needed for our "quotient projection" algorithm . An important point of this section is the understanding that the imaginary part of the quotient $Q_R(t,w)$, a simple measure of "phase locking" between the two measurements, can be used to increase the probability of finding areas where signals are isolated.

Section 3 presents the main steps of the algorithm and in it we discuss the limits of the method.

The fundamental role of separation of sources in time frequency domain to achieve reconstruction was already underlined by Rickard and collaborators in [RD], [RBR], [RY], as we became aware of in the final stages of our work. In this paper we stress the use of suitable thresholding of the imaginary part of $Q_R(t,w)$ as important in proving the theoretical soundness of the method. The possibility of choosing the most efficient exploratory matrix is also emphasised here in line with the idea of choosing the best non-linear projection.

2 Quotients Projections

We need to state several conditions to assure that $Q_R(t,w)$ is an effective tool in detecting sources. We start with a condition on the coefficients in the mixtures x_1 and x_2:

Condition (1): Assume that $\frac{b_i}{a_i} \neq \frac{b_j}{a_j}$ when $i \neq j$.We call $\frac{b_i}{a_i}$ the slope of the source s_i.

Denote by $\Im(f)$ and $\Re(f)$ the imaginary and real parts of a complex function f. Note that if $S_i(t_0,w_0) \neq 0$ for a single $i = i_1$ at a given (t_0,w_0) then

$$Q_R(t_0,w_0) = \frac{R(1,1)a_{i_1}S_{i_1}(t_0,w_0) + R(1,2)b_{i_1}S_{i_1}(t_0,w_0)}{R(2,1)a_{i_1}S_{i_1}(t_0,w_0) + R(2,2)b_{i_1}S_{i_1}(t_0,w_0)} = \frac{R(1,1)a_{i_1} + R(1,2)b_{i_1}}{R(2,1)a_{i_1} + R(2,2)b_{i_1}}$$

and therefore $\Im(Q_R(t_0,w_0)) = 0$.

Thus, we can approximately identify the regions of the time-frequency plane where the different sources are isolated, by retaining only the elements of the matrix $Q_R(t,w)$ that have imaginary part very near to zero. Clearly, this is a necessary condition to be verified for points (t,w) where the sources are isolated, but it is not sufficient.

Let $T(t,w) = \frac{\Im(Q_R(t,w))}{\Re(Q_R(t,w))}$, we enforce $\Im(Q_R(t_0,w_0)) \approx 0$ asking that $|T(t_0,w_0)| < \epsilon$. This is a computationally simple way to make sure that the relative magnitude of the imaginary part is taken in account rather than the absolute one, note that if $\Re(Q_R(t_0,w_0)) = 0$ then $T(t_0,w_0)$ is not defined, on the other hand we will see in the following discussion that a point for which $\Re(Q_R(t_0,w_0)) = 0$ and $\Im(Q_R(t_0,w_0)) \approx 0$ can be ignored. The case $\Re(Q_R(t_0,w_0)) >> 0$, $\Im(Q_R(t_0.w_0)) \neq 0$ may lead to $|T| < \epsilon$ too, but this case is unlikely. if at (t_0,w_0) there are several non-zero sources This claim will be made precise in lemma 2.1. Our choice of $T(t,w)$ is clearly not the only possibility, any other continuous function $T(t,w)$ such that $|T(t,w)| = 0$ implies $|\Im(Q_R(t,w))| = 0$ would be suitable.

Let us call $Q_{(R,\epsilon)}(t,w)$ the function obtained by thresholding $Q_R(t,w)$ in the following way:

$$Q_{(R,\epsilon)}(t,w) = Q_R(t,w) \ if \ |\frac{\Im(Q_R(t,w))}{\Re(Q_R(t,w))}| < \epsilon \ \ , Q_{(R,\epsilon)}(t,w) = 0 \ otherwise.$$

Ideally we claim that the distribution of the values of $\Re(Q_{(R,\epsilon)})$, is, as ϵ goes to zero, made of several delta functions of different weigth centered at 0 (due to the regions of the time frequency domain where there is no contribution from any signal) and at the value of the quotients $q_R(i) = \frac{R(1,1)a_i+R(1,2)b_i}{R(2,1)a_i+R(2,2)b_i}$, $i = 1,...,n$

that we can assume finite for a generic choice of R.

Therefore the number of "dominant" peaks (see remark 3.3) of the value distribution of the non zero values of $\Re(Q_{(R,\epsilon)})$ gives us generically an estimate of the number of sources and their positions will give the values of the $q_R(i)$'s and therefore also the values of the slopes $\frac{b_i}{a_i}$'s.

Definition 2.1 *We call silence a positive area region in time frequency domain where both X_1 and X_2 are identically zero.*

The restriction to non zero values of $\Re(Q_{(R,\epsilon)})$ makes the case in which $q_R(i) = 0$ for some i degenerate, since the removal of the zero values would also remove the contribution of signal s_i to the value distribution. A generic perturbation of R avoids this problem, in section 3 we indicate how to reduce the chance of choosing such degenerate exploratory matrix.

We assume throughout this section that R is choosen so that $q_R(i)$ is finite and non-zero for all $i = 1, ..., n$.

The possibility to separate sources using our idea rests on the following assumption:

Condition (2): the sources s_i, $i = 1, ..., n$ are separated in some regions of our chosen time-frequency representation.

Remark 2.1: Any complex-valued frame that achieves this objective for the class of sources that we are interested in would be suitable for the quotient signal decomposition.

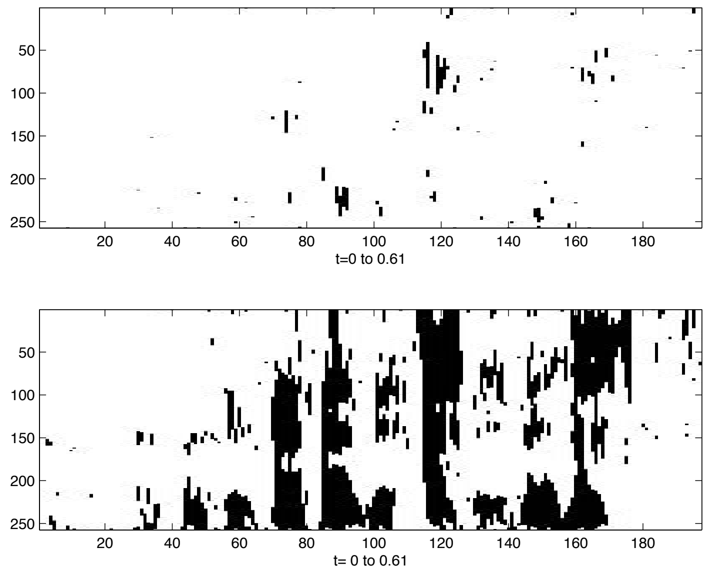

Figure 1: Dominant regions as defined in the text for $\epsilon = 10^{-3}$ (top) and for $\epsilon = 10^{-2}$ (bottom)

In this paper we perform our experimental work on linear mixtures of speech signals, using spectrograms as complex linear transformation, therefore it is interesting to verify to which extent is condition (2) true for this class of signals. As figure 1 shows, given two speech signals s_1 and s_2 on a time interval of 0.61 seconds, the region in time-frequency domain where we have $\frac{|S_i|}{|S_j|} < \epsilon$ $i \neq j$, $i, j = 1, 2$ is marginal for small ϵ ($\epsilon = 10^{-3}$), but it is sizable when we consider higher values ($\epsilon = 10^{-2}$).

This suggests that we must develop the theory in the context of small, but not insignificant, perturbations that can arise where each source is dominant. Such perturbations can be caused by noise or by low energy contribution by the other sources, as in this case, therefore the analysis of the problem should not require a specific knowledge of the probability distribution of the perturbations therefore a non-parametric approach to the theoretical stability of the algorithm introduced in this paper is needed. For the time being we consider the ideal case in which signals are truly separated to build a simple version of our algorithm. To prove that this basic algorithm is theoretically sound we need to give some mild conditions on the probability distribution of the sources in the transformed domain to make sure that the contribution to the value distribution of $\Re(Q_{(R,\epsilon)})$ due to values of (t,w) where we do have mixtures of sources is minimal. Certainly the following condition has to be satisfied:

Condition (3): Sources must be linearly independent on positive measure regions of the spectrogram, i.e. given any positive area region $\mathcal{B}$, we cannot find real numbers $(p_1, ..., p_n)$ such that $p_1 S_1(t,w) + ... + p_n S_n(t,w) = 0$ for all $(t,w) \in \mathcal{B}$.

If condition (3) is not verified, we can have degenerate situations in which ghost sources are detected. Assume for example that $S_j = pS_i$, p constant for some i and j in a region $\mathcal{M}$, and that this dependence happens in some region where the contribution of other signals is marginal. Then we have that, on $\mathcal{M}$:

$$Q_R = \frac{R(1,1)(a_i S_i + a_j S_j) + R(1,2)(b_i S_i + b_j S_j)}{R(2,1)(a_i S_i + a_j S_j) + R(2,2)(b_i S_i + b_j S_j)} = \frac{R(1,1)(a_i + p a_j) + R(1,2)(b_i + p b_j)}{R(2,1)(a_i + p a_j) + R(2,2)(b_i + p b_j)}$$

The slope $\frac{b_i + p b_j}{a_i + p a_j}$ of a "source" that does not exist as physical entity would be detected and, the quotient projection algorithm would need $n+1$ measurements to give complete reconstruction of the n physical sources and the virtual one.

On the other hand it is reasonable that sources can be very similar in some cases (think of Gregorian chant).

Thus it is expected that there will always be limit cases that lead the algorithm into detecting ghost sources. Our own auditory system is not immune from illusions.

We can enforce (3) assuming that:

Condition (3') S_i, $i = 1, ..., n$ are spatially distributed realizations of random complex variables $\hat{S}_i$ with supports $Supp(\hat{S}_i)$ not totally overlapping, i.e. there exist positive area regions $B_i \subset Supp(\hat{S}_i)$ such that B_i does not belong to any $Supp(\hat{S}_j)$ for $j \neq i$.

In other words we fix the geometrical supports and we assume that the values of S_i on each value of $(t, w) \in Supp(\hat{S}_i)$ is a realization of the corresponding random variable $\hat{S}_i$.

Let now $\mathcal{Z} = \bigcap Supp(\hat{S}_i)$, $\mathcal{D} = \bigcup Supp(\hat{S}_i)$, $\mathcal{I}_{S_i} = Supp(\hat{S}_i) \backslash [\bigcup (Supp(\hat{S}_i) \cap Supp(\hat{S}_j))]$, $j \neq i$ and $\mathcal{I} = \bigcup \mathcal{I}_{S_i}$.

Remark 2.2 For simplicity let $\mathcal{Z} \bigcup \mathcal{I} = \mathcal{D}$, the following discussion would work even if there are regions where only the support of some sources overlap, at least when the number of sources is finite. Moreover note that silence in the time frequency domain is not included in $\mathcal{D}$.

From condition (2) we know that $\mathcal{I}$ is of positive measure.

Denote $\hat{S}_i = \Re(\hat{S}_i) + i\Im(\hat{S}_i)$ and let $\hat{P}_1 = a_1\Re(\hat{S}_1) + ...a_n\Re(\hat{S}_n)$, $\hat{P}_2 = a_1\Im(\hat{S}_1) + ...a_n\Im(\hat{S}_n)$, $\hat{M}_1 = b_1\Re(\hat{S}_1) + ... + b_n\Re(\hat{S}_n)$, $\hat{M}_2 = b_1\Im(\hat{S}_1) + ... + b_n\Im(\hat{S}_n)$.

As random variables, let $\hat{F}_{\Re} = \Re(Q_R)$ and $\hat{F}_{\Im} = \frac{\Im(Q_R)}{\Re(Q_R)}$, where we drop the explicit dependence from (t, w) while assuming that (t, w) is a random point uniformly distributed in $\mathcal{D}$.

A simple computation shows that,

$$\hat{F}_{\Im} = \frac{\Im(Q_R)}{\Re(Q_R)} = \frac{(R(1,1)R(2,2) - R(1,2)R(2,1))(\hat{P}_1\hat{M}_2 - \hat{P}_2\hat{M}_1)}{T(\hat{P}_1, \hat{M}_1)B(\hat{P}_2, \hat{M}_2) + T(\hat{P}_2, \hat{M}_2)B(\hat{P}_1, \hat{M}_1)}$$

where $T(\hat{P}_i, \hat{M}_i) = R(1,1)\hat{P}_i + R(1,2)\hat{M}_i$ and $B(\hat{P}_i, \hat{M}_i) = R(2,1)\hat{P}_i + R(2,2)\hat{M}_i$.

To prove the following lemma and theorem we need one more technical condition:

Condition (4): the joint probability density function of $\hat{P}_i$ and $\hat{M}_i$, $i = 1, 2$ is a continuous function.

Let $f_{\mathcal{X}}$ be the underlining probability density function of the values of $\hat{F}_{\Im}|_{\mathcal{X}}$ on a region $\mathcal{X}$. Then:

Lemma 2.1 *If conditions (1), (2), (3') and (4) are satisfied, then the probability density function $f_{\mathcal{D},\epsilon}$ of $\hat{F}_{\Im}$ knowing that $|\hat{F}_{\Im}| < \epsilon$ converges to a delta function centered at the origin as ϵ goes to zero. More specifically $f_{\mathcal{D},0} = f_{\mathcal{I}}$.*

Proof: Condition (4), and the fact that

$$\mathcal{S}_q = \{(\hat{M}_1, \hat{M}_2, \hat{P}_1, \hat{P}_2) \mid \hat{F}_{\Im}(\hat{M}_1, \hat{M}_2, \hat{P}_1, \hat{P}_2) = q\}$$

is a set of measure zero in $\mathbb{R}^4$ for every $q \in \mathbb{R}$, tell us that $f_{\mathcal{Z}}$ is a non-atomic probability density function. As regards $f_{\mathcal{I}}$, we know it is centered at the origin, since signals are isolated on $\mathcal{I}$ and we expect $\hat{F}_{\Im} = 0$ for all values of $(t, w) \in \mathcal{I}$, therefore $f_{\mathcal{I}}$ is a delta function centered at the origin.

The probability density function that we observe, before imposing the thresholding, is $f_{\mathcal{D}} = \frac{A(\mathcal{Z})}{A(\mathcal{D})} f_{\mathcal{Z}} + \frac{A(\mathcal{I})}{A(\mathcal{D})} f_{\mathcal{I}}$ where $A(*)$ is the area function. After imposing that $|\hat{F}_{\Im}| < \epsilon$ we observe the new probability density function:

$$f_{\mathcal{D},\epsilon} = \frac{\sigma A(\mathcal{Z})}{\sigma A(\mathcal{Z}) + A(\mathcal{I})} f_{\mathcal{Z},\epsilon} + \frac{A(\mathcal{I})}{\sigma A(\mathcal{Z}) + A(\mathcal{I})} f_{\mathcal{I}}$$

where $\sigma = \int_{-\epsilon}^{\epsilon} f_{\mathcal{Z}}$ and $f_{\mathcal{Z},\epsilon}$ is the restriction of $f_{\mathcal{Z}}$ to the interval $[-\epsilon, \epsilon]$. Since σ converges to zero as ϵ goes to zero, we have that $f_{\mathcal{D},0} = f_{\mathcal{I}}$.

The previous lemma shows that most non zero values of $Q_{(R,\epsilon)}$ will likely be in the regions $\mathcal{I}_{S_i}$ for ϵ small enough. Therefore we have the following theorem. Denote by $\delta_{q_R(i)}$ the delta function centered at $q_R(i)$, then:

Theorem 2.1 *Under the same conditions as lemma 2.1, the probability density function $g_{\mathcal{D},0}$ of $\hat{F}_{\Re}$ knowing that $\hat{F}_{\Im} = 0$ is $\sum_i \frac{A(\mathcal{I}_i)}{A(\mathcal{I})} \delta_{q_R(i)}$.*

Proof: As a consequence of lemma 2.1, $\hat{F}_{\Im}(t_0, w_0) = 0$ implies with probability 1 that $(t_0, w_0) \in \mathcal{I}$, which in turn implies that $\hat{F}_{\Re}(t_0, w_0) \in \{q_R(1), ..., q_R(n)\}$. The probability that $\hat{F}_{\Re}(t_0, w_0)$ corresponds to any specific one of the $q_R(i)$'s depends from the area of each region $\mathcal{I}_i$. More specifically the probability density function of $\hat{F}_{\Re}|_{\mathcal{I}}$ is $g_{\mathcal{I}} = \sum_i \frac{A(\mathcal{I}_i)}{A(\mathcal{I})} \delta_{q_R(i)}$, therefore $g_{\mathcal{D},0} = g_{\mathcal{I}} = \sum_i \frac{A(\mathcal{I}_i)}{A(\mathcal{I})} \delta_{q_R(i)}$.

Note that the interpretation of the sources as spatially distributed random variables (that is conditions (3') and (4)), is not essential to our method, but it gives a possible theoretical basis to show why the independence of the sources is not needed.

Example 1: To see in practice the basic idea in the ideal setting we applied the algorithm to the mixture of five speech signals that were each set to zero on some small not overlapping time intervals of length 0.0061 seconds to make sure that conditions (2) and (3) were satisfied. This ideal setting is not unlikely in practice, as it will happen if only one of the speech sources is active at some given time interval. Let $s_1, ..., s_5$ *be the sources and set* $x_1 = s_1 + 10s_2 + 1.4s_3 + s_4 + 0.3s_5$, $x_1 = s_1 + 3s_2 + 1.4s_3 + s_4 + 0.3s_5$, $x_2 = s_1 + 3.03s_2 + 1.03s_3 - 4.99s_4 + 0.5s_5$,

the resulting mixtures are observed on a time interval of 1.22 seconds. Consider the choice of $R = \begin{bmatrix} 0.5, 1 \\ 0.8, 1 \end{bmatrix}$ *(For a preliminary analysis on how to choose* R *see next section).*

The true values of the $q_R(i)$*'s are: 0.8333, 0.8342, 0.8046, 1.0715, 0.8783. Note that* $q_R(1)$ *and* $q_R(2)$ *are very close to each other.*

An histogram of the value distribution of $\Re(Q_{(R,\epsilon)})$ *selecting* $\epsilon = 10^{-2}$ *is shown in figure 2 top left, a detail is shown in figure 2 bottom left.*

Similarly figure 2 top right shows the value distribution with a choice of $\epsilon = 10^{-3}$ *and we see a detail of the distribution at the right bottom.*

We can see that as ϵ *becomes smaller, the value distribution approximate the sum of the delta functions centered at the* $q_R(i)$*'s.*

Note that even very minor separations in the slopes of s_1 *and* s_2 *are resolved in this ideal situations as the details at the bottom of figure 2 show.*

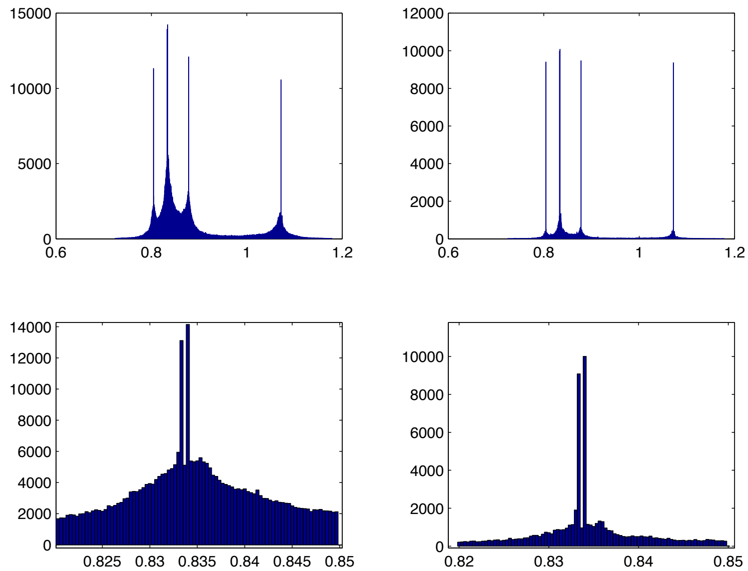

Figure 2: Value distribution of $\Re(Q_{R,\epsilon})$ and detail of the distribution for $\epsilon = 10^{-2}$ (left) and for $\epsilon = 10^{-3}$ (right)

3 Cluster Detection Algorithm

In this section we estimate the number of sources given two observations and we lay down the basis of the reconstruction algorithm. As we justify the assumptions and heuristic behind our method, we will deduce several steps of the algorithm that will be labeled as **(A1)**, **(A2)** and so on.

First of all, we have to choose the matrix R in such a way that the peaks of the value distribution of non zero values of $\Re(Q_{(R,\epsilon)})$ are enhanced when they correspond to the values of the $q_R(i)$'s.

If conditions (1), (2), (3) are fully satisfied and there is no noise, then any choice of a non singular R such that its rows are not orthogonal to any of the (a_i, b_i) is suitable, as this is sufficient to assure that all $q_R(i)$'s are finite, such choice is the generic case, therefore we can expect that almost any R will allow the detection of the sources.

In practice, when signals are not fully isolated anywhere, we need to select R so that, for each (t, w), the relative contribution in the numerator and denominator of $Q_R(t, w)$ due to each signal is as big as possible, since we do not want to reduce, with our choice of R, the area of the regions where each signal is "dominant".

This already implies that the direction of *both* the row vectors of R must be "far" from the direction of the vectors $(b_i, -a_i)$, $i = 1, ..., n$, the orthogonal vectors of the (a_i, b_i).

The specific value of each S_i will change from point to point, therefore we can only try to optimize the contribution of each *direction* determined by (b_i, a_i), $i = 1, ..., n$. To make such statement rigorous, let v_i, $i = 1, ..., n$ be unit vectors parallel to (b_i, a_i), and let $r_1 = (R(1,1), R(1,2))$ $r_2 = (R(2,1), R(2,2))$. Consider the function $F(r) = \sum_{i,j} \frac{(|<v_i,r>| - |<v_j,r>|)^2}{<v_i,r>^2}$, where $< a, b >$ denotes the inner product of a and b.

Then the choice of the second row of the exploratory matrix can be reduced to the solution of the following minimization problem:

$$(a) \;\; min_{r_2} F(r_2), \;\; |r_2| = 1.$$

We can then choose r_1 to be any slight perturbation of r_2, such that $|r_1 - r_2| < c|v_i - v_j|$, for all choices of $i \neq j$ with $c << 1$.

Clearly the exact directions associated to the (b_i, a_i)'s are not available as they are what we are looking for, so in general the choice of such "optimal" matrix cannot be determined and the use of several exploratory matrices, each enhancing a different source, is needed. This procedure can be done, but it would complicate considerably the algorithm.

As this paper is meant to be an introduction to the basic ideas behind such techniques, we restrict our attention to the case in which all (a_i, b_i) are in the positive quadrant, as this case correspond to the most relevant applications in speech processing in which the coefficients of the mixing matrix are positive attenuation coefficients of the energy intensity.

Given the previous restriction, any fixed choice of R such that r_1, r_2 are properly contained in the positive quadrant does assure that there is a lower bound on $< v_i, r_j >$, $j = 1, 2$ for any possible v_i in the positive quadrant, this in turn gives an upper bound on the possible value of $F(r_j)$ in the optimization problem (a). Note that the choice of $R = \begin{bmatrix} 1,0 \\ 0,1 \end{bmatrix}$, that is the simple quotient $\frac{X_1}{X_2}$, would reduce the resolution of any source s_i such that $(a_i, b_i) \approx (1, 0)$ or $(a_i, b_i) \approx (0, 1)$ since in the first case s_i would have marginal contribution in the denominator, and in the second case the signal would be marginal in the numerator.

Definition: Given a data set F, let F_β be the histogram of the values of F with bin size β. A measure of the roughness of F_β is:

$$I(F, \beta) = \sum_{n=-\infty}^{\infty} \frac{[\frac{1}{\beta}(F_\beta(n\beta) - F_\beta((n-1)\beta))]^2}{|F_\beta(n\beta)| + |F_\beta((n-1)\beta)| + J}$$

J is a parameter that has the effect of reducing the contribution of values of F_β

that are of the order of J. We can see in $I(F, \beta)$ a discrete modified version of the roughness penalty integral $\int \frac{f'^2}{|f|}$ used by Good and Gaskins in [GG].

With a slight abuse of notation denote by $\Re(Q_{(R,\epsilon)})$ the data set given by non-zero values of $\Re(Q_{(R,\epsilon)})(t, w)$ with (t, w) in the given time frequency domain.

We are ready now to write down the first step of our algorithm:

(A1) Slope Detection: Consider an exploratory matrix R_0 with positive rows bounded away from the vectors $(1, 0)$, $(0, 1)$. Compute $Q_{(R_0,\epsilon)}$ for some ϵ. Build a best estimation of the value distribution of $\Re(Q_{(R_0,\epsilon)})$ choosing the width $\bar{\beta}$ of the bins of the histogram of the values of $\Re(Q_{(R_0,\epsilon)})$ so that the roughness index $I(\Re(Q_{(R_0,\epsilon)}), \beta)$ is minimized for $\beta = \bar{\beta}$. Detect the position of the $q_{R_0}(i)$'s (*see next remark*) and compute the corresponding slopes $\frac{b_i}{a_i}$. The number E of slopes detected is our estimation of the number of distinct sources (in practice two sources withh very close slope may not be detected as distinct, see example 2 and the discussion that follows it).

Remark 3.2: Because of the presence of J, "large" peaks of $\Re(Q_{(R_0,\epsilon)})_{\bar{\beta}}$ will be quite smooth. In practice we see that, when the speech time series are sufficiently long (order of few seconds with sampling rate of 8192 Hz), a choice of $J \approx 100$ is often sufficient to smoothen the major peaks. The smoothness of the main peaks is actually so high when enough data are used, that we can detect the position of the $q_{R_0}(i)$'s simply by the following procedure that detect "large" local maxima:

Let $F = \Re(Q_{(R_0,\epsilon)})$ and consider the discrete function $D(F_{\bar\beta}) = |F_{\bar\beta}(n\bar\beta) - F_{\bar\beta}((n-1)\bar\beta)|$. Let $L = max\, D(F_{\bar\beta})$ be the maximum local displacement. We assume that a value x corresponds to a true $q_R(i)$ if $F_{\bar\beta}(x)$ is a local maximum and if we can find y_1 and y_2 such that $y_1 < x < y_2$ with $|F_{\bar\beta}(x) - F_{\bar\beta}(y_i)| > \phi L$, $i = 1, 2$ and $F_{\bar\beta}(y) < F_{\bar\beta}(x)$ for $y \in [y_1, y_2]$, $y \neq x$, $\phi >> 1$.

This strategy would work only if the smoothness of the histogram is relatively uniform on its domain, otherwise a very sharp main feature can produce a value of L that is too large for less pronunced peaks.

There is an element of indetermination in the choice of ϕ, we want at least $\phi > 1$, but one may need larger values of ϕ.

In any case we left aside this issue in the description of step A1 since there are several ways to choose the main features of an histogram and such choice is part of a more general problem than the one treated in this paper.

We can now identify the regions where the identified sources are isolated.

(A2) Cluster Construction: For $j = 1, ...n$, compute functions Q_j such that $Q_j(t, w) = Q_{(R_0,\epsilon)}(t, w)$ if $|\Re(Q_{(R_0,\epsilon)}(t, w)) - q_{R_0}(j)| < \bar\beta$ and $Q_j(t, w) = 0$ otherwise. Let $G_j = Supp(Q_j)$ be the support of the Q_j's.

The G_j's are our estimates of the regions of the time-frequency plane where the S_j's are isolated.

Suppose now that a total of n observations x_i were available and let, to make notation uniform, $x = Ms$ where $x = (x_1, ..., x_n)^t$, $s = (s_1, ..., s_n)^t$ and M is an $n \times n$ invertible matrix. Note that for each point in the time-frequency domain $M^{-1}X(t,w) = S(t,w)$ where $X(t,w) = (X_1(t,w),...,X_n(t,w))^t$ and $S(t,w) = (S_1(t,w),...,S_n(t,w))^t$. Assume that $E = n$, that is, assume that we are able to identify isolated regions in time-frequency domain for *all* signals.

Then, given values (t_i, w_i) such that $(t_i, w_i) \in G_i$, $i = 1,...,n$, the following equalities hold:

$$(i) \quad M^{-1}X(t_i, w_i) = (0, 0, ..., S_i(t_i, w_i), ..., 0)^t,$$

since at each (t_i, w_i) only S_i is non-zero.

Each of the equalities (i) for $i = 1,...,n$ gives some conditions on the coefficients of M^{-1}, therefore we can now write the last two steps of our reconstruction algorithm.

(A3) Constraints on Inverse Mixing Matrix: Choose points $(t_i, w_i) \in G_i$, $i = 1,...,k_2$, let $(p_1,...,p_n)$ be the rows of M^{-1}, for each p_k $k = 1,...,n$ consider the systems:

$$(E_k) \ \{< p_k, X(t_i, w_i) >= 0, \ \ i \neq k\}.$$

Find non zero solutions $\bar{p}_k$ of (E_k). Build the matrix $\tilde{M}^{-1} = [\bar{p}_1, ..., \bar{p}_n]$.

(A4) Reconstruction of the Sources: Apply $\tilde{M}^{-1}$ to $(x_1,...,x_n)^t$, then $(\tilde{s}_1,...,\tilde{s}_n)^t = \tilde{M}^{-1}(x_1,...,x_n)^t$ is our estimate of the sources.

Each system (E_k) is an underdetermined system of $n-1$ equations in n unknows (the coefficients of each p_k). Therefore each specific solution $\bar{p}_k$ of (E_k) is a multiple of some row of M^{-1} and $\tilde{M}^{-1}$ is a rescaled permutation of the rows of M^{-1}.

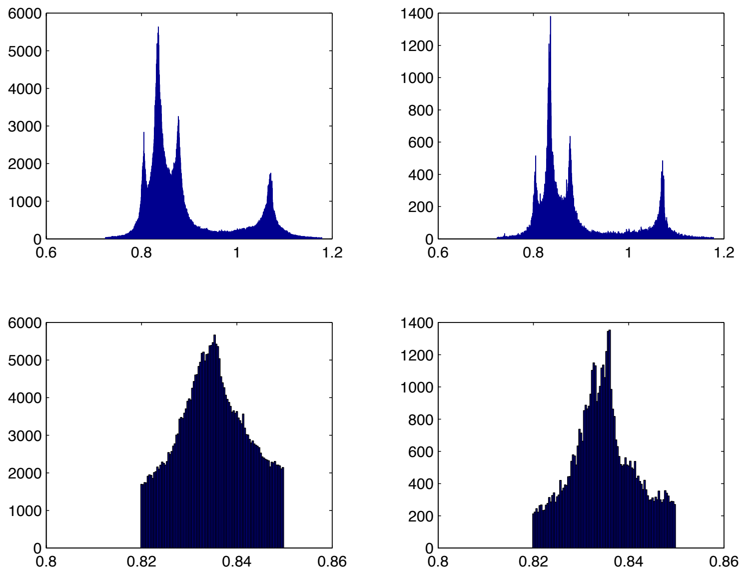

Figure 3: Value distribution of $\Re(Q_{R,\epsilon})$ and detail of the distribution for $\epsilon = 10^{-2}$ (left) and for $\epsilon = 10^{-3}$ (right).

In other words $\tilde{M}^{-1} = \Lambda S(M^{-1})$ where Λ is a non-singular diagonal matrix and $S(M^{-1})$ is a permutation of the rows of M^{-1}.

We mentioned several times up to now that in general signals may be dominant in some regions of the transformed domain, but not fully separated. Experimental work shows that for real speech signals we cannot achieve separation of two sources if their corresponding slopes are very close unlike the case when there are fully isolated regions, as we can see in the following example:

Example 2: let us work out example 1 again without setting artificially the signals equal to zero on small windows.

Figure 3 shows that the histogram of the value distribution with optimal bin size does not allow to distinguish $q_R(1)$ and $q_R(2)$ for $\epsilon = 10^{-2}$, but it achieves

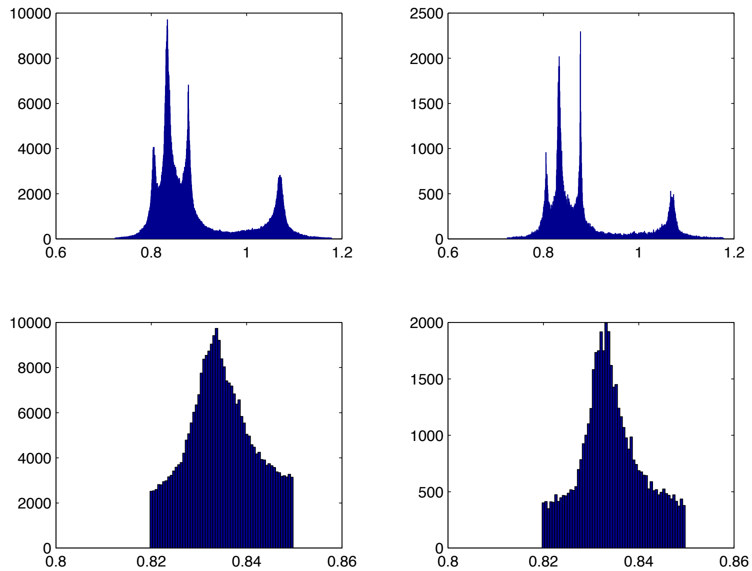

Figure 4: Value distribution of $\Re(Q_{R,\epsilon})$ and detail of the distribution for $\epsilon = 10^{-2}$ (left) and for $\epsilon = 10^{-3}$ (right) when the intensity of the sources is changed.

some separation for $\epsilon = 10^{-3}$ (see detail at the bottom right).

On the other hand the same analysis in which s_1 and s_2 are replaced by $s_1' = 4s_1$ and $s_2' = \frac{s_2}{8}$ fails to show any discrimination between the two sources as shown by the histograms in figure 4. Note that the directions of the slopes of the sources was not changed.

Remark 3.3: Example 2 suggests that closeness of the slopes and the degree of dominance of signals in small regions are related in achieving separation, and we wonder to which extent the imaginary part threshold can help in enhancing the resolving power of the algorithm. Clearly the specific choice of the threshold is a crucial factor in achieving optimal resolution. But the possibility of using very small thresholds is limited in practice also by the finite sample size. We

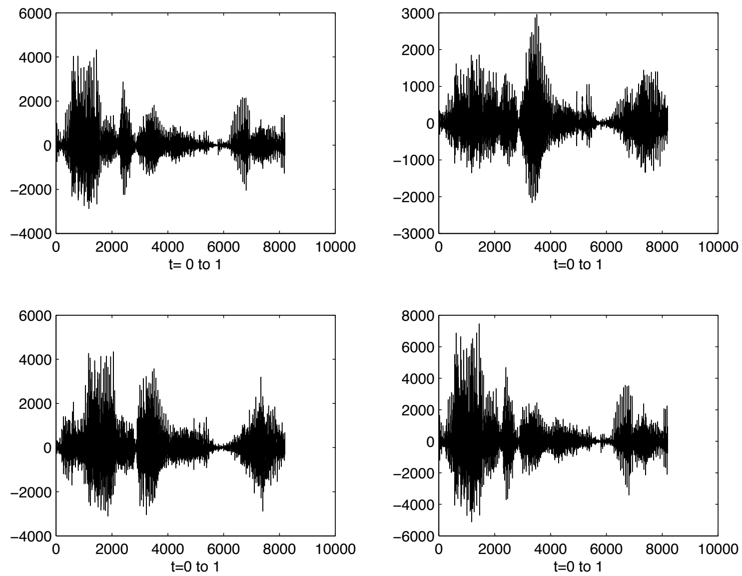

Figure 5: Observations.

need in this case to control ϵ so that the optimal roughness index $I(\Re(Q_{(R,\epsilon)},\bar{\beta})$ is always below some given constant. Further work is in progress on these issues.

It turns out that the previous simple algorithm performs well also for real speech signals when the slopes of the sources are well separated.

Example 3: Consider the case in which the four speech signals s_1, s_2, s_3 *and* s_4 *are mixed with the mixing matrix* $M = \begin{bmatrix} 0.2, 0.6, 0.18, 0.3 \\ 0.2, 0.22, 0.25, 0.5 \\ 0.65, 0.2, 0.4, 0.6 \\ 0.5, 0.99, 0.3, 0.4 \end{bmatrix}$ *on a time interval of one second. Let* $s_2(t) = s_1(t+\tau)$ *with* $\tau = 0.073$*, that is two of the sources are shifted versions of each other. We show the resulting mixtures* x_i*,* $i = 1, ..., 4$ *in figure 5.*

Assume we know that all coefficients of M *are positive, then we can use the exploratory matrix* $R = \begin{bmatrix} 2, 7 \\ 1, 1 \end{bmatrix}$ *and apply directly step A1 to the first*

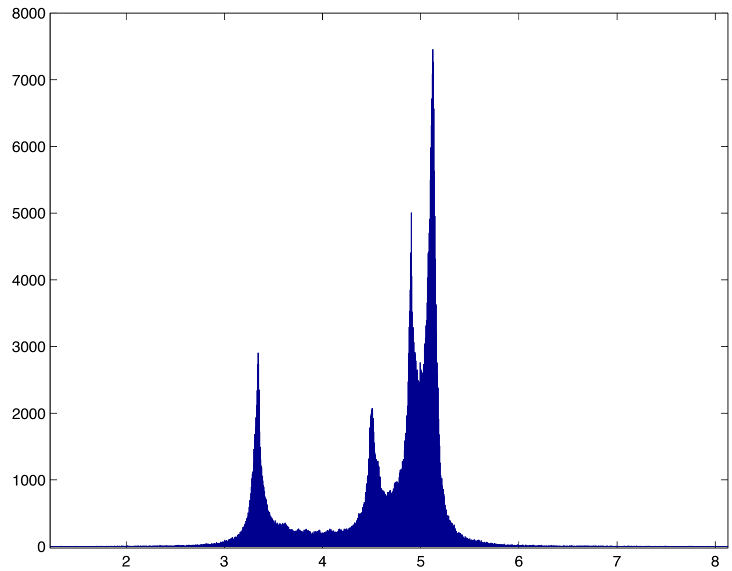

Figure 6: Value distribution of $\Re(Q_{R,\epsilon})$

two mixtures with a choice of threshold $\epsilon = 10^{-2}$. *The value distribution with optimized bin size* $\bar{\beta} = 0.0054$ *is shown in figure 6, in figure 7 we show the graph of the index computed for decreasing values of* $\beta = \frac{1}{n}$, $n = 10, ..., 1000$.

The estimates of $q_R(i)$*'s corresponding to the major peaks of the histograms are in increasing order, 3.3475, 4.5020, 4.9025, 5.1275. The true values are: 3.3415, 4.5000, 4.9070, 5.1250, the relative error is less than* $2 * 10^{-3}$ *for all sources. Step A2 gives us the clusters whose details are shown in figure 8*

Before applying step A3 we preprocessed the clusters so that we retain only the non zero values of the G_i*'s that are in small rectangular regions where there is very high density of non zero elements, since we expect regions where each signal is isolated to have positive area.*

This heuristic adjustment reduces in practice the chance of selecting in step

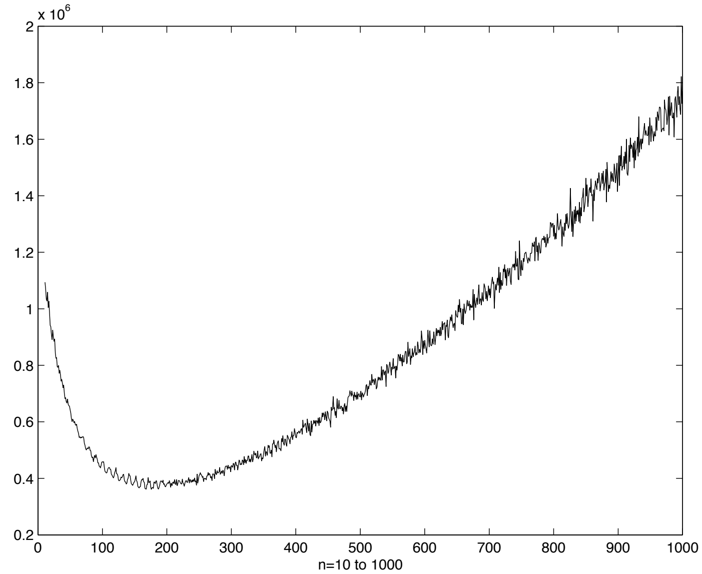

Figure 7: Computation of the roughness index $I(\Re(Q_{R,\epsilon}), \frac{1}{n})$ for $n = 10...1000$

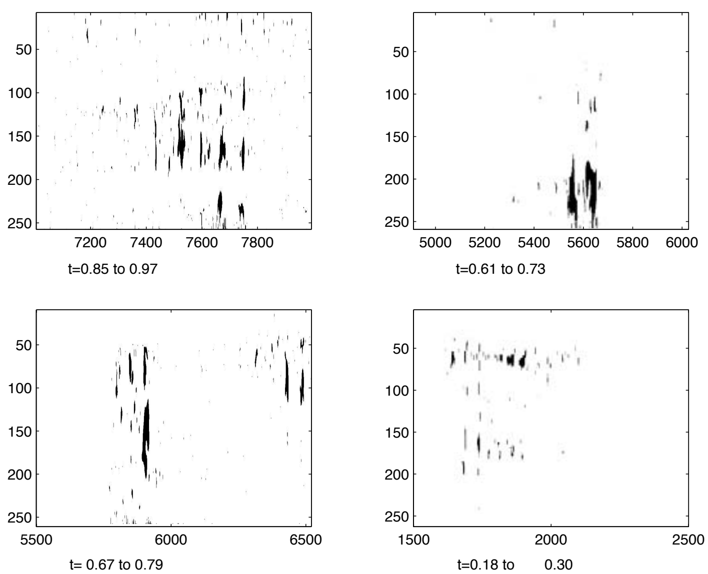

Figure 8: Details of clusters where sources are isolated

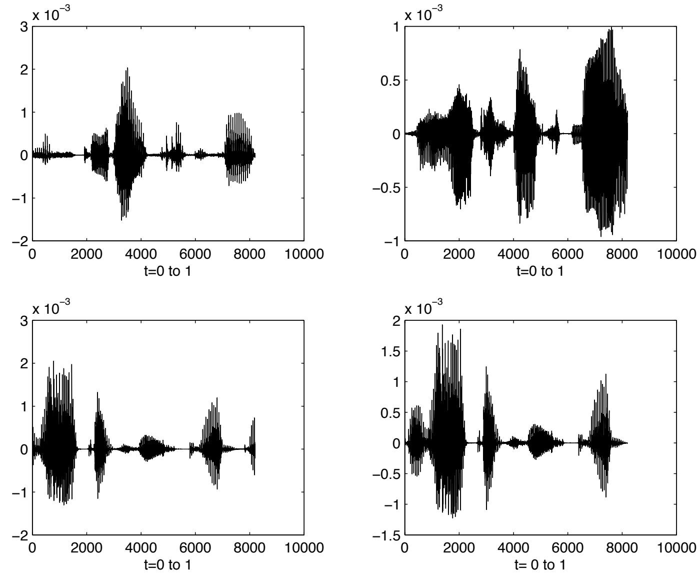

Figure 9: Estimates of, clockwise from top left, s_4, s_3, s_2 and s_1

A3 points that do not belong to positive area regions where signals are isolated.

Then we apply step A3 to randomly chosen points in these dense regions of the clusters G_i *and we get* $\tilde{M}^{-1} = \begin{bmatrix} -1.8097, -1.8046, -1.6412, -1.8021 \\ 0.8726, 0.8553, 0.7969, 0.8327 \\ -0.4809, -0.4771, -0.5094, -0.4594 \\ 1.0000, 1.0000, 1.0000, 1.0000 \end{bmatrix}$

where all last components of each row were chosen equal to one. Finally step A4 gives the reconstructions shown in figure 9, rescaled so that they have unit l^1 *norm .*

Compare these reconstructions to the rescaled original sources shown in figure 10. All reconstructions have a signal to noise ratio between 16 and 22 decibel.

We would like to stress that, while on one hand the choice of mixing matrix is somehow non degenerate since the slopes of the sources are "far" from each

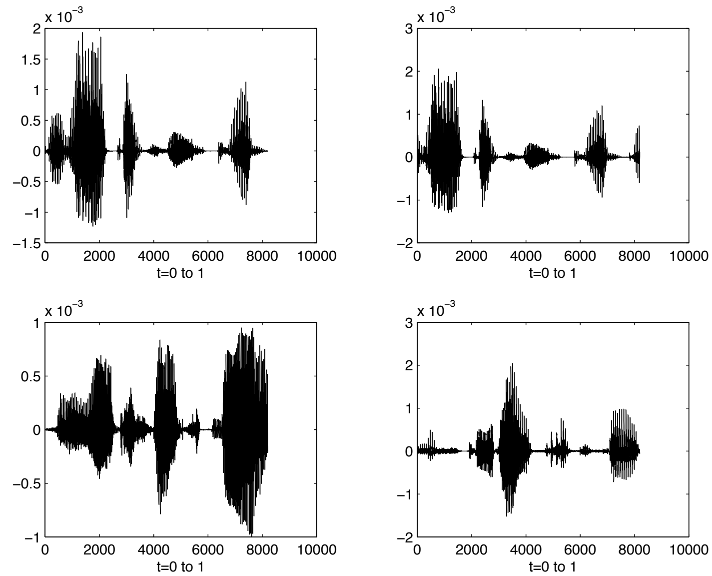

Figure 10: Original Sources, clockwise from top left: s_1, s_2, s_3, s_4

other and the norm of $a_i s_i$, $b_i s_i$ *is of the same order of magnitude for* $i = 1, ..., 4$, *on the other hand the length of the time interval used for our analysis is only one second as opposed to several hundred seconds in traditional ICA algorithms. It seems likely that a long time interval will benefit the accuracy of our algorithm as well, since a long time interval increases the chance of having areas where signals are almost isolated as in example 1 (for our case study of speech signals, different people, hopefully, start to speak at different times...).*

We believe that the quotient projections algorithm is only the first step of a class of non-linear projection algorithms that make full use of the interplay of real and imaginary parts of homogeneous quotients. In a forthcoming paper we discuss such more general quotient projections algorithms and we expand the theoretical basis of the method.

Acknowledgements

This research was supported in part by the Office of Naval Research under the ODDRE MURI97 Program Grant No. N000149710501EE to the Center for Auditory and Acoustics Research and by the Army Research Office under the ODDRE MURI01 Program Grant No. DAAD19-01-1-0465 to the Center for Communicating Networked Control Systems (through Boston University). The second author was also supported in part by the National Science Foundation under grant No. NSF-DMS-0070044.

We would like to thank Shihab Shamma and Daniele Carlo Struppa for valuable and stimulating discussions. The first author would also like to thank Struppa for his support and for reviewing a preliminary version of this paper.

References

[**CC**] S. Choi, A. Cichoki, Blind Separation of Nonstationary Sources in Noisy Mixtures. Electronic Letters, vol. 36, n. 9, pp. 848-849, 2000.

[**D**] D. Donoho, Sparse Components of Images and Optimal Atomic Decompositions. Available at www-stat.stanford.edu/ donoho/Reports/1998/SCA.pdf

[**GG**] I.J. Good, R.A. Gaskins, Nonparametric roughness penalties for probability densities. Biometrika(1971), **58**, 2, pp. 255-277.

[**H**] P. J. Huber, Projection pursuit. With discussion. Ann. Statist. 13 (1985), pp. 435-525.

[**HR**] J.H. van Hateren, D.L. Ruderman, Independent component analysis of natural image sequences yields spatiotemporal filters similar to simple cells in primary visual cortex. Proc. R. Soc. Lond. B 265, 1998.

[**L**] T.-W. Lee, *Independent Component Analysis. Theory and Applications*, Kluwer, Boston, 1998.

[**OF**] B.A. Olshausen, D.J. Field, Sparse coding with an overcomplete basis set: a strategy employed by V1? Vision Research, 37: 3311-3325, 1997.

[**QKS**] Y. Qi, P. S. Krishnaprasad, S. Shamma, The Subband-based Independent Component Analysis. Proceedings of ICA2000, 19-22 June 2000, Helsinki, Finland.

[**RD**] S. Rickard, F. Dietrich, DOA Estimation of Many W-Disjoint Orthogonal Sources From Two Mixtures Using Duet. Statistical Signal and Array Processing, 2000. Proceedings of the Tenth IEEE Workshop on , 2000 pp. 311-314.

[**RBR**] S. Rickard, R. Balan, J. Rosca, Real-Time Time-Frequecy Based Blind Source Separation. Proceedings of ICA2001, 9-12 December 2001, San Diego, California, USA.

[**RY**] S. Rickard, O. Yilmaz, On The Approximate W-Disjoint Orthogonality of Speech.Acoustics, Speech and Signal Processing, 2002 IEEE International Conference on , Volume: 1 , 2002 pp. 529-532.