Nonlinear Independent Component Analysis: Existence and Uniqueness Results

Aapo Hyvärinen* and Petteri Pajunen

Helsinki University of Technology

Laboratory of Computer and Information Science

P.O. Box 5400, FIN-02015 HUT, Finland

Email: aapo.hyvarinen,petteri.pajunen@hut.fi

Tel: +358-9-4513278, Fax: +358-9-4513277

September 21, 1998

Running title: Nonlinear ICA

Revised submission to *Neural Networks*

*Corresponding author

Nonlinear Independent Component Analysis: Existence and Uniqueness Results

Abstract

The question of existence and uniqueness of solutions for nonlinear independent component analysis is addressed. It is shown that if the space of mixing functions is not limited, there exists always an infinity of solutions. In particular, it is shown how to construct parameterized families of solutions. The indeterminacies involved are not trivial, as in the linear case. Next, it is shown how to utilize some results of complex analysis to obtain uniqueness of solutions. We show that for two dimensions, the solution is unique up to a rotation, if the mixing function is constrained to be a conformal mapping, together with some other assumptions. We also conjecture that the solution is strictly unique except in some degenerate cases, since the indeterminacy implied by the rotation is essentially similar to estimating the model of linear independent component analysis.

Keywords: independent component analysis, blind source separation, redundancy reduction, feature extraction.

1 Introduction

Independent Component Analysis (ICA) (Comon, 1994; Jutten and Herault, 1991) is a statistical technique whose goal is to represent a set of random variables as linear functions of statistically independent component variables. ICA can be applied to blind source separation (Jutten and Herault, 1991) as well as to feature extraction (Bell and Sejnowski, 1997; Olshausen and Field, 1996). More generally, it has been proposed that the goal of sensory coding is redundancy reduction (Barlow, 1961; Barlow, 1972). This corresponds to obtaining a factorial code of the observed data or equivalently, a representation with independent components, as in ICA. A nonlinear factorial code or nonlinear ICA can be formulated as the estimation of the following generative model for the data:

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

where $\mathbf{x} = [x_1, x_2, \ldots, x_n]^T$ is the vector of observed random variables, $\mathbf{s} = [s_1, s_2, \ldots, s_n]^T$ is the vector of the latent variables called the independent components, and $\mathbf{f}$ is an unknown function from $\mathbb{R}^n$ to $\mathbb{R}^n$. (The dimensions of $\mathbf{x}$ and $\mathbf{s}$ need not be equal, but we make this assumption here for simplicity.) For a linear $\mathbf{f}$, the above definition coincides with linear ICA. The fundamental assumption is that the components s_i are mutually statistically independent. The basic problem of ICA is then to estimate the realizations of the independent components s_i together with the mixing function $\mathbf{f}$, using only observations of the mixtures x_j.

In most cases, it is assumed that the mixing function $\mathbf{f}$ is linear (Comon, 1994; Jutten and Herault, 1991). Then a solution to the ICA estimation problem exists, and this solution is unique up to some trivial indeterminacies (permutation and multiplication of the s_i by constants). To obtain this result, it is also necessary to assume that the components s_i, except perhaps one, are non-Gaussian, and that the data follows the linear generative model (Comon, 1994). In other words, the linear generative model is essentially identifiable under these assumptions.

The purpose of this paper is to investigate the existence and uniqueness of nonlinear solutions of the problem of decomposing a random vector nonlinearly into components that are statistically independent. This can be interpreted as estimation of the nonlinear ICA model in Eq. (1). To our knowledge, these problems have not been treated in the literature, although some algorithms for nonlinear ICA have already been proposed (Burel, 1992; Deco and Brauer, 1995; Deco and Obradovic, 1995; Lee et al., 1997; Pajunen et al., 1996; Pajunen and Karhunen, 1997; Yang et al., 1998). In a special case, the problem has been treated in (Taleb and Jutten, 1997). The questions of existence and uniqueness of solutions are, however, of fundamental importance even in the construction of algorithms for nonlinear ICA.

We present two results in this paper. In Section 2, we show explicitly how to construct a function $\mathbf{g}$ from $\mathbb{R}^n$ to $\mathbb{R}^n$ so that the components of $\mathbf{y} = \mathbf{g}(\mathbf{x})$ are independent. We also show using this construction that such a decomposition into independent components is by no means unique in general. In Section 3, we show that for $n = 2$, the solution is unique up to a rotation, if the mixing function $\mathbf{f}$ is constrained to be a conformal mapping (Ahlfors, 1979), together with some other assumptions. We also conjecture that the solution is strictly unique except in some degenerate cases, since the indeterminacy implied by the rotation is essentially similar to solving the linear ICA problem. Section 4 gives some simulation results and Section 5 discusses the significance of the results.

2 Existence

2.1 Construction of a solution

In this subsection, we show constructively that the nonlinear ICA problem always has at least one solution. That is, given a random vector $\mathbf{x}$, there is always a function $\mathbf{g}$ so that the components of $\mathbf{y} = [y_1, \ldots, y_n]^T$ given by $\mathbf{y} = \mathbf{g}(\mathbf{x})$ are independent. In the next subsection, we show that this solution is highly non-unique.

The construction used here might be considered as a generalization of Gram-Schmidt orthogonalization. Given m independent variables $y_1, \ldots, y_m$, and a variable x, one constructs a new variable $y_{m+1} = g(y_1, \ldots, y_m, x)$ so that the set $y_1, \ldots, y_{m+1}$ is mutually independent.

The construction is defined recursively as follows. Assume that we have already m independent random variables $y_1, \ldots, y_m$ which follow a joint uniform distribution in $[0,1]^m$. (It is not a restriction to assume that the distributions of the y_i are uniform: this follows directly from the recursion, as will be seen below.) Denote by x any random variable, and by $a_1, \ldots, a_m, b$ some non-random scalars. Define

$$g(a_1, \ldots, a_m, b; p_{y,x}) = P(x \leq b | y_1 = a_1, \ldots, y_m = a_m) \tag{2}$$
$$= \frac{\int_{-\infty}^{b} p_{y,x}(a_1, \ldots, a_m, \xi) d\xi}{p_y(a_1, \ldots, a_m)}$$

where $p_y(\cdot)$ and $p_{y,x}(\cdot)$ are the (marginal) probability densities of $(y_1, \ldots, y_m)$ and $(y_1, \ldots, y_m, x)$, respectively (it is assumed here implicitly that such densities exist), and $P(\cdot|\cdot)$ denotes the conditional probability. The $p_{y,x}$ in the argument of g is to remind that g depends on the joint probability distribution of $y_1, \ldots, y_m$ and x. For $m = 0$, g is simply the cumulative distribution function of x. Now, g as defined above gives a nonlinear decomposition, as stated in the following theorem:

Theorem 1 *Assume that $y_1, \ldots, y_m$ are independent scalar random variables which follow a joint uniform distribution in the unit cube $[0,1]^m$. Let x be any scalar random variable (such that the joint distribution of $y_1, \ldots, y_m, x$ has a probability density with respect to the Lebesgue measure of $\mathbb{R}^{m+1}$). Define g as in (2), and set*

$$y_{m+1} = g(y_1, \ldots, y_m, x; p_{y,x}). \tag{3}$$

Then y_{m+1} is independent from the $y_1, \ldots, y_m$. In particular, the variables $y_1, \ldots, y_{m+1}$ are jointly uniformly distributed in the unit cube $[0,1]^{m+1}$.

Proof. Denote by

$$F(v_1, \ldots, v_m, \xi) = (v_1, \ldots, v_m, g(v_1, \ldots, v_m, \xi; p_{y,x})) \tag{4}$$

the transformation made on the vector $(y_1, \ldots, y_m, x)$ to obtain $(y_1, \ldots, y_{m+1})$. The Jacobian of this transformation equals

$$JF(v_1, \ldots, v_m, \xi) = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \\ c_1 & c_2 & \ldots & K \end{bmatrix} \tag{5}$$

where $c_1, c_2, \ldots$ are some irrelevant quantities, and

$$K = \frac{p_{y,x}(v_1, \ldots, v_m, \xi)}{p_y(v_1, \ldots, v_m)}. \tag{6}$$

The determinant of JF equals K. Thus, one obtains the density p_{y+} of the vector $(y_1, \ldots, y_m, y_{m+1})$ as

$$\begin{aligned} p_{y+}(v_1, \ldots, v_{m+1}) &= p_{y,x}(v_1, \ldots, v_m, \xi) \left[\frac{p_{y,x}(v_1, \ldots, v_m, \xi)}{p_y(v_1, \ldots, v_m)}\right]^{-1} \\ &= p_y(v_1, \ldots, v_m) \end{aligned} \tag{7}$$

From (2) it follows that $y_{m+1} \in [0, 1]$. Thus (7) implies that p_{y+} is a uniform density in $[0, 1]^{m+1}$, which implies that the $y_1, \ldots, y_{m+1}$ are mutually independent (Pajunen et al., 1996). $\square$

The construction can obviously be used to decompose n variables $x_1, \ldots, x_n$ into n independent components $y_1, \ldots, y_n$. Let m run from 0 to $n-1$, and obtain y_{m+1} using the above construction, where the observed variable x_{m+1} is used as x. In other words, let $y_{m+1} = g(y_1, \ldots, y_m, x_{m+1}; p_{y,x_{m+1}})$ for $m = 0, \ldots, n-1$. Using this Gram-Schmidt-like recursion, one obtains $y_1, \ldots, y_n$ which are mutually independent, and give a solution for the nonlinear ICA problem. The vector $\mathbf{y}$ could of course be expressed using a single function $\mathbf{g}$ as $\mathbf{y} = \mathbf{g}(\mathbf{x})$, but the closed-form expression for $\mathbf{g}$ would be very complicated. A related result in the two-dimensional case was given in (Darmois, 1951).

Note that our construction, when defined, does not depend on the number of the original independent components s_i in (1). In particular, the number of the original independent components could be larger than the dimension n of $\mathbf{x}$. The above construction would always give, however, exactly n variables $y_1, i = 1...n$. In contrast, if the number of independent components were smaller than n, the density of $\mathbf{x}$ would be degenerate, and the construction would not be defined.

2.2 Non-uniqueness of solutions

In the previous section a bijection was constructed which is a solution to nonlinear ICA. The mapping $\mathbf{g}$ transforms any random vector $\mathbf{x}$ into a uniformly distributed random vector $\mathbf{y} = \mathbf{g}(\mathbf{x})$. This construction also clearly shows that the decomposition in independent components is by no means unique. For example, we may apply first a linear transformation on the $\mathbf{x}$ to obtain a random vector $\mathbf{x}' = \mathbf{M}\mathbf{x}$, and then compute $\mathbf{y}' = \mathbf{g}'(\mathbf{x}')$ with $\mathbf{g}'$ being defined using the procedure given above, where $\mathbf{x}$ is replaced by $\mathbf{x}'$. Thus we obtain another decomposition of $\mathbf{x}$ into independent components. The resulting decomposition $\mathbf{y}' = \mathbf{g}'(\mathbf{M}\mathbf{x})$ is, in general, different from $\mathbf{y}$, and cannot be reduced to $\mathbf{y}$ by any simple transformations. Parameterizing the family of initial transformations (e.g., linear transformations), we obtain parametrized families of nonlinear ICA decompositions.

More rigorously, to show that the above construction is not a unique solution to nonlinear ICA, it is sufficient to construct a class of automorphisms, i.e. mappings $\mathbf{h} : [0,1]^n \rightarrow [0,1]^n$, that are measure-preserving, i.e., do not change the probability distribution of a random variable distributed uniformly in the unit cube. Then forming a mapping $\mathbf{h} \circ \mathbf{g}$ we get another solution to nonlinear ICA. Thus it can be seen that the class of measure-preserving automorphisms on $[0,1]^n$ determines the indeterminacy of the solutions to nonlinear ICA.

A concrete example of a measure-preserving automorphism in two dimensions is obtained by considering the two-dimensional random variable $\mathbf{y}$ as a complex random variable z, and defining the following mapping:

$$\mathbf{h}(z) = \begin{cases} z, & |z| > c; \\ z \exp(jf(|z|)), & |z| \leq c. \end{cases}$$

where the continuous scalar function f must fulfill $f(c) = 0$. The constant must be chosen so that $0 < c < 1$. Then h is continuous and measure-preserving. Since f and c can be chosen freely, this results in a large class of indeterminacies in the nonlinear ICA solutions. Clearly, the indeterminacy implied by the above function $\mathbf{h}$ is not trivial.

From the viewpoint of estimating the generative model (1), we thus we see that the constraint of independency of the s_i alone is not sufficient to make the elements in the generative model identifiable.

Finally, we would like to point out a simpler form of non-uniqueness: if s_1 and s_2 are independent, then any component-wise transformations $f_1(s_1)$ and $f_2(s_2)$ are also independent. This trivial indeterminacy can be compared with the scale indeterminacy in linear ICA, and is of little consequence compared to the more fundamental non-uniqueness shown above.

3 Uniqueness

3.1 Assumptions

As seen above, the solution of the nonlinear ICA problem is, in general, highly non-unique. In this section we show that under some restrictions, results on uniqueness can be obtained. Assume the following:

1. The dimension of the problem equals two, i.e., $n = 2$. This allows us to consider the data as complex variables $z = x_1 + ix_2$.

2. The mixing function $\mathbf{f}$ is a conformal mapping (Ahlfors, 1979; Churchill and Brown, 1990) (see below), and zero preserving, i.e. $\mathbf{f}(\mathbf{0}) = \mathbf{0}$.

3. The densities of the independent components have bounded support, and the supports are known.

Under these assumptions it can be shown that the ICA problem can be solved up to a transformation which is essentially a rotation. Moreover, we conjecture that it can be solved uniquely in most cases.

A bijective complex mapping is called conformal if it is analytic, i.e. the complex derivative exists, and its derivative is non-zero everywhere. Conformal mappings have applications in electrostatics, steady fluid flows, and in general all physical problems where Laplace equation appears, since a conformal map preserves solutions to Laplace equation. Intuitively, a conformal mapping can be thought of as a one-to-one nonlinear mapping which preserves the orthogonality of coordinates locally (Churchill and Brown, 1990).

As for the assumption of bounded support of densities, it is quite realistic in some applications, for example telecommunications, where the uniform distribution is often used to model the densities of the independent components (Cardoso and Laheld, 1996).

3.2 Using properties of conformal mappings

First, we prove that the nonlinear ICA problem can be solved up to a rotation. We observe a mixture vector $\mathbf{x} = \mathbf{f}(\mathbf{s})$ where $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$ is the conformal mixing mapping. Now choose a zero-preserving bijection $\mathbf{g}_0$ that is a conformal mapping, and maps the support of the densities of the observed variables to S, where S denotes the rectangular support of the joint density of the independent components: $\mathbf{g}_0 : \mathbf{f}(S) \to S$. Such a function can be approximated using the Schwarz-Christoffel transformation (Churchill and Brown, 1990), which is an integral formula that gives a conformal mapping from the upper half-plane onto a given polygon; the computation of the transformation usually requires numerical methods (Driscoll, 1996). The support of the mixture densities, on the other hand, can be estimated from the data. Now $\mathbf{h} = \mathbf{g}_0 \circ \mathbf{f}$ is a bijection $\mathbf{h} : S \to S$ which preserves zero. As a consequence of the Riemann mapping theorem, there exists a conformal mapping $\mathbf{r}$ which maps S onto the unit disk while preserving zero. Such a transformation can be found as a special case of the Schwarz-Christoffel transformation as well (Churchill and Brown, 1990; Driscoll, 1996). We can use this to construct $\mathbf{h}' = \mathbf{r} \circ \mathbf{h} \circ \mathbf{r}^{-1}$ which maps the unit disk onto itself and preserves the zero. To this mapping we can apply the following result, known as the Schwartz lemma (Ahlfors, 1979):

Theorem 2 *Let* $\mathbf{p}$ *be a conformal one-to-one mapping of the unit disk onto itself with a fixed point at* $\mathbf{0}$*. Then* $\mathbf{p}(r \exp(j\phi)) = r \exp(j(\phi + \phi_0))$*, i.e.* $\mathbf{p}$ *is a rotation.*

This implies that $\mathbf{h}'$ is a rotation in the unit circle, and that $\mathbf{h} = \mathbf{r}^{-1} \circ \mathbf{h}' \circ \mathbf{r}$ is a rotation transformed onto the rectangle S. Thus, using only the information on the bounds of the values of the independent component, we can determine the inverse transformation up to the function $\mathbf{h}$, which is essentially a rotation by the angle ϕ_0. This completes the proof.

3.3 Using independence

In the above proof, the independence of the s_i was not utilized at all. We only used the properties of the mixing function, together with the knowledge of the supports of the densities of the independent components, and an estimate of the support of the joint density of the mixtures. Taking the independence into account reduces the indeterminacy even further, and we conjecture that it can be completely eliminated in all but a few degenerate cases.

Since the indeterminacy left is essentially a rotation, it is enough to find the inverse rotation to obtain the independent components. Define

$$\mathbf{g}_\alpha = \mathbf{r}^{-1} \circ \mathrm{rot}_\alpha \circ \mathbf{r} \circ \mathbf{g}_0 \tag{8}$$

where rot_α means rotation with angle α, and $\mathbf{g}_0$ and $\mathbf{r}$ are as defined above (e.g., those given by the Schwarz-Christoffel transformation). The results in the preceding subsection imply that there exists an α_0 such that

$$\mathbf{f}^{-1} = \mathbf{g}_{\alpha_0} \tag{9}$$

Now, denote by $\hat{s}_i(\alpha), i = 1, 2$ the components obtained by

$$\hat{\mathbf{s}}_\alpha = \mathbf{g}_\alpha(\mathbf{x}) \tag{10}$$

Take some measure of the dependence of the obtained components, for example mutual information I (Comon, 1994; Hyvärinen, 1998). Then we can minimize the dependence with respect to α:

$$\min_\alpha I(\hat{s}_1(\alpha), \hat{s}_2(\alpha)). \tag{11}$$

For some value of α, the mutual information $I(\hat{s}_1(\alpha), \hat{s}_2(\alpha))$ vanishes. In one of such points, we have then obtained the original independent components. Furthermore, it seems very plausible that the mutual information vanishes only in the directions of the independent components (which means four points in the range $[0, 2\pi]$, corresponding to the four directions defined by the independent components). Thus we conjecture that the solution is unique except in some degenerate cases. By a degenerate distribution we mean here a distribution that is invariant with respect to the rotation defined above: this is the conterpart of a gaussian distribution in the linear case. Simulation results presented below give support to our conjecture.

4 Simulation results

4.1 Existence

First we present simulation results that illustrate the existence result of Section 2. Since Theorem 1 gives an explicit construction, we can use it to decompose a given n-dimensional signal into independent components.

In the simulations, we used two independent components of uniform distributions in $[0, 1]$. Three different mixing functions (mappings) were used:

$$\mathbf{f}_1(\mathbf{s}) = \begin{bmatrix} \tanh(4s_1 - 2) + s_1 + s_2/2 \\ \tanh(4s_2 - 2) + s_2 + s_1/2 \end{bmatrix}, \tag{12}$$

$$\mathbf{f}_2(\mathbf{s}) = \begin{bmatrix} \tanh(s_2)/2 + s_1 + s_1^2/2 \\ s_1^3 - s_1 + \tanh(s_2) \end{bmatrix}, \tag{13}$$

and

$$\mathbf{f}_3(\mathbf{s}) = \begin{bmatrix} s_2^3 + s_1 \\ \tanh(s_2) + s_1^3 \end{bmatrix}. \tag{14}$$

The function $\mathbf{f}_1$ is only moderately nonlinear, whereas $\mathbf{f}_2$ is rather strongly nonlinear. The function $\mathbf{f}_3$, on the other hand, is not even bijective, and thus quite problematic. Figs. 1–3 illustrate the functions by showing the images of a grid in the square $[0, 1] \times [0, 1]$, using these three functions.

Applying one of the above functions on the uniformly distributed independent components, a random vector $\mathbf{x}$ was obtained. The nonlinear ICA construction of

Section 2 was then used to obtain independent components y_1, y_2. The densities needed were estimated simply by histograms, dividing the values of x_1 into 100 bins. The function $\mathbf{g}$ was then computed nonparametrically, by direct application of (2). The joint density of the vector $\mathbf{y} = (y_1, y_2)$ is illustrated in Fig 4 for $\mathbf{f}_1$; it is a uniform distibution up to certain estimation errors, as expected. (For $\mathbf{f}_2$ and $\mathbf{f}_3$, the density was essentially similar). This shows that the obtained components were really independent, since variables with jointly uniform density are always mutually independent.

The theory in Section 2 showed that the independent components are by no means uniquely defined. To illustrate this fact, we computed the composite mappings $\mathbf{g}_i \circ \mathbf{f}_i$, where the $\mathbf{g}_i$ are the transformations defined in Theorem 1, using mixtures obtained by the $\mathbf{f}_i, i = 1, 2, 3$. The results are shown in Figs. 5–7. The composite mappings are not equal to identity; instead, they are some measure-preserving transformations on the unit square. Thus, the assumption of independency is not sufficient to determine the inverse transformations of the $\mathbf{f}_i$. For smooth mixing mappings ($\mathbf{f}_1$ and $\mathbf{f}_2$), the composite mapping is also smooth, whereas the composite mapping obtained from $\mathbf{f}_3$ folds together, like $\mathbf{f}_3$ itself. These results show that a simple application of the construction of Theorem 1 is not enough to estimate the original independent components.

4.2 Uniqueness

Next we give a simple illustration of the uniqueness result presented in Section 3. The independent components were uniformly distributed in $[-1, 1] \times [-1, 1]$. A conformal mixing mapping $\mathbf{f}_c$ was randomly constructed by generating a polygon with random vertices and constructing the corresponding Schwarz-Christoffel mapping, which is shown in Fig. 8. The Schwarz-Christoffel mapping was constructed using the SC MATLAB toolbox (see (Driscoll, 1996)). In the first part of the estimation of the independent components and the mixing mapping, the compact support of the mixture density was manually approximated using only the mixture samples. Fig. 9 shows the observed data, i.e., the obtained mixture samples, together with the estimated support of the (mapped) densities. The actual compact support of the mixture densities is depicted in Fig. 10. Our estimation of the support was very coarse, but this did not significantly deteriorate the results. Then, a conformal separating mapping $\mathbf{g}_0$ was computed using the Schwarz-Christoffel formula (Churchill and Brown, 1990; Driscoll, 1996) to the domain $[-1,1] \times [-1,1]$. This enabled the estimation of the mapping up to a rotation, as explained in Section 3.2. The composite mapping $\mathbf{h} = \mathbf{g}_0 \circ \mathbf{f}_c$ is depicted in Fig. 11. It is clearly a rotation. The second part of the estimation was to determine the (inverse) rotation needed to obtain the independent components, using some statistical criteria of independence, as explained in Section 3.3. Here, we determined directions in which the kurtosis in minimized; this criterion is closely related to minimization of mutual information (Comon, 1994; Hyvärinen, 1998). Thus we obtained a separating function $\mathbf{g}$, which gave a very good approximation of the inverse of the mixing mapping $\mathbf{f}_c$. This can be seen in Fig. 12, where the final composite function $\mathbf{g} \circ \mathbf{f}_c$ is depicted. This function is essentially equal to identity, up to some estimation errors, which appear near the boundaries.

5 Discussion

In this paper, we treated the problems of existence and uniqueness of the solutions for the nonlinear ICA problem, i.e., the problem of nonlinearly transforming an n-dimensional random vector into components that are statistically independent. This is the nonlinear generalization of the ICA problem treated in, e.g., (Comon, 1994; Jutten and Herault, 1991), and can also be interpreted as the estimation of a nonlinear generative model for the data.

First, we showed that there exists always a solution to the nonlinear ICA problem. In particular, we showed how to construct a solution. Using this construction, we showed that the solution is not unique, and that the non-uniqueness cannot be reduced to some trivial indeterminacies, like component-wise transformations. Two solutions for the problem may be completely different from each other. This result has important implications in the design of algorithms for nonlinear ICA: any algorithm for (general) nonlinear ICA should specify which solution it tries to find. At the same time, our construction gives a practical method for finding *one* of the solutions.

Second, we used some results from complex analysis to achieve uniqueness of the solution. If the dimension of the problem is constrained to be two, we conjectured that the solution is unique under some assumptions. In particular, we made the assumption that the mixing mapping is a *conformal* mapping. This is a strong assumption: Not all linear mappings $\mathbb{R}^2 \to \mathbb{R}^2$ are conformal mappings for the corresponding complex variables. The conformal mappings are, however, a rather large class of functions (Ahlfors, 1979). For example, any simply connected domain can be mapped onto another one by a conformal mapping. This result shows that it is possible to obtain uniqueness in nonlinear ICA by restricting the mixing function $\mathbf{f}$ to a certain class, thus complementing the uniqueness results in (Taleb and Jutten, 1997).

References

Ahlfors, L. (1979). *Complex Analysis.* McGraw-Hill, 3rd edition.

Barlow, H. B. (1961). Possible principles underlying the transformations of sensory messages. In Rosenblith, W. A., editor, *Sensory Communication*, pages 217–234. MIT Press.

Barlow, H. B. (1972). Single units and sensation: A neuron doctrine for perceptual psychology? *Perception*, 1:371–394.

Bell, A. and Sejnowski, T. (1997). The 'independent components' of natural scenes are edge filters. *Vision Research*, 37:3327–3338.

Burel, G. (1992). Blind separation of sources: A nonlinear neural algorithm. *Neural Networks*, 5(6):937–947.

Cardoso, J.-F. and Laheld, B. H. (1996). Equivariant adaptive source separation. *IEEE Trans. on Signal Processing*, 44(12):3017–3030.

Churchill, R. V. and Brown, J. W. (1990). *Complex Variables and Applications.* McGraw-Hill, 5th edition.

Comon, P. (1994). Independent component analysis – a new concept? *Signal Pro-*

cessing, 36:287–314.

Darmois, G. (1951). Analyse des liaisons de probabilité. In *Proc. Int. Stat. Conferences 1947*, volume III A, page 231, Washington, D. C.

Deco, G. and Brauer, W. (1995). Nonlinear Higher-order Statistical Decorrelation by Volume-Conserving Neural Networks. *Neural Networks*, 8:525–535.

Deco, G. and Obradovic, D. (1996). *An Information-Theoretic Approach to Neural Computing.* Springer-Verlag, New York.

Driscoll, T. (1996). A MATLAB toolbox for Schwarz-Christoffel mapping. *ACM Trans. on Mathematical Software*, 22:168–186.

Hyvärinen, A. (1998). New approximations of differential entropy for independent component analysis and projection pursuit. In *Advances in Neural Information Processing 10 (Proc. NIPS*97)*, pp. 273–279. MIT Press.

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

Lee, T.-W., Koehler, B., and Orglmeister, R. (1997). Blind source separation of nonlinear mixing models. In *Neural networks for Signal Processing VII.*

Olshausen, B. A. and Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609.

Pajunen, P., Hyvärinen, A., and Karhunen, J. (1996). Nonlinear blind source separation by self-organizing maps. In *Proc. Int. Conf. on Neural Information Processing*, pages 1207–1210, Hong Kong.

Pajunen, P. and Karhunen, J. (1997). A maximum likelihood approach to nonlinear blind source separation. In *Proceedings of the 1997 Int. Conf. on Artificial Neural Networks (ICANN'97)*, pages 541–546, Lausanne, Switzerland.

Taleb, A. and Jutten, C. (1997). Nonlinear source separation: The post-nonlinear mixtures. In *Proc. European Symposium on Artificial Neural Networks (ESANN97)*, pages 279–284, Bruges, Belgium.

Yang, H. H., Amari, S.-I., and Cichocki, A. (1998). Information-theoretic approach to blind separation of sources in non-linear mixture. *Signal Processing*, 64(3):291–300.

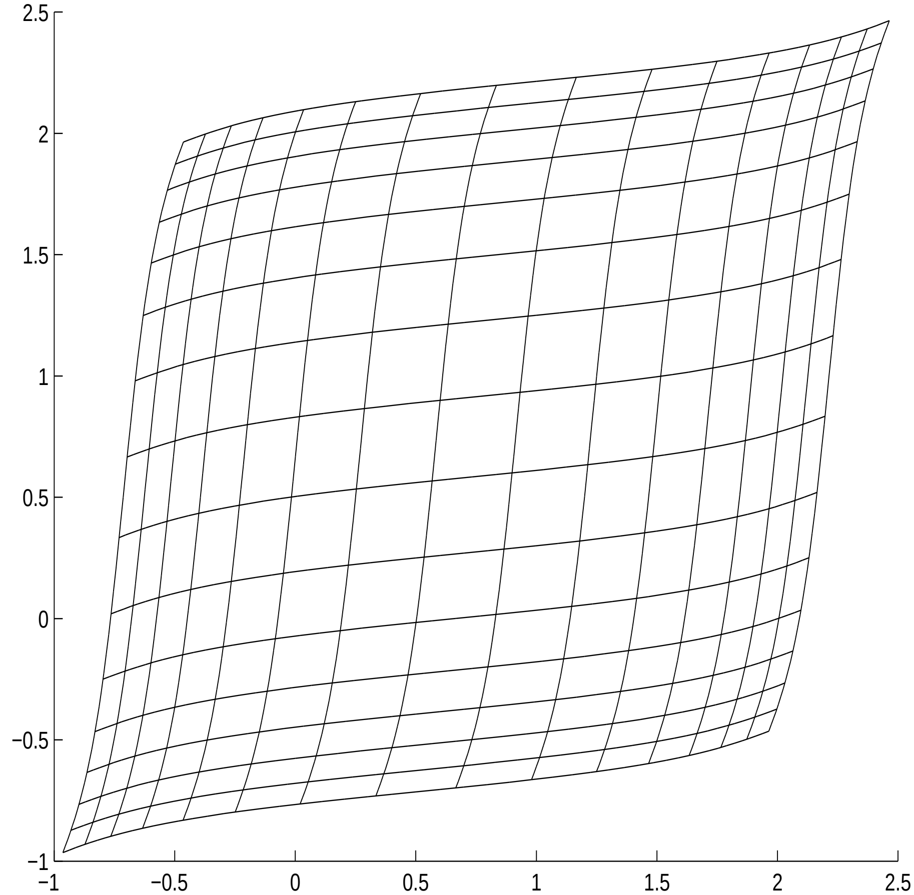

Figure 1: The nonlinear mixing mapping $\mathbf{f}_1$ used in the simulations. The mapping is only moderately nonlinear.

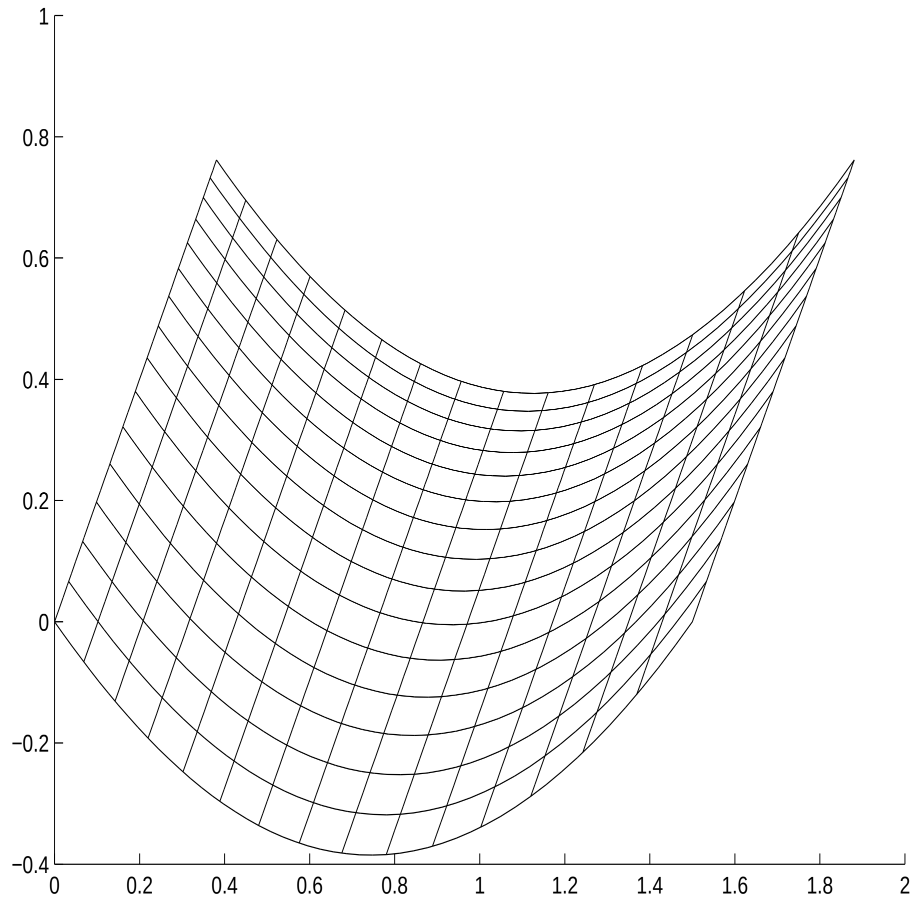

Figure 2: The nonlinear mapping $\mathbf{f}_2$ used in the simulations. The mapping rather strongly nonlinear.

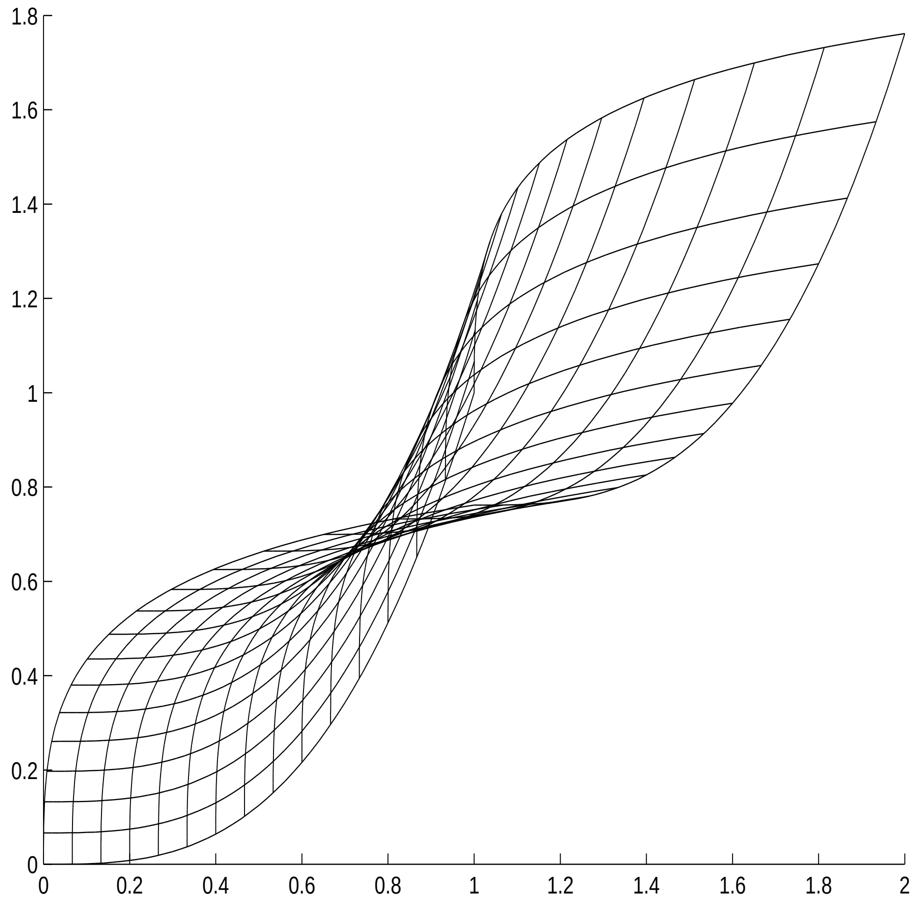

Figure 3: The nonlinear mapping $\mathbf{f}_3$ used in the simulations. The mapping is very nonlinear; in fact, it is not even bijective.

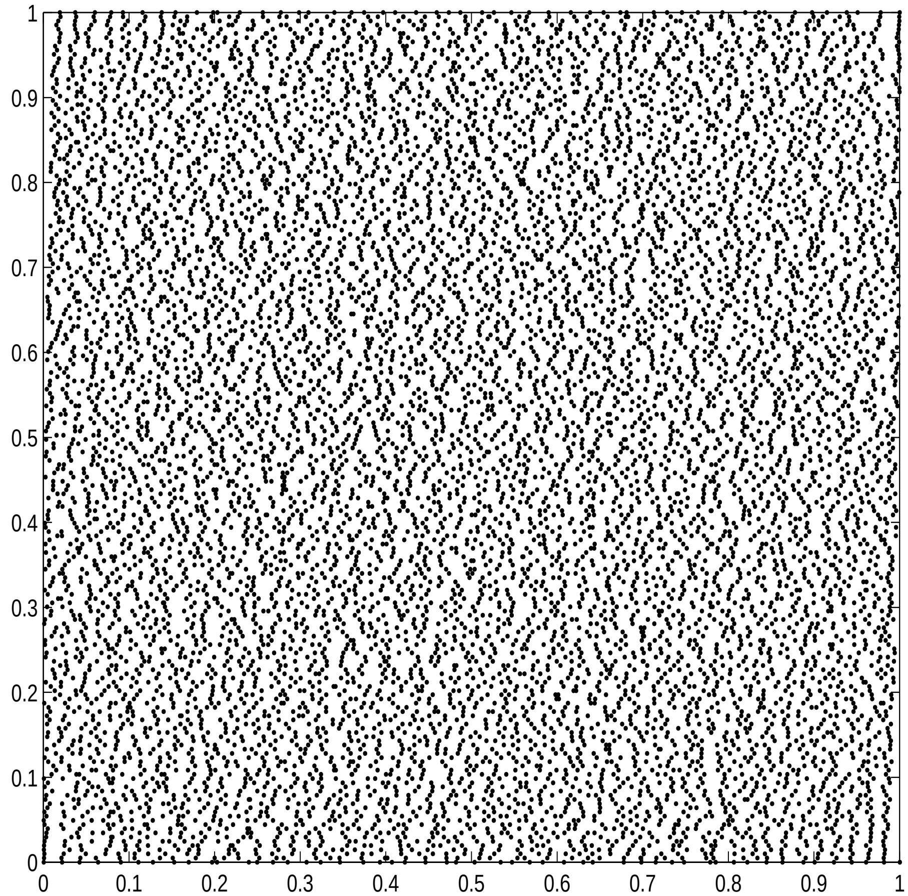

Figure 4: The joint density of the obtained independent components for mixtures obtained by $\mathbf{f}_1$.

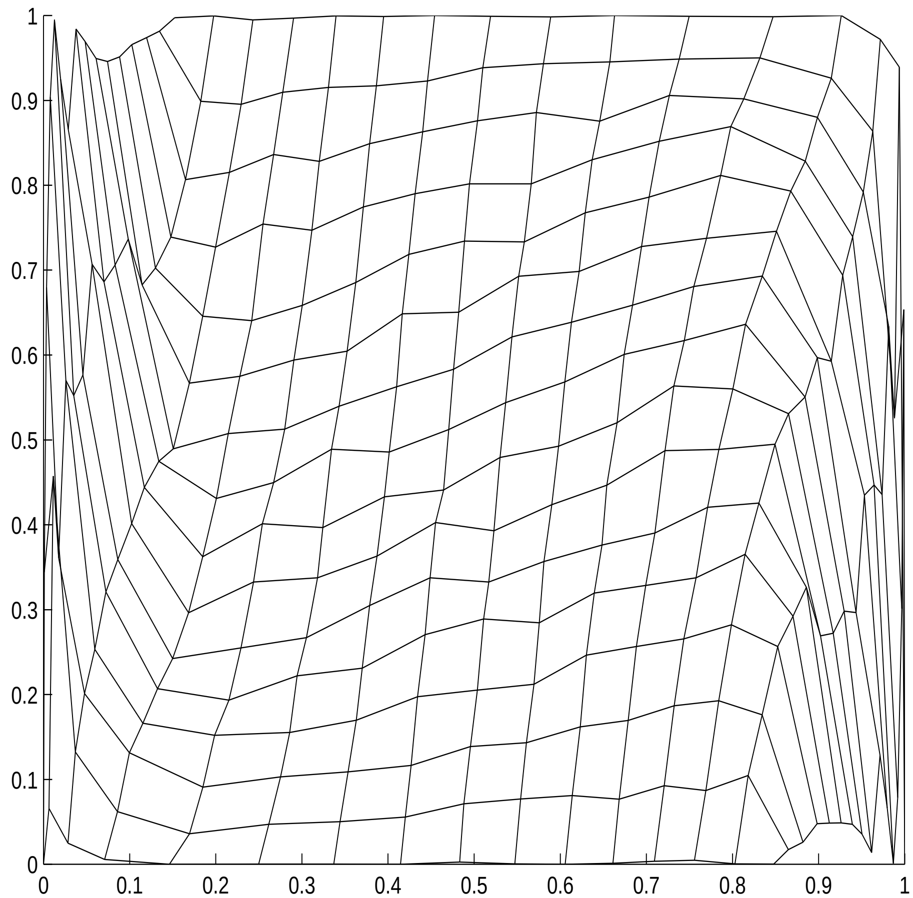

Figure 5: The composite mapping $\mathbf{g}_1 \circ \mathbf{f}_1$ obtained by finding independent components from mixtures defined by the mixing function $\mathbf{f}_1$, using the construction of Theorem 1.

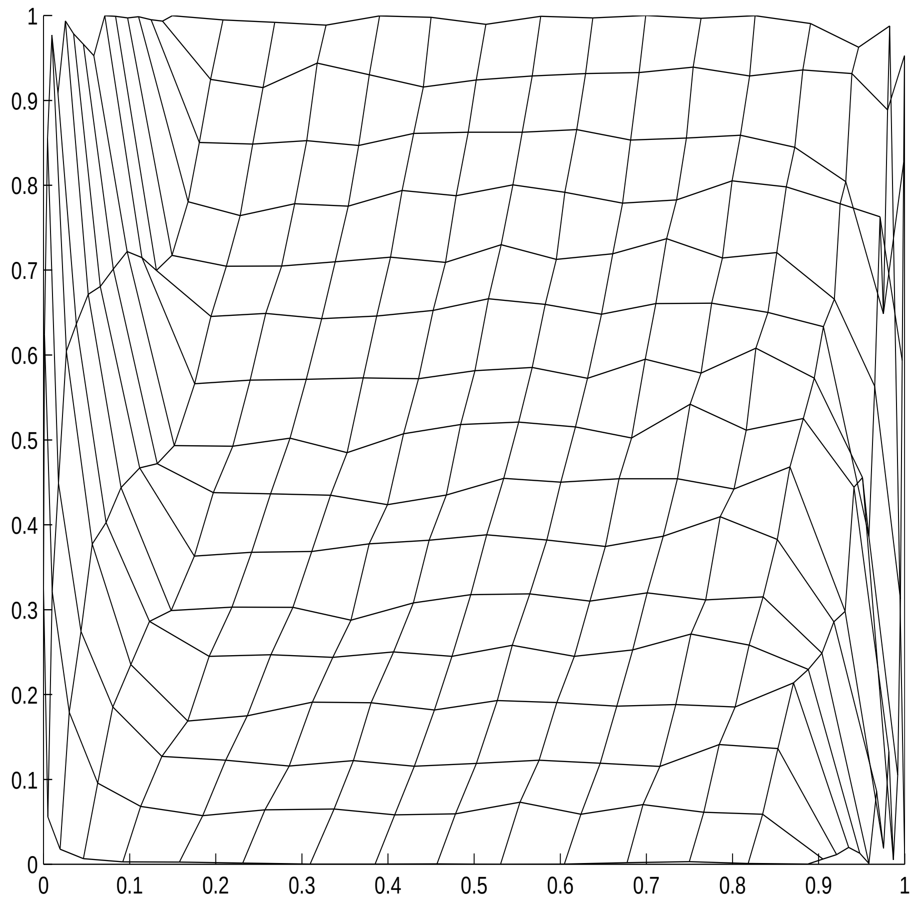

Figure 6: The composite mapping $\mathbf{g}_2 \circ \mathbf{f}_2$ obtained by finding independent components from mixtures defined by the mixing function $\mathbf{f}_2$, using the construction of Theorem 1.

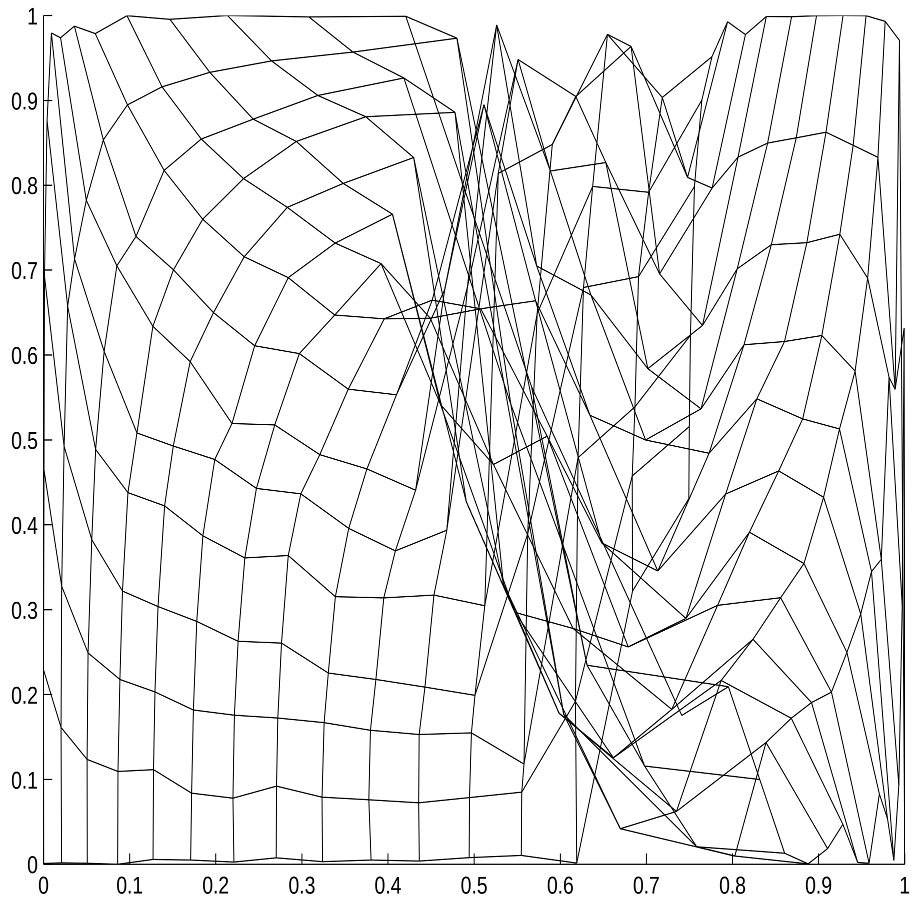

Figure 7: The composite mapping $\mathbf{g}_3 \circ \mathbf{f}_3$ obtained by finding independent components from mixtures defined by the mixing function $\mathbf{f}_3$, using the construction of Theorem 1.

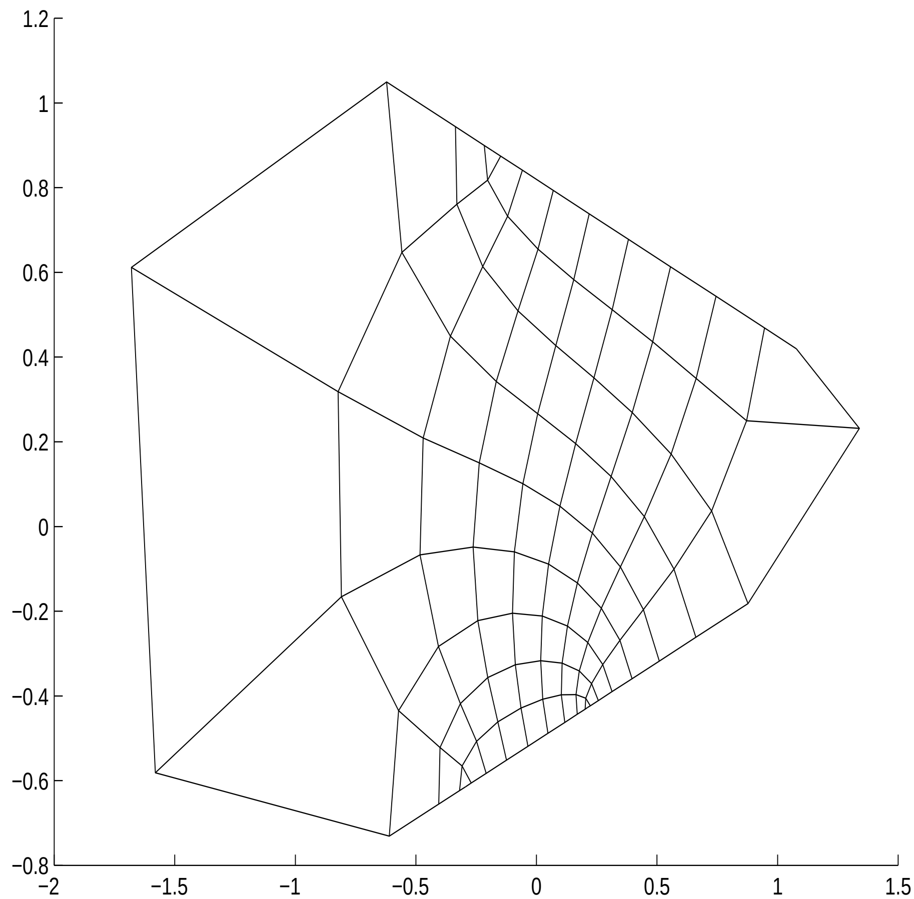

Figure 8: The conformal mixing mapping $\mathbf{f}_c$. The mixing mapping was clearly quite nonlinear.

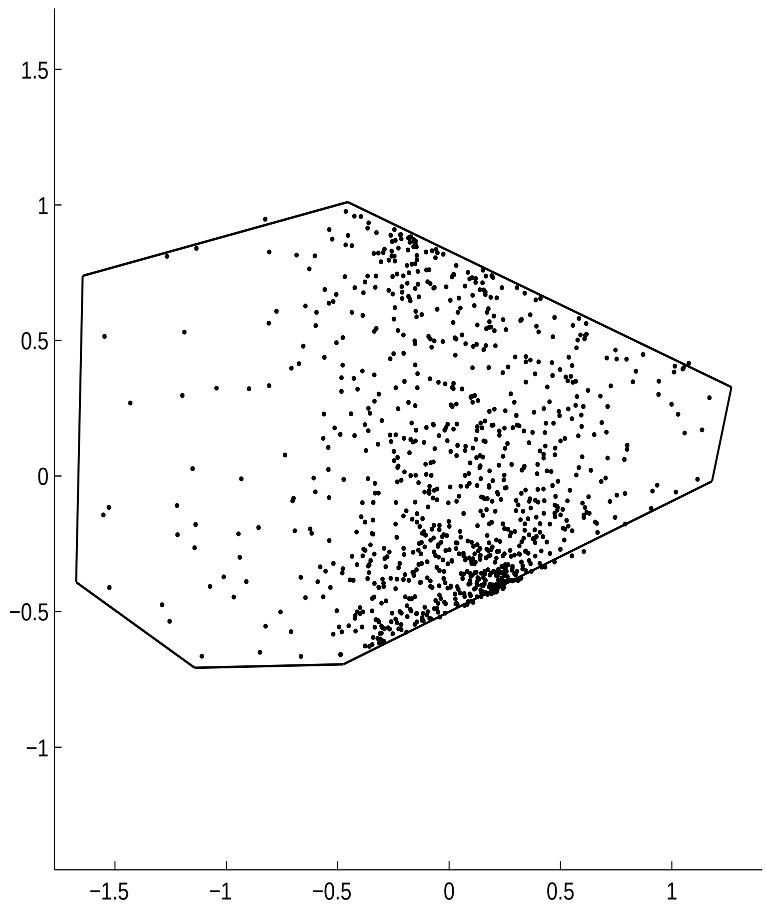

Figure 9: The mixture samples, i.e., the observed data used in determining the separating conformal mapping, together with our estimation of the support of the mixture densities.

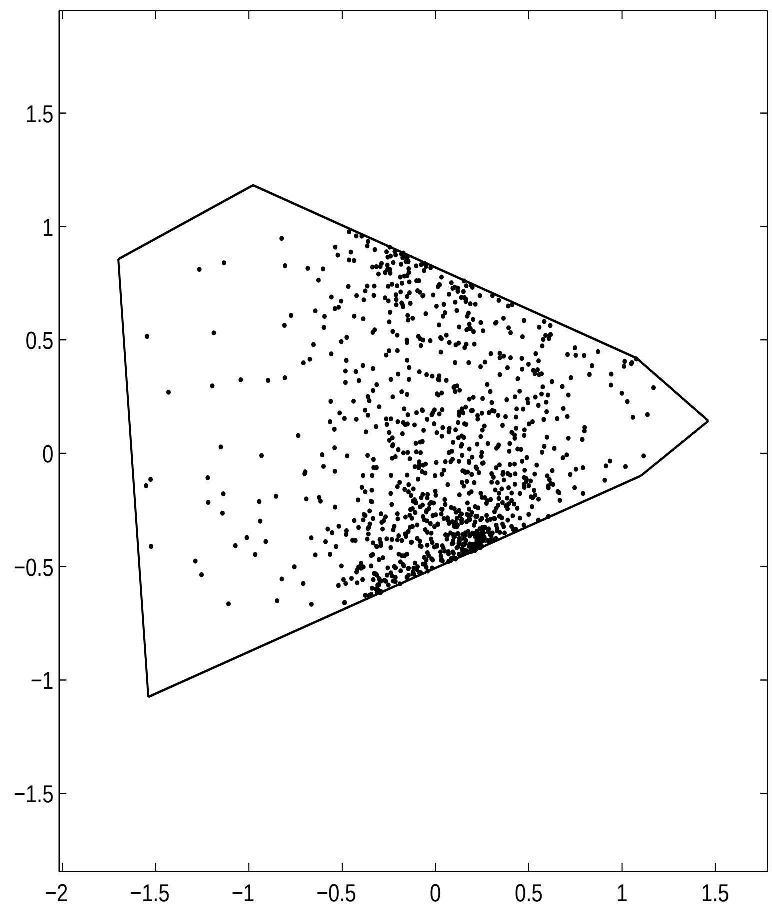

Figure 10: The true compact support of the densities of the mixtures.

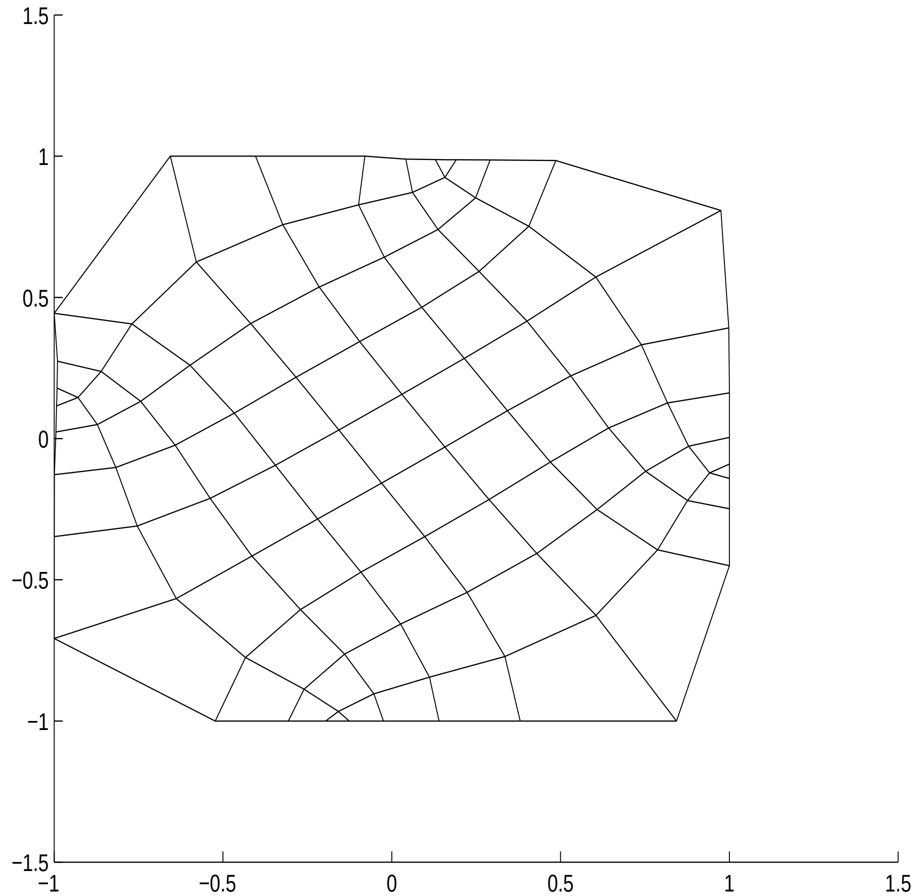

Figure 11: The mapping **h**, i.e., the composite mapping after the first part of the estimation procedure (determination of the inverse conformal mapping).

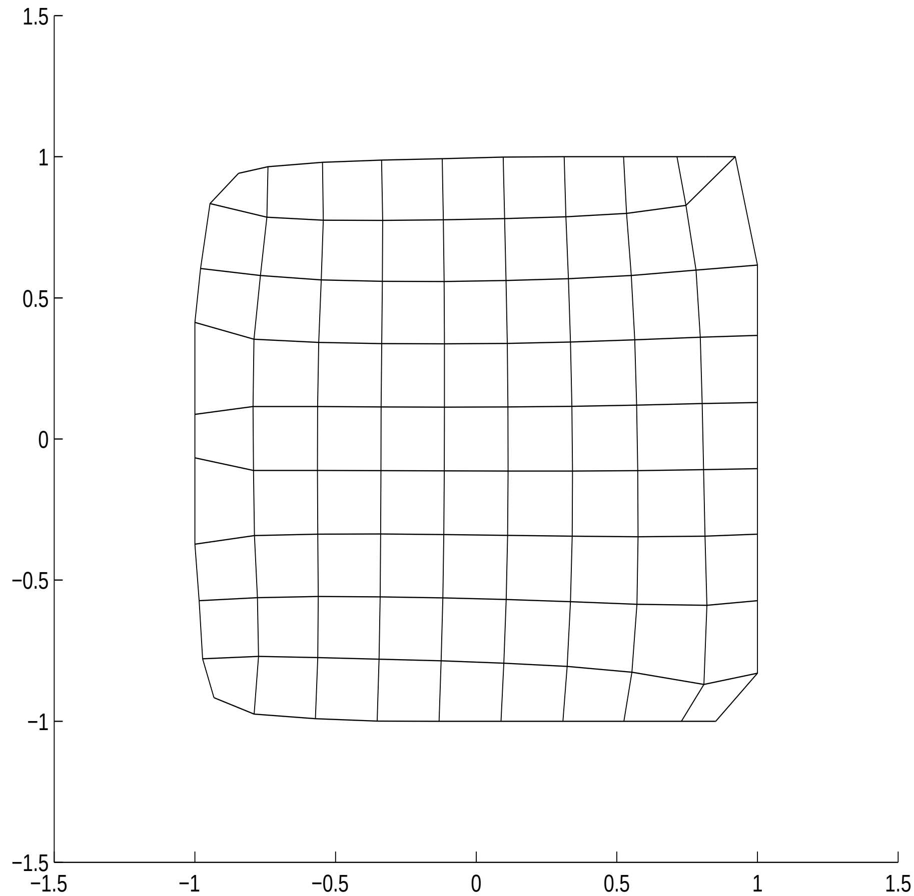

Figure 12: The final composite mapping, obtained by determining the correct rotation. Indeed, we found the original independent components.

Figure captions:

Fig. 1. The nonlinear mixing mapping $\mathbf{f}_1$ used in the simulations. The mapping is only moderately nonlinear.

Fig. 2. The nonlinear mapping $\mathbf{f}_2$ used in the simulations. The mapping rather strongly nonlinear.

Fig. 3. The nonlinear mapping $\mathbf{f}_3$ used in the simulations. The mapping is very nonlinear; in fact, it is not even bijective.

Fig. 4. The joint density of the obtained independent components for mixtures obtained by $\mathbf{f}_1$.

Fig. 5. The composite mapping $\mathbf{g}_1 \circ \mathbf{f}_1$ obtained by finding independent components from mixtures defined by the mixing function $\mathbf{f}_1$, using the construction of Theorem 1.

Fig. 6. The composite mapping $\mathbf{g}_2 \circ \mathbf{f}_2$ obtained by finding independent components from mixtures defined by the mixing function $\mathbf{f}_2$, using the construction of Theorem 1.

Fig. 7. The composite mapping $\mathbf{g}_3 \circ \mathbf{f}_3$ obtained by finding independent components from mixtures defined by the mixing function $\mathbf{f}_3$, using the construction of Theorem 1.

Fig. 8. The conformal mixing mapping $\mathbf{f}_c$. The mixing mapping was clearly quite nonlinear.

Fig. 9. The mixture samples, i.e., the observed data used in determining the separating conformal mapping, together with our estimation of the support of the mixture densities.

Fig. 10. The true compact support of the densities of the mixtures.

Fig. 11. The mapping $\mathbf{h}$, i.e., the composite mapping after the first part of the estimation procedure (determination of the inverse conformal mapping).

Fig. 12. The final composite mapping, obtained by determining the correct rotation. Indeed, we found the original independent components.