A FUNDAMENTAL RELATION BETWEEN BLIND AND SUPERVISED ADAPTIVE FILTERING ILLUSTRATED FOR BLIND SOURCE SEPARATION AND ACOUSTIC ECHO CANCELLATION

Herbert Buchner

Deutsche Telekom Laboratories
Technical University Berlin
Ernst-Reuter-Platz 7, 10587 Berlin, Germany

Walter Kellermann

Multimedia and Signal Processing
University of Erlangen-Nuremberg
Cauerstr. 7, 91058 Erlangen, Germany

ABSTRACT

In recent years broadband signal aquisition by sensor arrays, e.g., for speech and audio signals in a hands-free scenario, has become a popular research field in order to separate certain desired source signals from competing or interfering source signals ((blind) source separation or interference cancellation) and to possibly dereverberate them (blind deconvolution). In various practical scenarios, some or even all interfering source signals may be directly accessible and/or some side information on the propagation path is known. In these cases we can tackle the separation problem by supervised adaptation algorithms, e.g., the popular LMS- or RLS-type algorithms, rather than the more involved blind adaptation algorithms. In contrast, for blind estimation, such as in the blind source separation (BSS) scenario where both the propagation paths and the original source signals are unknown, the method of independent component analysis (ICA) is typically applied. Traditionally, the ICA method and supervised adaptation algorithms have been treated as different research areas. In this paper, we establish a conceptually simple, yet fundamental relation between these two worlds. This is made possible using the previously introduced generic broadband adaptive filtering framework, called TRINICON. As we will demonstrate, not only both the well-known blind and supervised adaptive filtering algorithms turn out as special cases of this generic framework, but we also gain various new insights and synergy effects for the development of new and improved adaptation algorithms.

Index Terms— adaptive signal processing, separation, echo suppression, MIMO systems, robustness

1. INTRODUCTION: ADAPTIVE MIMO FILTERING AND IDEAL SIGNAL SEPARATION SOLUTION

In broadband signal aquisition by sensor arrays, such as in hands-free speech communication scenarios, the original source signals $s_q(n)$, $q = 1, \ldots, Q$ are filtered by a linear multiple input and multiple output (MIMO) system (e.g., the reverberant room) before they are captured as sensor signals $x_p(n)$, $p = 1, \ldots, P$. In this paper, we describe this MIMO mixing system by length-M FIR filters, 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. Moreover, we assume throughout this paper that $Q \leq P$. According to a certain optimization criterion, we are interested in finding a corresponding length-L FIR demixing system with coefficients $w_{pq,\kappa}$ by adaptive signal processing. This yields the output signals $y_q(n)$. As a compact formulation of the set of demixing filter coefficients and mixing filter coefficients we form

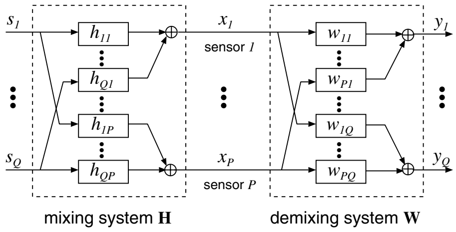

Fig. 1. General setup for MIMO signal processing.

the $PL \times Q$ demixing coefficient matrix

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

and the corresponding $QM \times P$ mixing coefficient matrix $\check{\mathbf{H}}$, respectively, where

$$\mathbf{h}_{qp} = [h_{qp,0}, \ldots, h_{qp,M-1}]^{\mathrm{T}}, \tag{2}$$

$$\mathbf{w}_{pq} = [w_{pq,0}, \ldots, w_{pq,L-1}]^{\mathrm{T}} \tag{3}$$

denote the coefficient vectors of the FIR subfilters of the MIMO systems, and superscript $^{\mathrm{T}}$ denotes transposition of a vector or a matrix. The downwards pointing hat symbol on top of $\mathbf{W}$ in (1) serves to distinguish this *condensed* matrix from the corresponding larger matrix structure $\mathbf{W}$ as introduced below in Sect. 2. The rigorous distinction between these different matrix structures is also an essential aspect of the general TRINICON framework, as shown later.

In this paper, we focus on *signal separation and system identification problems*. If both the propagation paths and the original source signals in Fig. 1 are unknown, the demixing system has to be estimated by *blind* source separation (BSS) for which the method of independent component analysis (ICA) is typically applied [1]. In other practical scenarios, in which some or even all interfering source signals are directly accessible and/or some side information on the propagation path is known, we can tackle the separation problem by *supervised* adaptation algorithms, such as the popular least-mean-square (LMS)- or the recursive least-squares (RLS)-type algorithms [2].

The *ideal MIMO separation filters* $\check{\mathbf{W}}_{\mathrm{ideal,sep}}$ were derived and discussed in detail in [3] for an arbitrary number of sources and sensors. In this paper we only need to consider the special case $Q = P = 2$. Then, the ideal separating filter matrix $\check{\mathbf{W}}_{\mathrm{ideal,sep}}$

reads [3]

$$\check{\mathbf{W}}_{\mathrm{ideal,sep}} = \begin{bmatrix} \mathbf{h}_{22} & -\mathbf{h}_{12} \\ -\mathbf{h}_{21} & \mathbf{h}_{11} \end{bmatrix} \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix}, \quad (4)$$

where due to the scaling ambiguity (in blind problems) each column is multiplied by an unknown scalar α_q.

For $L = L_{\mathrm{opt,sep}} = M$, this *ideal separation solution corresponds to a MIMO system identification up to an arbitrary scalar constant* [3] (independently of the adaptation method and the possible prior knowledge). This important link was also successfully exploited to obtain practical schemes for blind system identification and multiple source localization [3].

2. REVIEW OF TRINICON FOR THE CASE OF SEPARATION PROBLEMS

In this section we give a brief overview of the essential elements of TRINICON ('TRIple-N ICA for CONvolutive mixtures'), a generic concept for broadband adaptive MIMO filtering [4, 5, 6, 7]. Thereby, we restrict the presentation here *to the case of separation and identification problems* and to simple Euclidean gradient-based coefficient updates in the time domain.

Various approaches exist to estimate the demixing matrix $\mathbf{W}$ by utilizing the following source signal properties [1] which were all combined in TRINICON as a versatile framework:

(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 [1].

(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.

Throughout this section, we present the framework for $Q = P$ without loss of generality. In practice, the current number of simultaneously active sources is allowed to vary throughout the application and only the condition $Q \leq P$ has to be fulfilled.

2.1. OPTIMIZATION CRITERION

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 [7]:

$$\mathbf{y}(n) = \mathbf{W}^{\mathrm{T}}\mathbf{x}(n), \quad (5)$$

where n denotes the time index, and

$$\begin{aligned} \mathbf{x}(n) &= [\mathbf{x}_1^{\mathrm{T}}(n), \ldots, \mathbf{x}_P^{\mathrm{T}}(n)]^{\mathrm{T}}, & (6)\\ \mathbf{y}(n) &= [\mathbf{y}_1^{\mathrm{T}}(n), \ldots, \mathbf{y}_P^{\mathrm{T}}(n)]^{\mathrm{T}}, & (7)\\ \mathbf{x}_p(n) &= [x_p(n), \ldots, x_p(n-2L+1)]^{\mathrm{T}}, & (8)\\ \mathbf{y}_q(n) &= [y_q(n), \ldots, y_q(n-D+1)]^{\mathrm{T}}. & (9) \end{aligned}$$

The parameter D in (9), $1 \leq D < L$, 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 $\mathbf{W}_{pq}$ analogously to (1).

Based on a generalization of Shannon's mutual information, the following cost function for signal separation was introduced in [4] taking into account all three fundamental signal properties (i)-(iii):

$$\begin{aligned} \mathcal{J}(m, \mathbf{W}) &= -\sum_{i=0}^{\infty} \beta(i,m) \frac{1}{N} \\ &\cdot \sum_{j=iL}^{iL+N-1} \left\{ \sum_{p=1}^{P} \log(\hat{p}_{y_p,D}(\mathbf{y}_p(j))) - \log(\hat{p}_{y,PD}(\mathbf{y}(j))) \right\}, \quad (10) \end{aligned}$$

where $\hat{p}_{y_p,D}(\cdot)$ and $\hat{p}_{y,PD}(\cdot)$ are assumed or estimated multivariate pdfs of dimensions D and PD, 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 [5].

2.2. GRADIENT-BASED COEFFICIENT UPDATE

For brevity and simplicity we concentrate in this paper on iterative Euclidean gradient-based block-online coefficient updates which can be written in the general form (according to our already performed investigations, all results of this paper also carry over to other known strategies, such as the natural gradient- or Newton-based updates)

$$\begin{aligned} \check{\mathbf{W}}^0(m) &:= \check{\mathbf{W}}(m-1), & (11a)\\ \check{\mathbf{W}}^{\ell}(m) &= \check{\mathbf{W}}^{\ell-1}(m) - \mu\Delta\check{\mathbf{W}}^{\ell}(m), \quad \ell = 1, \ldots, \ell_{\max}, & (11b)\\ \check{\mathbf{W}}(m) &:= \check{\mathbf{W}}^{\ell_{\max}}(m), & (11c) \end{aligned}$$

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.

It can be shown that by taking the gradient of $\mathcal{J}(m)$ with respect to the demixing filter matrix $\check{\mathbf{W}}(m)$, we obtain the following generic gradient descent-based TRINICON update rule [4, 5]:

$$\begin{aligned} \Delta\check{\mathbf{W}}^{\ell}(m) &= \frac{1}{N} \sum_{i=0}^{\infty} \beta(i,m)\, \mathcal{SC}\left\{ \sum_{j=iL}^{iL+N-1} \left[\mathbf{x}(j)\boldsymbol{\Phi}_{s,PD}^{\mathrm{T}}(\mathbf{y}(j)) \right.\right. \\ &\left.\left. -\mathbf{V}\left(\left(\mathbf{W}^{\ell-1}(m)\right)^{\mathrm{T}}\mathbf{V}\right)^{-1}\right]\right\}, \quad (12a) \end{aligned}$$

with the window matrix

$$\begin{aligned} \mathbf{V} &= \mathrm{Bdiag}\{\tilde{\mathbf{V}}, \ldots, \tilde{\mathbf{V}}\}, & (12b)\\ \tilde{\mathbf{V}} &= \left[\mathbf{I}_{D\times D},\ \mathbf{0}_{D\times(2L-D)}\right]^{\mathrm{T}}, & (12c) \end{aligned}$$

and the multivariate score function (here for separation problems)

$$\begin{aligned} \boldsymbol{\Phi}_{s,PD}(\mathbf{y}(j)) &= \left[\boldsymbol{\Phi}_{y_1,D}^{\mathrm{T}}(\mathbf{y}_1(j)), \ldots, \boldsymbol{\Phi}_{y_P,D}^{\mathrm{T}}(\mathbf{y}_P(j))\right]^{\mathrm{T}}, & (12d)\\ \boldsymbol{\Phi}_{y_p,D}(\mathbf{y}_p(j)) &= -\frac{\partial \log \hat{p}_{y_p,D}(\mathbf{y}_p(j))}{\partial \mathbf{y}_p(j)}. & (12e) \end{aligned}$$

The so-called *Sylvester constraint* $\mathcal{SC}\{\bullet\}$ in (12a), formally introduced in [5, 7], is an important building block linking the two different coefficient matrix formulations $\mathbf{W}$ (in the cost function) and $\check{\mathbf{W}}$ (in the optimization procedure). In [3] the explicit formulation of the *generic* Sylvester constraint was derived. There it was shown in a rigorous way that the generic Sylvester constraint corresponds to a *channel-wise summation* of elements according to Fig. 2. Seemingly different previously introduced approaches in the literature each correspond to certain approximations of ($\mathcal{SC}$) by neglecting some of the elements within this summation, as illustrated in [3]. Various links from the general equations (12) to existing and also to novel BSS algorithms have been discussed, e.g., in [3, 5].

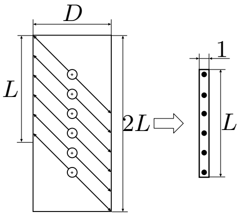

Fig. 2. Illustration of the generic Sylvester constraint ($\mathcal{SC}$) after [3] for one channel.

3. TRINICON FOR SUPERVISED SYSTEM IDENTIFICATION AND INTERFERENCE CANCELLATION

So far in this paper, we have not made any explicit assumption whether or not some of the source signals are directly accessible, i.e., if our scenario is actually blind, supervised, or semi-blind. In fact, as we will see in this section, the rigorous formulation of blind adaptive filtering algorithms, as given by the TRINICON framework, may be seen as a generic version, from which the supervised and semi-blind versions may be deduced as special cases.

3.1. ILLUSTRATION BY SINGLE-CHANNEL ECHO CANCELLATION FROM A SPECIALIZED MIXING MODEL

As a popular example for supervised adaptive filtering we consider acoustic echo cancellation (AEC), e.g., [8], where all interfering source signals are directly accessible. In the case of one unknown source $s_1 = s$ and one hard-wired interfering source signal $s_2 = r$, we obtain the *specialized mixing model* shown on the left side of Fig. 3. According to the right-hand side of (4) the corresponding ideal *demixing system* taking into account this prior knowledge reads

$$\begin{bmatrix} \mathbf{w}_{11} & \mathbf{w}_{12} \\ \mathbf{w}_{21} & \mathbf{w}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_1 & \mathbf{0} \\ -\mathbf{h} & \mathbf{1}_1 \end{bmatrix}. \tag{13}$$

By comparing both sides of this equation, we immediately obtain the corresponding demixing system structure shown on the right side in Fig. 3. This is indeed the well-known single-channel AEC approach, which in this way follows rigorously from the general equation (4) together with the prior knowledge on the specialized mixing system.

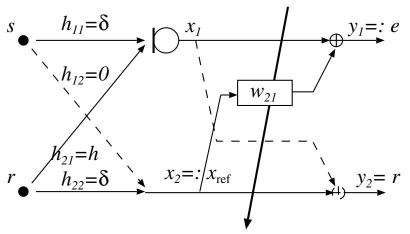

Fig. 3. Single-channel echo cancellation: specialized mixing system and corresponding specialized demixing system after (13).

3.2. CORRESPONDING SPECIAL CASE OF THE GRADIENT-BASED TRINICON COEFFICIENT UPDATE

From the specialized form (13) of the demixing system follows that we need to adapt only the lower left submatrix of $\check{\mathbf{W}}$. All other "subfilters" are already fixed to $\mathbf{1}_1$ and $\mathbf{0}$, respectively, according to Fig. 3. In the following we study the implications of this specialization on the TRINICON-based coefficient update (12a).

We first consider the specialization of the *second term* of (12a) using (12c) and a formulation of (13) in terms of Sylvester matrices:

$$\begin{aligned}
&\mathbf{V}\left(\mathbf{W}^{\mathrm{T}}\mathbf{V}\right)^{-1} = \\
&= \begin{bmatrix} \tilde{\mathbf{V}} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{V}} \end{bmatrix} \left(\begin{bmatrix} \tilde{\mathbf{V}}^{\mathrm{T}} & -\hat{\mathbf{H}}^{\mathrm{T}} \\ \mathbf{0} & \tilde{\mathbf{V}}^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{V}} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{V}} \end{bmatrix} \right)^{-1} \\
&= \begin{bmatrix} \tilde{\mathbf{V}} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{V}} \end{bmatrix} \begin{bmatrix} \mathbf{I}_{D\times D} & -\hat{\mathbf{H}}^{\mathrm{T}}\tilde{\mathbf{V}} \\ \mathbf{0} & \mathbf{I}_{D\times D} \end{bmatrix}^{-1} = \begin{bmatrix} \tilde{\mathbf{V}} & \tilde{\mathbf{V}}\hat{\mathbf{H}}^{\mathrm{T}}\tilde{\mathbf{V}} \\ \mathbf{0} & \tilde{\mathbf{V}} \end{bmatrix}.
\end{aligned} \tag{14}$$

The last expression follows from the inversion of partitioned matrices, e.g., [9]. The simple structure of (14) is a very important result as it shows that *in the supervised case, the lower left submatrix of the second term of (12a) disappears without loss of generality*.

Next, we note that the lower left submatrix of the *first term* $\mathbf{x}(j)\mathbf{\Phi}_{s,PD}^{\mathrm{T}}(\mathbf{y}(j))$ in the coefficient update (12a) obviously reads $\mathbf{x}_2(j)\mathbf{\Phi}_{y_1,D}^{\mathrm{T}}(\mathbf{y}_1(j))$. We now perform the following formal substitutions in order to be in accordance with the literature on supervised adaptive filtering, e.g., [2] (see also Fig. 3):

$$\mathbf{x}_2 \to \mathbf{x}_{\mathrm{ref}}, \qquad \mathbf{y}_1 \to \mathbf{e}, \qquad \mathbf{w}_{21} \to -\hat{\mathbf{h}}. \tag{15}$$

Hence, the lower left submatrix of the first term is finally expressed as $\mathbf{x}_{\mathrm{ref}}(j)\mathbf{\Phi}_{e,D}^{\mathrm{T}}(\mathbf{e}(j))$. Note that $\mathbf{w}_{21} \to -\hat{\mathbf{h}}$ is justified by (13).

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

$$\begin{aligned}
\hat{\mathbf{h}}^{\ell}(m) &= \hat{\mathbf{h}}^{\ell-1}(m) \\
&+ \frac{\mu}{N}\sum_{i=0}^{\infty}\beta(i,m)\,\mathcal{SC}\left\{\sum_{j=iL}^{iL+N-1}\mathbf{x}_{\mathrm{ref}}(j)\mathbf{\Phi}_{e,D}^{\mathrm{T}}(\mathbf{e}(j))\right\}.
\end{aligned} \tag{16}$$

This is the *triple-N-generalization of the Least-Mean-Squares (LMS) algorithm* from supervised adaptive filtering theory which in its well-known original form exhibits the simple update [2]

$$\hat{\mathbf{h}}(n) = \hat{\mathbf{h}}(n-1) + \mu\,\check{\mathbf{x}}_{\mathrm{ref}}(n)e(n), \tag{17}$$

where the length-L vector $\check{\mathbf{x}}_{\mathrm{ref}}$ is a truncated version of $\mathbf{x}_{\mathrm{ref}}$ (formally, this truncation is obtained by ($\mathcal{SC}$) for $D = 1$, see Fig. 2). Although not shown in this paper due to space limitations, we may 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., [10]) by choosing a Newton-type TRINICON coefficient update instead of the gradient descent-type update.

From the generalized LMS update (16) above we can make the following observations in comparison with the simple case (17): Due to the generalized approach, we inherently obtain

- block online adaptation, possibly with multiple iterations ℓ to speed up the convergence [5].
- block averaging by $N > 1$ for more uniform convergence
- an error nonlinearity to take into account the *nongaussianity* of the near-end signals (see discussion of the criterion above)
- multivariate error $\mathbf{e}$ to take into account the *nonwhiteness* of the signals (see discussion of the criterion above)

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 Gaussian source signals at the near end (source s in Fig. 3) and at the far end (source r in Fig. 3), and additionally, stationarity and whiteness at the near-end. The general update resulting from TRINICON does not have these restrictions.

3.3. INCORPORATING SPHERICALLY INVARIANT RANDOM PROCESSES

In general, the estimation and handling of high-dimensional multivariate pdfs $\hat{p}_{e,D}(\mathbf{e})$ is a very challenging task. In [4] we proposed an efficient solution by assuming spherically invariant random processes (SIRPs). The SIRP models are representative for a wide class of stochastic processes. It has been shown that speech signals in particular can very accurately be represented by SIRPs [11]. The general model of a correlated SIRP of D-th order is given with a properly chosen function $f_D(\cdot)$ by [11]

$$\hat{p}_{e,D}(\mathbf{e}) = \frac{1}{\sqrt{\pi^D \det(\mathbf{R_{ee}})}} f_D\left(\mathbf{e}^{\mathrm{T}}\mathbf{R}_{\mathbf{ee}}^{-1}\mathbf{e}\right) \tag{18}$$

with the $D \times D$ correlation matrix $\mathbf{R_{ee}}$. As the best known example, the multivariate Gaussian can be viewed as a special case of the class of SIRPs, where $f_D(u) = \frac{1}{\sqrt{2^D}}\exp(-\frac{1}{2}u)$. To calculate the score function for SIRPs in general, we employ the chain rule to (18)

$$-\frac{\partial \log \hat{p}_{e,D}(\mathbf{e})}{\partial \mathbf{e}} = -\frac{\frac{\partial \hat{p}_{e,D}(\mathbf{e})}{\partial \mathbf{e}}}{\hat{p}_{e,D}(\mathbf{e})} = \underbrace{\left[-\frac{1}{f_D(u)}\frac{\partial f_D(u)}{\partial u}\right]}_{:=\phi_{e,D}(u)} \mathbf{R}_{\mathbf{ee}}^{-1}\mathbf{e}, \tag{19}$$

where $u = \mathbf{e}^{\mathrm{T}}\mathbf{R}_{\mathbf{ee}}^{-1}\mathbf{e}$. For convenience, we call the scalar function $\phi_{e,D}(u)$ the *SIRP score*. A great advantage of SIRPs is that the required function $f_D(u)$ can actually be derived analytically from the corresponding *univariate* pdf [11]. Following the procedure in [11], we exemplarily presented in [4] the optimum SIRP score for a univariate Laplacian pdf.

Having derived the multivariate score function for the SIRP model (19), we can now introduce it into the generalized LMS update (16)

$$\hat{\mathbf{h}}^{\ell}(m) = \hat{\mathbf{h}}^{\ell-1}(m) + \frac{\mu}{N}\sum_{i=0}^{\infty}\beta(i,m)$$
$$\mathcal{SC}\left\{\sum_{j=iL}^{iL+N-1}\mathbf{x}_{\mathrm{ref}}(j)\mathbf{e}^{\mathrm{T}}(j)\mathbf{R}_{\mathbf{ee}}^{-1}(i)\phi_{e,D}\left(\mathbf{e}^{\mathrm{T}}(j)\mathbf{R}_{\mathbf{ee}}^{-1}(i)\mathbf{e}(j)\right)\right\}. \tag{20}$$

3.4. RELATION TO SECOND-ORDER STATISTICS

The case of second-order statistics corresponds to a Gaussian model for $\hat{p}_{e,D}(\mathbf{e})$. We easily derive from (19) that in this case $\phi_{e,D}(u) = 1/2$ so that (20) simplifies accordingly. In any case, both from the resulting update equation and from (20), we see that the SIRP model leads to an inherent normalization by the auto-correlation matrix. This normalization in conjunction with $N > 1$ may be interpreted as an *inherent stepsize control* and it also illustrates why BSS does not require a separate double-talk detector (DTD), such as the traditional AEC algorithms do [8]. Conversely, with a suitable estimation procedure for $\mathbf{R_{ee}}$ the triple-N AEC algorithms may be applied without additional double-talk detector.

Further approximating the near-end signal s (and e) as *stationary white* Gaussian noise with *unit variance*, i.e., $\mathbf{e} = e$, $\mathbf{R_{ee}} = 1$, finally leads to the traditional LMS (17). However, this stationarity assumption brings about the necessity of an additional adaptation control.

3.5. RELATION TO MULTIVARIATE ROBUST STATISTICS

By introducing a so-called scaling matrix $\mathbf{S}$ as a multivariate generalization of the *scaling factor* in, e.g., [10, 12, 13], so that $\mathbf{R_{ee}} = \mathbf{SS}^{\mathrm{T}}$, it is straightforward to equivalently formulate (20) as

$$\hat{\mathbf{h}}^{\ell}(m) = \hat{\mathbf{h}}^{\ell-1}(m) + \frac{\mu}{N}\sum_{i=0}^{\infty}\beta(i,m)$$
$$\cdot\mathcal{SC}\left\{\sum_{j=iL}^{iL+N-1}\mathbf{x}_{\mathrm{ref}}(j)\mathbf{\Psi}^{\mathrm{T}}\left(\mathbf{S}^{-1}(i)\mathbf{e}(j)\right)\mathbf{S}^{-1}(i)\right\}, \tag{21}$$

where $$\mathbf{\Psi}(\mathbf{z}) := \mathbf{z}\,\phi_{e,D}\left(\mathbf{z}^{\mathrm{T}}\mathbf{z}\right), \quad \mathbf{z} = \mathbf{S}^{-1}\mathbf{e}. \tag{22}$$

This is the multivariate, i.e., *triple-N-generalization of the well-known robust LMS algorithm*

$$\hat{\mathbf{h}}(n) = \hat{\mathbf{h}}(n-1) + \frac{\mu}{s(n)}\check{\mathbf{x}}_{\mathrm{ref}}(n)\psi\left(\frac{e(n)}{s(n)}\right) \tag{23}$$

after, e.g, [13] which is based on univariate statistics (i.e., white near-end signal s) and designed to tolerate a certain amount of DTD failures [12]. But even though (23) is more robust during double talk, a DTD is still necessary, while for a suitable parameter choice, (21) does not necessarily require a separate DTD when accompanied with a suitable estimation procedure of the scaling matrix $\mathbf{S}(m)$.

4. CONCLUSIONS

In this paper we established a systematic connection between the blind and supervised adaptive filter theories. It turned out that this opens up a great potential for various synergy effects and improved algorithms. As a by-product, the new viewpoint developed in this paper also leads to a simple information-theoretic interpretation of supervised algorithms. We have already extended the unified treatment of these algorithms further in several directions to be presented in forthcoming publications, such as to the multichannel cases and to semi-blind scenarios.

5. REFERENCES

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

[2] S. Haykin, *Adaptive Filter Theory*, 4th ed., Prentice-Hall, 2002.

[3] 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-Verlag, Berlin/Heidelberg, pp. 101-147, Sept. 2007.

[4] 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*, Kyoto, pp. 223-226, Sept. 2003.

[5] 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.

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

[7] 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.

[8] C. Breining et al., "Acoustic Echo Control - An Application of Very-High-Order Adaptive Filters," *IEEE Signal Processing Magazine*, pp. 42-69, July 1999.

[9] W.H. Press et al., *Numerical Recipes in C*, Cambridge University Press, 1995.

[10] 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.

[11] H. Brehm and W. Stammler, "Description and generation of spherically invariant speech-model signals," *Signal Processing*, vol. 12, pp. 119-141, 1987.

[12] P.J. Huber, *Robust Statistics*, Wiley, New York, 1981.

[13] T. Gänsler, S.L. Gay, M.M. Sondhi, and J. Benesty, "Double-talk robust fast converging algorithms for network echo cancellation," *IEEE Trans. Speech Audio Processing*, vol. 8, pp. 656-663, Nov. 2000.