Diagonalization of time-delayed covariance matrices does not guarantee statistical independence in high-dimensional feature space

Tiziano Zito
Institute for Theoretical Biology
Humboldt University Berlin
Invalidenstraße 43, D-10115 Berlin, Germany
t.zito@biologie.hu-berlin.de

Laurenz Wiskott
Institute for Theoretical Biology
Humboldt University Berlin
Invalidenstraße 43, D-10115 Berlin, Germany
l.wiskott@biologie.hu-berlin.de

ABSTRACT

Independent Slow Feature Analysis (ISFA) is an algorithm for performing nonlinear blind source separation, which combines linear ICA with Slow Feature Analysis (SFA). In its current form the objective function is based on time-delayed covariance matrices. While the algorithm performs well in general, we occasionally encountered cases in which the estimated sources are highly statistically dependent. Here we present a detailed analysis of these cases, which has revealed that second-order covariance matrices do not guarantee statistical independence of a few signals extracted from a high-dimensional feature space.

Keywords: nonlinear blind source separation, ICA, SFA, ISFA, time-delayed covariance matrices

1 INTRODUCTION

Independent component analysis (ICA) is a method for blind source separation (BSS) by transforming a multi-dimensional input signal $\mathbf{x}(t)$ such that the output signal components $u_i(t)$ are statistically independent of each other (for an overview see Hyvärinen, 1999). One approach to ICA is to (approximately) diagonalize several time-delayed covariance matrices of the normalized output signals by minimizing, e.g. by Givens rotations, the objective function

$$\Psi_{\mathrm{ICA}} := \sum_{\tau \in \mathcal{T}} \sum_{\substack{i,j=1 \\ i \neq j}}^{N} \frac{1}{4} \langle u_i(t) u_j(t+\tau) + u_i(t+\tau) u_j(t) \rangle^2 \quad (1)$$

where N is the number of output components and $\mathcal{T}$ indicates the set of time-delays, under the constraint of $\mathbf{u}(t)$ having unit covariance matrix (Belouchrani et al., 1997; Ziehe and Müller, 1998).

In the nonlinear case ICA is an ill-posed problem since the solution is not unique (Jutten and Karhunen, 2003).

2 ISFA

We have proposed earlier to use temporal slowness as an additional criterion to make the solution unique and have demonstrated that nonlinear BSS is possible if statistical independence and slowness are combined. The respective algorithm is based on slow feature analysis (SFA) (Wiskott and Sejnowski, 2002) and second-order ICA and is referred to as independent slow feature analysis (ISFA) (Blaschke et al., 2006). In ISFA the nonlinearity is realized by mapping the input signal into a high-dimensional feature space, within which a low-dimensional subspace is searched for with components that are statistically independent and slowly varying. The objective function to be minimized by ISFA reads:

$$\begin{aligned} \Psi_{\mathrm{ISFA}} &:= \Psi_{\mathrm{ICA}} + \Psi_{\mathrm{SFA}} \\ &= \Psi_{\mathrm{ICA}} - \sum_{i=1}^{N} \langle u_i(t) u_i(t+1) \rangle^2 \end{aligned} \quad (2)$$

which has to be minimized under the constraint of $\mathbf{u}(t)$ having unit covariance matrix. The second term captures the slowness objective, since a slow signal is highly correlated with itself if time-shifted by one time step.

3 FAILURE CASES

Although ISFA works well most of the time, we found that it fails in more than 10% of the cases, because the ICA-criterion based on time-delayed covariances is misleading. As an example consider Figure 1. The left panel shows a scatter plot of two statistically independent audio signals $x_1(t)$ and $x_2(t)$ of CD-music, which were 2^{21} samples long. Together they form a series of two-dimensional input vectors. We mapped these vectors into a nine-dimensional feature space of polynomials of degree three, i.e. x_1, x_2, x_1^2, $x_1 x_2$, x_2^2, x_1^3, $x_1^2 x_2$, $x_1 x_2^2$, and x_2^3, with successive normalization to zero mean and unit covariance matrix. We then searched for two orthogonal directions that (approximately) diagonalize covariance matrices with 50 different time-delays evenly spaced within 1 second.

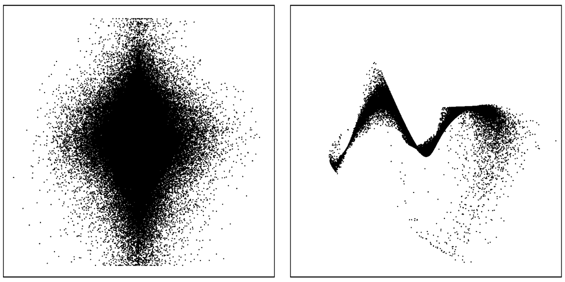

Figure 1: Scatter plot of two original sources (left) and of the two most independent signals in the nine-dimensional feature space of polynomials of degree three (right), for which (1) actually assumes a smaller value than for the original sources.

In this simple test case, we did not mix the input signals, so that the preferred directions should correspond to the original sources, i.e. the first two components. The next panel shows the result of the optimization of Ψ_{ISFA}. It is clear by visual inspection that the output signal components found are not statistically independent but that one component is largely a function of the other one. The smaller panels in Figure 2 to the right show similar failure examples with other music excerpts.

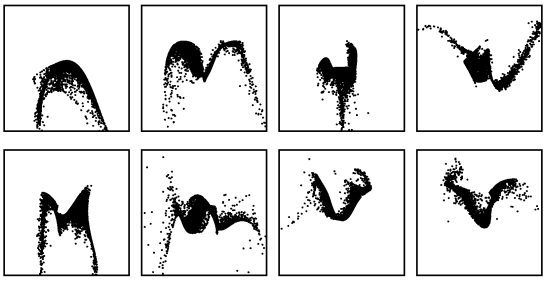

Figure 2: Scatter plots of most independent signals in failure examples.

In all these cases, not only the value of objective function Ψ_{ISFA} is smaller than for the original sources, but both Ψ_{SFA} and Ψ_{ICA} are smaller. This indicates that the signals found are slower *and* appear more mutually independent than the original sources, according to the second-order measure of statistical independence used here. However, as visual inspection clearly shows, the estimated sources are not statistically independent and therefore Ψ_{ICA} fails to guarantee statistical independence.

The cross-correlation functions shown in Figure 3 indicate that these failures are not a result of the particular time-delays we have chosen, since the cross-correlation function for the output signal components is overall smaller than for the original sources. Even using different or more time delays, such datasets would have been processed incorrectly. Furthermore, since the objective function is smaller than for the original sources, these failures are not due to local optima.

The failures must be attributed to the weakness of the ICA-term in the objective function. If the SFA-term were too weak, one would observe output signal components which are truly statistically independent but at least some of them are too quickly varying, so that they are not correlated to the sources but to some nonlinearly distorted version of the sources, something we did not observe. Moreover, we were actually able to detect the failure cases in an unsupervised manner resorting to a measure of statistical independence based on higher-order cumulants, which clearly indicates that the ICA-term is responsible for the failures.

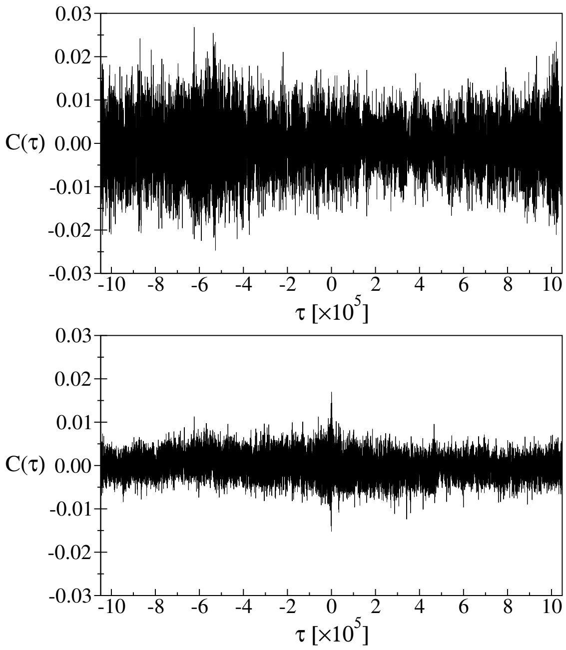

Figure 3: Cross-correlation functions of a failure case: (above) cross-correlation function of the original sources, (below) cross-correlation function of the estimated sources. Same dataset as in Figure 1.

For a possible theoretical account of the failure of second-order ICA in our context consider the following example. Given a symmetrically distributed source x_1 the correlation between, for instance, x_1 and x_1^2 vanishes (Harmeling et al., 2003, sec. 4.1). To the extent that this also holds for time-shifted versions $x_1(t)$ and $x_1^2(t+\tau)$ (cf. Harmeling et al., 2003, sec. 5.4), the statistical dependence between x_1 and x_1^2 does not manifest itself in the time-delayed correlations. Thus, second-order ICA cannot be expected to prevent extraction of x_1 and x_1^2 as the estimated sources, which can easily lead to a failure case, if x_1^2 is more slowly varying than, e.g., x_2, but see (Harmeling et al., 2003).

4 CONCLUSIONS

We have analyzed ISFA, which is an algorithm for nonlinear blind source separation based on the combination of the slowness objective with the statistical independence objective. Using a second-order measure in the ICA objective occasionally leads to failures. Signals are selected as being the most mutually independent signals within a high-dimensional feature space, which are in fact highly dependent on each other. Those failure cases are not due to local optima or to the slowness objective. We conclude that no measure of statistical independence based on time-delayed covariance matrices is appropriate to reliably determine a few statistically independent dimensions in a high-dimensional feature spaces.

This is an important insight for the development of nonlinear blind source separation algorithms.

ACKNOWLEDGEMENTS

This work has been supported by the Volkswagen Foundation through a grant to LW for a junior research group. We thank Barak Pearlmutter for helpful comments. All examples have been implemented in Python using the Modular Toolkit for Data Processing (Berkes and Zito, 2005, 2006).

References

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

P. Berkes and T. Zito. Modular toolkit for data processing version 1.1. `http://mdp-toolkit.sf.net`, 2005.

P. Berkes and T. Zito. MDP 2.0 - a data processing framework for scientific development and education. In *Proc. Europython 2006*, Geneva, July 3–5, 2006.

T. Blaschke, T. Zito, and L. Wiskott. Independent slow feature analysis and nonlinear blind source separation. *Neural Computation (accepted)*, 2006.

S. Harmeling, A. Ziehe, M. Kawanabe, and K.-R. Müller. Kernel-based nonlinear blind source separation. *Neural Computation*, 15:1089–1124, 2003.

A. Hyvärinen. Survey on independent component analysis. *Neural Computing Surveys*, 2:94–128, 1999.

C. Jutten and J. Karhunen. Advances in nonlinear blind source separation. In *Proc. of the 4th Int. Symposium on Independent Component Analysis and Blind Signal Separation, Nara, Japan, (ICA 2003)*, pages 245–256, 2003.

L. Wiskott and T. Sejnowski. Slow feature analysis: Unsupervised learning of invariances. *Neural Computation*, 14(4):715–770, 2002.

A. Ziehe and K.-R. Müller. TDSEP – an efficient algorithm for blind separation using time structure. In *Proc. of the 8th Int. Conference on Artificial Neural Networks (ICANN'98)*, pages 675 – 680, Berlin, 1998. Springer Verlag.