RELATION BETWEEN BLIND SYSTEM IDENTIFICATION AND CONVOLUTIVE BLIND SOURCE SEPARATION

Herbert Buchner, Robert Aichner, and *Walter Kellermann*

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

1. INTRODUCTION. Traditionally blind source separation (BSS) has often been considered as an inverse problem. In this paper we show that the theoretically optimum convolutive BSS solution corresponds to blind multiple-input multiple-output (MIMO) system identification. By choosing an appropriate filter length we show that for broadband algorithms the well-known ambiguities can be avoided. Ambiguities in instantaneous BSS algorithms are scaling and permutation [1]. In narrowband convolutive BSS these ambiguities occur independently in each frequency bin so that arbitrary scaling becomes arbitrary filtering [2]. For additional measures to solve the internal permutation problem see, e.g., [2] and for the arbitrary filtering, e.g., [3]. On the other hand broadband time-domain BSS approaches are known to avoid the bin-wise permutation ambiguity. However, traditionally, multichannel blind deconvolution (MCBD) algorithms are often used in the literature [3, 4], which have the drawback of whitening the output signals when applied to acoustic scenarios. Repair measures for this problem have been proposed in [3] (minimum distortion principle) and [4] (linear prediction). In the following we consider the optimum broadband solution of mere separation approaches (MIMO systems, see Fig. 1b) as presented, e.g., in [5], and relate it to the known blind system identification approach based on single-input multiple-output (SIMO) models [6, 7, 8] (Fig. 1a).

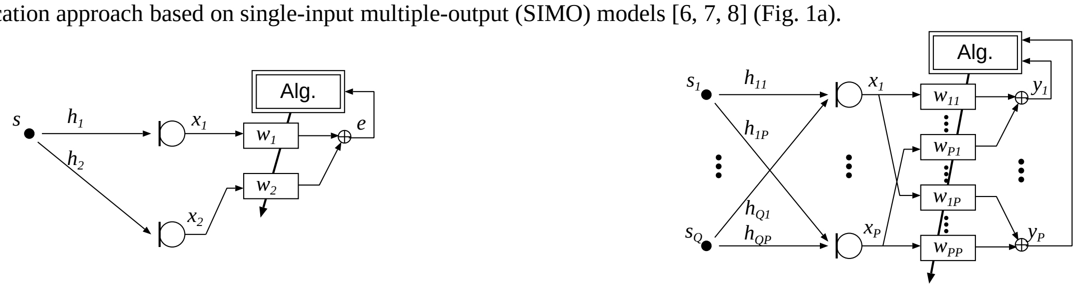

Figure 1: Blind system identification based on (a) SIMO and (b) MIMO models.

2. RELATION BETWEEN BLIND SYSTEM IDENTIFICATION AND BLIND SOURCE SEPARATION. From Fig. 1a and for $e(n) = 0$ it follows for sufficient excitation $s(n)$ that

$$h_1(n) * w_1(n) = -h_2(n) * w_2(n) \tag{1}$$

This can be expressed in the z-domain as $H_1(z)W_1(z) = -H_2(z)W_2(z)$. Without loss of generality we assume that $H_1(z)$, $H_2(z)$ and the adaptive filters $W_i(z)$ exhibit an FIR structure. Thus, the z-domain representations can be expressed by the zeros $z_{0H_i,\nu}$, $z_{0W_i,\mu}$ and the gains A_{H_i}, A_{W_i} of the filters $H_i(z)$ and $W_i(z)$ respectively:

$$A_{H_1}\prod_{\nu=1}^{M-1}(z - z_{0H_1,\nu})A_{W_1}\prod_{\mu=1}^{L-1}(z - z_{0W_1,\mu}) = -A_{H_2}\prod_{\nu=1}^{M-1}(z - z_{0H_2,\nu})A_{W_2}\prod_{\mu=1}^{L-1}(z - z_{0W_2,\mu}) \tag{2}$$

We assume that $H_1(z)$ and $H_2(z)$ have no common zeros. Then the equality of (2) can only hold if the filter length is chosen as $L = M$ and if the zeros $z_{0W_1,\mu} = z_{0H_2,\mu}$ and $z_{0W_2,\mu} = z_{0H_1,\mu}$ for $\mu = 1, \ldots, L-1$. This leads to the optimum filters $W_1(z) = \alpha H_2(z)$ and $W_2(z) = -\alpha H_1(z)$. It can be seen that the optimum filters can only be determined up to an arbitrary scaling by a *scalar* constant $\alpha := \frac{A_{W_1}}{A_{H_2}} = \frac{A_{W_2}}{A_{H_1}}$. Note that this holds only for $L = M$. For $L > M$ the scaling ambiguity would result in arbitrary *filtering*. For the SIMO case this scaling ambiguity was similarily derived in [7]. Adaptive algorithms for this SIMO structure have been proposed in the context of blind deconvolution, e.g., in [6, 7] and for blind identification used for passive source localization, e.g., in [8, 9]. In [6, 9] this SIMO approach was also generalized to more than two microphone channels.

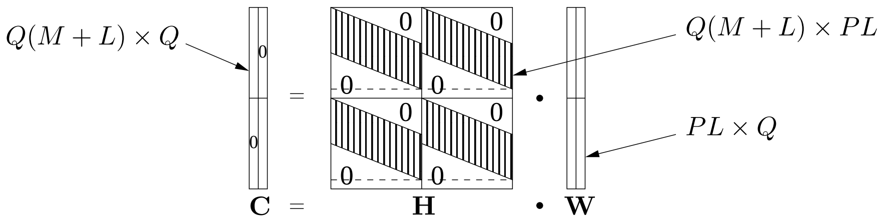

Figure 2: Overall system $\mathbf{C}$ for the MIMO case, illustrated for $P = Q = 2$.

Now let us study the relationship of the SIMO-based method to MIMO systems. As shown in Fig. 2 the Sylvester matrix $\mathbf{H}$ is defined. The filter coefficients of the subfilters h_{ij} in Fig. 1 are represented in the columns of the submatrices of $\mathbf{H}$ [5]. With compatible dimensions, the corresponding matrix for the overall system $\mathbf{C}$ can then be expressed as $\mathbf{C} = \mathbf{HW}$. Obviously, to ideally separate the sources, the cross-channels of the overall system $\mathbf{C}$ must be zero, i.e., $\text{boff}\{\mathbf{C}\} = \mathbf{0}$ together with an uncritical external permutation (which can be easily resolved by using geometrical information). This separation condition has also been rigorously derived from the broadband BSS update equation in [5]. For simplicity, we rewrite $\text{boff}\{\mathbf{C}\} = \mathbf{0}$ for the case of two sources and two microphones. Expressed by the convolution operator, we obtain

$$h_{11} * w_{12} = -h_{12} * w_{22} \quad (3)$$

$$h_{21} * w_{11} = -h_{22} * w_{21} \quad (4)$$

By comparing Fig 1 (a) and (b), we see that (3) and (4) can in fact directly be considered as the generalization of the SIMO identification condition (1) to multiple sources. The similarity of the equations (3), (4) and (1) indicates that BSS performs MIMO system identification. Similarly to the SIMO case we investigate the solutions of the MIMO conditions (3), (4) with respect to the optimum demixing filter length L and a potential filtering ambiguity (the solution of the permutation ambiguity has been addressed in [5]). The optimum filter length L can be determined using the dimensions of the matrix notation in Fig. 2. This was done similarly for the case of deconvolution in [10]. However, since there are no constraints on the direct paths of $\mathbf{C}$ in mere BSS due to the condition $\text{boff}\{\mathbf{C}\} = \mathbf{0}$ we can exclude the respective elements of $\mathbf{H}$ for determining the optimum filter length, i.e., of Q rows of submatrices we only consider $Q-1$ rows of submatrices. This leads to $(Q-1)(M+L) = PL$, i.e., $L_{\text{opt,BSS}} = \frac{Q-1}{P-Q+1}M$. For $P = Q = 2$ we obtain $L_{\text{opt,BSS}} = M$ as in the SIMO case. Interestingly, the general matrix formulation in Fig. 2 can be linked to the results for the deconvolution shown in [10] where the optimum filter length is $L_{\text{opt,MCBD}} = \frac{Q}{P-Q}(M-1)$ and $P > Q$ is required: In [10] the submatrices of $\mathbf{H}$ do not contain the additional rows of zeros (marked by the dashed line in Fig. 2). This results in the modified factor $(M-1)$ in $L_{\text{opt,MCBD}}$. However, it can be shown by taking the pseudo-inverse $\mathbf{H}^+ = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T$ instead of the inverse that the exact deconvolution given in [10] follows rigorously when choosing $L = L_{\text{opt,MCBD}}$. For the case of broadband BSS with correct filter length $L = L_{\text{opt,BSS}}$ we can apply the same reasoning with (3), (4) as in the SIMO case above, showing that the arbitrary filtering reduces to an arbitrary *scaling* only.

3. CONCLUSIONS. In this paper we have shown that convolutive BSS corresponds to blind MIMO system identification. From this we can draw the conclusions that (1) for a suitable choice of the filter length arbitrary filtering is prevented with broadband approaches, (2) the known whitening problem is avoided, and (3) the BSS framework also allows for several new applications, such as simultaneous localization of multiple sources, e.g. [11].

References

[1] A. Hyvaerinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. John Wiley & Sons, 2001.

[2] H. Sawada, R. Mukai, S. Araki, and S. Makino. A robust and precise method for solving the permutation problem of frequency-domain blind source separation. *IEEE Trans. Speech Audio Proc.*, 12(8), September 2004.

[3] K. Matsuoka and S. Nakashima. Minimal distortion principle for blind source separation. In *Proc. Int. Symposium on Independent Component Analysis and Blind Signal Separation (ICA)*, San Diego, CA, USA, Dec. 2001.

[4] S.C. Douglas. Blind separation of acoustic signals. In M. Brandstein and D. Ward, editors, *Microphone Arrays: Signal Processing Techniques and Applications*, pages 355–380. Springer, Berlin, 2001.

[5] H. Buchner, R. Aichner, and W. Kellermann. A generalization of blind source separation algorithms for convolutive mixtures based on second-order statistics. *IEEE Trans. Speech and Audio Proc.*, 13(1), Jan. 2005.

[6] M.I. Gürelli and C.L. Nikias. EVAM: An eigenvector-based algorithm for multichannel blind deconvolution of input colored signals. *Trans. Signal Processing*, 43(1):134–149, Jan. 1995.

[7] K. Furuya and Y. Kaneda. Two-channel blind deconvolution of nonminimum phase FIR systems. *IEICE Trans. Fundamentals*, E80-A(5):804–808, May. 1997.

[8] J. Benesty. Adaptive eigenvalue decomposition algorithm for passive acoustic source localization. *JASA*, 107(1):384–391, Jan. 2000.

[9] J. Chen, Y. Huang, and J. Benesty. Time delay estimation. In Y. Huang and J. Benesty, editors, *Audio Signal Processing for Next-Generation Multimedia Communication Systems*, pages 197–228. Kluwer Academic Publishers, Boston, April 2004.

[10] M. Miyoshi and Y. Kaneda. Inverse filtering of room acoustics. *IEEE Trans. Acoust., Speech, Signal Proc.*, 36(2):145–152, Feb. 1988.

[11] 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*, Philadelphia, PA, USA, Mar. 2005.