Smooth signal extraction from instantaneous mixtures

Nikolaos Mitianoudis, Tania Stathaki and Anthony G. Constantinides

Abstract

The problem of blind separation of statistically independent sources from instantaneous mixtures, using the efficient framework of Independent Component Analysis (ICA), has been widely addressed in the literature. In this paper, the authors propose a sequential blind signal extraction algorithm that attempts to identify smooth sources in instantaneous mixtures. The approach incorporates smoothness constraints in the traditional negentropy cost function to extract smooth components, using an approximate second-order optimisation method.

Index Terms

Independent Component Analysis, Constrained ICA, FastICA, smoothness constraint.

EDICS Category: SAS-ICAB

I. INTRODUCTION

Let $\underline{x}(n) = [x_1(n), x_2(n), \dots, x_M(n)]^T$ model M observation signals that monitor a phenomenon and also $\underline{s}(n) = [s_1(n), s_2(n), \dots, s_N(n)]^T$ model N factors that influence the monitored phenomenon. In this study, each factor influences the observing signals instantaneously and possible corruption by additive noise is considered insignificant. The following model connects the observed with the source signals.

$$\underline{x}(n) = A\underline{s}(n) \tag{1}$$

where A is a full-rank mixing matrix denoting instantaneous transmission. The mixing matrix will be considered square (i.e. $N = M$) in our analysis.

Manuscript received 15 June, 2006; revised

The authors are with the Electrical and Electronic Engineering Department, Imperial College London, Exhibition Road, London SW7 2AZ, UK (e-mail: n.mitianoudis@imperial.ac.uk, tel: +44 (0)207 594 6229, fax: +44 (0)207 594 6234).

The source separation problem is to estimate the underlying factors $\underline{s}(n)$, given the observation signals and some general statistical priors for these factors. Introducing the assumption of statistical independence between the source signals, led to the development of *Independent Component Analysis* (ICA) [4]. One can separate nonGaussian sources, using different interpretations of statistical independence. Some methods interpret statistical independence as maxima of nonGaussianity and perform separation by estimating the directions of the most nonGaussian components using *kurtosis* or *negentropy*.

In several applications, it is essential to identify signals with specific temporal properties along with independence. *Smoothness* can be imposed as a desired signal property, especially when identifying components in an environment with interference. In ECG and EEG signals, the electrodes pick up several interfering signals that might satisfy the nonGaussianity criteria of ICA algorithms, however, may not convey useful information. Instead, components that have some "smooth" temporal structure (slowly varying profile) may need to be identified in the mixture.

The identification of signals with special properties can always be performed as a post-processing step after a conventional source separation algorithm. However, adding a constraint to the estimation criterion can identify the desired components with some additional computational overhead, but without separating and testing all present independent components, especially when N is large. The concept of constrained ICA has been previously introduced in [6], however, in the concept of global search for independent components that resembles several reference signals. In [1], the extraction of components with temporal periodic structure is introduced assuming a priori knowledge of the sources' autocorrelation function.

In this paper, we derive a negentropy-based deflation signal extraction algorithm [5] with smoothness constraints. The extra smoothness constraint is incorporated in the original negentropy cost function, using Lagrange multipliers, in order to identify smooth components. The proposed approach derives an approximate Newton-step algorithm, following the approximation and derivation of FastICA [5].

II. Source separation of smooth signals

A. *Definition of smoothness*

In mathematics, a function is defined as *smooth* when it is infinitely (indefinitely) differentiable, i.e., has derivatives of all finite orders. Here, the definition of smoothness is extended to describe signals that feature a slowly-varying temporal structure. This can also be described by absense of high-frequency content in the Fourier transform domain. In order to avoid possible processing of complex numbers and the additional computational complexity of a transformation, temporal criteria are defined to describe smoothness.

The consecutive samples of a slowly varying ("smooth") signal should not be much different on average. Equivalently, the rate of change (usually described by the first derivative) should be relatively small. Consequently, a signal $u(n)$ can be considered smooth, if the following condition holds:

$$\mathcal{E}\{(u(n) - u(n-1))^2\} \leq \rho\mathcal{E}\{u^2(n)\} \tag{2}$$

where $0 < \rho < 1$ is empirically set. This definition simply states that the squared difference between two successive samples can not exceed a portion ρ of the average squared amplitude of the signal. Similarly, the average rate of change (first derivative) in square should not exceed a percentage of the signal's energy (variance, in the case of zero-mean signals).

However, this definition may not necessarily be an adequate criterion to identify smooth components in a linear mixture. If there are many smooth components, any linear combination of them may also satisfy the smoothness condition, as well as the individual components. Therefore, the extra assumption of independence is needed, so as to identify only the original smooth components.

B. Principal Component Analysis

A Principal Component Analysis (PCA) step will orthogonalise (decorrelate) and normalise the data to unit variance [4]. Assuming zero-mean signals, the prewhitening matrix V is formed by the eigenvectors of the covariance matrix $C = \mathcal{E}\{\underline{x}\underline{x}^T\}$ and scaled by the inverse square root of their corresponding eigenvalues. The "whitened" data $\underline{z}$ are given by:

$$\underline{z} = V\underline{x} \tag{3}$$

where $\mathcal{E}\{\underline{z}\underline{z}^T\} = I$. Decorrelation is not sufficient to separate independent signals [4]. In order to extract L ($L \leq N$) smooth components u_i from the mixture, we need to estimate L projection operators $\underline{w}_i$:

$$u_i = \underline{w}_i^T \underline{z} \quad \forall\ i = 1, \ldots, L \tag{4}$$

C. Smooth signal extraction

The next step is to formulate a cost function $J(\underline{w})$ that will be able to identify smooth orthogonal projections. The main criterion for the separation will be nonGaussianity, expressed in terms of negentropy. The smoothness criterion will be added as a constraint to the optimisation problem. The nonGaussianity cost function is defined [5]:

$$J_1(\underline{w}) \propto |\mathcal{E}\{G(u)\} - \mathcal{E}\{G(v)\}| \tag{5}$$

where G is a non-quadratic function and v is a zero-mean unit-norm Gaussian variable. The smoothness criterion is defined by the following cost function:

$$J_2(\underline{w}) = \mathcal{E}\{(u(n) - u(n-1))^2\} - \rho\mathcal{E}\{u^2(n)\} \tag{6}$$

Defining $\underline{\Delta z} = \underline{z}(n) - \underline{z}(n-1)$, the above equation is written as follows:

$$J_2(\underline{w}) = \mathcal{E}\{(\underline{w}^T \underline{\Delta z})^2\} - \rho\mathcal{E}\{(\underline{w}^T \underline{z})^2\} \tag{7}$$

In addition, we have to ensure that we are only performing rotation and not any scaling deformation. Therefore, we impose the following unit-norm constraint to the estimated components u.

$$J_3(\underline{w}) = ||u||^2 - 1 = \underline{w}^T \underline{w} - 1 \tag{8}$$

Consequently, the inequality constrained optimisation problem is the following:

$$\max_{w} J_1(\underline{w}) \tag{9}$$

$$\text{subject to} \quad J_2(\underline{w}) \leq 0 \tag{10}$$

$$J_3(\underline{w}) = 0 \tag{11}$$

The inequality constraint in (10) can be replaced with the equality constraint $\max(J_2(\underline{w}), 0) = 0$, as introduced for solving Zero Tolerance Problems (ZTP) [2]. To solve this equality constrained optimisation problem, the method of Lagrange Multipliers is employed. The objective is to formulate an approximate Newton-type method, following a derivation similar to the original FastICA algorithm [5] and [6]. As traditionally performed by these methods, the unit-norm constraint (11) is enforced by projection of the estimated $\underline{w}$ on the unit-sphere in each iteration (12), therefore, it is not considered in the optimisation cost function.

$$\underline{w}^+ \leftarrow \underline{w}^+ / ||\underline{w}^+|| \tag{12}$$

The constrained optimisation problem is transformed to an unconstrained maximisation problem, using the Lagrangian function:

$$J(\underline{w}, \lambda) = J_1(\underline{w}) + \lambda \max(J_2(\underline{w}), 0) \tag{13}$$

The above optimisation problem is addressed using alternate optimisation. Hence, estimates for $\underline{w}$ and λ are updated in an alternating manner. That is, given the current estimate for λ, a new estimate for $\underline{w}$ is calculated and given the estimate for $\underline{w}$, we update λ. Following the strategy proposed in [8], we perform gradient ascent to update $\underline{w}$ and gradient descent to estimate λ.

$$\underline{w}^+ \leftarrow \underline{w} + \eta_1 \frac{\partial J(\underline{w})}{\partial \underline{w}} \tag{14}$$

$$\lambda^+ \leftarrow \lambda - \eta_2 \max(J_2(\underline{w}), 0) \tag{15}$$

where η_1, η_2 are the corresponding learning rates and $\underline{w}^+$ denotes the new estimate of $\underline{w}$. To accelerate the estimation of $\underline{w}$, we can use an approximate Newton step to replace the gradient ascent optimisation method. A Newton-step is given by the following update:

$$\underline{w}^+ \leftarrow \underline{w} - \left[\frac{\partial^2 J}{\partial \underline{w}^2}\right]^{-1} \frac{\partial J}{\partial \underline{w}} \tag{16}$$

where, in this case, the gradient vector and the Hessian matrix are estimated, using the following updates (see Appendix):

$$\begin{aligned} \frac{\partial J}{\partial \underline{w}} &= \mu \mathcal{E}\{\underline{z} G'(\underline{w}^T \underline{z})\} \\ &+ \lambda(\mathcal{E}\{(\underline{w}^T \underline{\Delta z})\underline{\Delta z} - \rho(\underline{w}^T \underline{z})\underline{z}\})(\text{sgn}(J_2) + 1) \end{aligned} \tag{17}$$

$$\begin{aligned} \frac{\partial^2 J}{\partial \underline{w}^2} &= \mu \mathcal{E}\{G''(\underline{w}^T \underline{z})\} I \\ &+ \lambda(C_{\Delta z} - \rho I)(\text{sgn}(J_2) + 1) \end{aligned} \tag{18}$$

where $\mu = \text{sgn}(\mathcal{E}\{G(u)\} - \mathcal{E}\{G(v)\})$. After calculating the estimate for $\underline{w}$, we calculate estimates for the Lagrange multiplier λ via (15) and then normalise the unmixing vector to unit-norm via (12).

The same procedure can be used to extract other smooth components that exist in the linear mixture. The above rule is randomly re-initialised but should not converge to the same component. As all solutions lie in an orthonormal structure, due to prewhitening, the new components should always be orthogonal to the already estimated components [5]. Hence, the update for the $i-th$ component $\underline{w}_i^+$ should always be orthogonal to the space spanned by the vectors $\underline{w}_1, \underline{w}_2, \dots, \underline{w}_{i-1}$.

$$\underline{w}_i^+ \leftarrow \underline{w}_i^+ - BB^T \underline{w}_i^+ \tag{19}$$

where $B = [\underline{w}_1 \ \underline{w}_2 \dots \underline{w}_{i-1}]$.

The proposed method resembles Constrained ICA, proposed in [6]. The difference is that the inequality constraint is treated as a ZTP problem that results into a simpler approach and seems to be stable. In addition, the proposed constraint aims at separating "smooth" components, instead of components that follow some reference signals. Moreover, the proposed smoothness constraint can also be formulated as one-lag autocorrelation and therefore one can find connections with second-order methods that exploit time-structure, such as AMUSE [7]. Nonetheless, the proposed combination of negentropy and second-order information generally avoids several shortcomings of second-order methods, such as susceptibility to Gaussian noise and non-distinct eigenvectors of the lag-covariance matrix for certain time-lags [4].

III. Experiments

To evaluate the performance of the proposed algorithm, we create an artificial mixture of four signals, using the following random mixing matrix A. We used two Laplacian noise signals and two "smooth" signals, a sinusoidal signal of normalised frequency 0.04π rad/sample and an exponentially decaying saw-tooth signal (Fig. 1), providing a valid example of slowly varying temporal profile, compared to the Laplacian noise signals. The exponential saw-tooth signal demonstrates that the source signals may not necessarily be periodic. The dataset consisted of 2000 samples of each source.

$$A = \begin{bmatrix} 2.0128 & 0.7539 & 1.1149 & 0.5482 \\ -0.5079 & -0.3650 & 1.5789 & -1.4763 \\ 0.1326 & 0.9526 & 2.0719 & -0.5666 \\ 0.9986 & -0.4279 & -0.2595 & 1.1143 \end{bmatrix} \tag{20}$$

The proposed algorithm is used to identify and separate the two "smooth" signals from the mixture. We used a random initialisation of $\underline{w}$, a learning rate $\eta_2 = 0.9$, $\mu = 1$ for subGaussian sources, $\rho = 0.02$ and an initial value $\lambda = 60$ for the Lagrange multiplier. In addition, the non-quadratic function $G(\cdot) = \log\cosh(\cdot)$ was used in the adaptation of $\underline{w}$. The algorithm managed to isolate the two desired signals (Fig. 1) with *Signal-to-Noise Ratio* of $36.08dB$ and $35.3dB$ respectively. The convergence performance of the algorithm is depicted in Fig. 2 (top), in terms of $|\underline{w}^T\underline{w}^+|$. The algorithm features similar performance with random selection of the mixing matrix A and several smooth synthetic signals. The performance is similar to the FastICA algorithm, despite the addition of another constraint. The estimation of the Lagrange multiplier λ is slower, due to the gradient update. The algorithm converges even for random initialisations of λ, however, different choices for λ will influence the speed of convergence. The algorithm also converges, in the case that μ is estimated during the adaptation, however, the convergence seems to be less smooth and fast, compared to the previous case.

The choice of ρ can also influence the selection of different smooth components. Some values of ρ may not fulfill the smoothness constraint for several desired components. In the previous synthetic example, we estimated the "smoothness" of the two smooth components $\rho = \mathcal{E}\{\Delta u^2\}/\mathcal{E}\{u^2\}$. We measured

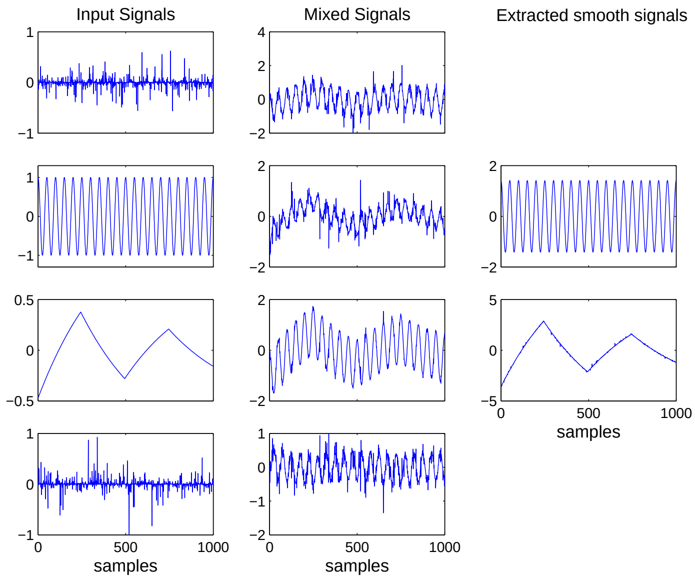

Fig. 1. Four artificial input sources (left), the four mixed signals (middle) mixed using the matrix in (20) and the two "smooth" sources (sine-wave and exponential sawtooth) identified by the proposed algorithm (right).

$\rho_1 = 0.0168$ for the sine wave and $\rho_2 = 0.000265$ for the exponential saw-tooth. Using the same initial values for λ and η_2 as previously, a choice for $\rho > 0.02$ will extract both smooth components. A choice for $\rho < 0.02$ will extract only the exponential saw-tooth, as the sine-wave does not approximately fulfill the smoothness constraint. This shows that selectivity of smooth components is possible, using different values for ρ.

The algorithm is also applied on some real biomedical data. The electrocardiogram (ECG) of a pregnant woman, used in [1] and initially distributed by De Moor [3], is processed by the proposed algorithm. The source separation problem is to separate smooth ECG components, belonging to the mother or the foetus, because they are mixed in the observation signals. The eight input signals are depicted in Fig. 3. The same initial values for λ and η_2 were used, however, ρ was set to 0.4, due to previous smoothness measurements of the independent components. The algorithm managed to identify two components, describing the ventricular activity of the mother and a third component capturing both the ventricular and the atrial activity of the foetus (Fig. 4).

Finally, in the case that the number of requested "smooth" components is greater than their actual number, the algorithm tended to identify as extra "smooth" components either the Laplacian noise sources in the synthetic example (as the negentropy term dominated the constraint) or synthetic mixture components that were not independent in the biomedical example. For more accurate results, the actual

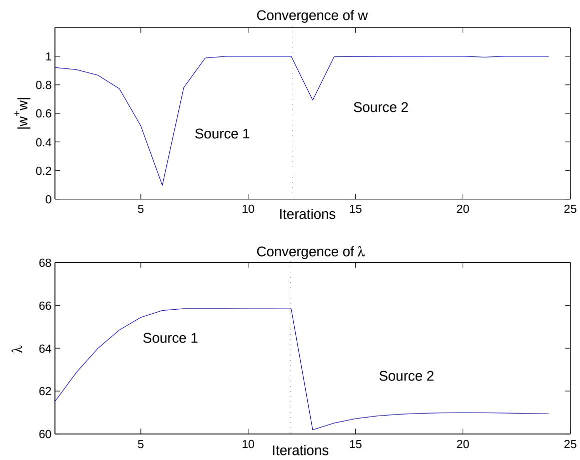

Fig. 2. Convergence of the estimated unmixing vectors $\underline{w}$ (top) and the Lagrange multiplier λ (bottom) for the two sources.

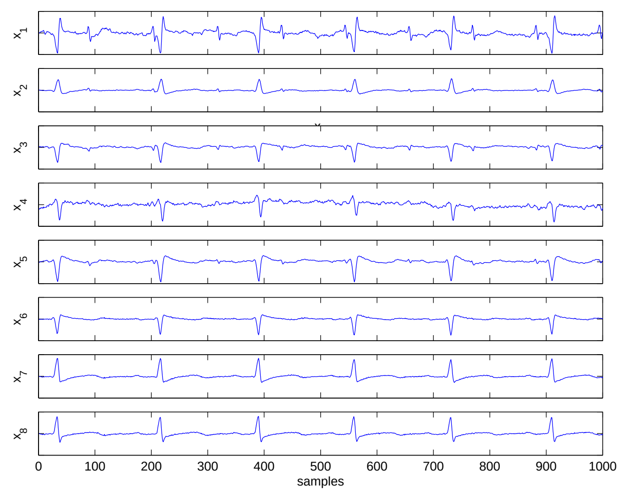

Fig. 3. ECG signals taken by abdominal ($x_1 - x_5$) and thoracic ($x_6 - x_8$) measurements from a pregnant woman [1], [3].

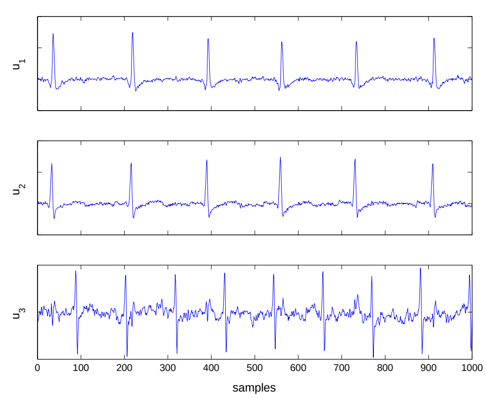

Fig. 4. Three smooth components extracted, the first two describe mainly the ventricular activity of the mother and the third describes the heart activity of the foetus.

number of "smooth" sources should be known.

IV. Conclusion

In this paper, the authors have extended current work in the field of constrained Independent Component Analysis. A novel constraint was imposed on the negentropy cost function, employed in the original FastICA algorithm, in order to identify "smooth" components (components of slowly-varying temporal profile) in instantaneous mixtures. The proposed algorithm imposes the constraint using Lagrange multipliers, leading to an approximate Newton method that features the stability and performance of the original FastICA algorithm, with a small delay in convergence due to the adaptation of the Lagrange multiplier.

Acknowledgments

The authors would like to thank the reviewers for their insightful comments.

Appendix

The first and second-order derivatives of $J_1(\underline{w})$ and $J_2(\underline{w})$ are calculated, as follows:

$$\frac{\partial J_1}{\partial \underline{w}} = \mu\mathcal{E}\{\underline{z}G'(\underline{w}^T\underline{z})\} \tag{21}$$

$$\begin{aligned}\frac{\partial^2 J_1}{\partial \underline{w}^2} &= \mu\mathcal{E}\{G''(\underline{w}^T\underline{z})\underline{z}\underline{z}^T\} \approx \mu\mathcal{E}\{G''(\underline{w}^T\underline{z})\}\mathcal{E}\{\underline{z}\underline{z}^T\} \\ &= \mu\mathcal{E}\{G''(\underline{w}^T\underline{z})\}I \end{aligned} \tag{22}$$

where μ is the sign of the expression $\mathcal{E}\{G(u)\} - \mathcal{E}\{G(v)\}$, which can be either manually set or estimated during the adaptation. The above approximation is viable, due to the prewhitened solution space [5]. The $\max(\cdot)$ function can be expressed, as follows:

$$\max(J_2, 0) = \begin{cases} J_2, & \text{if } J_2 > 0 \\ 0, & \text{if } J_2 \leq 0 \end{cases} \tag{23}$$

Therefore, one can derive the following derivatives:

$$\frac{\partial}{\partial \underline{w}}\max(J_2, 0) = \frac{\partial J_2}{\partial \underline{w}}(\text{sgn}(J_2) + 1)/2 \tag{24}$$

$$\frac{\partial^2}{\partial \underline{w}^2}\max(J_2, 0) = \frac{\partial^2 J_2}{\partial \underline{w}^2}(\text{sgn}(J_2) + 1)/2 \tag{25}$$

where

$$\frac{\partial J_2}{\partial \underline{w}} = 2\mathcal{E}\{(\underline{w}^T\underline{\Delta z})\underline{\Delta z}\} - 2\rho\mathcal{E}\{(\underline{w}^T\underline{z})\underline{z}\} \tag{26}$$

$$\frac{\partial^2 J_2}{\partial \underline{w}^2} = 2\mathcal{E}\{\underline{\Delta z}\underline{\Delta z}^T\} - 2\rho\mathcal{E}\{\underline{z}\underline{z}^T\} = 2C_{\Delta z} - 2\rho I \tag{27}$$

REFERENCES

[1] A.K. Barros, A. Cichocki, *Extraction of specific signals with temporal structure*, Neural Computation, vol. 13, no. 9, pp. 1995-2003, Sept. 2001.

[2] R.K. Brayton, S.W. Director, G.D. Hachtel, L.M. Vidigal, *A New Algorithm for Statistical Circuit Design Based on Quasi-Newton Methods and Function Splitting*, IEEE Trans. on Circuits and Systems, Vol. CAS-26, No. 9, September 1979.

[3] D. DeMoor (Ed.), *Daisy:Database for the identification of systems*, http://www.esat.kuleuven.ac.be/sista/daisy.

[4] A. Hyvärinen, J. Karhunen, E. Oja, *Independent Component Analysis*, John Wiley & Sons, 2001.

[5] A. Hyvärinen, *Fast and robust fixed-point algorithms for independent component analysis*, IEEE Trans. Neural Networks, vol. 10, no. 3, pp. 626-634, May 1999.

[6] W. Lu, J.C. Rajapakse, *Approach and applications of constrained ICA*, IEEE Trans. Neural Networks, Vol. 16, No. 1, January 2005, pp. 203-212.

[7] L. Tong, V.C. Soon, R. Liu and Y. Huang, *AMUSE: a new blind identification algorithm*, Proc. ISCAS, New Orleans, 1990.

[8] S. Zhang, A.G. Constantinides, *Lagrange Programming Neural Networks*, IEEE Trans. Circuits and Systems-II: Analog and Digital Signal Processing, Vol. 39, No. 7, 1992.