Filtering-Free Blind Separation of Correlated Images

Frédéric Vrins, John A. Lee, and Michel Verleysen*

Université catholique de Louvain, Machine Learning Group,
Place du Levant, 3, 1380 Louvain-la-Neuve, Belgium
{vrins, lee, verleysen}@dice.ucl.ac.be
www.ucl.ac.be/mlg

Abstract. When using ICA for image separation, a well-known problem is that most often a large correlation exists between the sources. Because of this dependence, there is no more guarantee that the global maximum of the ICA contrast matches the outputs to the sources. In order to overcome this problem, some preprocessing can be used, like e.g. band-pass filtering. However, those processings involve parameters, for which the optimal values could be tedious to adjust. In this paper, it is shown that a simple ICA algorithm can recover the sources, without any other preprocessing than whitening, when they are correlated in a specific way. First, a single source is extracted, and next, a parameter-free postprocessing is applied for optimizing the extraction of the remaining sources.

1 Introduction

In the recent past years, Independent Component Analysis (ICA) has been a fast growing research topic; many algorithms have been developed to solve the ICA problem. Though they share the same goal, all algorithms approach the problem differently and may have different performances on specific applications. This explains why new ICA contrasts are still developed nowadays. For instance, the support width measure (SWM) has been recently suggested as cost function for ICA; its main advantages are its theoretical convexity for bounded sources, its geometrical interpretation and its simplicity [1].

A particular application of ICA is the blind separation of mixed images. This application does not entirely fulfill the assumptions of the canonical ICA problem, since natural images can be highly correlated, i.e. the sources are not independent anymore. Consequently, it can be expected that usual ICA algorithms would fail to recover the source images given only linear mixtures of them. Indeed, even if exceptions seem to exist [2], two mixed images can be

* M.V. is a Senior Research Associate of the Belgian F.N.R.S. The authors are grateful to A. Cichocki from BSI RIKEN for discussion on MDS-ICA. This work was partially supported by the Belgian F.M.S.R. (project 3.4590.02).

J. Cabestany, A. Prieto, and D.F. Sandoval (Eds.): IWANN 2005, LNCS 3512, pp. 1091–1099, 2005.

more independent than the dependent sources. However, several tools have been designed to address this issue. The most efficient one seems to be filtering. In that case, it is assumed that a frequency band exists, for which the source images are statistically independent. By filtering the raw image mixtures outside this frequency band, new mixtures are obtained and processed by a usual ICA algorithm. Once the latter has converged, the computed unmixing matrix is used for separating the initial (unfiltered) mixed images [3].

Even if the previous method looks very efficient, additional parameters appear, like the cutoff frequencies and the order of the filter, which may be difficult to adjust. For instance, finding the frequency band that makes unknown images fully independent, starting from mixtures of them may be a tedious task. To our knowledge, no simple and automatic method exists to solve this problem.

In this work, we propose to use the SWICA algorithm [4] for optimizing the SWM cost function and for solving image separation problems without any other preprocessing than whitening. The method is tested on a simple example, involving two correlated source images (landscapes). SWICA is first applied to extract a single source image. Next, it is modified to optimize the separation of the second source image. Paradoxically, it is an apparently weakness of the SWM criterion that allows us to separate correlated images without any filtering.

The remainder of this paper is organized as follows. Section 2 introduces the SWM contrast and the corresponding algorithm (SWICA). Section 3 deals with additional theoretical issues, whereas experiments are detailed in Section 4. Finally, conclusions are drawn in Section 5.

2 ICA by Support Width Minimization

In order to make this paper self-contained, this section summarizes the SWM criterion [1] and the related ICA algorithm [4] (SWICA).

2.1 The SWM Cost Function

Consider the linear mixing model $\mathbf{x} = \mathbf{A}\mathbf{s}$, where $\mathbf{s} = [s_1, \ldots, s_n]^T$ is a source vector made of n independent and zero-mean random variables, $\mathbf{A}$ is a square mixing matrix and $\mathbf{x} = [x_1, \ldots, x_n]^T$ is the vector of the mixtures. In general, the mixtures are dependent and correlated, but it is easy to find a decorrelation transformation $\mathbf{V}$ such that $\mathbf{z} = \mathbf{V}\mathbf{x}$ satisfies the whiteness condition $E\{\mathbf{z}\mathbf{z}^T\} = \mathbf{I}_n$. Next, if the sources are white, then an ICA algorithm can be run on the whitened mixtures $\mathbf{z}$ in order to recover the sources. More formally, the ICA algorithm identifies the orthogonal matrix $\mathbf{W}$ in the unmixing model $\mathbf{s} \approx \mathbf{y} = \mathbf{W}\mathbf{z}$. The symbol '$\approx$' means that the transfer matrix $\mathbf{C} \triangleq \mathbf{W}\mathbf{V}\mathbf{A}$ is equal to $\mathbf{P}\mathbf{D}$, where $\mathbf{P}$ and $\mathbf{D}$ are permutation and scaling matrices, respectively.

Several methods exist to find $\mathbf{W}$. One of them consists in minimizing the support width $\Omega(y_i)$ of each estimated source y_i, provided the sources are bounded (i.e. the support of the source distribution $\Omega(s_i)$ is finite). In this case, it has been shown that if the number of sample points is large enough, the SWM criterion is convex. In other words, each local minimum of $\Omega(y_i)$ satisfies $\mathbf{c}_i = \pm\mathbf{e}_k$, where $\mathbf{c}_i$ is the i-th row of $\mathbf{C}$ and $\mathbf{e}_k$ the k-th row of $\mathbf{I}_n$. This can be understood by looking at the theoretical expression of $\Omega(y_i)$, expressed as a function of the source supports:

$$\Omega(y_i) = \sum_{j=1}^{n} \mathbf{c}_i(j)\Omega(s_j) \ , \tag{1}$$

where the source supports $\Omega(s_1), \ldots \Omega(s_n)$ are positive and finite. The function $\Omega(y_i)$ can have a minimum only if $\mathbf{c}_i = \pm\mathbf{e}_k$, i.e. when y_i is proportional to a source s_j.

The support minimization of the output is illustrated in Fig. 1. The first angle ϕ correspond to the rotation of the source resulting from $\mathbf{VA}$ while angle φ is associated with the unmixing matrix $\mathbf{W}$. The width D is shown when $\mathbf{W}$ has not converged yet to a satisfactory unmixing matrix, i.e. $\Omega(y_1) = \Omega(\mathbf{w}_1\mathbf{z})$ is not minimized yet. After convergence of the algorithm, we observe $\Omega(y_1) = D'$, which corresponds to $y_1 \propto s_1$; this minimum is global since in this example, $\Omega(s_1) < \Omega(s_2)$. A local minimum of $\Omega(y_1)$ would be obtained if $y_1 \propto s_2$.

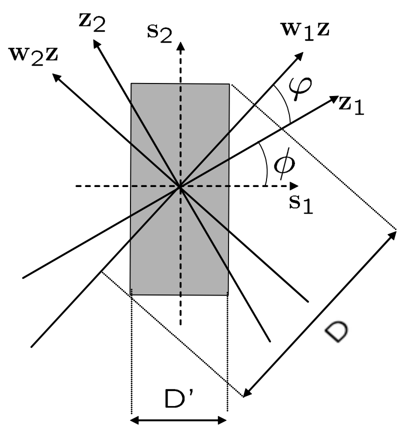


Fig. 1. SWM: principle of ICA by finding directions with minimum support width

2.2 The SWICA Algorithm

SWM is a single-unit ICA criterion. For this reason, when several sources have to be extracted, an algorithm based on a deflation approach must be used. In addition, it must be mentioned that the SWM contrast $(-\Omega)$ is not differentiable, thus making all traditional optimization procedures (fixed point, gradient ascent, etc.) unusable. The algorithm we propose takes as input the whitened mixtures and extracts the sources one after the other, by determining the corresponding row of $\mathbf{W}$. In order to keep $\mathbf{W}$ orthogonal, rows of $\mathbf{W}$ are seen as directions and are updated accordingly. For this purpose, angular variations of the current row $\mathbf{w}_i$ towards another row $\mathbf{w}_j$ are defined and noted as

$$\mathbf{w}_{i\uparrow j} = \cos(\alpha)\,\mathbf{w}_i + \sin(\alpha)\,\mathbf{w}_j \ \text{ and } \ \mathbf{w}_{i\downarrow j} = \cos(\alpha)\,\mathbf{w}_i - \sin(\alpha)\,\mathbf{w}_j \ . \tag{2}$$

Such angular variations allow comparing the current value of the contrast ($-\Omega(\mathbf{w}_i\mathbf{z})$) to surrounding values ($-\Omega(\mathbf{w}_{i\uparrow j}\mathbf{z})$ or $-\Omega(\mathbf{w}_{i\downarrow j}\mathbf{z})$). As the contrast is not differentiable, a very simple optimization procedure is proposed. Briefly put, for each row $\mathbf{w}_i$, the algorithm looks at the contrast value in each perpendicular direction ($\mathbf{w}_j$ with $i+1<j<n$), for both positive and negative angular variations. Then it updates $\mathbf{W}$ by rotating both $\mathbf{w}_i$ and $\mathbf{w}_j$ according to the highest contrast value. Consequently, the algorithm keeps $\mathbf{W}$ orthogonal.

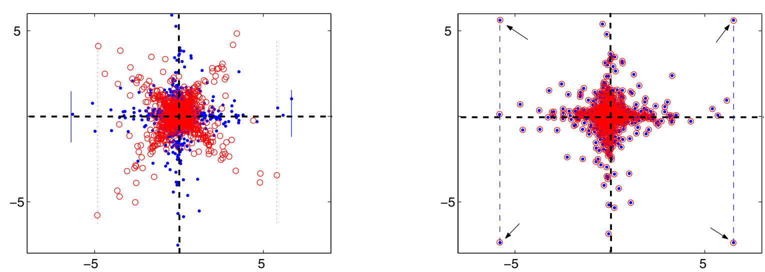
The only parameters of the algorithm are the convergence rate and the number of iterations. By construction, the algorithm is monotonic: the contrast is either increased or kept constant. Similar algorithms using more sophisticated techniques, like discrete gradient approximations (based on a 2^{nd}-order Taylor expansion) have been tried too. Unfortunately, they lead to worse results than the simple proposed algorithm. In addition, they involve a larger number of parameters that are tedious to adjust. More details can be found in [4].

3 SWM: An Extreme Statistics Contrast

Maximizing the contrast requires an estimation of $-\Omega(y_i)$. If the number of observations is large enough, the following estimator measures the support width of a random variable u:

$$\hat{\Omega}(u) = \max(u) - \min(u) \ . \tag{3}$$

This estimator works best for abruptly bounded variables. When tails of the distribution are longer and less dense, as for platykurtic variables, the estimator may fail to give a good approximation of the support, as illustrated in Fig. 2(a),



(a) Essential points for the estimation of $\Omega(\mathbf{y}_1)$ are missing: the source extraction failed

(b) Four points have been added at the borders of the source JPDF (located by the arrows); these 'artificial' points allow SWICA to extract the sources

Fig. 2. SWICA applied on super-Gaussian signals: scatter plots of the source signals (dots) and of the outputs (circle)

because there are not enough observations in 'critical areas' (the 'corners of the square' in the figure). In this case, SWICA may be completely mislead: SWICA has minimized the support of y_1, but this output does not correspond to a source. Of course, if four well-chosen observations are available (see arrows in Fig. 2(b)), the problem disappears. This means that the SWM contrast may be very sensitive to a small number of observations. Fortunately, the distributions of the pixel intensities in an image are usually abruptly bounded, due to particular implementation choices (small encoding range) and image properties (because neighboring pixels often have similar values, there are usually few outliers).

4 Separation of Correlated Images

In this section, we show results of three ICA algorithms (FastICA [5], JADE [6] and SWICA) used to separate two correlated images (correlation: 26%). The source images are two gray-level landscapes (Fig. 3(a) and 3(d)); pixel intensities range from 0 to 255. A random linear mixing matrix and a decorrelation transformation are then applied; mixed images are shown in Fig. 3(b) and 3(e) (after translation and scaling to map the mixtures in the full range $[0, 255]$, for readability purposes).

When looking at the scatter plot of the source images in Fig. 5(a), it becomes clear that they are not independent because the joint probability density function cannot be factorized (for instance, look at several horizontal – or vertical – conditional pdfs in the scatter plot: they are not equal to each other). As could be feared, both JADE and FastICA fails to recover the source images (see Fig. 4(b) and 4(c)).

In order to assess the extraction quality of the i-th source s_i, the 'i-th performance index' is defined as $\mathrm{PI}(i) \triangleq \frac{\sum_{j=1}^{n} \mathbf{c}_i(j)^2}{\max_j \mathbf{c}_i(j)^2} - 1$; a zero PI(i) indicates that y_i is proportional to a source, while a high PI(i) means that y_i results from the superimposition of several sources.

4.1 Extraction of the First Source

SWICA behaves rather differently than JADE and FastICA and recovers one of the sources (Fig. 4(d)). The output scatter plot is a parallelogram and two of its edges are parallel to the vertical axis: values of y_1 computed by SWICA nearly equal those of s_1, since both marginal pdfs (initial sources and estimated ones) along the horizontal axis coincide. Unfortunately, the two other edges of the parallelogram are not parallel to the horizontal axis, meaning that y_2 does not correspond to the second source.

Understanding why both JADE and FastICA fail in this example is straightforward. Because source images are correlated, estimating independent sources amounts to extracting their common but independent components. In the case of two landscapes, these components are not the source images but new images (e.g. comp. 1 could account for the shared soil/sky contrast whereas comp. 2 could account for varying trees, mountain and clouds).

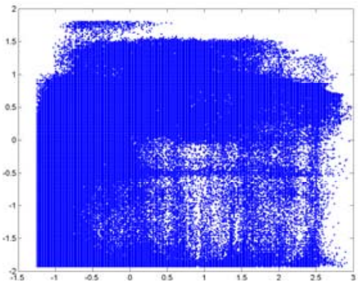
(a) Source 1 (b) Mixture 1 (c) Output 1

(d) Source 2 (e) Mixture 2 (f) Output 2

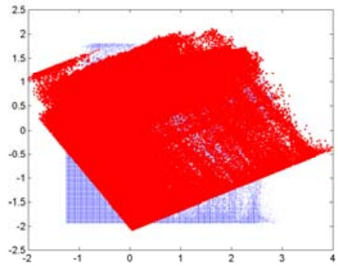
Fig. 3. Example of images separation using SWICA (cov=0.26); source images (a,d), rescaled mixed images (b,e), rescaled extracted images (c,f)

Contrarily to other ICA contrasts, the SWM extracts a very limited piece of information out of the marginal pdf of the currently estimated source y_1: the bounds. In our image application, bounds are particularly interesting parts of the image distribution. Indeed, images may be assumed to involve three more or less important parts: (i) a global shared shape, (ii) local independent details and (iii) encoding techniques. The global shared shape leads to highly correlated and dense spots in the scatter plot. On the other hand, local independent details contributes to fill the scatter plot in a uniform but very sparse way. Finally, encoding technique generally produce saturation effects (towards full white and/or black), which are independent (source images are independently encoded). Usual ICA contrasts are especially sensitive to the global shape, which is dominating in correlated images, and thus try to make the images independent. On the other hand, SWM focuses on the bounds of the scatter plot: these bounds are generally well drawn due to parts (ii) and/or (iii) of the images and contains most of the independent features of the images. Of course, if parts (ii) and (iii) in the image are negligible, the edge of the scatter plots disappear and SWICA is likely to fail.

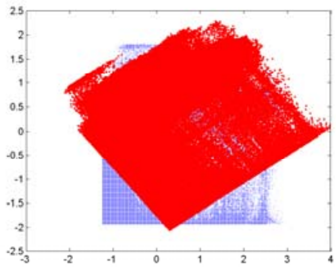
This explains why SWICA can match its first output y_1 to one of the source (see Figure 1). But why does SWICA fail to recover the second source image? Actually, as for all ICA orthogonal contrasts, we whiten the mixtures beforehand and then constrain the unmixing matrix $\mathbf{W}$ to be orthogonal. Unfortunately, this constraint is too restrictive in our case and amounts to recovering sources that are not correlated. More precisely, on one hand we know that



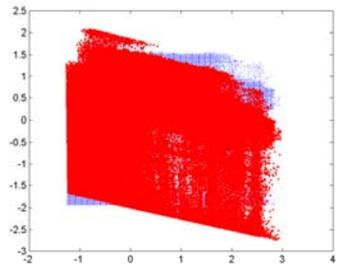
(a) Normalized sources



(b) FastICA (y_1 vs y_2); PI(1) = 0.246 and PI(2) = 0.449



(c) Jade (y_1 vs y_2); PI(1) = 0.497 and PI(2) = 0.687



(d) Swica (y_1 vs y_2); PI(1) = 1.3×10^{-4} and PI(2) = 0.063

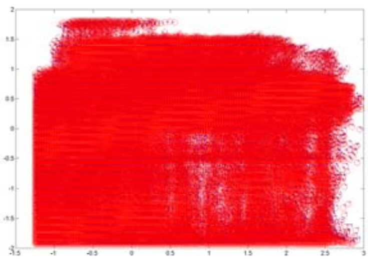
Fig. 4. Scatter plots between normalized images. In (b)-(d), the plots of y_1 vs y_2 (dark) are superimposed to the plot of the normalized sources (light gray)

$E\{\mathbf{y}\mathbf{y}^T\} = \mathbf{W}E\{\mathbf{z}\mathbf{z}^T\}\mathbf{W}^T = \mathbf{W}\mathbf{W}^T = \mathbf{I}_n$ because of the whiteness property. On the other hand, we know that $E\{\mathbf{s}\mathbf{s}^T\}$ is not diagonal, which is contradictory.

4.2 Extraction of the Second Source

The above-mentioned arguments explain why one source can be recovered, whereas the other ones cannot be recovered when the source images are correlated and $\mathbf{W}$ is constrained to be orthogonal.

In the easy case where $n = 2$, the second line of the $\mathbf{W}$ matrix given by SWICA must be modified. Therefore, we minimize $\hat{\Omega}(y_2') = \mathbf{w}_2'\mathbf{z}$, where $\mathbf{w}_2'$ is not constrained to be orthogonal to $\mathbf{w}_1$ anymore. In order to avoid converging to y_1, we take $\mathbf{w}_2$ as first guess for $\mathbf{w}_2'$. This procedure is applied only on the second output, without changing the first one, and allows separating the second source, as shown in Fig. 5. This procedure can be extended for a larger number



(a) Swica (y_1 vs y'_2); PI(1) $= 1.3\times10^{-4}$ and PI(2) $= 1.4 \times 10^{-4}$

(b) Second output y'_2 after mapping to $(0, 255)$

Fig. 5. Results after extraction of the second source image: the low PI(.) indicate that both sources are recovered correctly

of source images, by deriving a deflation algorithm to correct the bias due to the orthogonality constraint.

5 Conclusion and Future Work

In this paper, the SWM contrast is used to solve the ICA problem involving correlated images. This approach is motivated by the unsatisfying results of JADE and FastICA for the same problem, when no other preprocessing than whitening, like filtering is used.

Regarding the separation of two correlated images, SWICA succeeds in recovering one of the images but prewhitening of the mixtures settles on $\mathbf{W}$ an orthogonality constraint that jeopardizes the retrieval of the second source image. To circumvent this problem, the orthogonality constraint is relaxed after the first deflation step. In this case, both source images are recovered correctly.

Future work will compare the proposed approach to MSD ICA (multiresolution subband decomposition), a method that preprocesses the image mixtures using filters. A second target is the development of a new version of SWICA, without orthogonality constraint.

References

1. Vrins, F., Verleysen, M.: Swm : A class of convex contrasts for source separation. In: proceedings of ICASSP'05 (accepted), Philadelphia (USA) (2005)
2. Yang, H., Amari, S.I.: Adaptive on-line learning algorithms for blind separation-maximization entropy and minimum mutual information. Neural Computation **9** (1997) 1457–1482
3. Cichocki, A., Georgiev, P.: Blind separation algorithms with matrix constraints. IEICE Trans. fundamentals **E86-A** (2003) 522–531

4. Lee, J., Vrins, F., Verleysen, M.: A simple ica algorithm for non-differentiable contrasts. In: proceedings of EUSIPCO'05 (submitted), Antalya (Turkey) (2005)
5. Hyvärinen, A.: Fast and robust fixed-point algorithms for independent component analysis. IEEE Trans. on Neural Networks **10** (1999) 626–634
6. Cardoso, J.F.: Source separation using higher order moments. Proceedings of the IEEE, Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP'89) (1989, Glasgow, England) 2109–2112