Contents

4 TRINICON-based Blind System Identification with Application to Multiple-Source Localization and Separation

Herbert Buchner[1]*, Robert Aichner[2], and Walter Kellermann[2]

[1] Deutsche Telekom Laboratories,
Technical University Berlin,
Ernst-Reuter-Platz 7,
D-10587 Berlin, Germany
E-mail: hb@buchner-net.com

[2] Multimedia Communications and Signal Processing,
University of Erlangen-Nuremberg,
Cauerstr. 7,
D-91058 Erlangen, Germany
E-mail: {aichner, wk}@LNT.de

Abstract. This contribution treats blind system identification approaches and how they can be used to localize multiple sources in environments where multipath propagation cannot be neglected, e.g., acoustic sources in reverberant environments. Based on TRINICON, a general framework for broadband adaptive MIMO signal processing, we first derive a versatile blind MIMO system identification method. For this purpose, the basics of TRINICON will be reviewed to the extent needed for this application, and some new algorithmic aspects will be emphasized. The generic approach then allows us to study various illustrative relations to other algorithms and applications. In particular, it is shown that the optimization criteria used for blind system identification allow a generalization of the well-known Adaptive Eigenvalue Decomposition (AED) algorithm for source localization: Instead of one source as with AED, several sources can be localized simultaneously. Performance evaluation in realistic scenarios will show that this method compares favourably with other state-of-the-art methods for source localization.

4.1 Introduction

4.1.1 Overview

The area of broadband signal aquisition by sensor arrays in multipath or convolutive environments can be divided into two general tasks: the acquisition of clean source signals, and the analysis of the scene, e.g., in order to extract the source positions or the reverberation time of the environment. A challenging and important example for such environments are 'natural' acoustic

* This work was mainly performed while the first author was with Multimedia Communications and Signal Processing, University of Erlangen-Nuremberg.

human/machine interfaces which use multiple microphones to support sound signal acquisition so that the users should be untethered and mobile. Due to the distance between the sources and the sensors, the sensor signal processing generally has to cope with two basic problems, namely the presence of additive noise and interferers, and the disturbing effect of reflections and scattering of the desired source signals in the recordings. Intuitively, if all propagation paths from the desired and interfering sources to all sensors were known exactly we would in principle be able to ideally solve all the above mentioned tasks and associated problems. However, since both the original source signals and the propagation signals are generally unknown in practice, a blind estimation of the propagation paths, i.e., a *blind system identification* (BSI) of the multiple-input multiple-output (MIMO) system is desirable in order to analyze the scene with the given sensor signals.

This chapter consists of two parts. In the first part, consisting of Sect. 4.2 and Sect. 4.3, a general treatment of BSI for MIMO systems is presented, based on TRINICON, a previously introduced versatile framework for broadband adaptive MIMO signal processing [1–4], which is especially well suited for speech and audio signals. We also show a practically important relation between BSI and blind source separation (BSS) for convolutive mixtures. In addition to the inherent broadband structure necessary for a proper system identification, the top-down, i.e., *deductive* approach of the TRINICON framework also allows us to present both relations to already known and new efficient algorithms. Section 4.2 follows the ideas outlined in [5,6]. Some of these ideas were also developed independently in [7] in a slightly different way.

An important and particularly illustrative application of broadband MIMO BSI considered in the second part of this chapter, Sections 4.4 to 4.6, is the acoustic localization of multiple simultaneously active sources in reverberant environments. A popular method to the estimation of the position of an acoustic source in a room is to apply a two-stage approach, consisting of the estimation of *time differences of arrival* (TDOAs) between microphone pairs, followed by the (possibly multidimensional) determination of the position by a purely geometrical calculation. In contrast to another method, based on a farfield assumption and the estimation of *directions of arrival* (DOAs), the TDOA-based method also allows an accurate localization of sources in the nearfield. For the signal processing part of these methods, there are already some popular and conceptually simple approaches in the literature both for a single source, such as the generalized cross-correlation method with its numerous variants [8,9], and for multiple sources, such as the subspace methods, known as, e.g., MUSIC [10] or ESPRIT [11] and their variants, e.g., [9]. However, most of the source localization methods were originally designed only for freefield propagation and/or narrowband applications so that none of the above-mentioned approaches takes multipath propagation and dispersion such as room reverberation in acoustic scenarios into account.

Each of the unmodeled reverberation paths causes an additional peak in the correlation function like an additional source which in turn causes ambiguities in these methods [12–14]. A considerable advantage of the BSI-based source localization method over the conventional correlation-based methods is that due to the explicit multipath model the reverberation does no longer act as a disturbance to the position estimates so that the above-mentioned ambiguity is inherently solved by this method. So far the literature on efficient algorithmic solutions for blind adaptive system identification and their application to source localization has mainly focused on single-input multiple-output (SIMO) systems, i.e., for a single source [15,16]. As we will see in this chapter, the broadband MIMO solution based on TRINICON results in a general multidimensional localization scheme for multiple sources in reverberant environments. The TRINICON-based TDOA estimation for multiple sources was first demonstrated in [6]. Due to the system identification, this approach is also suitable for an accurate localization in the nearfield of the sources. Moreover, a further differentiating and practically important feature of the TRINICON-based approach [6] is that its signal-separating property also inherently resolves a fundamental spatial ambiguity. This ambiguity generally arises in the *multidimensional* case of any multiple-source localization task where the multiple TDOAs/DOAs corresponding to the multiple sources in each dimension must be assigned to the corresponding multiple TDOAs/DOAs of the same sources for the other dimensions. In [17] it was demonstrated that this assignment is made possible due to the inherent blind source separation ability of this approach, i.e., because of the availability of the separated signals due to the relation between BSI and BSS as mentioned above.

4.1.2 Blind adaptive MIMO Filtering Tasks and Matrix Formulation

Blind signal processing on convolutive mixtures of unknown time series is desirable for several application domains, a prominent example being the so-called cocktail party problem in acoustics, where we want to recover the speech signals of multiple speakers who are simultaneously talking in a real room. The room may be very reverberant due to reflections on the walls, i.e., the original source signals $s_q(n)$, $q = 1, \dots, Q$ are filtered by a linear multiple input and multiple output (MIMO) system before they are picked up by the sensors yielding the sensor signals $x_p(n)$, $p = 1, \dots, P$. In this chapter, we describe this MIMO mixing system by length-M finite impulse response (FIR) filters, i.e.,

$$x_p(n) = \sum_{q=1}^{Q} \sum_{\kappa=0}^{M-1} h_{qp,\kappa} s_q(n-\kappa), \tag{4.1}$$

where $h_{qp,\kappa}$, $\kappa = 0, \ldots, M-1$ denote the coefficients of the FIR filter model from the q-th source signal $s_q(n)$ to the p-th sensor signal $x_p(n)$ according to Fig. 4.1. Moreover, we assume throughout this chapter that the number Q of sources is less or equal to the number P of sensors. These cases $Q \leq P$ are of particular interest in the context of blind system identification as detailed below, and they are commonly known as *overdetermined* and *determined*, respectively. Note that in general, the sources $s_q(n)$ may or may not be all simultaneously active at a particular instant of time.

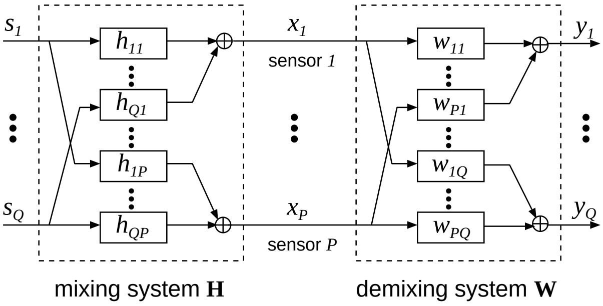


Fig. 4.1. Setup for blind MIMO signal processing.

Obviously, since only the sensor signals, i.e., the output signals of the mixing system are accessible by the blind signal processing, *any* type of linear blind adaptive MIMO signal processing may be described by the serial structure shown in Fig. 4.1. Thus, according to a certain optimization criterion, we are interested in finding a corresponding demixing system by the blind adaptive signal processing whose output signals $y_q(n)$ are described by

$$y_q(n) = \sum_{p=1}^{P} \sum_{\kappa=0}^{L-1} w_{pq,\kappa} x_p(n-\kappa). \tag{4.2}$$

The parameter L denotes the FIR filter length of the demixing filters $w_{pq,\kappa}$.

Depending on the chosen coefficient optimization criterion, we distinguish two general classes of blind signal processing problems[1]:

- **"Direct blind adaptive filtering problems":** This class summarizes here blind system identification (BSI) and blind source separation (BSS)/blind interference cancellation for convolutive mixtures.
  In the BSS approach, we want to determine a MIMO FIR demixing filter which separates the signals *up to an – in general arbitrary – filtering and*

[1] Note that in supervised adaptive filtering we may distinguish the analogous general classes of problems. In this case we classify system identification and interference cancellation after [18] as the "direct supervised adaptive filtering problems", whereas inverse modeling and linear prediction after [18] may be classified as the "inverse supervised adaptive filtering problems".

permutation by forcing the output signals to be mutually independent. Traditionally, and perhaps somewhat misleadingly, BSS has often been considered to be an inverse problem in the literature, e.g., in [19,20]. In another interpretation, BSS may be considered as a set of *blind beamformers* [21,22] under certain restricting conditions, most notably the fulfilment of the spatial sampling theorem by the microphone array. Moreover, under the farfield assumption, the DOAs may be extracted from the corresponding array patterns, which in turn may be calculated from the BSS filter coefficients.

In this chapter we will see that more generally, a properly designed and configured broadband BSS system actually performs blind MIMO system identification (which is independent of the spatial sampling theorem). The general broadband approach shown in this chapter thus allows us to unify the BSS and BSI concepts and provides various algorithmic synergy effects and new applications. This general class of direct blind adaptive signal processing problems is the main focus of this chapter.

- **"Inverse blind adaptive filtering problems":** This class stands here for multichannel blind deconvolution (MCBD)[2] w.r.t. the mixing system $\mathbf{H}$.

  Here, in addition to the separation, we want to recover the original signals *up to an arbitrary* (frequency-independent) *scaling, possibly a time shift, and a permutation*, i.e., in the acoustic applications we want to dereverberate the signals. In terms of the MIMO system description, for this task, effectively, an inversion of (long and usually non-minimum phase) room impulse responses is necessary. However, using the multiple-input/output inverse theorem (MINT) [23] any MIMO system $\mathbf{H}$ can exactly be inverted if P, Q, and L are suitably chosen, and if h_{qp} $\forall$ $p \in \{1, \ldots, P\}$ do not have common zeros in the z-plane. Therefore, in principle, there is a general solution to the MCBD problem by using multiple sensors. Obviously, to realize MCBD two different fundamental approaches are conceivable. One approach is to first perform blind MIMO system identification as mentioned above, followed by a (MINT-based) inversion of the estimated mixing system, e.g., [25,26]. The other, theoretically equivalent but in practice often more reliable approach is to perform directly a blind estimation of the actual inverse of the MIMO mixing system, e.g., [3,27–29]. As we may expect, in any case there are various relations between the general classes of direct and inverse blind adaptive filtering problems, i.e., BSI and MCBD, and the corresponding algorithms. As a side aspect, this chapter also tries to highlight some of these relations.

[2] Later in Sect. 4.3.2 we will see that in practical systems for the blind deconvolution tasks it is important to take the spectral characteristics of the source signals into account. The method of multichannel blind partial deconvolution (MCBPD), introduced in Sect. 4.3.2 to address this issue also belongs to the class of inverse blind adaptive filtering problems.

Matrix formulation. To analyze and to formulate the above-mentioned blind adaptive MIMO filtering problems compactly, we introduce the following matrix formulation of the overall system consisting of the mixing and demixing systems. Moreover, this matrix formulation is also used directly in the TRINICON framework described later in Sect. 4.3 in order to blindly estimate the adaptive demixing filter coefficients.

As a compact formulation of the mixing filter coefficients $h_{qp,\kappa}$, $\kappa = 0,\ldots,M-1$ and the demixing filter coefficients $w_{pq,\kappa}$, $\kappa = 0,\ldots,L-1$, $p = 1,\ldots,P$, $q = 1,\ldots,Q$, we form the $QM \times P$ mixing coefficient matrix

$$\check{\mathbf{H}} = \begin{bmatrix} \mathbf{h}_{11} & \cdots & \mathbf{h}_{1P} \\ \vdots & \ddots & \vdots \\ \mathbf{h}_{Q1} & \cdots & \mathbf{h}_{QP} \end{bmatrix} \tag{4.3}$$

and the $PL \times Q$ demixing coefficient matrix

$$\check{\mathbf{W}} = \begin{bmatrix} \mathbf{w}_{11} & \cdots & \mathbf{w}_{1Q} \\ \vdots & \ddots & \vdots \\ \mathbf{w}_{P1} & \cdots & \mathbf{w}_{PQ} \end{bmatrix}, \tag{4.4}$$

respectively, where

$$\mathbf{h}_{qp} = [h_{qp,0},\ldots,h_{qp,M-1}]^{\mathrm{T}}, \tag{4.5}$$

$$\mathbf{w}_{pq} = [w_{pq,0},\ldots,w_{pq,L-1}]^{\mathrm{T}} \tag{4.6}$$

denote the coefficient vectors of the FIR subfilters of the MIMO systems, and superscript $^{\mathrm{T}}$ denotes transposition of a vector or a matrix. The downwards pointing hat symbol on top of $\mathbf{H}$ and $\mathbf{W}$ in (4.3) and (4.4) serves to distinguish these *condensed* matrices from the corresponding larger matrix structures as introduced below in (4.10) for the case of the mixing system. The rigorous distinction between these different matrix structures is also an essential aspect of the general TRINICON framework as shown later.

Analogously, the coefficients $c_{qr,\kappa}$, $q = 1,\ldots,Q$, $r = 1,\ldots,Q$, $\kappa = 0,\ldots,M+L-2$ of the overall system of length $M+L-1$ from the sources to the adaptive filter outputs are combined into the $Q(M+L-1)\times Q$ matrix

$$\check{\mathbf{C}} = \begin{bmatrix} \mathbf{c}_{11} & \cdots & \mathbf{c}_{1Q} \\ \vdots & \ddots & \vdots \\ \mathbf{c}_{Q1} & \cdots & \mathbf{c}_{QQ} \end{bmatrix}, \tag{4.7}$$

where

$$\mathbf{c}_{qr} = [c_{qr,0},\ldots,c_{qr,M+L-2}]^{\mathrm{T}}. \tag{4.8}$$

All these subfilter coefficients $c_{qr,\kappa}$ are obtained by convolving the mixing filter coefficients with the demixing filter coefficients. In general, a convolution of two such finite-length sequences can also be written as a matrix-vector product so that the coefficient vector for the model from the q-th source to the r-th output reads here

$$\mathbf{c}_{qr} = \sum_{p=1}^{P} \mathbf{H}_{qp,[L]} \mathbf{w}_{pr}. \tag{4.9}$$

The so-called *convolution* or *Sylvester matrix* $\mathbf{H}_{qp,[L]}$ of size $M + L - 1 \times L$ in this equation exhibits a special structure, containing the M filter taps in each column,

$$\mathbf{H}_{qp,[L]} = \begin{bmatrix} h_{qp,0} & 0 & \cdots & 0 \\ h_{qp,1} & h_{qp,0} & \ddots & \vdots \\ \vdots & h_{qp,1} & \ddots & 0 \\ h_{qp,M-1} & \vdots & \ddots & h_{qp,0} \\ 0 & h_{qp,M-1} & \ddots & h_{qp,1} \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & h_{qp,M-1} \end{bmatrix}. \tag{4.10}$$

The additional third index in brackets denotes the width of the Sylvester matrix which has to correspond to the length of the column vector $\mathbf{w}_{pr}$ in (4.9) so that the matrix-vector product is equivalent to a linear convolution. The brackets serve to emphasize this fact and to clearly distinguish the meaning of this index from the meaning of the third index of the individual elements of the matrices, e.g., in (4.10).

We may now express the overall system matrix $\check{\mathbf{C}}$ after (4.7) compactly using this Sylvester matrix formulation to finally obtain

$$\check{\mathbf{C}} = \mathbf{H}_{[L]} \check{\mathbf{W}}, \tag{4.11}$$

where $\mathbf{H}_{[L]}$ denotes the $Q(M + L - 1) \times PL$ MIMO block Sylvester-matrix combining all channels,

$$\mathbf{H}_{[L]} = \begin{bmatrix} \mathbf{H}_{11,[L]} & \cdots & \mathbf{H}_{1P,[L]} \\ \vdots & \ddots & \vdots \\ \mathbf{H}_{Q1,[L]} & \cdots & \mathbf{H}_{QP,[L]} \end{bmatrix}. \tag{4.12}$$

Based on this matrix formulation, we are now in a position to formulate the blind adaptive MIMO filtering tasks compactly, and to discuss the corresponding ideal solutions, regardless of how the adaptation is actually performed in practice (note that this also means that the results are valid for both blind and supervised adaptation). The blind adaptation of the coefficients towards these ideal solutions will be treated later in Sect. 4.3.

Condition for Ideal Solution of Inverse Adaptive Filtering Problems (*Ideal Inversion Solution*). As mentioned above, the aim of the inverse adaptive filtering problem is to recover the original signals $s_q(n)$, $q = 1, \ldots, Q$, as shown in Fig. 4.1, up to an arbitrary frequency-independent scaling, time shift, and possibly a permutation of the demixing filter outputs. Disregarding the potential permutation among the output signals[3], this condition may be expressed in terms of an *ideal overall system matrix*

$$\check{\mathbf{C}}_{\text{ideal,inv}} = \text{Bdiag}\left\{[0,\ldots,0,1,0,\ldots,0]^{\text{T}},\ldots,[0,\ldots,0,1,0,\ldots,0]^{\text{T}}\right\}\boldsymbol{\Lambda}_{\alpha}, \tag{4.13}$$

where the Bdiag$\{\cdot\}$ operator describes a block-diagonal matrix containing the listed vectors on the main diagonal. Here, these target vectors represent pure delays. The diagonal matrix $\boldsymbol{\Lambda}_{\alpha} = \text{Diag}\left\{[\alpha_1,\ldots,\alpha_P]^{\text{T}}\right\}$ accounts for the scaling ambiguity. The *condition for the ideal inversion solution* thus reads

$$\mathbf{H}_{[L]}\check{\mathbf{W}} = \check{\mathbf{C}}_{\text{ideal,inv}}. \tag{4.14}$$

This system of equations may generally be solved exactly or approximately by the Moore-Penrose pseudoinverse, denoted by $\cdot^{+}$, so that

$$\begin{aligned}\check{\mathbf{W}}_{\text{LS,inv}} &= \mathbf{H}_{[L]}^{+}\check{\mathbf{C}}_{\text{ideal,inv}}\\ &= \left[\mathbf{H}_{[L]}^{\text{T}}\mathbf{H}_{[L]}\right]^{-1}\mathbf{H}_{[L]}^{\text{T}}\check{\mathbf{C}}_{\text{ideal,inv}}.\end{aligned} \tag{4.15}$$

Note that this expression corresponds to the least-squares (LS) solution

$$\check{\mathbf{W}}_{\text{LS,inv}} = \arg\min_{\check{\mathbf{W}}}\|\mathbf{H}_{[L]}\check{\mathbf{W}} - \check{\mathbf{C}}_{\text{ideal,inv}}\|^2. \tag{4.16}$$

It can be shown that under certain practically realizable conditions this solution becomes the ideal inversion solution, i.e., the pseudoinverse in (4.15) turns into the true matrix inverse,

$$\check{\mathbf{W}}_{\text{ideal,inv}} = \mathbf{H}_{[L]}^{-1}\check{\mathbf{C}}_{\text{ideal,inv}}. \tag{4.17}$$

This method is known as the Multiple-input/output INverse Theorem (MINT) [23] and is applicable even for mixing systems with nonminimum phase. The basic requirement for $\mathbf{H}_{[L]}$ in order to be invertible is that it is of full rank. This assumption can be interpreted such that the FIR acoustic impulse responses contained in $\mathbf{H}_{[L]}$ do not possess any common zeros in the z-domain, which usually holds in practice for a sufficient number of sensors

[3] It could formally be described by an additional permutation matrix in the ideal solution. However, since in many practical cases this ambiguity may easily be resolved (e.g., by a correlation analysis), we renounced on this formal treatment for clarity.

[23]. Another requirement for invertibility of $\mathbf{H}_{[L]}$ is that the number of its rows equals the number of its columns, i.e., $Q(M+L-1) = PL$ according to the dimensions noted above (4.12). From this condition, we immediately obtain the *optimum filter length for inversion* [24]:

$$L_{\text{opt,inv}} = \frac{Q}{P-Q}(M-1). \tag{4.18}$$

An important conclusion of this consideration is that the MIMO mixing system can be inverted exactly even with a finite-length MIMO demixing system, as long as $P > Q$, i.e., the number of sensors is greater than the number of sources. Note that P, Q, M must be such that $L_{\text{opt,inv}}$ is an integer number in order to allow the matrix inversion in (4.17). Otherwise, we have to resort to the LS approximation (4.15) with $L_{\text{opt,inv}} = \lceil Q(M-1)/(P-Q) \rceil$.

Conditions for Ideal Solution of Signal Separation Problems (*Ideal Separation Solution*). The goal of any separation algorithm, such as BSS or conventional beamforming, is to eliminate the crosstalk between the different sources $s_q(n)$, $q = 1, \ldots, Q$, as shown in Fig. 4.1, in the output signals $y_q(n)$, $q = 1, \ldots, Q$ of the demixing system. Disregarding again a potential permutation among the output signals as above, this condition may be expressed in terms of the overall system matrix $\check{\mathbf{C}}$ as

$$\check{\mathbf{C}} - \text{bdiag}\left\{\check{\mathbf{C}}\right\} = \text{boff}\left\{\check{\mathbf{C}}\right\} = \mathbf{0}. \tag{4.19}$$

Here, the operator bdiag$\{\cdot\}$ applied to a block matrix consisting of several submatrices or vectors sets all submatrices or vectors on the off-diagonals to zero. Analogously, the boff$\{\cdot\}$ operation sets all submatrices or vectors on the diagonal to zero.

With the overall system matrix (4.11), the condition for the ideal separation is expressed as

$$\text{boff}\left\{\mathbf{H}_{[L]}\check{\mathbf{W}}\right\} = \mathbf{0}. \tag{4.20}$$

This relation for the ideal solution of the *direct blind adaptive filtering problems* is the analogous expression to the relation (4.14) for the ideal solution of the inverse blind adaptive filtering problems.

As we will see in the next section, the relation (4.20) allows us

- to derive an explicit expression of the ideal separation solution analogously to (4.17)
- to establish a link between BSS and BSI
- to establish the conditions for BSI
- to derive the optimum separating FIR filter length $L_{\text{opt,sep}}$ analogously to (4.18) for which the ideal separation solution (4.19) can be achieved.

If we are only interested in separation with certain other constraints to the output signals (e.g., minimal signal distortion between sensor signals and output signals), but not in system identification, we may impose further explicit conditions to the block-diagonal elements of $\mathbf{H}_{[L]}\check{\mathbf{W}}$ in addition to the condition (4.20) on the block-offdiagonals. For instance, the so-called *minimum distortion principle* after [30] may in fact be regarded as such an additional condition. However, since this is not within the scope of system identification we will not discuss these conditions further in this chapter.

4.2 Blind MIMO System Identification and Relation to Blind Source Separation

Traditionally, blind source separation (BSS) has often been considered as an inverse problem. In this section we show that the theoretically ideal convolutive (blind) source separation solution corresponds to blind MIMO system identification. By choosing an appropriate filter length we show that for broadband algorithms the well-known filtering ambiguity can be avoided. Ambiguities in instantaneous BSS algorithms are scaling and permutation [19]. In narrowband convolutive BSS these ambiguities occur independently in each frequency bin so that arbitrary scaling becomes arbitrary filtering, as mentioned above. For additional measures to solve the internal permutation problem appearing independently in each frequency bin, see, e.g., [31] and for the arbitrary filtering, e.g., [30]. 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 [20,30], which have the drawback of whitening the output signals when applied to acoustic scenarios. Repair measures for this problem have been proposed in [30] (minimum distortion principle) and in [20] (linear prediction). In the following we consider the ideal broadband solution of mere MIMO separation approaches and relate it to the known blind system identification approach based on single-input multiple-output (SIMO) models [15,25,26]. This section follows the ideas outlined in [5,6]. Some of these ideas were also developed independently in [7] in a slightly different way.

This section discusses the ideal separation condition $\operatorname{boff}\left\{\mathbf{H}_{[L]}\check{\mathbf{W}}\right\} = \mathbf{0}$ illustrated in Fig. 4.2 for the case $Q = P = 3$. Since in this equation we impose explicit constraints only on the block-offdiagonal elements of $\check{\mathbf{C}}$, this is equivalent to establishing a set of homogeneous systems of linear equations

$$\mathbf{H}_{(:\backslash q):,[L]}\check{\mathbf{W}}_{:q} = \mathbf{0}, \quad q = 1, \ldots, Q \tag{4.21}$$

to be solved. Each of these systems of equations results from the constraints on one column of $\check{\mathbf{C}}$, as illustrated in Fig. 4.2 for the first column. The notation in the indices in (4.21) indicates that for the q-th column $\check{\mathbf{W}}_{:q}$ of the demixing

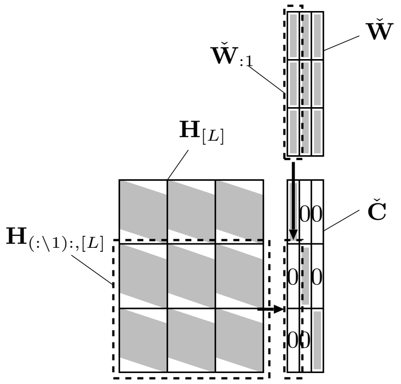


Fig. 4.2. Overall system $\check{\mathbf{C}}$ for the ideal separation, illustrated for $P = Q = 3$.

filter matrix $\check{\mathbf{W}}$, we form a submatrix $\mathbf{H}_{(:\backslash q):,[L]}$ of $\mathbf{H}_{[L]}$ by striking out the q-th row $\mathbf{H}_{q:,[L]}$ of Sylvester-submatrices of the original matrix $\mathbf{H}_{[L]}$.

For homogeneous systems of linear equations such as (4.21) it is known that non-trivial solutions $\check{\mathbf{W}}_{:q} \not\equiv \mathbf{0}$ are indeed obtained if the rank of $\mathbf{H}_{(:\backslash q):,[L]}$ is smaller than the number of elements of $\check{\mathbf{W}}_{:q}$. Later in this section, we will also derive an expression of the optimum separation filter length $L_{\mathrm{opt,sep}}$ for an arbitrary number of sensors and sources analogously to the optimum inversion filter length $L_{\mathrm{opt,inv}}$ in (4.18). This derivation will be based on this observation.

In the following subsections, we first discuss the solution of (4.21) for the case $P = Q = 2$, and then generalize the results to more than two sources and sensors.

4.2.1 Square Case for Two Sources and Two Sensors

For the case $Q = P = 2$, the set of homogeneous linear systems of equations (4.21) reads

$$\mathbf{H}_{11,[L]}\mathbf{w}_{12} + \mathbf{H}_{12,[L]}\mathbf{w}_{22} = \mathbf{0}, \tag{4.22a}$$

$$\mathbf{H}_{21,[L]}\mathbf{w}_{11} + \mathbf{H}_{22,[L]}\mathbf{w}_{21} = \mathbf{0}. \tag{4.22b}$$

Since the matrix-vector products in these equations represent convolutions of FIR filters they can equivalently be written as a multiplication in the z-domain:

$$H_{11}(z)W_{12}(z) + H_{12}(z)W_{22}(z) = 0, \tag{4.23a}$$

$$H_{21}(z)W_{11}(z) + H_{22}(z)W_{21}(z) = 0. \tag{4.23b}$$

Due to the FIR filter structure the z-domain representations can be expressed by the zeros $z_{0H_{qp},\nu}$, $z_{0W_{pq},\mu}$ and the gains $A_{H_{qp}}$, $A_{H_{pq}}$ of the filters $H_{qp}(z)$

and $W_{pq}(z)$, respectively:

$$A_{H_{11}} \prod_{\nu=1}^{M-1} (z - z_{0H_{11},\nu}) A_{W_{12}} \prod_{\mu=1}^{L-1} (z - z_{0W_{12},\mu}) = \\ - A_{H_{12}} \prod_{\nu=1}^{M-1} (z - z_{0H_{12},\nu}) A_{W_{22}} \prod_{\mu=1}^{L-1} (z - z_{0W_{22},\mu}), \tag{4.24a}$$

$$A_{H_{21}} \prod_{\nu=1}^{M-1} (z - z_{0H_{21},\nu}) A_{W_{11}} \prod_{\mu=1}^{L-1} (z - z_{0W_{11},\mu}) = \\ - A_{H_{22}} \prod_{\nu=1}^{M-1} (z - z_{0H_{22},\nu}) A_{W_{21}} \prod_{\mu=1}^{L-1} (z - z_{0W_{21},\mu}). \tag{4.24b}$$

Analogously to the case of MINT [23] described in the previous section, we assume that the impulse responses contained in $\mathbf{H}_{(:\backslash q):,[L]}$, i.e., $H_{11}(z)$ and $H_{12}(z)$ in (4.24a) do not share common zeros. In the same way, we assume that $H_{21}(z)$ and $H_{22}(z)$ in (4.24b) do not share common zeros. If no common zeros exist and if we choose the *optimum filter length for the case* $Q = P = 2$ as $L_{\text{opt,sep}} = M$, then the equality in (4.24a) can only hold if the zeros of the demixing filters are chosen as $z_{0W_{12},\mu} = z_{0H_{12},\mu}$ and $z_{0W_{22},\mu} = z_{0H_{11},\mu}$ for $\mu = 1, \ldots, M-1$. Analogously, the equality in (4.24b) can only hold if $z_{0W_{11},\mu} = z_{0H_{22},\mu}$ and $z_{0W_{21},\mu} = z_{0H_{21},\mu}$ for $\mu = 1, \ldots, M-1$. Additionally, to fulfill the equality, the gains of the demixing filters in (4.24a) have to be chosen as $A_{W_{22}} = \alpha_2 A_{H_{11}}$ and $A_{W_{12}} = -\alpha_2 A_{H_{12}}$, where α_2 is an arbitrary scalar constant. Thus, the demixing filters are only determined up to a scalar factor α_2. Analogously, for the equality (4.24b) the gains of the demixing filters are given as $A_{W_{11}} = \alpha_1 A_{H_{22}}$ and $A_{W_{21}} = -\alpha_1 A_{H_{21}}$ with the scalar constant α_1.

In summary, this leads to the ideal separating filter matrix $\check{\mathbf{W}}_{\text{ideal,sep}}$ given in the time domain as

$$\check{\mathbf{W}}_{\text{ideal,sep}} = \begin{bmatrix} \alpha_1 \mathbf{h}_{22} & -\alpha_2 \mathbf{h}_{12} \\ -\alpha_1 \mathbf{h}_{21} & \alpha_2 \mathbf{h}_{11} \end{bmatrix} = \begin{bmatrix} \mathbf{h}_{22} & -\mathbf{h}_{12} \\ -\mathbf{h}_{21} & \mathbf{h}_{11} \end{bmatrix} \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix}, \tag{4.25}$$

where due to the scaling ambiguity each column is multiplied by an unknown scalar α_q.

From (4.25) we see that under the conditions put on the zeros of the mixing system in the z-domain, and for $L = L_{\text{opt,sep}}$, this *ideal separation solution corresponds to a MIMO system identification up to an arbitrary scalar constant.* Thus, a suitable algorithm which is able to perform *broadband* BSS *under these conditions* can be used for blind MIMO system identification (if the source signals provide sufficient spectral and temporal support for exciting the mixing system). In Section 4.3, we present such a suitable algorithmic framework for this task. Moreover, as we will see in the following subsection, this approach may be seen as a generalization of the state-of-the-art method for the blind identification of SIMO systems.

Finally, since we did not impose an explicit constraint on the block-diagonal elements of the overall system $\check{\mathbf{C}}$ in the original separation condition (4.20), we are now interested in the resulting overall system in the case of the ideal separating solution (4.25). By inserting this solution into (4.11), we readily obtain

$$\begin{aligned}
\check{\mathbf{C}}_{\text{ideal,sep}} &= \mathbf{H}_{[M]}\check{\mathbf{W}}_{\text{ideal,sep}} \\
&= \begin{bmatrix} \mathbf{H}_{11,[M]} & \mathbf{H}_{12,[M]} \\ \mathbf{H}_{21,[M]} & \mathbf{H}_{22,[M]} \end{bmatrix} \begin{bmatrix} \alpha_1\mathbf{h}_{22} & -\alpha_2\mathbf{h}_{12} \\ -\alpha_1\mathbf{h}_{21} & \alpha_2\mathbf{h}_{11} \end{bmatrix} \\
&= \begin{bmatrix} \alpha_1\left(\mathbf{H}_{11,[M]}\mathbf{h}_{22} - \mathbf{H}_{12,[M]}\mathbf{h}_{21}\right) & \alpha_2\left(\mathbf{H}_{12,[M]}\mathbf{h}_{11} - \mathbf{H}_{11,[M]}\mathbf{h}_{12}\right) \\ \alpha_1\left(\mathbf{H}_{12,[M]}\mathbf{h}_{22} - \mathbf{H}_{22,[M]}\mathbf{h}_{21}\right) & \alpha_2\left(\mathbf{H}_{22,[M]}\mathbf{h}_{11} - \mathbf{H}_{21,[M]}\mathbf{h}_{12}\right) \end{bmatrix} \\
&= \begin{bmatrix} \alpha_1\left(\mathbf{H}_{11,[M]}\mathbf{h}_{22} - \mathbf{H}_{12,[M]}\mathbf{h}_{21}\right) & \mathbf{0} \\ \mathbf{0} & \alpha_2\left(\mathbf{H}_{22,[M]}\mathbf{h}_{11} - \mathbf{H}_{21,[M]}\mathbf{h}_{12}\right) \end{bmatrix},
\end{aligned} \tag{4.26}$$

where in the last line the commutativity of the convolution has been exploited so that the crosstalk between the channels is cancelled out perfectly. The output signals of the overall system are filtered (but not arbitrarily filtered) versions of the original source signals.

4.2.2 Relation to SIMO System Identification

BSS algorithms aiming at the ideal solution (4.25) can be interpreted as a generalization of the popular class of blind SIMO system identification approaches, e.g., [25,26,32], as illustrated in Fig. 4.3a. The main reason for

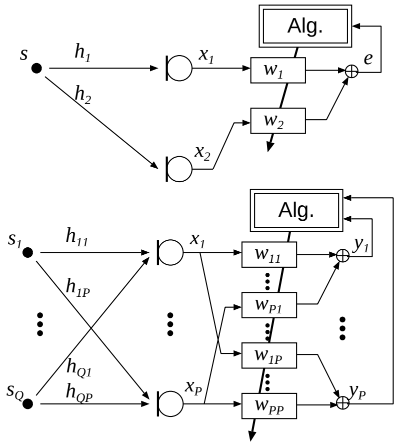


Fig. 4.3. Blind system identification based on (a) SIMO and (b) MIMO models.

the popularity of this SIMO approach is that it can be implemented as a relatively simple least-squares error minimization. From Fig. 4.3a and for $e(n) = 0$ it follows for sufficient excitation $s(n)$ that

$$h_1(n) * w_1(n) = -h_2(n) * w_2(n). \tag{4.27}$$

This can be expressed in the z-domain as $H_1(z)W_1(z) = -H_2(z)W_2(z)$. Comparing this error cancelling condition with the ideal separation conditions (4.23a)/(4.23b), we immediately see that the SIMO-based approach indeed corresponds exactly to one of the separation conditions, and for deriving the ideal solution, we may apply exactly the same reasoning as in the MIMO case above. Thus, assuming that $H_1(z)$ and $H_2(z)$ have no common zeros, the equality of (4.27) can only hold if the filter length is chosen again as $L = M$. Then, this leads to the ideal cancellation filters $W_1(z) = \alpha H_2(z)$ and $W_2(z) = -\alpha H_1(z)$ which can be determined up to an arbitrary scaling by the factor α as in the MIMO case. For $L > M$ the scaling ambiguity would result in arbitrary *filtering*. For the SIMO case this scaling ambiguity was derived similarily in [26].

Note that the SIMO case may also be interpreted as a special 2 × 2 MIMO case according to Fig. 4.3b with the specialization being that one of the sources is always identical to zero so that the BSS output corresponding to this (virtual) source also must always be identical to zero, whereas the other BSS output signal is not of interest in this case. This leads again to the cancellation condition (4.27), and illustrates that the relation between broadband BSS and SIMO-based BSI will also hold from an algorithmic point of view, i.e., known adaptive solutions for SIMO BSI may also be derived as special cases of the algorithmic framework for the MIMO case.

Adaptive algorithms performing the error minimization mentioned above for the SIMO structure have been proposed in the context of blind deconvolution, e.g., in [25,26] and for blind identification used for passive source localization, e.g., in [15,16]. It is also known as the *adaptive eigenvalue decomposition* (AED) algorithm. This name comes from the fact that in the SIMO case, the homogeneous system of equations (4.21) may be reformulated into an analogous signal-dependent homogeneous system of equations containing the sensor-signal correlation matrix instead of the mixing filter matrix. The solution vector (in the SIMO case the matrix $\check{\mathbf{W}}$ reduces to a vector) of the homogeneous system can then be interpreted as the *eigenvector corresponding to the zero-valued (or smallest) eigenvalue* of the sensor correlation matrix. In [16,25] this SIMO approach, i.e., the single-source case, was also generalized to more than two microphone channels.

4.2.3 Ideal Separation Solution in the General Square Case for More than Two Sources and Sensors

The factorized formulation of the ideal separation solution for the case $Q = P = 2$ in the second part of (4.25) suggests that it may be expressed more generally by the adjoint of the matrix $\check{\mathbf{H}}$ where the entries $\mathbf{h}_{qp}$ are treated like scalar values. We formalize this operation and call it the *block-adjoint operator* $\mathrm{badj}_P\{\cdot\}$, where the index P denotes the number of submatrices in each row or column of the argument. Using the block-adjoint operator the general form of (4.25) for an arbitrary number $Q = P$ of sensors or sources reads [5]

$$\check{\mathbf{W}}_{\text{ideal,sep}} = \mathrm{badj}_P \left\{\check{\mathbf{H}}\right\} \ \boldsymbol{\Lambda}_\alpha, \tag{4.28}$$

where the diagonal matrix $\boldsymbol{\Lambda}_\alpha = \mathrm{Diag}\left\{[\alpha_1, \dots, \alpha_P]^{\mathrm{T}}\right\}$ again describes the scaling ambiguity. Note that the size of $\mathrm{badj}_P \left\{\check{\mathbf{H}}\right\}$ is determined for $P, Q > 2$ by the internal convolutions of the FIR filters contained in $\check{\mathbf{H}}$. We may easily verify that the resulting size after the convolutions is $[P(P-1)(M-1)+1] \times P$.

To verify that the approach (4.28) is indeed the ideal separating solution for $P, Q \geq 2$, we may calculate the overall system matrix as in the 2×2 case above. Extending the well-known property of the conventional adjoint of a square matrix $\mathbf{A}$

$$\mathbf{A}\mathrm{adj}\left\{\mathbf{A}\right\} = \det\left\{\mathbf{A}\right\} \cdot \mathbf{I} = \mathrm{Diag}\left\{\det\left\{\mathbf{A}\right\}, \dots, \det\left\{\mathbf{A}\right\}\right\} \tag{4.29}$$

to the analogous formulation of the block-adjoint it may be shown that

$$\begin{aligned}\check{\mathbf{C}}_{\text{ideal,sep}} &= \mathbf{H}_{[L]}\mathrm{badj}_P \left\{\check{\mathbf{H}}\right\} \ \boldsymbol{\Lambda}_\alpha \\ &= \mathrm{Bdiag}\left\{\mathrm{bdet}_P \left\{\check{\mathbf{H}}\right\}, \dots, \mathrm{bdet}_P \left\{\check{\mathbf{H}}\right\}\right\} \ \boldsymbol{\Lambda}_\alpha,\end{aligned} \tag{4.30}$$

where the operator $\mathrm{bdet}_P \left\{\check{\mathbf{H}}\right\}$ denotes a *block-determinant operator* on the mixing system. Similarly to the block-adjoint, the block-determinant operator generalizes the conventional determinant operator so that we work with submatrices as its entries rather than scalar values. Thus, in contrast to the conventional determinant the block-determinant $\mathrm{bdet}_P \left\{\check{\mathbf{H}}\right\}$ is still a matrix and can be interpreted as a MIMO system with FIR filters of length $P(M-1)+1$ due to the P internal convolutions. The submatrices on the block-diagonal in (4.26) represent this operation for the 2×2 case.

4.2.4 Ideal Separation Solution and Optimum Separating Filter Length for an Arbitrary Number of Sources and Sensors

As mentioned above, for homogeneous systems of linear equations such as the ideal separation conditions (4.21) it is known that non-trivial solutions $\check{\mathbf{W}}_{:q} \not\equiv \mathbf{0}$ are obtained if the rank of $\mathbf{H}_{(:\backslash q):,[L]}$ is smaller than the number of elements of $\check{\mathbf{W}}_{:q}$. Additionally, as in the case of MINT [23] described in the previous section, we assume that the impulse responses contained in $\mathbf{H}_{(:\backslash q):,[L]}$ do not share common zeros in the z-domain so that $\mathbf{H}_{(:\backslash q):,[L]}$ is assumed to have full row rank. Thus, combining these conditions leads to the requirement that the matrix $\mathbf{H}_{(:\backslash q):,[L]}$ is *wide*, i.e., the number PL of its columns must be greater than the number $(Q-1)(M+L-1)$ of its rows to obtain non-trivial solutions, i.e., $PL > (Q-1)(M+L-1)$. Solving this inequality for L yields the lower bound for the separating filter length as

$$L_{\mathrm{sep}} > \frac{Q-1}{P-Q+1}(M-1). \tag{4.31}$$

The difference between the number of columns of $\mathbf{H}_{(:\backslash q):,[L]}$ and the number of rows further specifies the dimension of the space of possible non-trivial solutions $\check{\mathbf{W}}_{:q}$, i.e., the number of linearly independent solutions spanning the solution space. Obviously, due to the bound derived above, the best choice we can make to narrow down the solutions is a one-dimensional solution space, i.e., $PL = (Q-1)(M+L-1)+1$. Solving now this *equality* for L and choosing the integer value to be strictly larger than the above bound finally results in the *optimum separating filter length* as

$$L_{\mathrm{opt,sep}} = \frac{(Q-1)(M-1)+1}{P-Q+1}. \tag{4.32}$$

Note that narrowing down the solution space to a one-dimensional space by this choice of filter length precisely means that in this case the *filtering ambiguity of BSS reduces to an arbitrary scaling.* These considerations show that this is possible even for an arbitrary number P of sensors and an arbitrary number Q of sources, where $P \geq Q$. However, the parameters P, Q, M must be such that $L_{\mathrm{opt,sep}}$ is an integer number in order to allow the ideal separation solution. Otherwise, we have to resort to approximations by choosing, e.g., $L_{\mathrm{opt,sep}} = \lceil[(Q-1)(M-1)+1]/(P-Q+1)\rceil$.

To actually obtain the ideal separation solution $\check{\mathbf{W}}_{\mathrm{ideal,sep}}$ with (4.32) for the general, i.e., not necessarily square case $P \geq Q$, we may not straightforwardly apply the block-adjoint and block-determinant operators introduced in the previous subsection. We therefore consider again the original set of homogeneous systems of linear equations (4.21). For the choice $L = L_{\mathrm{opt,sep}}$, we may easily augment the matrix $\mathbf{H}_{(:\backslash q):,[L]}$ to a square matrix $\tilde{\mathbf{H}}_{(:\backslash q):,[L]}$ by adding one row of zeros on both sides of (4.21). The corresponding augmented set of linear systems of equations

$$\tilde{\mathbf{H}}_{(:\backslash q):,[L]}\check{\mathbf{W}}_{:q} = \mathbf{0}, \quad q = 1, \ldots, Q. \tag{4.33}$$

is equivalent to the original set (4.21). However, we may now express the *general solution vector* $\check{\mathbf{W}}_{:q}$ *of (4.21) for the q-th column of* $\check{\mathbf{W}}$ as *the eigenvector corresponding to the zero-valued eigenvalue of the augmented matrix* $\tilde{\mathbf{H}}_{(:\backslash q):,[L]}$.

The general equation (4.32) for the optimum separation filter length is the expression which is analogous to the optimum inverse filter length considered earlier in (4.18). Comparing these two equations, we can verify that in contrast to the inversion, which requires $P > Q$ for the ideal solution using FIR filters, the ideal separation condition is already possible for $P = Q$. Moreover, for the special case $P = Q = 2$, the general expression (4.32) also confirms the choice $L_{\mathrm{opt,BSS}} = M$ as already obtained in Sect. 4.2.1. Figure 4.4 illustrates the different optimum filter lengths by an example.

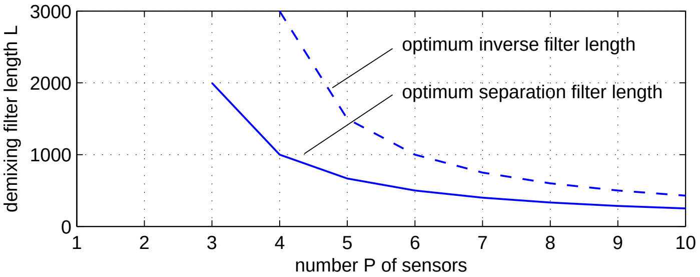


Fig. 4.4. Comparison of the optimum filter lengths for $M = 1000$ and $Q = 3$.

In practice it is obviously difficult to choose the optimum filter length in the blind applications precisely since the length M of the mixing system is generally unknown. Moreover, in many applications we do not require a complete identification of all reflections within the mixing system but only of the dominant ones (e.g., in the localization application considered in later in this chapter). Fortunately, this is in line with the above-mentioned requirement to avoid an overestimation of the filter length in order to narrow down the solution space, i.e., to prevent the filtering ambiguity. Thus, in any case, the choice $L \leq L_{\mathrm{opt,sep}}$ is preferable in practice.

4.2.5 General Scheme for Blind System Identification

In the previous Sections 4.2.1 and 4.2.2 we have explicitly shown the relation between the ideal separation solution and the mixing system for the two-sensor cases. These considerations did also result in a link to the well-known SIMO-based system identification method (note that for BSI with more than two sensors, one simple approach is to apply several of these schemes in parallel), and also to a generalization of this method to the MIMO case with two simultaneously active sources. In the case of more than two sources we may not directly pick the estimated mixing system coefficients from the separation solution $\check{\mathbf{W}}$. The previous Sections 4.2.3 and 4.2.4 generalized the considerations on the two-sensor cases for the *separation* task. In this section, we now outline the generalization of the two-sensor cases in Sections 4.2.1 and 4.2.2 for the *identification* task. The considerations so far suggest the following generic *two-step BSI scheme for an arbitrary number of sources* (where $P \geq Q$):

(1) Based on the available sensor signals, perform a properly designed broadband BSS (see Sect. 4.3) resulting in an estimate of the demixing system matrix.

(2) Analogously to the relation (4.21) between the mixing and demixing systems, and the associated considerations in Sect. 4.2.4 for the separation task, determine an *estimate of the mixing system matrix* using the estimated demixing system from the first step.

In general, to perform step (2) for more than two sources, some further considerations are required. First, an equivalent reformulation of the homogeneous system of equations (4.33) is necessary so that now the *demixing system matrix* instead of the mixing system matrix is formulated *as a blockwise Sylvester matrix.* Note that this corresponds to a *block-transposition* (which we denote here by superscript $\cdot^{\mathrm{bT}}$) of (4.21), i.e.,

$$\left(\mathbf{W}^{\mathrm{bT}}\right)_{(:\backslash q):,[M]}\left(\check{\mathbf{H}}^{\mathrm{bT}}\right)_{:q}=\mathbf{0},\ \ q=1,\ldots,Q. \tag{4.34}$$

The block-transposition is an extension of the conventional matrix transposition. It means that we keep the original form of the channel-wise submatrices but we may change the order of the mixing and demixing subfilters by exploiting the commutativity of the convolutions similarly as in (4.26). Note that the commutativity property does not hold for the MIMO system matrices as a whole, i.e., $\mathbf{W}_{(:\backslash q):,[M]}$ and $\check{\mathbf{H}}_{:q}$, so that they have to be block-transposed to change their order.

Similarly to Sect. 4.2.4, we may then calculate the corresponding estimate of the mixing system in terms of eigenvectors using the complementary form (4.34) of the homogeneous system of equations. Based on this system of equations, we can devise various powerful strategies for BSI in the general MIMO case.

4.2.6 Summary

We have defined and analyzed the signal separation and deconvolution problems using clear conditions for the involved linear mixing and demixing systems. For both problems, the ideal demixing filter coefficients have been derived. Thereby, signal separation was classified as a direct blind adaptive filtering problem, which is in contrast to deconvolution as an inverse adaptive filtering problem. Under certain conditions, such as a suitable filter length for the demixing system and sufficient excitation by the source signals, blind MIMO system identification can be achieved by blind signal separation. In this case, the solutions are unique up to a scaling factor (and a possible permutation of the output channels). From this uniqueness and the correspondence between BSS and BSI we can draw the conclusions that (1) arbitrary filtering may be 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, as shown later in this chapter.

4.3 TRINICON - A General Framework for Adaptive MIMO Signal Processing and Application to the Blind Adaptation Problems

For the blind estimation of the coefficients corresponding to the desired solutions discussed in the previous section, we have to consider and to exploit the properties of the excitation signals, such as their non-stationarity, their spectral characteristics, and their probability densities.

In the existing literature, the BSS problem has mostly been addressed for instantaneous mixtures or narrowband approaches in the frequency domain which adapt the coefficients independently in each DFT bin, e.g., [19,33,34]. On the other hand, in the case of MCBD, many approaches either aim at whitening the output signals as they are based on an i.i.d. model of the source signals (e.g., [27,28]), which is undesirable for speech and audio signals which should not be whitened, or are rather heuristically motivated, e.g., [29].

The aim of this section is to present an overview of the algorithmic part of broadband blind adaptive MIMO filtering based on TRINICON ('TRIple-N Independent component analysis for CONvolutive mixtures'), a generic concept for adaptive MIMO filtering which takes all the above mentioned signal properties (nonwhiteness, nonstationarity, and nongaussianity) into account, and allows a unified treatment of broadband BSS as needed for a proper BSI, and MCBD algorithms applicable to speech and audio signals in real acoustic environments [1–4]. This framework generally uses multivariate stochastic signal models in the cost function to describe the temporal structure of the source signals. This versatile approach provides a powerful cost function for both, BSS/BSI and MCBD, and, for the latter, also leads to improved solutions for speech dereverberation.

As in the previous sections, we will again mainly focus on the direct blind adaptive filtering problems, such as BSS. In [4], a direct relation between the BSS adaptation mechanism and the ideal separation solution (4.20) was established. Moreover, although both time-domain and equivalent broadband frequency-domain formulations of TRINICON have been developed with the corresponding multivariate models in both the time domain and the frequency domain [2,4], we consider in this chapter mainly the time-domain formulation. We discuss here only gradient-based coefficient updates for clarity of presentation. The algorithmic TRINICON framework is directly based on the matrix notation developed above.

Throughout this section, we regard the standard BSS model where the number Q of *maximum simultaneously active source signals* $s_q(n)$ is equal to the number of sensor signals $x_p(n)$, i.e., $Q = P$. However, it should be noted that in contrast to other BSS algorithms we do not assume prior knowledge about the exact number of active sources. Thus, even if the algorithms will be derived for $Q = P$, the number of simultaneously active sources may change throughout the application of the BSS algorithm and only the condition $Q \leq P$ has to be fulfilled.

4.3.1 Cost Function and Gradient-Based Coefficient Optimization

Matrix notation for convolutive mixtures. To introduce an algorithm for broadband processing of convolutive mixtures, we first need to formulate the convolution of the FIR demixing system of length L in the following matrix form [4]:

$$\mathbf{y}^{\mathrm{T}}(n) = \mathbf{x}^{\mathrm{T}}(n)\mathbf{W}, \tag{4.35}$$

where n denotes the time index, and

$$\mathbf{x}^{\mathrm{T}}(n) = [\mathbf{x}_1^{\mathrm{T}}(n), \ldots, \mathbf{x}_P^{\mathrm{T}}(n)], \tag{4.36}$$

$$\mathbf{y}^{\mathrm{T}}(n) = [\mathbf{y}_1^{\mathrm{T}}(n), \ldots, \mathbf{y}_P^{\mathrm{T}}(n)], \tag{4.37}$$

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_{11} & \cdots & \mathbf{W}_{1P} \\ \vdots & \ddots & \vdots \\ \mathbf{W}_{P1} & \cdots & \mathbf{W}_{PP} \end{bmatrix}, \tag{4.38}$$

$$\mathbf{x}_p^{\mathrm{T}}(n) = [x_p(n), \ldots, x_p(n - 2L + 1)], \tag{4.39}$$

$$\mathbf{y}_q^{\mathrm{T}}(n) = [y_q(n), \ldots, y_q(n - D + 1)] \tag{4.40}$$

$$= \sum_{p=1}^{P} \mathbf{x}_p^{\mathrm{T}}(n)\mathbf{W}_{pq}. \tag{4.41}$$

The parameter D in (4.40), $1 \leq D < L$, denotes the number of lags taken into account to exploit the nonwhiteness of the source signals as shown below. $\mathbf{W}_{pq}$, $p = 1, \ldots, P$, $q = 1, \ldots, P$ denote $2L \times D$ Sylvester matrices that contain all coefficients of the respective filters:

$$\mathbf{W}_{pq} = \begin{bmatrix} w_{pq,0} & 0 & \cdots & 0 \\ w_{pq,1} & w_{pq,0} & \ddots & \vdots \\ \vdots & w_{pq,1} & \ddots & 0 \\ w_{pq,L-1} & \vdots & \ddots & w_{pq,0} \\ 0 & w_{pq,L-1} & \ddots & w_{pq,1} \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & w_{pq,L-1} \\ 0 & \cdots & 0 & 0 \\ \vdots & & \vdots & \vdots \\ 0 & \cdots & 0 & 0 \end{bmatrix}. \tag{4.42}$$

Note that for $D = 1$, (4.35) simplifies to the well-known vector formulation of a convolution, as it is used extensively in the literature on supervised adaptive filtering, e.g., [18].

Optimization Criterion. Various approaches exist to blindly estimate the demixing matrix $\mathbf{W}$ for the above-mentioned tasks by utilizing the following source signal properties [19] which we all combine into an efficient and versatile algorithmic framework [1–3]:
(i) Nongaussianity is exploited by using higher-order statistics for independent component analysis (ICA). ICA approaches can be divided into several classes. Although they all lead to similar update rules, the minimization of the mutual information (MMI) among the output channels can be regarded as the most general approach for BSS [19]. To obtain an estimator not only allowing spatial separation but also temporal separation for MCBD, we use the Kullback-Leibler divergence (KLD) [35] between a certain *desired* joint pdf (essentially representing a hypothesized stochastic source model) and the joint pdf of the actually estimated output signals. The desired pdf is factorized w.r.t. the different sources (for BSS) and possibly also w.r.t. certain temporal dependencies (for MCBD) as shown below. The KLD is guaranteed to be positive [35], which is a necessary condition for a useful cost function.
(ii) Nonwhiteness is exploited by simultaneous minimization of output cross-relations over multiple time-lags. We therefore consider 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.

Based on the KLD, we now define the following general cost function taking into account all three fundamental signal properties (i)-(iii):

$$\mathcal{J}(m, \mathbf{W}) = -\sum_{i=0}^{\infty} \beta(i,m) \frac{1}{N} \sum_{j=iL}^{iL+N-1} \left\{ \log(\hat{p}_{s,PD}(\mathbf{y}(j))) - \log(\hat{p}_{y,PD}(\mathbf{y}(j))) \right\} , \tag{4.43}$$

where $\hat{p}_{s,PD}(\cdot)$ and $\hat{p}_{y,PD}(\cdot)$ are the assumed or estimated PD-variate source model (i.e., desired) pdf and output pdf, respectively. The index m denotes the block time index for a block of N output samples shifted by L samples relatively to the previous block. Furthermore, D is the memory length, i.e., the number of time-lags to model the nonwhiteness of the P signals as above. β is a window function with finite support that is normalized so that $\sum_{i=0}^{m} \beta(i,m) = 1$, allowing for online, offline, and block-online algorithms [2,36].

Gradient-Based Coefficient Update. In this chapter we concentrate on iterative gradient-based block-online coefficient updates which can be written in the general form

$$\check{\mathbf{W}}^0(m) := \check{\mathbf{W}}(m-1), \tag{4.44a}$$

$$\check{\mathbf{W}}^\ell(m) = \check{\mathbf{W}}^{\ell-1}(m) - \mu \Delta\check{\mathbf{W}}^\ell(m), \ \ \ell = 1, \dots, \ell_{\max}, \tag{4.44b}$$

$$\check{\mathbf{W}}(m) := \check{\mathbf{W}}^{\ell_{\max}}(m), \tag{4.44c}$$

where μ is a stepsize parameter, and the superscript index ℓ denotes an iteration parameter to allow for multiple iterations ($\ell = 1, \dots, \ell_{\max}$) within each block m. The $LP \times P$ coefficient matrix $\check{\mathbf{W}}$ (defined in (4.4)) to be optimized is smaller than the $2LP \times DP$ Sylvester matrix $\mathbf{W}$ used above for the formulation of the cost function, and it contains only the non-redundant elements of $\mathbf{W}$.

The simplest case of the above procedure (4.44a)-(4.44c) is the gradient descent update, which is defined by

$$\Delta\check{\mathbf{W}}^\ell(m) = \nabla_{\check{\mathbf{W}}} \mathcal{J}(m, \mathbf{W})|_{\check{\mathbf{W}}=\check{\mathbf{W}}^\ell(m)} . \tag{4.45}$$

Obviously, when calculating this gradient explicitly, we are confronted with the problem of the different coefficient matrix formulations $\mathbf{W}$ and $\check{\mathbf{W}}$ in the cost function and in the optimization procedure, respectively. This is a direct consequence of taking into account the nonwhiteness signal property, as mentioned above, and – although it may seem less obvious at this point – it leads to an important building block whose actual implementation is fundamental to the properties of the resulting algorithm, the so-called *Sylvester constraint* ($\mathcal{SC}$) on the coefficient update [2,4]. Using the Sylvester constraint operator the gradient descent update (4.45) can be rewritten as

$$\Delta\check{\mathbf{W}}^\ell(m) = \mathcal{SC}\left\{\nabla_{\mathbf{W}} \mathcal{J}(m, \mathbf{W})\right\}|_{\mathbf{W}=\mathbf{W}^\ell(m)} . \tag{4.46}$$

Depending on the particular realization of ($\mathcal{SC}$), we are able to select both, well known and also novel improved adaptation algorithms [36]. As discussed in [36] there are two particularly simple and popular realizations of ($\mathcal{SC}$) leading to two different classes of algorithms:

(1) Computing only the *first column* of each channel of the update matrix to obtain the new coefficient matrix $\check{\mathbf{W}}$. This method is denoted as ($\mathcal{SC}_\mathrm{C}$).
(2) Computing only the *L-th row* of each channel of the update matrix to obtain the new coefficient matrix $\check{\mathbf{W}}$. This method is denoted as ($\mathcal{SC}_\mathrm{R}$).

It can be shown that in both cases the update process is significantly simplified [36]. However, in general, both choices require some tradeoff in the algorithm performance. While $\mathcal{SC}_\mathrm{C}$ may provide a potentially more robust convergence behaviour, it will not work for arbitrary source positions (e.g., in the case of two sources, they are required to be located in different half-planes w.r.t. the orientation of the microphone array), which is in contrast to the more versatile $\mathcal{SC}_\mathrm{R}$ [36]. Note that the choice of $\mathcal{SC}$ also determines the appropriate coefficient initialization [36].

Next, in this chapter, we derive a novel *generic* Sylvester constraint to further formalize and clarify this concept.

Let $W_{kj}^{KJ} = [\mathbf{W}]_{kj}^{KJ}$ denote the kj-th component of the *Sylvester matrix* after (4.42) for the KJ-th channel corresponding to the KJ-th submatrix in (4.38). According to [2,4], the gradient of $\mathcal{J}$ w.r.t. these components is transformed by a certain choice of $(\mathcal{SC})$ to the gradient w.r.t. the components $\check{W}_m^{MN} = \left[\check{\mathbf{W}}\right]_m^{MN}$ of the *condensed matrix* as used above in (4.45). This can be expressed concisely by applying the chain rule for matrix derivatives in the following form:

$$\begin{aligned}
\frac{\partial \mathcal{J}}{\partial \check{W}_m^{MN}} &= \sum_{k,j,K,J} \frac{\partial \mathcal{J}}{\partial W_{kj}^{KJ}} \frac{\partial W_{kj}^{KJ}}{\partial \check{W}_m^{MN}} \\
&= \sum_{k,j,K,J} \frac{\partial \mathcal{J}}{\partial W_{kj}^{KJ}} \delta_{KM} \delta_{JN} \delta_{k,(m+j-1)} \\
&= \sum_{k,j} \frac{\partial \mathcal{J}}{\partial W_{kj}^{MN}} \delta_{k,(m+j-1)} ,
\end{aligned} \tag{4.47}$$

where

$$\delta_{ij} = \begin{cases} 1 & \text{for} \quad i = j \\ 0 & \text{for} \quad i \neq j \end{cases} \tag{4.48}$$

denotes the Kronecker symbol. Hence, we have the simple linear relation

$$[\nabla_{\check{\mathbf{W}}} \mathcal{J}]_m^{MN} = \sum_{k,j} [\nabla_{\mathbf{W}} \mathcal{J}]_{kj}^{MN} \, \delta_{k,(m+j-1)} \tag{4.49}$$

between the MN-th submatrices of $\nabla_{\mathbf{W}} \mathcal{J}$ and $\nabla_{\check{\mathbf{W}}} \mathcal{J}$. If we consider now for illustration of (4.49) the individual elements of $\nabla_{\check{\mathbf{W}}} \mathcal{J}$ for one channel in more detail, i.e.,

$$\begin{aligned}
[\nabla_{\check{\mathbf{W}}} \mathcal{J}]_1^{MN} &= \sum_{j} [\nabla_{\mathbf{W}} \mathcal{J}]_{jj}^{MN} \\
[\nabla_{\check{\mathbf{W}}} \mathcal{J}]_2^{MN} &= \sum_{k,j} [\nabla_{\mathbf{W}} \mathcal{J}]_{kj}^{MN} \, \delta_{k,(j+1)} = \sum_{j} [\nabla_{\mathbf{W}} \mathcal{J}]_{j+1,j}^{MN} \\
&\vdots \\
[\nabla_{\check{\mathbf{W}}} \mathcal{J}]_L^{MN} &= \sum_{k,j} [\nabla_{\mathbf{W}} \mathcal{J}]_{kj}^{MN} \, \delta_{k,(j+L-1)} = \sum_{j} [\nabla_{\mathbf{W}} \mathcal{J}]_{j+L-1,j}^{MN} ,
\end{aligned}$$

we can readily see that the generic Sylvester constraint corresponds – up to the constant D denoting the width of the submatrices – to a *channel-wise arithmetic averaging* of elements according to Fig. 4.5.

Note that the previously introduced approaches, classified by the choice $(\mathcal{SC}_{\mathrm{C}})$ or $(\mathcal{SC}_{\mathrm{R}})$ as mentioned above, thus correspond to certain approximations by neglecting some of the elements within this averaging process, as illustrated in Fig. 4.6.

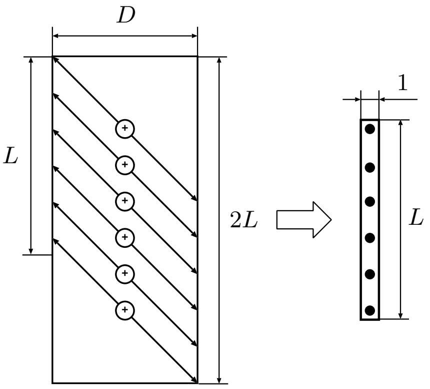
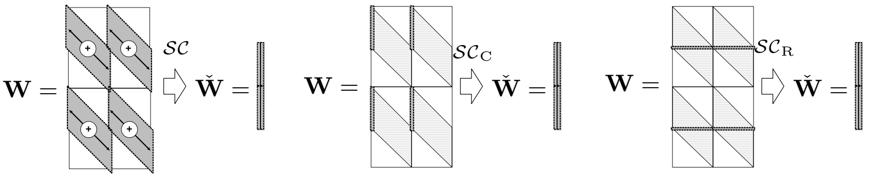


Fig. 4.5. Illustration of the generic Sylvester constraint ($\mathcal{SC}$) for one channel, i.e., the MN-th submatrix $[\nabla_{\check{\mathbf{W}}}\mathcal{J}]^{MN}$ of $\nabla_{\check{\mathbf{W}}}\mathcal{J}$.



(a) (b) (c)

Fig. 4.6. Illustration of two efficient approximations of (a) the generic Sylvester constraint $\mathcal{SC}$: (b) the column Sylvester constraint $\mathcal{SC}_\mathrm{C}$ and (c) the row Sylvester constraint $\mathcal{SC}_\mathrm{R}$.

Natural Gradient-Based Coefficient Update. It can be shown (after a somewhat tedious but straightforward derivation) that by taking the *natural gradient* [19] of $\mathcal{J}(m)$ with respect to the demixing filter matrix $\mathbf{W}(m)$ [4],

$$\Delta\check{\mathbf{W}} \propto \mathcal{SC}\left\{\mathbf{W}\mathbf{W}^\mathrm{T}\frac{\partial\mathcal{J}}{\partial\mathbf{W}}\right\}, \tag{4.50}$$

we obtain the following generic TRINICON-based update rule:

$$\check{\mathbf{W}}(m) = \check{\mathbf{W}}(m-1) - \mu\Delta\check{\mathbf{W}}(m), \tag{4.51a}$$

$$\begin{aligned}\Delta\check{\mathbf{W}}(m) = \frac{1}{N}\sum_{i=0}^{\infty}\beta(i,m)\,\mathcal{SC}\Bigg\{&\sum_{j=iL}^{iL+N-1}\mathbf{W}(i)\mathbf{y}(j)\\ &\cdot\left[\boldsymbol{\Phi}_{s,PD}^\mathrm{T}(\mathbf{y}(j)) - \boldsymbol{\Phi}_{y,PD}^\mathrm{T}(\mathbf{y}(j))\right]\Bigg\},\end{aligned} \tag{4.51b}$$

with the *desired* score function

$$\boldsymbol{\Phi}_{s,PD}(\mathbf{y}(j)) = -\frac{\partial\log\hat{p}_{s,PD}(\mathbf{y}(j))}{\partial\mathbf{y}(j)} \tag{4.51c}$$

resulting from the hypothesized source model, and the actual score function

$$\boldsymbol{\Phi}_{y,PD}(\mathbf{y}(j)) = -\frac{\partial \log \hat{p}_{y,PD}(\mathbf{y}(j))}{\partial \mathbf{y}(j)}. \tag{4.51d}$$

The hypothesized source model $\hat{p}_{s,PD}(\cdot)$ in (4.51c) is chosen according to the class of signal processing problem to be solved. For instance, a factorization of $\hat{p}_{s,PD}(\cdot)$ among the sources yields BSS, i.e.,

$$\hat{p}_{s,PD}(\mathbf{y}(j)) \overset{\text{(BSS)}}{=} \prod_{q=1}^{P} \hat{p}_{y_q,D}(\mathbf{y}_q(j)), \tag{4.52}$$

while a complete factorization leads to the traditional MCBD approach,

$$\hat{p}_{s,PD}(\mathbf{y}(j)) \overset{\text{(MCBD)}}{=} \prod_{q=1}^{P} \prod_{d=1}^{D} \hat{p}_{y_q,1}(\mathbf{y}_q(j-d)). \tag{4.53}$$

4.3.2 Special Cases and Illustration in the Time Domain

Besides the various options to design the Sylvester constraint, there are many further interesting known and novel practical approximations within the framework. To begin with, we first consider algorithms based on second-order statistics (SOS) as they are particularly illustrative.

Realizations based on Second-Order Statistics. Here, the source models are simplified to sequences of multivariate Gaussian functions described by $PD \times PD$ correlation matrices $\mathbf{R}_{\cdot\cdot}$ within the length-N signal blocks. This leads to the coefficient update [3]

$$\begin{aligned} \Delta\check{\mathbf{W}}(m) &= \sum_{i=0}^{\infty} \beta(i,m)\, \mathcal{SC}\left\{\mathbf{W}(i)\hat{\mathbf{R}}_{\mathbf{yy}}\left[\hat{\mathbf{R}}_{\mathbf{ss}}^{-1} - \hat{\mathbf{R}}_{\mathbf{yy}}^{-1}\right]\right\} \\ &= \sum_{i=0}^{\infty} \beta(i,m)\, \mathcal{SC}\left\{\mathbf{W}(i)\left[\hat{\mathbf{R}}_{\mathbf{yy}} - \hat{\mathbf{R}}_{\mathbf{ss}}\right]\hat{\mathbf{R}}_{\mathbf{ss}}^{-1}\right\}. \end{aligned} \tag{4.54}$$

Generic SOS-based BSS. The BSS variant of the generic SOS natural gradient update (4.54) follows immediately by setting

$$\hat{\mathbf{R}}_{\mathbf{ss}}(i) = \operatorname{bdiag} \hat{\mathbf{R}}_{\mathbf{yy}}(i). \tag{4.55}$$

The update (4.54) together with (4.55) was originally obtained independently in [4] as a generalization of the cost function of [37]:

$$\mathcal{J}_{\mathrm{SOS}}(m) = \sum_{i=0}^{\infty} \beta(i,m)\left\{\log\det \hat{\mathbf{R}}_{\mathbf{ss}}(i) - \log\det \hat{\mathbf{R}}_{\mathbf{yy}}(i)\right\}. \tag{4.56}$$

In Fig. 4.7 the mechanism of (4.54) based on the model (4.55) is illustrated. By minimizing $\mathcal{J}_{\mathrm{SOS}}(m)$, all cross-correlations for D time-lags are reduced and will ideally vanish, while the auto-correlations are untouched to preserve the structure of the individual signals. This class of algorithms leads

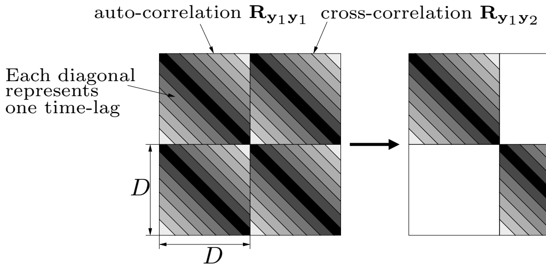

auto-correlation $\mathbf{R}_{\mathbf{y}_1\mathbf{y}_1}$
cross-correlation $\mathbf{R}_{\mathbf{y}_1\mathbf{y}_2}$
Each diagonal represents one time-lag
D
D

Fig. 4.7. Illustration of SOS-based BSS.

to very robust practical solutions even for a large number of filter taps due to an inherent normalization by the auto-correlation matrices, reflected by the inverse in (4.54) of bdiag $\hat{\mathbf{R}}_{\mathbf{yy}}$. Note that there are also various efficient approximations of this broadband algorithm, e.g, [36,38,39], with a reduced computational complexity allowing already real-time operation on a regular PC platform. These efficient implementations also form a powerful basis for blind system identification and for simultaneous localization of multiple acoustic sources, as shown later in this chapter. Moreover, a close link has been established [2,4] to various popular frequency-domain algorithms, as we discuss in more detail in Sect. 4.3.3.

Inverse blind adaptive filtering problems. To illustrate that TRINICON also offers a powerful framework for *inverse* blind adaptive filtering problems in the same way as it does for the *direct* blind adaptive filtering problems, we give a brief overview of some of the most important ideas in this context in the following two paragraphs, based on [3].

MCBD based on SOS. Traditionally, ICA-based MCBD algorithms assume i.i.d. source models, e.g., [27,28]. In the SOS case, this corresponds to a complete whitening of the output signals by not only applying a joint de-cross-correlation, but also a joint de-auto-correlation, i.e., $\hat{\mathbf{R}}_{\mathbf{ss}} = \operatorname{diag} \hat{\mathbf{R}}_{\mathbf{yy}}$ over multiple time-instants, as illustrated in Fig. 4.9 (b).

MCBPD based on SOS. Signal sources which are non i.i.d. should not become i.i.d. at the output of the blind adaptive filtering stage. Therefore, their statistical dependencies should be preserved. In other words, the adaptation algorithm has to distinguish between the statistical dependencies within the source signals, and the statistical dependencies introduced by the mixing system $\check{\mathbf{H}}$. We denote the corresponding generalization of the traditional MCBD technique as *MultiChannel Blind Partial Deconvolution* (MCBPD) [3]. Equations (4.51b)-(4.51d) inherently contain a statistical source model (signal properties (i)-(iii) in Sect. 4.3.1), expressed by the multivariate densities, and thus provide all necessary requirements for the MCBPD approach.

A typical example for MCBPD applications is speech dereverberation, which is especially important for distant-talking automatic speech recognition (ASR), as there is a very strong need for speech dereverberation without introducing artifacts to the signals. In this application, MCBPD allows to distinguish between the actual speech production system, i.e., the vocal tract, and the reverberant room (Fig. 4.8). Ideally, only the influence of the room acoustics should be minimized. In the *SOS case*, the auto-correlation structure of the speech signals can be taken into account, as shown in Fig. 4.9 (c). While the room acoustics influences all off-diagonals, the effect of the vocal tract is concentrated in the first few off-diagonals around the main diagonal. These first off-diagonals of $\hat{\mathbf{R}}_{\mathbf{yy}}$ are now taken over into $\hat{\mathbf{R}}_{\mathbf{ss}}$, as shown in Fig. 4.9 (c). Alternatively, the structure in Fig. 4.9 (c) may be approximated by small sub-matrices making its handling somewhat more efficient. Note that there is a close link to linear prediction techniques which gives guidelines for the number of lags to be preserved.

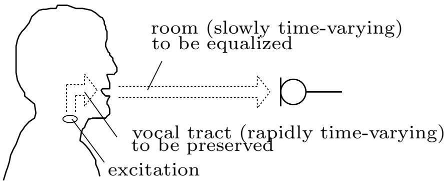
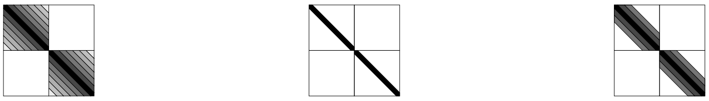

room (slowly time-varying)
to be equalized
vocal tract (rapidly time-varying)
to be preserved
excitation

Fig. 4.8. Illustration of speech dereverberation as an MCBPD application.



(a) BSS (b) MCBD (c) MCBPD

Fig. 4.9. Desired correlation matrices $\hat{\mathbf{R}}_{\mathbf{ss}}$ for BSS, MCBD, and MCBPD with TRINICON in the SOS case.

Realizations based on Higher-Order Statistics. The general HOS approach (4.51b)-(4.51d) provides the possibility to take into account all available information on the statistical properties of the desired source signals. This provides an increased flexibility and improved performance of BSS relative to the SOS case. Moreover, the more accurate modeling of the desired source signals yields also an improved MCBPD.

To apply the general approach in a real-world scenario, appropriate multivariate score functions (4.51c) and (4.51d) have to be determined. Fortunately, there is an efficient solution to this problem by assuming so-called spherically invariant random processes (SIRPs) [40–42]. The general form of correlated SIRPs of D-th order is given with a properly chosen function $f_D(\cdot)$ by

$$\hat{p}_D(\mathbf{y}_p(j)) = \frac{1}{\sqrt{\pi^D \det(\mathbf{R}_{\mathbf{y}_p\mathbf{y}_p}(i))}} f_D\left(\mathbf{y}_p^{\mathrm{T}}(j)\mathbf{R}_{\mathbf{y}_p\mathbf{y}_p}^{-1}(i)\mathbf{y}_p(j)\right) \tag{4.57}$$

for the p-th channel, where $\mathbf{R}_{\mathbf{y}_p\mathbf{y}_p}$ denotes the corresponding auto-correlation matrix with the corresponding number of lags. These models are representative for a wide class of stochastic processes. Speech signals in particular can very accurately be represented by SIRPs [42]. A great advantage arising from the SIRP model is that multivariate pdfs can be derived analytically from the corresponding univariate pdf together with the (lagged) correlation matrices. The function $f_D(\cdot)$ can thus be calculated from the well-known univariate models for speech, e.g., the Laplacian density. Using the chain rule, the corresponding score function (4.51c) can be derived from (4.57), as shown in [1,2] in more detail.

The calculation of the other score function (4.51d) becomes particularly simple in most practical realizations by transforming the output pdf $\hat{p}_{y,PD}(\cdot)$ into the corresponding multivariate input signal pdf using $\mathbf{W}$, which is considered as a mapping matrix of a linear transformation (see [1,2] for the general broadband case where $\mathbf{W}$ exhibits a blockwise-Sylvester structure). The derivative of the input signal pdf vanishes as it is independent of the demixing system.

Note that the multivariate Gaussian pdf is a special case of a SIRP and thus, the above described SOS-based algorithms represent special cases of the corresponding algorithms based on SIRPs [1,2]. As in the SOS case, by transforming the model into the DFT domain, various links to novel and existing popular frequency-domain algorithms can be established [2], as we discuss in more detail in Sect. 4.3.3.

4.3.3 On Frequency-Domain Realizations

For convolutive mixtures, the classical approach of frequency-domain BSS appears to be an attractive alternative where all techniques originally developed for instantaneous BSS are typically applied independently in each frequency bin, e.g., [19]. Unfortunately, this traditional narrowband approach exhibits several limitations as identified in, e.g., [43–45]. In particular, the permutation problem, which is inherent in BSS, may then also appear independently in each frequency bin so that extra repair measures have to be taken to address this *internal* permutation. Problems caused by circular convolution effects due to the narrowband approximation are reported in, e.g., [44].

In [2] it is shown how the equations of the TRINICON framework can be transformed into the frequency domain in a rigorous way (i.e., without any approximations) in order to avoid the above-mentioned problems. As in the case of the time-domain algorithms, the resulting generic DFT-domain BSS may serve both as a unifying framework for existing algorithms, and also as a guideline for developing new improved algorithms by certain suitable *selective* approximations as shown in, e.g., [2] or [38]. Figure 4.10 gives an overview on the most important classes of DFT-domain BSS algorithms known so far (various more special cases may be developed in the future). A very important observation from this framework using multivariate pdfs is that, in general, all frequency components are linked together so that the internal permutation problem is avoided (the following elements are reflected in Fig. 4.10 by the different approximations of the generic SIRP-based BSS):

1. Constraint matrices appearing in the generic frequency-domain formulation (see, e.g., [2]) describe the inter-frequency correlation between DFT components.
2. The multivariate score function, derived from the multivariate pdf is a broadband score function. As an example, for SIRPs the argument of the multivariate score function (which is a nonlinear function in the higher-order case) is $\mathbf{y}_p^{\mathrm{T}}(j)\mathbf{R}_{\mathbf{y}_p\mathbf{y}_p}^{-1}(i)\mathbf{y}_p(j)$ according to (4.57). Even for the simple case $\mathbf{R}_{\mathbf{y}_p\mathbf{y}_p}^{-1}(i) = \mathbf{I}$ where we have $\mathbf{y}_p^{\mathrm{T}}(j)\mathbf{y}_p(j) = \|\mathbf{y}_p(j)\|^2$, i.e., the quadratic norm, and - due to the Parseval theorem - the same in the frequency domain, i.e., the quadratic norm over all DFT components, we immediately see that all DFT-bins are taken into account simultaneously so that the internal permutation problem is avoided. Note that the traditional narrowband approach (with the internal permutation problem) would result as a special case if we assumed all DFT components to be statistically independent from each other (which is of course not the case for real-world broadband signals such as speech and audio signals). In contrast to this independence approximation the dependencies among all frequency components (including higher-order dependencies) are inherently taken into account in TRINICON in an optimal way. Actually, in the traditional narrowband approach, the additionally required repair mechanisms for permutation alignment try to exploit such inter-frequency dependencies.

From the viewpoint of the blind system identification the *broadband algorithms with constraint matrices* (i.e., the algorithms represented in the first column of Fig. 4.10) are of particular interest. Among these algorihms, the system described in [38] has turned out to be very efficient in this context and

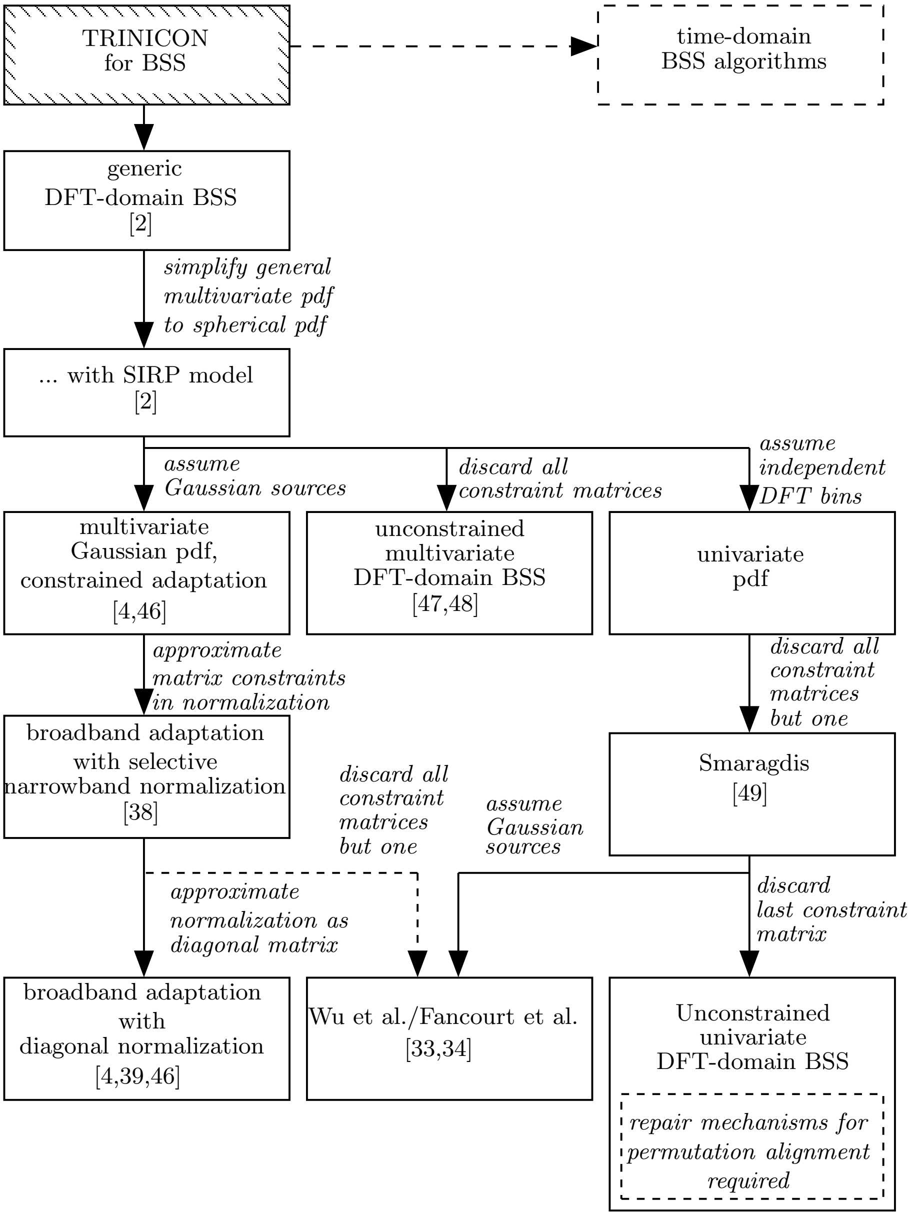


Fig. 4.10. Overview of BSS algorithms in the DFT domain.

for multiple source localization as described later in this chapter. A pseudocode of this algorithm is also included in [38].

Another important consideration for the practical implementation of BSI is the proper choice of the Sylvester constraint. Since the *column constraint* $\mathcal{SC}_\mathrm{C}$ is not suited for arbitrary source positions, it is generally *not appropriate* for BSI and the source localization application discussed next in this chapter. Thus, in the implementations discussed below the *row constraint* $\mathcal{SC}_\mathrm{R}$ is used.

4.4 Acoustic Source Localization: An Overview of the Approaches and TRINICON as a Generic Source Localization Scheme

The precision of acoustic source localization is determined by several factors. Critical parameters of the acoustic environment itself are the Signal-to-Noise Ratio (SNR, additive distortion) and the reverberation time (T_{60}) or the power ratio between the direct path and the reverberation of the room (Signal-to-Reverberation Ratio, SRR). Further important conditions are determined by the sources (number of sources, spatial diversity, velocity of the motion, signal statistics) and the sensors (number of sensors, array geometry, temporal sampling rate).

In the literature (e.g., [9]) existing approaches for acoustic source localization are often roughly divided into three categories:

(1) Maximization of the output power of steered beamformers (SRP, *steered response power*)
(2) Approaches based on high-resolution spectral estimation (also called *subspace* approaches)
(3) Approaches based on the estimation of *time differences of arrival* (TDOA) as an intermediate step

All of these methods may essentially be seen as two-step methods consisting of a certain signal processing stage, based on the available sensor signals, and a certain mapping from the signal processing output to the geometrical source position(s) by taking into account the sensor array geometry.

The first category of source localization approaches is based on the variation of the spatial alignment of a beamformer and results in systematic scanning of the acoustic environment so that basically this method may provide a very accurate localization even for multiple sources. Note that in this method, both of the above-mentioned two steps, i.e., the signal processing, and the geometric mapping step are included in the scanning process. In principle, the number of microphones is easily scalable in this concept. It also allows a relatively high robustness to additive interferences [9]. However, known disadvantages of this technique are that due to the (fixed) beamformer design one normally has to assume an ideal freefield propagation of the acoustic signals, and the search process necessarily becomes computationally very demanding if the desired spatial resolution increases.

The second category includes a class of algorithms which can be considered as an advancement and a systematization of the first category. The corresponding approach is based on the $P \times P$ correlation matrix of the P microphone signals (without time lags) and it allows also inherently a simultaneous localization of multiple active sources. If the number Q of sources is less or equal to the number P of sensors then it can be shown [50] that the eigenvectors for the Q largest eigenvalues of the correlation matrix span a certain subspace ('signal subspace'). This subspace corresponds to the one resulting from the direction vectors of the sources. From these direction vectors, we can extract the corresponding directions of arrival (DOA) in a separate mapping step. The remaining $P - Q$ eigenvectors, i.e., the eigenvectors corresponding to the $P - Q$ smallest eigenvalues of the correlation matrix, constitute the subspace of the background noise which is orthogonal to the signal subspace. This concept forms the basis for several well-known algorithms proposed in the literature, e.g., MUSIC [10] and ESPRIT [11]. Unfortunately, however, these algorithms were originally developed only for narrowband signals and therefore they are not immediately applicable to acoustic broadband signals, such as speech. One of the problems is that, in general, each frequency component yields a different signal subspace. Moreover, just as above in category (1), the room reverberation is not modeled by this method. Therefore, numerous modifications of these algorithms have been proposed in the literature. In order to solve the problem due to the narrowband assumption, [51] proposes an introduction of a *focussing* to a certain center frequency so that only a single signal subspace is obtained which ideally contains the complete information on the source positions. Unfortunately, in practice this is often problematic due to robustness issues and due to the necessity of a good initial guess which is difficult to obtain. Therefore, so far this approach has not been widely used for audio signals [9]. In addition, if there is multipath propagation due to spatial reflections, then these reflected signal components act on the microphones as additional correlated sources. In order to solve the narrowband problem, and thus the focussing problem in an optimal way, a new approach has been proposed in [52] which takes into account the underlying physics of wave propagation. There, the sound field for freefield propagation is decomposed into eigenfunctions. Thereby even the scattering on the microphone array itself can be efficiently taken into account, and by using a circular or spherical array, a full 360 degrees field-of-view is possible.

Finally, according to the above discussions in this section and in Sect. 4.2, broadband BSS may be also regarded as a generalized subspace approach based on the block-diagonalization of the signal correlation matrix in the case of second-order statistics. Due to the systematic incorporation of time lags into the correlation matrix in contrast to the instantaneous correlation matrix used in the conventional subspace methods, we are now able to take into account the room reverberation. Thus, as illustrated in Fig. 4.11, the blind broadband adaptive MIMO filtering approach generalizes and unifies both the traditional subspace methods, and the SIMO-based BSI. Note that both of these traditional methods are based on the calculation of the eigenvector(s) corresponding to the smallest eigenvalue(s) of the (lagged or instantaneous) sensor correlation matrix (see also the note at the end of Sect. 4.2.2 for the case of SIMO-based BSI).

Category (3) is by far the most widely used. As in the previous category, we split here the determination of the source position using multiple microphone signals into two separate steps. In contrast to the first two categories, the first step is here the *explicit* estimation of the temporal signal delays between different pairs of microphones (*time difference of arrival*, TDOA). The second step constitutes the calculation of the position in the three-dimensional space or in the two-dimensional plane using these estimates. Under the assumption that the relative microphone positions are known a-priori, the problem of source localization from a given set of TDOAs can be reduced to a purely geometrical problem.

For the explicit determination of TDOAs many different techniques have been proposed in the literature [9,16]. Of particular interest is the fact that there are some more recent and powerful TDOA estimation techniques based directly on the blind system identification methods presented earlier in this chapter. Thus, since these techniques inherently take the room reflections into account, they promise a high robustness even in real reverberant environments.

Another advantage of the TDOA-based method is that it also allows for an accurate localization of sources in the *nearfield*.

In summary, the TDOA-based method using a blind system identification technique may be considered as the most general and versatile source localization approach. However, most of the BSI techniques known so far from the literature are based on SIMO systems, as shown in Sect. 4.2.2, i.e., the localization systems based on these techniques are only suitable for one source. The MIMO BSI technique described earlier in this chapter thus also generalizes this type of TDOA-based method to allow a *simultaneous localization of multiple sources in reverberant environments* even in the nearfield. In the following we therefore consider this two-step TDOA-based approach in more detail. To begin with, the geometrical considerations in Section 4.5.1 concentrate on the second step for the actual localization, particularly on the necessary number of different TDOA measurements and the array geometry. In contrast, for the first step, i.e., the TDOA estimation, the acoustic conditions of the room play a very important role. For that, we also consider the popular *generalized cross-correlation* (GCC) method as a reference in Sect. 4.5.3.

Figure 4.11 summarizes the above considerations and illustrates that the TRINICON-based MIMO BSI scheme may be considered as a *generic source localization approach*. Note also that in principle, broadband BSS can be applied to all three of the above-mentioned categories of acoustic source lo-

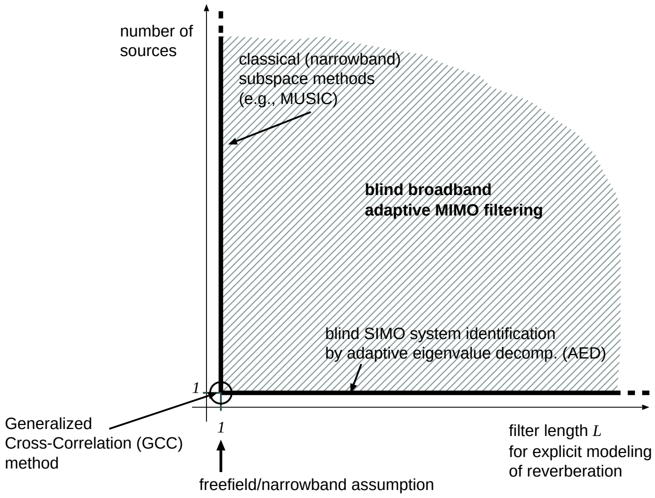

number of sources
classical (narrowband) subspace methods (e.g., MUSIC)
blind broadband adaptive MIMO filtering
blind SIMO system identification by adaptive eigenvalue decomp. (AED)
1
1
Generalized Cross-Correlation (GCC) method
freefield/narrowband assumption
filter length L for explicit modeling of reverberation

Fig. 4.11. Relations between various localization approaches.

calization approaches since broadband BSS may be considered as (1) multiple blind beamformers, (2) a generalized subspace approach based on the block-diagonalization of the signal correlation matrix, and (3) a method for MIMO-based BSI, as discussed above.

4.5 Acoustic Source Localization based on Time-Differences of Arrival

4.5.1 Basic Geometric Considerations

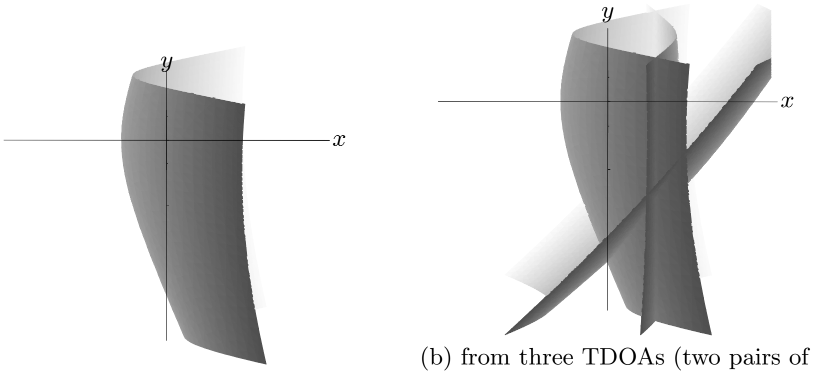
The obtained information on the source position $\hat{\mathbf{r}}_{\mathrm{s}}$ from an estimated TDOA $\hat{\tau}_{ij}$ between the signals of microphones i and j can be expressed as

$$c\hat{\tau}_{ij} = \|\hat{\mathbf{r}}_{\mathrm{s}} - \mathbf{r}_i\| - \|\hat{\mathbf{r}}_{\mathrm{s}} - \mathbf{r}_j\| \,, \tag{4.58}$$

where c denotes the velocity of sound and the vectors $\mathbf{r}_i$ and $\mathbf{r}_j$ denote the three-dimensional (or two-dimensional) positions of microphones i and j, respectively.

In three-dimensional space such an equation describes a hyperboloïd, as exemplarily shown in Fig. 4.12 (a).

From information on another time difference one obtains a second hyperboloïd. If the two pairs of microphones are placed on one straight line, the points fulfilling both conditions are describing a circle. To provide information on the position of the source on the circle, the TDOA of a third microphone pair has to be taken into account. To avoid linear dependencies, this microphone pair must not be placed on the same straight line as the former two.



(a) from one TDOA (both microphones on the x-axis)

(b) from three TDOAs (two pairs of microphones on the x-axis, third pair diagonally in the xy-plane)

Fig. 4.12. Determination of the possible source positions in the room.

For example, if we want to restrict ourselves for practical reasons to a two-dimensional microphone array in the xy-plane, the third microphone pair will be in the same plane. This results in two unique intersection points, one with positive, and one with negative z-coordinate, as shown in Fig. 4.12 (b). In many scenarios one can already uniquely determine the position with such a setup since either the negative or positive z-coordinates may be excluded (e.g., if the array is mounted on a wall).

4.5.2 Microphone Array Geometry and Grid of Potential Source Positions

Equation (4.58) represents a nonlinear set of equations. Considering this set of equations, we readily see that the geometry of the microphone array has a major influence on the calculation of the positions. In the following, we exemplarily study this influence for two different arrays for the two-dimensional case where two TDOA estimates are needed (see Section 4.5.1).

The setup in Fig. 4.13 uses the signal from the center microphone simultaneously for two equations by estimating the TDOA between microphones 1 and 2, and between 2 and 3, respectively. This has two advantages. It not only reduces the necessary number of microphones, but the calculation of the source positions, based on the nonlinear set of equations (4.58) will become much easier as there are fewer geometrical parameters if we set the origin on the position of the center microphone. Note that in general, the solution of this set of equations is not trivial and in some cases there is no closed-form solution. Thus, for more complicated array geometries, we have to resort to numerical and/or approximate solutions, rather than exact closed-form solutions. There is a rich literature on this problem which also includes the consideration of overdetermined sets of equations, i.e., taking into account more TDOA estimates in order to further improve the robustness to measurement errors. Important distinctions between these methods include likelihood-based [53–56] together with iterative optimization (e.g., using the Newton-Raphson [57] or Gauss-Newton [58] methods) versus least-squares and linear approximation versus closed-form algorithms [59–65].

On the other hand, there may also be a major disadvantage with the microphone setup after Fig. 4.13, depending on the chosen TDOA estimation method, as we will discuss in the following.

The TDOAs, estimated by the methods discussed in Section 4.5.3 are ordinarily represented by *integer* numbers, corresponding to discrete sampling instants along the time axis. Therefore the estimates of the potential source positions are restricted to a grid of *discrete* positions. The density of this grid depends on the sampling rate and on the positions of the microphones. Note that in principle this density is *independent* of the chosen TDOA estimation method for a given array geometry.

Figure 4.15 shows the possible positions that can be obtained using the microphone array after Fig. 4.13 with a spacing of $d = 16$cm and a temporal sampling rate of $f_\mathrm{s} = 48$kHz. As we can see, the spatial resolution decreases with increasing distance (*range*) between the source and the microphone array.

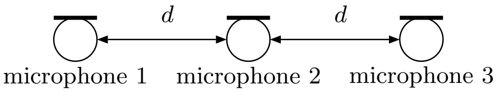


Fig. 4.13. Microphone array 1.

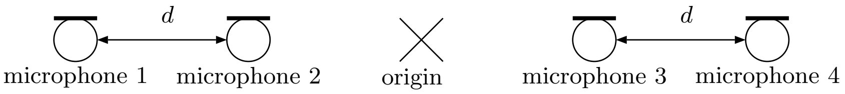


Fig. 4.14. Microphone array 2.

To obtain a better resolution with a fixed sampling rate and integer TDOAs, we can either place more pairs of microphones in the room that are closer to the respective source positions, or we change the geometric parameters of the setup in Fig. 4.13. A first possibility is to increase the distance d, i.e., the *spatial diversity* between the microphones. For $d = 50$cm the resolution is already dramatically improved, as can be seen in Fig. 4.16, and in the case of, e.g., $d = 200$cm, we would obtain a very dense grid, and a much wider coverage. This is due to the fact that the maximum time difference of arrival is increased ($\tau_\mathrm{max} = \frac{cd}{f_\mathrm{s}}$). With increasing τ_max the number of potential TDOA values for each microphone pair is also increased, and thus the number of potential positions is increased significantly. The full potential of this method can be exploited with *TDOA estimators based on blind system identification* (Section 4.5.3). With other TDOA estimators, one drawback of this method is that the precision of the TDOA estimation itself may be affected due to spatial aliasing if the microphone spacing is too large. This ambiguity problem typically occurs with narrowband implementations, i.e., the signal processing is carried out independently for each frequency bin in that implementations. The GCC is often implemented in this way. Therefore, an alternative for that case would be to use the modified setup after Fig. 4.14. Instead of increasing the spacing within the microphone pairs, we only increase the distance between the individual pairs as this affects the geometrical calculation. The TDOA estimation is only affected by the spacing d. Note, however, that a small spacing generally increases the error variance, i.e., small TDOA deviations will have a larger influence on the final position estimate. Therefore, in any case a broadband implementation of the blind system identification is recommended for accurate source localization. Moreover, to further improve the spatial resolution of the localizer at a low computational cost, *fractional* delays can be obtained with the BSI-based method by performing a sinc interpolation [66] on the filters of the unmixing system $\check{\mathbf{W}}$ before performing the effective TDOA estimations, given in (4.60) and (4.61a)/(4.61b), without further increasing the sampling rate for the BSS operations.

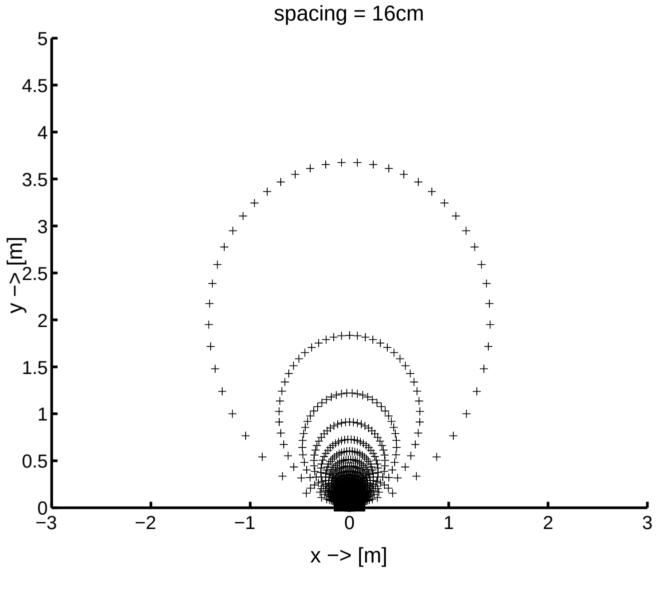


Fig. 4.15. Potential source positions using three equidistant microphones with a spacing of 16cm at a sampling rate of 48kHz.

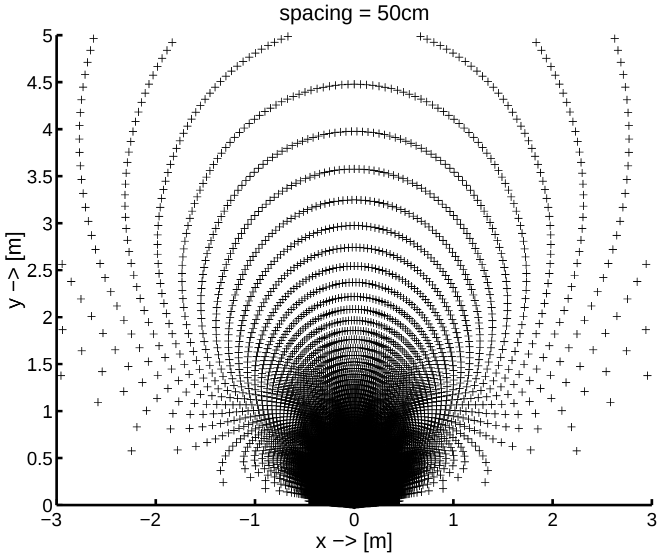


Fig. 4.16. Potential source positions using three equidistant microphones with a spacing of 50cm at a sampling rate of 48kHz.

4.5.3 Estimation of Time Differences of Arrival

The key for an effective localization with the two-step approach is an accurate and robust TDOA estimator. In the following we examine different possibilities for TDOA estimation. Thereby, as in the previous section, we concentrate here on the use of microphone pairs for clarity, although it has recently become possible to simultaneously take into account multiple microphone signals [16].

The Method of the Generalized Cross-Correlation. The method of the generalized cross-correlation (GCC) [8] is based on the ideal free-field propagation model $x_i(t) = \alpha_i s(t - \tau_i) + b_i(t)$, where b_i is an additive noise signal on the i-th microphone, and α_i denotes an attenuation factor. Due to its simplicity the method is nevertheless often applied in reverberant environments so that to date it is still the most widely used method for single-source localization.

The basic principle of this technique consists of the maximization of the inverse Fourier transformation of a weighted cross-power spectral density, i.e.,

$$\hat{\tau}_{ij} = \arg\max_{\tau_{ij}} \mathcal{F}^{-1}\left\{\Phi(f) S_{x_i x_j}(f)\right\}, \tag{4.59}$$

where $\Phi(f)$ is the weighting function, and $S_{x_i x_j}(f)$ denotes the estimated cross-power spectral density between x_i and x_j.

The numerous variants of the GCC method [8,67] differ mainly in their weighting functions and in the estimation procedure for $S_{x_i x_j}(f)$. The classical cross-correlation method (CC) uses $\Phi(f) = 1$. The weighting function $\Phi(f) = \frac{1}{|S_{x_i x_j}(f)|}$ yields the *Phase Transform* (PHAT) technique in which a behaviour independent of the spectral support is achieved by the normalization by the magnitude of the power spectral density [8].

Other improvements concentrate on the pre-filtering of the input signal, such as, e.g., the cepstral processing in [68].

However, these methods suffer from the fact that the underlying signal model of the GCC does not reflect the conditions in real acoustic environments. This becomes particularly obvious in very reverberant environments. It can be shown that here the robustness can break down abruptly with increasing reverberation time T_{60} [12,13] as will also be confirmed by our experimental evaluation in sect. 4.6.

The Blind System Identification Method. To treat this reverberation problem, a completely different approach was presented for single-source localization in [15] which is based on blind adaptive filtering using the adaptive eigenvalue decomposition algorithm. According to Sect. 4.2.2, the AED algorithm adapts itself directly to the impulse responses h_1 and h_2, i.e., the SIMO model between a source s and the microphones. Therefore, this approach is inherently based on the realistic convolutive propagation model. Note that the scaling ambiguity in blind system identification is uncritical for the TDOA estimation (as can be easily seen by Eq. (4.60) below).

To perform the adaptation, a wide range of adaptation algorithms, such as the Least-Mean-Squares (LMS) algorithm [18] (in modified form [15]), realized in the time domain, or in efficient frequency-domain realization [18] can be used.

Based on the estimated filter coefficients, the TDOA can then be calculated after each coefficient update according to

$$\begin{aligned}\hat{\tau} &= \arg\max_n |\hat{h}_{2,n}| - \arg\max_n |\hat{h}_{1,n}| \\ &= \arg\max_n |w_{1,n}| - \arg\max_n |w_{2,n}|. \end{aligned} \tag{4.60}$$

Note that here, as above in the case of GCC, we consider only one microphone pair. However, there are generalizations, both of GCC [16] and AED [16] for more than two sensors, in order to further increase the robustness by spatial redundancy.

Motivated by the high accuracy of the above-mentioned adaptive SIMO filter approach for localizing only one source, the more general approach of blind adaptive MIMO filtering for simultanteous localization of *multiple simultaneously active sources* was proposed in [6], based on the considerations discussed in Sect. 4.2 of this chapter. Thereby, the objective was to maintain the realistic convolutive propagation model for the localization, as in the case of AED. As with AED, we may calculate the Q TDOAs for the Q sources from the FIR filters w_{pq} once they are estimated by a TRINICON-based broadband adaptation algorithm, such as [38], as discussed in Sect. 4.3.3. Thereby we make the reasonable assumption that the sources are mutually uncorrelated. The extraction of the multiple TDOAs from the estimated MIMO filter coefficients is based on the relationship between the broadband BSS framework and the AED, as discussed in Sect. 4.2.2. For instance, in the case of two simultaneously active sources, (4.22a) is the corresponding equation to estimate the TDOA of source 1, while (4.22b) gives the TDOA of source 2. Moreover, since the coefficient initialization in the case of Sylvester constraint ($\mathcal{SC}_{\mathrm{R}}$), described in [6], also corresponds to the one recommended for the AED in [15], we can expect similar steady-state performances due to this close link. This is verified in Section 4.6. From these findings, we can express the TDOA estimates immediately in the same way as in (4.60) as

$$\hat{\tau}_1 = \arg\max_n |w_{12,n}| - \arg\max_n |w_{22,n}|, \tag{4.61a}$$

$$\hat{\tau}_2 = \arg\max_n |w_{11,n}| - \arg\max_n |w_{21,n}|. \tag{4.61b}$$

4.6 Simultaneous Localization of Multiple Sound Sources in Reverberant Environments using Blind Adaptive MIMO System Identification

The audio data used for the evaluation have been recorded at a sampling rate of 48 kHz in a TV studio with a reverberation time of $T_{60} \approx 700$ ms. These data are made available as part of an audio-visual database [70]. This database also includes reference data of the speaker positions measured using infrared sensors. From the reference positions reference TDOAs are calculated by geometric considerations. This allows us to consider both, fixed and moving speakers in a real acoustic environment. From the database, we chose two scenes in the same environment with one fixed and one moving source, respectively. Those are used separately for the SIMO-based approaches, and a superposition (Fig. 4.17) is used for the MIMO-based approach. The distance between the two microphones was 16 cm. For the adaptation algorithms, the filter lengths were chosen to 1024 (Obviously, this length is shorter than $L_{\mathrm{opt,sep}}$, given the above mentioned reverberation time. However, for the localization application in the given scenario, it turned out to be sufficient in order to capture the dominant reflections. Moreover, as discussed in Sect. 4.2.4, the disturbing filtering ambiguity is still avoided in this case.). The block length for the GCC (using a phase-transform (PHAT) weighting rule [8]) has been set to 1024. GCC and AED have been complemented by a signal power-based voice-activity detector. Figures 4.18 (a) and (b) show the reference and estimated TDOAs for the fixed and the moving speakers, respectively. In these first experiments, only one speaker was active (also in case of the MIMO-based approach). Subplot (a) confirms that both of the

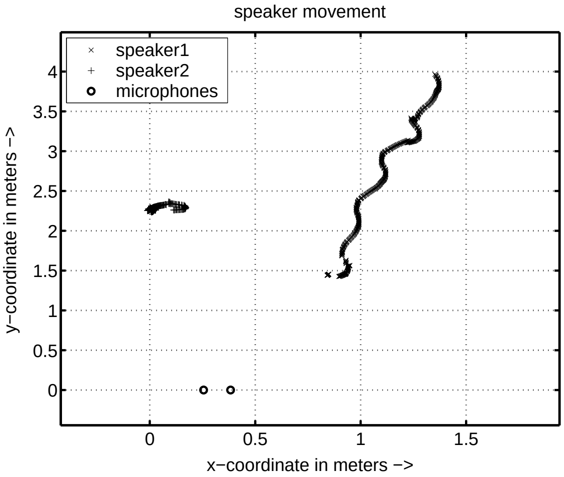
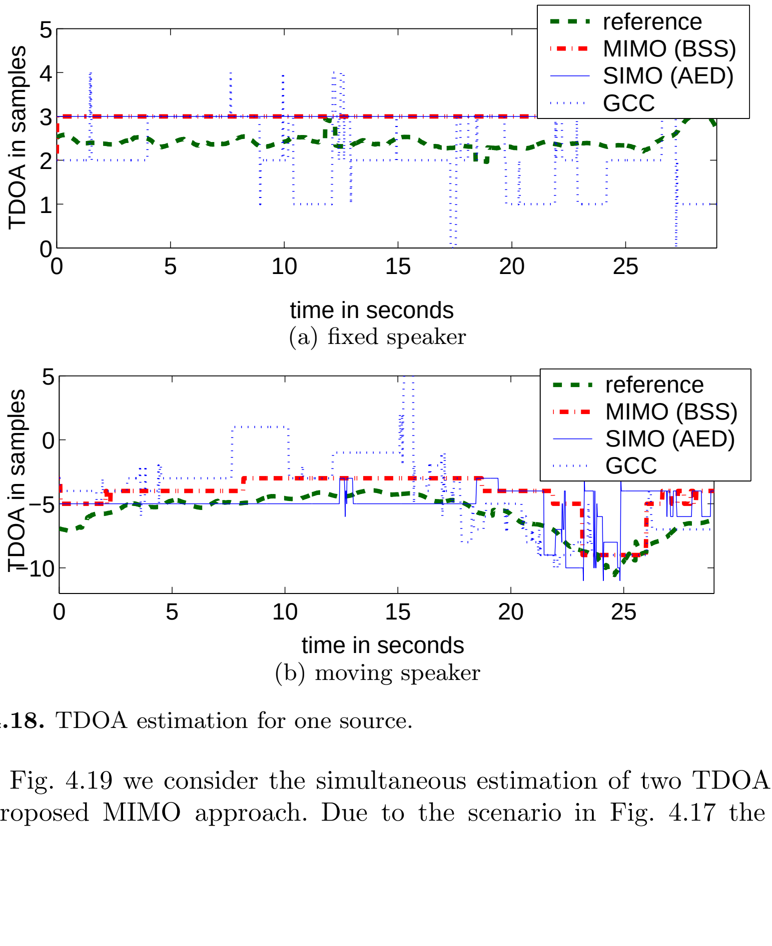


Fig. 4.17. Scenario used for the simulations.

blind adaptation algorithms lead to the same accurate TDOA estimates in this static case, as expected from the considerations in Sect. 4.5. Note that the TDOA estimates can only attain integer values.



(a) fixed speaker

(b) moving speaker

Fig. 4.18. TDOA estimation for one source.

In Fig. 4.19 we consider the simultaneous estimation of two TDOAs by the proposed MIMO approach. Due to the scenario in Fig. 4.17 the two

TDOAs exhibit different signs. The estimates deviate only slightly from the corresponding results of the MIMO-based approach in Figs. 4.18 (a) and (b) during some very short time intervals. This may be explained by the different speech activity of the two sources which is typical and inevitable for realistic situations. However, the short peaks in Fig. 4.19 may be easily removed by appropriate post-processing. Further experimental results, which also illustrate the robustness of the multiple TDOA estimates with respect to background noise and shadowing effects caused by objects placed between the sensors, may be found in [69].

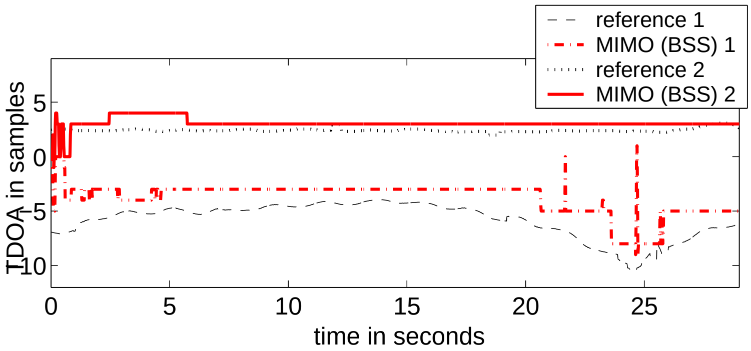


Fig. 4.19. Simultaneous TDOA estimation for two sources.

4.7 Conclusions

In this chapter we have shown the relation between convolutive broadband BSS and 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 the simultaneous localization of multiple sources. Based on the relationship between MIMO BSI and broadband BSS we were also able to further clarify the relations between various source localization approaches. As a side aspect, also some relations and similarities to the inverse problems, such as blind dereverberation of speech signals have been illuminated. For all of these applications it became obvious that a proper broadband adaptation of the filter coefficients is desirable and in some cases even absolutely necessary. In many ways the TRINICON framework turned out to be a useful tool to solve the associated problems. On the algorithmic side, we have derived a generic Sylvester constraint which unifies previous algorithmic results and may serve as a guideline for the development of new efficient algorithms.

References

1. H. Buchner, R. Aichner, and W. Kellermann, “Blind source separation for convolutive mixtures exploiting nongaussianity, nonwhiteness, and nonstationarity,” in *Proc. Int. Workshop Acoustic Echo and Noise Control (IWAENC)*, Kyoto, Japan, pp. 223-226, Sept. 2003.
2. H. Buchner, R. Aichner, and W. Kellermann, “Blind source separation for convolutive mixtures: A unified treatment,” in Y. Huang and J. Benesty (eds.), *Audio Signal Processing for Next-Generation Multimedia Communication Systems*, Kluwer Academic Publishers, Boston, pp. 255-293, Feb. 2004.
3. H. Buchner, R. Aichner, and W. Kellermann, “TRINICON: A versatile framework for multichannel blind signal processing,” in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Montreal, Canada, vol. 3, pp. 889-892, May 2004.
4. 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 Audio Process.*, vol. 13, no. 1, pp. 120-134, Jan. 2005.
5. H. Buchner, R. Aichner, and W. Kellermann, “Relation between blind system identification and convolutive blind source separation,” in *Proc. Joint Workshop Hands-Free Speech Communication and Microphone Arrays (HSCMA)*, Piscataway, NJ, USA, Mar. 2005 (additional presentation slides with more results downloadable from the web site `www.LNT.de/lms/`).
6. 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 Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Philadelphia, PA, USA, Mar. 2005.
7. M. Hofbauer, *Optimal Linear Separation and Deconvolution of Acoustical Convolutive Mixtures*, Dissertation, Hartung-Gorre Verlag, Konstanz, May 2005.
8. C.H. Knapp and G.C. Carter, ”The generalized correlation method for estimation of time delay,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-24, pp. 320-327, Aug. 1976.
9. M.S. Brandstein and D.B. Ward, *Microphone Arrays: Signal Processing Techniques and Applications*, Springer, Berlin, 2001.
10. R. O. Schmidt, “Multiple emitter location and signal parameter estimation,” *IEEE Trans. Antennas Propagation,* vol. AP-34, no. 3, pp. 276-280, March 1986.
11. R. Roy and T. Kailath, “ESPRIT - estimation of signal parameters via rotational invariance techniques,” *IEEE Trans. Acoust., Speech, Signal Processing,* vol. 37. no. 7, pp. 984-995, July 1989.
12. B. Champagne, S. Bedard, and A. Stéphenne,“Performance of time-delay estimation in the presence of room reverberation,” *IEEE Trans. Speech Audio Process.,* vol. 4, pp. 148-152, Mar. 1996.
13. J.P. Ianniello, “Time delay estimation via cross-correlation in the presence of large estimation errors,” *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-30, no. 6, pp. 998-1003, Dec. 1982.
14. J. Scheuing and B. Yang, “Disambiguation of TDOA estimates in multi-path multi-source environments (DATEMM),” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, Toulouse, France, 2006.

15. J. Benesty, “Adaptive eigenvalue decomposition algorithm for passive acoustic source localization,” *J. Acoust. Soc. Am.*, vol. 107, pp. 384-391, Jan. 2000.
16. J. Chen, Y. Huang, and J. Benesty, “Time delay estimation” in Y. Huang and J. Benesty (eds.), *Audio Signal Processing for Next-Generation Multimedia Communication Systems*, Kluwer Academic Publishers, Boston, pp. 197-227, Feb. 2004.
17. A. Lombard, H. Buchner, and W. Kellermann, “Multidimensional localization of multiple sound sources using blind adaptive MIMO system identification,” in *Proc. IEEE Int. Conf. Multisensor Fusion and Integration for Intelligent Systems (MFI)*, Heidelberg, Germany, Sept. 2006.
18. S. Haykin, *Adaptive Filter Theory*, 4th ed., Prentice-Hall, Englewood Cliffs, NJ, 2002.
19. A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, Wiley & Sons, Inc., New York, 2001.
20. S.C. Douglas, “Blind separation of acoustic signals” in M. Brandstein and D. Ward (eds.), *Microphone Arrays: Signal Processing Techniques and Applications*, pp. 355–380, Springer, Berlin, 2001.
21. J.-F. Cardoso and A. Souloumiac, “Blind beamforming for non gaussian signals,” *IEE Proceedings-F*, vol. 140, no. 6, pp. 362-370, Dec. 1993.
22. S. Araki et al., “Equivalence between frequency-domain blind source separation and frequency-domain adaptive beamforming,” in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Orlando, FL, USA, pp. 1785-1788, May 2002.
23. M. Miyoshi and Y. Kaneda, “Inverse filtering of room acoustics,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol 36, no. 2, pp. 145-152, Feb. 1988.
24. K. Furuya, “Noise reduction and dereverberation using correlation matrix based on the multiple-input/output inverse-filtering theorem (MINT),” in *Proc. Int. Workshop Hands-Free Speech Communication (HSC)*, Kyoto, Japan, pp. 59-62, Apr. 2001.
25. M.I. Gürelli and C.L. Nikias, “EVAM: An eigenvector-based algorithm for multichannel blind deconvolution of input colored signals,” *IEEE Trans. Signal Process.*, vol. 43, no. 1, pp. 134–149, Jan. 1995.
26. K. Furuya and Y. Kaneda, “Two-channel blind deconvolution of nonminimum phase FIR systems,” *IEICE Trans. Fundamentals*, vol. E80-A, no. 5, pp. 804–808, May 1997.
27. S. Amari et al.,“Multichannel blind deconvolution and equalization using the natural gradient,” in *Proc. IEEE Int. Workshop Signal Processing Advances in Wireless Communications*, pp. 101-107, 1997.
28. S. Choi et al., “Natural gradient learning with a nonholonomic constraint for blind deconvolution of multiple channels,” in *Proc. Int. Symp. Independent Component Analysis Blind Source Separation (ICA)*, pp. 371-376, 1999.
29. B.W. Gillespie and L. Atlas, “Strategies for improving audible quality and speech recognition accuracy of reverberant speech,” in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Hongkong, China, Apr. 2003.
30. K. Matsuoka and S. Nakashima, “Minimal distortion principle for blind source separation,” in *Proc. Int. Symp. Independent Component Analysis Blind Signal Separation (ICA)*, San Diego, CA, USA, Dec. 2001.
31. 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 Process.*, vol. 12, no. 8, Sept. 2004.

32. H. Liu, G. Xu, and L. Tong, "A deterministic approach to blind identification of multi-channel FIR systems," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Adelaide, Australia, Apr. 1994.
33. H.-C. Wu and J. C. Principe,"Simultaneous diagonalization in the frequency domain (SDIF) for source separation," in *Proc. Int. Symp. Independent Component Analysis Blind Signal Separation (ICA)*, pp. 245-250, 1999.
34. C.L. Fancourt and L. Parra, "The coherence function in blind source separation of convolutive mixtures of non-stationary signals," in *Proc. Int. Workshop Neural Networks Signal Processing (NNSP)*, 2001, pp. 303-312.
35. T.M. Cover and J.A. Thomas, *Elements of Information Theory*, Wiley & Sons, New York, 1991.
36. R. Aichner, H. Buchner, F. Yan, and W. Kellermann, "A real-time blind source separation scheme and its application to reverberant and noisy acoustic environments," *Signal Processing*, vol. 86, no. 6, pp.1260–1277, 2006.
37. M. Kawamoto, K. Matsuoka, and N. Ohnishi, "A method of blind separation for convolved non-stationary signals," *Neurocomputing*, vol. 22, pp. 157-171, 1998.
38. R. Aichner, H. Buchner, and W. Kellermann, "Exploiting narrowband efficiency for broadband convolutive blind source separation," *EURASIP Journal on Applied Signal Processing*, vol. 2007, pp. 1-9, Sept. 2006.
39. T. Nishikawa, H. Saruwatari, and K. Shikano, "Comparison of time-domain ICA, frequency-domain ICA and multistage ICA for blind source separation," in *Proc. European Signal Processing Conference (EUSIPCO),* vol. 2, pp. 15-18, Sep. 2002.
40. K. Yao, "A representation theorem and its applications to spherically-invariant random processes," *IEEE Trans. Inform. Theor.*, vol. 19, no. 5, pp. 600-608, Sept. 1973.
41. J. Goldman, "Detection in the presence of spherically symmetric random vectors," *IEEE Trans. Inform. Theor.*, vol. 22, no. 1, pp. 52-59, Jan. 1976.
42. H. Brehm and W. Stammler, "Description and generation of spherically invariant speech-model signals," *Signal Processing*, vol. 12, pp. 119-141, 1987.
43. S. Araki et al., "The fundamental limitation of frequency-domain blind source separation for convolutive mixtures of speech," *IEEE Trans. Speech Audio Process.,* vol. 11, no. 2, pp. 109-116, Mar. 2003.
44. H. Sawada et al., "Spectral smoothing for frequency-domain blind source separation," in *Proc. Int. Workshop Acoustic Echo and Noise Control (IWAENC)*, Kyoto, Japan, Sept. 2003, pp. 311-314.
45. M.Z. Ikram and D.R. Morgan, "Exploring permutation inconsistency in blind separation of speech signals in a reverberant environment," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Istanbul, Turkey, June 2000, vol. 2, pp. 1041-1044.
46. H. Buchner, R. Aichner, and W. Kellermann, "A generalization of a class of blind source separation algorithms for convolutive mixtures," in *Proc. Int. Symp. Independent Component Analysis Blind Signal Separation (ICA)*, Nara, Japan, Apr. 2003.
47. T. Kim, T. Eltoft, and T.-W. Lee, "Independent vector analysis: an extension of ICA to multivariate components," in *Proc. Int. Conf. Independent Component Analysis Blind Signal Separation (ICA)*, Mar. 2006.

48. A. Hiroe, "Solution of permutation problem in frequency domain ICA using multivariate probability density functions," in *Proc. Int. Conf. Independent Component Analysis Blind Signal Separation (ICA)*, pp. 601-608, Mar. 2006.
49. P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, vol. 22, pp. 21-34, Jul. 1998.
50. D. H. Johnson, D. E. Dudgeon, *Array Signal Processing*, Prentice Hall, New Jersey, 1993.
51. H. Wang, M. Kaveh, "Coherent Signal-Subspace Processing for the Detection and Estimation of Angles of Arrival of Multiple Wide-Band Sources," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, no. 4, pp.823-831, Aug. 1985.
52. H. Teutsch, W. Kellermann, "Acoustic source detection and localization based on wavefield decomposition using circular microphone arrays," *J. Acoust. Soc. Am.*, vol. 120, no. 5, Nov. 2006.
53. W.R. Hahn and S.A. Tretter, "Optimum processing for delay-vector estimation in passive signal arrays," *IEEE Trans. Inform. Theory,* vol. IT-19, pp. 608-614, May 1973.
54. M. Wax and T. Kailath, "Optimum localization of multiple sources by passive arrays," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-31, no. 5, pp. 1210-1218, Oct. 1983.
55. P.E. Stoica and A. Nehorai, "MUSIC, maximum likelihood and Cramer-Rao bound," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. 37, pp. 720-740, May 1989.
56. J.C. Chen, R.E. Hudson, and K. Yao, "Maximum-likelihood source localization and unknown sensor location estimation for wideband signals in the near-field," *IEEE Trans. Signal Process.,* vol. 50, pp. 1843-1854, Aug. 2002.
57. Y. Bard, *Nonlinear Parameter Estimation*, Academic Press, New York, 1974.
58. W.H. Foy, "Position-location solutions by Taylor-series estimation," *IEEE Trans. Aerosp. Electron. Syst.,* vol. AES-12, pp. 187-194, Mar. 1976.
59. R.O. Schmidt, "A new approach to geometry of range difference location," *IEEE Trans. Aerosp. Electron.,* vol. AES-8, pp. 821-835, Nov. 1972.
60. H.C. Schau and A.Z. Robinson, "Passive source localization employing intersecting spherical surfaces from time-of-arrival differences," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-35, no. 8, pp. 1223-1225, Aug. 1987.
61. J.O. Smith and J.S. Abel, "Closed-form least-squares source location estimation from range-difference measurements," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-35, no. 12, pp. 1661-1669, Dec. 1987.
62. Y.T. Chan and K.C. Ho, "A simple and efficient estimator for hyperbolic location," *IEEE Trans. Signal Process.,* vol. 42, no. 8, pp. 1905-1915, Aug. 1994.
63. Y.T. Chan and K.C. Ho, "An efficient closed-form localization solution from time difference of arrival measurements," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP),* 1994, vol. II, pp. 393-396.
64. Y. Huang, J. Benesty, G.W. Elko, and R.M. Mersereau, "Real-time passice source localization: an unbiased linear-correction least-squares approach," *IEEE Trans. Speech Audio Process.,* vol. 9, no. 8, pp. 943-956, Nov. 2001.
65. J.S. Abel and J.O. Smith, "The spherical interpolation method for closed-form passive source localization using range difference measurements," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP),* vol. 1, pp. 471-474, 1987.

66. T.I. Laakso et al., “Splitting the unit delay,” *IEEE Signal Processing Mag.,* vol. 13, pp. 30-60, 1996.
67. M.S. Brandstein and H.F. Silverman, “A robust method for speech signal time-delay estimation in reverberant rooms,” in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Munich, Apr. 1997.
68. A. Stéphenne and B. Champagne, “A new cepstral prefiltering technique for estimating time delay under reverberant conditions,” *Signal Processing*, vol. 59, pp. 253-266, 1997.
69. R. Aichner, H. Buchner, S. Wehr, and W. Kellermann, ”Robustness of acoustic multiple-source localization in adverse environments,” in *Proc. ITG Fachtagung Sprachkommunication*, Kiel, Germany, Apr. 2006.
70. M. Krinidis, et al., ”An audio-visual database for evaluating person tracking algorithms,” in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Philadelphia, PA, USA, Mar. 2005.