Contents

References . . . 74

TRINICON for Dereverberation of Speech and Audio Signals

Herbert Buchner[1] and Walter Kellermann[2]

[1] Deutsche Telekom Laboratories,
Berlin University of Technology,
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: wk@LNT.de

In this chapter, we develop an analytical top-down approach to the problem of blind dereverberation of speech and audio signals based on TRINICON, a general framework for broadband adaptive MIMO signal processing. Two fundamentally different approaches to the dereverberation problem for realistic scenarios can be distinguished: The "identification-and-inversion approach" which results in a two-step procedure consisting of blind identification of the acoustic MIMO mixing system, followed by an inversion of the identified system. As an alternative, the "direct-inverse approach" blindly estimates the inverse of the acoustic mixing system directly. As shown in this chapter, for both cases TRINICON yields the information-theoretically optimum estimation procedures in a unified way and allows for a direct comparison between the approaches, paves the way to synergies, and yields various useful insights for practical realizations. This chapter also relates other known algorithms, and presents novel improved algorithms as special cases of the generic concept.

1 Introduction

Blind signal processing of convolutive mixtures of unknown time series is an important building block in modern systems involving broadband signal acquisition by sensor arrays in multipath or convolutive environments. A challenging and important example for such environments is given by 'natural' acoustic human/machine interfaces using multiple microphones to support sound signal acquisition so that the users may be untethered and mobile in real rooms. To obtain the desired source signals, the signal processing generally has to cope with two fundamental problems due to the distance between the sources and the sensors: (i) the presence of additive noise and interferers, e.g., competing speakers, and (ii) the disturbing effect of reflections and scattering of the desired source signals in the recordings. In this chapter we tackle these problems by blind adaptive multiple input/multiple output (MIMO) filtering.

In this introductory section, we first formulate the fundamental adaptive filtering problems and distinguish 'direct' and 'inverse' problems in Sect. 1.1. Moreover, we introduce a classification into two different generic approaches to blind deconvolution which are fundamental to the dereverberation approaches for speech and audio signals. In Sect. 1.2 we introduce a compact matrix notation which we will use throughout this chapter. Section 1.3 provides an overview of our analysis of the two generic approaches to blind deconvolution as useful for blind dereverberation.

1.1 Generic Tasks for Blind Adaptive MIMO Filtering

The signal acquisition scenario mentioned above is modeled such that the original source signals $s_q(n)$, $q = 1, \ldots, Q$ are filtered by a linear MIMO system before they are picked up by the sensors yielding the sensor signals $x_p(n)$, $p = 1, \ldots, P$. In this chapter, we describe this MIMO mixing system by length-M finite impulse response (FIR) filters, i.e.,

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

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. 1. Throughout this chapter, we assume that the number Q of sources is less or equal to the number P of sensors. The cases $Q < P$ and $Q = P$ are of particular interest as detailed below, and they are commonly known as *overdetermined* and *(fully) 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.

Obviously, since only the sensor signals, i.e., the output signals of the mixing system, are assumed to be accessible to the blind signal processing, *any*

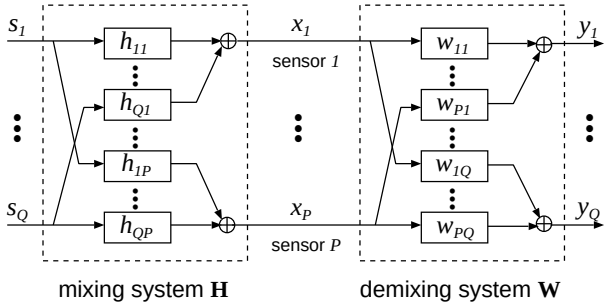


Fig. 1. Setup for blind MIMO signal processing.

type of linear blind adaptive MIMO signal processing may be described by the structure shown in Fig. 1. Thus, with respect to a yet undefined 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

$$y_q(n) = \sum_{p=1}^{P} \sum_{\kappa=0}^{L-1} w_{pq,\kappa} x_p(n-\kappa) \tag{2}$$

and where the parameter L denotes the FIR filter length of the demixing filters with coefficients $w_{pq,\kappa}$.

Depending on the optimization criterion for determining the coefficients $w_{pq,\kappa}$, we distinguish two general classes of blind signal processing problems as summarized in Tab. 1 along with the corresponding supervised problems[3,4]:

- **"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 permutation ambiguity* 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 [2, 3]. In another interpretation, BSS may be considered as a set of *blind beamformers* [4, 5] under certain restricting conditions, most notably the

[3] Note that in supervised adaptive filtering one may distinguish the analogous general classes of problems. There, we classify system identification and interference cancellation after [1] as (there may be others, or at least other terms) "direct supervised adaptive filtering problems", whereas inverse modeling and linear prediction after [1] may be classified as "inverse supervised adaptive filtering problems".

[4] The TRINICON framework for broadband adaptive MIMO filtering presented in Sect. 4 of this chapter is applicable to all of the problems listed in Tab. 1 and yields corresponding generic adaptation algorithms.

fulfilment of the spatial sampling theorem by the microphone array. Furthermore, under the farfield assumption, the directions of arrival (DOAs) can be extracted from the corresponding array patterns, which in turn can be calculated from the BSS filter coefficients, e.g., [6].
In this chapter (Sect. 3) we will see that, more generally, a properly designed broadband BSS system actually performs blind MIMO system identification (which is independent of the spatial sampling theorem). The general broadband approach presented here unifies the BSS and BSI concepts and provides various algorithmic synergy effects and new applications. One important and particularly illustrative application of the general broadband approach to MIMO BSI is the acoustic localization of multiple simultaneously active sources even in reverberant environments as detailed in [7, 8]. In this chapter, we utilize the general MIMO BSI approach for deconvolution and especially to dereverberation of acoustic signals (see below) as another new application.

- **"Inverse blind adaptive filtering problems":** This class stands here for multichannel blind deconvolution (MCBD) and the so-called multichannel blind partial deconvolution (MCBPD)[5] w.r.t. the mixing system $\mathbf{H}$ and forms the main part of this chapter. Furthermore, the linear prediction problem as known from the literature on supervised adaptive filtering may also be considered as an inverse blind adaptive filtering problem, as we show in this chapter. The relation between linear prediction and MCBD/MCBPD will also be shown later in this chapter.
  The goal of any *blind deconvolution* approach is to recover the original signals *up to an arbitrary* (frequency-independent) *scaling and possibly a time shift*. In the general MIMO case, i.e., for multiple simultaneously active sources, blind deconvolution also includes separation of the source signals (up to a permutation ambiguity). MCBD and MCBPD provide adaptive methods to the blind deconvolution problem for independent identically distributed (i.i.d.) sources and for general nonwhite sources, respectively.
  For the intended acoustic applications, i.e., for speech and audio source signals, the problem of blind deconvolution means that we want to *dereverberate* the signals by inverting the effect of the convolutive mixture matrix $\mathbf{H}$. In this case, blind deconvolution is denoted by *blind dereverberation*.
  Furthermore, for blind dereverberation, i.e., in acoustic applications, we typically have to deal with nonwhite sources. Hence, for a direct adaptive approach to blind dereverberation the more general MCBPD method has to be used, as we will discuss later in more detail.
  In terms of the MIMO system description, for the task of blind deconvo-

[5] Later in Sect. 6 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. 6 of this chapter to address this issue, also belongs to the class of inverse blind adaptive filtering problems.

lution/blind dereverberation, strictly speaking, an inversion of (long and usually nonminimum-phase) room impulse responses is necessary. However, using the multiple-input/output inverse theorem (MINT) [9], any MIMO FIR system $\mathbf{H}$ can exactly be inverted by a MIMO FIR system $\mathbf{W}$ if P, Q, and L are suitably chosen, and if the impulse responses 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. In this chapter we present adaptive blind deconvolution algorithms which should ideally converge to the ideal MINT solution.

	supervised adaptive filtering problems (after [1])	blind adaptive filtering problems (treated in this chapter)
"direct adaptive filtering problems"	system identification	blind system identification
	interference cancellation	blind source separation/ blind interference cancellation
"inverse adaptive filtering problems"	inverse modeling/equalization	blind (partial) deconvolution
	linear prediction	linear prediction

Table 1. Classification of the linear adaptive filtering problems.

From the two classes of blind adaptive filtering problems shown in Tab. 1, it becomes obvious that two different fundamental approaches to effective blind deconvolution – and thus to dereverberation – 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., [10, 11]. In this chapter we refer to this approach as the *identification-and-inversion approach* (II approach) to blind deconvolution.

The other, theoretically equivalent but, as we will see later, in practice often more reliable approach is to perform directly a blind estimation of the actual inverse of the MIMO mixing system, e.g., [12, 13, 14, 15]. In this chapter we refer to this approach as the *direct-inverse approach* (DI approach) to blind deconvolution. Note that for blind *dereverberation*, the DI approach implies the application of MCBPD for nonwhite signals.

1.2 A Compact Matrix Formulation for MIMO Filtering Problems

To compactly formulate and analyze the blind adaptive MIMO filtering problems in Sections 2 and 3, respectively, we introduce the following matrix formulation of the overall system in Fig. 1 consisting of the mixing and demixing systems. This matrix formulation is also used in the TRINICON framework described later in Sect. 4 in order to blindly estimate the adaptive demixing filter coefficients.

For capturing the mixing system with coefficients $h_{qp,\kappa}$, $\kappa = 0, \ldots, M-1$ and the demixing system with 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

$$\check{\mathbf{H}} = \begin{bmatrix} \mathbf{h}_{11} & \cdots & \mathbf{h}_{1P} \\ \vdots & \ddots & \vdots \\ \mathbf{h}_{Q1} & \cdots & \mathbf{h}_{QP} \end{bmatrix} \tag{3}$$

and the $PL \times Q$ demixing coefficient matrix

$$\check{\mathbf{W}} = \begin{bmatrix} \mathbf{w}_{11} & \cdots & \mathbf{w}_{1Q} \\ \vdots & \ddots & \vdots \\ \mathbf{w}_{P1} & \cdots & \mathbf{w}_{PQ} \end{bmatrix}, \tag{4}$$

respectively, where

$$\mathbf{h}_{qp} = [h_{qp,0}, \ldots, h_{qp,M-1}]^{\mathrm{T}}, \tag{5}$$

$$\mathbf{w}_{pq} = [w_{pq,0}, \ldots, w_{pq,L-1}]^{\mathrm{T}} \tag{6}$$

denote the coefficient vectors of the individual FIR filters of the MIMO systems, and where superscript $^{\mathrm{T}}$ denotes transposition of a vector or a matrix. The downwards pointing hat symbol ('check') on top of $\mathbf{H}$ and $\mathbf{W}$ in (3) and (4) serves to distinguish these *condensed* matrices from the corresponding larger matrix structures as introduced below in (10). Although seemingly a merely formal peculiarity, the rigorous distinction between these different matrix structures is an essential tool for the development 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 demixing filter outputs are combined into the $Q(M+L-1) \times Q$ matrix

$$\check{\mathbf{C}} = \begin{bmatrix} \mathbf{c}_{11} & \cdots & \mathbf{c}_{1Q} \\ \vdots & \ddots & \vdots \\ \mathbf{c}_{Q1} & \cdots & \mathbf{c}_{QQ} \end{bmatrix}, \tag{7}$$

where

$$\mathbf{c}_{qr} = [c_{qr,0}, \ldots, c_{qr,M+L-2}]^{\mathrm{T}}. \tag{8}$$

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

$$\mathbf{c}_{qr} = \sum_{p=1}^{P} \mathbf{H}_{qp,[L]} \mathbf{w}_{pr}. \tag{9}$$

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

$$\mathbf{H}_{qp,[L]} = \begin{bmatrix} h_{qp,0} & 0 & \cdots & 0 \\ h_{qp,1} & h_{qp,0} & \ddots & \vdots \\ \vdots & h_{qp,1} & \ddots & 0 \\ h_{qp,M-1} & \vdots & \ddots & h_{qp,0} \\ 0 & h_{qp,M-1} & \ddots & h_{qp,1} \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & h_{qp,M-1} \end{bmatrix}. \tag{10}$$

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 (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 matrix elements, e.g., i of $h_{qp,i}$ in (10).

We may now compactly express the overall system matrix $\check{\mathbf{C}}$ after (7) using this Sylvester matrix formulation to finally obtain

$$\check{\mathbf{C}} = \mathbf{H}_{[L]} \check{\mathbf{W}}, \tag{11}$$

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

$$\mathbf{H}_{[L]} = \begin{bmatrix} \mathbf{H}_{11,[L]} & \cdots & \mathbf{H}_{1P,[L]} \\ \vdots & \ddots & \vdots \\ \mathbf{H}_{Q1,[L]} & \cdots & \mathbf{H}_{QP,[L]} \end{bmatrix}. \tag{12}$$

Based on this matrix formulation, we are now able to compactly formulate the blind adaptive MIMO filtering problems in the upcoming Sections 2 and 3, and to discuss the corresponding ideal solutions, regardless of how the adaptation is actually performed in practice (note that this also implies that the results are valid for both blind and supervised adaptation). The blind adaptation of the demixing filter coefficients towards these ideal solutions will be treated later in Sections 4 to 6.

1.3 Overview of this Chapter

This chapter consists of three parts. Based on the matrix notation in Sect. 1.2, we formulate and analyze both the above-mentioned inverse and the direct blind adaptive MIMO filtering problems in Sections 2 and 3 respectively, and we relate these categories of adaptive MIMO filtering problems to the two fundamental approaches to blind deconvolution, i.e., the DI approach and the II approach. As it turns out, the explicit formulation and analysis of the theoretically ideal solution of the direct filtering problems is somewhat more involved and less well known than the one of the inverse filtering problem. Accordingly, Sect. 3 gives a detailed review of a recent comprehensive treatment [8] of the direct filtering problems. Thereby, a fundamental relation between BSI and BSS for convolutive mixtures is of particular practical importance. The resulting practical scheme for BSI serves as a basis for the identification-and-inversion approach to blind deconvolution in the general MIMO case. Section 3 follows in this regard the ideas first outlined in [7, 16].

Section 4 constitutes the second major part of this chapter and is devoted to the adaptation of the MIMO demixing system towards the ideal solutions discussed in Sections 2 and 3. Our considerations are based on TRINICON, a previously introduced versatile framework for broadband adaptive MIMO signal processing [12, 18, 19, 20], which is especially well suited for speech and audio signals. The general information-theoretic optimization criterion of TRINICON allows to exploit all fundamental properties of the excitation signals, such as their nonstationarity, their spectral characteristics (nonwhiteness), and their probability densities (nongaussianity). Moreover, in addition to the inherent broadband structure necessary for a proper system identification and deconvolution, the top-down, i.e., *deductive* approach of the TRINICON framework also allows us to present relations to both already known and new efficient algorithms. So far, this deductive approach has already led to various new insights into several classes of adaptive filtering problems shown in Table 1, most notably blind source separation [8, 19], blind system identification including a generic framework for source localization [8], and the corresponding supervised adaptive problems [21]. Based on the ideas first outlined in [12], the aim of this chapter is to consider TRINICON for the inverse blind adaptive problems in more detail.

In the third part of this chapter we first apply TRINICON to BSS and the identification-and-inversion approach to blind deconvolution/blind dereverberation in Sect. 5, followed by the application to the direct-inverse approach in Sect. 6. As in the previously studied classes of adaptive filtering problems, we will see that the general framework again allows us to relate various known and seemingly different algorithms for dereverberation, and it also yields improvements beyond the current state of the art. Section 7 presents results for both the II approach and the DI approach.

2 Ideal Inversion Solution and Direct-Inverse Approach to Blind Deconvolution

This section presents a concise summary on the ideal inversion solution for MIMO FIR systems. This inversion solution represents the ideal solution of the DI approach to blind deconvolution. Hence, its discussion also yields important guidelines for the design of the adaptive system based on the DI approach.

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. 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[6], this condition may be expressed in terms of an *ideal* $Q(M+L-1) \times Q$ *overall system matrix*

$$\check{\mathbf{C}}_{\text{ideal,inv}} = \text{Bdiag}\left\{[0,\ldots,0,1,0,\ldots,0]^{\text{T}},\ldots,[0,\ldots,0,1,0,\ldots,0]^{\text{T}}\right\}\boldsymbol{\Lambda}_{\alpha}, \tag{13}$$

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

$$\mathbf{H}_{[L]}\check{\mathbf{W}} = \check{\mathbf{C}}_{\text{ideal,inv}}. \tag{14}$$

This system of linear equations may generally be solved exactly or approximately by the Moore-Penrose pseudoinverse (e.g., [22]), denoted by $\cdot^{+}$, so that

$$\begin{aligned}\check{\mathbf{W}}_{\text{LS,inv}} &= \mathbf{H}_{[L]}^{+}\check{\mathbf{C}}_{\text{ideal,inv}}\\ &= \left[\mathbf{H}_{[L]}^{\text{T}}\mathbf{H}_{[L]}\right]^{-1}\mathbf{H}_{[L]}^{\text{T}}\check{\mathbf{C}}_{\text{ideal,inv}}.\end{aligned} \tag{15}$$

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

$$\check{\mathbf{W}}_{\text{LS,inv}} = \arg\min_{\check{\mathbf{W}}} \|\mathbf{H}_{[L]}\check{\mathbf{W}} - \check{\mathbf{C}}_{\text{ideal,inv}}\|^2. \tag{16}$$

It can be shown that under certain conditions which can be fulfilled in practice and are described below, this solution becomes the ideal inversion solution, i.e., the pseudoinverse in (15) turns into the true matrix inverse,

[6] It could formally be described by an additional permutation matrix in the ideal solution. However, since in many practical cases this ambiguity may be resolved by a signal classification approach or other prior information, we renounced on this formal treatment for clarity.

$$\check{\mathbf{W}}_{\text{ideal,inv}} = \mathbf{H}_{[L]}^{-1}\check{\mathbf{C}}_{\text{ideal,inv}}. \tag{17}$$

The principle to calculate the exact inverse using (17) is known as the Multiple-input/output INverse Theorem (MINT) [9] 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 row 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 [9]. 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 in conjunction with (12). From this condition, we immediately obtain the *optimum filter length for inversion* [23]:

$$L_{\text{opt,inv}} = \frac{Q}{P-Q}(M-1). \tag{18}$$

As an important consequence 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 (17). Otherwise, we have to resort to the general LS approximation (15) with $L_{\text{opt,inv}} = \lceil Q(M-1)/(P-Q) \rceil$.

Based on the generic TRINICON framework for adaptive MIMO filtering in Sect. 4, we will present in Sect. 6 a coherent overview of blind deconvolution algorithms which aim at the ideal inversion solution (15) or the general LS solution (17) for a suitable choice of parameters, respectively.

3 Ideal Solution of Direct Adaptive Filtering Problems, and Identification-and-Inversion Approach to Blind Deconvolution

As an alternative deconvolution approach, the "identification-and-inversion approach" to blind deconvolution is based on a two-step procedure: first, the acoustic MIMO mixing system is blindly identified, and then the identified system is inverted in a separate step. Obviously, for the latter step the results of the previous section can be applied, preferably the MINT solution. In this section, we therefore concentrate on the ideal solution of the system identification step. As we shall see in this section, the relation between source separation and MIMO system identification is of fundamental importance for the practical realization of blind system identification.

In contrast to the inversion problem, the goal of any separation algorithm, such as BSS or conventional beamforming, is to eliminate only the crosstalk between the different sources $s_q(n)$, $q = 1, \ldots, Q$ in the output signals $y_q(n)$,

$q = 1, \ldots, Q$ of the demixing system (see Fig. 1). Disregarding again a potential permutation among the output signals, this condition may be expressed in terms of the overall system matrix $\check{\mathbf{C}}$ as

$$\check{\mathbf{C}} - \text{bdiag}\left\{\check{\mathbf{C}}\right\} = \text{boff}\left\{\check{\mathbf{C}}\right\} = \mathbf{0}. \tag{19}$$

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 (11), the condition for the ideal separation is expressed as

$$\text{boff}\left\{\mathbf{H}_{[L]}\check{\mathbf{W}}\right\} = \mathbf{0}. \tag{20}$$

This relation for the ideal solution of the *direct blind adaptive filtering problems* is the analogous expression to the relation (14) for the ideal solution of the inverse blind adaptive filtering problems.

As we will see in this section, the relation (20) allows us

- to derive an explicit expression of the ideal separation solution analogously to (17)
- to establish a link between BSS and BSI which will serve as an important basis to the identification-and-inversion approach to blind dereverberation in the general MIMO case
- to establish the conditions for ideal BSI
- to derive the optimum separation FIR filter length $L_{\text{opt,sep}}$ analogously to (18) for which the ideal separation solution (19) can be achieved.

If we are only interested in separation with certain other constraints to the 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 (20) on the block-offdiagonals. For instance, the so-called *minimum distortion principle* after [24] can 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.

Traditionally, blind source separation (BSS) has often been considered as an inverse problem (e.g., [2, 3]). 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 L we show that for broadband algorithms the well-known filtering ambiguity (e.g., [50]) can be avoided. 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 [10, 11, 26]. This section follows the ideas outlined in [7, 16]. Some of these ideas were also developed independently in [17] in a slightly different way.

This section discusses the ideal separation condition $\text{boff}\left\{\mathbf{H}_{[L]}\check{\mathbf{W}}\right\} = \mathbf{0}$ as illustrated in Fig. 2 for the case $Q = P = 3$. Since in this equation we

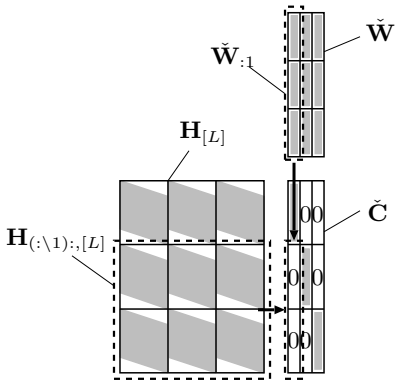


Fig. 2. Overall system $\check{\mathbf{C}}$ for the ideal separation, illustrated for $P = Q = 3$.

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

$$\mathbf{H}_{(:\backslash q):,[L]}\check{\mathbf{W}}_{:q} = \mathbf{0}, \quad q = 1, \ldots, Q \tag{21}$$

to be solved. Each of these systems of equations results from the constraints on one column of $\check{\mathbf{C}}$, as illustrated in Fig. 2 for the first column. The notation in the indices in (21) indicates that for the q-th column $\check{\mathbf{W}}_{:q}$ of the demixing filter matrix $\check{\mathbf{W}}$, we form a submatrix $\mathbf{H}_{(:\backslash q):,[L]}$ of $\mathbf{H}_{[L]}$ by removing 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 (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 $\check{\mathbf{W}}_{:q}$. Based on this and later in this section, we will also derive an expression of the optimum separation filter length $L_{\text{opt,sep}}$ for an arbitrary number of sensors and sources analogously to the optimum inversion filter length $L_{\text{opt,inv}}$ in (18).

In the following subsections, we first discuss the solution of (21) for the case $P = Q = 2$, and then generalize the results to more than two sources and sensors.

3.1 Ideal Separation Solution for Two Sources and Two Sensors

For the case $Q = P = 2$, the set of homogeneous linear systems of equations (21) reads

$$\mathbf{H}_{11,[L]}\mathbf{w}_{12} + \mathbf{H}_{12,[L]}\mathbf{w}_{22} = \mathbf{0}, \tag{22a}$$
$$\mathbf{H}_{21,[L]}\mathbf{w}_{11} + \mathbf{H}_{22,[L]}\mathbf{w}_{21} = \mathbf{0}. \tag{22b}$$

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:

$$H_{11}(z)W_{12}(z) + H_{12}(z)W_{22}(z) = 0, \tag{23a}$$
$$H_{21}(z)W_{11}(z) + H_{22}(z)W_{21}(z) = 0. \tag{23b}$$

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:

$$A_{H_{11}} \prod_{\nu=1}^{M-1} (z - z_{0H_{11},\nu}) \cdot A_{W_{12}} \prod_{\mu=1}^{L-1} (z - z_{0W_{12},\mu}) = - A_{H_{12}} \prod_{\nu=1}^{M-1} (z - z_{0H_{12},\nu}) \cdot A_{W_{22}} \prod_{\mu=1}^{L-1} (z - z_{0W_{22},\mu}), \tag{24a}$$

$$A_{H_{21}} \prod_{\nu=1}^{M-1} (z - z_{0H_{21},\nu}) \cdot A_{W_{11}} \prod_{\mu=1}^{L-1} (z - z_{0W_{11},\mu}) = - A_{H_{22}} \prod_{\nu=1}^{M-1} (z - z_{0H_{22},\nu}) \cdot A_{W_{21}} \prod_{\mu=1}^{L-1} (z - z_{0W_{21},\mu}). \tag{24b}$$

Analogously to the case of MINT [9] 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 (24a) and $H_{21}(z)$ and $H_{22}(z)$ in (24b), respectively, do not share common zeros. If no common zeros exist and if we choose the *optimum*[7] *filter length for the case* $Q = P = 2$ as $L_{\text{opt,sep}} = M$, then the equality in (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 (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 (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 (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 separation filter matrix $\check{\mathbf{W}}_{\text{ideal,sep}}$ given in the time domain as

$$\check{\mathbf{W}}_{\text{ideal,sep}} = \begin{bmatrix} \alpha_1 \mathbf{h}_{22} & -\alpha_2 \mathbf{h}_{12} \\ -\alpha_1 \mathbf{h}_{21} & \alpha_2 \mathbf{h}_{11} \end{bmatrix} = \begin{bmatrix} \mathbf{h}_{22} & -\mathbf{h}_{12} \\ -\mathbf{h}_{21} & \mathbf{h}_{11} \end{bmatrix} \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix}, \tag{25}$$

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

[7] Note that for $L < L_{\text{opt,sep}} = M$ it is obviously not possible to compensate all zeros of $H_{11}(z)$ and $H_{12}(z)$ by $W_{22}(z)$ and $W_{12}(z)$, respectively. On the other hand, in the case $L > L_{\text{opt,sep}} = M$, the filters $W_{12}(z)$ and $W_{22}(z)$ will exhibit $L - M$ arbitrary common zeros which are undesired. We will consider the practically important issue of order-overestimation in Sect. 3.5.

From (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 support for exciting the mixing system). In Section 4, a suitable algorithmic framework for this task will be presented. Moreover, as we will see in the following subsection, this approach can be seen as a generalization of the state-of-the-art method for the blind identification of SIMO systems.

In practice, the difficulty of finding the correct filter length $L_{\text{opt,sep}}$ is obviously another important issue since the length M of the mixing system is generally unknown. In Sect. 3.5 we will address this problem and the consequences of overestimation and underestimation, respectively.

3.2 Relation to MIMO and SIMO System Identification

From a system-theoretic point of view, the BSS approach aiming at the ideal solution (25) can be interpreted as a generalization of the popular class of blind SIMO system identification approaches, e.g., [10, 11, 27], as illustrated in Fig. 3a. The main reason for the popularity of this SIMO approach is

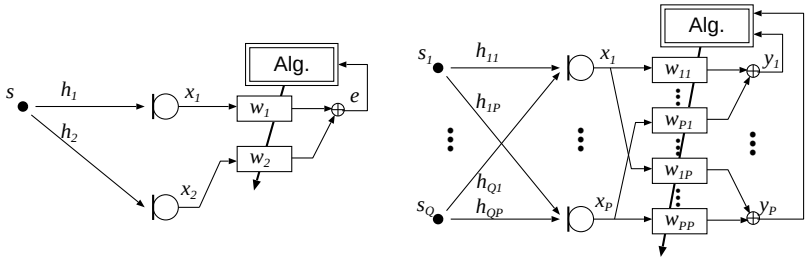


Fig. 3. Blind system identification based on (a) SIMO and (b) MIMO models.

that the optimum filters can be found as the result of a relatively simple least-squares error minimization. From Fig. 3a and for $e(n) = 0$ it follows for sufficient excitation $s(n)$ that

$$h_1(n) * w_1(n) = -h_2(n) * w_2(n). \tag{26}$$

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 (23a) and (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 (26) 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 similarly in [11].

Note that the SIMO case may also be interpreted as a special 2×2 MIMO case according to Fig. 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 must also be identical to zero, whereas the other BSS output signal is not of interest in this case. This leads again to the cancellation condition (26), 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 can 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 [10, 11], and blind system identification for passive source localization, e.g., in [26, 28]. In the latter case, this algorithm is also known as the *adaptive eigenvalue decomposition* (AED) algorithm which points to the fact that, in the SIMO case, the homogeneous system of equations (21) may be reformulated as 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 [10, 28] this SIMO approach, i.e., the single-source case, was also generalized to more than $P = 2$ microphone channels. In Sect. 5 we will present how - from an algorithmic point of view - the AED indeed directly follows from the general TRINICON framework for broadband adaptive MIMO filtering. Moreover, this will lead to a generalization of the original least-squares-based AED algorithm so that it is able to additionally exploit higher-order statistics and also contains an inherent adaptation control. This algorithmic link between the SIMO and MIMO cases will also lead to important insights for the direct-inverse approach to blind deconvolution later in Sect. 6.

3.3 Ideal Separation Solution and Optimum Separation 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 (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 [9] 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 separation filter length as

$$L_{\text{sep}} > \frac{Q-1}{P-Q+1}(M-1). \tag{27}$$

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 separation filter length* as

$$L_{\text{opt,sep}} = \frac{(Q-1)(M-1)+1}{P-Q+1}. \tag{28}$$

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 $L_{\text{opt,sep}}$ is an integer number in order to allow the ideal separation solution. Otherwise, we have to resort to approximations by choosing, e.g., the next higher integer, i.e., $L_{\text{opt,sep}} = \lceil [(Q-1)(M-1)+1]/(P-Q+1) \rceil$.

To actually obtain the ideal separation solution $\check{\mathbf{W}}_{\text{ideal,sep}}$ with (28) for the general, i.e., not necessarily square case $P \geq Q$, we consider again the original set of homogeneous systems of linear equations (21). For the choice $L = L_{\text{opt,sep}}$, 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 (21). The corresponding augmented set of linear systems of equations

$$\tilde{\mathbf{H}}_{(:\backslash q):,[L]} \check{\mathbf{W}}_{:q} = \mathbf{0}, \quad q = 1, \ldots, Q. \tag{29}$$

is equivalent to the original set (21). However, we may now interpret the *general solution vector* $\check{\mathbf{W}}_{:q}$ *of (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 (28) for the optimum separation filter length plays the same role for BSI as (18) for inversion. 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 can be met for $P = Q$. Moreover, for the special case $P = Q = 2$, the general expression (28) also confirms the choice $L_{\text{opt,BSS}} = M$ as already obtained in Sect. 3.1. Figure 4 compares the different optimum filter lengths by an example.

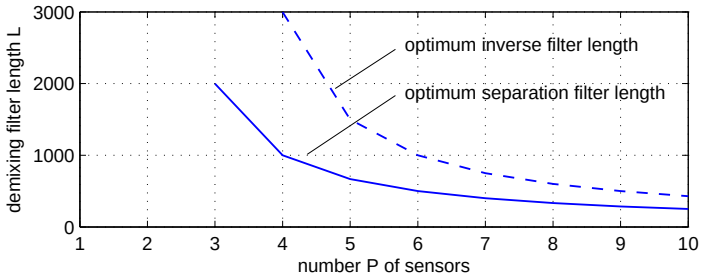


Fig. 4. Comparison of the optimum filter lengths for inversion and separation for $M = 1000$ and $Q = 3$.

3.4 General Scheme for Blind System Identification

In Sections 3.1 and 3.2 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, a simple approach is to apply several of these schemes in parallel, e.g., [29]), and also showed that the MIMO case with two simultaneously active sources is a generalization of the SIMO system identification method. In the case of more than two sources we cannot directly extract the estimated mixing system coefficients $h_{qp,\kappa}$ from the separation solution $\check{\mathbf{W}}$. The previous Section 3.3 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 3.1 and 3.2 for the *identification* task which is the first step of the identification-and-inversion approach to blind deconvolution, as detailed in Sect. 3.5. 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) resulting in an estimate of the demixing system matrix.
(2) Analogously to the relation (21) between the mixing and demixing systems, and the associated considerations in Sect. 3.3 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 (21) 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 (21), i.e.,

$$\left(\mathbf{W}^{\mathrm{bT}}\right)_{(:\backslash q):,[M]} \left(\check{\mathbf{H}}^{\mathrm{bT}}\right)_{:q} = \mathbf{0}, \;\; q = 1, \ldots, Q. \tag{30}$$

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. 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. 3.3, we may then calculate the corresponding estimate of the mixing system in terms of eigenvectors using the complementary form (30) 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.

3.5 Application of Blind System Identification to Blind Deconvolution

In order to obtain a complete blind dereverberation system after the identification-and-inversion approach, the considerations in the previous sections suggest the structure shown in Fig. 5. As discussed above, the acoustic MIMO mixing system can be blindly identified by means of an adaptive broadband BSS algorithm. Algorithmic solutions will be detailed in Sect. 5 based on the TRINICON framework outlined in Sect. 4. For the subsequent inversion of the estimated mixing system we refer to Sect. 2.

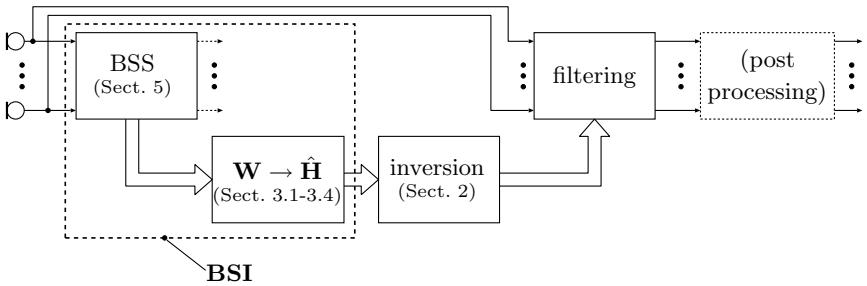


Fig. 5. Identification-and-inversion approach to blind dereverberation.

Attractive features of the identification-and-inversion approach to blind dereverberation are that (1) it is relatively easy to deal with an increased

number of microphone channels (the so-called overdetermined case for blind adaptive filtering) by simple parallelization of BSI algorithms, and (2) the approach is applicable for nearly arbitrary audio source signals, as long as they exhibit sufficient spectral support.

Based on the blind SIMO system identification mentioned in Sect. 3.2 (i.e., the estimate of the channel impulse responses is the eigenvector corresponding to the minimum eigenvalue of the correlation matrix), the identification-and-inversion approach to blind dereverberation was proposed, e.g., in [10, 11] for one acoustic source signal.

Using the general scheme for blind MIMO system identification from the previous Sect. 3.1-3.4 and the TRINICON framework shown below, we are now in a position to generalize the identification-and-inversion approach to multiple simultaneously active sources, i.e., to the MIMO case. Note that the MINT after Sect. 2 is already capable of handling the general MIMO case for $P < Q$. As in the SIMO case, the blind MIMO system identification approach has already been successfully applied in the context of passive source localization in reverberant environments, e.g., in [7, 8].

Note that previously, in [29], the identification-and-inversion approach was discussed for the MIMO case under the assumption that from time to time each source signal occupies a time interval exclusively. Then, during every single-talk interval, a SIMO system was blindly identified and its channel impulse responses were saved for later dereverberation when more than one source was active. Obviously, in practice, the applicability of this approach will be very limited in time-varying environments and with increasing numbers of independent sources (consider, e.g., a cocktail party scenario). In addition, a sophisticated multichannel sound source detection algorithm that distinguishes single and multiple speaker activity would be needed in practice. Such a required multichannel adaptation control is inherently available in TRINICON-based BSS/BSI algorithms for the general MIMO case.

However, both in the SIMO case and in the general MIMO case, there are still some fundamental challenges in the context of this dereverberation approach:

- The channel impulse responses must not exhibit common zeros in the z-domain (both for the system identification (see Sections 3.1 and 3.3) and also for the subsequent system inversion (see Sect. 2)).
- The filter length must be known exactly (both for the system identification (see Sections 3.1 and 3.3) and for the subsequent system inversion (see Sect. 2)).

The first problem can be mitigated in practice by increasing the number of microphones so that the probability for common zeros is reduced [9]. Hence, the choice of the correct filter length $L_{\text{opt,sep}}$ is the major remaining difficulty in this approach[8].

[8] Note that in some other applications of blind adaptive filtering we do not require a complete identification of the mixing system. For instance, for acoustic

The consequences of overestimation and underestimation of the filter order can be seen, e.g., from (24a) and (24b): In the case of *underestimation*, i.e., for $L < L_{\text{opt,sep}} = M$ it is obviously not possible to compensate all zeros of $H_{11}(z)$ and $H_{12}(z)$ by $W_{22}(z)$ and $W_{12}(z)$, respectively. The case of *overestimation*, i.e., $L > L_{\text{opt,sep}} = M$, is by far more problematic. In this case, the filters $W_{12}(z)$ and $W_{22}(z)$ will exhibit $L - M$ arbitrary common zeros which are undesired. This corresponds to the requirement to narrow down the solution space addressed in Sect 3.3, by avoiding an overestimation of the filter length in order to prevent a filtering ambiguity. In other words, in the overestimated case, the ideal blind identification solution $\hat{H}_1(z) = \alpha H_1(z)$ and $\hat{H}_2(z) = \alpha H_2(z)$ turns into $\hat{H}_1(z) = C_{\min}(z)H_1(z)$ and $\hat{H}_2(z) = C_{\min}(z)H_2(z)$ with the *common polynomial* $C_{\min}(z)$ corresponding to an arbitrary filtering. Consequently, after the inverse filtering in Fig. 5, the overestimation of the filter length would result in a remaining filtering $1/C_{\min}(z)$ of the original source signals.

Various ways exist to solve the filtering ambiguity problem caused by the overestimation of the filter order. The transfer function order could be obtained if the dimension of the nullspace in the autocorrelation matrix of the observed signals is precisely calculated [10, 30], i.e., by counting the number of very small eigenvalues. Another way to find the optimum order is to use a suitable cost function, e.g., [11, 31, 32]. Unfortunately, these blind system order estimation approaches are often unreliable (particularly in noisy environments) and computationally too complex (especially the latter ones, i.e., [11, 31, 32]). An alternative approach proposed, e.g., in [33] is to *compensate* for the remaining filtering $1/C_{\min}(z)$ using a post filter (Fig. 5) by estimating the common polynomial with a multichannel linear prediction scheme. This approach seems to be numerically very sensitive for large filter lengths. Note also that this latter approach slightly limits the application domain by assuming sources that can be modeled by AR processes, such as speech signals.

A fundamentally different alternative to the identification-and-inversion approach to blind dereverberation is the direct-inverse approach. Here, the aim is to directly estimate the inverse MIMO filter after Sect. 2 based on a dereverberation cost function. It is therefore inherently more robust to the order-overestimation problem. However, as we will see later in this chapter, this comes at the cost of the requirement for a more precise stochastic modeling of the source signals which again specializes the application domain, e.g., to speech signals. Moreover, the direct-inverse approach requires to take into account all microphone channels at once which renders the adaptation more complex.

---

source localization only the positions of the dominant components are required. Fortunately, this is in line with the requirement to avoid an overestimation of the filter length. Thus, in these applications the choice $L \leq L_{\text{opt,sep}}$ is preferable in practice.

Similar to the adaptation aspects of the identification-and-inversion approach in Sect. 5, we will treat the algorithmic aspects of the direct-inverse approach in Sect. 6. Both approaches are presented in a unified way based on TRINICON as outlined next in Sect. 4. The unified treatment also allows for an illuminating comparison.

4 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 nonstationarity, their spectral characteristics, and their probability densities.

In the existing literature, the known algorithms for blind system identification, blind source separation, and blind deconvolution were introduced independently. The BSS problem has mostly been addressed for instantaneous mixtures or by narrowband approaches in the frequency domain which adapt the coefficients independently in each DFT bin, e.g., [2, 34, 35]. 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. (independent identically distributed) model of the source signals (e.g., [13, 14]), which is undesirable for the generally nonwhite speech and audio signals as these should not be whitened, or are rather heuristically motivated, e.g., [15].

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 the signal properties of speech and audio signals (nonwhiteness, nonstationarity, and nongaussianity) into account, and allows a unified treatment of broadband BSS (as needed for a proper BSI) and MCBD algorithms as applicable to speech and audio signals in real acoustic environments [12, 18, 19, 20]. This framework generally uses multivariate stochastic signal models in the cost function to describe the temporal structure of the source signals and thereby provides a powerful cost function for both, BSS/BSI and MCBD, and, for the latter, also leads to improved algorithms for speech dereverberation.

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 [19, 20], we consider in this chapter mainly the time-domain formulation. Furthermore, we restrict ourselves here to gradient-based coefficient updates and disregard Newton-type adaptation algorithms for clarity and brevity. The algorithmic TRINICON framework is directly based on the matrix notation developed above.

Throughout this section, we regard the symmetric case 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 blind algorithms in the ICA literature, 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 TRINICON-based algorithm and only the condition $Q \leq P$ has to be fulfilled.

4.1 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 [20]:

$$\mathbf{y}^{\mathrm{T}}(n) = \mathbf{x}^{\mathrm{T}}(n)\mathbf{W}, \tag{31}$$

where n denotes the time index, and

$$\mathbf{x}^{\mathrm{T}}(n) = [\mathbf{x}_1^{\mathrm{T}}(n), \ldots, \mathbf{x}_P^{\mathrm{T}}(n)], \tag{32}$$

$$\mathbf{y}^{\mathrm{T}}(n) = [\mathbf{y}_1^{\mathrm{T}}(n), \ldots, \mathbf{y}_P^{\mathrm{T}}(n)], \tag{33}$$

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_{11} & \cdots & \mathbf{W}_{1P} \\ \vdots & \ddots & \vdots \\ \mathbf{W}_{P1} & \cdots & \mathbf{W}_{PP} \end{bmatrix}, \tag{34}$$

$$\mathbf{x}_p^{\mathrm{T}}(n) = [x_p(n), \ldots, x_p(n-2L+1)], \tag{35}$$

$$\mathbf{y}_q^{\mathrm{T}}(n) = [y_q(n), \ldots, y_q(n-D+1)] \tag{36}$$

$$= \sum_{p=1}^{P} \mathbf{x}_p^{\mathrm{T}}(n)\mathbf{W}_{pq}. \tag{37}$$

The parameter D in (36), $1 \leq D < L$, 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:

$$\mathbf{W}_{pq} = \begin{bmatrix} w_{pq,0} & 0 & \cdots & 0 \\ w_{pq,1} & w_{pq,0} & \ddots & \vdots \\ \vdots & w_{pq,1} & \ddots & 0 \\ w_{pq,L-1} & \vdots & \ddots & w_{pq,0} \\ 0 & w_{pq,L-1} & \ddots & w_{pq,1} \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & w_{pq,L-1} \\ 0 & \cdots & 0 & 0 \\ \vdots & & \vdots & \vdots \\ 0 & \cdots & 0 & 0 \end{bmatrix}. \tag{38}$$

Note that for $D = 1$, (31) simplifies to the well-known vector formulation of a convolution, as it is used extensively in the literature on supervised adaptive filtering, e.g., [1].

4.2 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 [2] which we all combine into an efficient and versatile algorithmic framework [18, 19, 12]:

(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 to solve the direct adaptive filtering problems after Table 1, such as source separation [2, 19] and system identification [8, 21]. To obtain an even more versatile estimator not only allowing spatial separation but also temporal separation for dereverberation and inverse adaptive filtering problems in general, we use the Kullback-Leibler divergence (KLD) [36] between a certain *desired* joint pdf (essentially representing a hypothesized stochastic source model) and the joint pdf of the actually estimated output signals [12]. Note that the mutual information is a special case of the KLD [36]. The desired pdf in the KLD is factorized w.r.t. the different sources (for the direct adaptive filtering problems, such as source separation) and possibly also w.r.t. certain temporal dependencies (for inverse adaptive filtering problems, such as dereverberation) as shown below. The KLD is guaranteed to be positive [36], 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):

$$\mathcal{J}(m,\mathbf{W}) = -\sum_{i=0}^{\infty}\beta(i,m)\frac{1}{N}\sum_{j=iN_L}^{iN_L+N-1}\left\{\log(\hat{p}_{s,PD}(\mathbf{y}(j))) - \log(\hat{p}_{y,PD}(\mathbf{y}(j)))\right\}, \tag{39}$$

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. In this chapter we assume that these pdfs are generally described by certain data-dependent parameterizations, so that we can write in more detail

$$\hat{p}_{s,PD} = \hat{p}_{s,PD}\left(\mathbf{y}, \boldsymbol{\mathcal{Q}}_s^{(1)}, \boldsymbol{\mathcal{Q}}_s^{(2)}, \ldots\right) \tag{40a}$$

and

$$\hat{p}_{y,PD} = \hat{p}_{y,PD}\left(\mathbf{y}, \boldsymbol{\mathcal{Q}}_y^{(1)}, \boldsymbol{\mathcal{Q}}_y^{(2)}, \ldots\right), \tag{40b}$$

respectively. We further assume that the model parameter estimates are given by the generic form

$$\boldsymbol{\mathcal{Q}}_s^{(r)}(i) = \frac{1}{N}\sum_{j=iN_L}^{iN_L+N-1}\left\{\boldsymbol{\mathcal{G}}_s^{(r)}(\mathbf{y}(j))\right\}, \quad r = 1,2,\ldots, \tag{41a}$$

$$\boldsymbol{\mathcal{Q}}_y^{(r)}(i) = \frac{1}{N}\sum_{j=iN_L}^{iN_L+N-1}\left\{\boldsymbol{\mathcal{G}}_y^{(r)}(\mathbf{y}(j))\right\}, \quad r = 1,2,\ldots, \tag{41b}$$

where $\boldsymbol{\mathcal{G}}_s^{(r)}$ and $\boldsymbol{\mathcal{G}}_y^{(r)}$ are suitable functions of the observation vectors $\mathbf{y}$, and $\boldsymbol{\mathcal{Q}}_s^{(r)}$ and $\boldsymbol{\mathcal{Q}}_y^{(r)}$ represent block-averages of $\boldsymbol{\mathcal{G}}_s^{(r)}(\mathbf{y})$ and $\boldsymbol{\mathcal{G}}_y^{(r)}(\mathbf{y})$, respectively. In general, the bold calligraphic symbols denote multidimensional arrays, or in other words, tensorial quantities. The elements of $\boldsymbol{\mathcal{Q}}_s^{(r)}$, $\boldsymbol{\mathcal{Q}}_y^{(r)}$, $\boldsymbol{\mathcal{G}}_s^{(r)}$, and $\boldsymbol{\mathcal{G}}_y^{(r)}$ are denoted by $\mathcal{Q}_{s,i_1,i_2,\ldots}^{(r)}$, $\mathcal{Q}_{y,i_1,i_2,\ldots}^{(r)}$, $\mathcal{G}_{s,i_1,i_2,\ldots}^{(r)}$, and $\mathcal{G}_{y,i_1,i_2,\ldots}^{(r)}$, respectively, where $i_1, i_2, \ldots$ are the indices in the corresponding tensor dimensions. Well-known special cases of such parameterizations are estimates of the variance $\hat{\sigma}_y^2(i) = \frac{1}{N}\sum_{j=iN_L}^{iN_L+N-1}\left\{y^2(j)\right\}$ and the correlation matrix $\mathbf{R}_{\mathbf{yy}}(i) = \frac{1}{N}\sum_{j=iN_L}^{iN_L+N-1}\left\{\mathbf{y}(j)\mathbf{y}^T(j)\right\}$ in the multivariate case $PD > 1$. The index m denotes the block time index for a block of N output samples shifted by N_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 [19, 37].

4.3 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

$$\check{\mathbf{W}}^{0}(m) := \check{\mathbf{W}}(m-1), \tag{42a}$$

$$\check{\mathbf{W}}^{\ell}(m) = \check{\mathbf{W}}^{\ell-1}(m) - \mu \Delta\check{\mathbf{W}}^{\ell}(m), \;\; \ell = 1, \ldots, \ell_{\max}, \tag{42b}$$

$$\check{\mathbf{W}}(m) := \check{\mathbf{W}}^{\ell_{\max}}(m), \tag{42c}$$

where μ is a stepsize parameter, and the superscript index ℓ denotes an iteration parameter to allow for multiple iterations ($\ell = 1, \ldots, \ell_{\max}$) within each block m. The $LP \times P$ coefficient matrix $\check{\mathbf{W}}$ (defined in (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 $\mathbf{W}$.

Obviously, when calculating the gradient of $\mathcal{J}(m, \mathbf{W})$ w.r.t. $\check{\mathbf{W}}$ explicitly, we are confronted with the problem of the different matrix formulations $\mathbf{W}$ and $\check{\mathbf{W}}$. The larger dimensions of $\mathbf{W}$ (see, e.g., (38)) are a direct consequence of taking into account the nonwhiteness signal property by choosing $D > 1$. As noted above, the rigorous distinction between these different matrix structures is an essential aspect of the general TRINICON framework and 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, formally introduced in [19, 20]. Using the Sylvester constraint operator the gradient descent update can be written as

$$\Delta\check{\mathbf{W}}^{\ell}(m) = \mathcal{SC}\left\{\nabla_{\mathbf{W}}\mathcal{J}(m, \mathbf{W})\right\}|_{\mathbf{W}=\mathbf{W}^{\ell}(m)}. \tag{43}$$

Depending on the particular realization of ($\mathcal{SC}$), we are able to select both, well-known and novel improved adaptation algorithms [37]. As discussed in [37] there are two particularly simple and popular realizations of ($\mathcal{SC}$) leading to two different classes of algorithms (see Fig. 7):

(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 [37]. However, in general, both choices require some tradeoff regarding algorithm performance. While $\mathcal{SC}_{\mathrm{C}}$ may provide a potentially more robust convergence behaviour, it will not work for arbitrary source positions, which is in contrast to the more versatile $\mathcal{SC}_{\mathrm{R}}$ [37]. Specifically, $\mathcal{SC}_{\mathrm{C}}$ allows to adapt only *causal* demixing systems. In geometrical terms this means that in the case of separating two sources using $\mathcal{SC}_{\mathrm{C}}$, they are required to be located in different half-planes w.r.t. the orientation of the microphone array [37].

For separating sources located in the same half-plane, or for more than two sources, noncausal demixing filters are required. With $\mathcal{SC}_{\mathrm{R}}$ it is possible to initialize $\check{\mathbf{W}}_{pp}$, $p = 1, \ldots, P$ with *shifted* unit impulses to allow noncausal filter taps [37]. Since acoustic scenarios exhibit nonminimum phase impulse responses, the need for noncausal demixing filters is further amplified in the dereverberation application.

In [8] an explicit formulation of a *generic* Sylvester constraint was derived to further formalize and clarify this concept, and to combine the versatility of $\mathcal{SC}_{\mathrm{R}}$ with the robust performance of $\mathcal{SC}_{\mathrm{C}}$ [38]. It turns out 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. 6.

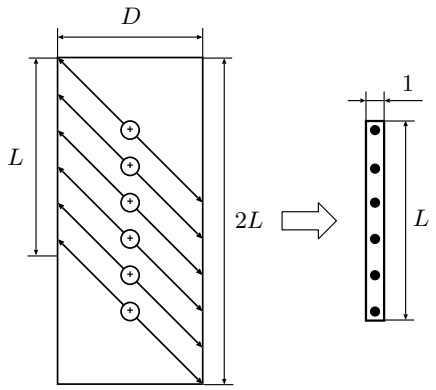


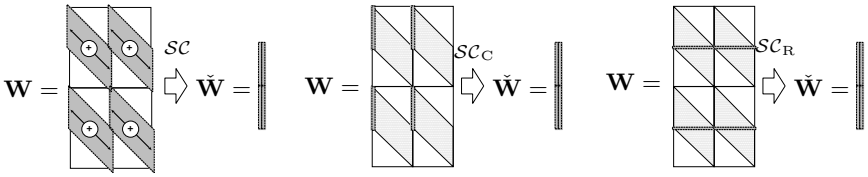
Fig. 6. Illustration of the generic Sylvester constraint ($\mathcal{SC}$) after [8] for one channel.

Note that the previously introduced approaches, classified by the choice of ($\mathcal{SC}_{\mathrm{C}}$) or ($\mathcal{SC}_{\mathrm{R}}$) as mentioned above, thus correspond to approximations of ($\mathcal{SC}$) by neglecting most of the elements within this averaging process, as illustrated in Fig. 7. In Sect. 6 of this chapter, we will see that by choosing the different Sylvester constraints, we are also able to establish relations to various known multichannel blind deconvolution algorithms from the literature.

It can be shown (see Appendix A) that by taking the gradient of $\mathcal{J}(m)$ with respect to the demixing filter matrix $\check{\mathbf{W}}(m)$ according to (43), we obtain the following generic gradient descent-based TRINICON update rule:

$$\Delta\check{\mathbf{W}}^{\ell}(m) = \frac{1}{N}\sum_{i=0}^{\infty}\beta(i,m)\,\mathcal{SC}\left\{\sum_{j=iN_L}^{iN_L+N-1}\mathbf{x}(j)\left[\boldsymbol{\Phi}^{\mathrm{T}}_{s,PD}(\mathbf{y}(j)) - \boldsymbol{\Phi}^{\mathrm{T}}_{y,PD}(\mathbf{y}(j))\right]\right\} \tag{44a}$$

with the *desired* generalized score function



(a) (b) (c)

Fig. 7. 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 $\mathcal{SC}_{\mathrm{R}}$.

$$\boldsymbol{\Phi}_{s,PD}(\mathbf{y}(j)) = -\frac{\partial \log \hat{p}_{s,PD}(\mathbf{y}(j))}{\partial \mathbf{y}(j)} - \frac{1}{N} \sum_{r} \sum_{i_1,i_2,\ldots} \frac{\partial \mathcal{G}^{(r)}_{s,i_1,i_2,\ldots}}{\partial \mathbf{y}^T} \sum_{j=iN_L}^{iN_L+N-1} \frac{\partial \hat{p}_{s,PD}}{\partial \mathcal{Q}^{(r)}_{s,i_1,i_2,\ldots}} \quad (44b)$$

resulting from the hypothesized source model $\hat{p}_{s,PD}$, and the actual generalized score function

$$\boldsymbol{\Phi}_{y,PD}(\mathbf{y}(j)) = -\frac{\partial \log \hat{p}_{y,PD}(\mathbf{y}(j))}{\partial \mathbf{y}(j)} - \frac{1}{N} \sum_{r} \sum_{i_1,i_2,\ldots} \frac{\partial \mathcal{G}^{(r)}_{y,i_1,i_2,\ldots}}{\partial \mathbf{y}^T} \sum_{j=iN_L}^{iN_L+N-1} \frac{\partial \hat{p}_{y,PD}}{\partial \mathcal{Q}^{(r)}_{y,i_1,i_2,\ldots}}, \quad (44c)$$

where the stochastic model parameters are given by (41), and $\mathcal{G}^{(r)}_{s,i_1,i_2,\ldots}$, $\mathcal{G}^{(r)}_{y,i_1,i_2,\ldots}$, $\mathcal{Q}^{(r)}_{s,i_1,i_2,\ldots}$, and $\mathcal{Q}^{(r)}_{y,i_1,i_2,\ldots}$ are the elements of $\boldsymbol{\mathcal{G}}^{(r)}_{s}$, $\boldsymbol{\mathcal{G}}^{(r)}_{y}$, $\boldsymbol{\mathcal{Q}}^{(r)}_{s}$, and $\boldsymbol{\mathcal{Q}}^{(r)}_{y}$, respectively, as explained below (41). The form of the coefficient update (44a) with the generalized score functions (44b) and (44c) also fits well into the theory of so-called estimating functions [39].

The hypothesized source model $\hat{p}_{s,PD}(\cdot)$ in (44b) is chosen according to the class of signal processing problem to be solved (see Table 1). For instance, a factorization of $\hat{p}_{s,PD}(\cdot)$ among the sources yields BSS (or BSI via the scheme described in Sect. 3.4), i.e.,

$$\hat{p}_{s,PD}(\mathbf{y}(j)) \overset{(\mathrm{BSS})}{=} \prod_{q=1}^{P} \hat{p}_{y_q,D}(\mathbf{y}_q(j)), \quad (45a)$$

while a complete factorization leads to the traditional MCBD approach,

$$\hat{p}_{s,PD}(\mathbf{y}(j)) \overset{(\mathrm{MCBD})}{=} \prod_{q=1}^{P} \prod_{d=1}^{D} \hat{p}_{y_q,1}(y_q(j-d+1)). \quad (45b)$$

Additionally, in Sect. 6 we will introduce another, more general class, called the multichannel blind partial deconvolution (MCBPD) approach.

Alternative formulation of the gradient-based coefficient update

Both for practical realizations and also for some theoretical considerations, an equivalent reformulation of the gradient-based update (44a) is often useful. This alternative formulation is obtained by transforming the output signal pdf $\hat{p}_{y,PD}(\mathbf{y})$ in the cost function into the PD-dimensional input signal pdf using $\mathbf{W}$ as a mapping matrix for this linear transformation. The relation (134) in Appendix B shows this pdf transformation. (Note that the result of Appendix B is needed again later in this chapter). Gradient calculation as above leads to the alternative formulation of the gradient-based update,

$$\Delta\check{\mathbf{W}}^{\ell}(m) = \frac{1}{N}\sum_{i=0}^{\infty}\beta(i,m)\,\mathcal{SC}\left\{\sum_{j=iN_L}^{iN_L+N-1}\left[\mathbf{x}(j)\boldsymbol{\Phi}_{s,PD}^{\mathrm{T}}(\mathbf{y}(j)) - \mathbf{V}\left(\left(\mathbf{W}^{\ell-1}(m)\right)^{\mathrm{T}}\mathbf{V}\right)^{-1}\right]\right\}, \tag{46a}$$

with the window matrix

$$\mathbf{V} = \mathrm{Bdiag}\{\tilde{\mathbf{V}},\ldots,\tilde{\mathbf{V}}\}, \tag{46b}$$

$$\tilde{\mathbf{V}} = \left[\mathbf{I}_{D\times D},\ \mathbf{0}_{D\times(2L-D)}\right]^{\mathrm{T}}. \tag{46c}$$

4.4 Natural Gradient-Based Coefficient Update

It is known that stochastic gradient descent generally suffers from slow convergence in many practical problems due to statistical dependencies in the data being processed. A modification of the ordinary gradient which is especially popular in the field of ICA and BSS due to its computational efficiency, is the so-called *natural gradient* [2]. It can be shown that by taking the natural gradient of $\mathcal{J}(m)$ with respect to the demixing filter matrix $\mathbf{W}(m)$ [20],

$$\Delta\check{\mathbf{W}} \propto \mathcal{SC}\left\{\mathbf{W}\mathbf{W}^{\mathrm{T}}\frac{\partial\mathcal{J}}{\partial\mathbf{W}}\right\}, \tag{47}$$

we obtain the following generic TRINICON-based update rule:

$$\Delta\check{\mathbf{W}}^{\ell}(m) = \frac{1}{N}\sum_{i=0}^{\infty}\beta(i,m)\,\mathcal{SC}\left\{\sum_{j=iN_L}^{iN_L+N-1}\mathbf{W}^{\ell}(i)\mathbf{y}(j)\left[\boldsymbol{\Phi}_{s,PD}^{\mathrm{T}}(\mathbf{y}(j)) - \boldsymbol{\Phi}_{y,PD}^{\mathrm{T}}(\mathbf{y}(j))\right]\right\}. \tag{48}$$

Moreover, from (46a) we obtain an alternative formulation of (48):

$$\Delta\check{\mathbf{W}}^{\ell}(m) = \sum_{i=0}^{\infty}\beta(i,m)\,\mathcal{SC}\left\{\mathbf{W}^{\ell}(i)\left[\frac{1}{N}\sum_{j=iN_L}^{iN_L+N-1}\mathbf{y}(j)\boldsymbol{\Phi}_{s,PD}^{\mathrm{T}}(\mathbf{y}(j)) - \mathbf{I}\right]\right\}, \tag{49}$$

which exhibits an especially simple – and thus computationally efficient – structure. An important feature of this natural gradient update is that its adaptation performance is largely independent of the conditioning of the acoustic mixing system matrix [20].

4.5 Incorporation of Stochastic Source Models

The general update equations (42) with (44), (46), (48), (49) offer the possibility to account for all available information on the statistical properties of the desired source signals. To apply this general approach in a real-world scenario, appropriate multivariate score functions $\boldsymbol{\Phi}^{\mathrm{T}}_{s,PD}(\mathbf{y})$ (and $\boldsymbol{\Phi}^{\mathrm{T}}_{y,PD}(\mathbf{y})$ where required) in the update equations have to be determined, based on appropriate multivarate stochastic signal models.

The selection of the stochastic signal models is based on several different considerations. As already illustrated by (45a) and (45b), the design of the signal model is instrumental in defining the class of the adaptive filtering problem according to Tab. 1. This aspect will be detailed in Sect. 5 and 6. Another important aspect is that many of the different adaptation techniques in the literature represent different approximations of the probability density functions.

For estimating pdfs a distinction between parametric and non-parametric techniques is common (see, e.g., [40]).

A *parametric* technique defines a family of density functions in terms of a set of parameters as in (40a) and (40b). The parameters are then optimized so that the density function corresponds to the observed samples. In the context of ICA different parametric representations have been used. Examples include Gaussian models in the simplest case, Gaussian mixture models, and generalized Gaussian. The important class of spherically-invariant random processes, as detailed below, may also be understood as a parametric approach. Other parametric techniques are based on higher moments [41], e.g., Gram-Charlier expansion, Parson densities, or on higher cumulants [41], e.g., the Edgeworth expansion. As an important representative of these techniques, we consider the Gram-Charlier expansion for TRINICON, as detailed below.

The *non-parametric* techniques usually define the estimated density directly in terms of the observed samples. The best known non-parametric estimate is the histogram, which is very data intensive. Somewhat less data is required by the Parzen-Windows method [40]. Note that sometimes also the above-mentioned techniques based on series with higher moments are classified as non-parametric in the literature [41]. Obviously, the incorporation of various assumptions about the densities by truncating these series expansions in practice provides a smooth transition to powerful parametric techniques which require less data than the simpler non-parametric techniques.

Another important aspect in the choice of stochastic models is their robustness. According to [42], robustness denotes insensitivity to a certain amount of deviations from the statistical modeling assumptions due to some fraction

of outliers with some arbitrary probability density. Unfortunately, many of the traditional estimation techniques, such as least-squares estimation, or the higher-order techniques mentioned above turn out to be fairly sensitive in this sense. The theory of *robust statistics* [42] provides a systematic framework to robustify the various techniques and it has been very successfully applied to adaptive filtering, e.g., [43]. In [21] the theory of multivariate robust statistics was introduced in TRINICON. Although in this chapter we will not consider the robustness extensions in detail, it is important to note that they fit well into the general class of spherically-invariant random processes detailed below.

Finally, it should be noted that in addition to the model selection the choice of estimation procedure for the corresponding *stochastic model parameters* (e.g., correlation matrices in (50) below, higher-order moments, scaling parameter for robust statistics in [21], etc.), in other words, the practical realization of (41), is another important design consideration. The estimation of the stochastic model parameters and the TRINICON-based updates of the adaptive filter coefficients are performed in an alternating way.

Similar to the estimation of correlation matrices in linear prediction problems [68] we have to distinguish in actual implementations between the more accurate so-called *covariance method* and the approximative *correlation method* leading to a lower complexity, e.g., [37]. As we will see later in this chapter, based on these different estimation methods for the correlation matrices and on the above-mentioned approximations $\mathcal{SC}_{\mathrm{R}}\{\cdot\}$ and $\mathcal{SC}_{\mathrm{C}}\{\cdot\}$ of the Sylvester constraint $\mathcal{SC}\{\cdot\}$ we can establish an illustrative classification scheme for BSI and deconvolution algorithms.

Spherically Invariant Random Processes as Signal Model

An efficient and fairly general solution to the problem of determining the high-dimensional score functions in broadband adaptive MIMO filtering is to assume so-called *spherically invariant random processes* (SIRPs), e.g., [44, 45, 46], as proposed in [18, 19]. The general form of correlated SIRPs of D-th order is given with a properly chosen function $f_{p,D}(\cdot)$ for the p-th output channel of the MIMO system by

$$\hat{p}_{y_p,D}(\mathbf{y}_p(j)) = \frac{1}{\sqrt{\pi^D \det(\mathbf{R}_{\mathbf{y}_p\mathbf{y}_p}(i))}} f_{p,D}\left(\mathbf{y}_p^T(j)\mathbf{R}^{-1}_{\mathbf{y}_p\mathbf{y}_p}(i)\mathbf{y}_p(j)\right), \qquad (50)$$

where $\mathbf{R}_{\mathbf{y}_p\mathbf{y}_p}$ denotes the corresponding $D \times D$ autocorrelation 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 [46]. A major 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 $f_{p,D}(\cdot)$ 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, e.g., (44b) can be derived from (50), as detailed in [18, 19].

To calculate the score function for SIRPs in general, we employ the chain rule to (50) so that the first term in (44b) reads

$$-\frac{\partial \log \hat{p}_{y_p,D}(\mathbf{y}_p)}{\partial \mathbf{y}_p} = -\frac{\frac{\partial \hat{p}_{y_p,D}(\mathbf{y}_p)}{\partial \mathbf{y}_p}}{\hat{p}_{y_p,D}(\mathbf{y}_p)} = \underbrace{\left[-\frac{1}{f_{p,D}(u_p)}\frac{\partial f_{p,D}(u_p)}{\partial u_p}\right]}_{:=\phi_{y_p,D}(u_p)} \mathbf{R}^{-1}_{\mathbf{y}_p\mathbf{y}_p}(i)\mathbf{y}_p(j), \tag{51}$$

where $u_p = \mathbf{y}_p^T \mathbf{R}^{-1}_{\mathbf{y}_p\mathbf{y}_p}\mathbf{y}_p$. For convenience, we call the scalar function $\phi_{y_p,D}(u_p)$ the *SIRP score*. It can be shown (after a somewhat tedious but straightforward derivation) that for SIRPs in general, the second term in (44b) is equal to zero so that the general score function is given by the simple expression (51). A great advantage of SIRPs is that the required function $f_D(u)$ can actually be derived analytically from the corresponding *univariate* pdf [46]. As a practical important example, following the procedure in [46], we obtain, e.g., as the *optimum SIRP score for univariate Laplacian pdfs* [18]:

$$\phi_{y_q,D}(u_q) = -\frac{1}{D - \sqrt{2u_q}\frac{K_{D/2+1}(\sqrt{2u_q})}{K_{D/2}(\sqrt{2u_q})}}, \tag{52}$$

where $K_\nu(\cdot)$ denotes the ν-th order modified Bessel function of the second kind.

Multivariate Gaussians as Signal Model: Second-Order Statistics

To see the link to adaptation algorithms that are based purely on second-order statistics (SOS), we use the model of *multivariate Gaussian* pdfs

$$\hat{p}_{y_p,D}(\mathbf{y}_p(j)) = \frac{1}{\sqrt{(2\pi)^D \det \mathbf{R}_{\mathbf{y}_p\mathbf{y}_p}(i)}} \mathrm{e}^{-\frac{1}{2}\mathbf{y}_p^T(j)\mathbf{R}^{-1}_{\mathbf{y}_p\mathbf{y}_p}(i)\mathbf{y}_p(j)} \tag{53}$$

as a special case of a SIRP with $f_{q,D}(u_q) = \frac{1}{\sqrt{2^D}} \exp(-\frac{1}{2}u_q)$. Hence, the score function for the generic SOS case is obtained straightforwardly from (51) for the constant SIRP score $\phi_{y_p,D}(u_p) = 1/2$, and it can be shown that most of the popular SOS-based adaptation algorithms represent special cases of the corresponding algorithms based on SIRPs, e.g., [12, 18, 19, 21]. Moreover, by transforming the model into the DFT domain, this relation also carries over to various links to novel and existing popular frequency-domain algorithms [8, 19].

It is interesting to note that the generic SOS-based update was originally obtained independently in [20] (first for the BSS application) as a generalization of the cost function of [47]:

$$\mathcal{J}_{\mathrm{SOS}}\left(m,\mathbf{W}\right)=\sum_{i=0}^{\infty}\beta(i,m)\left\{\log\det\mathbf{R}_{\mathbf{ss}}(i)-\log\det\mathbf{R}_{\mathbf{yy}}(i)\right\}. \tag{54}$$

This cost function can be interpreted as a distance measure between the actual time-varying output-correlation matrix $\mathbf{R}_{\mathbf{yy}}$ and a certain *desired* output-correlation matrix $\mathbf{R}_{\mathbf{ss}}$.

Nearly Gaussian Densities as Signal Model

Two different expansions are commonly used to obtain a parameterized representation of probability density functions which only slightly deviate from the Gaussian density (often called *nearly Gaussian densities*): the Edgeworth and the Gram-Charlier expansions, e.g., [2]. They lead to very similar approximations, so we only consider here the Gram-Charlier expansion. As explained in Appendix C, these expansions are based on the so-called Chebyshev-Hermite polynomials $P_{\mathrm{H},n}(\cdot)$.

We first illustrate the idea in the univariate case. A fourth-order expansion of a univariate, zero-mean, and nearly Gaussian pdf is given in (140) in Appendix C with the estimates of *skewness* $\hat{\kappa}_3=\hat{E}\left\{y^3\right\}$ and the *kurtosis* $\hat{\kappa}_4=\hat{E}\left\{y^4\right\}-3\hat{\sigma}^4$, the latter one being the most important higher-order statistical quantity in our context. Generally, speech signals exhibit supergaussian densities whose third-order cumulants are negligible compared to its fourth-order cumulants. Therefore, we are particularly interested in the approximation

$$\hat{p}(y)\approx\frac{1}{\sqrt{2\pi}\hat{\sigma}}\mathrm{e}^{-\frac{y^2}{2\hat{\sigma}^2}}\left(1+\frac{\hat{\kappa}_4}{4!\,\hat{\sigma}^4}P_{\mathrm{H},4}\left(\frac{y}{\hat{\sigma}}\right)\right). \tag{55}$$

Similar to the specialization (54) of the TRINICON optimization criterion for the case of SOS, the Gram-Charlier-based model also allows for an interesting illustration of the criterion. By exploiting the near-gaussianity by the approximation $\log(1+\epsilon)\approx\epsilon$ for $\log\left(1+\frac{\hat{\kappa}_4}{4!\,\hat{\sigma}^4}P_{\mathrm{H},4}\left(\frac{y}{\hat{\sigma}}\right)\right)$ in the logarithmized respresentation of (55), and noting that $P_{\mathrm{H},4}\left(\frac{y}{\hat{\sigma}}\right)=\left(\frac{y}{\hat{\sigma}}\right)^4-6\left(\frac{y}{\hat{\sigma}}\right)^2+3$ we can develop the following expression appearing in the TRINICON criterion (39):

$$\begin{aligned}
&\frac{1}{N}\sum_{j=iN_L}^{iN_L+N-1}\log\hat{p}(y)\\
&\qquad\approx\frac{1}{N}\left(\sum_{j=iN_L}^{iN_L+N-1}\log\frac{1}{\sqrt{2\pi}\hat{\sigma}}\mathrm{e}^{-\frac{y^2}{2\hat{\sigma}^2}}\right)+\frac{1}{N}\left(\sum_{j=iN_L}^{iN_L+N-1}\frac{\hat{\kappa}_4}{4!\,\hat{\sigma}^4}P_{\mathrm{H},4}\left(\frac{y}{\hat{\sigma}}\right)\right)\\
&\qquad=\frac{1}{N}\left(\sum_{j=iN_L}^{iN_L+N-1}\log\frac{1}{\sqrt{2\pi}\hat{\sigma}}\mathrm{e}^{-\frac{y^2}{2\hat{\sigma}^2}}\right)+\frac{\hat{\kappa}_4^2}{4!\,(\hat{\sigma}^2)^4},
\end{aligned} \tag{56}$$

where $\hat{\kappa}_4 = \frac{1}{N}\sum_{j=iN_L}^{iN_L+N-1} y^4 - 3\hat{\sigma}^4$ represents an estimate for the kurtosis based on block averaging. As we can see, in addition to the SOS, the optimization is directly based on the normalized kurtosis, which is a widely-used measure of *nongaussianity*. This additive representation will play a particularly important role in the application to the direct-inverse approach to blind dereverberation in Sect. 6.

To obtain general coefficient update rules based on this representation, we finally consider the multivariate formulation of the Gram-Charlier expansion after (146a) in Appendix C. To calculate the multivariate Chebyshev-Hermite polynomials, we apply the relation

$$P_{\mathrm{H},\mathbf{n}}(\mathbf{y}_p) = \prod_{d=1}^{D} P_{\mathrm{H},n_d}(y_{d,p}) \tag{57}$$

after (144) so that

$$\hat{p}_{y_p,D}(\mathbf{y}_p(j)) = \frac{1}{\sqrt{(2\pi)^D \det \mathbf{R}_{\mathbf{y}_p\mathbf{y}_p}(i)}} \mathrm{e}^{-\frac{1}{2}\mathbf{y}_p^T(j)\mathbf{R}^{-1}_{\mathbf{y}_p\mathbf{y}_p}(i)\mathbf{y}_p(j)}$$

$$\cdot \sum_{n_1=0}^{\infty} \cdots \sum_{n_D=0}^{\infty} a_{n_1\cdots n_D,p}\; P_{\mathrm{H},n_1}\left(\left[\mathbf{L}_p^{-1}(i)\mathbf{y}_p(j)\right]_1\right) \cdot \cdots \cdot P_{\mathrm{H},n_D}\left(\left[\mathbf{L}_p^{-1}(i)\mathbf{y}_p(j)\right]_D\right)$$

with the coefficients according to (146b),

$$a_{n_1\cdots n_D,p} = \frac{\hat{E}\left\{P_{\mathrm{H},n_1}\left(\left[\mathbf{L}_p^{-1}(i)\mathbf{y}_p(j)\right]_1\right) \cdot \cdots \cdot P_{\mathrm{H},n_D}\left(\left[\mathbf{L}_p^{-1}(i)\mathbf{y}_p(j)\right]_D\right)\right\}}{n_1! \cdot \cdots \cdot n_D!}. \tag{58}$$

Multivariate generalizations of the skewness and the kurtosis were introduced by Mardia in [48]. In our context the corresponding multivariate generalization of the kurtosis can be written as

$$\hat{\kappa}^{(D)}_{4,\mathrm{norm}} = \hat{E}\left\{\left[\mathbf{y}_p^T(j)\mathbf{R}^{-1}_{\mathbf{y}_p\mathbf{y}_p}(i)\mathbf{y}_p(j)\right]^2\right\} - D(D+2). \tag{59}$$

Similar to the univariate case, this quantity can be related to our formulation of the multivariate probability density. Note that for $D = 1$ it corresponds to the traditional normalized kurtosis $\hat{\kappa}_4/\hat{\sigma}^4 = \hat{E}\{y_p^4\}/\hat{\sigma}^4 - 3$, as it appears in, e.g., (55).

In this chapter, we further consider an important special case of this general multivariate model, which is particularly useful for speech processing. In this case, the inverse covariance matrix $\mathbf{R}^{-1}_{\mathbf{y}_p\mathbf{y}_p} = (\mathbf{L}_p^T\mathbf{L}_p)^{-1}$ is first factorized as [49]

$$\mathbf{R}^{-1}_{\mathbf{y}_p\mathbf{y}_p}(i) = \mathbf{A}_p(i)\boldsymbol{\Sigma}^{-1}_{\tilde{\mathbf{y}}_p\tilde{\mathbf{y}}_p}(i)\mathbf{A}_p^T(i), \tag{60}$$

where $\mathbf{A}_p(i)$ and $\boldsymbol{\Sigma}_{\tilde{\mathbf{y}}_p\tilde{\mathbf{y}}_p}(i)$ denote a $D\times D$ unit lower triangular matrix (i.e., its elements on the main diagonal are equal to 1) and a diagonal matrix, respectively [49]. The $D \times D$ unit lower triangular matrix $\mathbf{A}_p(i)$ can be interpreted

as a (time-varying) convolution matrix of a whitening filter. It is therefore convenient for computational reasons to model the signal y_p as an autoregressive (AR) process of order $n_A = D-1$, with time-varying AR coefficients $a_{p,k}(n)$, and residual signal $\tilde{y}_p(n)$, i.e.,

$$y_p(n) = -\sum_{k=1}^{D-1} a_{p,k}(n) y_p(n-k) + \tilde{y}_p(n). \tag{61}$$

The matrices $\mathbf{A}_p$ and $\boldsymbol{\Sigma}_{\tilde{\mathbf{y}}_p\tilde{\mathbf{y}}_p}$ can then be written as

$$\mathbf{A}_p = \begin{bmatrix} 1 & a_{p,1}(n) & a_{p,2}(n) & \cdots & \cdots & \cdots & \cdots & a_{p,D-1}(n) \\ 0 & 1 & a_{p,1}(n-1) & \cdots & \cdots & \cdots & \cdots & a_{p,D-2}(n-1) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 1 \end{bmatrix}^T \tag{62}$$

and

$$\begin{aligned} \boldsymbol{\Sigma}_{\tilde{\mathbf{y}}_p\tilde{\mathbf{y}}_p} &= \mathrm{Diag}\left\{\hat{\sigma}^2_{\tilde{y}_p}(n), \ldots, \hat{\sigma}^2_{\tilde{y}_p}(n-D+1)\right\} \\ &= \hat{E}\left\{ \begin{bmatrix} \tilde{y}_p(n) \\ \vdots \\ \tilde{y}_p(n-D+1) \end{bmatrix} [\tilde{y}_p(n), \ldots, \tilde{y}_p(n-D+1)] \right\}. \end{aligned} \tag{63}$$

Now, the multivariate stochastic signal model can be rewritten by shifting the *prefiltering matrix* $\mathbf{A}_p$ into the data terms, i.e.,

$$\tilde{\mathbf{y}}_p := \mathbf{A}_p^T \mathbf{y}_p = [\tilde{y}_p(n), \tilde{y}_p(n-1), \ldots, \tilde{y}_p(n-D+1)]^T. \tag{64}$$

Moreover, by assuming the whitened elements of vector $\tilde{\mathbf{y}}_p$ to be i.i.d. (which in practice is a widely used assumption in AR modeling), so that the expansion coefficients $a_{n_1 \cdots n_D, p}$ are factorized, we obtain thanks to (57) with $\mathbf{L}_p(i) = \mathrm{Diag}\left\{\frac{1}{\hat{\sigma}_{\tilde{y}_p}(j)}, \ldots, \frac{1}{\hat{\sigma}_{\tilde{y}_p}(j-D+1)}\right\} \mathbf{A}^T(i)$ and (64) the following model representation:

$$\begin{aligned} \hat{p}_{y_p,D}(\mathbf{y}_p(j)) = \prod_{d=1}^{D} \frac{1}{\sqrt{2\pi\ \hat{\sigma}^2_{\tilde{y}_p}(j-d+1)}} \mathrm{e}^{-\frac{\tilde{y}_p^2(j-d+1)}{2\hat{\sigma}^2_{\tilde{y}_p}(j-d+1)}} \\ \cdot \sum_{n_d=0}^{\infty} \frac{\hat{E}\left\{P_{\mathrm{H},n_d}\left(\frac{\tilde{y}_p(j-d+1)}{\hat{\sigma}_{\tilde{y}_p}(j-d+1)}\right)\right\}}{n_d!}\ P_{\mathrm{H},n_d}\left(\frac{\tilde{y}_p(j-d+1)}{\hat{\sigma}_{\tilde{y}_p}(j-d+1)}\right). \end{aligned}$$

By considering only the fourth-order term in addition to SOS again, i.e.,

$$\begin{aligned} \hat{p}_{y_p,D}(\mathbf{y}_p(j)) = \prod_{d=1}^{D} \frac{1}{\sqrt{2\pi\ \hat{\sigma}^2_{\tilde{y}_p}(j-d+1)}} \mathrm{e}^{-\frac{\tilde{y}_p^2(j-d+1)}{2\hat{\sigma}^2_{\tilde{y}_p}(j-d+1)}} \\ \cdot \left(1 + \frac{\hat{\kappa}_{4,\tilde{y}_p}}{4!\sigma^4_{\tilde{y}_p}(j-d+1)}\ P_{\mathrm{H},n_d}\left(\frac{\tilde{y}_p(j-d+1)}{\hat{\sigma}_{\tilde{y}_p}(j-d+1)}\right)\right), \end{aligned}$$

and by exploiting the near-gaussianity using the approximation $\log(1+\epsilon) \approx \epsilon$, we obtain after a straightforward calculation the following expression for the score function (44c):

$$\begin{aligned}\boldsymbol{\Phi}_{y,PD}(\mathbf{y}(j)) &= \\ &= \mathbf{A}(i) \left[\frac{\tilde{y}_p(j-d+1)}{2\hat{\sigma}^2_{\tilde{y}_p}(j-d+1)} - \left(\frac{\sum_{j=iN_L}^{iN_L+N-1} \tilde{y}_p^4(j-d+1)}{3\left(\sum_{j=iN_L}^{iN_L+N-1} \tilde{y}_p^2(j-d+1)\right)^2} - 1 \right) \right. \\ &\quad \left. \cdot \left(\frac{\tilde{y}_p^3(j-d+1)}{\hat{\sigma}^4_{\tilde{y}_p}(j-d+1)} - \frac{\tilde{y}_p(j-d+1)\sum_{j=iN_L}^{iN_L+N-1} \tilde{y}_p^4(j-d+1)}{\hat{\sigma}^6_{\tilde{y}_p}(j-d+1)} \right) \right], \end{aligned} \tag{65}$$

where the expression in brackets denotes a column vector composed of the elements for $d = 1, \ldots, D$ and $p = 1, \ldots, P$, and $\mathbf{A}(i) = [\mathbf{A}_1(i), \ldots, \mathbf{A}_P(i)]$ after (62). Note that the first term corresponds to the SOS as in (51), while the second term is related to the multivariate normalized kurtosis. This expression will play an important role in Sect. 6.

5 Application of TRINICON to Blind System Identification and the Identification-and-Inversion Approach to Blind Deconvolution

In Sect. 3 we developed the identification-and-inversion approach to blind deconvolution from a system-theoretic point of view. We have seen that in the general MIMO case its practical (i.e., adaptive) realization can be traced back to the problem of blind source separation for convolutive mixtures with appropriately chosen filter length L and subsequent inversion, e.g., using MINT (Fig. 5). Both signal separation and system identification belong to the class of direct adaptive filtering problems according to Tab. 1. On the other hand, it was shown that in the SIMO case this approach leads to a well-known class of realizations for which the AED algorithm in its various versions is known from the literature. Hence, as the two main aspects in this section

- we discuss the specialization of the TRINICON framework to practical algorithms that are suitable for adaptive MIMO BSI. Various different BSS algorithms have been proposed in recent years (e.g., [50]), and many of them can be related to TRINICON [8, 19]. However, of special importance for BSI and the identification-and-inversion approach to dereverberation are efficient realizations of *broadband* BSS algorithms.
- we develop the relation to the SIMO case explicitly from an algorithmic point of view. This will lead to various new insights and also to some generalizations of the AED.

Both of these main aspects will also serve as important starting points for the developments in Sect. 6. An experimental comparison of the identification-and-inversion approach with the direct-inverse approach to blind dereverberation also follows in Sect. 6.

5.1 Generic Gradient-Based Algorithm for Direct Adaptive Filtering Problems

To begin with, we specialize TRINICON to the case of direct adaptive filtering problems, i.e., signal separation and system identification. Again, for simplicity of the presentation, we concentrate here on iterative Euclidean gradient-based and natural gradient-based block-online coefficient updates. As mentioned in Sect. 4, the class of signal separation and system identification algorithms is specified by the factorization of the hypothesized source model $\hat{p}_{s,PD}(\cdot)$ among the sources according to (45a). The desired multivariate score function then becomes the partitioned vector

$$\boldsymbol{\Phi}_{s,PD}(\mathbf{y}(j)) = \left[\boldsymbol{\Phi}^{\mathrm{T}}_{y_1,D}(\mathbf{y}_1(j)), \ldots, \boldsymbol{\Phi}^{\mathrm{T}}_{y_P,D}(\mathbf{y}_P(j))\right]^{\mathrm{T}}, \tag{66a}$$

$$\boldsymbol{\Phi}_{y_p,D}(\mathbf{y}_p(j)) = -\frac{\partial \log \hat{p}_{y_p,D}(\mathbf{y}_p(j))}{\partial \mathbf{y}_p(j)}. \tag{66b}$$

The corresponding generic coefficient update rules are then directly given by (44a), (46a), (48), and (49).

In this section, our considerations are based on the SIRP model (including SOS as a special case). Accordingly, each partition of the vector (66a) is given by (51). The resulting general class of broadband BSS algorithms was first presented in [18] and has led to various efficient realizations so far (see Sect. 5.3). The idea of using a SIRP model was also adopted, e.g., in the approximate DFT-domain realizations [51, 52].

Illustration for Second-Order Statistics

By setting the SIRP scores $\phi_{y_p,D}(\cdot) = 1/2$, $p = 1, \ldots, P$, we obtain the particularly illustrative case of SOS-based adaptation algorithms. Here, the source models are simplified to multivariate Gaussian functions described by $PD \times PD$ correlation matrices $\mathbf{R}_{..}$ estimated from the length-N signal blocks, so that the update rules (44a) and (48) lead to [12]

$$\Delta\check{\mathbf{W}}(m) = \sum_{i=0}^{\infty} \beta(i,m)\, \mathcal{SC}\left\{\mathbf{R}_{\mathbf{xy}}(i)\left[\mathbf{R}^{-1}_{\mathbf{ss}}(i) - \mathbf{R}^{-1}_{\mathbf{yy}}(i)\right]\right\} \tag{67}$$

and

$$\Delta\check{\mathbf{W}}(m) = \sum_{i=0}^{\infty} \beta(i,m)\, \mathcal{SC}\left\{\mathbf{W}(i)\mathbf{R}_{\mathbf{yy}}(i)\left[\mathbf{R}_{\mathbf{ss}}^{-1}(i) - \mathbf{R}_{\mathbf{yy}}^{-1}(i)\right]\right\}$$
$$= \sum_{i=0}^{\infty} \beta(i,m)\, \mathcal{SC}\left\{\mathbf{W}(i)\left[\mathbf{R}_{\mathbf{yy}}(i) - \mathbf{R}_{\mathbf{ss}}(i)\right]\mathbf{R}_{\mathbf{ss}}^{-1}(i)\right\}, \qquad (68)$$

respectively. The BSS versions of these generic SOS natural gradient updates follow immediately by setting

$$\mathbf{R}_{\mathbf{ss}}(i) = \mathrm{bdiag}\, \mathbf{R}_{\mathbf{yy}}(i). \qquad (69)$$

The update (68) together with (69) was originally obtained independently in [20] from the cost function (54). In Fig. 8 the mechanism of (68) based on the model (69) 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.

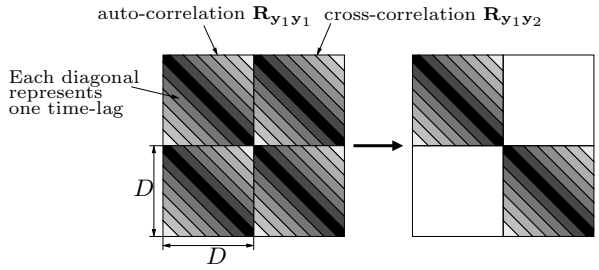


Fig. 8. Illustration of SOS-based broadband BSS.

A very important feature of the TRINICON-based coefficient updates is the inherent normalization by the auto-correlation matrices, reflected by the inverse of $\mathbf{R}_{\mathbf{ss}}(i) = \mathrm{bdiag}\, \mathbf{R}_{\mathbf{yy}}(i)$ in (68). As we will see in Sect. 5.2, this normalization can in fact be interpreted as an *adaptive stepsize control*. In fact, as was shown in [19], the update equations of another very popular subclass of second-order BSS algorithms, based on a cost function using the Frobenius norm[9] $\|\mathbf{A}\|_{\mathrm{F}}^2 = \sum_{i,j} a_{ij}^2$ of a matrix $\mathbf{A} = (a_{ij})$, e.g., [2],[53]-[57], differ from the more general TRINICON-based updates mainly in the inherent normalization. The gradient-based update resulting from the Frobenius norm can be regarded as an analogon to the traditional least mean square (LMS) algorithm [1] in supervised adaptive filtering without stepsize control. Indeed, many simulation results have shown that for large filter lengths L, this Frobenius-based updates are prone to instability, while the properly normalized updates show

[9] Analogously to the TRINICON-based $\mathcal{J}_{\mathrm{SOS}}$ this approach may be generalized for convolutive mixtures to $\mathcal{J}_{\mathrm{F}}(m) = \sum_{i=0}^{\infty} \beta(i,m)\, \|\mathbf{R}_{\mathbf{yy}}(i) - \mathrm{bdiag}\, \mathbf{R}_{\mathbf{yy}}(i)\|_{\mathrm{F}}^2$.

a very robust convergence behaviour even for hundreds or thousands of filter coefficients for the application in real acoustic environments, e.g., [20]. As we will see in Sect. 6, an analogous consideration concerning the inherent normalization is also possible for dereverberation algorithms of the direct-inverse-type.

The realization of this normalization is also an important aspect in various efficient approximations of the generic broadband algorithms, e.g., [37, 58, 59], with a reduced computational complexity for real-time operation. Moreover, a close link has been established [19, 20] to various popular frequency-domain algorithms, as we discuss in more detail in Sect. 5.3.

In the following Sect. 5.2 we show that taking into account the nongaussianity (in addition to the SOS) can be regarded as a further improvement of the inherent adaptation control.

5.2 Realizations for the SIMO Case

As mentioned in Sect. 3.5, most of the existing literature on the identification-and-inversion approach to blind deconvolution is based on the SIMO mixing model, e.g., [10, 11, 29, 30, 31, 32, 33]. Using the TRINICON framework, the approach has been developed rigorously for the more general MIMO case based on first principles.

In this section we show how to deduce the class of SIMO-based algorithms from TRINICON. Besides a generalization of these algorithms, this consideration will also serve as an important background for the later developments in Sect. 6.

As a starting point, we consider the gradient-based update (46a) of the MIMO demixing system $\check{\mathbf{W}}$ with the specialized score function (66) for separation and identification problems.

The ideal separation filter matrix $\check{\mathbf{W}}_{\text{ideal,sep}}$ in the 2×2 case is given by (25), i.e.,

$$\check{\mathbf{W}}_{\text{ideal,sep}} = \begin{bmatrix} \mathbf{h}_{22} & -\mathbf{h}_{12} \\ -\mathbf{h}_{21} & \mathbf{h}_{11} \end{bmatrix} \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix}, \tag{70}$$

where due to the scaling ambiguity (in blind problems) each column is multiplied by an unknown scalar α_q. For $L = L_{\text{opt,sep}} = M$, this ideal separation solution corresponds to a MIMO system identification up to an arbitrary scalar constant (independently of the adaptation method and the possible prior knowledge).

We now consider the SIMO mixing model in Fig. 3(a) as a specialization of the MIMO mixing model in Fig. 3(b), i.e., $\mathbf{h}_{11} \rightarrow \mathbf{h}_1$, $\mathbf{h}_{12} \rightarrow \mathbf{h}_2$, $\mathbf{h}_{21} \rightarrow \mathbf{0}$, $\mathbf{h}_{22} \rightarrow \mathbf{0}$.

According to the right-hand side of (70) the corresponding ideal *demixing system* taking into account this prior knowledge reads

$$\begin{bmatrix} \mathbf{w}_{11} & \mathbf{w}_{12} \\ \mathbf{w}_{21} & \mathbf{w}_{22} \end{bmatrix} = \alpha \begin{bmatrix} \mathbf{0} & -\mathbf{h}_2 \\ \mathbf{0} & \mathbf{h}_1 \end{bmatrix}. \tag{71}$$

By comparing both sides of this equation, we immediately obtain the corresponding demixing system structure shown on the right side in Fig. 3(a). This is indeed the well-known SIMO BSI/AED approach, which in this way follows rigorously from the general equation (70) together with the prior knowledge on the specialized mixing system. Moreover, we see that only the second column of the demixing matrix is relevant for the adaptation process. The elements of the first column can be regarded as *don't cares*.

We now consider the *second term* of the coefficient update (46a). From the relation (134) in Appendix B immediately follows

$$\log \hat{p}_{\mathbf{y},PD}(\mathbf{y}(n)) = \text{const. } \forall\ \mathbf{W} \Rightarrow \log \left|\det\left\{\mathbf{V}^{\mathrm{T}}\mathbf{W}\right\}\right| = \text{const. } \forall\ \mathbf{W}. \tag{72}$$

Specifically, in the case of SOS (e.g., (54)) this leads to

$$\log |\det \mathbf{R}_{\mathbf{yy}}| = \text{const. } \forall\ \mathbf{W} \Rightarrow \log \left|\det\left\{\mathbf{V}^{\mathrm{T}}\mathbf{W}\right\}\right| = \text{const. } \forall\ \mathbf{W}. \tag{73}$$

As the second term in the update (46a) respresents the *gradient* of the expression $\log \left|\det\left\{\mathbf{V}^{\mathrm{T}}\mathbf{W}\right\}\right|$ w.r.t. $\mathbf{W}$, we conclude that *the second term in the coefficient update is equal to zero if* $\det\mathbf{R}_{\mathbf{yy}}$ *is independent of* $\mathbf{W}$. We therefore consider now the dependence of $\det\mathbf{R}_{\mathbf{yy}}$ on $\mathbf{W}$ in more detail. Since $\mathbf{R}_{\mathbf{yy}} = \hat{E}\{\mathbf{y}\mathbf{y}^T\} = \mathbf{W}^T\mathbf{H}^T\mathbf{R}_{\mathbf{ss}}\mathbf{H}\mathbf{W}$, we have

$$\log |\det \mathbf{R}_{\mathbf{yy}}| = \underbrace{\log |\det \mathbf{R}_{\mathbf{ss}}|}_{=\text{const. } \forall\ \mathbf{W}} + 2 \log \left|\det\{\mathbf{W}^T\mathbf{H}^T\}\right|. \tag{74}$$

Now let $\mathbf{W} = \left[\mathbf{W}_1^T, \ldots, \mathbf{W}_P^T\right]^T$ and $\mathbf{H} = [\mathbf{H}_1, \ldots, \mathbf{H}_P]$ be MISO and SIMO, respectively, as special case of the MIMO definition (12). In this special case, the input-output-relation of the overall system reads

$$\mathbf{y} = \mathbf{W}^T\mathbf{H}^T\mathbf{s} = \left(\sum_{p=1}^{P} \mathbf{W}_p^T\mathbf{H}_p^T\right)\mathbf{s}, \tag{75}$$

and $\sum_{p=1}^{P} \mathbf{W}_p^T\mathbf{H}_p^T$ represents an upper triangular matrix with diagonal elements $\sum_{p=1}^{P} w_{p,0}h_{p,0}$. Hence, in the SIMO case, (74) simplifies to

$$\log |\det \mathbf{R}_{\mathbf{yy}}| = \text{const.} + 2N \log \left|\sum_{p=1}^{P} w_{p,0}h_{p,0}\right|. \tag{76}$$

Again, in the special case of only one active source, we can formulate an interesting statement concerning the first taps $w_{p,0}$ of the demixing subfilters. As the demixing subfilters ideally compensate for the individual time-differences of arrival at the microphones, only the subfilter $\mathbf{w}_{p_{\mathrm{far}}}$ connected to the microphone which has the greatest distance to the source, may exhibit a nonzero value at its first tap weight, i.e.,

$$w_{p,0} = \alpha \cdot \delta_{p,p_{\text{far}}}, \tag{77}$$

where δ_{ij} denotes the Kronecker symbol. Introducing this property finally leads to

$$\begin{aligned} \log|\det\mathbf{R}_{\mathbf{yy}}| &= \text{const.} + 2N \log|\alpha h_{p_{\text{far}},0}| \\ &= \text{const.}. \end{aligned} \tag{78}$$

Hence, together with (73), we can draw the conclusion that *in the SIMO case, the second term of the coefficient update (46a) disappears without loss of generality.*

Next, we consider the *first term* $\mathbf{x}(j)\boldsymbol{\Phi}^T_{s,PD}(\mathbf{y}(j))$ in the coefficient update (46a) for the SIMO case and note that its second (block-)column reads $\mathbf{x}(j)\boldsymbol{\Phi}^T_{y_2,D}(\mathbf{y_2}(j))$. We now perform the following formal substitutions in order to be in accordance with the literature on blind SIMO identification and supervised adaptive filtering, e.g., [1] (see Fig. 3(a) and Fig. 3(b)):

$$\mathbf{y}_2 \rightarrow \mathbf{e}, \qquad \begin{bmatrix} \mathbf{w}_{12} \\ \mathbf{w}_{22} \end{bmatrix} = \begin{bmatrix} -\hat{\mathbf{h}}_2 \\ \hat{\mathbf{h}}_1 \end{bmatrix} \rightarrow \mathbf{w} = \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \end{bmatrix}. \tag{79}$$

Hence, the second column of the first term of the coefficient update is finally expressed as $\mathbf{x}(j)\boldsymbol{\Phi}^T_{e,D}(\mathbf{e}(j))$. Note that the substitution of the coefficient notation in (79) is justified by (71).

Thus, we obtain the following sub-matrix of the specialized gradient-based TRINICON update:

$$\mathbf{w}^{\ell}(m) = \mathbf{w}^{\ell-1}(m) + \frac{\mu}{N} \sum_{i=0}^{\infty} \beta(i,m)\, \mathcal{SC} \left\{ \sum_{j=iN_L}^{iN_L+N-1} \mathbf{x}(j)\boldsymbol{\Phi}^T_{e,D}(\mathbf{e}(j)) \right\}. \tag{80}$$

This formally represents the *triple-N-generalization of the Least-Mean-Squares (LMS) algorithm* from supervised adaptive filtering theory (see also [21]) which in its well-known original form exhibits the simple update [1]

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \mu\, \check{\mathbf{x}}(n) e(n), \tag{81}$$

where the length-L vector $\check{\mathbf{x}}$ is a truncated version of $\mathbf{x}$ (formally, this truncation is obtained by $(\mathcal{SC})$ for $D = 1$, see Fig. 6). Although not shown in this chapter, it is possible to analogously derive the corresponding generalizations of other supervised algorithms (NLMS, RLS, etc., which may essentially be seen as special cases of a Newton-type update, e.g., [60]) by choosing a Newton-type TRINICON coefficient update instead of the gradient descent-type update.

From the generalized LMS update (80) above we can make the following observations in comparison with the simple case (81): Due to the generalized approach, we inherently obtain

- block online adaptation, possibly with multiple iterations ℓ to speed up the convergence [19],
- block averaging by $N > 1$ for a more uniform convergence,
- an error nonlinearity to take into account the *nongaussianity* of the signals (by a proper choice of $\boldsymbol{\Phi}_{e,D}^T(\cdot)$),
- multivariate error $\mathbf{e}$ to take into account the *nonwhiteness* of the signals (by choosing $D > 1$).

Note that in various ways, the RLS algorithm can be seen as the optimal supervised adaptation algorithm. However, the RLS is optimum only in the case of a Gaussian source signal and Gaussian additive noise on the microphones, with the noise being additionally stationary and white. The general update resulting from TRINICON does not have these restrictions.

Coefficient initialization

The general relation between MIMO BSI and SIMO BSI also leads to an important guideline for the initialization of the filter coefficients. In particular, we consider the question whether the algorithm can converge to the (undesired) trivial solution $\mathbf{w} = \mathbf{0}$. As we will show, the answer is no, as long as the initialization $\mathbf{w}(0)$ is not orthogonal to the ideal solution $\mathbf{w}_{\text{ideal}} = \left[-\mathbf{h}_2^T\ \mathbf{h}_1^T\right]^T$.

To prove this condition, we pre-multiply the update (80) with $\mathbf{w}_{\text{ideal}}^T$ on both sides of the update equation:

$$\begin{aligned}\mathbf{w}_{\text{ideal}}^T\mathbf{w}^{\ell}(m) &= \mathbf{w}_{\text{ideal}}^T\mathbf{w}^{\ell-1}(m)\\ &+ \frac{\mu}{N}\sum_{i=0}^{\infty}\beta(i,m)\ \left[-\mathbf{h}_2^T\ \mathbf{h}_1^T\right]\ \mathcal{SC}\left\{\sum_{j=iN_L}^{iN_L+N-1}\begin{bmatrix}\mathbf{x}_1(j)\\ \mathbf{x}_2(j)\end{bmatrix}\boldsymbol{\Phi}_{e,D}^T(\mathbf{e}(j))\right\},\end{aligned} \tag{82}$$

$$\begin{aligned}\mathbf{w}_{\text{ideal}}^T\mathbf{w}^{\ell}(m) &= \mathbf{w}_{\text{ideal}}^T\mathbf{w}^{\ell-1}(m) + \frac{\mu}{N}\sum_{i=0}^{\infty}\beta(i,m)\\ &\sum_{j=iN_L}^{iN_L+N-1}\left(\mathbf{h}_1^T\mathcal{SC}\left\{\mathbf{x}_2(j)\boldsymbol{\Phi}_{e,D}^T(\mathbf{e}(j))\right\} - \mathbf{h}_2^T\mathcal{SC}\left\{\mathbf{x}_1(j)\boldsymbol{\Phi}_{e,D}^T(\mathbf{e}(j))\right\}\right).\end{aligned} \tag{83}$$

With (148) from Appendix D this expression can be expanded to

$$\begin{aligned}\mathbf{w}_{\text{ideal}}^T\mathbf{w}^{\ell}(m) &= \mathbf{w}_{\text{ideal}}^T\mathbf{w}^{\ell-1}(m)\\ &+ \frac{\mu}{N}\sum_{i=0}^{\infty}\beta(i,m)\sum_{j=iN_L}^{iN_L+N-1}\sum_{l=1}^{D}\left(\mathbf{h}_1^T\check{\mathbf{x}}_2(j-l+1) - \mathbf{h}_2^T\check{\mathbf{x}}_1(j-l+1)\right)\Phi_{e,l}(\mathbf{e}(j)).\end{aligned} \tag{84}$$

Since $\mathbf{h}_1^T\check{\mathbf{x}}_2(\cdot) - \mathbf{h}_2^T\check{\mathbf{x}}_1(\cdot) \equiv 0$ is fixed due to the acoustic model, we have $\mathbf{w}_{\text{ideal}}^T\mathbf{w}^{\ell}(m) = \mathbf{w}_{\text{ideal}}^T\mathbf{w}^{\ell-1}(m) = \text{const.}$, i.e., *provided that* $\mathbf{w}_{\text{ideal}}^T\mathbf{w}(0) \neq 0$, *the coefficient vector* $\mathbf{w}$ *will not converge to zero.* $\diamond$

Efficient implementation of the Sylvester Constraint for the special case of SIMO models

As already explained for the general MIMO case, we also further specialize the generalized LMS update (80) by incorporating the SIRP model. Introducing the score function (51) immediately leads to SIRPs-based generalized LMS update analogously to [21]

$$\mathbf{w}^{\ell}(m) = \mathbf{w}^{\ell-1}(m) + \frac{\mu}{N} \sum_{i=0}^{\infty} \beta(i,m) \sum_{j=iN_L}^{iN_L+N-1} \mathcal{SC}\left\{\mathbf{x}(j)\mathbf{e}^T(j)\mathbf{R}_{\mathbf{ee}}^{-1}(i)\right\} \phi_{e,D}\left(\mathbf{e}^T(j)\mathbf{R}_{\mathbf{ee}}^{-1}(i)\mathbf{e}(j)\right). \quad (85)$$

As in the general MIMO case, we see that the SIRP model leads to an inherent normalization by the auto-correlation matrix. Note that the SOS case follows for $\phi_{e,D}(\cdot) = 1/2$. In both the SOS case and for general SIRPs the normalization by the correlation matrix in conjunction with $N > 1$ may be interpreted as an *inherent stepsize control*. (It also illustrates why BSS does not require a separate double-talk detector, such as traditional supervised algorithms do, e.g., for acoustic echo cancellation or adaptive beamforming.) Moreover, in [21] it was shown that for a suitable choice of parameters, the general SIRP-based update (85) can be interpreted as a multivariate, i.e., *triple-N generalization of the robust LMS algorithm* based on robust statistics [42], as mentioned in Sect. 4.5.

To further simplify the realization, we next study the expression

$$\mathcal{SC}\left\{\mathbf{x}(j)\mathbf{e}^T(j)\mathbf{R}_{\mathbf{ee}}^{-1}(i)\right\} \quad (86)$$

appearing in (85). According to the structure of the generic Sylvester constraint in Fig. 6 and [8] (see also Appendix D), the l-th element of the p-th subvector (contributing to the p-th channel impulse response) can be expanded to

$$\sum_{d=1}^{D} [\mathbf{x}_p(j)]_{l+d-1} \left[\mathbf{R}_{\mathbf{ee}}^{-1}(i)\mathbf{e}(j)\right]_d = \check{\mathbf{x}}_{p,D}^T(j-l+1)\mathbf{R}_{\mathbf{ee}}^{-1}(i)\mathbf{e}(j), \quad (87)$$

where $\check{\mathbf{x}}_{p,D}$ denotes the length-D vector

$$\check{\mathbf{x}}_{p,D}(n) = [x_p(n), x_p(n-1), \ldots, x_p(n-D+1)]^T. \quad (88)$$

With this expansion, the expression (86) reads

$$\mathcal{SC}\left\{\mathbf{x}(j)\mathbf{e}^T(j)\mathbf{R}_{\mathbf{ee}}^{-1}(i)\right\} = \begin{bmatrix} \check{\mathbf{x}}_{1,D}^T(j) \\ \vdots \\ \check{\mathbf{x}}_{1,D}^T(j-L+1) \\ \check{\mathbf{x}}_{2,D}^T(j) \\ \vdots \\ \check{\mathbf{x}}_{2,D}^T(j-L+1) \end{bmatrix} \mathbf{R}_{\mathbf{ee}}^{-1}(i)\mathbf{e}(j). \quad (89)$$

In the same way as shown in Sect. 4.5 in the context of nearly Gaussian source models, we now factorize the inverse covariance matrix $\mathbf{R}_{\mathbf{ee}}^{-1}$ as [49]

$$\mathbf{R}_{\mathbf{ee}}^{-1}(i) = \mathbf{A}(i)\boldsymbol{\Sigma}_{\tilde{\mathbf{e}}\tilde{\mathbf{e}}}^{-1}(i)\mathbf{A}^T(i), \tag{90}$$

where $\mathbf{A}(i)$ and $\boldsymbol{\Sigma}_{\tilde{\mathbf{e}}\tilde{\mathbf{e}}}(i)$ denote again a $D \times D$ unit lower triangular matrix and a diagonal matrix, respectively [49].

By interpreting $\mathbf{A}(i)$ as a time-varying convolution matrix of a whitening filter, we model the signal e as an AR process of order $D-1$, with time-varying AR coefficients $a_k(n)$, and residual signal $\tilde{e}(n)$, i.e.,

$$e(n) = -\sum_{k=1}^{D-1} a_k(n)e(n-k) + \tilde{e}(n). \tag{91}$$

Now, the expression (89) can be rewritten by shifting the *prefiltering matrix* $\mathbf{A}$ into the data terms, i.e.,

$$\tilde{\mathbf{e}} := \mathbf{A}^T\mathbf{e} = [\tilde{e}(n), \tilde{e}(n-1), \ldots, \tilde{e}(n-D+1)]^T, \tag{92}$$

$$\check{\tilde{\mathbf{x}}}_{p,D} := \mathbf{A}^T\check{\mathbf{x}}_{p,D} = [\tilde{x}_p(n), \tilde{x}_p(n-1), \ldots, \tilde{x}_p(n-D+1)]^T, \tag{93}$$

so that

$$\begin{aligned}
\mathcal{SC}\left\{\mathbf{x}(j)\mathbf{e}^T(j)\mathbf{R}_{\mathbf{ee}}^{-1}(i)\right\} &= \begin{bmatrix} \check{\tilde{\mathbf{x}}}_{1,D}^T(j) \\ \vdots \\ \check{\tilde{\mathbf{x}}}_{1,D}^T(j-L+1) \\ \check{\tilde{\mathbf{x}}}_{2,D}^T(j) \\ \vdots \\ \check{\tilde{\mathbf{x}}}_{2,D}^T(j-L+1) \end{bmatrix} \boldsymbol{\Sigma}_{\tilde{\mathbf{e}}\tilde{\mathbf{e}}}^{-1}(i)\tilde{\mathbf{e}}(j) \\
&= \left[\check{\tilde{\mathbf{x}}}(j), \ldots, \check{\tilde{\mathbf{x}}}(j-D+1)\right] \begin{bmatrix} \frac{\tilde{e}(j)}{\sigma_{\tilde{e}}^2(j)} \\ \vdots \\ \frac{\tilde{e}(j-D+1)}{\sigma_{\tilde{e}}^2(j-D+1)} \end{bmatrix} \\
&= \sum_{d=0}^{D-1} \check{\tilde{\mathbf{x}}}(j-d)\frac{\tilde{e}(j-d)}{\sigma_{\tilde{e}}^2(j-d)}.
\end{aligned} \tag{94}$$

Finally, (85) becomes

$$\begin{aligned}
\mathbf{w}^{\ell}(m) = \mathbf{w}^{\ell-1}(m) + \frac{\mu}{N}\sum_{i=0}^{\infty}\beta(i,m) \\
\sum_{j=iN_L}^{iN_L+N-1}\sum_{d=0}^{D-1}\check{\tilde{\mathbf{x}}}(j-d)\frac{\tilde{e}(j-d)}{\sigma_{\tilde{e}}^2(j-d)}\phi_{e,D}\left(\tilde{\mathbf{e}}^T(j)\boldsymbol{\Sigma}_{\tilde{\mathbf{e}}\tilde{\mathbf{e}}}^{-1}(i)\tilde{\mathbf{e}}(j)\right).
\end{aligned} \tag{95}$$

Note that this formulation provides a computationally efficient realization of the generic Sylvester constraint.

Moreover, it is interesting to note that both the error signal e and the input (i.e., microphone) signal vector $\check{\mathbf{x}}$ appear as filtered versions in the update. After interpreting $\mathbf{A}$ in (90) as a whitening filter, this adaptation algorithm can in fact be interpreted as a so-called *filtered-x*-type algorithm [61]. As shown in Fig. 9, this type of algorithms typically appears whenever there is another filter between the adaptive filter and the position of the error calculation. This cascade structure will also be of fundamental importance in the direct-inverse approach in Sect. 6.

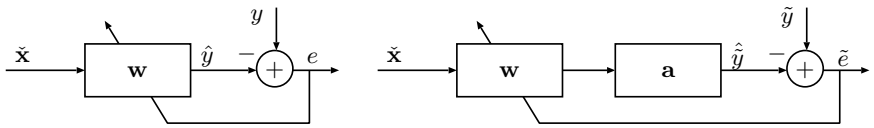


Fig. 9. Supervised adaptive filtering in (a) conventional and (b) filtered-x configuration.

5.3 Efficient Frequency-Domain Realizations for the MIMO Case

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., [2]. However, this traditional narrowband approach exhibits several limitations as identified in, e.g., [62, 63, 64]. In particular, the permutation problem, which is inherent to BSS, may then also appear independently in each frequency bin so that extra repair measures are needed to address this *internal* permutation. Problems caused by circular convolution effects due to the narrowband approximation are reported in, e.g., [63].

In [19] 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., [19] or [58]. Figure 10 gives an overview on the most important classes of DFT-domain BSS algorithms known so far. 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. 10 by different approximations of the generic SIRP-based BSS):

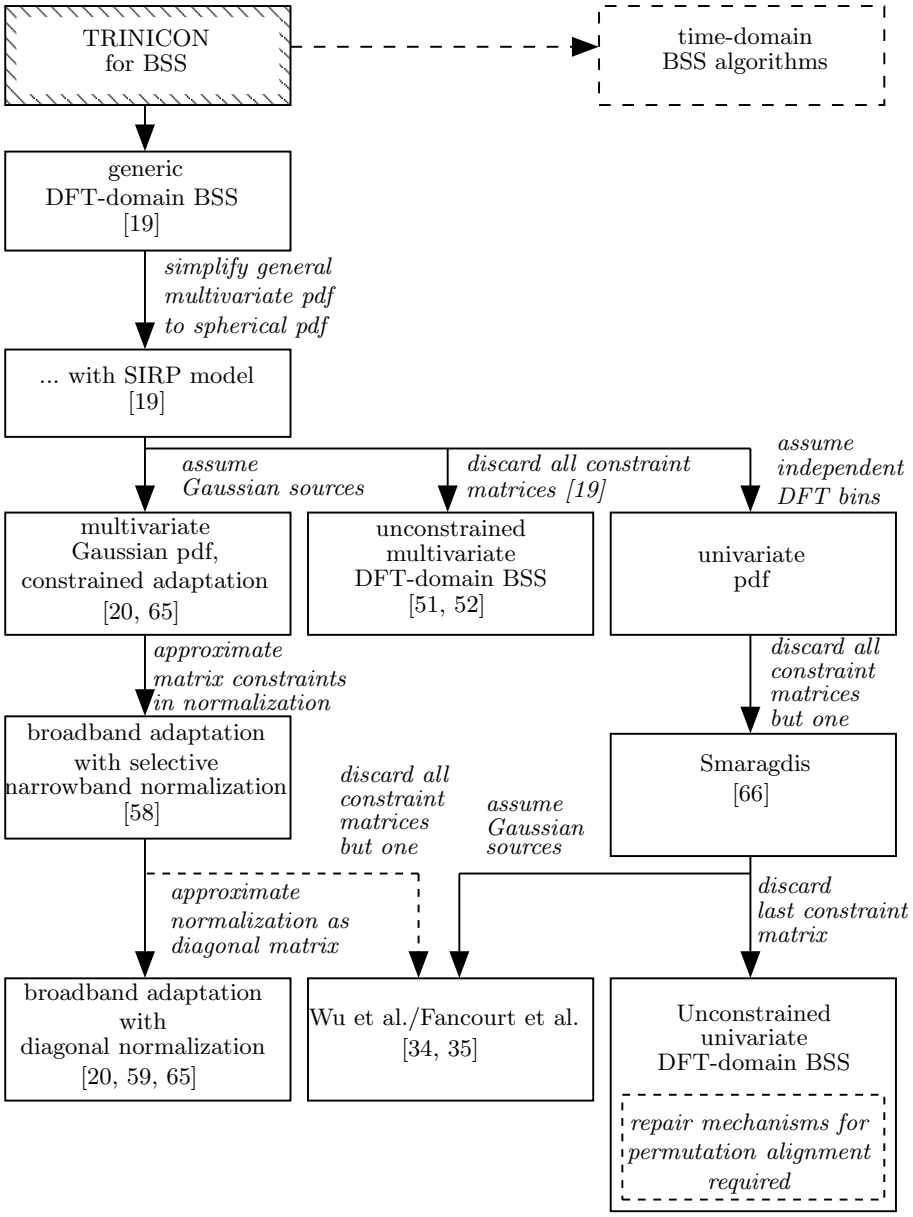


Fig. 10. Overview of BSS algorithms in the DFT domain. Note that the broadband algorithms on the left column are also suitable for BSI, and thus, for the identification-and-inversion approach to blind deconvolution/blind dereverberation.

1. Constraint matrices appearing in the generic frequency-domain formulation (see, e.g., [19]) 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 (50). 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 by considering the joint densities as the most comprehensive description of random signals. 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 blind system identification, the *broadband algorithms with constraint matrices* (i.e., the algorithms represented in the first column of Fig. 10) are of particular interest. Among these algorithms, the system described in [58] has turned out to be very efficient in this context. A pseudo-code of this algorithm is also included in [58].

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 configurations, it is generally *not appropriate* for BSI and deconvolution. Thus, for the implementations discussed in this chapter the *row constraint* $\mathcal{SC}_{\mathrm{R}}$ is used.

6 Application of TRINICON to the Direct-Inverse Approach to Blind Deconvolution

In this section we discuss multichannel blind adaptation algorithms with the aim to solve the inverse adaptive filtering problem (see Table 1) directly without BSI as an intermediate step. This section mainly follows and extends the concept first presented in [12].

The two main aspects in this section are as follows:

- First, we briefly discuss traditional ICA-based multichannel blind deconvolution (MCBD) algorithms from the literature. Unfortunately, as we will see, these algorithms are not well suited for speech and audio signals. However, our considerations lead to various insights and to a classification scheme which is also useful for both the pure separation/identification algorithms from the previous section, and also to the multichannel blind partial deconvolution (MCBPD) algorithms considered afterwards.
- The main aspect in this section is the discussion of the MCBPD algorithms. These algorithms can be regarded as advanced versions of MCBD so that they are also suitable for speech and audio signals. As already mentioned at the end of Sect. 3.5, these algorithms are not just based on the spatial diversity and the statistical independence of the different source signals, but they require more precise stochastic source models. Based on the results of Sect. 4, and to some extent of Sect. 5, we present a general framework which unifies the treatment of many of the known algorithms for the direct-inverse approach to blind dereverberation of speech signals, and also leads to various new algorithms.

6.1 Multichannel Blind Deconvolution (MCBD)

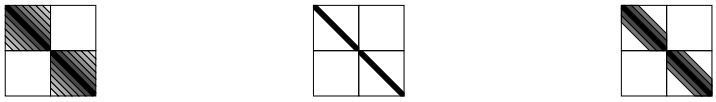
Analogously to the Sect. 5.1, we now specialize TRINICON to the case of traditional MCBD algorithms. As shown by (45b), this class of algorithms is specified by a complete factorization of the hypothesized source model $\hat{p}_{s,PD}(\cdot)$, i.e., traditionally, ICA-based MCBD algorithms assume i.i.d. source models, e.g., [13, 14]. In other words, in addition to the separation of statistically independent sources, MCBD algorithms also temporally whiten the output signals, so that this approach is not directly suitable for audio signals. Nevertheless, studying these algorithms leads to some important insights, as in contrast to some BSS algorithms they are inherently broadband algorithms. Their popularity results from the fact that due to the complete factorization of the source model, they only require univariate pdfs. Thereby, the multivariate score function (44b) reduces to a vector of univariate score functions each representing a scalar nonlinearity. As, additionally, the second term in (44b) is commonly neglected in most of these algorithms, the scalar nonlinearity reads

$$\Phi_{y_p,1}(y_p(j-d+1)) = -\frac{\partial \log \hat{p}_{y_p,1}(y_p(j-d+1))}{\partial y_p(j-d+1)}. \tag{96}$$

The corresponding generic coefficient update rules are then given by (44a), (46a), (48), and (49).

In the SOS case, analogously to the representation in Sect. 5.1, the complete factorization of the output pdf corresponds to the desired correlation matrix $\mathbf{R_{ss}} = \operatorname{diag} \mathbf{R_{yy}}$, as illustrated in Fig. 11(b).

Using (96) several relationships between the generic HOS natural gradient update rule (49) and well-known MCBD algorithms in the literature can



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

Fig. 11. Desired correlation matrices $\mathbf{R_{ss}}$ for BSS (Sect. 5), MCBD (Sect. 6.1), and MCBPD (Sect. 6.2) with TRINICON in the SOS case.

be established [67]. As noted in Sect. 4.5, these links are obtained by the application of different implementations of the Sylvester constraint $\mathcal{SC}$, the distinction between the correlation and covariance method [68] for the estimation of the cross-relation

$$\mathbf{R}_{\mathbf{y}\boldsymbol{\Phi}(\mathbf{y})}(i) = \frac{1}{N} \sum_{j=iN_L}^{iN_L+N-1} \mathbf{y}(j)\boldsymbol{\Phi}_{s,PD}^{\mathrm{T}}(\mathbf{y}(j)) \tag{97}$$

in (49), and the different approximations of the multivariate pdfs. This altogether spans a whole tree of algorithms as depicted in Fig. 12. Here, the most general algorithm is given as the generic HOS natural gradient algorithm (49) which is based on multivariate pdfs. A distinction with respect to the implementation of the Sylvester constraint $\mathcal{SC}$ leads to two branches which can again be split up with respect to the method used for the estimation of the cross-relation matrices. Approximating the multivariate pdfs by univariate ones, neglecting the nonstationarity, and using the Sylvester constraint $\mathcal{SC}_{\mathrm{R}}$ yields two block-based MCBD algorithms presented in [69, 70]. By changing the block-based adaptation to a sample-by-sample algorithm, a link to the popular MCBD algorithm in [13] and [71] can be established. (It should be noted that also the so-called nonholonomic extension [19] of [13] presented in [14] can be derived from the framework.) By using the Sylvester constraint $\mathcal{SC}_{\mathrm{C}}$ a link to the MCBD algorithm in [72] is obtained. However, it should be remembered that algorithms based on $\mathcal{SC}_{\mathrm{C}}$ are less general as only causal filters can be adapted and thus for MCBD algorithms only minimum-phase systems can be treated as was pointed out in [72].

Note that by using the general Sylvester constraint without approximations, a performance gain both over $\mathcal{SC}_{\mathrm{R}}$ and $\mathcal{SC}_{\mathrm{C}}$ is possible [38].

6.2 Multichannel Blind Partial Deconvolution (MCBPD)

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 $\check{\mathbf{H}}$, i.e., the reverberant room. We denote the corresponding generalization of the traditional MCBD

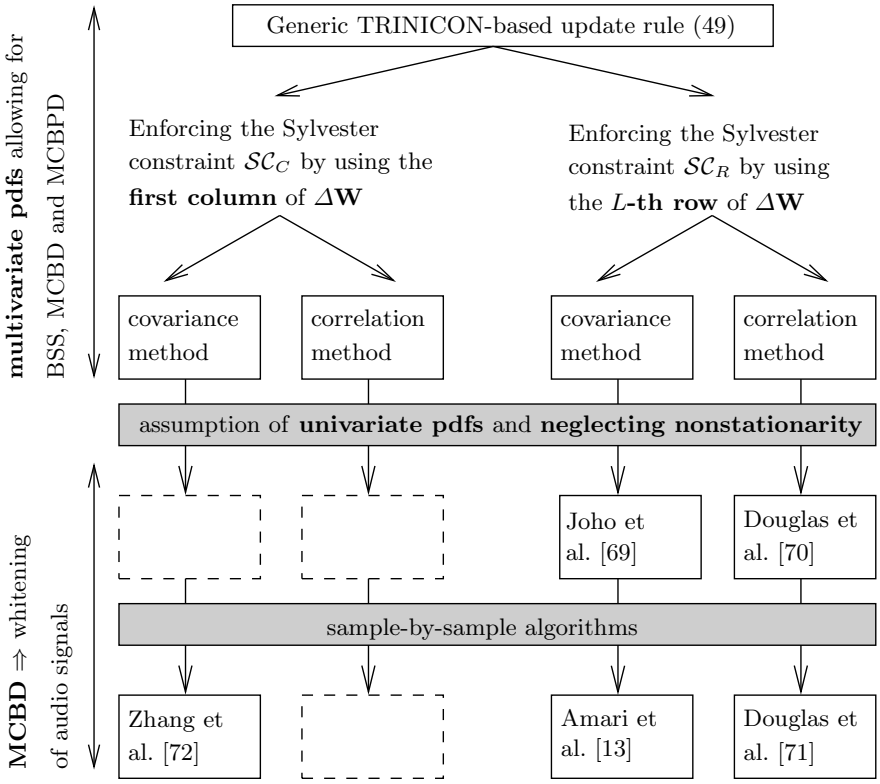


Fig. 12. Overview of links between the generic algorithm (49) and existing MCBD algorithms after [67].

technique as *MultiChannel Blind Partial Deconvolution* (MCBPD) [12]. Equations (44)-(49) inherently contain a statistical source model (signal properties (i)-(iii) in Sect. 4.2), expressed by the multivariate densities, and thus provide all necessary requirements for the MCBPD approach.

Ideally, only the influence of the room acoustics should be minimized. A typical example for MCBPD applications is speech dereverberation, which is especially important for distant-talking automatic speech recognition (ASR), where there is a strong demand 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. 13).

For the distinction between the production system of the source signals (e.g., the speech production system) and the room acoustics we can again exploit all three fundamental signal properties already mentioned in Sect. 4.2:

(i) Nonwhiteness. The auto-correlation structure of the speech signals can be taken into account, as illustrated in Fig. 11(c) after [12]. While the

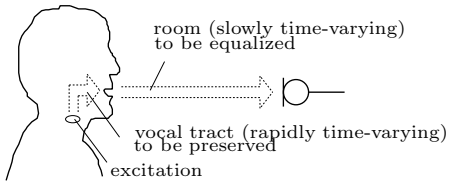


Fig. 13. Illustration of speech dereverberation as an MCBPD application (after [12]).

room acoustics influences all off-diagonals, the effect of the vocal tract is concentrated in the first few off-diagonals around the main diagonal. In the simplest case, these first Z off-diagonals of $\mathbf{R_{yy}}$ are now taken over into the banded matrix

$$\mathbf{R_{ss}} = \text{bandbdiag}_Z \, \mathbf{R_{yy}}, \tag{98}$$

as illustrated in Fig. 11(c). Note that there is a close link to linear prediction techniques as detailed below which gives guidelines for the number of lags to be preserved.

(ii) Nonstationarity. The speech production system and the room acoustics also differ in their time-variance according to Fig. 13 after [12]. While the room acoustics is assumed to be constant during the adaptation process, the speech signal is only short-time stationary [68], modeled by the time-varying speech production model. Typically, the duration of the stationarity intervals is assumed to be approximately 20ms [68]. We therefore adjust the block length N and in practice preferably also the block shift N_L in the criterion (39) with the model parameter estimates (41) and in the corresponding updates (44)-(49) to the assumed duration of the stationarity interval. Note that for a block-based adaptation (typically performed by exploiting the efficiency of the FFT, cf. Sect. 5.3 for the case of BSS) and $N = N_L < L$, this corresponds to a *partitioned* block formulation as known from supervised adaptive filtering, e.g., [60].

(iii) Nongaussianity. Speech is a well-known example for supergaussian signals. Due to a convolutive sum – describing in our application the filtering by the room acoustics – the pdfs of the recorded sensor/microphone signals tend to be somewhat closer to Gaussians. Hence, another strategy is to maximize the nongaussianity of the output signals of the demixing system (as far as possible by the MIMO FIR filters), e.g., [73, 74, 75, 76]. This strategy is addressed, e.g., using the kurtosis as a widely-used distance measure of nongaussianity as in the second term in (56). It can be shown that this second term can indeed be identified as an estimate of the so-called *negentropy* which is an information-theoretic distance measure to the Gaussian [2].

Formally, the above-mentioned exploitation of the nonwhiteness to distinguish between the coloration of the sources and the mixing system is achieved by decoupling the prediction order n_A in (61) from the dimension D of the correlation matrix $\mathbf{R}_{\mathbf{yy}}$, i.e.,

$$\tilde{y}_p(n) = \sum_{k=0}^{n_A} a_{p,k}(n) y_p(n-k) \tag{99}$$

with $0 \leq n_A \leq D-1$ and $a_{p,0}(n) \equiv 1$. This corresponds to a generalization of the upper triangular matrix structure (62) in the factorization (60) to the *banded* matrix

$$\mathbf{A}_p = \begin{bmatrix} 1 & a_{p,1}(n) & a_{p,2}(n) & \cdots & a_{p,n_A}(n) & 0 & \cdots & 0 \\ 0 & 1 & a_{p,1}(n-1) & \cdots & a_{p,n_A-1}(n-1) & a_{p,n_A}(n-1) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 1 \end{bmatrix}^T \tag{100}$$

so that we can again apply the compact notation

$$\tilde{\mathbf{y}}_p = \mathbf{A}_p^T \mathbf{y}_p = [\tilde{y}_p(n), \tilde{y}_p(n-1), \ldots, \tilde{y}_p(n-D+1)]^T , \tag{101}$$

$$\tilde{\check{\mathbf{x}}}_{p,D} = \mathbf{A}_p^T \check{\mathbf{x}}_{p,D} = [\tilde{x}_p(n), \tilde{x}_p(n-1), \ldots, \tilde{x}_p(n-D+1)]^T . \tag{102}$$

Hence, the resulting formulation of the generalized score function (65) carries over to MCBPD, as well as to the traditional MCBD and to broadband BSS/BSI, depending on the parameter n_A. In other words, the different modes in Fig. 11 are selected by certain choices of the order n_A. This is further illustrated in Fig. 14.

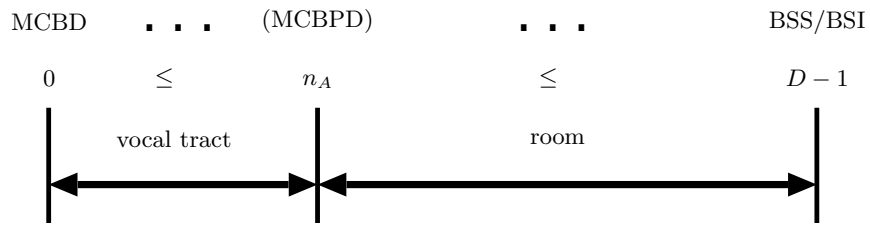


Fig. 14. Illustration of the parameter n_A.

The corresponding general gradient descent-based coefficient update for nearly Gaussian sources is then obtained by introducing the score function (65) into the generic update (46a). Note that for an efficient implementation of the Sylvester constraint of the first term in (46a) we can apply the same procedure as demonstrated in (87) and (89). With (102) we then obtain

$$\check{\mathbf{W}}^{\ell}(m) = \check{\mathbf{W}}^{\ell-1}(m) - \frac{\mu}{N}\sum_{i=0}^{\infty}\beta(i,m)\sum_{j=iN_L}^{iN_L+N-1}\sum_{d=0}^{D-1}\check{\tilde{\mathbf{x}}}(j-d)$$
$$\cdot\left[\frac{\tilde{y}_p(j-d)}{2\hat{\sigma}^2_{\tilde{y}_p}(j-d)} - \left(\frac{\sum_{j=iN_L}^{iN_L+N-1}\tilde{y}_p^4(j-d)}{3\hat{\sigma}^4_{\tilde{y}_p}(j-d)} - 1\right)\right.$$
$$\left.\cdot\left(\frac{\tilde{y}_p^3(j-d)}{\hat{\sigma}^4_{\tilde{y}_p}(j-d)} - \frac{\tilde{y}_p(j-d)\sum_{j=iN_L}^{iN_L+N-1}\tilde{y}_p^4(j-d)}{\hat{\sigma}^6_{\tilde{y}_p}(j-d)}\right)\right]$$
$$+\mu\sum_{i=0}^{\infty}\beta(i,m)\mathcal{SC}\left\{\mathbf{V}\left(\left(\mathbf{W}^{\ell-1}(m)\right)^{\mathrm{T}}\mathbf{V}\right)^{-1}\right\}, \qquad (103)$$

where the expression in brackets denotes a row vector composed of the elements for $p = 1, \ldots, P$. *This general TRINICON-based MIMO coefficient update for nearly Gaussian sources leads both to blind separation and dereverberation of the signals.*

Analogously to the considerations at the end of Sect. 5.2 we see that this update rule can again be interpreted as a so-called *filtered-x*-type algorithm since both the input (i.e., microphone) signal vector and the output signals appear as filtered versions in the update. Analogously to Fig. 9 we immediately obtain Fig. 15 for the dereverberation application as a consequence of this filtered-x interpretation. While $\mathbf{W}$, driven by the filtered-x-type coefficient update, ideally inverts the room acoustic mixing system $\mathbf{H}$, the (set of) linear prediction filter(s) $\mathbf{A}$ from the stochastic source model ideally inverts the (set of) speech production system(s) of the source(s). The coefficient updates of $\mathbf{W}$ and the estimation of $\mathbf{A}$ are carried out in an alternating fashion like the estimation of the other stochastic model parameters, as mentioned in Sect. 4.5. Note that (in accordance with the known filtered-x concept) the filtered input vector $\check{\tilde{\mathbf{x}}}$ in (103) is obtained using the filter coefficients from the linear prediction (LP) analysis of the *output* signals y_p. In other words, the coefficients of the output LP analysis filters are copied to the input transformation filters according to (102).

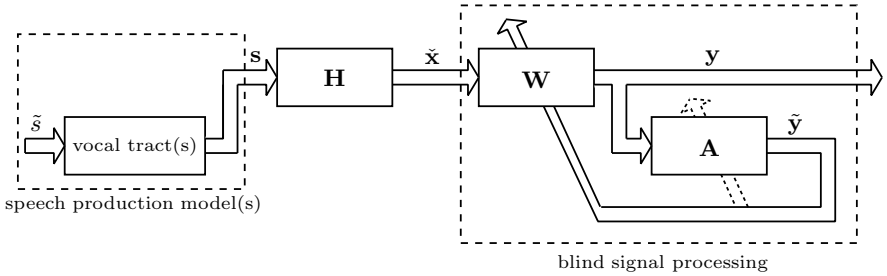


Fig. 15. Inversion of the speech production models within the blind signal processing and filtered-x-type interpretation.

It should be mentioned that the linear prediction is also classified as a (blind) inverse adaptive filtering problem in Table 1, and hence, the estimate of the prediction coefficients can also be obtained directly from the TRINICON optimization criterion (39). Assuming a single-source scenario and SOS-based estimation of the prediction coefficients, we obtain for this inverse adaptive filtering problem as a special case of (39) according to (54) and the considerations in Sect. 5.2 for the single-source case

$$\mathcal{J}_{\text{pred}}\left(m,\mathbf{A}\right)=\sum_{i=0}^{\infty}\beta(i,m)\log\det\operatorname{diag}\mathbf{R}_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}}(i)\propto\sum_{i=0}^{\infty}\beta(i,m)\log\hat{\sigma}_{\tilde{y},i}^{2}. \quad (104)$$

Furthermore, assuming stationarity, this criterion is equivalent to the traditional least-squares-based estimate $\mathcal{J}_{\text{pred,LS}}\left(m,\mathbf{A}\right)\propto\hat{\sigma}_{\tilde{y},m}^{2}$ due to the monotonicity of the logarithm, while for non-stationary signals, it is more general. Nevertheless, for the practical experiments in Sect. 7 we will apply the Levinson-Durbin algorithm as an efficient realization of the LS-based estimation using the so-called correlation method [68].

6.3 Special Cases and Links to Known Algoritms

According to Fig. 14, all of the previously discussed algorithms from the various classes according to Tab. 1 can be regarded as special cases of the MCBPD concept. In this section, we only discuss algorithms that are specifically designed for dereverberation using the direct-inverse approach. Moreover, we focus here on algorithms based on the Gram-Charlier model, i.e., we discuss special cases of (103) and relations to some known algorithms.

SIMO vs. MIMO mixing systems

Similar to the considerations in Sect. 5.2 for SIMO-based BSI, we deduce now the specialized coefficient update for the case of SIMO mixing systems, i.e., for the case of only one source signal. Again, we first consider the last term of the generic gradient-based update (103). According to the corresponding steps of the derivation in Sect. 5.2 (Eqs. (72)-(78)) we can see that in the same way the last term also disappears for MCBPD in the SIMO case. Next, we pick the filter coefficients of interest for the SIMO case. Assuming the active source signal will appear on the first output of the demixing filter, it is straightforward to pick $\mathbf{w}$ as the first column of the general MIMO coefficient matrix $\check{\mathbf{W}}$. This way we immediately obtain

$$\mathbf{w}^{\ell}(m) = \mathbf{w}^{\ell-1}(m) - \frac{\mu}{N}\sum_{i=0}^{\infty}\beta(i,m)\sum_{j=iN_L}^{iN_L+N-1}\sum_{d=0}^{D-1}\check{\tilde{\mathbf{x}}}(j-d) \cdot\left(\frac{\tilde{y}(j-d)}{2\hat{\sigma}_{\tilde{y}}^2(j-d)} - \left(\frac{\sum_{j=iN_L}^{iN_L+N-1}\tilde{y}^4(j-d)}{3\hat{\sigma}_{\tilde{y}}^4(j-d)} - 1\right)\right. \cdot\left.\left(\frac{\tilde{y}^3(j-d)}{\hat{\sigma}_{\tilde{y}}^4(j-d)} - \frac{\tilde{y}(j-d)\sum_{j=iN_L}^{iN_L+N-1}\tilde{y}^4(j-d)}{\hat{\sigma}_{\tilde{y}}^6(j-d)}\right)\right). \tag{105}$$

Note that the structure of the resulting algorithm is very similar to the one of the generalized AED (95) in Sect. 5.2. The main differences are the different sign of the update term and the fact that we now pick the *first* column of $\check{\mathbf{W}}$ since we are now interested in obtaining the enhanced signal rather than in minimizing an error signal for the signal cancellation in the AED.

Efficient implementation using the correlation method

An efficient implementation which still exploits all three fundamental signal properties as discussed in Sect. 6.2 is obtained by assuming a global nonstationarity of the source signals but short-time stationarity in each block as known from linear prediction. As a first step to obtain a simplified update equation, we integrate the sum over d into the sum over j. Next, we replace the time-varying output prediction error variances by blockwise constant values $\hat{\sigma}_{\tilde{y}_p,i}$ for the i-th block. This finally allows us to move the sum over j into the numerators in the brackets in order to obtain the compact expression

$$\check{\mathbf{W}}^{\ell}(m) = \check{\mathbf{W}}^{\ell-1}(m) - \frac{\mu}{N}\sum_{i=0}^{\infty}\beta'(i,m) \cdot\left[\frac{\sum_{j=iN'_L}^{iN'_L+N-1}\check{\tilde{\mathbf{x}}}(j)\tilde{y}_p(j)}{2\hat{\sigma}_{\tilde{y}_p,i}^2} - \left(\frac{\sum_{j=iN'_L}^{iN'_L+N-1}\tilde{y}_p^4(j)}{3\hat{\sigma}_{\tilde{y}_p,i}^4} - 1\right)\right. \cdot\left.\left(\frac{\sum_{j=iN'_L}^{iN'_L+N-1}\check{\tilde{\mathbf{x}}}(j)\tilde{y}_p^3(j)}{\hat{\sigma}_{\tilde{y}_p,i}^4} - \frac{\sum_{j=iN'_L}^{iN'_L+N-1}\check{\tilde{\mathbf{x}}}(j)\tilde{y}_p(j)\sum_{j=iN_L}^{iN_L+N-1}\tilde{y}_p^4(j)}{\hat{\sigma}_{\tilde{y}_p,i}^6}\right)\right] + \mu\sum_{i=0}^{\infty}\beta(i,m)\mathcal{SC}\left\{\mathbf{V}\left(\left(\mathbf{W}^{\ell-1}(m)\right)^{\mathrm{T}}\mathbf{V}\right)^{-1}\right\}. \tag{106}$$

Note that for the SIMO case this expression is simplified in a straightforward way as mentioned in the previous paragraph so that the last term again disappears. This efficient version is also used for the experiments in Sect. 7.

Relations to some known HOS approaches

As already mentioned in Sect. 6.2 most of the HOS-based blind deconvolution approaches aim at finding deconvolution filters that render the output signals

as nongaussian as possible [73, 74, 75] with kurtosis being the most common measure for nongaussianity.

In [76] an approach to speech dereverberation by kurtosis maximization was presented. It is based on the idea of performing the whole adaptation and filtering procedure on LP residuals as a heuristic extension of the ideas in [77, 78]. Hence, the main structural difference of this approach to the general TRINICON-based update rule is that the LP analysis is carried out using the microphone signals, i.e., the input signals of the blind adaptive filter rather than on its output signals as in the above-mentioned and systematically obtained filtered-x structure. Nevertheless, the resulting algorithm also exhibits several remarkable similarities to the generic update. The adaptation rule in [76] is based directly on the kurtosis, i.e., the *square root* of only the second term in (56). The update therefore structurally corresponds to the part in the second parentheses of the second term in the brackets in (103). (The first term in (103) results from the SOS and the expression in the first parentheses in the second term results from the application of the chain rule due to the *square* of the kurtosis in the Gram-Charlier expansion.)

The same approximate expression of the update rule, i.e., the gradient descent directly based on the kurtosis is also used in [79]. Note that these approaches are based on the acoustic SIMO model.

Relations to some known SOS approaches

It is known that linear filtering of a source signal increases the temporal predictability of the observed signal. A deconvolution filter which makes its output less predictable may thus be able to recover the source signal. This observation is the key to most of the SOS-based linear deconvolution methods, i.e., in essence they aim at finding deconvolution filters which minimize a measure of predictability of the output signal, e.g., [80]. Hence, in a certain sense, blind deconvolution may also be interpreted as the application of a very long linear prediction error filter. Note that this is also reflected by the symmetric structure in Fig. 15.

As a simple approach, the optimization criterion in [80] is directly based on the variance of the long-term prediction error at the output of the deconvolution filter. In order to avoid trivial solutions and to preserve some of the temporal structure of the source signals, this long-term prediction error variance is normalized by a short-term prediction error variance, and finally the logarithm of this ratio is taken. Although this approach does not explicitly exploit the nonstationarity of the signals in the sense as outlined in Sect. 6.2, this logarithm of the ratio between the prediction error variances – which can be expressed as a difference between two logarithmic prediction error variances – can still roughly be related to the generic SOS-based optimization criterion (54) considering the link with linear prediction at the end of Sect. 6.2, and the short-term prediction error variance in the normalization as a special case of the *desired* correlation matrix $\mathbf{R}_{\mathbf{ss}}$.

Another related approach to preserve the temporal structure of the original source signal is called correlation shaping in [15]. The heuristically introduced optimization criterion after Gillespie and Atlas in [15] for the SIMO case reads

$$\mathcal{J}_{\mathrm{GA}} = \sum_{\kappa} \gamma(\kappa) \left(r_{yy}(\kappa) - r_{ss}(\kappa) \right)^2 , \tag{107}$$

where κ denotes the lag of the output correlation sequence $r_{yy}(\kappa)$ and a certain desired correlation sequence $r_{ss}(\kappa)$. The factor $\gamma(\kappa)$ allows for an individual weighting of the lags. As a preferred embodiment of this concept, it is proposed in [15] to choose $\gamma(\kappa)$ and $r_{ss}(\kappa)$ such that $r_{yy}(\kappa)$ is minimized for all lags outside of the *don't care* region $-Z \leq \kappa \leq Z$. Obviously, this approach is equivalent to the minimization of the Frobenius norm $\mathcal{J}_{\mathrm{F,GA}} = \|\mathbf{R_{yy}} - \mathbf{R_{ss}}\|_{\mathrm{F}}^2$ with the banded matrix $\mathbf{R_{ss}} = \mathrm{bandbdiag}_Z \, \mathbf{R_{yy}}$ after (98) and Fig. 11(c) if the so-called *correlation method* is used for the estimation of $\mathbf{R_{yy}}$ (i.e., this matrix is assumed to be Toeplitz). Hence, in the context of dereverberation the approach [15] can be seen directly as an analogon to the Frobenius-based approaches for BSS/BSI mentioned in Sect. 5.1 (e.g., [2],[53]-[57]). The main differences between [15] and the generic SOS-based MCBPD are:
(i) The criterion (107) does not exploit the nonstationarity of the signals in the sense as outlined in Sect. 6.2.
(ii) As already explained in Sect. 5.1, in contrast to the generic SOS criterion (54) the minimization of the Frobenius-based criterion does not lead to the inherent normalization of the coefficient update which can be interpreted as an inherent stepsize control according to Sect. 5.2, and hence is an important feature for a robust adaptation performance. Similar to the BSS/BSI case, many simulation results have shown that for large filter lengths L, the Frobenius-based adaptation is prone to instability, while the generic MCBPD adaptation shows a very robust convergence behavior for real acoustic environments, as we will see in Sect. 7.

In [23, 81] a third related SOS-based approach was presented. As in the previously described SOS-based algorithms, this approach distinguishes between the speech production system and the room acoustics by exploiting only the nonwhiteness. It explicitly takes into account an estimate of the *long-term* power spectral density of the speech signal. Moreover, an interesting aspect of this approach is that it was originally derived directly from MINT (see Sect. 2) describing the ideal inversion solution at the equilibrium of the adaptation. Indeed, it can be shown (analogously to the analysis of the equilibria for BSS in [20] in the SOS case) that ideally the equilibrium of the SOS-based update (67) in the case of MCBPD with (98) corresponds to the MINT solution according to Sect. 2. We now show how this approach can be derived rigorously from the TRINICON-based coefficient update (67). Under the stationarity assumption we have in the equilibrium

$$\Delta\mathbf{W} = \mathbf{R_{xy}} \left[\mathbf{R_{ss}}^{-1} - \mathbf{R_{yy}}^{-1} \right] = \mathbf{0}, \tag{108}$$

i.e.,

$$\mathbf{R_{xy}} = \mathbf{R_{xy}}\mathbf{R_{yy}^{-1}}\mathbf{R_{ss}}. \tag{109}$$

Developing the left hand side of this equation as $\mathbf{R_{xx}}\mathbf{W}$ and the right hand side of this equation using Sylvester matrices and corresponding data matrices $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{S}$ of compatible dimensions as in [20] as $\mathbf{R_{xy}}\mathbf{R_{yy}^{-1}}\mathbf{R_{ss}} = \mathbf{X}^T\mathbf{Y}(\mathbf{Y}^T\mathbf{Y})^{-1}\mathbf{S}^T\mathbf{S} = \mathbf{X}^T(\mathbf{Y}^T)^+\mathbf{S}^T\mathbf{S} = \mathbf{X}^T(\mathbf{S}^T)^+(\mathbf{C}^T)^+\mathbf{S}^T\mathbf{S} = \mathbf{X}^T\mathbf{S} = \mathbf{R_{xs}}$, where $\cdot^+$ denotes the Moore-Penrose pseudoinverse, we obtain

$$\mathbf{R_{xx}}\mathbf{W} = \mathbf{R_{xs}}. \tag{110}$$

Note that this relation is in fact the Wiener-Hopf equation for the inverse filtering configuration. (This again reflects the equivalence to the traditional LS approach for inverse adaptive filtering problems in the stationary case, as mentioned at the end of Sect. 6.2 for the linear prediction problem.) Next, a filter $\mathbf{B}$ in Sylvester structure modeling the vocal tract is introduced so that $\mathbf{S} = \mathbf{S}_0\mathbf{B}$, where $\mathbf{S}_0$ denotes a corresponding data matrix of the i.i.d. excitation signal. Hence

$$\mathbf{R_{ss}} = \mathbf{S}^T\mathbf{S} = \mathbf{B}^T\mathbf{R_{s_0s_0}}\mathbf{B} = \mathbf{B}^T\mathbf{B}. \tag{111}$$

Using this model, we can rewrite (110) as

$$\mathbf{R_{xx}}\mathbf{W} = \mathbf{H}^T\mathbf{R_{ss}} = \mathbf{H}^T\mathbf{B}^T\mathbf{B}. \tag{112}$$

Multiplication by the pseudo inverse of $\mathbf{B}$ on both sides, and exploiting the commutation property of the convolution ($\mathbf{B}$ denotes a SISO system), we can write

$$\mathbf{R_{xx}}\mathbf{B}^+\mathbf{W} = \mathbf{B}^T\mathbf{H}^T \tag{113}$$

or

$$\left(\mathbf{B}^+\right)^T\mathbf{R_{xx}}\mathbf{B}^+\mathbf{W} = \mathbf{H}^T. \tag{114}$$

Let us denote the inverse filter of the vocal tract similarly as in the previous sections as $\mathbf{A} := \mathbf{B}^+$. Using this filter the correlation matrix $\mathbf{R_{xx}}$ is transformed into $\mathbf{R_{\tilde{x}\tilde{x}}} = \mathbf{A}^T\mathbf{R_{xx}}\mathbf{A}$ so that

$$\mathbf{R_{\tilde{x}\tilde{x}}}\mathbf{W} = \mathbf{H}^T. \tag{115}$$

We now pick only the first columns of the Sylvester matrices for the SIMO case on both sides. Moreover, it is important to assume that the first microphone is the one that is closest to the source [81]. Using this assumption we finally obtain

$$\mathbf{w} = h_{1,0}\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}^{-1}\mathbf{1}, \tag{116}$$

where $\mathbf{1} = [1, 0, \ldots, 0]^T$ and $h_{1,0}$ denotes the first coefficient of the acoustic model from the source to the first microphone which acts as an arbitrary scaling factor. This expression exactly corresponds to the algorithm presented in [81] including the whitening procedure, originally introduced in a heuristic way. We can see from this derivation that this algorithm indeed follows from

TRINICON for the SOS case and stationarity assumption. Moreover, we see that in contrast to the previously presented approaches, this algorithm requires some prior knowledge on the source position. In other words, it may in fact be regarded as a *semi-blind* deconvolution algorithm. Furthermore, it becomes obvious that extending this approach to the general MIMO case raises the problem of estimating the relative positions of multiple simultaneously active sound sources.

7 Experiments

In this section, we evaluate the dereverberation performance for both the SIMO case (i.e., one source) and the MIMO case (two sources) using measured data. In the first set of experiments in the SIMO case, we compare the convergence properties based on the exploitation of the different stochastic signal properties (SOS, HOS) for the ideal demixing filter length. We then compare the DI approach with the II approach and investigate the sensitivity of both approaches w.r.t. the overestimation of the filter lengths. Finally, by extending the scenario to the MIMO case, we consider both the separation performance and the dereverberation performance. For illustration, we also compare the results in the MIMO case with the corresponding results of pure separation algorithms.

7.1 SIMO case

The experiments have been conducted using speech data convolved with impulse responses of length $M = 9000$ of a real room with a reverberation time $T_{60} \approx 700$ms and a sampling frequency of 16kHz. To begin with, we consider an acoustic SIMO scenario, i.e., there is only $Q = 1$ active sound source in the room. A linear four-element microphone array ($P = 4$) with an inter-element spacing of 16cm was used. Preliminary experiments using MINT (see Sect. 2) applied to the measured impulse responses have shown that for the choice $P = 4$ the ideal inversion solution indeed exists for the given acoustic scenario, i.e., the mixing system is invertible according to Sect. 2. The speech signal arrived from 24^o relative to the normal plane of the array axis and the distance between the speaker and the center of the microphone array was 165cm.

As already mentioned, according to MINT the overdetermined scenario $P > Q$ is required for dereverberation. From a practical point of view it is thus interesting to consider the required degrees of freedom depending on the number of sensors. The total number of filter coefficients is $C := LP$. According to (18), we obtain as the optimal number of filter coefficients in the SIMO case

$$C = P \cdot \frac{M-1}{P-1} = \frac{P}{P-1} \cdot (M-1). \tag{117}$$

We see that for the minimum number $P = 2$ of sensors we require $C = 2 \cdot (M-1)$ coefficients. For $P \to \infty$ it follows $C \to M-1$. It turns out that the total number of required filter coefficients *decreases* with increasing number of microphones. Hence, the framework is well suitable and efficient for the overdetermined case.

To evaluate our simulation results there are various possible quality measures for dereverberation of speech and audio signals (e.g., [82, 83, 84, 85]), such as the reverberation time (T_{60}), the definition (D_{50}), the clarity index (C_{80}), the (rapid) speech transmission index (STI/RASTI), or the spectral distortion (SD). While the first three quantities are system-based and are defined in the context of room acoustics, the latter two are signal-dependent distortion measures. Another signal-dependent quantity which is commonly used in the signal processing literature for the evaluation of dereverberation approaches is the *signal-to-reverberation ratio* (SRR, see, e.g., [86]). Similarly to the quantities D_{50} and C_{80} it measures the power ratio between the direct sound and the contribution by the reverberation. However, since the SRR is signal-based, it also takes into account the excitation of the adaptation algorithm. It is measured in decibel (dB) and is defined for a signal s_q at a sensor with signal x_p as

$$SRR_{p,s_q} = 10 \log \frac{\sum_n \left(\sum_{\kappa=0}^{n_\Delta} h_{qp,\kappa} s_q(n-\kappa)\right)^2}{\sum_n \left(\sum_{\kappa=n_\Delta}^{M-1} h_{qp,\kappa} s_q(n-\kappa)\right)^2} \mathrm{dB}, \tag{118}$$

where n_Δ is a discrete-time index defining the boundary between the direct signal path and the contribution by the reverberation. Note that usually, in the case of speech signals, the first 50ms after the main peak of the impulse responses are also added to the contribution of the direct path, i.e., n_Δ is replaced by the so-called critical delay time n_{50}, which is known to contribute to the speech intelligibility [82]. In the following simulation results this perceptual effect is taken into account. The SRR after (118) also forms the basis for the definition of the so-called *segmental SRR* (e.g., [86]), which is usually preferred in practice due to the nonstationarity of speech and audio signals and the higher correlation to the quality perceived by auditory measurements. The segmental SRR is based on time-varying local SRR estimates which are obtained by decomposing the signals into K_S segments of length N_S, i.e., the averaging in (118) is performed only over these short intervals. The segmental SRR is then defined as the average of the local SRR estimates over the K_S segments. In our simulations, we use $N_\mathrm{S} = 320$. This corresponds to the typical stationarity interval for speech (20ms for a sampling rate of 16kHz).

Furthermore, in the context of adaptive signal processing, another interesting aspect of the SRR is that formally, it corresponds directly to the definition of the so-called *signal-to-interference ratio* (SIR) which is usually used in the literature for the evaluation of signal separation approaches, such as BSS. If we consider the MCBPD optimization criterion, which can also be regarded

as a *contrast function* for signal separation and dereverberation, we may hypothesize that in practice, the potential SRR improvement will generally be upper-bounded by the potential SIR improvement in the MIMO case. The same consideration also applies to the segmental SRR and the segmental SIR.

We first consider the Direct-Inverse approach to SIMO-based dereverberation. Our simulations are based on the coefficient update (106) (without the last term in the SIMO case) using the correlation method. We chose $L = 3000$ according to (18), the block length $N = N'_L = 320$ corresponding to a stationarity interval of 20ms, and $n_A = 32$. Figure 16 shows the SRR improvement for offline (batch) adaptation, i.e., $\beta(i,m) = \beta(i)$ in (39) (and thus $\beta'(i,m) = \beta'(i)$ in (106)) corresponds to a rectangular window function over the entire available signal length, and the outer sum in (39) and (106) turns into a summation of the contributions from all blocks with equal weights. The left subplot illustrates the convergence over the number of iterations.

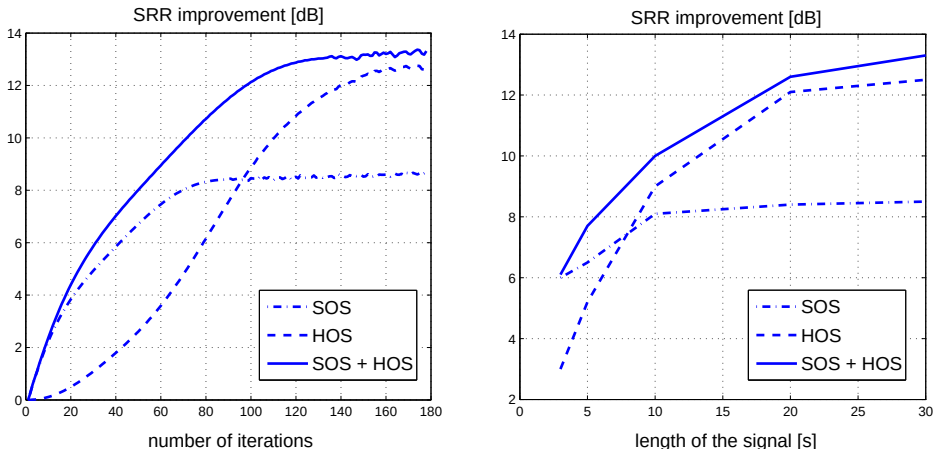


Fig. 16. SRR performance of SIMO-based MCBPD for (a) increasing number of offline-iterations, (b) different overall signal lengths.

We see that the optimization based purely on second-order statistics (SOS, dash-dot line, only the first term in the brackets in (106) was used) exhibits a rapid initial convergence, while the kurtosis-based approach (HOS, dashed line, only the second term in the brackets in (106) was used) finally achieves a higher level of SRR improvement at the cost of a slower initial convergence. By exploiting all the available statistical signal properties (SOS+HOS, solid line, both terms in the brackets in (106) were used), the TRINICON framework combines the advantages of the former two approaches. The higher data requirement for HOS-based estimation is also reflected in the right subplot. Here, we performed the offline adaptation for various overall signal lengths. It can be seen that the SOS-based contribution of the optimization already

provides reasonable performance for relatively short signal lengths. Hence, in practice, where online adaptation is required due to potential changes of the room impulse responses, the synergy effects provided by TRINICON appear to be attractive.

Figure 17 shows the first 5000 taps of one of the room impulse responses of the measured mixing system and of the overall system (i.e., between the source and the MCBPD output) after dereverberation, based on the combined (SOS+HOS) TRINICON approach and 180 iterations with a signal length of 30s (see Fig. 16). The same parameters were used for the spectrograms for

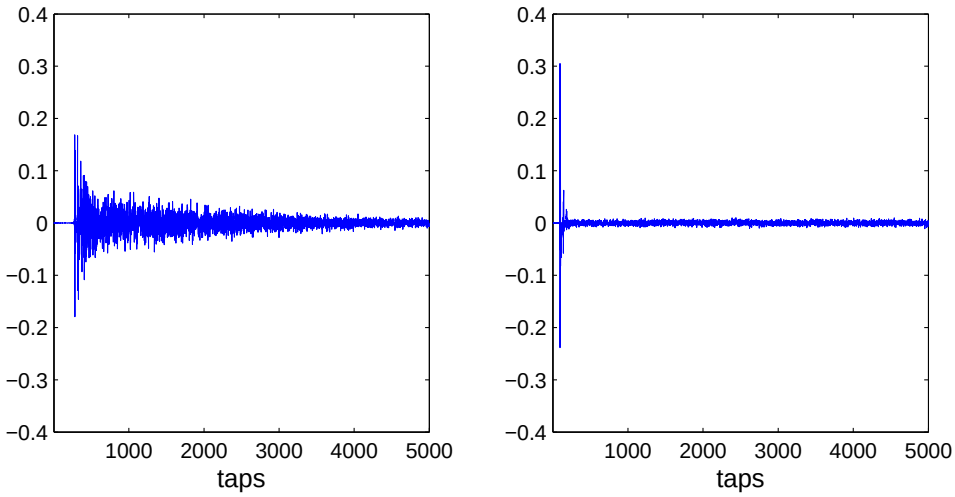


Fig. 17. First 5000 taps of (a) one of the measured room impulse responses of the mixing system $\mathbf{H}$ and (b) impulse response of the overall system $\mathbf{C}$ after convergence.

the first three seconds of the signals in Fig. 18. Both representations illustrate a significant enhancement of the speech signals. The spectrograms were computed as sequences of DFTs of windowed data segments. In this example, the Hamming window length was chosen to be 20 ms, as it is typical in speech analysis. This is short enough so that any single 20 ms frame will typically contain data from only one phoneme, yet long enough that it will include at least two periods of the fundamental frequency during voiced speech assuming the lowest voiced pitch to be around 100 Hz.

As mentioned in Sections 2 and 3, the correct choice of the filter length is an important issue in blind dereverberation, especially in the application of the identification-and-inversion approach. Hence, we now compare the DI and II approaches with respect to the sensitivity of overestimation of the filter lengths. Note that formally, according to Sect. 5.2, the TRINICON-based adaptation algorithm for blind system identification differs only slightly from

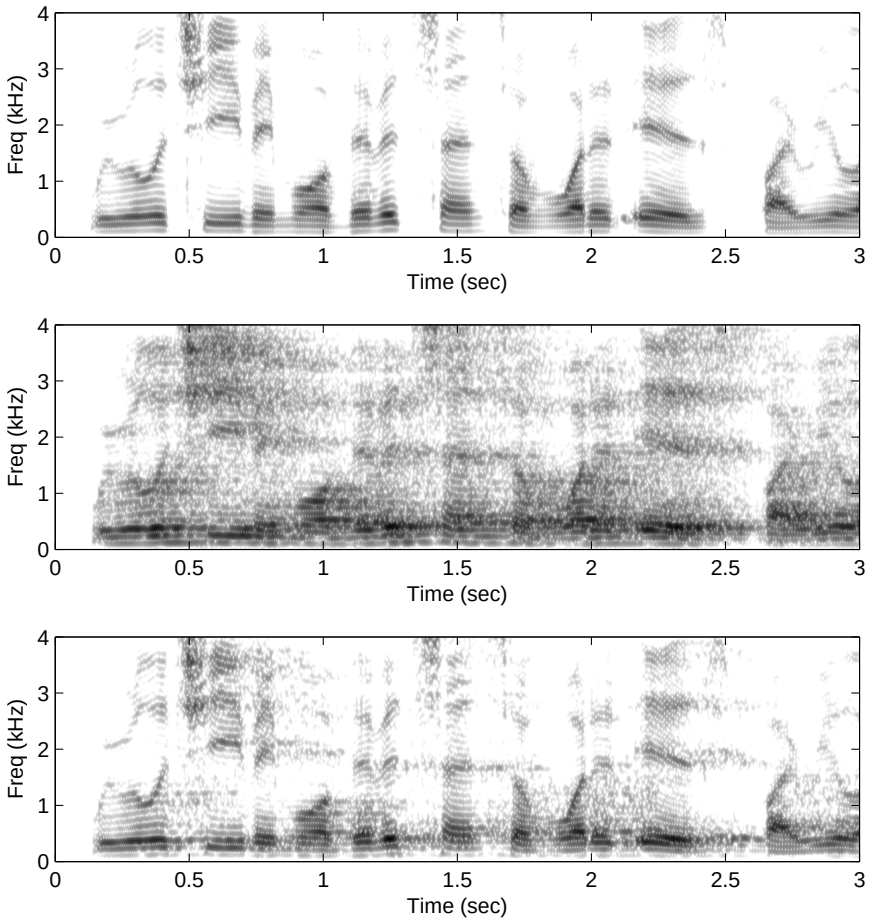


Fig. 18. Spectrograms for $0 \ldots 4$kHz of the first three seconds of (a) original source signal $s(n)$ (b) received signal $x_1(n)$ at microphone 1 and (c) output signal $y(n)$ after convergence.

the corresponding MCBPD algorithm (e.g., (105)): The sign of the update term is changed and the relation between the filter coefficients and the estimates of the mixing system, i.e., (79), has to be taken into account. Moreover, in the II approach to dereverberation, additionally, the application of MINT (17) is required to calculate the demixing system based on the estimated mixing system. These modifications were made in (106) for our next experiment

comparing the II approach with the DI approach (using (106) without modifications).

To allow for a fair comparison between the two different approaches, we assumed the same mixing system with only two sensor channels in both cases. For this experiment, the mixing system was composed of two very simple artificially created impulse responses in order to guarantee the avoidance of common zeros (or even near common zeros), as shown in Fig. 19. Hence, as long as the optimal filter length is chosen, this SIMO system is guaranteed to be invertible, which we also confirmed by applying MINT in a supervised manner. Table 2 shows the results of the blind estimation in terms of SRR

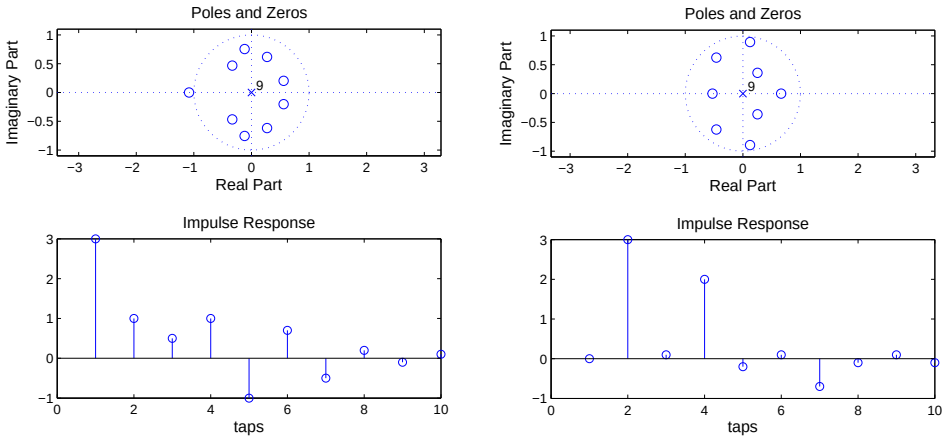


Fig. 19. Poles and zeros in the z-domain of subfilters $\mathbf{h}_1$ (left) and $\mathbf{h}_2$ (right) of a simple SIMO mixing system without common zeros.

improvement for both the DI and II approaches for different demixing filter lengths, and without any additional repair measures mentioned in Sect. 3.5. Note that in this experiment we chose n_Δ in the above SRR definition equal to the delay of the main peaks of the impulse responses due to their short lengths. Obviously, the numerical results confirm that with both approaches the best performance is obtained by choosing the optimal filter length according to Sections 2 and 3. Moreover, the results clearly show that the direct-inverse approach is significantly more robust to overestimation of the filter length. On the other hand, however, we have to note that the potential applicability of the identification-and-inversion approach is more general as the distinction between the speech production system and the room acoustics is not required in this case.

	$L \approx 80\% L_{\text{opt}}$	$L = L_{\text{opt}}$	$L \approx 120\% L_{\text{opt}}$	$L \approx 140\% L_{\text{opt}}$	$L \approx 150\% L_{\text{opt}}$
DI	29.8dB	31.2dB	27.3dB	24.1dB	22.4dB
II	22.0dB	25.4dB	9.6dB	4.5dB	0.2dB

Table 2. Comparison of the Direct-Inverse (DI) approach to blind dereverberation with the Identification-and-Inversion (II) approach with respect to the sensitivity of overestimation of the filter length for the simple example $M = 10$, $P = 2$, $L_{\text{DI,opt}} = 9$, $L_{\text{II,opt}} = 10$.

7.2 MIMO case

Finally, we extend the investigation of MCBPD for the direct-inverse approach to the MIMO case. We again consider the same acoustic scenario with $T_{60} \approx 700$ms, as described above for the SIMO case. In the following experiment there are two active speakers (one male speaker and one female speaker). The configuration is symmetric w.r.t. to the linear microphone array. We again apply the coefficient update (106) using the correlation method and the same parameter settings as described for the SIMO case. Figure 20 shows both

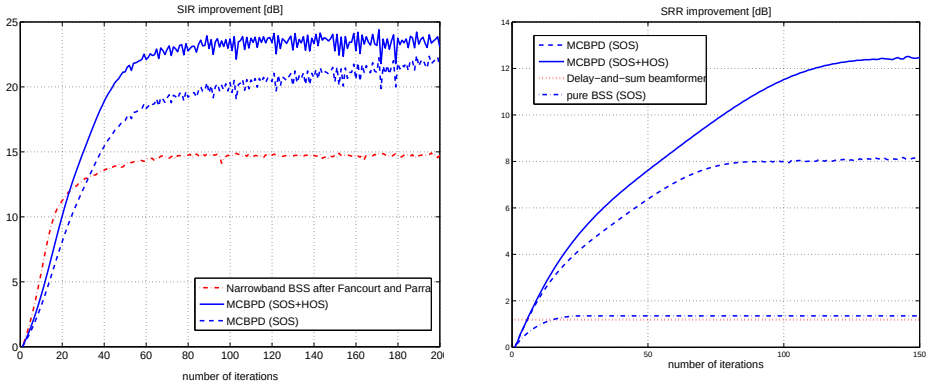


Fig. 20. SIR and SRR performance of MIMO-based MCBPD.

the improvement of the signal-to-interference ratio (i.e., source separation at the ouputs) and the improvement of the signal-to-reverberation ratio. The SIR and SRR curves were averaged between the contributions from the two sources. Similar to the SIMO case, TRINICON provides synergies between the SOS-based adaptation and the HOS-based adaptation. This advantage

can be seen in both the separation and the dereverberation performances. We also confirm that the SRR improvement is generally upper bounded by the SIR improvement. It is remarkable that the SRR improvements in the MIMO case are only slightly lower than those in the SIMO case. As reference, we also included the SIR convergence curve of the popular narrowband BSS algorithm after Fancourt and Parra [35] which is based on SOS (see also Sect. 5.3). We see that the initial convergence of the rigorously derived broadband approach is well comparable with that of the narrowband algorithm, while the final SIR performance is significantly higher. The reference curve for a pure separation algorithm based on SOS ([20] as a special case of (106) with $n_A = L - 1$ according to Fig. 14, $N = L$, and using only the first term in the brackets) in the SRR plot, and the comparison with a conventional delay-and-sum beamformer confirms the high efficiency of the MCBPD extension presented in this chapter.

8 Conclusions

Based on the TRINICON framework for broadband adaptive MIMO filtering, we developed in this chapter a strictly analytical top-down approach to the problem of blind dereverberation of speech and audio signals. It was shown that this provides both a common framework for various existing and novel powerful blind dereverberation algorithms and allows for a direct comparison between the various algorithms and the different existing approaches to blind dereverberation.

Comparing the two fundamental approaches to blind dereverberation, i.e., the identification-and-inversion approach and the direct-inverse approach, we can summarize that in principle the II approach is suitable for arbitrary audio signals, however, on the downside, this flexibility w.r.t. the source signals implies a high sensitivity to overestimation of the optimum filter length and common zeros in z-domain representation of the mixing system paths, so that additional repair mechanisms are necessary. Moreover, the explicit MINT-based inversion of the estimated mixing matrix in the II approach increases the computational complexity. On the other hand, the direct-inverse approach avoids the two-step procedure and the related problems of the II approach, but requires more stringent stochastic model assumptions on the source signals in order to avoid whitening effects. Fortunately, the TRINICON framework inherently allows the incorporation of powerful source models leading to a high separation and dereverberation performance without distortions for signals like speech.

A Compact Derivation of the Gradient-Based Coefficient Update

For the following compact derivation, we formulate the TRINICON coefficient optimization criterion (39) in the following way:

$$\mathcal{J} = \hat{E}_{\text{long}} \left\{ \hat{E}_{\text{block}} \left\{ f \left(\mathbf{y}, \boldsymbol{\mathcal{Q}}^{(1)}, \boldsymbol{\mathcal{Q}}^{(2)}, \ldots \right) \right\} \right\} \tag{119}$$

with

$$f = -\left(\log \hat{p}_{\text{s},PD}(\mathbf{y}) - \log \hat{p}_{\text{y},PD}(\mathbf{y}) \right) \tag{120}$$

and the operators $\hat{E}_{\text{block}}\{a\} = \frac{1}{N} \sum_{j=iN_L}^{iN_L+N-1} a(j)$ for averaging within each block, and $\hat{E}_{\text{long}}\{b\} = \sum_{i=0}^{\infty} \beta(i,m) \cdot b(i)$ over multiple blocks depending on the choice of the function β. The set of quantities

$$\boldsymbol{\mathcal{Q}}^{(r)} = \hat{E}_{\text{block}} \left\{ \boldsymbol{\mathcal{G}}^{(r)}(\mathbf{y}) \right\}, \quad r = 1, 2, \ldots, \tag{121}$$

(where $\boldsymbol{\mathcal{G}}^{(r)}$ are suitable functions of the observation vectors $\mathbf{y}$) contains all stochastic model parameters $\boldsymbol{\mathcal{Q}}_s^{(\cdot)}$ and $\boldsymbol{\mathcal{Q}}_y^{(\cdot)}$ according to (41) determining $\hat{p}_{s,PD}(\cdot)$ and $\hat{p}_{y,PD}(\cdot)$, respectively.

The gradient of (119) w.r.t. $\check{\mathbf{W}}$ reads according to (43) (omitting the iteration index here for simplicity):

$$\Delta\check{\mathbf{W}} = \hat{E}_{\text{long}} \left\{ \mathcal{SC} \left\{ \hat{E}_{\text{block}} \left\{ \frac{\partial}{\partial \mathbf{W}} f \left(\mathbf{y}, \boldsymbol{\mathcal{Q}}^{(1)}, \boldsymbol{\mathcal{Q}}^{(2)}, \ldots \right) \right\} \right\} \right\}. \tag{122}$$

We now apply the general multivariate chain rule:

$$\frac{\partial}{\partial \mathbf{W}} f \left(\mathbf{y}, \boldsymbol{\mathcal{Q}}^{(1)}, \boldsymbol{\mathcal{Q}}^{(2)}, \ldots \right) = \sum_i \frac{\partial \left[\mathbf{y}\right]_i}{\partial \mathbf{W}} \frac{\partial f}{\partial \left[\mathbf{y}\right]_i} + \sum_r \sum_{i_1,i_2,\ldots} \frac{\partial \mathcal{Q}_{i_1,i_2,\ldots}^{(r)}}{\partial \mathbf{W}} \frac{\partial f}{\partial \mathcal{Q}_{i_1,i_2,\ldots}^{(r)}}, \tag{123}$$

where $\mathcal{Q}_{i_1,i_2,\ldots}^{(r)}$ denote the elements of $\boldsymbol{\mathcal{Q}}^{(r)}$. Analogously $\mathcal{G}_{i_1,i_2,\ldots}^{(r)}$ denote the elements of $\boldsymbol{\mathcal{G}}^{(r)}$. The derivatives in the second term w.r.t. $\mathbf{W}$ can be expressed as

$$\frac{\partial \mathcal{Q}_{i_1,i_2,\ldots}^{(r)}}{\partial \mathbf{W}} = \hat{E}_{\text{block}} \left\{ \frac{\partial}{\partial \mathbf{W}} \mathcal{G}_{i_1,i_2,\ldots}^{(r)}(\mathbf{y}) \right\} = \hat{E}_{\text{block}} \left\{ \sum_i \frac{\partial \mathcal{G}_{i_1,i_2,\ldots}^{(r)}}{\partial \left[\mathbf{y}\right]_i} \frac{\partial \left[\mathbf{y}\right]_i}{\partial \mathbf{W}} \right\}. \tag{124}$$

With the MIMO relation $\mathbf{y} = \mathbf{W}^T\mathbf{x}$ and with (124) we obtain[10] from (123)

[10] Since in element-wise formulation, $[\mathbf{y}]_i = \sum_\ell [\mathbf{x}]_\ell [\mathbf{W}]_{\ell i}$, we obtain $\frac{\partial [\mathbf{y}]_i}{\partial [\mathbf{W}]_{jk}} = \sum_\ell [\mathbf{x}]_\ell \delta_{j\ell}\delta_{ki} = [\mathbf{x}]_j \delta_{ki}$, and thus $\left[\sum_i \frac{\partial [\mathbf{y}]_i}{\partial [\mathbf{W}]_{jk}} \frac{\partial f}{\partial [\mathbf{y}]_i}\right] = \left[\sum_i [\mathbf{x}]_j \delta_{ki} \frac{\partial f}{\partial [\mathbf{y}]_i}\right] = \left[[\mathbf{x}]_j \frac{\partial f}{\partial [\mathbf{y}]_k}\right] = \mathbf{x}\frac{\partial f}{\partial \mathbf{y}^T}$.

$$\frac{\partial}{\partial \mathbf{W}} f\left(\mathbf{y}, \boldsymbol{\mathcal{Q}}^{(1)}, \boldsymbol{\mathcal{Q}}^{(2)}, \ldots\right) = \mathbf{x}\frac{\partial f}{\partial \mathbf{y}^T} + \sum_r \sum_{i_1,i_2,\ldots} \hat{E}_{\text{block}}\left\{\mathbf{x}\frac{\partial \mathcal{G}^{(r)}_{i_1,i_2,\ldots}}{\partial \mathbf{y}^T}\right\} \frac{\partial f}{\partial \mathcal{Q}^{(r)}_{i_1,i_2,\ldots}}. \tag{125}$$

By introducing this equation into (122), we obtain

$$\begin{aligned}
\Delta\check{\mathbf{W}} = \hat{E}_{\text{long}}\Bigg\{\mathcal{SC}\Bigg\{&\hat{E}_{\text{block}}\left\{\mathbf{x}\frac{\partial f}{\partial \mathbf{y}^T}\right\} \\
&+ \sum_r \sum_{i_1,i_2,\ldots} \hat{E}_{\text{block}}\left\{\mathbf{x}\frac{\partial \mathcal{G}^{(r)}_{i_1,i_2,\ldots}}{\partial \mathbf{y}^T}\right\} \hat{E}_{\text{block}}\left\{\frac{\partial f}{\partial \mathcal{Q}^{(r)}_{i_1,i_2,\ldots}}\right\}\Bigg\}\Bigg\} \\
= \hat{E}_{\text{long}}\Bigg\{\mathcal{SC}\Bigg\{&\hat{E}_{\text{block}}\Bigg\{\mathbf{x}\Bigg(\frac{\partial f}{\partial \mathbf{y}^T} \\
&+ \sum_r \sum_{i_1,i_2,\ldots} \frac{\partial \mathcal{G}^{(r)}_{i_1,i_2,\ldots}}{\partial \mathbf{y}^T} \hat{E}_{\text{block}}\left\{\frac{\partial f}{\partial \mathcal{Q}^{(r)}_{i_1,i_2,\ldots}}\right\}\Bigg)\Bigg\}\Bigg\}\Bigg\}.
\end{aligned} \tag{126}$$

With (120) the last expression finally leads to the gradient-based update (44).

B Transformation of the Multivariate Output Signal PDF in (39) by Blockwise Sylvester Matrix

Due to the linear MIMO relation

$$\mathbf{y}^{\mathrm{T}}(n) = \mathbf{x}^{\mathrm{T}}(n)\mathbf{W}(n) \tag{127}$$

after (31) we express the PD-variate output log-likelihood $\log(\hat{p}_{\mathbf{y},PD}(\mathbf{y}(n)))$ in (39) in terms of the $2PL \times PD$ MIMO coefficient matrix $\mathbf{W}$ and the corresponding multivariate input pdf.

Since in general, $\mathbf{W}$ is not quadratic ($D \leq L$), we cannot immediately apply the well-known relation between the pdfs of two linearly related vectors via the determinant of a quadratic mapping matrix [87]. However, in our case $2PL > PD$, i.e., for 'tall' matrices $\mathbf{W}$ we can form a joint pdf $\hat{p}_{\mathbf{y}\tilde{\mathbf{x}},2LP}(\mathbf{y}(n), \tilde{\mathbf{x}}(n))$ of the output vector $\mathbf{y}$ and certain elements $\tilde{\mathbf{x}}$ of the input vector $\mathbf{x}$ so that this joint pdf exhibits the same dimensionality as the input pdf $\hat{p}_{\mathbf{x},2LP}(\mathbf{x}(n))$. Then, after the transformation

$$\hat{p}_{\mathbf{y}\tilde{\mathbf{x}},2LP}(\mathbf{y}(n), \tilde{\mathbf{x}}(n)) = \frac{\hat{p}_{\mathbf{x},2LP}(\mathbf{x}(n))}{\left|\det\tilde{\mathbf{W}}\right|} \tag{128}$$

with a quadratic $2LP \times 2LP$ matrix $\tilde{\mathbf{W}}$, the desired multivariate output pdf $\hat{p}_{\mathbf{y},PD}(\mathbf{y}(n))$ is obtained without loss of generality as a marginal density by integration for $\tilde{\mathbf{x}}(n)$ [87].

In our application a *channel-wise* extension of matrix $\mathbf{W}$ is desirable so that the MIMO relation (127)

$$\left[\mathbf{y}_1^{\mathrm{T}}, \ldots, \mathbf{y}_P^{\mathrm{T}}\right] = \left[\mathbf{x}_1^{\mathrm{T}}, \ldots, \mathbf{x}_P^{\mathrm{T}}\right] \begin{bmatrix} \mathbf{W}_{11} & \cdots & \mathbf{W}_{1P} \\ \vdots & \ddots & \vdots \\ \mathbf{W}_{P1} & \cdots & \mathbf{W}_{PP} \end{bmatrix}$$

may be extended to

$$\left[\mathbf{y}_1^{\mathrm{T}}, \tilde{\mathbf{x}}_1^{\mathrm{T}}, \ldots, \mathbf{y}_P^{\mathrm{T}}, \tilde{\mathbf{x}}_P^{\mathrm{T}}\right] = \left[\mathbf{x}_1^{\mathrm{T}}, \ldots, \mathbf{x}_P^{\mathrm{T}}\right] \tilde{\mathbf{W}}, \tag{129}$$

where $\tilde{\mathbf{x}}_p$, $p = 1, \ldots, P$ denote vectors containing the $2L - D$ last elements of $\mathbf{x}_p$ and

$$\tilde{\mathbf{W}} = \begin{bmatrix} \mathbf{W}_{11} & \begin{bmatrix} \mathbf{0}_{D\times 2L-D} \\ \mathbf{I}_{2L-D\times 2L-D} \end{bmatrix} & \cdots & \mathbf{W}_{1P} & \mathbf{0}_{2L\times 2L-D} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{W}_{P1} & \mathbf{0}_{2L\times 2L-D} & \cdots & \mathbf{W}_{PP} & \begin{bmatrix} \mathbf{0}_{D\times 2L-D} \\ \mathbf{I}_{2L-D\times 2L-D} \end{bmatrix} \end{bmatrix}. \tag{130}$$

With (128) we obtain

$$\begin{aligned} \hat{p}_{\mathbf{y},PD}(\mathbf{y}(n)) &= \frac{1}{\left|\det\tilde{\mathbf{W}}\right|} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \hat{p}_{\mathbf{x},2LP}(\mathbf{x}(n)) \mathrm{d}\tilde{\mathbf{x}}_1 \cdot \ldots \cdot \mathrm{d}\tilde{\mathbf{x}}_P \\ &= \frac{1}{\left|\det\tilde{\mathbf{W}}\right|} \hat{p}_{\mathbf{x}_{PD},PD}(\mathbf{x}_{PD}(n)), \end{aligned} \tag{131}$$

which leads to the following simple expression for the desired log-likelihood:

$$\log \hat{p}_{\mathbf{y},PD}(\mathbf{y}(n)) = \log \hat{p}_{\mathbf{x}_{PD},PD}(\mathbf{x}_{PD}(n)) - \log \left|\det\tilde{\mathbf{W}}\right|. \tag{132}$$

Since the first term on the right hand side of (132) does not depend on the filter coefficients, it does not need to be considered further for the gradient of the optimization criterion (39). To simplify the important second term in (132) together with $\tilde{\mathbf{W}}$ from (130) we exploit the fact that we can exchange colums or rows of $\tilde{\mathbf{W}}$ without changing the value of $|\det\tilde{\mathbf{W}}|$. Application of the general matrix relation

$$\det \begin{bmatrix} \mathbf{A}_1 & \mathbf{0} \\ \mathbf{A}_2 & \mathbf{I} \end{bmatrix} = \det\mathbf{A}_1 \tag{133}$$

immediately leads then to the compact formulation

$$\log \hat{p}_{\mathbf{y},PD}(\mathbf{y}(n)) = \log \hat{p}_{\mathbf{x}_{PD},PD}(\mathbf{x}_{PD}(n)) - \log \left|\det\left\{\mathbf{V}^{\mathrm{T}}\mathbf{W}\right\}\right| \tag{134}$$

with the window matrix $\mathbf{V}$ defined in (46). Note that $\mathbf{V}^{\mathrm{T}}\mathbf{W}$ is only of dimension $DP \times DP$.

C Polynomial Expansions for Nearly Gaussian Probability Densities

C.1 Orthogonal Polynomials

Let I be a finite or infinite interval and $r(x)$ be a continuous and positive function (which we call here *weighting function*) on the interval such that $\int_I f(x)r(x)\mathrm{d}x$ exists for every *polynomial* $f(x)$. Then there is a unique set of polynomials $P_n(x)$, $n = 0, 1, ...$ of order n such that

$$\int_I P_k(x)P_n(x)r(x)\mathrm{d}x := \langle P_k, P_n\rangle_r = c_n\,\delta_{kn} \tag{135}$$

with a predefined constant c_n. These polynomials $P_n(x)$ are called *orthogonal polynomials*. The operation $\langle\cdot,\cdot\rangle_r$ denotes the *inner product* in the vector space of the polynomials.

An important class of orthogonal polynomials in our context are the so-called Chebyshev-Hermite polynomials $P_{\mathrm{H},n}(x)$ which are specified by $I = (-\infty, \infty)$, the weighting function $r(x) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2}$, and $c_n = n!$, e.g., [41].

For the orthogonal polynomials considered here there is an important *proposition* stating that they even form a basis in a Hilbert space so that any quadratically integrable function $f(x)$ w.r.t. $r(x)$ on I can be expressed by the expansion, e.g., [41]

$$f(x) = \sum_{n=0}^{\infty} \frac{1}{c_n}\,\langle f, P_n\rangle_r\;\; P_n(x). \tag{136}$$

C.2 Polynomial Expansion for Univariate Densities

The two different expansions that are usually used to obtain a parameterized representation of nearly Gaussian probability density functions are the Edgeworth and the Gram-Charlier expansions, e.g., [2]. They lead to very similar approximations, so we only consider in this chapter the Gram-Charlier expansion. These expansions are based on the above-mentioned Chebyshev-Hermite polynomials $P_{\mathrm{H},n}(x)$.

Let $p(x) = \frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{x^2}{2\sigma^2}}\tilde{p}\left(\frac{x}{\sigma}\right)$ represent an arbitrary univariate probability density, where $\tilde{p}(\cdot)$ contains the higher-order contributions. According to (136) the higher-order statistics contribution $\tilde{p}$ can readily be expanded as

$$\tilde{p}(x) = \sum_{n=0}^{\infty} a_n P_{\mathrm{H},n}(x), \tag{137a}$$

$$a_n = \frac{1}{n!}\int_{-\infty}^{\infty} \tilde{p}(x')P_{\mathrm{H},n}(x')\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x'^2/2}\mathrm{d}x'. \tag{137b}$$

Hence, the complete density function $p(x)$ is finally expressed as

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{x^2}{2\sigma^2}} \sum_{n=0}^{\infty} a_n P_{\mathrm{H},n}\left(\frac{x}{\sigma}\right). \tag{138a}$$

The coefficients a_n after (137b) can be compactly written using the expectation operator:

$$a_n = \frac{1}{n!} E\left\{P_{\mathrm{H},n}\left(\frac{x}{\sigma}\right)\right\}. \tag{138b}$$

Example: Fourth-order approximation for a zero-mean process.

To obtain explicit expressions for the coefficients (138b), the Chebyshev-Hermite can be calculated using the derivatives of the standardized Gaussian probability density function (corresponding to the weighting function $r(x)$):

$$P_{\mathrm{H},n}(x) = (-1)^n \frac{1}{r(x)} \frac{\partial^n r(x)}{\partial x^n} \tag{139}$$

so that $P_{\mathrm{H},0}(x) = 1$, $P_{\mathrm{H},1}(x) = x$, $P_{\mathrm{H},2}(x) = x^2 - 1$, $P_{\mathrm{H},3}(x) = x^3 - 3x$, $P_{\mathrm{H},4}(x) = x^4 - 6x^2 + 3$. The resulting expansion coefficients for zero-mean processes are $a_0 = 1$, $a_1 = a_2 = 0$, $a_3 = \frac{E\{x^3\}}{3!\sigma^3}$, $a_4 = \frac{1}{4!}\left(\frac{E\{x^4\}}{\sigma^4} - 3\right)$ so that

$$p(x) \approx \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{x^2}{2\sigma^2}} \left(1 + \frac{\kappa_3}{3!\,\sigma^3} P_{\mathrm{H},3}\left(\frac{x}{\sigma}\right) + \frac{\kappa_4}{4!\,\sigma^4} P_{\mathrm{H},4}\left(\frac{x}{\sigma}\right)\right) \tag{140}$$

with [88] the *skewness* $\kappa_3 = E\{x^3\}$ and the *kurtosis* $\kappa_4 = E\{x^4\} - 3\sigma^4$. In the context of higher-order statistics-based estimation the kurtosis plays a particularly prominent role since it indicates whether a pdf is supergaussian ($\kappa_4 > 0$) or subgaussian ($\kappa_4 < 0$).

C.3 Multivariate Orthogonal Polynomials

Based on the previous subsection we may now generalize the Gram-Charlier expansion to multivariate probability density functions for a vector $\mathbf{x}$ of length D.

We formulate the orthogonality relation analogously to (135),

$$\int_{I^D} P_{\mathbf{k}}(\mathbf{x}) P_{\mathbf{n}}(\mathbf{x}) r(\mathbf{x}) \mathrm{d}\mathbf{x} = c_{\mathbf{n}}\, \delta_{\mathbf{kn}} \tag{141}$$

and the inner product

$$\langle f, g \rangle_r := \int_{I^D} f(\mathbf{x}) g(\mathbf{x}) r(\mathbf{x}) \mathrm{d}\mathbf{x}. \tag{142}$$

The D-variate Chebyshev-Hermite polynomials are specified by the D-variate weighting function [89]

$$r(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^D}} \mathrm{e}^{-\|\mathbf{x}\|_2^2/2} = \prod_{i=1}^{D} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x_i^2/2} = \prod_{i=1}^{D} r_1(x_i). \tag{143}$$

As we can see, in this case we have a *product weighting function*. It can be shown [89] that this has the very advantageous consequence that it also leads to corresponding *product polynomials*

$$P_{\mathbf{n}}(\mathbf{x}) = \prod_{i=1}^{D} P_{i,n_i}(x_i). \tag{144}$$

Note that $\mathbf{n}$ denotes a vector of indices n_i, $i = 1, \ldots, D$. The expansion of a multivariate function $f(\mathbf{x})$ is then given as

$$f(\mathbf{x}) = \sum_{\mathbf{n}=\mathbf{0}}^{\infty} \frac{1}{c_{\mathbf{n}}} \langle f, P_{\mathbf{n}} \rangle_r \; P_{\mathbf{n}}(\mathbf{x}). \tag{145}$$

C.4 Polynomial Expansion for Multivariate Densities

Let $p(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^D \det \mathbf{R}_{\mathbf{xx}}}} \mathrm{e}^{-\frac{1}{2}\mathbf{x}^T {\mathbf{R}_{\mathbf{xx}}}^{-1} \mathbf{x}} \tilde{p}\left(\mathbf{L}^{-1}\mathbf{x}\right)$ represent an arbitrary D-variate probability density, where $\tilde{p}(\cdot)$ again contains the higher-order contributions, and $\mathbf{L}$ is obtained by the Cholesky decomposition $\mathbf{R}_{\mathbf{xx}} = \mathbf{L}^T\mathbf{L}$ (note that $\sqrt{\mathbf{x}^T {\mathbf{R}_{\mathbf{xx}}}^{-1} \mathbf{x}} = \|\mathbf{L}^{-1}\mathbf{x}\|_2$).

In the same way as in the univariate case, we now obtain the following representation of a multivariate probability density function $p(\mathbf{x})$:

$$p(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^D \det \mathbf{R}_{\mathbf{xx}}}} \mathrm{e}^{-\frac{1}{2}\mathbf{x}^T {\mathbf{R}_{\mathbf{xx}}}^{-1} \mathbf{x}} \sum_{\mathbf{n}=\mathbf{0}}^{\infty} a_{\mathbf{n}} P_{\mathrm{H},\mathbf{n}}\left(\mathbf{L}^{-1}\mathbf{x}\right) \tag{146a}$$

with the coefficients

$$a_{\mathbf{n}} = \frac{1}{\prod_{i=1}^{D} n_i!} E\left\{P_{\mathrm{H},\mathbf{n}}(\mathbf{L}^{-1}\mathbf{x})\right\}. \tag{146b}$$

Note that $P_{\mathrm{H},\mathbf{n}}(\cdot)$ in (146a) and (146b) is given by (144).

D Expansion of the Sylvester Constraints in (83)

We consider here an expression with the Sylvester Constraint for one channel of the form

$$\mathbf{a}^T \mathcal{SC}\left\{\mathbf{b}\mathbf{c}^T\right\},$$

where $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ denote column vectors of length L, $2L$, and D, respectively. With the explicit expression of the generic Sylvester Constraint for one channel after Fig. 6 and [8],

$$[\mathbf{w}]_m = \sum_{k=1}^{2L} \sum_{\ell=1}^{D} [\mathbf{W}]_{k\ell}\, \delta_{k,(m+\ell-1)},$$

where δ_{ij} denotes the Kronecker symbol, the above expression reads

$$\sum_{m=1}^{L} a_m \sum_{k=1}^{2L} \sum_{\ell=1}^{D} b_k c_\ell \delta_{k,(m+\ell-1)} = \sum_{\ell=1}^{D} \sum_{m=1}^{L} a_m b_{m+\ell-1} c_\ell. \tag{147}$$

From the linearity of the operations, we easily deduce

$$\begin{aligned} &\mathbf{a}_1^T \mathcal{SC}\left\{\mathbf{b}_1 \mathbf{c}^T\right\} + \mathbf{a}_2^T \mathcal{SC}\left\{\mathbf{b}_2 \mathbf{c}^T\right\} \\ &= \sum_{\ell=1}^{D} \left(\sum_{m=1}^{L} a_{1,m} b_{1,m+\ell-1} + \sum_{m=1}^{L} a_{2,m} b_{2,m+\ell-1} \right) c_\ell. \end{aligned} \tag{148}$$

References

1. S. Haykin, *Adaptive Filter Theory*, 4th ed., Prentice-Hall, Englewood Cliffs, NJ, 2002.
2. A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, Wiley & Sons, Inc., New York, 2001.
3. 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.
4. J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non gaussian signals," *IEE Proceedings-F*, vol. 140, no. 6, pp. 362-370, Dec. 1993.
5. 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.
6. A. Lombard, T. Rosenkranz, H. Buchner, and W. Kellermann, "Multidimensional localization of multiple sound sources using averaged directivity patterns of blind source separation systems," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Taipei, Taiwan, April 2009.
7. 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.
8. H. Buchner, R. Aichner, and W. Kellermann, "TRINICON-based blind system identification with application to multiple-source localization and separation," in S. Makino, T.-W. Lee, and S. Sawada (eds.), *Blind Speech Separation*, Springer, Berlin, pp. 101-147, Sept. 2007.
9. 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.

10. 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.
11. 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.
12. 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.
13. 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.
14. 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, Aussois, France, Jan. 1999.
15. 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.
16. 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.
17. M. Hofbauer, *Optimal Linear Separation and Deconvolution of Acoustical Convolutive Mixtures*, Dissertation, Hartung-Gorre Verlag, Konstanz, May 2005.
18. 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.
19. 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.
20. 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.
21. H. Buchner and W. Kellermann, "A fundamental relation between blind and supervised adaptive filtering illustrated for blind source separation and acoustic echo cancellation," in *Proc. Joint Workshop Hands-Free Speech Communication and Microphone Arrays (HSCMA)*, Trento, Italy, May 2008.
22. D.A. Harville, *Matrix Algebra From A Statistician's Perspective,* Springer, New York, 1997.
23. 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.
24. 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.

25. 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.
26. J. Benesty, "Adaptive eigenvalue decomposition algorithm for passive acoustic source localization," *J. Acoust. Soc. Am.*, vol. 107, pp. 384-391, Jan. 2000.
27. 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.
28. 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.
29. Y. Huang, J. Benesty, and J. Chen, "Separation and dereverberation of speech signals with multiple microphones" in J. Benesty, S. Makino, and J. Chen (eds.), *Speech Enhancement*, Springer, Berlin, pp. 271-298, 2005.
30. S. Gannot and M. Moonen, "Subspace methods for multi-microphone speech dereverberation," *EURASIP J. Applied Signal Processing,* vol. 2003, no. 11, pp. 1074-1090, 2003.
31. W. Bobillet, E. Grivel, R. Guidorzi, and M. Najim, "Cancelling convolutive and additive coloured noises for speeech enhancement," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Montreal, Canada, vol. 2, pp. 777-780, Apr. 2004.
32. I. Santamaria, J. Via, and C.C.Gaudes, "Robust blind identification of SIMO channels: a support vector regression approach," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Montreal, Canada, vol. 5, pp. 673-676, Apr. 2004.
33. T. Hikichi and M. Miyoshi, "Blind algorithm for calculating the common poles based on linear prediction," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Montreal, Canada, vol. 4, pp. 89-92, Apr. 2004.
34. 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.
35. C.L. Fancourt and L. Parra, "The coherence function in blind source separation of convolutive mixtures of nonstationary signals," in *Proc. Int. Workshop Neural Networks Signal Processing (NNSP)*, 2001, pp. 303-312.
36. T.M. Cover and J.A. Thomas, *Elements of Information Theory*, Wiley & Sons, New York, 1991.
37. 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.
38. H. Buchner, R. Aichner, and W. Kellermann, "The TRINICON framework for adaptive MIMO signal processing with focus on the generic Sylvester constraint," *Proc. ITG Conf. on Speech Communication*, Aachen, Germany, Oct. 2008.
39. S. Amari and M. Kawanabe, "Information geometry of estimating functions in semiparametric statistical models," *Bernoulli*, vol. 2, no. 3, 1996.
40. R. Duda and P. Hart, *Pattern Classification and Scene Analysis*, John Wiley & Sons, 1973.
41. M.G. Kendal and A. Stuart, *The Advanced Theory of Statistics*, Vol. 1, 2nd Ed., Hafner Publishing Company, New York, NY, 1963.

42. P.J. Huber, *Robust Statistics*, Wiley, New York, 1981.
43. T. Gänsler, S.L. Gay, M.M. Sondhi, and J. Benesty, "Double-talk robust fast converging algorithms for network echo cancellation," in *IEEE Trans. Acoustics, Speech, and Language Processing*, Vol. 8, pp. 656-663, Nov. 2000.
44. 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.
45. J. Goldman, "Detection in the presence of spherically symmetric random vectors," *IEEE Trans. Inform. Theor.*, vol. 22, no. 1, pp. 52-59, Jan. 1976.
46. H. Brehm and W. Stammler, "Description and generation of spherically invariant speech-model signals," *Signal Processing*, vol. 12, pp. 119-141, 1987.
47. M. Kawamoto, K. Matsuoka, and N. Ohnishi, "A method of blind separation for convolved non-stationary signals," *Neurocomputing*, vol. 22, pp. 157-171, 1998.
48. K.V. Mardia, "Measures of multivariate skewness and kurtosis with applications," in *Biometrika*, vol. 57, pp. 519-530.
49. L. Ljung, *System Identification: Theory for the User*, Prentice-Hall, Englewood Cliffs, NJ, 1987.
50. S. Makino, T.-W. Lee, and S. Sawada (eds.), *Blind Speech Separation*, Springer, Berlin, pp. 101-147, Sept. 2007.
51. 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.
52. 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.
53. L. Molgedey and H.G. Schuster, "Separation of a mixture of independent signals using time delayed correlations," *Physical Review Letters,* vol. 72, pp. 3634-3636, 1994.
54. J.-F. Cardoso and A. Souloumiac, "Jacobi angles for simultaneous diagonalization," *SIAM J. Mat. Anal. Appl.,* vol. 17, no. 1, pp. 161-164, Jan. 1996.
55. S. Ikeda and N. Murata, "An approach to blind source separation of speech signals," *Proc. Int. Symposium on Nonlinear Theory and its Applications,* Crans-Montana, Switzerland, 1998.
56. L. Parra and C. Spence, "Convolutive blind source separation of non-stationary sources," *IEEE Trans. Speech and Audio Processing*, pp. 320-327, May 2000.
57. D.W.E. Schobben and P.C.W. Sommen, "A frequency-domain blind signal separation method based on decorrelation," *IEEE Trans on Signal Processing,* vol. 50, no. 8, pp. 1855-1865, Aug. 2002.
58. 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.
59. 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.
60. H. Buchner, J. Benesty, T. Gänsler, and W. Kellermann, "Robust extended multidelay filter and double-talk detector for acoustic echo cancellation," in *IEEE Trans. Acoustics, Speech, and Language Processing*, Sept. 2006.
61. J.C. Burgess, "Active adaptive sound control in a duct: A computer simulation," *J. Acoust. Soc. Am.*, vol. 70, no. 3, pp. 715-726, Sept. 1981.

62. 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.
63. 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.
64. 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.
65. 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.
66. P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, vol. 22, pp. 21-34, Jul. 1998.
67. R. Aichner, H. Buchner, and W. Kellermann, "On the causality problem in time-domain blind source separation and deconvolution algorithms," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, vol. 5, pp. 181-184, Philadelphia, PA, USA, Mar. 2005.
68. J.D. Markel and A.H.Gray, *Linear Prediction of Speech*, Springer, Berlin, 3rd edition, 1976.
69. M. Joho and P. Schniter, "Frequency domain realization of a multichannel blind deconvolution algorithms based on the natural gradient," in *Proc. Int. Symp. Independent Component Analysis and Blind Source Separation (ICA)*, pp. 15-26, Nara, Japan, Apr. 2003.
70. S. Douglas, H. Sawada, and S. Makino, "A causal frequency-domain implementation of a natural gradient multichannel blind deconvolution and source separation algorithm," in *Proc. Int. Congr. on Acoustics,* vol. 1, pp. 85-88, Kyoto, Japan, Apr. 2004.
71. S. Douglas, H. Sawada, and S. Makino, "Natural gradient multichannel blind deconvolution and source separation using causal FIR filters," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)* vol. 5, pp. 477-480, Montreal, Canada, May 2004.
72. L.-Q. Zhang, A. Cichocki, and S.-I. Amari, "Geometrical structures of FIR manifold and their application to multichannel blind deconvolution," in *Proc. IEEE Int. Workshop Neural Networks for Signal Processing (NNSP)*, pp. 303-312, Madison, WI, USA, Aug. 1999.
73. R.A. Wiggins, "Minimum entropy deconvolution," *Geoexploration*, vol 16, pp. 21-35, 1978.
74. R.H. Lambert, *Multichannel Blind Deconvolution: FIR Matrix Algebra and Separation of Multipath Mixtures*, Ph.D. dissertation, Univ. of Southern California, Los Angeles, CA, May 1996.
75. M.K. Broadhead and L.A. Pflug, "Performance of some sparseness criterion blind deconvolution methods in the presence of noise," *J. Acoust. Soc. Am.,* vol. 102, no. 2, pp. 885-893, Feb. 2000.
76. B.W. Gillespie, H.S. Malvar, and D.A.F. Florêncio,"Speech dereverberation via maximum-kurtosis subband adaptive filtering," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)* vol. 6, Salt Lake City, UT, USA, May 2001.

77. M. Brandstein, "On the use of explicit speech modeling in microphone array applications," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Seattle, WA, USA, pp. 3613-3616, May 1998.
78. B. Yegnanarayana and P.S. Murthy, "Enhancement of reverberant speech using LP residual signal," *IEEE Trans Speech Audio Processing,* vol. 8, no. 3, pp. 267-281, May 2000.
79. T. Yoshioka, T. Hikichi, and M. Miyoshi, "Dereverberation by using time-variant nature of speech production system," *EURASIP Journal on Advances in Signal Processing,* Vol. 2007.
80. J.V. Stone, "Blind deconvolution using temporal predictability," *Neurocomputing*, vol. 49, pp. 79-86, Dec. 2002.
81. K. Furuya and A. Kataoka, "Robust speech dereverberation using multichannel blind deconvolution with spectral subtraction," *IEEE Trans on Audio, Speech, Language Processing,* vol. 15, no. 5, pp. 1579-1591, Jul. 2007.
82. H. Kuttruff, *Room Acoustics*, Spon Press, London, 4th edition, 2000.
83. W. Reichardt, A. Alim, and W. Schmidt, "Definition und Messgrundlage eines objektiven Masses zur Ermittlung der Grenze zwischen brauchbarer und unbrauchbarer Durchsichtigkeit bei Musikdarbietung," *Acoustica*, no. 32, pp. 126-137, 1975. In German.
84. W.B. Kleijn and K.K. Paliwal, Eds., *Speech Coding and Synthesis,* Elsevier Science, Amsterdam, The Netherlands, 1995.
85. L. Rabiner and B.-H. Juang, *Fundamentals of Speech Recognition,* Prentice Hall, 1993.
86. P. Naylor and N.D. Gaubitch, "Speech dereverberation," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Eindhoven, The Netherlands, Sept. 2005.
87. A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 3rd edition, 1991.
88. Ch.L. Nikias and J.M. Mendel, "Signal processing with higher-order spectra," in *IEEE Signal Processing Magazine,* vol. 10, no. 3, pp. 10-37, July 1993.
89. Y. Xu, "Lecture notes on orthogonal polynomials of several variables" in W. zu Castell, F. Filbir, B. Forster (eds.), *Advances in the Theory of Spectral Functions and Orthogonal Polynomials*, Nova Science Publishers, Vol. 2, pp. 135-188, 2004.