
Quotient Signal Estimation

D. Napoletani¹, C. A. Berenstein², P. Krishnaprasad³, and D. C. Struppa⁴

¹ School of Computational Sciences, George Mason University, Fairfax, VA 22030
dnapolet@gmu.edu

² Institute for Systems Research, University of Maryland, College Park, MD 20742
carlos@glue.umd.edu

³ Institute for Systems Research, University of Maryland, College Park, MD 20742
krishna@glue.umd.edu

⁴ Department of Mathematical Sciences, George Mason University, Fairfax, VA
22030 dstruppa@gmu.edu

Summary. In this paper we propose a method for blind signal decomposition that does not require the independence or stationarity of the sources. We define suitable quotients of linear combinations of the images of the mixtures in a given frame and we show experimentally that such quotients can be used to recursively extract three sources from only two measurements. A general strategy to extract more than three sources from two measurements is proposed.

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 more than two sources given at least two mixtures. More specifically let $x_1[i] = a_1 s_1[i] + \dots + a_N s_N[i]$, $x_2[i] = b_1 s_1[i] + \dots + b_N s_N[i]$, $i = 1, \dots, L$ be the two mixtures of N real-valued discrete sources $s_n[i]$, $n = 1, \dots, N$, $i = 1, \dots, L$, with L the length of the signals and $a_i, b_i \in \mathbb{R}$. Compute the expansion of x_1 and x_2 in some suitable frame dictionary $\mathcal{D} = \{g_1, \dots, g_M\}$, $M > N$ for L -dimensional discrete signals, that is, compute the inner products $X_u[k] = \langle x_u, g_k \rangle$, $k = 1, \dots, M$

and $u = 1, 2$.

Clearly, if we denote the expansion of s_n in \mathcal{D} by S_n , we have:

$$X_1 = a_1 S_1 + \dots + a_N S_N, \quad X_2 = b_1 S_1 + \dots + b_N S_N.$$

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

$$Q_R[k] = \frac{R(1,1)X_1[k] + R(1,2)X_2[k]}{R(2,1)X_1[k] + R(2,2)X_2[k]}, \quad k = 1, \dots, M$$

In this paper we use the collection of $Q_R[k]$ for some choice of R to find atoms in the dictionary where each source is dominant. This in turn reduces the search for the estimation of the sources to a type of projection pursuit. The reasoning behind our method is that, if at k_0 there is only one source, say S_{n_1} , then

$$Q_R[k_0] = \frac{R(1,1)a_{n_1}S_{n_1}[k_0] + R(1,2)b_{n_1}S_{n_1}[k_0]}{R(2,1)a_{n_1}S_{n_1}[k_0] + R(2,2)b_{n_1}S_{n_1}[k_0]} = \frac{R(1,1)a_{n_1} + R(1,2)b_{n_1}}{R(2,1)a_{n_1} + R(2,2)b_{n_1}}$$

that is, *the value of the quotient at k_0 is independent of the value of the measurements at that point, but it depends only on the coefficients a_{n_1} , b_{n_1} of S_{n_1}* . Ideally, if at each k only one source is dominant, the distribution of the values of Q_R , is made up of several delta functions of different weight centered at 0 (due to the elements in the dictionary where there is no contribution from any signal), and at the value of the quotients

$$q_R(n) = \frac{R(1,1)a_n + R(1,2)b_n}{R(2,1)a_n + R(2,2)b_n},$$

$n = 1, \dots, N$ that we can assume finite for a generic choice of R . To be able to discriminate the sources in the histogram of the value distribution of Q_R , it is therefore necessary that $\frac{b_n}{a_n} \neq \frac{b_m}{a_m}$ when $n \neq m$, we call $\frac{b_n}{a_n}$ *the slope* of the source s_n . In this ideal setting, the number of ‘dominant’ peaks of the value distribution of the non zero values of Q_R gives us an estimate of the number of sources and their positions will give the values of the $q_R(n)$ and therefore also the values of the slopes $\frac{b_n}{a_n}$. Unfortunately in practice it is not possible to enforce a total separation of the sources in \mathcal{D} , therefore the value distribution of Q_R will at best give the position of the few most prominent peaks. So, any effective algorithm based on the previous insights must be able to recursively reduce the influence of the sources on the shape of the histogram of Q_R . Even then, we need additional restrictions to avoid having degenerate situations in which ghost sources are detected. Assume for example that $S_m = pS_n$, p constant, $0 < m, n \leq N$ in some region $\mathcal{M} \subset \mathcal{D}$ where the contribution of other signals is marginal. Then we have that, on \mathcal{M} :

$$Q_R = \frac{R(1,1)(a_n S_n + a_m S_m) + R(1,2)(b_n S_n + b_m S_m)}{R(2,1)(a_n S_n + a_m S_m) + R(2,2)(b_n S_n + b_m S_m)} =$$

$$\frac{R(1,1)(a_n + pa_m) + R(1,2)(b_n + pb_m)}{R(2,1)(a_n + pa_m) + R(2,2)(b_n + pb_m)}$$

The slope $\frac{b_n+pb_m}{a_n+pa_m}$ of a ‘source’ that does not exist as physical entity would be detected. Therefore we must require that sources are linearly independent on most atoms of the dictionary. On the other hand it is indeed possible that sources can be very similar in some cases, 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 summarize now the previous discussion in a series of conditions that seem to be necessary for the histogram of the value distribution of Q_R to help in detecting the sources :

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

Condition (2): Each source is dominant on a subset of elements of \mathcal{D} .

Condition (3): Sources must be linearly independent for ‘most’ values of $k = 1, \dots, M$, i.e., we should not be able to find real numbers (p_1, \dots, p_n) such that $p_1S_1[k] + \dots + p_nS_n[k] = 0$.

These conditions are not fully rigorous, moreover conditions (2) and (3) are dependent on the choice of \mathcal{D} , which, as we will see, is changing in time in the algorithm we developed, nevertheless such restrictions are the background against which we can assess the practical applicability of our method to specific problems. In the examples of next section something stronger than condition (1) is enforced, namely we select the mixing matrix in a set of matrices such that the angle between any two vectors (a_i, b_i) is larger than a given constant. We produce in this way matrices that are ‘non degenerate’ with respect to our method. Note that we do not require the sources to be independent or stationary, but rather we have geometrical separation conditions of the sources in the dictionary \mathcal{D} . In essence, two related data sets (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). In a way, we can say that our approach relaxes the requirements on the sources while it imposes more stringent conditions on the mixing matrices.

The second section of this paper introduces the general strategy of the ‘quotient projection’ algorithm. In section three we perform experimental separations for the case of two mixtures and three sources. Moreover we show that, for the case of speech sources, a more sophisticated technique is needed if we want to extract four sources from two measurements. The strategy of such

generalized algorithm is outlined.

The possibility of identifying and separating sources through the use of the simple quotient $\frac{X_1}{X_2}$ in time frequency domain was already underlined by Rickard and collaborators in [RD], [RBR], [RY], and [BR], but see also our preliminary unpublished work [NBK]. In this paper on the other hand we stress the possibility of choosing *several* exploratory matrices to build a robust algorithm that can be implemented in *general* dictionaries, especially the results in section three show that any one single choice of exploratory matrix is unable to properly identify all sources for most mixing matrices.

It should be noted that a major problem for speech (audio) signals is to find the correct model of their mixing. Our assumption here (implicit in the way in which the mathematical model is constructed) corresponds to instantaneous mixing; while this choice is common in testing the basic performance of blind signal separation algorithms, it is very restrictive for real world signals, where one needs to take into account echoes and asynchrony of measurements.

2 Quotients Projections

We directly give a description of the basic steps of our algorithm (based on the insights of the previous section), we then explain the heuristics behind it.

- (a) Given x_1 and x_2 , choose a dictionary \mathcal{D} and an exploratory matrix R .
- (b) Compute X_1 and X_2 , compute the quotient Q_R and find the first N ‘significant’ maxima q_n , $n = 1, \dots, N$ of the histogram of the value distribution of Q_R .
- (c) Set the initial estimate of sources \bar{s}_n as $\bar{s}_n[i] = 0$, $i = 1, \dots, L$, $n = 1, \dots, N$.
- (d) Choose positive numbers ϵ_n and let \mathcal{T}_n be the collections of all g_k in \mathcal{D} such that $|Q_R[k] - q_n| < \epsilon_n$.
- (e) Let $Z_{n,u}[k] = X_u[k]$, $u = 1, 2$ and $n = 1, \dots, N$ if $g_k \in \mathcal{T}$ and $Z_{n,u}[k] = 0$ otherwise. Compute the inverses $z_{n,u}$, of $Z_{n,u}$, $u = 1, 2$, $n = 1, \dots, N$.
- (f) Let $\bar{s}_n = \bar{s}_n + z_{n,1}$, $n = 1, \dots, N$. and set $x_u = x_u - \sum z_{n,u}$, $u = 1, 2$. Go to step (a).

The first thing to note is that (a)-(f), as it is, does not have a termination rule. In the experiments performed in next section we iterated the core loop (a)-(f) a fixed large number of times. Note also that the selection of \mathcal{D} and R in (a) does not need to be the same at every iteration. Actually we will see in

the examples of the following section that there is an advantage in selecting several dictionaries to find all atoms that contribute to a given source. The same reasoning can be applied to the choice of R : we will give evidence in the next section that the most general strategy based on (a)-(f) uses several choices of R . The algorithm (a)-(f) first identifies the dominant sources in step (b), with the inspection of the histogram of Q_R , then it uses the localization of the peaks associated to these sources in steps (d) to (e) to ‘extract’ all the structure belonging to such dominant sources as visible in the dictionary \mathcal{D} . The whole process is repeated with another dictionary in a fashion similar to matching pursuit algorithms [M]. Step (b) is ambiguous as we do not explain how to find the ‘significant’ sources, the fact is that the number of sources that can be extracted is limited in any case when the sources are speech signals (our main case of study). In our implementation we assume we know a priori the maximum number N_0 of sources and we build a robust estimator of the first significant $N < N_0$ maxima of the histogram of the value distribution of Q_R . The specific implementation of step (b) of our procedure is the following:

(b’) Compute Q_R . Build a best estimation of the value distribution of Q_R between Q_5 and Q_{95} , respectively the 5 and 95 percentiles (to avoid the selection of large domains due to extreme outliers), Choose the width of the bins of the histogram to be $\frac{Q_{95}-Q_5}{500}$, this bin size is fine enough provided R is close to a singular matrix, so that the true quotients are effectively close to each other. Let H_R be the resulting histogram. We then use smoothing splines to find the best approximation of H_R that has no more than N_0 maxima. Let q_n $n = 1, \dots, N < N_0$ be the ordinate of such maxima, and (l_n, r_n) the inflection points left and right of q_n . Let $m_n = \min(|H_R(q_n) - H_R(l_n)|, |H_R(q_n) - H_R(r_n)|)$, and $M = \max_n m_n$, we consider significant only the maxima such that the $m_n > 0.05M$. Moreover we set $\epsilon_n = \frac{1}{2} \min(|q_n - l_n|, |q_n - r_n|)$.

It is immediately evident that there are several parameters chosen in (b’) in a somewhat arbitrary way, nevertheless (b’) as stated seems to work for large classes of mixing matrices and for most speech sources we tested. Further experimental and theoretical work would be necessary to establish if (b’) is indeed the most efficient implementation of (b). Here we only say that often the true quotient $q_R(n)$ will not be detected as a maximum of H_R , but it will appear as sizeable asymmetrical ‘enlargement’ of one of the other peaks (see Figure 6 in next section), these enlargements are associated to inflection points of the smoothed histogram, so our choice of ϵ_n as half the distance of the computed quotients q_n from the nearest inflection point prevents the most common cause of clustering two separate sources in a single reconstruction.

Remark: we restrict our attention to the case in which all (a_n, b_n) are in the positive quadrant, as this case correspond to the most relevant applications

in speech processing where 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 $\langle v_i, r_j \rangle, j = 1, 2$ for any possible v_i in the positive quadrant. 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 ability of identifying any source s_n such that $(a_n, b_n) \approx (1, 0)$ or $(a_n, b_n) \approx (0, 1)$ since in the first case s_n would have marginal contribution in the denominator, and in the second case the signal would be marginal in the numerator and there would be no significant peak associated to $q_R(n)$ in the histogram of the value distribution of Q_R .

Remark: in [BR] the authors offer an interesting stochastic model for mixing of speech signals, that in the case of quotient $\frac{X_1}{X_2}$ provides an optimal estimation of the position of the peaks. It is conceivable that similar ideas could be used in our case, instead of the heuristic in (b'), to find the position of the peaks, the identification of ϵ_n in (d) above, and possibly to select the optimal exploratory matrix as well.

3 Experiments and Further Developments.

We treat in this section two distinct cases: the case of three sources of interest and two measurements, and the case in which there are four sources and two measurements. In this second case we give evidence that no single choice of exploratory matrix is enough to find all four quotients.

In the following two experiments we apply the algorithm to speech signals from the TIMIT database. Because of this choice, the dictionary in (a) is selected to be a cosine packets or wavelet packets basis (see [M]) of the sum of the current x_1 and x_2 . These types of bases are known to be effective in approximating speech signals and therefore they increase the chance of having condition 2 satisfied. Moreover to increase the robustness of the iterative structure extraction, we attenuate the coefficients selected in (e) and we alternate between wavelet packets bases and cosine packets bases in (a). We explicitly write these modifications of steps (a) and (e):

(a') Given x_1 and x_2 , compute $x = x_1 + x_2$. Choose a 'random' cosine or wavelet packets basis \mathcal{D} for x . Choose an exploratory matrix R .

(e') If the basis used is a cosine packets basis set $\alpha = 0.4$, otherwise if it is a wavelet packets basis set $\alpha = 0.1$. Let $Z_{n,u}[k] = \alpha X_u[k]$, $u = 1, 2$ and $n = 1, \dots, N$ if $g_k \in \mathcal{T}$ and $Z_{n,u}[k] = 0$ otherwise. Compute the inverses $z_{n,u}$, of $Z_{n,u}$, $u = 1, 2, n = 1, \dots, N$.

Remark: The values of α for cosine and wavelets packets were heuristically adjusted to obtain good reconstruction when the algorithm alternates one iteration using cosine packets and one using wavelets packets. As regards the ‘random’ choice of basis in (a’), in this paper we simply selected a small set of basis trees and alternated among them uniformly to generate our cosines and wavelets packets bases. The issue of basis selection deserves certainly a more detailed study and in a subsequent paper we will compare the performance of several different ways to choose dictionaries in (a’). We are now ready to write down the basic implementation of our algorithm:

(A1) Apply (a) and (b’) once with the windowed Fourier frame (see [S]) as choice of dictionary \mathcal{D} .

(A2) Apply (a’), (c)-(d), (e’) and (f) a fixed number ($T = 60$) of times, alternating in (a’) a choice of cosine packets and wavelet packets basis.

We first expand in the windowed Fourier frame (rather than a basis) to assure the generation of a relatively smooth histogram determined by *local* time frequency structure. Note that the identification of the dominant regions in the windowed Fourier frame are used only to ‘track’ the positions of the quotients $q_R(n)$: the sources do not need to be totally separated in any given basis. Rather we use several bases in (A2) to recursively separate the structure belonging to different sources.

Example 1: In our first set of experiments, we asks for the estimation of three unknown speech sources (of a duration of about one second) given two measurements, a situation that, in the setting of this paper, can be described by the model:

$$x_1 = a_1 s_1 + a_2 s_2 + a_3 s_3 \quad x_2 = b_1 s_1 + b_2 s_2 + b_3 s_3.$$

We apply the algorithm to 30 instances of the model with $a_u, b_u, u = 1, 2, 3$ chosen as coordinates of unit vectors in the positive quadrant such that the angle between any two of them is at least $\pi/16$. The specific way we generated such mixing matrices is the following: we repeatedly compute three angles $\theta_n, n = 1, 2, 3$, chosen uniformly in $[0, \frac{\pi}{2}]$ until their minimum angular distance is bigger than $\frac{\pi}{16}$. We then set $a_n = \sin(\theta_n)$ and $b_n = \cos(\theta_n)$. The choice of the exploratory matrix is taken as $R = \begin{bmatrix} \sin(\frac{\pi}{4}) & \cos(\frac{\pi}{4}) \\ \sin(\frac{\pi}{4} + \delta) & \cos(\frac{\pi}{4} + \delta) \end{bmatrix}$, for some $\delta \ll \frac{\pi}{16}$. The reason we take both numerator and denominator close to $(\sin(\frac{\pi}{4}), \cos(\frac{\pi}{4}))$ is that for this choice we have the largest possible lower bound on the reduction of the norm of the projection of any vector in the

positive quadrant, so avoid, as much as we can, to unwittingly reduce the contribution of one source to the overall value distribution of Q_R .

Figure 1 shows the generic smoothened histogram generated by (b'). The attenuation coefficient α in (e') prevents a full recovery of the energy of the sources, but what we are really interested in is of course the overall shape of the reconstruction, so in Figures 2, 3 and 4 we show from top to bottom the original sources and the reconstructions (for one representative choice of mixing matrix) scaled to have norm one. In Figure 5 we show the two original mixtures. Since the comparison between sources and reconstructions is meaningful only modulo a rescaling of them, we define the scaled Signal-to-Noise Ratio (scaled SNR) to be the usual SNR (measured in decibel) where signal and reconstruction are scaled to have norm one.

We can summarize our results by noting that: in all cases except one we find three maxima and therefore we obtained reconstructions of all sources s_1, s_2 and s_3 96% of the time. The mean scaled SNR was: 7.6db for reconstructions of s_1 , 6.4db for reconstructions of s_2 and 6.3db for reconstructions of s_3 . The perceptual quality was very good for most reconstructions.

Incidentally the same set of 30 mixtures, if analyzed with the exploratory matrix $R = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, leads to lower scaled SNR and lower proportion of reconstructions: we reconstruct exactly one source 20% of the time, exactly two sources 43% of the time and just 37% of the time we recover exactly three sources. On the other hand, the mean scaled SNR for the cases in which we do have reconstruction is only slightly worse than for the exploratory matrix of example 1. We have scaled SNR of :7.6db for the nonzeros reconstructions of s_1 , 5.8db for the nonzeros reconstructions of s_2 and 6.2db for the nonzeros reconstructions of s_3

Example 2: The generalization to the recovery of more than three sources in the setting of algorithm (A1)-(A2) is problematic. Assume for example that we have 4 specific speech sources s_i , $i = 1, \dots, 4$ (of a duration of about one second) and two mixtures,

$$x_1 = a_1s_1 + a_2s_2 + a_3s_3 + a_4s_4 \quad x_2 = b_1s_1 + b_2s_2 + b_3s_3 + b_4s_4.$$

All other details are as in example 1, modulo the increase in the number of sources. The result of applying algorithm (A1)-(A2) to this setting is somewhat disappointing. We reconstruct exactly one source 20% of the time, exactly two sources 54% of the time, exactly three sources 23% of the time and just 3% of the time we recover exactly four sources. The mean scaled SNR is: 4db for reconstructions of s_1 , 4.2db for reconstructions of s_2 , 3.7db for reconstructions of s_3 and 4.2db for reconstructions of s_4 . In Figure 6 we show a typical histogram generated by (b'), we see clearly that the procedure fails to

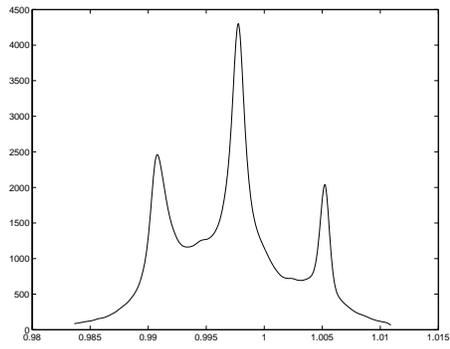


Fig. 1. Generic smoothed histogram from step (b') as described in the text with three sources.

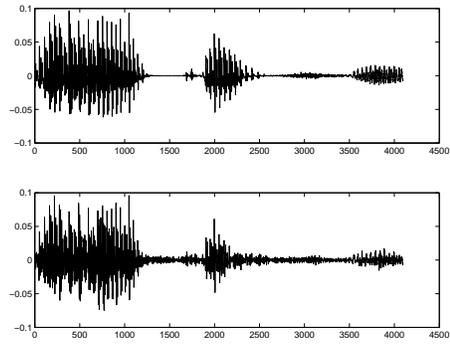


Fig. 2. From top to bottom, signal s_1 and one instance of reconstruction \bar{s}_1 , both scaled to have norm 1.

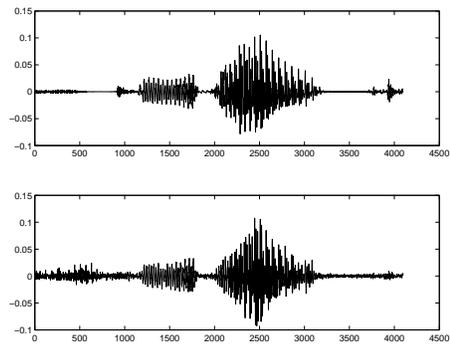


Fig. 3. From top to bottom, signal s_2 and one instance of reconstruction \bar{s}_2 , both scaled to have norm 1.

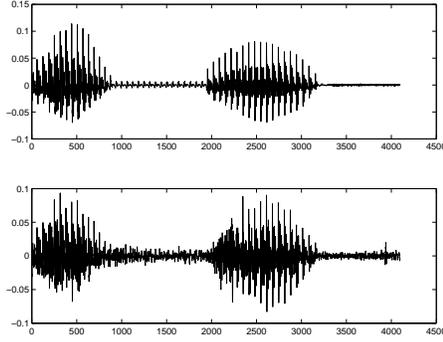


Fig. 4. From top to bottom, signal s_3 and one instance of reconstruction \bar{s}_3 , both scaled to have norm 1.

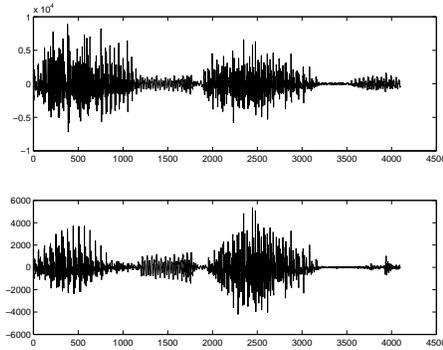


Fig. 5. Original measurements.

detect one of the quotients, which appears as an inflection point at the right of largest maxima.

We may think that a diminishing quality of the reconstructions is unavoidable as the number of sources increases, after all we are dealing with more and more underdetermined systems of vector equations. On the other hand we do believe that at least the attempt of reconstructing 4 speech sources from 2 measurements is within the reach of our method. In Figure 7 we show the position of true ('o' curves) and computed ('*' curves) quotients as we change a particular unidimensional family of exploratory matrices for a specific fixed choice of mixing matrix. We assume that the angular distance between coefficients of the sources is larger than $\frac{\pi}{16}$ (as we enforced in Examples 1 and 2). The family of exploratory matrices has the form: $R_\beta = \begin{bmatrix} \sin(\beta + \frac{\pi}{100}) & \cos(\beta + \frac{\pi}{100}) \\ \sin(\beta) & \cos(\beta) \end{bmatrix}$

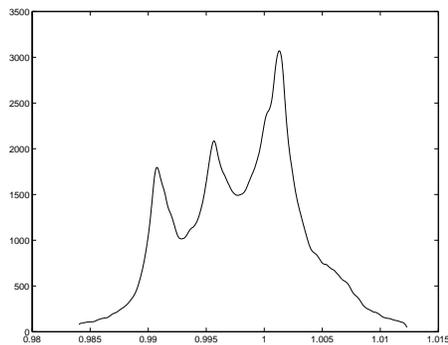


Fig. 6. Generic smoothed histogram from step (b') as described in the text with four sources, *but only three detected maxima*.

where β goes from 0 to π . That is R_β is made of two vectors that are much closer to each other than any two vectors (a_n, b_n) .

Note that each of the true quotient curves γ_n is defined exactly modulo π and that it has a very regular monotone shape, with divergence corresponding to the value of β that makes $(\sin(\beta), \cos(\beta))$ perpendicular to (a_n, b_n) . Also note that there is at least one value of R_β for which the computed quotient curve $\bar{\gamma}_n$ is locally well approximating the corresponding γ_n . While *there is no value of R_β for which all curves are locally approximated*. This observation is the key of a refinement of algorithm (A1)-(A2) that we sketch in (B1)-(B4).

(B1) Apply (a) once with the windowed Fourier frame as choice of dictionary \mathcal{D} .

(B2) Apply (b') for the entire family of exploratory matrices R_β . For each computed significant quotient q , let $\bar{\gamma}_n(\beta, q) = H_R(q)$ be the intensity of the histogram at the quotient itself.

(B3) For each curve $\bar{\gamma}_n$, find the value of $\beta = \beta_n$ that makes $\bar{\gamma}_n(\beta_n, q)$ maximum.

(B4) For each n , choose $R = R_{\beta_n}$ and apply (a'), (c)-(d), (e') and (f) a fixed number ($T = 60$) only to the estimation of source \bar{s}_n . Alternate in (a') a selection of cosine and wavelet packets best basis.

Preliminary results show that (B1)-(B4) is effective in extracting 4 speech sources from 2 measurements most of the time, we will come back in a subsequent paper on a complete implementation and analysis of this 'moving'

quotient algorithm.

Data files of the sources, mixtures and reconstructions computed in examples 1 and 2 are available upon request for direct evaluation of their perceptual quality.

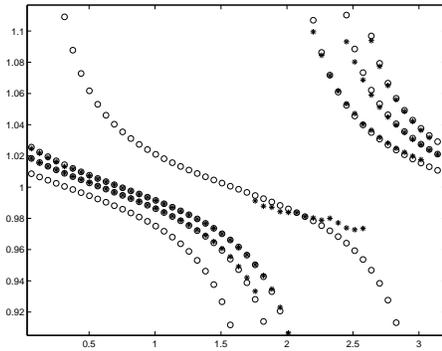


Fig. 7. Curves of true ('o' curves) and computed ('*' curves) quotients as the parameter β goes from 0 to π .

Acknowledgements

This research was supported in part by the Office of Naval Research under the ODDR&E MURI97 Program Grant No. N000149710501EE to the Center for Auditory and Acoustics Research and by the Army Research Office under the ODDR&E 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 Marco Panza for valuable and stimulating discussions. We are also grateful to the referee for constructive and valuable suggestions.

References

- BR] R. Balan, J. Rosca, Statistical properties of STFT ratios for two channel systems and applications to BSS. Proceedings ICA2000, 19-22 June 2000, Helsinki, Finland.
- [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.
- [M] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 1998.
- [NBK] D. Napoletani, C. A. Berenstein, P. S. Krishnaprasad, Quotient Signal Decomposition and Order Estimation, Technical report N. 2002-47 available at <http://techreports.isr.umd.edu/ARCHIVE/> , 2002.
- [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-Frequency 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. *2002 IEEE International Conference on Acoustics, Speech and Signal Processing*, Volume: 1 , 2002 pp. 529-532.
- [S] T. Strohmer, Numerical Algorithms for Discrete Gabor Expansions, in *Gabor Analysis and Algorithms. Theory and Applications*, H. G. Feichtinger, T. Strohmer editors. Birkhauser, 1998.