On the Relationships between Blind Equalization and Blind Source Separation – Part II: Relationships

Romis Attux, Aline Neves, Leonardo T. Duarte, Ricardo Suyama, Cynthia Junqueira,
Leandro Rangel, Tiago M. Dias, João M. T. Romano

Abstract— The objective of this two-part work is to present and discuss the relationships between the problems of blind equalization and blind source separation. Both tasks appear, at first sight, to be essentially distinct, since equalization theory was developed mainly under single-input / single-output (SISO) and single-input/ multiple-output (SIMO) models, whereas the very idea of source separation strongly suggests the need for considering models with multiple inputs and multiple outputs (MIMO). However, in this second part, equivalences between the Benveniste-Goursat-Ruget theorem and the approach to blind source separation based on maximum-likelihood, between the Shalvi-Weinstein techniques and the separation methods that employ kurtosis and, finally, between the Bussgang algorithms and the ICA tools built from concepts such as negentropy and nonlinear principal component analysis are indicated. Finally, some connections previously unexplored in the literature are presented that are derived from ideas such as that of temporality and that of considering the parallels existing between a two-stage (magnitude and phase) equalization procedure and the classical pair PCA / ICA.

Index Terms— Adaptive Filtering, Blind Equalization, Blind Source Separation, Independent Component Analysis and Convolutive Mixtures.

I. INTRODUCTION

In this second part, we turn our attention to the relationships between the problems of blind equalization and blind source separation. Our exposition establishes a number of specific links and analyzes each of them separately, not, however, without indicating more general connections when it is relevant. After discussing the connections between the theorems of Benveniste-Goursat-Ruget (BGR) and Shalvi-Weinstein (SW), the class of Bussgang algorithms and the concepts of nongaussianity, nonlinear principal component analysis and kurtosis-based methods, our discussion enters a final stage in which previously unexplored ideas play a key role. This final discussion gravitates around two poles: the notion of *temporality* and the idea of relating the twofold procedure exemplified by the operation of a magnitude-phase equalizer to the sequential (two-stage) use of principal and independent component analysis that is well-established in blind source separation.

II. RELATIONSHIPS

After having defined in part I the problems of blind equalization and blind source separation and the classical methods to solve them, it is time to start our study of their multifaceted relationships. In this section, we will analyze how different approaches belonging to both research fields can be related. Our analysis will follow a general scheme: after exposing in intuitive terms the theoretical foundations of each approach, we shall proceed to the task of highlighting and discussing a number of links between them, as well as some of their implications to the future development of the area.

A. Maximum Likelihood and BGR

The BGR theorem, presented in section 2.4 of part I, represented a major breakthrough in blind channel equalization theory. The idea behind it is that, considering that both channel and equalizer are linear devices, if the probability density functions (pdfs) of the transmitted and estimated signals are equal, the combined channel+equalizer discrete-time impulse response must correspond to an impulse, which implies perfect recovery of the original signal. In a certain sense, it delineates a pdf fitting criterion that can be enunciated in the following terms: adapt the equalizer in order to render the pdf of the estimated signal as close as possible to that of the transmitted signal.

Not surprisingly, this line of reasoning reveals a clear

Romis Attux, is with the Laboratory of Signal Processing for Communications (DSPCom) – FEEC – University of Campinas (Unicamp) – CP 6101 – CEP 13083-970 – Campinas – SP – Brazil (phone: +55-1935213857; fax: +55-19-35213857; e-mail: attux@dca.fee.unicamp.br).

Ricardo Suyama, Leandro Elias Paiva Rangel, Tiago Macedo Dias and João Marcos Travassos Romano are with the Laboratory of Signal Processing for Communications (DSPCom) – FEEC – University of Campinas (Unicamp) – CP 6101 – CEP 13083-970 – Campinas – SP – Brazil (e-mails: {rsuyama, romano}@dmo.fee.unicamp.br, {leandroe, tdias}@decom.fee.unicamp.br)

Aline Neves is now with the Engineering, Modeling and Applied Social Sciences Center, Federal University of ABC, Brazil (e-mail: aline.neves@ufabc.edu.br)

Leonardo Tomazeli Duarte is with GIPSA Lab, Domaine Universitaire BP 46 38402 Saint Martin d'Hères Cedex (e-mail: leonardo.duarte@gipsa-lab.inpg.fr)

Cynthia Cristina Martins Junqueira is with the Institute of Aeronautics and Space (IAE) – Aerospace Technical Center (CTA) (e-mail: cynthia@iae.cta.br)

relationship with the maximum likelihood approach in blind signal separation. In the BSS context, if one assumes that the pdfs of the sources are known a priori, we saw in section 3.2.3 of part I that a valid criterion to attain the perfect separation of the sources is to maximize the log-likelihood function. Taking into account the particularities of the problem, it was shown that the maximization of the log-likelihood function is equivalent to minimizing the Kullback-Leibler divergence between the pdf of estimated signals and that of the sources. Since the divergence attains its minimum if both pdfs are equal, this criterion also expresses the idea of pdf matching provided by the BGR theorem.

B. Kurtosis and Shalvi-Weinstein

A close look at sections 2.6 and 3.2.2 of part I shows how two different developments, in different contexts, led to similar optimization problems, problems whose essence is the maximization of kurtosis. In spite of this similarity, the historical and conceptual motivations behind the use of this higher-order statistic in the problems of equalization and source separation are by no means identical. In fact, the contrast between how fourth-order statistics gained attention in these research fields is a very rich subject the fundamentals of which will occupy us in this section.

In the SISO equalization problem, Shalvi and Weinstein showed how, in a context in which a ZF solution is attainable, the straightforward relation between the channel input and equalizer output signals, together with an effective gain control, may set the scene for the establishment of a blind criterion [1]. However, although gain control is an important restriction, we have seen in section 2.3 of part I that second-order statistics are generally not sufficient to achieve equalization. The required additional ingredient is exactly the use of a higher-order statistic, wherefore it is possible to state that, in the Shalvi-Weinstein theorem, kurtosis plays the crucial role of selecting, among the various solutions capable of restoring the variance of the transmitted signal, one which *actually inverts the channel* (i.e. which corresponds to a ZF solution). In the case in which a ZF solution is not attainable, the parameters of the equalizer are chosen in accordance with the notion of maximizing the absolute value of the kurtosis.

On the other hand, the correspondent solution to the problem of source separation is obtained through the concept of nongaussianity, which was defined in section 3.2.2 of part I. Given the fact that a superposition of i.i.d. variables tends to be "more Gaussian" than any of them individually, a direct path to recover a source from a mixture is to force a condition "as nongaussian as possible". A natural choice to quantify the degree of nongaussianity of a signal is exactly the kurtosis, which is zero almost exclusively for Gaussian random variables.

1) Relationships

Even though these two approaches seem to be rather disparate at first sight, we will now attempt to show that they are, in fact, deeply interrelated. The principle of using the kurtosis as a higher-order statistic of reference in a "matching" process – the essence of the SW theorem – can be directly extended to the context of blind source separation. Firstly, due to the fact that, in the classical ICA problem, the mixing system is invertible, the separation scenario also allows, in a certain sense, "that a ZF solution be attained". Moreover, it is also possible to impose the power normalization to the separating system, which leads us to the same scenario of the classical SW theorem. The main difference lies in the nature of the mixture: in ICA, it is formed by an instantaneous superposition of independent sources, whereas, in SISO equalization, the received signal is composed of several delayed samples of a transmitted message. This difference will be relevant in some of our future discussion (see section 3.1), but, at this point, it is not for two reasons:

- When we exposed the SW theorem, we assumed that the samples of the transmitted signal were i.i.d., which means that the received signal, in the above case, is indeed a linear combination of independent sources (each corresponding to a delayed version of $s(n)$), in analogy with ICA.
- The kurtosis criterion is instantaneous in its nature, i.e., does not depend on samples taken at different time instants.

These arguments reveal that the SW theorem, which was conceived under the aegis of a typical equalization problem, can be directly extended to the ICA framework. As a consequence, a source can be extracted by linearly combining a number of mixtures at least equal to the number of sources and choosing the coefficients of this linear combination in order to force the kurtosis of the combiner output to be equal to the kurtosis of the sources (under the ubiquitous second-order power restriction).

It is also possible to show that the principle of using kurtosis as a nongaussianity measure, a well-known pillar of blind source separation theory, can be applied to the problem of SISO equalization. As seen in section 2.5 of part I, considering i.i.d. transmitted samples, the larger the error between the correct inverse filter and the approximate one, the more non-zero terms will be present within the impulse response $c_e(n)$ (see equation (19) of part I). If we assume, in accordance with the central limit theorem, that the distribution of a sum of i.i.d. variables tends toward a Gaussian distribution, we may conclude that, as the length of $c_e(n)$ increases, the closer to a Gaussian distribution will be the convolutional noise [2]. We could then understand the problem in the other sense and consider that, the more nongaussian is the signal, the smaller is the discrepancy between the two filters and, consequently, the better will be the inverse filter approximation. In this case, we would also look for a measure of nongaussianity, which naturally impels us to the use of the maximization of kurtosis as the desired criterion. Consequently, we have been led to the SW criterion by a different interpretation of the equalization problem, thus showing that the conceptions originated from the study of the problems of equalization and blind source separation are

equivalent, and, therefore, can be used by researchers of both fields to enrich their theoretical approaches.

2) Some Additional Comparisons and General Remarks

The nature and role of restrictions in these criteria also deserve some comments. As it is not possible to recover the variance of the independent components, it is assumed that the sources have unit variance. Therefore, ICA methods consider this condition as a restriction, while, in contrast, the SW criterion is based on the assumption of equality between the channel input and equalizer output variances. Actually, these conditions are the same: in the former approach, since it is established that the sources have unit variance, the output *y* is also forced to satisfy this condition, which gives a restriction very similar to the one proposed by Shalvi and Weinstein. Moreover, this similarity reveals itself again in the resulting adaptive algorithms, both of which require that **w** be divided by its own norm.

In addition, the sign ambiguity also exists in both cases. Notice that both criteria use the magnitude of the kurtosis value, what enables them to be used equivalently for super-Gaussian and sub-Gaussian sources.

Another interesting similarity concerns the pre-whitening operation. In the ICA method, this process is carried out before the application of the kurtosis criterion, thus simplifying the variance constraint for the deduction of the adaptive algorithm. In the SW case, pre-whitening is part of the adaptive algorithm, appearing as the inverse of a matrix given by the cross-cumulants of *y*. This matrix helps increase the convergence rate of the algorithm. It is then possible to observe that the pre-whitening task establishes a tradeoff (simplicity vs. speed of convergence) when analyzed from the typical BSS and equalization perspectives, a fact that may have practical implications for both fields.

C. Negentropy and Bussgang

Relevant points of intersection can be found when we compare two classes of techniques for blind equalization to the blind source separation algorithms derived from estimates of the negentropy. In this comparison, the existence of similarities may seem surprising at first sight, because the theoretical inspirations of these two classes of unsupervised methods are of a different nature. Motivated by that, we shall place the focus of our analysis on showing how this connection can be established and what implications it has for both fields. The two classes of blind equalization techniques with which we will be concerned in this section are those of the Bussgang algorithms, discussed in section 2.5 of the first part, and of super-exponential algorithms, the basis of which is the Shalvi-Weinstein formulation we discussed in sections 2.4 and 2.6 of the same part.

Bussgang techniques are characterized by their use of a memoryless estimator as a kind of replacement for a pilot signal that, naturally, must not be present in an unsupervised approach. With the aid of this nonlinear function, it is possible to produce a "pseudo-supervised" error to be employed in a classical LMS-like update expression. In summary, we may

state that nonlinearity emerges as a bridge between the familiar "supervised Wiener-like world" and the more complex blind domain.

The super-exponential algorithm (SEA) is obtained as an adaptive method for the optimization of the normalized SW criterion. If the existence of a power restriction is assumed, this criterion is equivalent to the kurtosis maximization [3] and can also be reduced to the minimum entropy criterion [4]. The SEA has the particular characteristic of taking into account the projection of the combined channel+equalizer response onto the so-called attainable set, that is, a set formed by all the equalizer solutions that, given the structure of the involved system, can be effectively attained. Thus, an important difference of the super-exponential algorithm when compared, for example, with Bussgang algorithms, is the realization of a pre-whitening operation, which is responsible for its fast convergence.

The source separation techniques that will be the object of our attention in this section are based on negentropy, which, as discussed in section 3.2.2 of the first part, is a robust measure of nongaussianity. As it is sometimes a hard task to obtain an estimate of the entropy from samples of a data signal, it is usual to employ approximations directly or indirectly based on higher-order moments. Perhaps the most widespread of them for real signals is [5][6]:

$$J = k_a \left\{ E\left[G_1\left(y(n) \right) \right] \right\}^2 + k_b \left\{ E\left[G_2\left(y(n) \right) \right] - E\left[G_2(v) \right] \right\}^2 \quad (1)$$

where $G_1(.)$ and $G_2(.)$ are zero-memory nonlinear functions and v is a Gaussian random variable with the same covariance as $y(n)$. When a single nonquadratic function is used, this expression becomes equation (44) of part I, which we reproduce here for the sake of convenience

$$J = \left\{ E\left[G\left(y(n) \right) \right] - E\left[G(v) \right] \right\}^2 \quad (2)$$

where $G(.)$ is a nonlinear function.

As shown in [6], it is possible to obtain a gradient-based algorithm to optimize the cost function presented in (2):

$$\Delta \mathbf{w} \propto \gamma E\left[\mathbf{x}(n) g\left(\mathbf{w}^T \mathbf{x}(n) \right) \right]$$
$$\mathbf{w} \leftarrow \mathbf{w} / \|\mathbf{w}\| \quad (3)$$

where $(.)^T$ denotes the transpose and $\|.\|$ is the Euclidean norm of a vector. Nevertheless, the most important method for finding the extrema of (2) is the FastICA algorithm:

$$\mathbf{w} \leftarrow E\left[\mathbf{x}(n) g\left(\mathbf{w}^T \mathbf{x}(n) \right) \right] - E\left[\mathbf{x}(n) g'\left(\mathbf{w}^T \mathbf{x}(n) \right) \right] \mathbf{w}$$
$$\mathbf{w} \leftarrow \mathbf{w} / \|\mathbf{w}\| \quad (4)$$

It must be emphasized that in both algorithms a pre-whitening stage is mandatory.

1) Relationship between Bussgang and Negentropy

Our study involves three techniques built from distinct motivations, and it is time for us to look for the means whereby they can be related. Firstly, let us analyze the gradient-based algorithm presented in (3); it shows that the update of the parameter vector associated with the recovery of a single component can be performed in the direction defined by the expected value of the product between a nonlinear function of the output *y* and the input $\mathbf{x}(n)$ (or, at least, an approximation thereof). If we compare it with the general form of a Bussgang algorithm, shown in equation (24) of part I, it becomes clear that the techniques will be related if gnegentropy($y(n)$) = gbussgang($y(n)$) – $y(n)$. This similarity is not obvious a priori, but neither is it incidental: its primary raison d'être is that both approaches are based on the use of nonlinear functions as estimators. The nonlinear mapping must, in either case, be carefully chosen in order to allow that the appropriate higher-order information be employed in the process of equalizing or separating. The curious thing is that, as we have just indicated, it is possible that identical update expressions be generated under very distinct estimation tasks, which suggests a connection between negentropy and sequence estimation that perhaps can be explored in future efforts.

2) Relationship between SEA and FastICA

As discussed in [6], there is also an interesting relationship between the super-exponential algorithm and the FastICA, which, as we have seen, is based on an approximation of negentropy. The connection can be understood in a straightforward manner if we consider the general expression of the FastICA:

$$\mathbf{w} \leftarrow E\left[\mathbf{x}(n)\, g\left(\mathbf{w}^T \mathbf{x}(n)\right)\right] - E\left[\mathbf{x}(n)\, g\,'\left(\mathbf{w}^T \mathbf{x}(n)\right)\right]\mathbf{w} \tag{5}$$
$$\mathbf{w} \leftarrow \mathbf{w}/\|\mathbf{w}\|$$

and choose g($y(n)$) = $2py(n)^{2p-1}$. This, as Kofidis [7] remarks, leads to an expression equivalent to that of the super-exponential algorithm (SEA), repeated here for convenience:

$$\tilde{\mathbf{w}} = \mathbf{R}^{-1}\mathbf{d}$$
$$\mathbf{w} = \frac{\tilde{\mathbf{w}}}{\tilde{\mathbf{w}}^T \mathbf{R}\tilde{\mathbf{w}}} \tag{6}$$

where $\mathbf{R}$ has elements of the form $\mathbf{R}_{ij} = \dfrac{E\left[x(n-j)x(n-i)\right]}{E\left[\left|a(n)\right|^2\right]}$ and $\mathbf{d}$ of the form $\mathbf{d}_i = \dfrac{cum(y(n):2p-1, x(n-i))}{C_{2p}^a}$. The first equation in (5) is equivalent to the first equation of (6), as a careful analysis of the vector $\mathbf{d}$ reveals, whereas the "second equations" of (5) and (6) are power normalizations. This

reveals the intimate relation between these techniques[1].

Interestingly, as shown by [7] and [8], the solutions obtained via the super-exponential algorithm when *p* = 2, i.e., the equilibrium points of the Shalvi-Weinstein criterion, are equivalent to those of the constant modulus criterion, a member of the family of Bussgang techniques. This brings us full circle and reinforces the intimate connection between these three ways of using nonlinearities to generate blind criteria.

D. Nonlinear PCA and Bussgang

The objective of Principal Components Analysis (PCA) is, as it was discussed in section 3.1 of part I, to project data into a set of orthogonal directions in accordance with a classical linear scheme. This is achieved in an elegant and direct manner through the use of second-order statistics and of a standard eigendecomposition[2].

The process of decomposing a given vector into a set of orthogonal components can be understood as a method for whitening its elements. Hence, since whitening is different from source separation, the linear decomposition on which the PCA process is founded does not suffice to allow the recovery of the independent sources. In this context, a very natural extension was proposed: to employ a *nonlinear data projection* to recover the underlying components. This is the rationale of the so-called *nonlinear principal component analysis* (NPCA) [5]. In NPCA, the objective is to minimize the cost function:

$$J_{NPCA} = E\left[\left\|\mathbf{x}(n) - \sum_{k=1}^{N} g_k\left(\mathbf{w}_k^T \mathbf{x}(n)\right)\mathbf{w}_k\right\|^2\right] \tag{7}$$

It is possible to demonstrate that, if the input data is prewhitened, this function becomes [6][7]:

$$J_{NPCA} = E\left[\left\|\mathbf{x}(n) - \mathbf{W}^T g\left(\mathbf{W}\mathbf{x}(n)\right)\right\|^2\right]$$
$$= E\left[\left\|\mathbf{y}(n) - g\left(\mathbf{y}(n)\right)\right\|\right] = \sum_{k=1}^{N} E\left\{\left[y_k(n) - g_k\left(y_k(n)\right)\right]^2\right\} \tag{8}$$

A comparison between (8) and equation (17) of part I shows that the NPCA criterion can be formulated as a sum of Bussgang-like cost functions up to the number of independent components.

The comparison in itself is complete, but some aspects deserve discussion. The emergence of a Bussgang-like cost function in nonlinear PCA should not surprise us, given the

---

[1] It is important to notice, however, that the FastICA algorithm given by (5) does not contain the inverse matrix $\mathbf{R}^{-1}$ - which is responsible for the pre-whitening operation - an important characteristic of the SEA. Nonetheless, this can be explained if we consider that the data has already been pre-whitened before the application of the algorithm.

[2] This general process, apropos, is present in the classical adaptive filtering theory as the Karhunen-Loève decomposition of a random process [9], a fact that, in itself, already reveals a connection.

line of reasoning we followed in section 2 of the first part of this article. There, we followed a path towards blind equalization techniques that went from the Wiener approach to blind algorithms via a second-order unsupervised technique (linear prediction) that proved itself rather limited in its applicability. Under the auspices of this sequence, the nonlinearities present in the Bussgang algorithms can be promptly viewed as an implicit source of higher-order statistics. Nonlinear PCA follows the same path, emerging as an efficient alternative to a second-order approach, which is not sufficient to perform ICA, based on the use of a nonlinear function that generates the required higher-order information in the course of a projection task. This situation offers us two interesting views on the problems of equalization and blind source separation:

1) It is possible that the problem of blind equalization be conceived as a problem of nonlinear PCA in a scenario in which the independent components are multiple delayed versions of the transmitted signals, versions that are superposed in the received message due to the existence of intersymbol interference.

2) The problem of blind separation via NPCA becomes analogous to a MIMO equalization task. In fact, we have a Bussgang-like cost function and the pre-whitening hypothesis, which guarantees that the multiple sources will be jointly recovered in a correct way. This closely resembles the blind multi-user detection techniques proposed, for instance, in [10][11], which are based on the constant modulus criterion.

III. EMERGENT TOPICS

After having discussed some relationships between well-established equalization and source separation approaches, it is convenient to analyze some connections associated with less visited notions. In this section, we will discuss two ideas that, to the best of our knowledge, have not been previously formulated in the literature: that of temporality and that of seeking parallels between the pair PCA / ICA and a two-stage amplitude / phase equalizer.

A. Temporality

In the classical problem of SISO equalization, a filter is introduced that must mitigate the effects of a band-limited channel on the transmitted pulses. This limitation reflects itself in the superposition of waveforms associated with different instants of time – a phenomenon we denominate intersymbol interference (ISI). The existence of linear ISI imposes to the received signal a model of the kind

$$x(n) = \sum_{k=0}^{M-1} h(k)s(n-k) \qquad (9)$$

where M is the channel length. In this case, an equalizer input vector with N elements obeys the formula:

$$\mathbf{x}(n) = \begin{bmatrix} h(0) & h(1) & \cdots & h(M-1) & 0 & 0 & \cdots & 0 \\ 0 & h(0) & h(1) & \cdots & h(M-1) & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & h(0) & h(1) & h(2) & \cdots & h(M-1) \end{bmatrix} \cdot \begin{bmatrix} s(n) \\ s(n-1) \\ \vdots \\ s(n-M-N+2) \end{bmatrix} = \mathbf{Hs}(n) \qquad (10)$$

where $\mathbf{H}$ is the NxM+N-1 convolution matrix. In (10), we have a mathematical expression that relates a number of measured samples to unknown versions of an information signal. As a rule, the transmitted message is supposed to be i.i.d., which means that the vector $\mathbf{s}(n)$ has independent elements. If each of these elements is supposed to be an independent component, it becomes possible to understand (10) as the statement of an ICA problem with "less sensors than sources" i.e. *an undermodeled ICA problem.*

When one attempts to solve a problem of this kind, an immediate difficulty arises: the impossibility of inverting the mixing matrix $\mathbf{H}$ (the convolution matrix). This limitation implies that perfect recovery of the independent components is not feasible - an ideal separating matrix simply is not attainable. These facts should not surprise us, for they are a direct consequence of what we discussed in section 2.1 of part I: a zero-forcing (i.e. an ideal) solution is beyond reach in a case in which both the channel and the equalizer are supposed to be linear non-trivial SISO FIR filters[3].

In the context outlined above, it is particularly attractive to consider methods that recover the sources in a "one-by-one" (deflation) basis, because the impracticability of extracting all components may, among other things, compromise any method founded on a joint measure of independence. Suitable methods (e.g. those based on kurtosis or even a version of the FastICA) can recover (with a non-null residual error) some sources and may be unable to recover others: this is something inexorably determined by the structure of the matrix $\mathbf{H}$ and the characteristics of the signal $s(n)$.

This is pretty much what we have to say about the subject from the standpoint of source separation when considering the problem of SISO baud-rate equalization. Nevertheless, our discussion would be incomplete without considering the problem from the point of view of modern equalization theory. The classical SISO unsupervised problem is commonly solved with the help of algorithms like those of the Bussgang class or those inspired in the SW formulation. Since the objective of the equalization task is to recover a version of the transmitted message, the entire effort is naturally directed towards the recovery of a single component (this corresponds to the "one-by-one" character discussed above). As indicated by the Wiener recipe, different equalization delays tend to produce solutions of distinct quality (this is the reason why we considered the Wiener problem a multimodal one): some delays can produce good solutions, while others can give rise to scenarios in which the desired signal cannot be satisfactorily recovered.

The situation we have just described, which is part of a supervised paradigm, is reflected in the behavior of blind equalization techniques. The cost functions associated with

---

[3] For SIMO/MISO channel models/equalizers, given certain conditions, it is still possible to perfectly recover the transmitted signals [12].

the Bussgang and SW algorithms contain multiple minima, among which we find configurations associated with certain "good" equalization delays [13][14] - this is simply another way of stating what we have already said: some sources are recoverable, some are not. A method for discovering all the relevant facts about this undermodeled scenario, by the way, is an exciting open research theme.

As we have just seen, the problems of SISO equalization and source separation are structurally related in a very consistent manner. Notwithstanding, there is a fact we must keep in mind: although the equalization problem can be formulated as a common ICA task, the involved sources are of a very special nature, since they are simply delayed versions of each other.

1) The idea of temporality

The emblematic ICA problem is spatial in its nature: a number of sensors are disposed in a certain environment and distinct mixtures of these sources are captured that, if properly combined, permit the restoration of the independent signals. The SISO equalization problem is essentially temporal: different time samples of a distorted signal are combined to reduce superposition in the time domain. In the beginning of section 3.1, we were able to fit the temporal problem of SISO equalization into the spatial mould of ICA. However, a question remains unanswered: is it possible that some trace of the temporal character of the equalization task be useful to facilitate the tackling of the undermodeled spatial problem?

Let us consider the mixing matrix $\mathbf{H}$ for a while. It is noticeable that it obeys a very specific form, which is a direct consequence of the convolution operator. Moreover, its particular form is responsible for an apparently innocuous property: its first and last columns possess a single non-zero element – the first and the last element, respectively. The consequence of this property is that the source $s(n)$ is present exclusively in the signal $x(n)$ and, analogously, that the source $s(n\text{-}N\text{-}M\text{+}2)$ is present exclusively in $x(n\text{-}N\text{+}1)$. Without loss of generality, let us turn our attention to the first of these facts and to its potential consequences.

If it is our wish to recover $s(n)$, and we are aware of the property we have just stated, what course of action could we follow? Keeping in mind that the delayed versions of $s(n)$ are mutually independent, we could combine all the elements of the input vector from $x(n\text{-}1)$ to $x(n\text{-}N\text{+}1)$ to eliminate all the sources present in $x(n)$, except for $s(n)$. Accordingly, we could build an error signal of the kind

$$e(n) = x(n) - \sum_{k=1}^{N-1} w_k x(n-k) \tag{11}$$

and attempt to minimize its mean-square value with respect to the parameters w_k of the combination. A question arises: and what if this combination eliminates the totality of $x(n)$, and not only the "undesirable" sources? This would certainly be a problem, but not in this particular case: since $s(n)$ is not present at the elements of the combination and, in addition to

that, is independent of all other sources, it is impossible that the error signal be zero. Furthermore, if the combination is ideally efficient, the error will tend to be exactly equal to $s(n)$! Therefore, the minimization of the error signal defined in (11) appears to be an efficient path to explore the structure of the matrix $\mathbf{H}$ in order to recover the source $s(n)$.

It is convenient that we consider for a while what has just been done. Our starting point was the ICA formulation of the equalization problem, which revealed to us an interesting matrix property. Thence, we proceeded to the next step: to build a criterion able to take advantage of the observed property. So far, we thread a familiar path. The important thing is that the obtained error signal, shown in (11), *corresponds to the prediction error we defined in section 2.3 of part I!* In other words, we have reached a familiar destination through a different path: our effort appears to be no more than a rediscovery of the classical prediction-based criterion for blind equalization.

Albeit it may seem that we have reached a dead end, it is possible that a fresh view on the subject may lead us to more interesting conclusions. As it was discussed in section 2.3 of part I, the prediction approach suffers from a serious limitation: it is successful exclusively when the channel is of minimum-phase. This restriction, the validity of which is beyond dispute, is, nonetheless, not so intuitively perceived when we analyze the expression of the error signal defined in (11). We are compelled to ask: why is not the linear combination of past input samples able to remove all sources distinct from $s(n)$ when the channel is of non-minimum phase? The answer can be only one: because, in this case, the very structure of the mixing matrix engenders a filtering problem (in this case, a prediction problem) that cannot be properly solved. Being the criterion soundly formulated in accordance with a Wiener-like mould, there remains a single conclusion: the limitation of the prediction-based approach must lie in the choice of a *linear filtering structure* to play the role of predictor.

This conclusion is undoubtedly encouraging: it means that the possibility of exploring the observed property of the matrix $\mathbf{H}$, i.e., of exploring the temporal character of the undermodeled ICA problem originated in SISO equalization, is not discarded. In fact, the crux of the entire question, as shown by Cavalcante et al. [15] and Ferrari et al. [16], *is the choice of a more powerful filtering structure*. These works reveal that a nonlinear predictor can overcome the classical "minimum-phase restriction" and establish an efficient nonlinear blind equalization approach. Let us translate this achievement into a language more suited to the course of our explanation: the use of a nonlinear prediction-error filter allows the temporal character of the "equalization ICA problem" to be explored and, as a consequence, that the source $s(n)$ be recovered[4]. This is very relevant, since such a condition could never be reached via linear ICA techniques: it

---

[4] The same line of reasoning is valid if a backward predictor is employed. In this case, the recovered source would be $s(n\text{-}N\text{+}1)$.

is the exploration of the structure of the matrix **H**, together with the use of a nonlinear predictor, that allows, in theory, a perfect recovery of *s(n)*.

We decided to assume the "didactical risk" of entering a discussion concerning the use of nonlinear prediction for blind equalization because the manner whereby the subject was reached finely illustrates two points of the greatest importance:

- Indeed, the problem of SISO equalization can be formulated as an undermodeled ICA problem;
- However, the fact that the sources are delayed versions of the same signal can be used to build a criterion that works more efficiently than any method based exclusively on the "spatial character" of the emblematic ICA problem.

The second of these items is the apex of our discussion, since it reveals that the *temporality* of the equalization problem must not be ignored when an ICA formulation thereof is established.

2) Temporality and Convolutive Mixtures

The problem of extracting sources from convolutive mixtures gives us an even more pungent example of the potential of application of the idea of temporality to the problem of source separation. A convolutive mixture obeys a model of the kind:

$$\mathbf{x}(n) = \sum_{k=0}^{L-1} \mathbf{A}_k \mathbf{s}(n-k) \tag{12}$$

where L indicates the model order and $\mathbf{A}_i$ is a mixing matrix associated with the time delay *i*. This model is both spatial and temporal in its character, since each sensor captures multiple sources (spatial character) and multiple delayed versions of them (temporal character). When a problem of this sort is studied, it is natural to attempt to formulate it in accordance the well-established ICA framework. This can be performed in a direct way by considering each delayed version of each source as an independent component, but this approach reveals, together with equation (12), the complexity of the problem: the convolutive model is, in a certain sense, formed by several SISO models; hence, the resulting ICA model is even "more undermodeled". There remains, notwithstanding, the possibility of resorting to the temporal character of the convolutive model in order to transform an unsolvable "purely spatial" ICA problem into a well-behaved "space-time" problem. A recently proposed approach to this problem can be found in [17], in which, once more, nonlinear prediction is employed to "cut the Gordian knot". The essence of this proposal is to use a bank of nonlinear predictors to transform the convolutive problem shown in (12) into an instantaneous (purely spatial) ICA problem that can be perfectly solved with the aid of conventional techniques. The rationale of this filter bank is simply an extension of the nonlinear prediction-based solution discussed above to the case in which there are delayed versions of *multiple signals*. The fact that a highly undermodeled problem in a conventional spatial view becomes perfectly solvable is yet another proof of the usefulness of the idea of temporality in the source separation framework.

B. PCA, ICA, Prediction and Equalization

It is always curious and inspiring to observe how proposals of a very similar spirit eventually emerge in separate research fields. In the literature, as a rule, the idea of performing *independent component analysis* is exposed as a kind of sequel to the mathematically simpler notion of *principal component analysis*. Furthermore, many of the existing algorithms indeed make use of a preliminary PCA pre-whitening stage to facilitate the task of recovering the sources.

This "two-stage" approach is quite intuitive and appealing. Since the sources are supposed to be i.i.d., they are also uncorrelated; therefore, the process of whitening the available input can be thought of as a device for restricting the possible solutions to the more select group of uncorrelated vectors. This should facilitate the entire search process and give rise to faster and simpler algorithms. This association, so strong in the field of blind source separation, obliges us to ask: does this "two-stage" approach have a counterpart in SISO equalization theory[5]? We think that it is possible to give a positive answer to this question in two different ways.

1) Super-Exponential

The process of pre-whitening is by no means absent from the classical SISO equalization theory. It is well-known that, in the context of the Wiener approach, described in section 2.2 of part I, the process of eliminating the correlation between elements of the input vector can modify the mean-square error cost function in a way that enhances the speed of convergence of first-order methods such as the *steepest descent* and *LMS algorithms* [18].

Although an analogy with supervised techniques is valid, it would be more interesting to establish parallels between the pair PCA / ICA and a blind equalization algorithm. In this case, the most representative candidate is probably the super-exponential algorithm (SEA), which was discussed in section 2.6 of part I. The key to relating both worlds is the presence of a pre-whitening matrix that is responsible for a higher speed of convergence in comparison, for example, with a conventional gradient-based Bussgang technique. As shown in section 3.1, the SISO equalization problem can be formulated as an undermodeled ICA problem in which the sources are delayed versions of the transmitted message. This "spatial view" of an inherently "temporal problem" relates in a quite direct manner the "temporal whitening" performed by the matrix $\mathbf{R}^{-1}$ and the spatial whitening performed in PCA. Therefore, the whitening stage of the SEA can be understood as an equivalent of the PCA stage of a typical blind source separation algorithm.

After this first analogy is accepted, the next natural step is

to consider if the complementary aspect follows the same pattern. In our opinion, the parallel is perfectly valid, since the SEA uses kurtosis as a reference measure to recover a delayed version of the transmitted signal, i.e., to extract one of the sources. Therefore, the second stage of this blind technique corresponds to ICA in a "one-by-one" basis, as it is typical in algorithms devised to operate in SISO environments.

Finally, the comparison can be extended to multiuser detection algorithms [10][11], in which the presence of a correlation-based penalty term can be considered to be a sort of "tacit PCA" that must be continually carried out.

2) Magnitude-Phase Structures

The relationships we have hitherto exposed were predominantly based on *criteria* for blind equalization and source separation. However, it is interesting to notice that there is an appealing connection between PCA and ICA and a filtering structure conceived to operate in an unsupervised fashion: the *magnitude-phase equalizer*[6] [21].

The magnitude-phase equalizer is a device composed of two parts: a linear predictor and an all-pass filter endowed with an additional decision-feedback stage. The role of the linear predictor is, in accordance with the background presented in section 2.3 of part I, to equalize the magnitude response of the communication channel without paying particular attention to the phase response; this is achieved through a conventional whitening process. The output of the prediction-error filter thus designed is the input of a second structure, a nonlinear recursive filter whose objective is to compensate for the phase distortions originated in the transmission process. The parameters of this second device are adapted with the aid of a decision-directed criterion.

Let us consider for a while the filtering structure we have just described. Its first stage, as we have seen, is responsible for performing a substantial part of the equalization task: to invert the magnitude response of the channel by means of a whitening process. A successful primary stage, therefore, simplifies the possible solutions to the second part – only all-pass filters need to be considered. The combination between pre-whitening and further simplification allows us to associate the first stage of a magnitude-phase equalizer to the typical PCA performed in source separation problems.

Interestingly, the second stage contains a nonlinear element, which is expected, since blind phase equalization, as seen in sections 2.3 and 2.4 of part I, demands that higher-order statistics of the input signal be somehow generated. By tacitly resorting to this additional statistical information, the task is fulfilled and the phase response of the communication channel is adequately compensated for. Thus, what characterizes this second stage is the use of a memoryless nonlinearity to generate higher-order statistics, a stratagem that, as seen in section 3.2.2 of part I, is not unusual in ICA. Therefore, our analogy is complete: the phase equalizer plays a role similar to

that of the essential ICA stage of a blind source separation algorithm[7].

IV. CONCLUSIONS AND PERSPECTIVES

We started this work from a discussion, which took place in part I, of the basic concepts, criteria and algorithms belonging to the fields of blind equalization and blind source separation. There, it was possible to notice how the development of both branches followed different paths, even though, in some cases, the obtained results were not essentially dissimilar. This careful exposition was essential for a good comprehension of part II.

In part II, we, in the first place, revisited relations as the equivalence between maximum-likelihood formulation in BSS and the BGR theorem, between nongaussianity measures such as kurtosis and negentropy and the super-exponential and Bussgang techniques and between nonlinear PCA and the Bussgang approach. Although many of these connections were already mentioned in the literature, it is not of our knowledge the existence of a work that has considered the subject in a unified and systematic manner.

In addition to these, if we may say, more direct equivalences, we have also treated new aspects of the relation between these two problems – the idea of temporality and the connection between PCA, ICA, magnitude and phase equalization. This discussion enabled us to understand the problems of equalization and ICA from points of view that can bring new interesting solutions and interpretations to these known problems.

Finally, we conclude that, although both problems seem, at first sight, to be different, it is possible to place them under the same framework and realize that the solutions in both cases are fundamentally very similar.

REFERENCES

[1] O. Shalvi, E. Weinstein, "New criteria for blind deconvolution of nonminimum phase systems (channels)", IEEE Transactions on Information Theory, No. 36 Vol. 2, pp. 312–321, 1990.
[2] S. Bellini, "Bussgang Techniques for Blind Deconvolution", in *S. Haykin, Blind Deconvolution*, Prentice Hall, 1994.
[3] M. Mboup, P. Regalia, "A gradient search interpretation of the super-exponential algorithm", IEEE Transactions on Information Theory, Vol. 46, No 7, pp. 2731-2734, 2000.
[4] D. Donoho, "On minimum entropy deconvolution", in *Applied Time Series Analysis II*, D. F. Findley, Ed. New York: Academic, 1981.
[5] A. Hyvarinen, J. Karhunen, E. Oja, *Independent Component Analysis*, Wiley Interscience, 2001.
[6] E. Kofidis, "Blind Source Separation: Fundamentals and Recent Advances", Course Material, XIX Brazilian Telecommunications Symposium, Fortaleza, Brazil.

---

[5] It should be noted that this approach has already been explored in a MIMO channel equalization context [19][20]

[6] This is, essentialy, the same idea presented in [19], but in a different context (MIMO equalization problem).

---

[7] It should be noted that after a preliminary PCA stage, the role of ICA is exactly to perform a phase rotation.

[7] Y. Li, Z. Ding, "Convergence Analysis of Finite Length Blind Adaptive Equalizers", IEEE Transactions on Signal Processing, Vol. 43, No. 9, pp. 2121-2129, 1995.

[8] P. Regalia, "On the equivalence between the Godard and Shalvi-Weinstein schemes of blind equalization", Signal Processing, Vol. 73, No. 1-2, pp. 185-190, 1999.

[9] S. Haykin, *Adaptive Filter Theory*, 3rd edition, Prentice Hall, 1996.

[10] C. B. Papadias, A. Paulraj, "A constant modulus algorithm for multiuser separation in the presence of delay spread using antenna arrays", IEEE Signal Processing Letters, Vol. 4, No. 6, pp. 171-181, 1997.

[11] F. R. P. Cavalcanti, J. M. T. Romano, A. L. Brandão, "Least-Squares CMA with Decorrelation for Fast Blind Multiuser Signal Separation", International Conference on Acoustics, Speech and Signal Processing (ICASSP), Phoenix, USA, 1999.

[12] L. Tong, G. Xu, T. Kailath, "Blind identification and equalization based on second-order statistics: a time domain approach", IEEE Trans. On Information Theory, vol. 40, No 2, pp. 340-349, 1994

[13] C. R. Johnson, P. Schniter, T. J. Endres, J. D. Behm, D. R. Brown, R. A. Casas, "Blind equalization using the constant modulus criterion: a review", Proceedings of the IEEE, Vol. 86, No. 10, pp. 1927 – 1950, 1998.

[14] R. Suyama, R. R. F. Attux, J. M. T. Romano, M. Bellanger, "Relations entre les critères du module constant et de Wiener", Proceedings of the 19e Colloque GRETSI, Paris, France, 2003.

[15] C. C. Cavalcante, J. R. Montalvão Filho, B. Dorizzi, J. C. M. Mota, "A Neural Predictor for Blind Equalization in Digital Communication", Adaptive Systems for Signal Processing, Communication and Control (AS-SPCC 2000), Lake Louise, Canada, 2000.

[16] R. Ferrari, C. M. Panazio, R. R. F. Attux, C. C. Cavalcante, F. J. Von Zuben, L. N. de Castro, J. M. T. Romano, "Unsupervised Channel Equalization Using Fuzzy Prediction-error Filters", IEEE Workshop on Neural Networks for Signal Processing, Toulouse, France, 2003.

[17] R. Suyama, L. T. Duarte, R. Ferrari, L. E. P. Rangel, R. R. F. Attux, C. C. Cavalcante, F. J. Von Zuben, J. M. T. Romano, "A Nonlinear Prediction Approach to the Blind Separation of Convolutive Mixtures", submitted to the EURASIP Journal on Applied Signal Processing, Special Issue on Advances in Blind Source Separation.

[18] P. S. R. Diniz, *Adaptive Filtering: Algorithms and Practical Implementation*, Springer, 1997.

[19] L. Rota, P. Comon, S. Icart, "Blind MIMO Paraunitary Equalizer", Proc. ICASSP'03, Hong-Kong, April 2003

[20] P. Comon, L. Rota, "Blind Separation of Independent Sources from Convolutive Mixtures", IEICE Trans. on Fundamentals of Elec. Com. Comput. Sciences, E86-A(3), March 2003

[21] C. A. F. da Rocha, J. M. T. Romano, O. Macchi, "Self-learning deconvolution using a cascade of magnitude and phase filters", 38th IEEE Midwest Symposium on Circuits and Systems, Rio de Janeiro, Brazil, 1995.