

BLIND EXTRACTION OF SPARSE IMAGES FROM UNDER-DETERMINED MIXTURES

Włodzimierz Kasprzak¹, Andrzej Cichocki², Adam F. Okazaki¹

¹*Warsaw University of Technology, Inst. of Control and Computation Eng., ul. Nowowiejska 15/19, PL-00-665 Warsaw*

W.Kasprzak@ia.pw.edu.pl, A.Okazaki@elka.pw.edu.pl

²*Brain Science Institute RIKEN, Lab. for Advanced Brain Signal Processing, Hirosawa 2-1, Wako-shi, 351-0198 Saitama, JAPAN*

cia@brain.riken.go.jp

Abstract We propose a blind signal extraction approach to the extraction of binary and sparse images from their under-determined mixtures, i.e. when the number of sensors is lower by one than the number of unknown sources. A practically feasible solution is proposed for constrained classes of images, i.e. sparse, binary-valued and dynamically-constrained sources.

Keywords: Blind signal processing, image enhancement, linear optimization, sparse signals

Introduction

The goal of *blind source separation* (BSS) is to extract (statistically independent) unknown source signals from their linear mixtures without knowing the mixing coefficients [1, 4, 6]. This blind signal processing technique has so far main applications in data mining and biomedical signal processing problems.

A precondition for the application of BSS solutions is that the number of (statistically independent) source signals is at most equal to the number of sources, known a priori. Typically it should be equal to the number of sensors and outputs. However, in practice these assumptions do not often hold. In this paper we consider the mixing case, where the number of independent sources is higher than the number of mixtures by one.

1. The blind source extraction problem

Denote by $\mathbf{x}(t) = [x_1(t), \dots, x_n(t)]^T$ the n -dimensional t -th data vector made up of the mixtures at discrete index value (usually time) t . The mixing

model in blind source separation (BSS) can then be written in the vector form

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t) \quad (1)$$

Here $\mathbf{s}(t) = [s_1(t), \dots, s_m(t)]^T$ is the source vector consisting of the m source signals at the index value t . Furthermore, each source signal $s_i(t)$ is assumed to be a stationary zero mean stochastic process. $\mathbf{n}(t)$ means Gaussian noise.

In standard neural and adaptive source separation approaches, an $m \times n$ separating matrix \mathbf{W} is updated so that the m -vector ($\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t)$) becomes an estimate ($\mathbf{y}(t) = \hat{\mathbf{s}}(t)$) of the original independent sources [1, 7].

A standard assumption in BSS is that the number m of the sources should be *known* in advance. Like in most neural BSS approaches, we have assumed up to now that the number m of the sources and outputs l are *equal* in the separating network. Generally, both these assumptions may not hold in practice.

Now, let us consider the difficult case where there are fewer mixtures than sources: $n < m$. Then the $n \times m$ mixing matrix \mathbf{A} in (1) has more columns than rows. In this case, complete separation is usually out of the question. However, some kind of separation may still be achievable in special instances at least. This is the goal of blind signal extraction (BSE).

The BSE problem has recently gained larger attention. Pajunen [9] has proposed an algorithm for binary source separation, that separates m binary sources from two or more mixtures. The restrictive assumptions about sources are that the mixture vectors must not overlap, and the mixing matrix must have *non-parallel* column vectors.

Chen & Donoho [3] have applied a so called *Basis Pursuit* approach for spectrum estimation. *Basis Pursuit* decomposes a signal into an optimal superposition of dictionary elements, where optimal means having the smallest l^1 norm of coefficients among all such decompositions. Their dictionary includes over-complete cosine and sine bases, and the Dirac basis. Hence this optimization principle leads to decompositions that can be very sparse.

This topic has been studied theoretically in [2]. The authors show that it is possible to separate the m sources into n disjoint groups if, and only if, \mathbf{A} has n linearly independent column vectors, and the remaining $m - n$ column vectors satisfy the special condition that each of them is parallel to one of these n column vectors.

Recently Li et al. [8] have proposed a multi-stage approach to sparse signal extraction. In the first stage the observed data is transformed into the time-frequency domain via wavelets. Next the use of a sophisticated hierarchical clustering technique allows to identify the mixing matrix. In the last step the sparse sources are estimated alternatively by linear-, quadratic- or semi-definite programming optimization. In case of images, their spectral distributions are very similar to each other, as their spectra are dominated by the first frequency coefficient. Therefore a different approach is proposed in our paper.

2. Proposed solution to BSE

In general the vectors $\mathbf{s}(k)$ and $\mathbf{x}(k)$ are correlated $\mathbf{R}_{\mathbf{x}\mathbf{s}} = E\{\mathbf{x}\mathbf{s}^T\} \neq 0$, but the noise vector is not correlated with \mathbf{s} . Our objective is to find the best estimation matrix $\hat{\mathbf{A}}_{best}$ such that the pair of vectors: $\mathbf{n} = \mathbf{x} - \mathbf{A}\mathbf{s}$ and \mathbf{s} are no longer correlated with each other:

$$0 = E\{(\mathbf{x} - \mathbf{A}\mathbf{s})\mathbf{s}^T\} = \mathbf{R}_{\mathbf{x}\mathbf{s}} - \mathbf{A}\mathbf{R}_{\mathbf{s}\mathbf{s}}. \quad (2)$$

and

$$\hat{\mathbf{A}}_{best} = \mathbf{R}_{\mathbf{x}\mathbf{s}}\mathbf{R}_{\mathbf{s}\mathbf{s}}^{-1}. \quad (3)$$

Assuming that the sensor signals are available only, our BSE approach consists of **two main steps**, that are iterated together (until convergence is achieved) and of a third (optional) final refinement step: (I.a) to estimate $\hat{\mathbf{A}}$ of the (unknown) mixing matrix \mathbf{A} from mixed (sensor) signal vector $\mathbf{x}(t)$; (I.b) to estimate source signals $\hat{\mathbf{s}}(t)$, for given $\hat{\mathbf{A}}$ and $\mathbf{x}(t)$; (II.) a final post-processing for specific signals (option).

Extraction of sources

After the estimate $\hat{\mathbf{A}}$ (or $\hat{\mathbf{A}}_1$) of the mixing matrix \mathbf{A} (or combined mixing matrix \mathbf{A}_1 , respectively) is known, there exist potentially many solutions to the under-determined source extraction problem. We can solve it in special cases, at least. When source signals are spiky and sparse signals, in the sense that they fluctuate mostly around zero and only occasionally have nonzero values, the problem of estimation of unknown signals can be converted to the extended linear programming problem, i.e. finding the optimal sequence of estimated source signals $\hat{s}_i(k)$ ($i = 1, \dots, n$), which minimize the l_1 norm:

$$\sum_k \sum_{i=0}^n |\hat{s}_i(t)|, \quad (4)$$

subject to the constraints:

$$\hat{\mathbf{A}}\hat{\mathbf{s}}(t) = \mathbf{x}(t), \quad \text{or} \quad \hat{\mathbf{A}}_1\hat{\mathbf{s}}(t) = \mathbf{v}(t), \quad \forall k. \quad (5)$$

A very efficient linear programming algorithm that allows one to minimize the l^1 norm is known, called the FOCUSS algorithm [5]. The author has applied this FOCUSS algorithm with the following iteration rule:

$$\hat{\mathbf{s}}(k+1) = \mathbf{D}(k+1) \text{inv}[\hat{\mathbf{A}}\mathbf{D}(k+1)]\mathbf{x}(k+1), \quad (6)$$

where the diagonal matrix \mathbf{D} is obtained by:

$$\mathbf{D}(k+1) = \text{diag}[|\hat{\mathbf{s}}(k)|^{1-p/2}], \quad (7)$$

and $inv[\cdot]$ means the pseudo-inverse operation:

$$inv[\mathbf{W}] = \mathbf{W}^T(\mathbf{W}\mathbf{W}^T)^{-1}. \quad (8)$$

The initial diagonal elements are set to $diag[\mathbf{D}(0)] = [1, \dots, 1]^T$ and the parameter $p = 0.5$. During the above iteration process a competition between the columns of $\hat{\mathbf{A}}$ appears, which of them should represent the vector \mathbf{x} . At the end some of the columns survive only to represent \mathbf{v} .

Mixing matrix estimation

With given source signal estimation $\hat{\mathbf{s}}$ estimate the mixing matrix $\hat{\mathbf{A}}$ by using an iterative rule:

$$\hat{\mathbf{A}}(k+1) = \hat{\mathbf{A}}(k) - \eta(k)[\hat{\mathbf{A}}(k)\mathbf{R}_{\hat{\mathbf{s}}\hat{\mathbf{s}}} - \mathbf{R}_{\mathbf{x}\hat{\mathbf{s}}} - \gamma(k)\hat{\mathbf{A}}(k)]. \quad (9)$$

The forgetting factor $\gamma(k) = tr[\hat{\mathbf{A}}(k)^T(\hat{\mathbf{A}}(k)\mathbf{R}_{\hat{\mathbf{s}}\hat{\mathbf{s}}} - \mathbf{R}_{\mathbf{x}\hat{\mathbf{s}}})]$ ensures that the Frobenius norm of the matrix $\hat{\mathbf{A}}$ is kept approximately constant during the iteration process, thus enforcing the stability of the algorithm.

Postprocessing for (n-1) sensors to (n) sources

After all the previous steps a proper estimation of the sources is usually done for all such vector samples, where at least one source sample is equal to zero. If all sources have non-zero values, then the estimated sample vector is estimated wrongly. At least in some specific signal cases this error could be corrected.

Let us assume that (n-1) sensors are available for (n) sources. Then the crossing section of (n-1) hyperplanes in the n-hyperspace determines a line in the n-dimensional space. A wrong signal vector corresponds to a point on this line, whereas the solution point is located somewhere else on this line. Thus the proper solution can be obtained by a linear shift of the estimated point. The direction cosine of the solution line is dependent on the known (estimated) mixing matrix and it is independent of the sources. Thus, if we find the proper correction value (at a given time sample) for one estimated source, we will be able properly to correct all the remaining outputs.

Usually there is no need for any post-correction if the sources are *spiky signals*, i.e. with high probability in each time sample at least one of the sources is equal to zero.

The ST constrain

The proposed correction mechanism can be applied to several types of source signals. The sources may be binary signals or three-valued positive signals. They may even be of general waveform but subject to our so called *ST-constrain*,

i.e. they should fluctuate rather slowly and smoothly in comparison with the sampling frequency – only *one source* is allowed to have an amplitude change for two consecutive pixels *at given image position*.

3. Test results

In our experiments, natural or synthetic grey-scale images (with 256 grey levels) are used. Their size is equal to 256×384 and 256×256 . Before the start of the learning procedure the image signals should be transformed to zero-mean signals, and for compatibility with the learning rate and initial weights they are also scaled to the interval $[-1.0, 1.0]$.

The obtained results can be assessed quantitatively by using suitable mathematical measures, like SNR (signal-to-noise ratio) between each reconstructed source and the corresponding original source.

In the first two experiments the reconstruction of binary images is shown, i.e. binary edge images (Fig. 1) or binary intensity images (Fig. 2). Obviously these are non-spiky signals and in some areas, in which all three sources take non-zero values, the separation fails.



(a) Three binary edge images



(b) Two edge image mixtures



(c) Reconstructed binary edge images



(a) Three binary source images



(b) Two mixtures



(c) Reconstructed binary sources

Figure 1. Example of edge image image extraction - binary edge images.

Figure 2. Example of a natural face image reconstruction - binary source images.

We have extended the testing to non-binary sources. If the sources are three-valued signals or they satisfy the specific ST-constrain, they can also be extracted from an under-determined mixture, i.e. if for n sources the number of mixtures is $(n-1)$. In such situations the third post-processing step is applied. Example of results for 3-valued image sources are shown on Fig. 3. For a ST-constrained intensity image set the results are shown in Fig. 4. Obviously, natural image sets usually do not satisfy our ST-constrain. Hence this

constrain was artificially satisfied by making 2 more copies of each pixel (and adding them to the image) and left shifting the second and third source by one or two pixels, respectively. As provided in Fig. 4, the extracted sources are disturbed by a low frequency error.

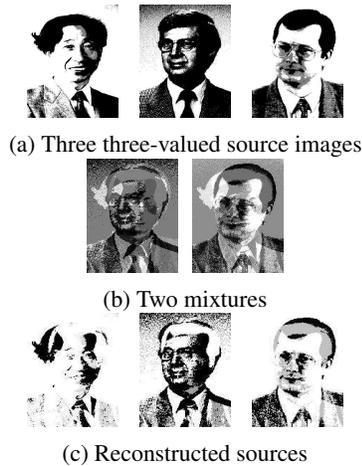


Figure 3. Example of a natural face image reconstruction - three-valued images.

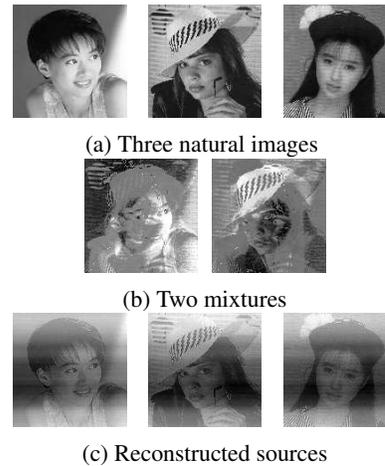


Figure 4. Example of a natural face image reconstruction, subject to the ST-restriction.

References

- [1] A.J. Bell, T.J. Sejnowski: An information maximization approach to blind separation and blind deconvolution, *Neural Computation*, vol. 7, 1995, 1129–1159.
- [2] X.R. Cao, R.W. Liu: A general approach to blind source separation, *IEEE Trans. on Signal Processing*, vol. 44(1996), March 1996, 562–571.
- [3] S.S. Chen, D.L. Donoho: Application of Basis Pursuit in Spectrum Estimation. *ICASSP'98, Proceedings*, vol. 3(1998), 1865–1868.
- [4] A. Cichocki, S. Amari: *Adaptive Blind Signal and Image Processing*, John Wiley, Chichester, UK, 2003 (II. corrected edition).
- [5] I.F. Gorodnitsky, B.D. Rao: Sparse Signal Reconstruction from Limited Data Using FOCUSS: A Re-weighted Minimum Norm Algorithm, *IEEE Transactions on Signal Processing*, vol. 45(1997), No.3, 600–616.
- [6] A. Hyvarinen, J. Karhunen, E. Oja: *Independent Component Analysis*, John Wiley, New York, 2001.
- [7] J. Karhunen, A. Cichocki, W. Kasprzak, P. Pajunen: On neural blind separation with noise suppression and redundancy reduction, *Int. J. of Neural Systems*, vol. 8 (1997), No. 2, 219–237, World Scientific Publ., London–Singapore.
- [8] Y. Li, A. Cichocki, S. Amari: Analysis of sparse representation and blind source separation, *Neural Computation*, vol. 16(2004), 1–42.
- [9] P. Pajunen: *An Algorithm for Binary Blind Source Separation*, Helsinki Univ. of Technology, Lab. of Computer and Information Science, Report A36, 1996.