Blind source separation by nonstationarity of variance: A cumulant-based approach

Aapo Hyvärinen

Neural Networks Research Centre
Helsinki University of Technology
P.O. Box 5400, FIN-02015 HUT, Finland
and
Department of Psychology, General Psychology Division,
University of Helsinki
P.O Box 13, FIN-00014 Univ. of Helsinki, Finland

To appear in IEEE Transactions on Neural Networks

June 19, 2001

Abstract

Blind separation of source signals usually relies either on the nongaussianity of the signals, or on their linear autocorrelations. A third approach was introduced by Matsuoka et al, who showed that source separation can be performed by using the nonstationarity of the signals, in particular the nonstationarity of their variances. In this paper, I show how to interpret the nonstationarity due to a smoothly changing variance in terms of higher-order cross-cumulants. This is based on considering the time-correlation of the squares (energies) of the signals, and leads to a simple optimization criterion. Using this criterion, I construct a fixed-point algorithm that is computationally very efficient.

1 Introduction

In blind source separation [8, 6], we observe linear mixtures $\mathbf{x}(t) = (x_1(t), x_2(t), ..., x_n(t))$ of unknown source signals $\mathbf{s}(t) = (s_1(t), s_2(t), ..., s_n(t))$. This can be expressed as a latent variables model:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \tag{1}$$

The fundamental assumption is that the source signals are statistically independent. The problem is then to estimate both the mixing matrix $\mathbf{A}$ and the source signals $s_i(t)$, using observations of the mixtures $x_i(t)$ alone. For simplicity, we assumed here that the dimension of $\mathbf{x}(t)$ equals the dimension of $\mathbf{s}(t)$, which need not necessarily be the case. Likewise, the matrix $\mathbf{A}$ is assumed to be nonsingular.

Estimation of this model is not possible in general: If the $s_i(t)$ are i.i.d. samples from a gaussian distribution, neither the matrix $\mathbf{A}$ nor the source signals $\mathbf{s}(t)$ can be properly estimated [3, 14]. However, if some assumptions on the $s_i(t)$ are made in addition to independence, the model can be estimated. Most source separation methods use one of the following two assumptions. First, one can assume that the source signals $s_i(t)$ have nongaussian (marginal) distributions [3, 7, 8]. Second, one can assume that the source signals have (linear) autocorrelations that are nonzero and distinct [14, 10, 2].

A third assumption that allows estimation of the model was introduced by Matsuoka et al [9]. This was the nonstationarity of the variances. If the variances of the source signals are nonstationary, and change

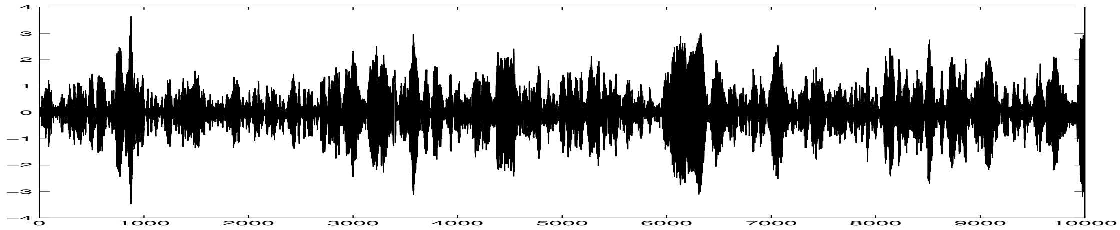

Figure 1: A signal with nonstationary variance.

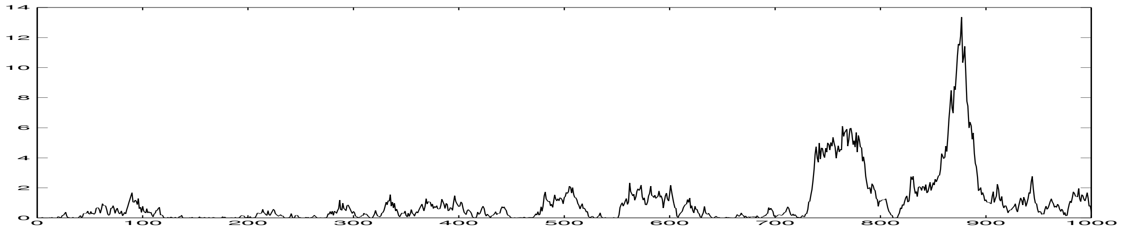

Figure 2: The energy (i.e. squares) of the initial part of the signal in Fig. 1. This is clearly correlated in time.

smoothly in time, the model can be estimated. Note that all these three assumptions are independent in the sense that none of them implies or presupposes any of the other assumptions.

In this paper, I propose a simple estimation method for nonstationary sources. First, I show how to formulate variance nonstationarity (where the variance is changing smoothly) using a higher-order cross-cumulant. Then I derive a fixed-point algorithm for maximization of this cross-cumulant following the derivation of the FastICA algorithm in [7]. The resulting estimation method is simpler and computationally more efficient than the one originally proposed by Matsuoka et al [9], and shows clearly the connection of nonstationarity to the cumulant-based methods widely used in many areas of signal processing.

2 Nonstationarity and cumulants

2.1 Nonstationarity and energy correlation

To illustrate the variance nonstationarity in its purest form, let us look at the source signal in Fig. 1. This source signal was created so that it has a gaussian marginal density, and no linear timecorrelation, i.e. $E\{s(t)s(t-\tau)\} = 0$ for any lag $\tau > 0$. (All signals here are assumed to have zero mean.) Thus, source signals of this kind could not be separated by ordinary source separation methods based on marginal nongaussianity or (linear) time correlations. On the other hand, the nonstationarity of the signal is clearly visible.

In this paper, I propose that nonstationarity due to smoothly changing variance could be measured by cumulants. To see how this works, consider the energy (i.e. the squares $s(t)^2$) of the signal in Fig. 1. The energies of the initial 1000 time points are shown in Fig. 2. The energy of the signal has a power spectrum mostly concentrated at lower frequencies. This is of course a consequence of the assumption that the variance changes smoothly in time. Thus, the variance nonstationarity is related to the time structure of the energies.

Before proceeding, note that the nonstationarity of a signal depends on the time-scale and the level of the detail in the model of the signal. If the nonstationarity of the variance is incorporated in the model (by hidden Markov models, for example), the signal no longer needs to be considered nonstationary [12]. This is the approach that we choose in the following. In particular, the energies are *not* considered nonstationary, but rather they are considered as stationary signals that are time-correlated. This is simply a question of changing the viewpoint.

2.2 A cumulant-based criterion

Inspired by the example above, we could thus measure the variance nonstationarity of a signal $y(t), t = 1, ...t$ using a measure based on the autocorrelation of energy: $E\{y(t)^2 y(t-\tau)^2\}$ where τ is some lag constant, often equal to one. For the sake of mathematical simplicity, it is often useful to use cumulants instead of such basic higher-order correlations. Cumulants also have useful algebraic properties that enable a powerful algorithm as will be seen below. The cumulant corresponding to the autocorrelation of energy is given by the fourth-order cross-cumulant

$$\mathrm{cum}(y(t), y(t), y(t-\tau), y(t-\tau)) = E\{y(t)^2 y(t-\tau)^2\} - E\{y(t)^2\}E\{y(t-\tau)^2\} - 2(E\{y(t)y(t-\tau)\})^2 \quad (2)$$

This could be considered as a normalized version of the autocorrelation of energy. In particular, if the signal has no linear time-correlations, we have $E\{y(t)y(t-\tau)\} = 0$, and this cumulant is equal to the covariance of the energies: $\mathrm{cov}(y(t)^2, y(t-\tau)^2)$.

In order for our approach to work, we must assume that the cross-cumulants $\mathrm{cum}(s_i(t), s_i(t), s_i(t-\tau), s_i(t-\tau))$ are not zero for the source signals. This is analogous to the case in basic ICA using kurtosis [4, 7], where it has to be assumed that the kurtoses of the sources are not zero. In our case where the variances are changing smoothly, this cumulant can actually be assumed positive for all source signals because the first term in (2) dominates the two normalizing terms (at least if the lag is not too large).

In particular, this assumption implies that a source signal must not have *jointly* gaussian statistics, that is, the vector $(s(t), s(t-\tau))$ must not have a jointly gaussian distribution. Note that although cross-cumulants are zero for random variables with jointly gaussian distributions [11], they need not be zero for variables that merely have gaussian *marginal* distributions. Thus, positive cross-cumulants do not imply nongaussian marginal distributions for the source signals, which shows that the property measured by this cross-cumulant is indeed completely different from the property of nongaussianity as measured, for example, by kurtosis.

A valid separation criterion can be easily obtained using this cross-cumulant, assuming that the corresponding cross-cumulants are nonzero. The idea is to use the cross-cumulant in exactly the same way as kurtosis in [4, 7]. Thus, we estimate the source signals by finding the linear combinations, say $\sum_i w_i x_i(t) = \mathbf{w}^T \mathbf{x}(t)$, such that *the absolute value of the cross-cumulant is maximized*. At the same time, we must constrain the scale, because otherwise the cross-cumulant could be made infinitely large by letting the norm of $\mathbf{w}$ grow infinitely. Just like in [4, 7], we do this by constraining the variance of $\mathbf{w}^T\mathbf{x}$ to be constant (say, equal to one). Thus, to estimate one source signal, we solve the following optimization problem:

$$\max_{\mathbf{w}} |\mathrm{cum}(\mathbf{w}^T\mathbf{x}(t), \mathbf{w}^T\mathbf{x}(t), \mathbf{w}^T\mathbf{x}(t-\tau), \mathbf{w}^T\mathbf{x}(t-\tau))| \quad (3)$$

$$\text{under constraint: } \mathrm{var}(\mathbf{w}^T\mathbf{x}(t)) = 1 \quad (4)$$

In fact, just as the kurtosis approach to ICA leads to the observation that the source signals are obtained as the most nongaussian (most kurtotic) linear combinations, we have here the result that the combinations with maximally nonstationary, smooth variance give the source signals. Proof of these properties are given next.

2.3 Mathematical analysis of criterion

Now we show in detail the validity of the criterion given above. This linear combination $\mathbf{w}^T\mathbf{x}(t)$ is a linear combination of the source signals $\mathbf{w}^T\mathbf{x}(t) = \mathbf{w}^T\mathbf{A}\mathbf{s}(t)$, say $\mathbf{q}^T\mathbf{s} = \sum_i q_i s_i(t)$. Using the basic properties of cumulants [11, 7], this particular cumulant of such a linear combination can be evaluated as

$$\mathrm{cum}(\mathbf{w}^T\mathbf{x}(t), \mathbf{w}^T\mathbf{x}(t), \mathbf{w}^T\mathbf{x}(t-\tau), \mathbf{w}^T\mathbf{x}(t-\tau)) = \sum_i q_i^4 \mathrm{cum}(s_i(t), s_i(t), s_i(t-\tau), s_i(t-\tau)). \quad (5)$$

Now, we can constrain the variance of $\mathbf{w}^T\mathbf{x}$ to be equal to unity to normalize the scale (cumulants are not scale-invariant). This implies $\mathrm{var}\sum_i q_i s_i = \|\mathbf{q}\|^2 = 1$. Let us consider what happens if we maximize the absolute value of the cumulants with respect to $\mathbf{w}$. This is equivalent to the optimization problem

$$\max_{\|\mathbf{q}\|^2=1} |\sum_i q_i^4 \mathrm{cum}(s_i(t), s_i(t), s_i(t-\tau), s_i(t-\tau))| \quad (6)$$

This optimization problem is formally identical to the one encountered when the absolute value of kurtosis is maximized to find the most nongaussian directions. The only difference is that the kurtoses of the source signals are replaced by the cumulants measuring nonstationarity. It was proven in [4] that such solutions to this optimization problem give the source signals. In other words, the maxima of (6) are obtained when only one of the q_i is nonzero. This proof applies directly in our case as well, and thus we see that the maximally nonstationary linear combinations give the source signals. Since the cross-cumulants are assumed to be all positive, the problem we have here is in fact slightly simpler since we can then simply maximize the cross-cumulant of the linear combinations, and need not necessarily consider its absolute value as is done with kurtosis [4, 7]. As with kurtosis, we must assume that the cumulants with the time lag τ are nonzero. Finding such a τ is a further problem in practice; taking $\tau = 1$ is the basic choice.

Thus we see that maximization of the nonstationarity, as measured by the cross-cumulant, of a linear combination of the observed mixtures allows for the estimation of one source signal. This also gives a one-unit approach [4, 7] to source separation by nonstationarity.

3 A fixed-point algorithm

To maximize the nonstationarity as measured by the cross-cumulant, I propose a fixed-point algorithm along the same lines as the FastICA algorithm [7] for maximizing nongaussianity.

To begin with, let us whiten the data. In other words, the data $\mathbf{x}(t)$ is linearly transformed into a vector $\mathbf{z}(t) = \mathbf{V}\mathbf{x}(t)$ so that the covariance of $\mathbf{z}(t)$ equals unity: $E\{\mathbf{z}(t)\mathbf{z}(t)^T\} = \mathbf{I}$. This is a well-known preprocessing in source separation [7]. This implies that the variance of $\mathbf{w}^T\mathbf{z}(t)$ is equal to its norm: Thus the normalization of its variance can be done by simply dividing $\mathbf{w}$ by its norm.

Now, using the principle of fixed-point iteration [7, 13], let us equate $\mathbf{w}$ to the gradient of the cross-cumulant of $\mathbf{w}^T\mathbf{z}(t)$. This gives, after straight-forward calculations, the following update formula $\mathbf{w}$:

$$\mathbf{w} \leftarrow E\{\mathbf{z}(t)\mathbf{w}^T\mathbf{z}(t)(\mathbf{w}^T\mathbf{z}(t-\tau))^2\} + E\{\mathbf{z}(t-\tau)\mathbf{w}^T\mathbf{z}(t-\tau)(\mathbf{w}^T\mathbf{z}(t))^2\} - 2\mathbf{w} - \mathbf{M}\mathbf{w}(\mathbf{w}^T\mathbf{M}\mathbf{w}) \quad (7)$$

where we have multiplied the gradient by $1/2$ for notational simplicity, and the matrix $\mathbf{M}$ is given by $\mathbf{M} = E\{\mathbf{z}(t)\mathbf{z}(t-\tau)^T\} + E\{\mathbf{z}(t-\tau)\mathbf{z}(t)^T\}$. The algorithm thus consists of iteratively computing the new value of $\mathbf{w}$ as in (7), normalizing $\mathbf{w}$ to unit norm after every step. Incidentally, if the lag τ is zero, the algorithm reduces to the FastICA algorithm using kurtosis [7]. (More sophisticated versions of the FastICA algorithm were introduced in [5].)

To estimate several source signals, different kinds of orthogonalization schemes need to be used [7, 6]. Basically, we use the update formula in (7) for several different vectors $\mathbf{w}$, and orthogonalize these vectors. Both deflationary and symmetric orthogonalization can be used, see [7, 5].

The convergence of the algorithm can be proven to be cubic, i.e. very fast. A detailed proof can be constructed as in [7]. To adapt the proof in [7] for our case, it is enough to express the algorithm with respect to the transformed variable $\mathbf{q}$, which can be simply obtained by computing the gradient of (5), to give

$$q \leftarrow q_i^3[4\mathrm{cum}(s_i(t), s_i(t), s_i(t-\tau), s_i(t-\tau))] \quad (8)$$

followed by normalization of the norm of $\mathbf{q}$. This can be easily seen [7, 13] to lead to convergence of $\mathbf{q}$ to a vector where only one of the q_i is nonzero.[1] The index i for which q_i will be nonzero depends on the initial

[1] In fact, this does not necessarily mean that we have found a maximum of the cumulant. As in the case of kurtosis [7], the algorithm actually finds a maximum of the absolute value of the cumulant. But in any case, this leads to separation of the signals, as discussed in Section 2.3.

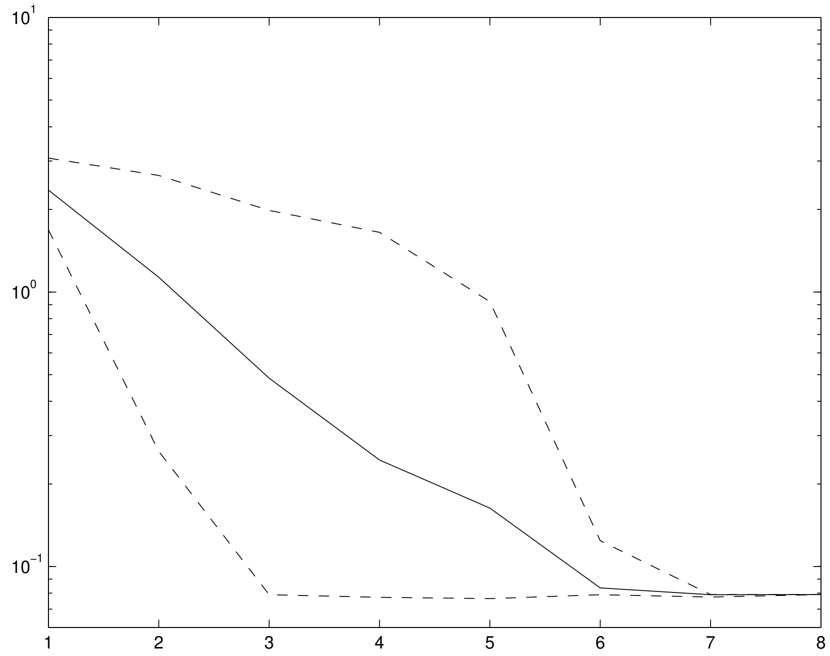

Figure 3: Simulations with the fixed-point algorithm. The figure shows a convergence index (see text for definition) in log-scale. Solid curve: mean over 10 runs. Dashed curves: maximum and minimum over 10 runs. The algorithm converged with a small number of iterations.

value of $\mathbf{q}$.[2]

4 Simulations

Artificial signals as in Fig. 1, i.e. with gaussian marginal distributions and zero linear autocovariance, were created. They were mixed with random matrices and the fixed-point algorithm was used to estimate the mixing matrix. The number of sources and mixtures was 4, and the data size was 10,000. The algorithm was run 10 times with different mixing matrices.

The convergence of the algorithm is shown in Fig. 3. The figure shows the values of a convergence index, defined as the sums of the absolute values of the matrix $\mathbf{WA}$, minus four. Thus a value of zero means that $\mathbf{A}$ was estimated perfectly. In most cases, convergence was obtained after 5 iterations.

5 Conclusion

I considered blind source separation using the less known principle of signal nonstationarity [9, 12]. I gave an interpretation of a particular cross-cumulant as a measure of nonstationarity (for an alternative interpretation of the cumulant, see [1]). This leads to a fast fixed-point algorithm that is similar to the FastICA algorithm [7, 5]. The convergence of the algorithm is cubic, i.e. very fast, like the convergence of the cumulant-based FastICA [7]. This computational efficiency is a major advantage when compared to other methods for separation of nonstationary signals [9, 12]. Moreover, the algorithm allows for sequential (deflationary) estimation of the source signals.

[2] There is a set of measure zero in the space of initial conditions in which convergence is not obtained [7], but we can safely assume that in practice, the initial conditions do not lie in that set.

The choice of the estimation principle for ICA depends, of course, on the data at hand. Marginal nongaussianity and linear autocorrelations are certainly fundamental properties found in many data sets. Nonstationarity due to smoothly changing variance has not been investigated as thoroughly but it is likely that it is a strong feature of data sets in certain applications. An interesting feature of this principle is that it takes into account the time structure of the data (unlike methods based on marginal nongaussianity), but it does not require the time-correlations to be different for each source signal (unlike methods based on linear time-correlations).

References

[1] A. K. Barros and N. Ohnishi. Removal of quasi-periodic sources from physiological measurements. In *Proc. Int. Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, pages 185–190, Aussois, France, 1999.

[2] A. Belouchrani, K. Abed Meraim, J.-F. Cardoso, and E. Moulines. A blind source separation technique based on second order statistics. *IEEE Trans. on Signal Processing*, 45(2):434–444, 1997.

[3] P. Comon. Independent component analysis—a new concept? *Signal Processing*, 36:287–314, 1994.

[4] N. Delfosse and P. Loubaton. Adaptive blind separation of independent sources: a deflation approach. *Signal Processing*, 45:59–83, 1995.

[5] A. Hyvärinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. on Neural Networks*, 10(3):626–634, 1999.

[6] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis.* Wiley Interscience, 2001.

[7] A. Hyvärinen and E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9(7):1483–1492, 1997.

[8] C. Jutten and J. Hérault. Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24:1–10, 1991.

[9] K. Matsuoka, M. Ohya, and M. Kawamoto. A neural net for blind separation of nonstationary signals. *Neural Networks*, 8(3):411–419, 1995.

[10] L. Molgedey and H. G. Schuster. Separation of a mixture of independent signals using time delayed correlations. *Physical Review Letters*, 72:3634–3636, 1994.

[11] C. Nikias and J. Mendel. Signal processing with higher-order spectra. *IEEE Signal Processing Magazine*, pages 10–37, July 1993.

[12] D.-T. Pham and J.-F. Cardoso. Blind separation of instantaneous mixtures of non-stationary sources. In *Proc. Int. Workshop on Independent Component Analysis and Blind Signal Separation (ICA2000)*, pages 187–193, Helsinki, Finland, 2000.

[13] O. Shalvi and E. Weinstein. Super-exponential methods for blind deconvolution. *IEEE Trans. on Information Theory*, 39(2):504:519, 1993.

[14] L. Tong, R.-W. Liu, V.C. Soon, and Y.-F. Huang. Indeterminacy and identifiability of blind identification. *IEEE Trans. on Circuits and Systems*, 38:499–509, 1991.