Minimum Entropy Blind Signal Deconvolution with Non Minimum Phase FIR Filters

Axel Röbel
Institute for Communication Sciences, Sekr. EN-8, Technical University of Berlin
Einsteinufer 17, 10587 Berlin, Germany
email: roebel@kgw.tu-berlin.de

ABSTRACT

In the following paper we investigate two algorithms for blind signal deconvolution that has been proposed in the literature. We derive a clear interpretation of the information theoretic objective function in terms of signal processing and show that only one is appropriate to solve the deconvolution problem, while the other will only work if the unknown filter is constrained to be minimum phase. Moreover we argue that the blind deconvolution task is more sensitive to a mismatch of the density model than currently expected. While there exist theoretical arguments and practical evidence that blind signal separation requires only a rough approximation of the signal density this is not the case for blind signal deconvolution. We give a simple example that supports our argumentation and formulate a sufficiently adaptive density model to properly solve that problem.

Keywords: Blind Signal Deconvolution, Infomax, Minimization of Entropy, Circular Matrices, Statistical Signal Processing.

1. INTRODUCTION

Recently information theoretic formulation of blind signal separation and blind signal deconvolution criteria have received much interest [8, 12]. The goal of blind deconvolution [6] is to recover a source signal $x(n)$ given only the output $y(n)$ of an unknown filter with impulse response $\{a_k\}$

$$y(n) = \sum_{k=0}^{N} a_k x(n-k). \tag{1}$$

The problem is to find the inverse filter $\{b_k\}$ that yields

$$x(n) = \sum_{k=0}^{M} b_k y(n-k) \tag{2}$$

given only $y(n)$. Because the knowledge of $y(n)$ is generally not sufficient to find the inverse filter we need to establish further constraints. In blind signal processing it is generally assumed that $x(n)$ is a white noise signal with non Gaussian density. Given this restriction the inverse filter has to remove all statistical dependencies across time that are introduced by the filter $\{a_k\}$. The infinitely many solutions of this problem differ only with respect to scaling and time shift. If we restrict the inverse filters to the class $\mathcal{B}$ of causal filters with $b_0 \neq 0$ and proper standardization the problem has a unique solution.

If $\{a_k\}$ and $\{b_k\}$ are restricted to be minimum phase with the standardization $b_0 = 1$, then the solution can be obtained by means of finding the filter $\{b_k\}$ that achieves the source signal $x(n)$ with minimum variance. This is the foundation of the well known and widely used linear prediction algorithm [10]. Without the restriction to minimum phase, however, there exist 2^M different filters $\{b_k\}$ with the same variance of the deconvolved signal. While one of these filters is the inverse of $\{a_k\}$, all the others include an additional all-pass component.

It has been shown earlier that many other objective functions may be used to find the inverse filter $\{b_k\}$, and that minimizing the entropy of $x(n)$

$$D(x) = -\int_x p(x) log(p(x)) dx, \tag{3}$$

where $p(x)$ is the distribution of the samples of $x(n)$, yields asymptotically optimal results [3]. A deconvolution algorithm that properly minimizes the signal entropy as defined in Eq. (3) is of special interest for data compression algorithms or source/filter signal models, which today use linear prediction to decorrelate the samples. Due to the restriction to minimum phase filtering and due to possible nonlinear dependencies in $x(n)$, however, the minimum variance objective of the linear prediction algorithm will generally fail to find the minimum entropy source signal, and, therefore, the results of linear prediction compression algorithms are suboptimal. Due to the fact, that the relation between the distribution $p(x)$ and the filter parameters $\{b_k\}$ is generally unknown, the use of the entropy as objective function has been rather crucial[3].

In a remarkable investigation on information theoretic objectives for blind signal processing it has been

shown recently that by means of a matrix formulation of the filter operation in Eq. (2) an approximate solution to the minimum entropy deconvolution can be obtained [2]. In that paper a triangular Toeplitz matrix has been used to express the filter operation. Later a different matrix formulation based on circular matrices has been proposed [7, 4]. However, the relation between both methods and the implications of the different approximations remain unclear. In the following investigation we will show that the matrix expression of Eq. (2) that is based on a Toeplitz matrix is only suitable if the unknown filter $\{a_k\}$ is constrained to be minimum phase, which is considerably more restrictive than originally stated. The use of the circular matrices, however, leads to a deconvolution algorithm that is able to solve the general deconvolution problem. Because the circular matrices are related to circular deconvolution some minor modifications have to be applied such that the solution obtained with the algorithm is suitable for the non circular deconvolution task.
By means of a simple experiment, we demonstrate that the fixed density model used in [2] is not sufficient for blind deconvolution even if the problem is constrained to super Gaussian sources. Due to the mismatch between the model and true signals distribution the inverse of the true convolution filter is related to a local and not the global optimum of the objective function. The algorithms that have been proposed recently to solve the general blind separation problem based on a sub and super Gaussian switch of the density model [9, 5] are not appropriate to solve this problem. Therefore, we propose the use of an adaptive bi-modal density model which is able to properly solve our example problem. Experimental results not presented in this paper show that our adaptive density model can also be used to deconvolve sub Gaussian sources.

It is interesting to note that our simple example that shows the limitations of the blind signal deconvolution with non adaptive densities appears to have no consequences for blind signal separation applications. While we can easily construct an equivalent signal separation problem that can not be solved with existing algorithms, this situations has not been observed in practical applications. Therefore, we conjecture that this problem is a consequence of the special symmetries of the matrices that describe the blind deconvolution problem which, however, are very unlikely to appear for a signal separation problem.

The following paper is organized as follows. In section 2 we shortly describe the blind deconvolution method introduced by Bell and Sejnowski. In section 3 we describe the alternative matrix formulation of the filtering process and argue that the methods differ only for non minimum phase problems, for which the second method achieves correct results. In section 4 we describe our adaptive bimodal source distribution model. Section 5 shortly explains some experimental results we have obtained for white noise test signals and section 6 concludes with a summary and an outlook on further work.

2. INFORMATION MAXIMIZATION AND MINIMUM ENTROPY

Bell and Sejnowski developed their deconvolution algorithm as an application of the minimum entropy blind signal separation algorithm they presented in the same paper [2]. In the following we give a short summary of the key idea of their algorithm, for detailed description see the original paper. In the following we adopt the original argumentation that is based on information maximization. Note, however, that the same algorithm can be derived also by means of a maximum likelihood approach [11]. Assume we are given an L-channel instantaneously mixed signal $\vec{y}(n)$ and are searching the original L source signals $x_i(n)$ that are assumed to be statistically independent. Formally, we are looking for the unmixing matrix B that achieves

$$\vec{x}(n) = B\vec{y}(n). \tag{4}$$

As Bell and Sejnowski has shown, the task can be addressed by maximizing the joint entropy of a nonlinearly transformed output signal $\vec{z}(n)$ with components

$$z_i(n) = f_i(x_i(n)),$$

with all f_i being constraint to be monotonically increasing with fixed range, i.e. $[-1, 1]$. Following [2] the joint entropy $D(\vec{z})$ can be approximately expressed as

$$\begin{aligned} D(\vec{z}) &= \log(|\det(B)|) \\ &+ \frac{1}{N}\left(\sum_{n=0}^{N-1}\sum_{i=1}^{L}\log(\frac{\partial z_i(n)}{\partial x_i(n)})\right) + C, \end{aligned} \tag{5}$$

where N is the length of the respective signal vectors and C is constant and equal to the joint entropy $D(\vec{y})$. The approximation is due to the calculation of a sample mean instead of the entropy integral. This term yields the expectation of the logarithm of the derivative $f_i'(x)$. Due to the special structure of f_i this derivative has the properties of a density, and, therefore the expectation is maximized if $f_i'(x)$ equals the density of x. In this case and with the same approximation as in Eq. (5) this second term equals the negative sum of the L entropies $D(x_i)$ and due to the basic relations between joint and scalar entropies we can rewrite Eq. (5) as

$$D(\vec{z}) = \log(|\det(B)|) - D(\vec{x}) - T(\vec{x}) + C, \tag{6}$$

where $T(\vec{x})$ is the mutual information between the channels. From the basic laws of variable transformation it is known that the joint entropy $D(\vec{x})$ equals the sum of the joint entropy $D(\vec{y})$ ($= C$), which is constant here, and a scaling term given by $\log(|\det(B)|)$. Therefore, we conclude that the first term in Eq. (5) compensates any scaling that is produced by means of the linear transformation

B. As long as the derivative of the nonlinearity f_i equals the density of the samples x_i we have

$$D(\vec{z}) = -T(\vec{x}), \tag{7}$$

and, therefore, under this constraint maximization of the joint entropy of $\vec{z}$ is equivalent to the minimization of the mutual information [12]. To simplify the algorithm Bell and Sejnowski proposed to use a fixed nonlinearity

$$f_i = \tanh(x_i) \tag{8}$$

which is equivalent to assume a fixed density model for the signals x_i. They conjecture that successful separation of super Gaussian sources is possible even if f_i is not equal to the source distribution (see Sec. **4.**).

To be able to apply the algorithm for blind deconvolution, Bell and Sejnowski formulate the deconvolution in Eq. (2) by means of matrix multiplication between an $L \times L$ matrix B and an L-dimensional vector $\vec{y}(n) = (y(n), y(n+1), \ldots, y(n+L-1))^T$. To construct B they set the matrix elements on the k-th diagonal to b_k, where the main diagonal is identified with $k = 0$ and the diagonals are counted from right to left. As an example we construct the matrix B for a causal filter of order $M = 3$

$$B = \begin{pmatrix} b_0 & 0 & 0 & 0 & 0 & \cdots \\ b_1 & b_0 & 0 & 0 & 0 & \cdots \\ b_2 & b_1 & b_0 & 0 & 0 & \cdots \\ 0 & b_2 & b_1 & b_0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}. \tag{9}$$

Multiplication of B with $\vec{y}$ from the right yields a vector representation of the output of the filter $x(n)$. Based on this matrix representation the blind separation algorithm summarized above may be applied. For causal filters B is lower triangular. Using the same nonlinearity f_i for all L channels i Eq. (5) becomes

$$D(\vec{z}) = L\log(|b_0|) + \frac{1}{N}\left(\sum_{n=0}^{N-1}\sum_{i=1}^{L}\log(\frac{\partial z_i(n)}{\partial x_i(n)})\right) + C. \tag{10}$$

Using this equation the gradient of $D(\vec{z})$ with respect to the filter parameters is easy to calculate

$$\frac{\partial D(\vec{z})}{\partial b_j} = \frac{L \cdot \delta(j)}{b_0} + \frac{1}{N}\sum_{n=0}^{N-1}\sum_{i=1}^{L}\frac{\frac{\partial}{\partial b_j}\frac{\partial z_i(n)}{\partial x_i(n)}}{\frac{\partial z_i(n)}{\partial x_i(n)}},$$

$$\text{with} \quad \delta(j) = \begin{cases} 1, & \text{for} \quad j = 0 \\ 0, & \text{for} \quad j \neq 0 \end{cases} \tag{11}$$

and can be employed for an adaptive algorithm for blind signal deconvolution.

While Bell and Sejnowski has successfully applied their algorithm to a number of blind deconvolution tasks, there exists a weak point in the above argumentation that restricts the usage of the algorithm to the case of minimum phase filters $\{b_k\}$. The assumption of equal sample distributions for all channels i that leads to the simple form of Eq. (10) is generally violated for the first $(M-1)$ channels. Given a sequence of vectors $\vec{y}(n)$ that are constructed from different segments of a signal $y'(n)$ of length $N' > N$ we find that the first $(M-1)$ channels always contain transients of the filter response and, therefore, obey different distributions. The impact of this deviation seems to be small for $M \ll L$, however, for non minimum phase filters the transient channels change the scaling behavior of the matrix B compared to the filter output such that an application of Eq. (10) yields incorrect results.

As explained above the first term in the above entropy equations has to compensate for the increase in entropy that is due to scaling. While the first term in Eq. (10) indicates that the scaling due to linear transformation Eq. (9) depends solely on b_0, this is not true for the output $x(n)$ of a non minimum phase filter. However, from the above reasoning it is difficult to develop the correct scaling compensation that has to be applied in Eq. (10).

3. CIRCULAR FILTERING

The formulation of the FIR-filtering as a matrix multiplication is not the only one possible. Therefore, we will now adopt a different argumentation and will show that Eq. (10) is correct only for minimum phase filters. As have been shown by Lambert [7] and [4] the deconvolution task can also be formulated using so called quadratic *circular matrices*[1] $\hat{B}$ instead of the Toeplitz matrices used so far. Compared to the earlier work on CM the following conduction gives a new interpretation of the objective function with respect to the filter transfer function. Moreover the interpretation enables us to understand the limitations of the Toeplitz filter matrix. In the following we will use a slightly different rule to construct the CM for a given filter than Lambert has proposed. As a consequence the analysis of the relation between matrix algebra and FFT FIR filter algebra is simplified, because the variable time shift that Lambert has to obey is fixed to zero.

To construct a circular matrix for a periodic sequence of length L we use as first row of the matrix the period of the sequence with time origin positioned at the first column. All following rows are built by circularly shifting the previous row to the right. The relation between a FIR filter of order $M < L$ and a CM is given by the filter response to a unit impulse train with period L. As an example we construct a CM $\hat{B}$ of size $L = 5$ for a FIR causal

[1] In the following denoted as CM.

filter of order $M = 3$

$$\hat{B}(b_k) = \begin{pmatrix} b_0 & b_1 & b_2 & 0 & 0 \\ 0 & b_0 & b_1 & b_2 & 0 \\ 0 & 0 & b_0 & b_1 & b_2 \\ b_2 & 0 & 0 & b_0 & b_1 \\ b_1 & b_2 & 0 & 0 & b_0 \end{pmatrix}. \qquad (12)$$

Using the circular matrices we are able to interpret all matrix operations required for blind signal separation in terms of operations on the periodic sequences used to construct the CM. Moreover all matrix operations can be solved by operating on the L-point discrete Fourier transform (DFT) of the sequence $\{b_k\}$ and constructing a CM from the inverse DFT of the result. The multiplication of two CM, $\hat{B}$ and $\hat{Y}$, that are constructed from sequences, $b(n)$ and $y(n)$, can be calculated equivalently by means of multiplication of the DFT of the sequences. Transposition of a CM is equivalent to reflecting the periodic sequence at time $n = 0$ and is equivalent to calculating the conjugate complex of the DFT of the related sequence $\{b_k\}$. Inversion of a CM yields again a CM and the result is equivalently obtained by inverting the elements of the DFT of $\{b_k\}$. The determinant of a CM is equal to the product of all elements of the L-point DFT of $\{b_k\}$.

With the circular matrices obtained from the filter coefficients $\{b_k\}$ and an N-periodic signal $y(n)$ we can express blind deconvolution for circular filtering without any error. Using the above relations between circular matrices and periodic sequences operations and neglecting the constant term C we formulate the joint entropy Eq. (5) using the CM of size $L = N$ as follows

$$D(\vec{z}) = \sum_{i=0}^{L-1} \log(|H_b(i)|) + \log(\frac{\partial z_0(i)}{\partial x_0(i)}). \qquad (13)$$

Here $H_b(i)$ is the L-point DFT of the filter impulse response $\{b_k\}$. Note that due to the symmetry of the CM in case of circular filtering and provided we use the same nonlinearity in all channels the assumption of equal channel distributions is correct. Therefore, for circular filtering the calculation of the mean in Eq. (5) can be neglected without error. However, we are interested in non circular deconvolution, and, therefore, we have to apply some corrections to the above equation.

First we consider the scaling compensation. If we consider $y(n)$ to be of finite length we may apply the circular deconvolution to its L-periodic continuation. If we increase L to infinity then the results of circular and non circular convolution agree. For increasing L the scaling term in Eq. (13) yields an increasingly dense sampling of the transfer function of the FIR filter, and, besides a factor L, the sum over $H_b(i)$ achieves an continually improved approximation of the integral of the log magnitude of the transfer function $H_b(jw)$

$$S_b = \frac{1}{2\pi} \int_w \log(|H_b(jw)|) dw. \qquad (14)$$

It is well known that Eq. (14) is related to the scaling properties of the filter $\{b_k\}$ [10, p. 130], and we conclude that Eq. (14) is a standardized (with respect to block length L) measure of the appropriate scaling compensation for FIR filtering operation. To be able to apply Eq. (13) to non circular deconvolution the mean of the first term in Eq. (13) should accurately approximate Eq. (14), and, therefore we shall choose L as large as possible. For large L and minimum phase filtering the scaling compensation obtained by the two formulas Eq. (13) and Eq. (10) agree. For non minimum phase filters, however, the scaling effect is under estimated in Eq. (10) such that the entropy is systematically to small, and, therefore, we expect that non minimum phase solutions can not be found. Consider now the second term in Eq. (13). From the previous section we know that this term approximates the negative channel entropy $D(x_i)$. As long as we achieve a sufficient sampling of $p(x)$ we may choose to sum over a sample subset of size K, and out weight the sub sampling by a additional factor L/K. Moreover, if we want to neglect the transients at the borders of the circular filtered $x(n)$ from the density adaptation, we may delete them from the summation with the same correction applied as above. Due to the possibility to use less than L samples to approximate the entropy $D(\vec{x}_i)$ we are free to select L as large as we need to achieve sufficient accuracy for the approximation of Eq. (14) by the first term in Eq. (13). For practical applications we propose to adapt L during the optimization such that the L-point sampling of the transfer function is always a close approximation to Eq. (14). Note, that L depends on the position of the roots of the FIR Filter $\{b_k\}$ and that L has to be increased if the roots of the Filter approach the unit circle. To be able to adapt L during optimization it is convenient to use a standardized entropy measure

$$D_s(\vec{z}) = \frac{D(\vec{z})}{L} \qquad (15)$$

Due to the algebraic relations stated above, the gradient of Eq. (13) with respect to the filter parameters $\{b_k\}$ can be calculated efficiently without any matrix operations. The gradient for the standardized joint entropy is simply

$$\begin{aligned} \frac{\partial D_s(\vec{z})}{b_i} &= DFT_i^{-1}(\{\frac{1}{\overline{DFT(\{b_k\})}}\}) + \\ &+ \frac{1}{K} \sum_{n=n_0}^{n_e} \frac{\frac{\partial}{\partial b_j} \frac{\partial z_0(n)}{\partial x_0(n)}}{\frac{\partial z_0(n)}{\partial x_0(n)}}. \end{aligned} \qquad (16)$$

Here $DFT(\{b_k\})$ denotes L-point DFT of the sequence $\{b_k\}$, DFT_i^{-1} denotes the ith element of the inverse L-point DFT, and $\overline{x_k}$ denotes the conjugate complex of a sequence $\{x_k\}$. Note, that the above algorithm considers batch processing of an entire block of samples. For stochastic update rules, for example in an non stationary environment, the use of the natural gradient should be considered [1, 4].

4. ADAPTIVE DISTRIBUTION MODEL

We now address a further weakness of the deconvolution algorithm that is related to the use of a fixed nonlinearity $f_i(x_i)$ which is equivalent to assume a fixed signal density. It has been shown that for zero mean signals the positions of the local maxima of Eq. (5) are unchanged if f_i does not match the signal distribution [12]. This is the theoretical foundation to use fixed density models for blind source separation. However, the decrease in entropy due to the mismatch between the density $p(x_i)$ and $f_i'(x_i)$ depends on the matrix B, and we can expect that there exist situations where a change of the model density $f_i(x_i)$ changes a global maximum of Eq. (5) into a local one. In this case, the global maximum does no longer reflect proper signal deconvolution or signal separation. To our knowledge this problem has never been reported for signal separation tasks. As the experimental results in Sec. **5.** demonstrate this situation is not that exceptional for deconvolution. This problem can not be addressed by means of using different nonlinearities for sub and super Gaussian sources as proposed in [9, 5], but only by means of a nonlinearity that models the source distribution with sufficient accuracy. Here we propose to use

$$\begin{aligned} f(x) &= \frac{1}{2}(w\tanh(a_1 x + b_1) \\ &+ (1-w)\tanh(a_2 x + b_2)) \\ w &= \frac{1}{1+\exp(-w_h)}, \end{aligned} \tag{17}$$

because the adaptive bi-modal density can be employed to model sub and super Gaussian densities. The mixture parameter w_h is transformed such that w is always in the interval $[0;1]$ The nonlinearity can be interpreted as a neural network with two hidden units [11], which can be adapted by gradient ascend of Eq. (13) with respect to the network parameters. The density model consists of a mixture of two densities of the form $\frac{a}{2\cosh(ax)^2}$. Using the Fourier Transform

$$\int_{-\infty}^{\infty} \frac{a}{2\cosh(ax)^2} e^{-jwx} dx = \frac{w\pi}{2a\sinh(\frac{w\pi}{2a})}$$

we have been able to calculate the moment generating function of this density and have found that its variance is

$$\sigma^2 = \frac{\pi^2}{12a^2}.$$

This result is used to initialize the distribution parameters, such that the model distribution matches the variance of x(n) for the initial filter matrix $\hat{B}$. We initialize the model distribution as follows

$$\begin{aligned} 1.1a_1 &= 0.9a_2 = \frac{12\sigma_x}{pi} \\ b_1 &= -b2 = 0.001 \\ w_h &= 0.0 \end{aligned}$$

such that the model is slightly non symmetric, however, with a variance that is close to the variance σ_x^2 of the signal x obtained from the initial CM $\hat{B}$.

5. EXPERIMENTAL RESULTS

To verify our reasoning we have applied the above algorithms to two deconvolution problems, with $\{a_k\}$ being minimum phase in the first and maximum phase in the second experiment. First we consider the algorithms with the fixed nonlinearity Eq. (8). To match the fixed density model a super Gaussian source signal $x(n)$ with exponential distribution and variance 1 has been selected. For the (unknown) filter $\{a_k\}$ we use the IIR filter transfer functions

$$H_1(z) = \frac{1}{1+0.5z^{-1}+0.2z^{-2}} \tag{18}$$

$$H_2(z) = \frac{1}{1+2z^{-1}+1.5z^{-2}}. \tag{19}$$

We realize the maximum phase filter $H_2(z)$ using a non causal filter. The inverse filter $\{b_k\}$ is provided with five coefficients, while the ideal deconvolution filter needs only three. We initialized the filter coefficients randomly with normal distribution and variance 1 and adapted the filters in batch mode with an epoch size of 10000 using the gradient calculated from the entropy equations explained above.

As expected the Bell and Sejnowski algorithm always converges to a minimum phase solution. In case of the minimum phase filter the solution is close to the inverse of $\{a_k\}$, however, for the maximum phase problem the algorithm have found solutions with roots of the transfer function that are reflected at the unit circle. The circular matrix algorithm with fixed density model finds the same results if the initial random filter is minimum phase, because for minimum phase filters both algorithms agree in their entropy estimation. With initial filters that have at least two roots of the transfer function on the proper side of the unit circle the CM algorithm converges to a filter with the correct roots.

If we analyze the global maximum of the objective function we find that in all cases the global maximum of the joint entropy Eq. (5) is not obtained for the correct inverse filter, but, for the filter with reflected roots. This is due to the fixed density model. While the all pass component that remains in the signal introduce slight statistical dependencies the entropy is maximal in this case because the density of the source (exponential distribution) is further apart from the fixed density model than the density of the all pass filtered signal (Fig. 1). We conclude that for general case blind signal deconvolution the nonlinearity f_i has to be adapted even in case of a super Gaussian signal. Otherwise the global maximum of the joint entropy does not indicate proper deconvolution. Using the adaptive nonlinearity proposed in Sec. **4.** the signal density can

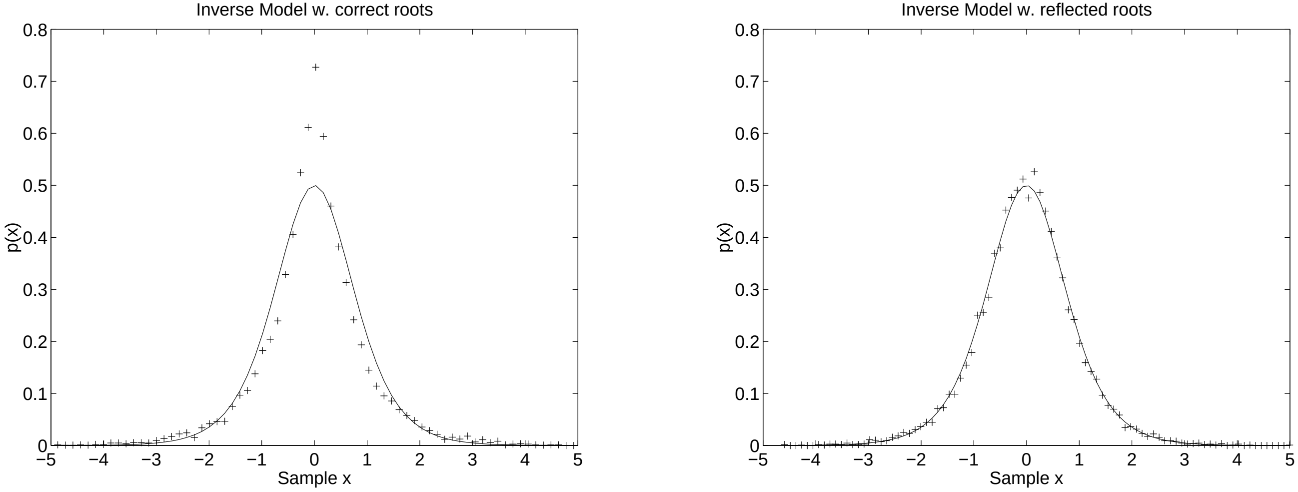

Figure 1: The sample histograms (marked $+$) of the deconvolved signals $p(x)$ in case of unknown filter $H_1(z)$ for the correct inverse model (left) and the inverse model with reflected roots (right) compared with the fixed model distribution $\frac{1}{\cosh^2(x)}$ (solid line). Due to distribution mismatch the incorrect inverse filter (right) yields the global maximum of the joint entropy $D(\vec{z})$.

be modeled more accurately. While the filter coefficients obtained with adaptive nonlinearity does not change significantly, the global maximum of the objective function is now obtained for the correct deconvolution filter.

6. OUTLOOK AND SUMMARY

In the present paper we have investigated into recent blind deconvolution algorithms and have shown, that only the circular matrix formulation of the filtering is appropriate to solve the deconvolution problem if the unknown filter is not minimum phase. Moreover, the experimental results demonstrate that the fixed density model has to be extended to an adaptive at least bi-modal distribution to be able to properly solve the deconvolution problem, even if the source distribution is constrained to be super Gaussian. We presented a simple example that shows that for insufficiently adapted nonlinearity the global optimum of the objective function is not achieved for the true deconvolution filter. Due to the close relations between signal deconvolution and signal separation a similar problem exists for the blind separation task. However, we conjecture that the problem, that to our knowledge has not been reported elsewhere, is related to the symmetries of the matrices that describe the deconvolution problem, and, therefore, may be of only marginal importance for signal separation applications.

Forthcoming investigations will consider applications of the algorithm to data compression of audio signals. Due to the explicit minimization of the entropy of the signal significant improvements of the actual algorithms based on linear prediction are expected. Initial investigations leads to the conclusion that the optimal deconvolution filter for audio signals in many cases requires maximum phase filtering. The compression improvements that are achieved with the new method are currently investigated.

REFERENCES

[1] S. Amari. Natural gradient works effciently in learning. *Neural Computation*, 10:251–276, 1998.

[2] A. J. Bell and T. J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1004–1034, 1995.

[3] David L. Donoho. On mimimum entropy deconvolution. In D. F. Findley, editor, *Proceedings of the Second Applied Time Series Symposium, 1980*, pages 565–608, 1981.

[4] S. C. Douglas and S. Haykin. On the relationship between blind deconvolution and blind source separation. In *Proc. 31st Asilomar Conf. Signals, Systems, and Computers*,, pages 1591–1595, 1997.

[5] M. Girolami. An alternative perspective on adaptive independent component analysis algorithms. *Neural Computation*, 10(8):2103–2114, 1998.

[6] S. Haykin, editor. *Blind Deconvolution*. Prentice-Hall, New Jersey, 1994.

[7] R. H. Lambert. *Multichannel Blind Deconvolution: FIR Matrix Algebra and Separation of Multipath*

Mixtures. PhD thesis, University of Southern California, Department of Electrical Engineering, 1996.

[8] T-W. Lee, M. Girolami, A. J. Bell, and T. J. Sejnowski. A unifying information-theoretic framework for independent component analysis. *International Journal on Mathematical and Computer Modeling*, 1998. In press.

[9] T-W. Lee and T.J. Sejnowski. Independent component analysis for sub-gaussian and super-gaussian mixtures. In *4th Joint Symposium on Neural Computation*, pages 132–139. Institute for Neural Computation, 1997.

[10] J. D. Markel and A. H. Gray. *Linear Prediction of Speech*. Springer Verlag, 1976.

[11] B. A. Pearlmutter and L. C. Parra. A context-sensitive generalization of ICA. In *International Conference on Neural Information Processing*, 1996.

[12] H. Yang and S. Amari. Adaptive online learning algorithms for blind separation: Maximum entropy and minimum mutual information. *Neural Computation*, 9:1457–1482, 1997.