On the Relation Between Blind System Identification and Subspace Tracking and Associated Generalizations

Herbert Buchner and Karim Helwani
Deutsche Telekom Laboratories, Technische Universität Berlin, Ernst-Reuter-Platz 7, 10587 Berlin, Germany
E-Mail: hb@buchner-net.com, Karim.Helwani@telekom.de

***Abstract*—Blind system identification and subspace tracking represent two important classes of signal processing problems with a variety of applications. Although originally seemingly independent from each other, the related algorithms exhibit various commonalities. In this paper, we present a novel unified derivation of the corresponding classes of adaptation algorithms. This top-down approach both clarifies the algorithmic relations and also leads to various powerful generalizations of the algorithms. Due to the rigorous approach, we obtain important practical design rules for an efficient system design. By exploiting multiple stochastic signal properties, the treatment also includes practically useful relations to blind signal extraction and blind source separation algorithms.**

I. Introduction

Subspace estimation and tracking plays an important role in various modern signal processing applications, e.g., in data compression, estimation of frequencies of sinusoids, denoising, direction-of-arrival estimation for multiple signals, etc. As a result, a variety of algorithms for subspace estimation have been proposed, e.g, [1], [2], [3], [4], [5]. These methods are typically based on eigenvalue decomposition of the correlation matrix of multichannel sensor signals or on singular value decomposition of the data matrix.

On the other hand, estimation methods exhibiting a very similar structure were proposed for the problem of blind system identification (BSI) for the case of single-input and multiple-output (SIMO) FIR systems in the context of blind deconvolution, e.g., [6], and of acoustic source localization in reverberant environments, e.g., [7]. Figure 1(a) shows a block diagram of this SIMO-based BSI approach. For this approach it can be shown that with sufficient excitation by the source signal $s(n)$ and with $e(n) \rightarrow 0$, i.e., the condition $h_1(n) * w_1(n) = -h_2(n) * w_2(n)$, the impulse responses $h_1(n)$ and $h_2(n)$ can be estimated uniquely up to a (frequency-independent) scaling factor α, so that, ideally,

$$w_1(n) = -\alpha \cdot h_2(n), \qquad w_2(n) = \alpha \cdot h_1(n), \tag{1}$$

as long as $h_1(n)$ and $h_2(n)$ do not share any common zeros in the z-domain and the filter length L is chosen correctly. In practice, the computation of the estimates of the filter coefficients $\mathbf{w} = \left[\mathbf{w}_1^{\mathrm{T}}, \mathbf{w}_2^{\mathrm{T}}\right]^{\mathrm{T}}$, $\mathbf{w}_p = [w_{p,0}, \ldots, w_{p,L-1}]^{\mathrm{T}}$, is typically based on the minimization of the error variance $\hat{\mathrm{E}}\left\{e^2\right\}$, where $e(n) = \check{\mathbf{x}}^{\mathrm{T}}(n)\mathbf{w}$ and $\check{\mathbf{x}}(n) = \left[\check{\mathbf{x}}_1^{\mathrm{T}}(n), \check{\mathbf{x}}_2^{\mathrm{T}}(n)\right]^{\mathrm{T}}$, $\check{\mathbf{x}}_p(n) = [x_p(n), \ldots, x_p(n-L+1)]^{\mathrm{T}}$. It is straightforward to show that this corresponds to the solution of the following homogeneous system of equations:

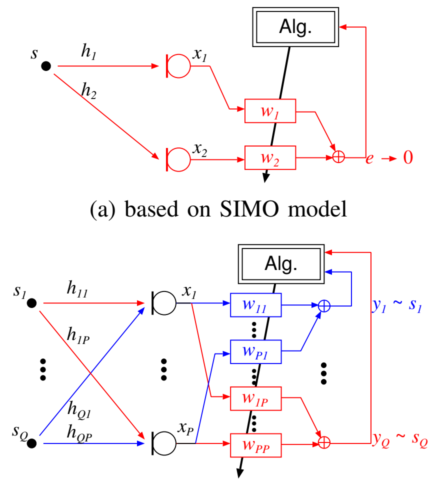
$$\frac{\partial}{\partial \mathbf{w}} \mathrm{E}\left\{e^2\right\} = \mathbf{R}_{\check{\mathbf{x}}\check{\mathbf{x}}}\mathbf{w} \overset{!}{=} \mathbf{0}, \tag{2}$$

where $\mathbf{R}_{\check{\mathbf{x}}\check{\mathbf{x}}}(n) = \hat{\mathrm{E}}\left\{\check{\mathbf{x}}(n)\check{\mathbf{x}}^{\mathrm{T}}(n)\right\}$ denotes the correlation matrix of the sensor signals.

Both the aforementioned subspace estimation methods and these SIMO-based BSI methods can thus be related to the calculation of eigenvectors $\mathbf{w}$ corresponding to *extreme eigenvalues* $\lambda_{\max}$ and $\lambda_{\min}$, respectively, of the correlation matrix $\mathbf{R}_{\check{\mathbf{x}}\check{\mathbf{x}}}(n)$ of multichannel sensor signals (Note that $\lambda \geq 0$ is guaranteed due to the positive-semidefiniteness of $\mathbf{R}_{\check{\mathbf{x}}\check{\mathbf{x}}}(n)$). In other words, the filter coefficients $\mathbf{w}$ can generally be expressed as solutions of the eigenvalue equation

$$\mathbf{R}_{\check{\mathbf{x}}\check{\mathbf{x}}}\mathbf{w} = \lambda \mathbf{w}. \tag{3}$$

Calculating the dominant eigenvectors of $\mathbf{R}_{\check{\mathbf{x}}\check{\mathbf{x}}}$ in this way is also well known as *principal component analysis (PCA)*, while the normal equation (2) (in the noiseless case) follows from (3) for $\lambda_{\min} = 0$. Calculating the eigenvectors corresponding to the minimum eigenvalues is also known as *minor component analysis (MCA)*.



(a) based on SIMO model

(b) based on MIMO model

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

In order to allow a *tracking* in time-varying scenarios and to reduce the complexity, various efficient *adaptive methods* were proposed independently in both areas (e.g., [1], [2] for basis estimation of a signal subspace (i.e., PCA) and, e.g., [7] for the estimation of SIMO system coefficients (i.e., MCA), respectively). Interestingly, in both cases, the estimation algorithms can be formulated similarly to the familiar form of recursive update equations as known from supervised adaptive filtering, e.g., the least-mean-square (LMS) algorithm[1]

$$\mathbf{w}(n) = \mathbf{w}(n-1) - \mu\check{\mathbf{x}}(n)e(n) \tag{4}$$

or recursive least-squares (RLS) algorithms [8].

Recently, the known approach of SIMO-BSI has been extended to the case of multiple-input and multiple-output (MIMO) systems by exploiting the close relationship between blind source separation (BSS) for broadband signals and blind system identification, e.g., [9],

[1]Here, the sign in (4) only formally differs from the LMS formulation for the supervised case in [8] due to a slightly different notation in the blind case.

see Fig. 1(b). Hence, using suitable broadband BSS algorithms, it is possible to perform MIMO-BSI. Similar to (1) in the SIMO case, we can express the ideal separation solution in the MIMO case [9] if we require the cancellation of all cross-channels of the overall system between the sources s_q and the outputs of the demixing system. For instance, in the 2×2 case, the ideal separation solution reads

$$\begin{bmatrix} \mathbf{w}_{11} & \mathbf{w}_{12} \\ \mathbf{w}_{21} & \mathbf{w}_{22} \end{bmatrix} = \begin{bmatrix} \alpha_1 \mathbf{h}_{22} & -\alpha_2 \mathbf{h}_{12} \\ -\alpha_1 \mathbf{h}_{21} & \alpha_2 \mathbf{h}_{11} \end{bmatrix}. \quad (5)$$

Analogously to (1) the separation solution is unique up to (frequency-independent) scaling (α_1, α_2), as long as h_{11} and h_{12} do not share any common zeros in the z-domain, and h_{21} and h_{22} do not share any common zeros in the z-domain, respectively, and the filter length L is not overestimated.

In order to blindly estimate the matrix $\check{\mathbf{W}}$ of demixing filter coefficients for BSS, the fundamental goal is to make the output signals y_1,...,y_P mutually statistically independent. Hence, in contrast to the principal component analysis, we need to apply the more general method of independent component analysis (ICA), e.g., [10]. To obtain suitable broadband algorithms, TRINICON ('TRIple-N ICA for CONvolutive mixtures'), e.g., [11], [12], [9], provides a generic concept for broadband adaptive MIMO filtering based on ICA. TRINICON simultaneously exploits all fundamental stochastic signal properties, i.e., nonwhiteness, nonstationarity, and nongaussianity of the source signals. Due to its generality, it is also a useful tool to deduce novel and improved algorithms, and to study the relations between various algorithms. Based on TRINICON, a simple relation between blind and supervised adaptive filtering was developed in [13].

In this paper, we develop explicit algorithmic relations between the coefficient update rules of broadband BSS/MIMO-BSI and (a.) SIMO-BSI and (b.) subspace tracking algorithms by suitably specializing the assumed MIMO mixing system. Similarly to [13], we will see that this top-down approach directly leads to both the known algorithms of the categories (a) and (b), but also novel and powerful generalizations of them which inherently exploit all fundamental stochastic signal properties, as mentioned above. These links also provide a practical avenue to the field of *blind signal extraction (BSE)* of dominant source signals from mixtures due to the exploitation of multiple stochastic properties, and clarify the relation between SIMO-BSI and subspace tracking. Due to the rigorous approach shown in this paper, we obtain various important design rules for practical system implementation (e.g., coefficient initialization) and for performance optimization.

II. A BRIEF RECAPITULATION OF TRINICON

In this section we first give a brief overview of the essential elements of TRINICON for the coefficient adaptation. Thereby, we restrict the presentation here to simple gradient-based coefficient updates in the time domain.

A. Optimization Criterion

Various approaches exist to estimate the demixing matrix $\check{\mathbf{W}}$ by utilizing the following fundamental source signal properties [10] which were all combined in TRINICON:

(i) Nongaussianity is exploited by using higher-order statistics for ICA. The minimization of the mutual information (MMI) among the output channels can be regarded as the most general approach to separation problems [10]. To obtain an estimator that is also suitable for inverse problem, TRINICON uses the Kullback-Leibler divergence (KLD) between a certain *desired* joint pdf (essentially representing a hypothesized stochastic source model as shown below) and the joint pdf of the actually estimated output signals.

(ii) Nonwhiteness is exploited by simultaneous minimization of output cross-relations over multiple time-lags. We therefore consider multivariate pdfs, i.e., 'densities covering 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.

Throughout this section, we formulate the framework for $Q = P$ sources without loss of generality. In practice, the current number of simultaneously active sources is allowed to vary throughout the application and only the conditions $Q \leq P$ (for separation only) and $Q < P$ (for deconvolution), respectively, have to be fulfilled.

To introduce an algorithm for broadband processing of convolutive mixtures, we first formulate the convolution of the FIR demixing system of length L in the following matrix form [11]:

$$\mathbf{y}(n) = \mathbf{W}^{\mathrm{T}} \mathbf{x}(n), \quad (6)$$

where n denotes the time index, and

$$\begin{aligned} \mathbf{x}(n) &= [\mathbf{x}_1^{\mathrm{T}}(n), \ldots, \mathbf{x}_P^{\mathrm{T}}(n)]^{\mathrm{T}}, && (7) \\ \mathbf{y}(n) &= [\mathbf{y}_1^{\mathrm{T}}(n), \ldots, \mathbf{y}_P^{\mathrm{T}}(n)]^{\mathrm{T}}, && (8) \\ \mathbf{x}_p(n) &= [x_p(n), \ldots, x_p(n-2L+1)]^{\mathrm{T}}, && (9) \\ \mathbf{y}_q(n) &= [y_q(n), \ldots, y_q(n-D+1)]^{\mathrm{T}}. && (10) \end{aligned}$$

The parameter D in (10), $1 \leq D < L$, denotes the number of time 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 in each column by successive shifting, i.e., the first column reads $\left[\mathbf{w}_{pq}^{\mathrm{T}}, 0, \ldots, 0\right]^{\mathrm{T}}$, the second column $\left[0, \mathbf{w}_{pq}^{\mathrm{T}}, 0, \ldots, 0\right]^{\mathrm{T}}$, etc. Finally, the $2PL \times PD$ matrix $\mathbf{W}$ combines all Sylvester matrices $\mathbf{W}_{pq}$.

Based on the KLD, the following cost function was introduced in [11] taking into account all three fundamental signal properties (i)-(iii):

$$\begin{aligned} \mathcal{J}(m, \mathbf{W}) &= -\sum_{i=0}^{\infty} \beta(i, m) \frac{1}{N} \\ &\cdot \sum_{j=iN_L}^{iN_L+N-1} \left\{ \log(\hat{p}_{s,PD}(\mathbf{y}(j))) - \log(\hat{p}_{y,PD}(\mathbf{y}(j))) \right\}, \quad (11) \end{aligned}$$

where $\hat{p}_{s,PD}(\cdot)$ and $\hat{p}_{y,PD}(\cdot)$ are assumed or estimated PD-variate source model (i.e., desired) pdf and output pdf, respectively. The index m denotes the block time index for a block of N output samples shifted by L samples relatively to the previous block. Furthermore, β is a window function allowing for online, offline, or block-online algorithms [12].

An *alternative formulation of the second term in the optimization criterion* (11) is obtained by using the mapping between the output pdf and the input pdf of the demixing filter which plays an important role for the following considerations in this paper. This mapping can be expressed as follows, e.g., [14]:

$$\hat{p}_{y,PD}(\mathbf{y}) = \frac{\hat{p}_{x_{PD},PD}(\mathbf{x}_{PD})}{|\det\{\mathbf{V}^{\mathrm{T}} \mathbf{W}\}|} \quad (12)$$

with the window matrix $\mathbf{V} = \mathrm{Bdiag}\{\tilde{\mathbf{V}}, \ldots, \tilde{\mathbf{V}}\}$, where $\tilde{\mathbf{V}} = \left[\mathbf{I}_{D \times D},\ \mathbf{0}_{D \times (2L-D)}\right]^{\mathrm{T}}$.

B. Gradient-Based Coefficient Update

For brevity and simplicity we concentrate in this subsection on iterative Euclidean gradient-based block-online coefficient updates which can be written in the general form

$$\check{\mathbf{W}}^0(m) := \check{\mathbf{W}}(m-1), \tag{13a}$$
$$\check{\mathbf{W}}^\ell(m) = \check{\mathbf{W}}^{\ell-1}(m) - \mu\Delta\check{\mathbf{W}}^\ell(m), \;\; \ell = 1,\ldots,\ell_{\max}, \tag{13b}$$
$$\check{\mathbf{W}}(m) := \check{\mathbf{W}}^{\ell_{\max}}(m), \tag{13c}$$

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 downwards pointing hat symbol on top of $\mathbf{W}$ in (13) serves to distinguish the *condensed* $PL \times Q$ demixing coefficient matrix $\check{\mathbf{W}}$ to be optimized, from the corresponding larger Sylvester matrix $\mathbf{W}$ in the cost function. The matrix $\check{\mathbf{W}}$ consists of the first column of each submatrix $\mathbf{W}_{pq}$ without the L zeros.

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 $\mathbf{W}$ are a direct consequence of taking into account the nonwhiteness signal property by choosing $D > 1$. The rigorous distinction between these different matrix structures is also 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 [12]. Using the Sylvester constraint operator the gradient descent update can be written as

$$\Delta\check{\mathbf{W}}^\ell(m) = \mathcal{SC}\left\{\nabla_{\mathbf{W}}\mathcal{J}(m,\mathbf{W})\right\}\big|_{\mathbf{W}=\mathbf{W}^\ell(m)}. \tag{14}$$

Depending on the particular realization of ($\mathcal{SC}$), we are able to select both, well known and also novel improved adaptation algorithms [14]. In [9] an explicit formulation of a *generic* Sylvester constraint was derived to further formalize and clarify this concept:

$$\left[\Delta\check{\mathbf{w}}^\ell_{pq}(m)\right]_i = \sum_{k,j}\left[\Delta\mathbf{W}^\ell_{pq}(m)\right]_{kj}\delta_{k,(i+j-1)}. \tag{15}$$

Here, δ_{ab} denotes the Kronecker symbol.

It can be shown [14] that by taking the gradient of $\mathcal{J}(m)$ with respect to the demixing filter matrix $\check{\mathbf{W}}(m)$ according to (14), we obtain the following generic gradient descent-based TRINICON update rule:

$$\Delta\check{\mathbf{W}}^\ell(m) = \frac{1}{N}\sum_{i=0}^{\infty}\beta(i,m)\,\mathcal{SC}\left\{\sum_{j=iN_L}^{iN_L+N-1}\left[\mathbf{x}(j)\mathbf{\Phi}^{\mathrm{T}}_{s,PD}(\mathbf{y}(j)) - \left(\left(\mathbf{W}^{\ell-1}(m)\right)^{\mathrm{T}}\right)^{+}\right]\right\}, \tag{16a}$$

with $\cdot^+$ denoting the pseudoinverse of a matrix, and with the generalized score function

$$\mathbf{\Phi}_{s,PD}(\mathbf{y}(j)) = -\frac{\partial\log\hat{p}_{s,PD}(\mathbf{y}(j))}{\partial\mathbf{y}(j)} - \frac{1}{N}\sum_r\sum_{i_1,i_2,\ldots}\frac{\partial\mathcal{G}^{(r)}_{s,i_1,i_2,\ldots}}{\partial\mathbf{y}}\sum_{j=iN_L}^{iN_L+N-1}\frac{\partial\log\hat{p}_{s,PD}}{\partial\mathcal{Q}^{(r)}_{s,i_1,i_2,\ldots}} \tag{16b}$$

resulting from the hypothesized source model $\hat{p}_{s,PD} = \hat{p}_{s,PD}(\mathbf{y},\boldsymbol{\mathcal{Q}}^{(1)}_s,\boldsymbol{\mathcal{Q}}^{(2)}_s,\ldots)$ with certain stochastic model parameters $\boldsymbol{\mathcal{Q}}^{(r)}_s, r = 1,2,\ldots$ (the calligraphic symbols denote multidimensional arrays) given by their elements $\mathcal{Q}^{(r)}_{s,i_1,i_2,\ldots}$ in the generic form $\mathcal{Q}^{(r)}_{s,i_1,i_2,\ldots}(i) = \frac{1}{N}\sum_{j=iN_L}^{iN_L+N-1}\left\{\mathcal{G}^{(r)}_{s,i_1,i_2,\ldots}(\mathbf{y}(j))\right\}$ with certain nonlinear functions $\mathcal{G}^{(r)}_{s,i_1,i_2,\ldots}(\mathbf{y})$, $r = 1,2,\ldots$. A well known special case of such a parameterization is the estimate of the correlation matrix $\mathbf{R}_{\mathbf{yy}}(i) = \frac{1}{N}\sum_{j=iN_L}^{iN_L+N-1}\left\{\mathbf{y}(j)\mathbf{y}^{\mathrm{T}}(j)\right\}$. The filter coefficients and the stochastic model parameters are estimated in an alternating way.

III. TRINICON FOR BSS AND MIMO-BSI

As already suggested in Sec. I, the case of broadband BSS/MIMO-BSI represents the most general setup considered in this paper so that it constitutes the common basis for the subsequent cases. In BSS, the aim is to achieve statistical independence between the output channels. Hence, the desired pdf is factorized w.r.t. the output channels, i.e.,

$$\hat{p}_{s,PD}(\mathbf{y}(j)) \overset{\text{(BSS)}}{=} \prod_{q=1}^{P}\hat{p}_{y_q,D}(\mathbf{y}_q(j)), \tag{17}$$

so that the desired score function simplifies to

$$\mathbf{\Phi}_{s,PD}(\mathbf{y}) \overset{\text{(BSS)}}{=} \left[\mathbf{\Phi}^{\mathrm{T}}_{1,D}(\mathbf{y}_1),\ldots,\mathbf{\Phi}^{\mathrm{T}}_{P,D}(\mathbf{y}_P)\right]^{\mathrm{T}}. \tag{18}$$

In other words, for each output channel the score function can be obtained individually from a certain choice of pdf. For illustration, the special case of algorithms based on second-order statistics (SOS) is obtained from choosing multivariate Gaussian source models leading to [12]

$$\mathbf{\Phi}_{q,D}(\mathbf{y}_q(j)) = \mathbf{R}^{-1}_{\mathbf{y}_q\mathbf{y}_q}(i)\mathbf{y}_q(j). \tag{19}$$

IV. RELATION TO MCA AND ADAPTIVE SIMO-BSI

In this section we show how to deduce the class of SIMO-based BSI (or minor component analysis) algorithms from TRINICON using the generic gradient-based update (16a). with the specialized score function (18) for separation and identification problems.

A. SIMO-BSI as a Specialization of the MIMO Case

The ideal separation filter matrix $\check{\mathbf{W}}_{\text{ideal,sep}}$ in the 2×2 case is given by (5). we now consider the SIMO mixing model in Fig. 1(a) as a specialization of the MIMO mixing model in Fig. 1(b), i.e., $\mathbf{h}_{11} \to \mathbf{h}_1$, $\mathbf{h}_{12} \to \mathbf{h}_2$, $\mathbf{h}_{21} \to \mathbf{0}$, $\mathbf{h}_{22} \to \mathbf{0}$.

According to the right-hand side of (5) the corresponding ideal *demixing system* taking into account this prior knowledge reads

$$\begin{bmatrix}\mathbf{w}_{11} & \mathbf{w}_{12}\\ \mathbf{w}_{21} & \mathbf{w}_{22}\end{bmatrix} = \alpha\begin{bmatrix}\mathbf{0} & -\mathbf{h}_2\\ \mathbf{0} & \mathbf{h}_1\end{bmatrix}. \tag{20}$$

By comparing both sides of this equation, we immediately obtain the corresponding demixing system structure shown on the right side in Fig. 1(a). This is indeed the well-known SIMO BSI/AED approach, which in this way follows rigorously from the general equation (5) 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 (16a).

Theorem 1: *In the case of SIMO mixing systems the expression* $\left(\mathbf{W}^{\mathrm{T}}\right)^+$ *in the second term of (16a) becomes equal to zero.*

Proof: From (12) immediately follows that if $\log\hat{p}_{\mathbf{y},PD}(\mathbf{y}(n)) = \text{const.}\;\forall\,\mathbf{W} \Rightarrow \log\left|\det\left\{\mathbf{V}^{\mathrm{T}}\mathbf{W}\right\}\right| = \text{const.}\;\forall\,\mathbf{W}$. Analogously, since $\mathbf{y} = \mathbf{W}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{s}$: if $\log\hat{p}_{\mathbf{y},PD}(\mathbf{y}(n)) = \text{const.}\;\forall\,\mathbf{W} \Rightarrow \log\left|\det\left\{\mathbf{W}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\right\}\right| = \text{const.}\;\forall\,\mathbf{W}$. Combining these two statements, we can say that if $\det\left\{\mathbf{W}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\right\} = \text{const.}\;\forall\,\mathbf{W} \Rightarrow \det\left\{\mathbf{V}^{\mathrm{T}}\mathbf{W}\right\} = \text{const.}\;\forall\,\mathbf{W}$, i.e., if $\det\left\{\mathbf{W}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\right\} = \text{const.}\;\forall\,\mathbf{W} \Rightarrow$ second term of the gradient is equal to zero.

Now, let $\mathbf{W} = \left[\mathbf{W}^{\mathrm{T}}_1,\ldots,\mathbf{W}^{\mathrm{T}}_P\right]^{\mathrm{T}}$ and $\mathbf{H} = [\mathbf{H}_1,\ldots,\mathbf{H}_P]$ be MISO and SIMO, respectively. Then, $\det\left\{\mathbf{W}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\right\} = \det\left\{\sum_{p=1}^{P}\mathbf{W}^{\mathrm{T}}_p\mathbf{H}^{\mathrm{T}}_p\right\}$. Since $\sum_{p=1}^{P}\mathbf{W}^{\mathrm{T}}_p\mathbf{H}^{\mathrm{T}}_p$ is upper triangular, we write $\det\left\{\mathbf{W}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\right\} = \left(\sum_{p=1}^{P} w_{p,0}h_{p,0}\right)^{N}$. For only one active source, only subfilter $\mathbf{w}_{p_{\mathrm{far}}}$ connected to the microphone with greatest distance to the source, may exhibit a nonzero value at its first tap weight, i.e., $w_{p,0} = \alpha\cdot\delta_{p,p_{\mathrm{far}}}$ This results in $\det\left\{\mathbf{W}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\right\} = (h_{p_{\mathrm{far}},0})^{N}$ = const. Hence, $\det\left\{\mathbf{V}^{\mathrm{T}}\mathbf{W}\right\}$ = const. $\forall\, \mathbf{W}$ and the second term in the update (16a) becomes equal to zero. ◇

Next, we consider the *first term* $\mathbf{x}(j)\mathbf{\Phi}^{\mathrm{T}}_{s,PD}(\mathbf{y}(j))$ in the coefficient update (16a) for the SIMO case and note that its second (block-)column reads $\mathbf{x}(j)\mathbf{\Phi}^{\mathrm{T}}_{y_2,D}(\mathbf{y_2}(j))$. 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., [8] (see Fig. 1(a) and Fig. 1(b)):

$$\mathbf{y}_2 \rightarrow \mathbf{e}, \qquad \begin{bmatrix} \mathbf{w}_{12} \\ \mathbf{w}_{22} \end{bmatrix} = \begin{bmatrix} -\hat{\mathbf{h}}_2 \\ \hat{\mathbf{h}}_1 \end{bmatrix} \rightarrow \mathbf{w} = \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \end{bmatrix}. \tag{21}$$

Hence, the second column of the first term of the coefficient update is finally expressed as $\mathbf{x}(j)\mathbf{\Phi}^{\mathrm{T}}_{e,D}(\mathbf{e}(j))$. Note that the substitution of the coefficient notation in (21) is justified by (20).

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

$$\mathbf{w}^{\ell}(m) = \\ = \mathbf{w}^{\ell-1}(m) - \frac{\mu}{N}\sum_{i=0}^{\infty}\beta(i,m)\,\mathcal{SC}\left\{\sum_{j=iN_L}^{iN_L+N-1}\mathbf{x}(j)\mathbf{\Phi}^{\mathrm{T}}_{e,D}(\mathbf{e}(j))\right\}. \tag{22}$$

This formally represents the *triple-N-generalization of the Least-Mean-Squares (LMS) algorithm* analogously to the supervised adaptive filtering theory, as discussed in [13]. Although we now performed different specializations of the generic TRINICON update, we formally obtained the same algorithmic structure. However, in the blind SIMO identification application, we additionally have to extract the final estimates $\hat{\mathbf{h}}_1$ and $\hat{\mathbf{h}}_2$ from the coefficient vector $\mathbf{w}(m)$ according to (21). This relation between LMS and blind SIMO identification corresponds to [7], where the adaptive eigenvalue decomposition (AED) problem was related to the traditional LMS algorithm of the simple form (4).

B. Normalization and Coefficient Initialization

The general relation between MIMO BSI and SIMO BSI 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 $\mathbf{w} = \mathbf{0}$ (Note that MCA was originally derived from the minimization of the error variance according to (2)).

In the literature, various constraints have been proposed to avoid the trivial solution, such as the unit norm constraint by renormalization after each step (e.g., [1], [7], [10]),

$$\mathbf{w}^{\ell}(m) = \frac{\mathbf{w}^{\ell-1}(m) - \mu\Delta\mathbf{w}^{\ell}(m)}{\|\mathbf{w}^{\ell-1}(m) - \mu\Delta\mathbf{w}^{\ell}(m)\|}. \tag{23}$$

In the following, we show how the convergence to the trivial solution can be avoided even without renormalization by suitable coefficient initialization.

Theorem 2: The algorithm (22) cannot converge to the trivial solution, as long as the initialization $\mathbf{w}(0)$ is not orthogonal to the ideal solution $\mathbf{w}_{\mathrm{ideal}} = \left[-\mathbf{h}_2^{\mathrm{T}}\ \mathbf{h}_1^{\mathrm{T}}\right]^{\mathrm{T}}$.

Proof: We pre-multiply the update (22) with $\mathbf{w}^{\mathrm{T}}_{\mathrm{ideal}}$ on both sides of the update equation:

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

Using (15) it can be shown that this expression can be expanded to

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

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

V. RELATION TO PCA, SUBSPACE TRACKING, AND BLIND SIGNAL EXTRACTION

A. Extraction of the First Principal Component or the First Independent Component

Assuming only one dominant source and starting the treatise again with a 2×2 BSS setup, we now cover the *complementary* case of Fig. 1(a). In other words, we are now interested in the first column of the demixing filter matrix

$$\check{\mathbf{W}} =: [\mathbf{w}\ *], \tag{24}$$

where $*$ denotes *don't cares*. According to the considerations in Sect. I, the selected matrix elements in the first column are related to the *signal subspace*, while the other column(s) span the *noise subspace*. To deduce an adaptation algorithm for this case, we apply again Theorem 1 to the second term of the generic gradient-based coefficient update. By picking now the first column of the demixing filter matrix according to (24), we again obtain the triple-N-generalization of the LMS algorithm:

$$\mathbf{w}^{\ell}(m) = \\ = \mathbf{w}^{\ell-1}(m) - \frac{\mu}{N}\sum_{i=0}^{\infty}\beta(i,m)\,\mathcal{SC}\left\{\sum_{j=iN_L}^{iN_L+N-1}\mathbf{x}(j)\mathbf{\Phi}^{\mathrm{T}}_{1,D}(\mathbf{y}_1(j))\right\}. \tag{25}$$

Note that due to the simultaneous exploitation of the stochastic signal properties (i)-(iii), this algorithm not only performs PCA but can be used for *blind signal extraction (BSE)*. Accordingly, in the latter case, the adaptive MISO system obtained by (25) can also be regarded as a self-steering beamformer.

Similar as in MCA, as discussed above, various constraints are discussed in the PCA literature, such as the unit norm constraint and unitarity constraints. For PCA, in some of the well known algorithms the unit norm constraint is also required to prevent divergence (e.g., variance *maximization* by gradient *ascent* adaptation), in contrast to MCA, as discussed above. Note that changing the sign of the TRINICON-based update in (25) will be absorbed into the arbitrary scaling factor α. Hence, to be in accordance with the PCA literature, we now consider the case with positive sign and norm constraint.

Approximation of the norm constraint, i.e., the denominator in (23) using a Taylor series expansion w.r.t. μ leads to the following expression according to [1], [10]:

$$\frac{\mathbf{w} - \mu\Delta\mathbf{w}}{\|\mathbf{w} - \mu\Delta\mathbf{w}\|} \approx \mathbf{w} - \mu\Delta\mathbf{w} + \mu\left(\Delta\mathbf{w}^{\mathrm{T}}\mathbf{w}\right)\mathbf{w}. \qquad (26)$$

Applying this approximation to (25) and changing the sign as mentioned above leads to

$$\begin{aligned}\mathbf{w}^{\ell}(m) &= \mathbf{w}^{\ell-1}(m) \\ &+ \frac{\mu}{N}\sum_{i=0}^{\infty}\beta(i,m)\sum_{j=iN_L}^{iN_L+N-1}\left\{\mathcal{SC}\left\{\mathbf{x}(j)\mathbf{\Phi}_{1,D}^{\mathrm{T}}(\mathbf{y}_1(j))\right\}\right. \\ &\left.-\mathbf{w}^{\ell-1}(m)\left[\left(\mathbf{w}^{\ell-1}(m)\right)^{\mathrm{T}}\mathcal{SC}\left\{\mathbf{x}(j)\mathbf{\Phi}_{1,D}^{\mathrm{T}}(\mathbf{y}_1(j))\right\}\right]\right\}. \qquad (27)\end{aligned}$$

This update rule represents the triple-N generalization of the so-called Oja Rule [1] which is well known in the field of PCA. This can be seen more clearly for $D = 1$ and by taking into account $y_1 = \mathbf{w}^{\mathrm{T}}\check{\mathbf{x}}$:

$$\begin{aligned}\mathbf{w}^{\ell}(m) &= \mathbf{w}^{\ell-1}(m) + \frac{\mu}{N}\sum_{i=0}^{\infty}\beta(i,m) \\ &\sum_{j=iN_L}^{iN_L+N-1}\left[\check{\mathbf{x}}(j) - \mathbf{w}^{\ell-1}(m)y_1(j)\right]\Phi_{1,1}(y_1(j)). \qquad (28)\end{aligned}$$

For $N = N_L = 1$, $\ell_{\max} = 1$, $\mathbf{\Phi}_{1,1}(y_1(j)) = y_1(j)$, $\beta(i,m) = \delta_{im}$ this corresponds exactly to the celebrated Oja rule for PCA.

B. Extraction of Multiple Principal Components or Multiple Independent Components

We now consider the case of multiple dominant sources. In general, for this case we have to apply (16a) with the BSS-specific score function (18).

As a practically important special case, we now consider an orthonormality (or, more generally, a unitarity) constraint on $\mathbf{W}$. This case is particularly popular, e.g., in communications engineering, and is motivated as follows. Consider a lossless mixing system. Then the ideal demixing system will also be lossless. For a lossless demxing system $\mathbf{W}$, we obtain the following condition for the signal powers from (6):

$$\mathbf{y}^{\mathrm{T}}\mathbf{y} = \mathbf{x}^{\mathrm{T}}\mathbf{W}\mathbf{W}^{\mathrm{T}}\mathbf{x} \stackrel{!}{=} \mathbf{x}^{\mathrm{T}}\mathbf{x} \quad \Rightarrow \quad \mathbf{W}\mathbf{W}^{\mathrm{T}} \stackrel{!}{=} \mathbf{I}, \qquad (29)$$

i.e., a lossless demixing system can be enforced by an orthonormality constraint, or, more generally, by a unitarity constraint.

Theorem 3: *For $\mathbf{W}$ constrained to be orthonormal (unitary), the expression $\left(\mathbf{W}^{\mathrm{T}}\right)^{+}$ in the second term of (16a) is again zero.*

Proof: Since both $\mathbf{H}$ and $\mathbf{W}$ are assumed to be lossless, the overall system $\mathbf{HW}$ is also lossless, i.e., $\mathbf{HW}$ is also orthonormal (unitary). Hence, its determinant is $\det\{\mathbf{HW}\} = \det\{\mathbf{W}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\} = 1 = \mathrm{const.}$ Together with the first part of the previous proof of Theorem 1, we conclude that in the lossless case the second term of the gradient becomes again equal to zero. ◇

Hence, we conclude that under the orthonormality (unitarity) assumption, the cases of blind signal extraction and principal component analysis for multiple dominant sources can be treated analogously to the previously considered case of only one dominant source.

In particular, by replacing the length-PL vector $\mathbf{w}^{\cdots}(m)$ in (27) by a $PL \times R$ matrix $\check{\mathbf{W}}^{\cdots}(m)$ for the R dominant components ($1 \leq R \leq P$), and $\mathbf{\Phi}_{1,D}(y_1(j))$ by $\mathbf{\Phi}_{\mathbf{y},D}(\mathbf{y}(j))$ accordingly, where the length-RD vector $\mathbf{y}$ contains the R dominant output signals with D time lags each, we directly obtain the corresponding triple-N generalization of the Oja rule for multiple dominant components. This can again be seen more clearly for $D = 1$, so that we obtain the generalization of (28) for multiple dominant sources. In [2] it was shown for the PCA case (i.e., stationary SOS case) that the resulting coefficient update also corresponds up to a normalization factor to the celebrated PAST (projection approximation subspace tracking) algorithm [2] with gradient-based adaptation. It should be noted that RLS-type algorithms, as in [2], correspond to the more advanced Newton-based adaptation instead of the simple gradient-based adaptation which we have considered in this paper for brevity.

VI. SUMMARY AND CONCLUSIONS

Based on TRINICON, a generic framework for adaptive signal processing, we have explicitly shown several algorithmic relations between BSS, MIMO-BSI, SIMO-BSI, MCA, PCA, subspace tracking, and blind signal extraction, as summarized in Fig. 2. Due to the top-down approach we also obtained various extensions of the algorithms and useful guidelines.

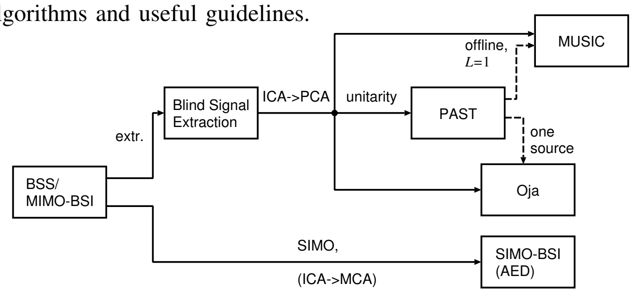


Fig. 2. Summary of algorithmic relations considered in this paper.

REFERENCES

[1] E. Oja, “A simplified neuron model as a principal component analyzer,” *J. Math. Bio.*, vol. 14, pp. 267–273, 1982.

[2] B. Yang, “Projection approximation subspace tracking,” *IEEE Trans. Signal Processing*, vol. 43, no. 1, pp. 95–107, Jan. 1995.

[3] R. Schmidt, “Multiple emitter location and signal parameter estimation,” *IEEE Trans. Antennas and Propagation*, vol. AP-34, no. 3, pp. 276–280, Mar. 1986.

[4] R. Roy and T. Kailath, “ESPRIT - estimation of signal parameters via rotational invariance techniques,” *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. 37, no. 7, pp. 984–995, Jul. 1989.

[5] M. Viberg, B. Ottersten, and T. Kailath, “Detection and estimation in sensor arrays using weighted subspace fitting,” *IEEE Trans. Signal Processing*, vol. 39, pp. 2436–2449, 1991.

[6] M. Gürelli and C. Nikias, “EVAM: an eigenvector-based algorithm for multichannel blind deconvolution of input colored signals,” *IEEE Trans. Signal Processing*, vol. 43, no. 1, pp. 134–149, Jan. 1995.

[7] J. Benesty, “Adaptive eigenvalue decomposition algorithm for passive acoustic source localization,” *J. Acoust. Soc. Am.*, vol. 107, pp. 384–391, Jan. 2000.

[8] S. Haykin, *Adaptive Filter Theory*, 4th ed. Englewood Cliffs, NJ: Prentice Hall Inc., 2002.

[9] H. Buchner, R. Aichner, and W. Kellermann, “TRINICON-based blind system identification with application to multiple-source localization and separation,” in *Blind Speech Separation*, S. Makino, T.-W. Lee, and S. Sawada, Eds. Berlin: Springer, Sept. 2007, pp. 101–147.

[10] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: John Wiley & Sons, 2001.

[11] H. Buchner, R. Aichner, and W. Kellermann, “TRINICON: A versatile framework for multichannel blind signal processing,” in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 3, Montreal, Canada, May 2004, pp. 889–892.

[12] ——, “Blind source separation for convolutive mixtures: A unified treatment,” in *Audio Signal Processing for Next-Generation Multimedia Communication Systems*, J. Benesty and Y. Huang, Eds. Boston: Kluwer Academic Publishers, Apr. 2004, pp. 255–293.

[13] 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. Workshop Hands-Free Speech Commun. and Microphone Arrays*, Trento, Italy, May 2008.

[14] ——, “TRINICON for dereverberation of speech and audio signals,” in *Speech Dereverberation*, P. Naylor and N. Gaubitch, Eds. London: Springer, Jul. 2010, pp. 311–385.