ON THE CAUSALITY PROBLEM IN TIME-DOMAIN BLIND SOURCE SEPARATION AND DECONVOLUTION ALGORITHMS

Robert Aichner, Herbert Buchner, and Walter Kellermann

Multimedia Communications and Signal Processing
University of Erlangen-Nuremberg
Cauerstr. 7, 91058 Erlangen, Germany
{aichner, buchner, wk}@LNT.de

ABSTRACT

Based on a recently presented generic framework for multichannel blind signal processing for convolutive mixtures we investigate in this paper the problem of incorporating acausal delays which are necessary with certain geometric constellations. Starting from a generic update equation which is applicable to blind source separation (BSS), multichannel blind deconvolution (MCBD), and multichannel blind partial deconvolution (MCBPD) for dereverberation of speech signals, two formulations of the natural gradient are derived. It is shown that one expression is applicable to mere causal filters whereas the other one also allows an implementation of noncausal filters. Moreover, proper initialization methods for both cases are given. For the implementation of the aforementioned algorithms cross-relation estimation techniques known from linear prediction are discussed. Based on these results, relationships between traditional MCBD algorithms can be established. Experimental results of different acoustic scenarios show the applicability of the presented algorithms.

1. INTRODUCTION

The task to perform blind signal processing on convolutive mixtures of unknown time series arises in several application domains. In this paper we deal with the so-called cocktail party problem, where we want to recover the speech signals of multiple speakers who are simultaneously talking in a room. The room may be very reverberant due to reflections on the walls, i.e., the original source signals $s_q(n)$, $q = 1, \ldots, Q$ are filtered by a linear multiple input and multiple output (MIMO) system $\mathbf{H}$ before they are picked up by the sensors. In the following, we assume that the number Q of source signals $s_q(n)$ equals the number of sensor signals $x_p(n)$, $p = 1, \ldots, P$.

As one technique to recover the speech signals, BSS can be seen as blind interference cancellation similarly to conventional adaptive beamforming. In the acoustic scenario in Fig. 1a causal filters are sufficient to achieve interference cancellation whereas for the source locations in Fig. 1b usually one noncausal demixing filter w_{12} or w_{21} is required as will be shown below. Due to the supervised filtering algorithms used in adaptive beamforming [1] the problem of acausality can there be solved by initializing the beamformer FIR filters with a unit impulse at the $L/2$-th tap (L denotes the filter length). However, unlike for supervised beamforming, accounting for the causality problem requires structural changes for the adaptation mechanism in unsupervised algorithms. The investigation of this problem is the main target of this paper and will be based on a generic framework presented in [2] which contains several different BSS, MCBD, and acoustic dereverberation algorithms. In this paper we will only consider time-domain algorithms as they are not exhibiting typical frequency-domain narrowband limitations as, e.g., circular convolution effects and permutation of the sources within individual frequency bins. Moreover, for estimation of the cross-relation matrices in the update equations we discuss the covariance method and correlation method as they are known from linear prediction problems [3]. We show that by the use of both methods several well-known algorithms can be incorporated in this framework and thus relationships between existing algorithms can be established.

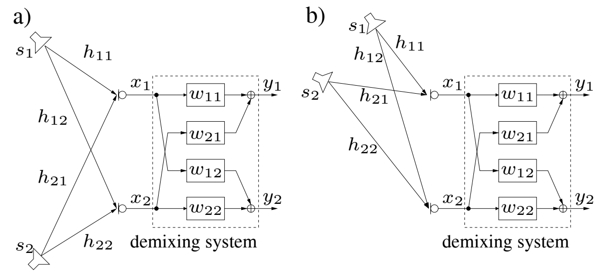

Fig. 1. Setups for BSS requiring (a) only causal delays and (b) causal and acausal delays for the demixing system $\mathbf{W}$.

2. GENERIC TRINICON-BASED UPDATE RULE

For generality, we investigate the acausality problem in the framework presented in [2] and thus, we briefly summarize it in the following. There, a versatile algorithm called TRINICON ('Triple-N ICA for convolutive mixtures) was presented which utilizes all of the following source signal properties to blindly estimate the demixing matrix $\mathbf{W}$ for the above-mentioned tasks of BSS, MCBD, and MCPBD:

(i) Nongaussianity is exploited by using higher-order statistics for independent component analysis (ICA). ICA approaches can be divided into several classes where the minimization of the mutual information (MMI) among the output channels can be regarded as the most general approach for BSS. To obtain an estimator not only allowing spatial separation but also temporal separation for MCBD, the Kullback-Leibler distance (KLD) [5] between a certain *desired* joint pdf (essentially representing a hypothesized stochastic source model) and the joint pdf of the actually estimated output signals is used [2].

(ii) Nonwhiteness is exploited by simultaneous minimization of output cross-relations over multiple time-lags. We therefore con-

sider *multivariate pdfs*, i.e., 'densities including D time-lags'.
(iii) Nonstationarity is exploited by simultaneous minimization of output cross-relations at different time-instants. We assume ergodicity within blocks of length N so that the ensemble average is replaced by time averages over these blocks.

For block processing we first need to formulate the convolution of the FIR demixing system of length L in the following matrix form [4]:

$$\mathbf{y}(m,j) = \mathbf{x}(m,j)\mathbf{W}(m), \tag{1}$$

where m denotes the block index, and $j = 0, \cdots, N-1$ is a time-shift index within a block of length N, and

$$\begin{aligned}
\mathbf{x}(m,j) &= [\mathbf{x}_1(m,j), \ldots, \mathbf{x}_P(m,j)], && (2)\\
\mathbf{y}(m,j) &= [\mathbf{y}_1(m,j), \ldots, \mathbf{y}_P(m,j)], && (3)\\
\mathbf{W}(m) &= \begin{bmatrix} \mathbf{W}_{11}(m) & \cdots & \mathbf{W}_{1P}(m) \\ \vdots & \ddots & \vdots \\ \mathbf{W}_{P1}(m) & \cdots & \mathbf{W}_{PP}(m) \end{bmatrix}, && (4)\\
\mathbf{x}_p(m,j) &= [x_p(mL+j), \ldots, x_p(mL-2L+1+j)], && (5)\\
\mathbf{y}_q(m,j) &= [y_q(mL+j), \ldots, y_q(mL-D+1+j)] && (6)\\
&= \sum_{p=1}^{P} \mathbf{x}_p(m,j)\mathbf{W}_{pq}(m). && (7)
\end{aligned}$$

D in (6) denotes the number of lags taken into account to exploit the nonwhiteness of the source signals as shown below. $\mathbf{W}_{pq}(m)$ denotes a $2L \times D$ Sylvester matrix that contains all coefficients of the respective filter:

$$\mathbf{W}_{pq}(m) = \begin{bmatrix}
w_{pq,0} & 0 & \cdots & 0 \\
w_{pq,1} & w_{pq,0} & \ddots & \vdots \\
\vdots & w_{pq,1} & \ddots & 0 \\
w_{pq,L-1} & \vdots & \ddots & w_{pq,0} \\
0 & w_{pq,L-1} & \ddots & w_{pq,1} \\
\vdots & & \ddots & \vdots \\
0 & \cdots & 0 & w_{pq,L-1} \\
0 & \cdots & 0 & 0 \\
\vdots & & \vdots & \vdots \\
0 & \cdots & 0 & 0
\end{bmatrix}. \tag{8}$$

In [2] the *natural gradient* of a cost function $\mathcal{J}(m)$ with respect to the demixing filter matrix $\mathbf{W}(m)$ was taken

$$\Delta\mathbf{W} \propto \mathbf{W}\mathbf{W}^H \frac{\partial \mathcal{J}}{\partial \mathbf{W}^*} =: \mathbf{W}\Delta\mathbf{T}, \tag{9}$$

where $\Delta\mathbf{T}$ corresponds to the gradient descent in the tangent search space which is projected back to the original manifold of the Euclidean space by the multiplication with $\mathbf{W}$ (for more details see, e.g., [6]). This leads to the following generic TRINICON-based update rule which can be written equivalently to [2] as:

$$\begin{aligned}
\mathbf{W}(m) &= \mathbf{W}(m-1) - \mu\Delta\mathbf{W}(m), && (10)\\
\Delta\mathbf{W}(m) &= \frac{2}{N}\sum_{i=0}^{\infty}\beta(i,m)\mathbf{W}(i) \\
&\quad \cdot \sum_{j=0}^{N-1}\left\{\mathbf{y}^H(i,j)\mathbf{\Phi}_{s,PD}(\mathbf{y}(i,j)) - \mathbf{I}\right\}, && (11)
\end{aligned}$$

where β is a window function with finite support that is normalized according to $\sum_{i=0}^{m}\beta(i,m) = 1$ allowing for online, offline, and block-online algorithms [4]. The *desired* score function

$$\mathbf{\Phi}_{s,PD}(\mathbf{y}(i,j)) = -\frac{\partial \log \hat{p}_{s,PD}(\mathbf{y}(i,j))}{\partial \mathbf{y}(i,j)}, \tag{12}$$

results from the hypothesized source model, where $\hat{p}_{s,PD}(\cdot)$ is the assumed or estimated PD-variate source model (i.e., desired) pdf and D denotes the memory length, i.e., the number of time-lags to model the nonwhiteness of the P signals as above. The generic update equation (11) can be applied to BSS if the desired source pdf $\hat{p}_{s,PD}(\cdot)$ is assumed to be the factorized output signal pdf $\hat{p}_{y_q,D}(\cdot)$ among the sources

$$\hat{p}_{s,PD}(\mathbf{y}(i,j)) \overset{\text{(BSS)}}{=} \prod_{q=1}^{P}\hat{p}_{y_q,D}(\mathbf{y}_q(i,j)). \tag{13}$$

As the factorization is only done among the channels, the source signals are only determined up to an unknown filtering, i.e. the algorithm is not dereverberating the signals picked up by the microphones. A complete factorization leads to an update equation with univariate pdfs

$$\hat{p}_{s,PD}(\mathbf{y}(i,j)) \overset{\text{(MCBD)}}{=} \prod_{q=1}^{P}\prod_{d=1}^{D-1}\hat{p}_{y_q,d}(y_q(iL-d+j)), \tag{14}$$

and thus to the traditional MCBD approach. It should be noted that due to the temporal whitening of the output signals, this approach is not suitable for audio signals.

In [2] it has been shown that also a partial factorization of $\hat{p}_{s,PD}(\cdot)$ is possible which was denoted as multichannel blind *partial* deconvolution (MCBPD) and allows to distinguish between the vocal tract and the reverberant room. Ideally, only the influence of the room acoustics should be minimized leading to a dereverberation without affecting the quality of the audio signals.

3. RELATED ALGORITHM CLASSES AND CAUSALITY PROBLEM IN TIME-DOMAIN BSS

In Fig. 1 it can be seen that depending on the source location acausal delays may be required for blind signal processing in acoustic environments. Based on Fig. 2 we will give an overview in the following sections on how different realizations of the generic algorithm deal with this causality problem. Moreover links between existing MCBD algorithms and also between existing approaches and the generic update (11) become apparent (see Fig. 2).

3.1. Enforcing the Sylvester Constraint $\mathcal{SC}$

When implementing the update rule (11) the Sylvester structure of $\Delta\mathbf{W}$ has to be ensured by using a Sylvester constraint $\mathcal{SC}$. This constraint can be enforced by selecting the L filter taps in the first column of $\Delta\mathbf{W}$ as a reference and generate the Sylvester structure in the form of (8) from it [4, 7]. This implementation is denoted by $\mathcal{SC}_C$. Another option for enforcing the Sylvester constraint is to select the L-th row of $\Delta\mathbf{W}$ as a reference (denoted by $\mathcal{SC}_R$) and generate the Sylvester structure in the form of (8).

The choice of $\mathcal{SC}$ affects the way how the matrix multiplication of $\mathbf{W}\Delta\mathbf{T}$ in (9) is implemented. This can be seen when

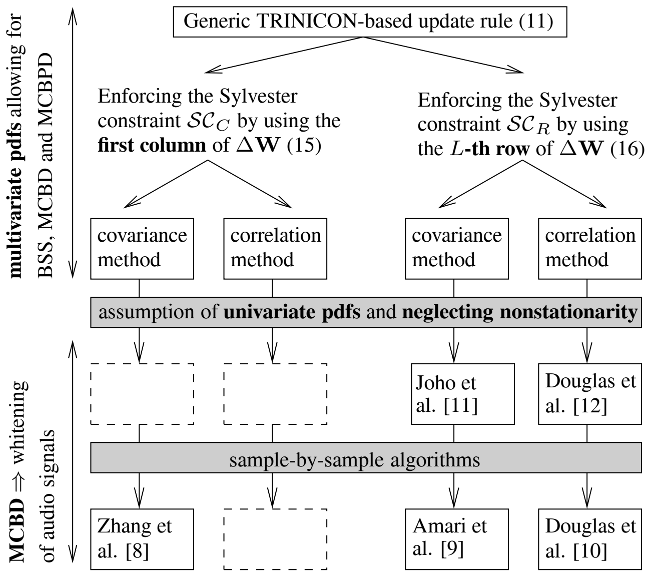

Fig. 2. Overview of links between the generic algorithm (11) and existing MCBD algorithms.

writing (9) element-wise. When using the first column, i.e., $\mathcal{SC}_C$, then the matrix multiplication (9) results in

$$\Delta\tilde{\mathbf{W}}_u = \sum_{v=0}^{u} \tilde{\mathbf{W}}_u \Delta\tilde{\mathbf{T}}_{u-v}, \qquad (15)$$

where $\tilde{\mathbf{W}}_u$ denotes the $P \times P$ demixing filter matrix for the u-th filter tap ($u = 0, \ldots, L-1$) and $\Delta\tilde{\mathbf{T}}_{u-v}$ denotes the $(u-v)$-th tap of the $P \times P$ gradient descent update in the tangent space. Eq. (15) is a convolution of purely causal sequences and thus the index $u - v$ lies within the range $0 \leq u - v \leq L - 1$. Therefore, the resulting elements of $\Delta\tilde{\mathbf{W}}$ are sequences of the same length as those of $\tilde{\mathbf{W}}$ and $\Delta\tilde{\mathbf{T}}$, respectively. Such an operation is called self-closed as the dimension of the manifold, i.e., the length of the sequences which are the elements of $\Delta\tilde{\mathbf{W}}$ does not change. The definition of the natural gradient given in (15) was also used in [8] for an MCBD algorithm based on univariate pdfs (see also Fig. 2).

The second option is to use the L-th row of $\Delta\mathbf{W}$ in (11) to generate the Sylvester structure, i.e., to enforce the Sylvester constraint $\mathcal{SC}_R$. This leads to the following element-wise natural gradient formulation of the matrix multiplication in (9):

$$\Delta\tilde{\mathbf{W}}_u = \sum_{v=0}^{L-1} \tilde{\mathbf{W}}_u \Delta\tilde{\mathbf{T}}_{u-v}. \qquad (16)$$

It can be seen that the only difference to (15) is the upper limit of the sum. As a consequence the index of $\Delta\tilde{\mathbf{T}}_{u-v}$ lies within $-L + 1 \leq u - v \leq L - 1$. Thus, the length of the sequence $\Delta\tilde{\mathbf{T}}$ does not correspond to that of $\tilde{\mathbf{W}}$ and $\Delta\tilde{\mathbf{W}}$. Therefore, with the convolution in (16) the sequence $\Delta\tilde{\mathbf{W}}$ is not self-closed.

The definition of the natural gradient in (16) can be traced back to [9] where it was used in the actual implementation of a sample-by-sample based MCBD algorithm. However, the derivation of the MCBD algorithm in [9] assumed doubly infinite filters and thus, the sequence length of all variables in (16) was doubly infinite resulting in a self-closed convolution. As shown in (16), the subsequent truncation to finite filters of length L violates the self-closedness. In experiments we could observe that algorithms based on $\mathcal{SC}_C$, i.e., (15) are more robust than algorithms based on $\mathcal{SC}_R$ (16).

3.2. Appropriate initialization methods

The choice of the Sylvester constraint $\mathcal{SC}$ also affects the applicable initialization methods. In scenarios where also acausal delays are necessary (see Fig.1b) it is desirable to shift the unit impulse similarly to adaptive beamforming [1] and use the initialization $w_{pp,L/2} = 1$. However, for the update (15), i.e., $\mathcal{SC}_C$ the coefficients of the demixing filter $\mathbf{W}(0)$ should be initialized with zeros except for $w_{pp,0} = 1$. Note that an initialization with $w_{pp,L/2} = 1$ would lead to the problem that due to the summation index in (15) all $\Delta\mathbf{W}_u$ for $0 \leq u \leq L/2 - 1$ would be equal to zero, i.e., these filter coefficients could not be adapted. Thus, the initialization of the first tap of the demixing filter is required. When algorithms based on (15) should be applied to scenarios as shown in Fig.1b the algorithm has to be extended using a filter decomposition approach to incorporate acausal delays as it was shown for traditional MCBD algorithms in [6]. A generalization of this filter decomposition approach to the generic update equation (11) based on multivariate pdfs is well possible.

When enforcing the Sylvester constraint by using $\mathcal{SC}_R$, the algorithm can also be initialized using $w_{pp,L/2} = 1$. Thus, similarly to adaptive beamforming also acausal delays are possible in this case as becomes obvious by simply evaluating (8) for successive iterations.

3.3. Covariance method vs. Correlation method

Similarly to linear prediction problems [3] we have to distinguish in actual implementations of the update equation (11) between two methods to estimate the cross-relation matrices defined by

$$\mathbf{R}_{\mathbf{y}\boldsymbol{\Phi}(\mathbf{y})}(i) = \sum_{j=0}^{N-1} \mathbf{y}^H(i,j) \boldsymbol{\Phi}_{s,PD}(\mathbf{y}(i,j)). \qquad (17)$$

The definition (17) corresponds to the so-called *covariance method*. The covariance method can be approximated by the *correlation method* when stationarity within each block of length $N + D - 1$ is assumed. This leads to a Toeplitz structure of the matrix $\mathbf{R}_{\mathbf{y}\boldsymbol{\Phi}(\mathbf{y})}$ and therefore, also to a lower complexity in calculating (17) and (11). Further details on the implementation of the correlation method for the BSS case of the generic update (11) can be found in [7].

By distinguishing between covariance and correlation method, we can establish a link to the algorithms presented in [11] and [12]. The method in [11] can be obtained from the generic update equation (11) when the pdf contained in the score function $\boldsymbol{\Phi}_{s,PD}$ is factorized according to (14) so that univariate pdfs are obtained (see also Fig. 2). Moreover, in [11] the covariance method is used with $N = 2L$, i.e., the $4L$ most recent x-values are used for the calculation of $\mathbf{R}_{\mathbf{y}\boldsymbol{\Phi}(\mathbf{y})}$. By enforcing the Sylvester structure of $\Delta\mathbf{W}$ with $\mathcal{SC}_R$ we obtain the identical update as in [11].

In [12] a similar block-based algorithm is proposed with the difference that a signal truncation is introduced and therefore only $2L$ x-values are used. In the pseudo-code in [12] it can be seen that this truncation corresponds to using the correlation method on a block of $2L$ x-values. Therefore, this algorithm can be obtained from the generic update equation by enforcing the Sylvester constraint $\mathcal{SC}_R$ in addition with univariate pdfs and the usage of the correlation method.

In Fig. 2 the relationship between existing MCBD algorithms based on univariate pdfs and the BSS, MCBD and MCBPD algorithms incorporating multivariate pdfs which are directly derived from the generic update equation (11) are depicted. Several examples of implementations for the latter class can be found in [2, 4, 7].

3.4. Algorithms based on second order statistics

From the generic update equation (11) also second order statistics (SOS) algorithms can be derived by inserting the Gaussian pdf in the score function (12). In the case of BSS algorithms (13) the multivariate Gaussian pdf has to be used:

$$\hat{p}_{y_q,D}(\mathbf{y}_q(i,j)) = \frac{1}{\sqrt{(2\pi)^D \det(\mathbf{R}_{\mathbf{y}_q\mathbf{y}_q})}} e^{-\frac{1}{2}\mathbf{y}_q \mathbf{R}_{\mathbf{y}_q\mathbf{y}_q}^{-1} \mathbf{y}_q^H} \quad (18)$$

where $\mathbf{R}_{\mathbf{y}_q\mathbf{y}_q}(i) = \frac{1}{N}\sum_{j=0}^{N-1} \mathbf{y}_q^H(i,j)\mathbf{y}_q(i,j)$. This leads to a BSS algorithm which is not whitening the output signals and with the following update equation

$$\Delta\mathbf{W}(m) = 2\sum_{i=0}^{\infty} \beta(i,m)\mathbf{W}(i)\left\{\mathbf{R}_{\mathbf{yy}}(i) - \text{bdiag}\,\mathbf{R}_{\mathbf{yy}}(i)\right\} \text{bdiag}^{-1}\,\mathbf{R}_{\mathbf{yy}}(i), \quad (19)$$

where $\mathbf{R}_{\mathbf{yy}}(i) = \frac{1}{N}\sum_{j=0}^{N-1} \mathbf{y}^H(i,j)\mathbf{y}(i,j)$. The structure in Fig. 2 is also maintained for second order statistics algorithms.

4. EXPERIMENTAL RESULTS

For our experiments we used a BSS algorithm based on SOS (19) which was implemented by using a block-on-line update rule as described in [7]. For the algorithm the correlation method was used to estimate $\mathbf{R}_{\mathbf{yy}}$ and the following parameters were chosen: $L = 1024$, $N = 2048$, and $D = 1024$ with a sampling frequency of 16 kHz. The experiments were conducted in a room with a reverberation time $T_{60} = 50$ ms. A two-element microphone array with a spacing of 21 cm was used. To evaluate the performance we used the signal-to-interference ratio (SIR) defined as the ratio of the signal power of the target signal to the signal power from the jammer signal. A scenario which requires only causal filters (source positions $\pm 70^o$) and one which requires causal and acausal filters (source positions $+45^o$ and $+90^o$) were evaluated. The two different realizations based on the Sylvester constraints $\mathcal{SC}_C$ and $\mathcal{SC}_R$ have been used. In Fig. 3 it can be seen that when only causal filters are required, both realizations lead to similar results. For the scenario requiring acausal delays only the algorithm using $\mathcal{SC}_R$ is applicable as discussed in Sect. 3. It can be seen that it also leads to good separation results for acausal filters. Moreover, the algorithm has been successfully applied to source localization of multiple sources [13].

5. CONCLUSIONS

We examined the problem of acausality in time-domain BSS and MCBD algorithms. Based on a generic framework we showed how different realizations can incorporate acausal demixing filters. Proper initialization methods and cross-relation methods have been discussed and experimental results have been presented. Moreover an overview of the relation to existing algorithms has been given.

6. ACKNOWLEDGMENTS

We thank Julien Bourgeois of DaimlerChrysler for fruitful discussions.

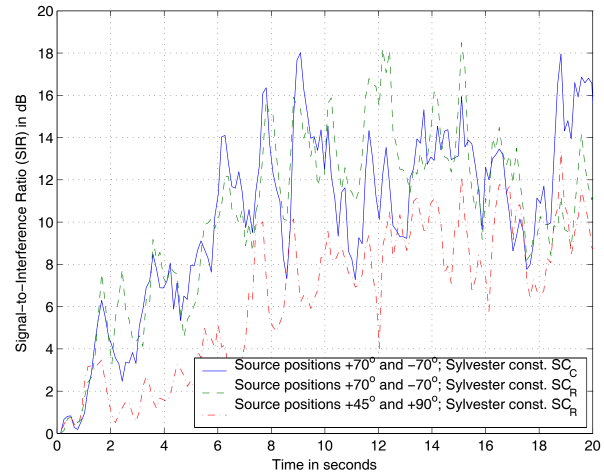

Fig. 3. Experimental results for scenario requiring only causal demixing filters (source pos. $\pm 70^o$) and for a scenario requiring causal and acausal filters (source pos. $+45^o$ and $+90^o$).

7. REFERENCES

[1] B. van Veen and K. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE Trans. Acoust., Speech, Signal Processing*, pp. 4–24, Apr. 1988.

[2] H. Buchner, R. Aichner, and W. Kellermann, "Trinicon: A versatile framework for multichannel blind signal processing," in *Proc. IEEE ICASSP*, Montreal, Canada, May 2004, vol. 3, pp. 889–892.

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

[4] H. Buchner, R. Aichner, and W. Kellermann, "Blind source separation for convolutive mixtures: A unified treatment," in *Audio Signal Processing for Next-Generation Multimedia Communication Systems*, J. Benesty and Y. Huang, Eds., pp. 255–293. Kluwer Academic Publishers, Boston, April 2004.

[5] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, Wiley & Sons, New York, 1991.

[6] L. Zhang, A. Cichocki, and S.-I. Amari, "Multichannel blind deconvolution of nonminimum-phase systems using filter decomposition," *IEEE Trans. Signal Proc.*, vol. 52, no. 5, pp. 1430–1442, May 2004.

[7] R. Aichner, H. Buchner, F. Yan, and W. Kellermann, "Real-time convolutive blind source separation based on a broadband approach," in *Proc. Int. Symp. ICA*, Granada, Spain, Sept. 2004, pp. 833–840.

[8] L. Zhang et al., "Geometrical structures of FIR manifold and their application to multichannel blind deconvolution," in *Proc. IEEE NNSP*, 1999, pp. 303–312.

[9] S.-I. Amari et al., "Multichannel blind deconvolution and equalization using the natural gradient," in *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications*, Paris, France, April 1997, pp. 101–104.

[10] S. Douglas, H. Sawada, and S. Makino, "Natural gradient multichannel blind deconvolution and source separation using causal fir filters," in *Proc. IEEE ICASSP*, May 2004, vol. 5, pp. 477–480.

[11] M. Joho and P. Schniter, "Frequency domain realization of a multichannel blind deconvolution algorithm based on the natural gradient," in *Proc. Int. Symp. ICA*, Japan, April 2003, pp. 543–548.

[12] S. Douglas, H. Sawada, and S. Makino, "A causal frequency-domain implementation of a natural gradient multichannel blind deconvolution and source separation algorithm," in *Proc. Int. Congr. on Acoustics*, Kyoto, Japan, Apr. 2004, vol. 1, pp. 85–88.

[13] H. Buchner, R. Aichner, J. Stenglein, H. Teutsch, and W. Kellermann, "Simultaneous localization of multiple sound sources using blind adaptive MIMO filtering," in *Proc. IEEE ICASSP*, 2005.