# A Projection Pursuit Approach To Blind Source Separation

*Fernando A. L. Moreto*

University of São Paulo
Signal Processing Laboratory – PSI - EPUSP
Av. Luciano Gualberto 158, tr.3, CEP 05508900
Butantã SP – São Paulo Brazil
fmoreto@lps.usp.br

*Miguel A. Ramírez*

University of São Paulo
Signal Processing Laboratory – PSI - EPUSP
Av. Luciano Gualberto 158, tr.3, CEP 05508900
Butantã SP – São Paulo Brazil
miguel@lps.usp.br

***Abstract*— The aim of this paper is to show that projection pursuit performed by searching non-Gaussian projections, can be used to separate signals and this is the same thing we do to estimate the ICA (Independent Component Analysis) model. We present a brief overview of ICA and projection pursuit. Finally we analyze a gradient ascent algorithm to optimize the projection index function based on kurtosis to perform the source separation, and present results and conclusions.**

***Index Terms*—Blind Source Separation, BSS, Higher-Order, Independent component analysis, ICA, Projection Pursuit, Signal Separation.**

## I. INTRODUCTION

Independent component Analysis (ICA) is a widely used method for Blind Source Separation (BSS), which is only based on the assumption of non-gaussianity and statistical independence of sources. ICA is an important tool when we have a signal generated by the combination of two or more source signals. Nowadays we can find a lot of applications in a wide range of areas such as bioengineering, telecommunications and financial applications. A description of ICA model is presented in section II.

Projection pursuit is a statistical technique developed to investigate data properties using low dimension projections that provide most revealing views of a high dimensional data [3], [4]. Projection pursuit was primary developed for clustering analysis, but these projections could be used for probability density estimation and regression analysis. Usually data structuring observed in the full dimension will be observable in a lower dimensional projection and each projection can provide additional insight. How this can be computed and a connection between projection pursuit and ICA can found in section III.

In section IV we present the results of experiments applying this method in a set of 2 sound sources mixed by a random matrix for extract the components (sources).

Finally we present an extension of this work in section V followed by a conclusion in section VI.

## II. INDEPENDENT COMPONENT ANALYSIS

Independent component analysis is an important tool when we have a signal generated by the combination of two or more source signals. For example, at a listening point in a room we can probe a signal which is a mixture of the audio signals generated by two people talking. This is the well-known cocktail-party problem. Independent component analysis was originally developed to deal with problems that are closely related to the cocktail-party problem. However nowadays we can find a lot of other interesting applications [7] such as: feature extraction, brain imaging applications and telecommunication applications. A brief historical overview showing the evolution of ICA technique can be found in [1].

ICA is a widely used method for BSS, which is so called because we haven't any previous knowledge of the number of sources and how they are mixed. It is only based on the assumption of non-gaussianity and statistical independence of sources. The ICA linear model can be defined by

$$\boldsymbol{x} = \mathbf{A}\boldsymbol{s} \tag{1}$$

where $\boldsymbol{x}$ is a vector of mixed signal, $\mathbf{A}$ is an unknown mixing matrix and $\boldsymbol{s}$ is an unknown vector of source signals.

For the sake of simplicity we will assume that our signals are zero mean and unity variance. After estimating $\mathbf{A}$, given mixed vector $\boldsymbol{x}$, one can compute the source by inverting $\mathbf{A}$ as shown in equation (2), assuming that $\mathbf{A}$ is invertible.

$$\boldsymbol{s} = \mathbf{A}^{-1}\boldsymbol{x} \tag{2}$$

In practice we estimate $\mathbf{A}^{-1}$ directly, called here $\mathbf{W}$. Therefore we can write the following equation $\boldsymbol{y} = \boldsymbol{w}^T\boldsymbol{x}$, where $\boldsymbol{y}$ is an estimated source, as similar as possible to $\boldsymbol{s}$ and $\boldsymbol{w}$ is a weighting vector to be determined that must be close to a row vector of the inverse of $\mathbf{A}$. In other words, the estimated source is a linear combination of the mixed signals.

The process of estimation is straightforward. As we do not have any previous knowledge about matrix $\boldsymbol{A}$, we cannot determine the exact value of $\boldsymbol{w}$, but we can find a good estimator. Let us make the transformation $\boldsymbol{z} = \mathbf{A}^T\mathbf{w}$. Having $\boldsymbol{y} = \boldsymbol{w}^T\boldsymbol{x}$, getting equation. (1) and considering only 2 sources without loss of generality we can easily write

$\boldsymbol{y} = z_1 s_1 + z_2 s_2$. Based on properties of kurtosis, see appendix A, we obtain $K(y) = z_1^4 K(s_1) + z_2^4 K(s_2)$. As said before $\boldsymbol{y}$ has been constrained to be unit variance then $\mathrm{E}\{y^2\} = z_1^2 + z_2^2 = 1$. Geometrically this means that $\boldsymbol{z}$ is constrained in the unit circle [7].

The optimization problem now will be to find maxima on the unit circle for the function represented in equation (3).

$$\left|K(y)\right| = \left|z_1^4 K(s_1) + z_2^4 K(s_2)\right| \quad (3)$$

Therefore this happens when one of the elements of $\boldsymbol{z}$ is different from zero and all the others are zero, so maximum kurtosis is exactly when $\boldsymbol{y}_i = \boldsymbol{s}_i$.

An algorithm is proposed in [1] to search for a linear transformation that minimizes the statistical dependence between its components by the expansion of mutual information as a function of cumulants of increasing orders.

ICA weights can be computed by a number of methods with differing mixes of advantage and disadvantage over each other methods. The present paper is based on the kurtosis exclusively because of its computational simplicity. In Section V we present an extension that can be applied in this work to obtain a algorithm that can be more robust than the present one based on kurtosis respect to outliers.

## III. Projection Pursuit

Projection pursuit is a statistical technique developed to investigate data properties using low dimension projections that provide most revealing views of a high dimensional data [3], [4]. These projections could be used for purpose of density estimation and regression analysis. Usually data structuring observed in the full dimension will be observable in a lower dimensional projection and each projection can provide additional insight.

It's not practical to map every possible projection so Friedman and Tukey (1974) propose an algorithm whose basic idea is to assign a numerical index, in 1D or 2D, to every projection characterizing the amount of structure present on it for that purpose. This index is then maximized related to the parameters defining the projections [4]. E.g. Principal component analysis can be regarded as a projection pursuit in which the interestingness is the total variation accounted for by the projection of data [8]. Friedman and Tukey developed an index for projection pursuit that emphasizes the clustering structure of data for exploratory data analysis.

It's well known that any linear combination $X = \boldsymbol{a}^T Z$ has unit variance if and only if $\boldsymbol{a}^T \boldsymbol{a} = 1$ and two linear combinations of orthogonal vectors are uncorrelated, thus let $Y$ be a random variable in p-dimensional space $\mathbb{R}^p$ if we "sphere" (remove all of the location, scale, and correlation structure) $Y$, called here by $Z$, performing an eigenvalue-eigenvector decomposition and assuring the constraint $\boldsymbol{a}^T \boldsymbol{a} = 1$ (by definition $Z$ variables are affine invariant), so the projection index based on them will be affine invariant. This avoids calculating variances in each projection, saving computational effort.

Let $X$ in $\mathbb{R}^d$ be a projection of $Z$ in $\mathbb{R}^p$ and A a matrix $d \times p$, therefore a linear projection from $\mathbb{R}^p$ to $\mathbb{R}^d$ can be defined as shown in equation (7).

$$X = \mathrm{A}\mathrm{Z} \quad (4)$$

Now in a one-dimension exploratory projection pursuit, by definition, we seek a vector row $a^T$ witch maximizing a certain projection index such that $p_a(X)$ is relatively highly structured [4].

The deep analysis of projection pursuit and projection index is out of scope of this paper. For more information see [3], [4] and [5].

It's been shown that the most interesting directions are those that are as non-Gaussian as possible [5], [8]. The density function that is most unpredictable, has maximum entropy [2], is the Gaussian density, so based on it, projections in the direction of least Gaussian distribution are desirable and this is exactly what is necessary to estimate ICA models. In other words we see that independent components can be found by finding several directions of maximum non-gaussianity using a measure of normality, e.g. kurtosis, so this approach is closely connected to projection pursuit, in which maximum non-Gaussian directions are considered interesting for visualization purpose and exploratory data analysis [3], [4], [7], [8].

## IV. Experiments

In this section we describe two experiments. In the first one we extract one source and in the second we explain how to extract more than one signal from the mixtures.

For practical reasons we assume that source signals are super-Gaussian and we also preprocess the set of mixing signals $\boldsymbol{x}$ to be centered by subtracting their means and by whitening them by singular value decomposition as mentioned before, so the yield components will be uncorrelated with unity variance [1], [7].

Now let $\boldsymbol{y} = \boldsymbol{w}^T \boldsymbol{z}$ be an extracted signal from a set of M transformed mixtures $\boldsymbol{z}$ by the weight vector $\boldsymbol{w}$, we must observe that rotating $\boldsymbol{w}$ around the origin the kurtosis of extracted signal is maximum exactly when $\boldsymbol{y} = \boldsymbol{s}$ and $\boldsymbol{w}$ is orthogonal to the projected axes[7]. So it's easy to derive a gradient ascent algorithm to find a $\boldsymbol{w}$ that maximizes the kurtosis of extracted signal $\boldsymbol{y}$ as follow.

The kurtosis of $\boldsymbol{y}$ can be seeing in equation (5) and the kurtosis gradient in equation (6). Note that the gradient changes both the length and angle of $\boldsymbol{w}$, whatever only the angle is important here because the length do not affect the form of extracted signal only the amplitude. We restricted the length of $\boldsymbol{w}$ to unity and shown the gradient algorithm in table 1.

$$K(\boldsymbol{y}) = K(\boldsymbol{w}^T \boldsymbol{z}) = E\{(\boldsymbol{w}^T \boldsymbol{z})^4\} - 3 \quad (5)$$

$$\frac{\partial K(\boldsymbol{w}^T \boldsymbol{z})}{\partial \boldsymbol{w}} \propto E\{\boldsymbol{z}(\boldsymbol{w}^T \boldsymbol{z})^3\} \quad (6)$$

To demonstrate this we got M=2 sources signals, a female speech signal and a music signal, and mixture them using a random mixture MxM matrix, then we apply the procedure above to separate one source from the set of mixture signals and presented the results in figure 1 and figure 2. One must see that this result is rather satisfactory, although the algorithm presented here can be improved especially because it is sensible to outliers as it is based on kurtosis. Read section V for more details.

TABLE1: PROJECTION PURSUIT GRADIENT ASCENT ALGORITHM.

1. Choose an initial vector $\boldsymbol{w}$ and a an initial value for $\mu$
2. $\boldsymbol{w}_{n+1} = \boldsymbol{w}_n + \mu \cdot E\{\boldsymbol{z}(\boldsymbol{w}_n^T \boldsymbol{z})^3\}$
3. $\boldsymbol{w}_{n+1} = \dfrac{\boldsymbol{w}_{n+1}}{|\boldsymbol{w}_{n+1}|}$
4. $\boldsymbol{w}_n = \boldsymbol{w}_{n+1}$
5. If convergence is not attained, go back to step 2.

Note that it extracts one source at time in contrast to conventional ICA methods that extract M sources at once. For separate multiple sources it's necessary to remove each recovered source from the set of remaining signal mixtures using the Gram-Schmidt Orthogonalization (GSO) and apply the procedure above again repeating this up to recovery of the last source.

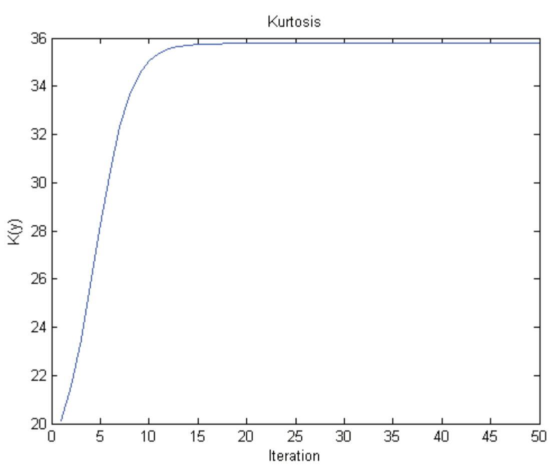


Figure 1: Kurtosis of extracted signal $y_1$ in each iteration.

GSO ensures that each extracted signal $y_i$ is orthogonal to every mixture of signal yet to be extracted. Let the original set of mixtures be $\boldsymbol{x}^0 = (x_1^0, \cdots, x_M^0)$ then a weight vector $\boldsymbol{w}_1$ is obtained which extracts a signal $y_1 = \boldsymbol{w}_1^T \boldsymbol{x}^0$. Where the number superscripted, 0, denote the original set of mixtures. Now we can effectively subtract $y_1$ from each signal mixture $x_i^0$ as follow,

$$x_i^1 = x_i^0 - \frac{\mathrm{E}\{y_1 x_i^0\} y_1}{\mathrm{E}\{y_1^2\}} \quad (7)$$

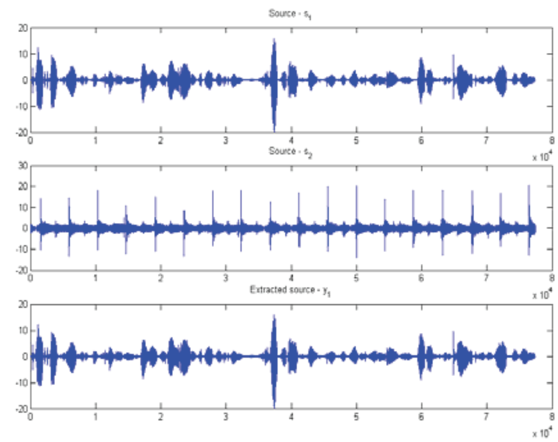


Figure 2: The graph at the top is the original source 1 and the middle is the source 2, finally the button graph is the first estimated extracted signal.

So as mention before $y_1$ is orthogonal to every mixture of signal $x_i^1$, in other words $\mathrm{E}\{x_i^1 y_1\} = 0$ for $i = \{1, \cdots, M\}$. If the projection pursuit is applied again to the modified set of mixtures $\boldsymbol{x}^1$ then the extracted signal $y_2$ can be subtracted from each mixture $x_i^1$ as similar manner as equation (7).

This procedure can be repeated until the last signal to be recovered. In the noise-free case we can stop when nothing to be extracted remains. Figure 4 show extracted signal $y_2$ and figure 3 the kurtosis of $y_2$ applying this procedure above.

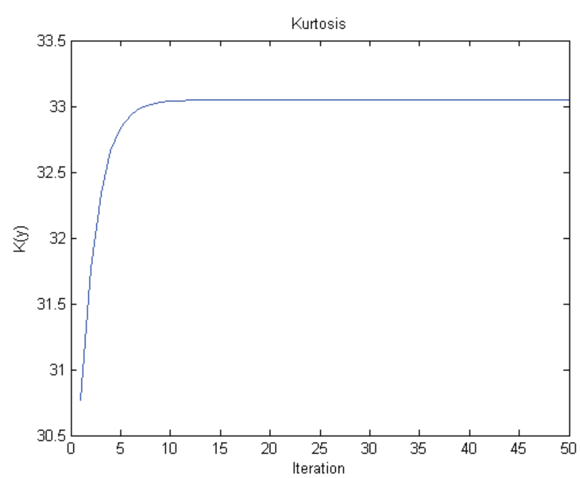


Figure 3: Kurtosis of extracted signal $y_2$ in each iteration.

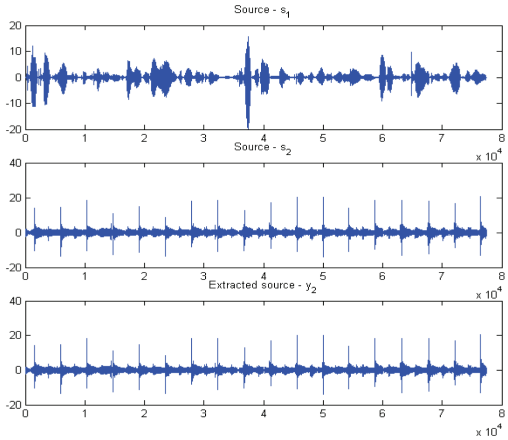


Figure 4: The graph at the top is the original source 1 and the middle is the source 2, finally the button graph is the $y_2$ extracted signal.

## V. EXTENSION WORKS

Another important measure of non-Gaussianity that can be used instead of kurtosis is given by negentropy, which is based on the information-theoretic quantily of differential entropy.

Entropy can be interpreted as the degree of information observed in a given random variable [6]. For a discrete random variable the entropy *H* is defined as

$$H(y) = -\sum_i P(y = a_i) \log P(y = a_i) \tag{8}$$

where $a_i$ are the possible values of *y*. [2] show the proof of the following fundamental result of information theory: “Gaussian variable has the largest entropy among all random variables of equal variance”.

To obtain a measure of non-Gaussianity that is non-negative for non-Gaussian variables and zero for Gaussian variables we can define Negentropy *J* as

$$J(\boldsymbol{y}) = H(\boldsymbol{y}_{gauss}) - H(\boldsymbol{y}) \tag{9}$$

where $\boldsymbol{y}_{gauss}$ is a Gaussian variable with the same covariance matrix as $\boldsymbol{y}$. As we can see to perform entropy calculations we may need to estimate the probability density function of y

We can use $J(\boldsymbol{y})$ instead of $K(\boldsymbol{y})$ with the advantage of $J(\boldsymbol{y})$ be more robust than $K(\boldsymbol{y})$. Unfortunately it’s computationally more complex and some approximations are necessary. There are several approximations that can be used to estimate the negentropy: [8] use higher-order moments, which have the same problem of kurtosis method, [1] use a method based in cumulants and [6] show an interesting method based on non-linear function that can be much better than the one given by higher-order moments.

From these results we can derive an algorithm for projection pursuit using negentropy as an objective non-Gaussian measure for optimization.

## VI. CONCLUDING REMARKS

The procedure described in this paper decomposes the mixed signals in much the same way as is done by ICA methods, considering the conditions of non-normality and independency. In fact, projection pursuit doesn’t consider any ICA model; rather, it only searches for a weight vector ***w*** such that the extracted signal is as non-Gaussian as possible. So, if the ICA model fails, the result is only the projection pursuit directions, resulting extracted signals which will differ significantly from those extracted by several ICA methods.

## VII. APPENDIX

### A. Kurtosis

The k-th moment $E\{x^k\}$ of the pdf of a random variable $x$ is defined as:

$$\mathrm{E}\{x^k\} = \int_{-\infty}^{\infty} f_x(x) x^k dx \tag{11}$$

Kurtosis is defined as the normalized version of the fourth central moment of a zero-mean random variable $x$ as follow:

$$K(x) = \frac{\mathrm{E}\{x^4\}}{\mathrm{E}\{x^2\}^2} - 3 \tag{12}$$

Kurtosis is zero for normal distributions and, in practice, it is nonzero for non-normal distributions, therefore kurtosis can be used as a measure of normality. Kurtosis can be either positive (Super-Gaussian) or negative (Sub-Gaussian), also know as platykurtic and leptykurtic respectively.

The main reason for using kurtosis as measure of non-normality is its computational and theoretical simplicity. Kurtosis can be estimated simply by using the fourth moment of sample data and theoretical analysis of kurtosis can be simplified because of the following properties: If we have two random variables, $x_1$ and $x_2$, statistically independent then holds the properties below:

$$K(x_1 + x_2) = K(x_1) + K(x_2) \tag{13}$$

and

$$K(\mathrm{a}x) = \mathrm{a}^4 K(x) \tag{14}$$

where a in equation (14) is constant.

## VIII. ACKNOWLEDGMENTS

This work is supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) under Grant no. 04/13206-7.

## REFERENCES

[1] Comon, P., “Independent Component Analysis – a new concept?” Signal Processing, 36:287-314, 1994

[2] Cover, T.M. and Thomas, J. A. “Elements of Information Theory” John Wiley and Sons, New York, 1991
[3] Friedman, J. H. and Tukey, J.W., “A Projection Pursuit Algorithm for exploratory data analysis”, IEEE Transactions on Computers, c-23, PP. 881-890, 1974
[4] Friedman, J. H., “Exploratory Projection Pursuit”, Journal of American Statistical Association, Vol. 82, No. 397, PP 249-266, 1987
[5] Huber, P., “Projection Pursuit”, The Annals of Statistics, 13(2), 435-475, 1985
[6] Hyvärinen, A., Karhunen, J. and Oja, E., “Independent Component Analysis: algorithms and applications”, Neural Networks, 13, PP 411-430, Helsinki, 2000
[7] Hyvärinen, A., Karhunen, J. and Oja, E., “Independent Component Analysis”, John Wiley and Sons, New York, 2001
[8] Jones, M.C. and Sibson, R., “What is Projection Pursuit”, Journal of Royal Statistical Society, Vol. 150, No.1, PP.1-37, 1987
[9] Papoulis, A., “Probability, Random Variables, and Stochastic Processes” McGraw-Hill, Tokio, 1965