

Independent Slow Feature Analysis and Nonlinear Blind Source Separation

Tobias Blaschke and Laurenz Wiskott

Institute for Theoretical Biology, Humboldt University Berlin
Invalidenstraße 43, D-10115 Berlin, Germany

{t.blaschke,l.wiskott}@biologie.hu-berlin.de

<http://itb.biologie.hu-berlin.de/~blaschke,wiskott>

Abstract. We present independent slow feature analysis as a new method for nonlinear blind source separation. It circumvents the indeterminacy of nonlinear independent component analysis by combining the objectives of statistical independence and temporal slowness. The principle of temporal slowness is adopted from slow feature analysis, an unsupervised method to extract slowly varying features from a given observed vectorial signal. The performance of the algorithm is demonstrated on nonlinearly mixed speech data.

1 Introduction

Unlike in the linear case the nonlinear Blind Source Separation (BSS) problem can not be solved solely based on the principle of statistical independence [Hyvärinen and Pajunen, 1999; Jutten and Karhunen, 2003]. Performing nonlinear BSS with Independent Component Analysis (ICA) requires additional information about the underlying sources or to regularize the nonlinearities. Since source signal components are usually more slowly varying than any nonlinear mixture of them we consider to require the estimated sources to be as slowly varying as possible. This can be achieved by incorporating ideas from Slow Feature Analysis (SFA) [Wiskott and Sejnowski, 2002] into ICA.

After a short introduction to linear BSS, nonlinear BSS and SFA we will show a way how to combine SFA and ICA to obtain an algorithm that solves the nonlinear BSS problem.

2 Linear Blind Source Separation

Let $\mathbf{x}(t) = [x_1(t), \dots, x_N(t)]^T$ be a linear mixture of a source signal $\mathbf{s}(t) = [s_1(t), \dots, s_N(t)]^T$ and defined by

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

with an invertible $N \times N$ mixing matrix \mathbf{A} . Finding a mapping

$$\mathbf{u}(t) = \mathbf{Q}\mathbf{W}\mathbf{x}(t), \quad (2)$$

This work has been supported by the Volkswagen Foundation through a grant to LW for a junior research group.

such that the components of \mathbf{u} are mutually statistically independent is called Independent Component Analysis (ICA). The mapping is often divided into a whitening mapping \mathbf{W} , resulting in uncorrelated signal components y_i with unit variance, and a successive orthogonal transformation \mathbf{Q} , because one can show [Comon, 1994] that after whitening an orthogonal transformation is sufficient to obtain independence. It is well known that ICA solves the linear BSS problem [Comon, 1994]. There exists a variety of algorithms performing ICA and therefore BSS (see e.g. [Cardoso and Souloumiac, 1993; Lee et al., 1999; Hyvärinen, 1999]). Here we focus on a method using only second-order statistics introduced by Molgedey and Schuster [1994]. The method consists of optimizing an objective function, subject to minimization, which can be written as

$$\Psi_{\text{ICA}}(\mathbf{Q}) = \sum_{\substack{\alpha, \beta=1 \\ \alpha \neq \beta}}^N \left(C_{\alpha\beta}^{(\mathbf{u})}(\tau) \right)^2 = \sum_{\substack{\alpha, \beta=1 \\ \alpha \neq \beta}}^N \left(\sum_{\gamma, \delta=1}^N Q_{\alpha\gamma} Q_{\beta\delta} C_{\gamma\delta}^{(\mathbf{y})}(\tau) \right)^2, \quad (3)$$

operating on the already whitened signal \mathbf{y} . $C_{\gamma\delta}^{(\mathbf{y})}(\tau)$ is an entry of a symmetrized time delayed covariance matrix defined by

$$\mathbf{C}^{(\mathbf{y})}(\tau) = \left\langle \mathbf{y}(t) \mathbf{y}(t+\tau)^T + \mathbf{y}(t+\tau) \mathbf{y}(t)^T \right\rangle, \quad (4)$$

and $\mathbf{C}^{(\mathbf{u})}(\tau)$ is defined correspondingly. $Q_{\alpha\beta}$ denotes an entry of \mathbf{Q} . Minimization of Ψ_{ICA} can be understood intuitively as finding an orthogonal matrix \mathbf{Q} that diagonalizes the covariance matrix with time delay τ . Since, because of the whitening, the instantaneous covariance matrix is already diagonal this results in signal components that are decorrelated instantaneously and at a given time delay τ . This can be sufficient to achieve statistical independence [Tong et al., 1991].

2.1 Nonlinear BSS and ICA

An obvious extension to the linear mixing model (1) has the form

$$\mathbf{x}(t) = F(\mathbf{s}(t)), \quad (5)$$

with a function $F(\cdot) : \mathbb{R}^N \rightarrow \mathbb{R}^M$, that maps N -dimensional source vectors \mathbf{s} onto M -dimensional signal vectors \mathbf{x} . The components x_i of the observable are a nonlinear mixture of the sources and like in the linear case source signal components s_i are assumed to be mutually statistically independent. Unmixing is in general only possible if $F(\cdot)$ is an invertible function, which we will assume from now on.

The equivalence of BSS and ICA in the linear case does in general not hold for a nonlinear function $F(\cdot)$ [Hyvärinen and Pajunen, 1999; Jutten and Karhunen, 2003]. To solve the nonlinear BSS problem additional constraints on the mixture or the estimated signals are needed to bridge the gap between ICA and BSS. Here we propose a new way to achieve this by adding a slowness objective to the independence objective of pure ICA. Assume for example a sinusoidal signal component $x_i = \sin(2\pi t)$ and a second component that is the square of the first $x_j = x_i^2 = 0.5(1 - \cos(4\pi t))$ is given. The

second component is more quickly varying due to the frequency doubling induced by the squaring. Typically nonlinear mixtures of signal components are more quickly varying than the original components. To extract the right source components one should therefore prefer the slowly varying ones. The concept of slowness is used in our approach to nonlinear BSS by combining an ICA part that provides the independence of the estimated source signal components with a part that prefers slowly varying signals over more quickly varying ones. In the next section we will give a short introduction to Slow Feature Analysis building the basis of the second part of our method.

3 Slow Feature Analysis

Assume a vectorial input signal $\mathbf{x}(t) = [x_1(t), \dots, x_M(t)]^T$ is given. The objective of SFA is to find an in general nonlinear input-output function $\mathbf{u}(t) = \mathbf{g}(\mathbf{x}(t))$ with $\mathbf{g}(\mathbf{x}(t)) = [g_1(\mathbf{x}(t)), \dots, g_R(\mathbf{x}(t))]^T$ such that the $u_i(t)$ are varying as slowly as possible. This can be achieved by successively minimizing the objective function

$$\Delta(u_i) := \langle \dot{u}_i^2 \rangle, \quad (6)$$

for each u_i under the constraints

$$\langle u_i \rangle = 0 \quad (\text{zero mean}), \quad (7)$$

$$\langle u_i^2 \rangle = 1 \quad (\text{unit variance}), \quad (8)$$

$$\langle u_i u_j \rangle = 0 \quad \forall j < i \quad (\text{decorrelation and order}). \quad (9)$$

Constraints (7) and (8) ensure that the solution will not be the trivial solution $u_i = \text{const}$. Constraint (9) provides uncorrelated output signal components and thus guarantees that different components carry different information. Intuitively we are searching for signal components u_i that have on average a small slope.

Interestingly Slow Feature Analysis (SFA) can be reformulated with an objective function similar to second-order ICA, subject to maximization [Blaschke et al., 2004],

$$\Psi_{\text{SFA}}(\mathbf{Q}) = \sum_{\alpha=1}^M \left(C_{\alpha\alpha}^{(\mathbf{u})}(\tau) \right)^2 = \sum_{\alpha=1}^M \left(\sum_{\beta,\gamma=1}^M Q_{\alpha\beta} Q_{\alpha\gamma} C_{\beta\gamma}^{(\mathbf{y})}(\tau) \right)^2. \quad (10)$$

To understand (10) intuitively we notice that slowly varying signal components are easier to predict, and should therefore have strong auto correlations in time. Thus, maximizing the time delayed variances produces slowly varying signal components.

4 Independent Slow Feature Analysis

If we combine ICA and SFA we obtain a method, we refer to as Independent Slow Feature Analysis (ISFA), that recovers independent components out of a nonlinear mixture using a combination of SFA and second-order ICA. As already explained, second-order ICA tends to make the output components independent and SFA tends to make them

slow. Since we are dealing with a nonlinear mixture we first compute a nonlinearly expanded signal $\mathbf{z} = \mathbf{h}(\mathbf{x})$ with $\mathbf{h}(\cdot) : \mathbb{R}^M \rightarrow \mathbb{R}^L$ being typically monomials up to a given degree, e.g. an expansion with monomials up to second degree can be written as

$$\mathbf{h}(\mathbf{x}(t)) = [x_1, \dots, x_N, x_1x_1, x_1x_2, \dots, x_Mx_M]^T - \mathbf{h}_0^T, \quad (11)$$

when given an M -dimensional signal \mathbf{x} . The constant vector \mathbf{h}_0^T is used to make the expanded signal mean free. In a second step \mathbf{z} is whitened to obtain $\mathbf{y} = \mathbf{Wz}$. Thirdly we apply linear ICA combined with linear SFA on \mathbf{y} in order to find the estimated source signal \mathbf{u} . Because of the whitening we know that ISFA, like ICA and SFA, is solved by finding an orthogonal $L \times L$ matrix \mathbf{Q} . We write the estimated source signal \mathbf{u} as

$$\mathbf{v} = \begin{pmatrix} \mathbf{u} \\ \tilde{\mathbf{u}} \end{pmatrix} = \mathbf{Qy} = \mathbf{QWz} = \mathbf{QWh}(\mathbf{x}), \quad (12)$$

where we introduced $\tilde{\mathbf{u}}$ since R , the dimension of the estimated source signal \mathbf{u} , is usually much smaller than L , the dimension of the expanded signal. While the u_i are statistically independent and slowly varying the components \tilde{u}_i are more quickly varying and may be statistically dependent on each other as well as on the selected components.

To summarize, we have an M dimensional input \mathbf{x} an L dimensional nonlinearly expanded and whitened \mathbf{y} and an R dimensional estimated source signal \mathbf{u} . ISFA searches an R dimensional subspace such that the u_i are independent and slowly varying. This is achieved at the expense of all \tilde{u}_i .

4.1 Objective function

To recover R source signal components u_i $i = 1, \dots, R$ out of an L -dimensional expanded and whitened signal \mathbf{y} the objective reads

$$\Psi_{\text{ISFA}}(u_1, \dots, u_R; \tau) = b_{\text{ICA}} \sum_{\substack{\alpha, \beta=1, \\ \alpha \neq \beta}}^R \left(C_{\alpha\beta}^{(\mathbf{u})}(\tau) \right)^2 - b_{\text{SFA}} \sum_{\alpha=1}^R \left(C_{\alpha\alpha}^{(\mathbf{u})}(\tau) \right)^2, \quad (13)$$

where we simply combine the ICA objective (3) and SFA objective (10) weighted by the factors b_{ICA} and b_{SFA} , respectively. Note that the ICA objective is usually applied to the linear case to unmix the linear whitened mixture \mathbf{y} whereas here it is used on the nonlinearly expanded whitened signal $\mathbf{y} = \mathbf{Wz}$. ISFA tries to minimize Ψ_{ISFA} which is the reason why the SFA part has a negative sign.

4.2 Optimization Procedure

From (12) we know that $\mathbf{C}^{(\mathbf{u})}(\tau)$ in (13) depends on the orthogonal matrix \mathbf{Q} . There are several ways to find the orthogonal matrix that minimizes the objective function. Here we apply successive Givens rotations to obtain \mathbf{Q} . A Givens rotation $\mathbf{Q}^{\mu\nu}$ is a rotation around the origin within the plane of two selected components μ and ν and has the matrix form

$$Q_{\alpha\beta}^{\mu\nu} := \begin{cases} \cos(\phi) & \text{for } (\alpha, \beta) \in \{(\mu, \mu), (\nu, \nu)\} \\ -\sin(\phi) & \text{for } (\alpha, \beta) \in \{(\mu, \nu)\} \\ \sin(\phi) & \text{for } (\alpha, \beta) \in \{(\nu, \mu)\} \\ \delta_{\alpha\beta} & \text{otherwise} \end{cases} \quad (14)$$

with Kronecker symbol $\delta_{\alpha\beta}$ and rotation angle ϕ . Any orthogonal $L \times L$ matrix such as \mathbf{Q} can be written as a product of $\frac{L(L-1)}{2}$ (or more) Givens rotation matrices $\mathbf{Q}^{\mu\nu}$ (for the rotation part) and a diagonal matrix with elements ± 1 (for the reflection part). Since reflections do not matter in our case we only consider the Givens rotations as is often used in second-order ICA algorithms (see e.g. [Cardoso and Souloumiac, 1996]).

We can therefore write the objective as a function of a Givens rotation $\mathbf{Q}^{\mu\nu}$ as

$$\Psi_{\text{ISFA}}(\mathbf{Q}^{\mu\nu}) = b_{\text{ICA}} \sum_{\substack{\alpha, \beta=1, \\ \alpha \neq \beta}}^R \left(\sum_{\substack{\gamma, \delta=1 \\ \gamma \neq \delta}}^L \mathcal{Q}_{\alpha\gamma}^{\mu\nu} \mathcal{Q}_{\beta\delta}^{\mu\nu} C_{\gamma\delta}^{(y)}(\tau) \right)^2 - b_{\text{SFA}} \sum_{\alpha=1}^R \left(\sum_{\beta, \gamma=1}^L \mathcal{Q}_{\alpha\beta}^{\mu\nu} \mathcal{Q}_{\alpha\gamma}^{\mu\nu} C_{\beta\gamma}^{(y)}(\tau) \right)^2, \quad (15)$$

Assume we want to minimize Ψ_{ISFA} for a given R , where R denotes the number of signal components we want to unmix. Applying a Givens rotation $\mathbf{Q}^{\mu\nu}$ we have to distinguish three cases

- **Case 1** Both axes u_μ and u_ν lie inside the subspace spanned by the first R axes ($\mu, \nu \leq R$): The sum over all squared cross correlations of all signal components that lie outside the subspace is constant as well as those of all signal components inside the subspace. There is no interaction between inside and outside, in fact the objective function is exactly the objective for an ICA algorithm based on second-order statistics e.g. TDSEP or SOBI [Ziehe and Müller, 1998; Belouchrani et al., 1997]. In [Blaschke et al., 2004] it has been shown that this is equivalent to SFA in the case of a single time delay.
- **Case 2** Only one axis, w.l.o.g. u_μ , lies inside the subspace, the other, u_ν , outside ($\mu \leq R < \nu$): Since one axis of the rotation plane lies outside the subspace, u_μ in the objective function can be optimized at the expense of u_ν outside the subspace. A rotation of $\pi/2$, for instance would simply exchange components u_μ and u_ν . This gives the possibility to find the slowest and most independent components in the whole space spanned by all u_i and u_j ($i = 1, \dots, R$, $j = R + 1, \dots, L$) in contrast to Case 1 where the minimum is searched within the subspace spanned by the R components in the objective function.
- **Case 3** Both axes lie outside the subspace ($R < \mu, \nu$): A Givens rotation with the two rotation axes outside the relevant subspace does not affect the objective function, and can therefore be disregarded.

It can be shown that like in [Blaschke and Wiskott, 2004] the objective function (15) as a function of ϕ can always be written in the form

$$\Psi_{\text{ISFA}}^{\mu\nu}(\phi) = A_0 + A_2 \cos(2\phi + \phi_2) + A_4 \cos(4\phi + \phi_4), \quad (16)$$

where the second term on the right hand side vanishes for Case 1. There exists a single minimum (if w.l.o.g. $\phi \in [-\frac{\pi}{2}, \frac{\pi}{2}]$) that can easily be calculated (see e.g. [Blaschke and Wiskott, 2004]). The derivation of (16) involves various trigonometric identities and, because of its length, is documented elsewhere¹.

¹ <http://itb.biologie.hu-berlin.de/~blaschke>

It is important to notice that the rotation planes of the Givens rotations are selected from the whole L -dimensional space whereas the objective function only uses information of correlations among the first R signal components u_i .

Successive application of Givens rotations \mathbf{Q}^{uv} leads to the final rotation matrix \mathbf{Q} which is in the ideal case such that $\mathbf{Q}^T \mathbf{C}^{(y)}(\tau) \mathbf{Q} = \mathbf{C}^{(y)}(\tau)$ has a diagonal $R \times R$ sub-matrix $\mathbf{C}^{(u)}(\tau)$, but it is not clear if the final minimum is also the global one. However, in various simulations no local minima have been found.

4.3 Incremental Extracting of Independent Components

It is possible to find the number of independent source signal components R by successively increasing the number of components to be extracted. In each step the objective function (13) is optimized for a fixed R . First a single signal component is extracted ($R = 1$) and then an additional one ($R = 2$) etc. The algorithm is stopped when no additional signal component can be extracted. As a stopping criterion every suitable measure of independence can be applied; we used the sum over squared cross-cumulants of fourth order. In our artificial examples, this value is typically small for independent components, and increases by two orders of magnitudes if the number of components to be extracted is greater than the number of original source signal components.

5 Simulation

Here we show a simple example, with two nonlinearly mixed signal components as shown in Figure 1. For comparison we chose a mixture from [Harmeling et al., 2003] defined by

$$\begin{aligned} x_1(t) &= (s_2(t) + 3s_1(t) + 6) \cos(1.5\pi s_1(t)), \\ x_2(t) &= (s_2(t) + 3s_1(t) + 6) \sin(1.5\pi s_1(t)). \end{aligned} \quad (17)$$

We used the ISFA algorithm with different nonlinearities (see Tab. 1). Already a nonlinear expansion with monomials up to degree three was sufficient to give good unmixing results. In all cases ISFA did find exactly two independent signal components. Using all monomials up to degree five led to results that showed virtually no difference between estimated and true source signal (see Fig. 1). A linear BSS method failed completely to find a good unmixing matrix.

6 Conclusion

We have shown that connecting the ideas of slow feature analysis and independent component analysis into ISFA is a possible way to solve the nonlinear blind source separation problem. SFA enforces the independent components of ICA to be slowly varying which seems to be a good way to discriminate between the original and nonlinearly distorted source signal components. A simple simulation showed that ISFA is able to extract the original source signal out of a nonlinear mixture. Furthermore ISFA can predict the number of source signal components via an incremental optimization scheme.

Table 1. Correlation coefficients of extracted (u_1 and u_2) and original (s_1 and s_2) source signal components

	linear		degree 2		degree 3		degree 4		degree 5		kTDSEP	
	u_1	u_2	u_1	u_2	u_1	u_2	u_1	u_2	u_1	u_2	u_1	u_2
s_1	-0.890	0.215	0.936	0.013	0.001	0.988	0.002	-0.996	0.998	-0.000	0.990	-
s_2	-0.011	-0.065	-0.027	0.149	-0.977	0.006	0.983	-0.000	-0.000	0.994	-	0.947

Correlation coefficients of extracted (u_1 and u_2) and original (s_1 and s_2) source signal components for linear ICA (first column) and ISFA with different nonlinearities (monomials up to degree 2,3,4, and 5). Note, that the source signal can only be estimated up to permutation and scaling, resulting in different signs and permutations of u_1 and u_2 . The correlation coefficients for kTDSEP were taken from [Harmeling et al., 2003] with same mixture but different source signal.

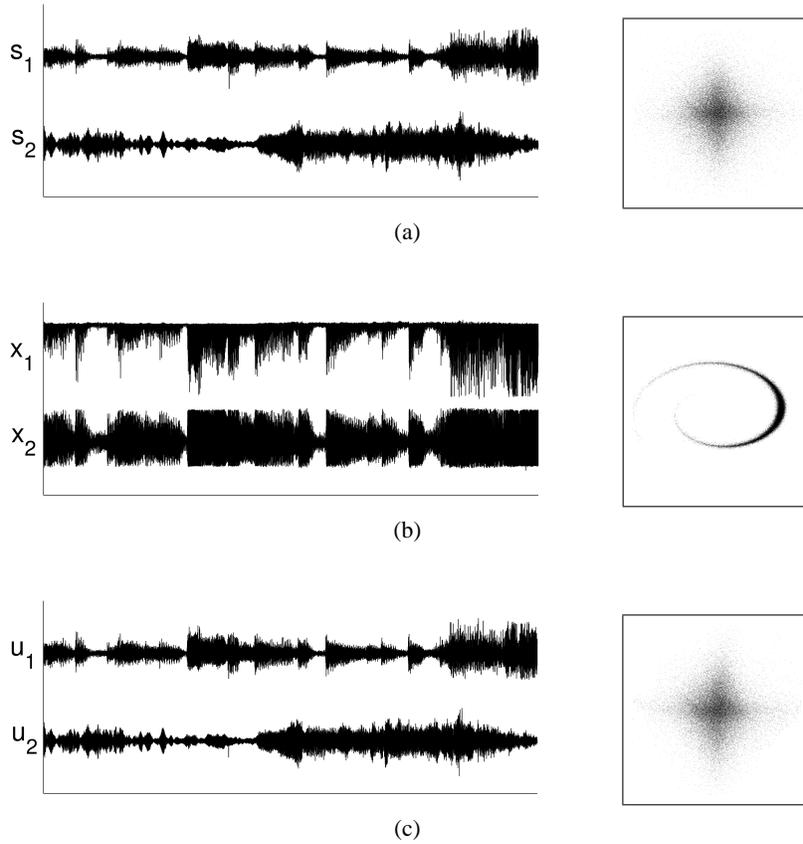


Fig. 1. Waveforms and Scatter-plots of (a) the original source signal components s_i , (b) the nonlinear mixture, and (c) recovered components with nonlinear ISFA (u_i). As a nonlinearity we used all monomials up to degree 5.

Bibliography

- Belouchrani, A., Abed Meraim, K., Cardoso, J.-F., and Éric Moulines (1997). **A blind source separation technique based on second order statistics**. *IEEE Transactions on Signal Processing*, 45(2):434–44.
- Blaschke, T. and Wiskott, L. (2004). **CuBICA: Independent component analysis by simultaneous third- and fourth-order cumulant diagonalization**. *IEEE Transactions on Signal Processing*, 52(5).
- Blaschke, T., Wiskott, L., and Berkes, P. (2004). What is the relation between independent component analysis and slow feature analysis? *in preparation*.
- Cardoso, J.-F. and Souloumiac, A. (1993). **Blind beamforming for non Gaussian signals**. *IEE Proceedings-F*, 140:362–370.
- Cardoso, J.-F. and Souloumiac, A. (1996). **Jacobi angles for simultaneous diagonalization**. *SIAM J. Mat. Anal. Appl.*, 17(1):161–164.
- Comon, P. (1994). **Independent component analysis, a new concept?** *Signal Processing, Elsevier*, 36(3):287–314. Special issue on Higher-Order Statistics.
- Harmeling, S., Ziehe, A., Kawanabe, M., and Müller, K.-R. (2003). Kernel-based nonlinear blind source separation. *Neural Computation*, 15:1089–1124.
- Hyvärinen, A. (1999). **Fast and robust fixed-point algorithms for independent component analysis**. *IEEE Transactions on Neural Networks*, 10(3):626–634.
- Hyvärinen, A. and Pajunen, P. (1999). **Nonlinear independent component analysis: existence and uniqueness results**. *Neural Networks*, 12(3):429–439.
- Jutten, C. and Karhunen, J. (2003). **Advances in nonlinear blind source separation**. In *Proc. of the 4th Int. Symp. on Independent Component Analysis and Blind Signal Separation (ICA2003)*, pages 245–256, Nara, Japan.
- Lee, T.-W., Girolami, M., and Sejnowski, T. (1999). **Independent component analysis using an extended Infomax algorithm for mixed sub-Gaussian and super-Gaussian sources**. *Neural Computation*, 11(2):409–433.
- Molgedey, L. and Schuster, G. (1994). **Separation of a mixture of independent signals using time delayed correlations**. *Physical Review Letters*, 72(23):3634–3637.
- Tong, L., Liu, R., Soon, V. C., and Huang, Y.-F. (1991). **Indeterminacy and identifiability of blind identification**. *IEEE Transactions on Circuits and Systems*, 38(5).
- Wiskott, L. and Sejnowski, T. (2002). **Slow feature analysis: Unsupervised learning of invariances**. *Neural Computation*, 14(4):715–770.
- Ziehe, A. and Müller, K.-R. (1998). **TDSEP – an efficient algorithm for blind separation using time structure**. In *8th International Conference on Artificial Neural Networks (ICANN'98)*, pages 675 – 680, Berlin. Springer Verlag.