Analysis of Modified 'Bussgang' Algorithms (MBA) for Channel Equalization

Simone Fiori

Abstract— **In our previous works we introduced two modified 'Bussgang' algorithms for blind channel equalization based on Bayesian iterative estimation of the source sequence. They were developed in order to reduce the computational complexity of the original 'Bussgang' algorithm as well as to make it more flexible by introducing a kind of source adaptivity. However, the previous work relied on some heuristic findings, validated by series of computer-based experiments. The aim of this paper is to present a theoretical investigation of some particular aspects of the adapting equations, namely, the steady-state conditions, in order to ameliorate the performances of the modified 'Bussgang' algorithms and to better explain their numerical behavior.**

Keywords— **Digital adaptive filtering; 'Bussgang' blind channel equalization; Bayesian estimation; Bernoulli numbers.**

I. Introduction

THE aim of blind channel equalization [3], [19] is to recover a source signal distorted by a transmission medium. Its most known applications are echo-graphic data focusing [5], optical memory-support storage and retrieval enhancement [6], equalization of communication channels [7] and seismic trace deconvolution for geophysical prospecting [21].

Among the most known blind equalization techniques, the 'Bussgang' one relies on the iterative Bayesian estimation of the source sequence. It is worth mentioning the existence of other Bayesian approaches to blind equalization, for example where the channel is modeled and the source and channel are jointly estimated and even tracked, recently including 'particle filtering' [4], [11], [15], [16], [18]. Some modifications to the basic 'Bussgang' algorithm have recently been proposed by the present Author in the previous contributions [8], [9].

In particular, the contribution [8] aimed at presenting a theoretical analysis and a numerical comparison of original 'Bussgang', of 'Bussgang' endowed with a flexible adaptive approximated estimator, and of the two algorithms modified by the help of the natural gradient theory, which allows accounting for the geometrical structure of filter parameters manifold. In our proposal, the original fixed-shape estimator was replaced by an adaptive function endowed with two adjustable parameters, which change during the filter adaptation phase in order to make the estimator matched to the adaptation progress. The corresponding algorithm will be hereafter referred to as modified 'Bussgang' algorithm with two adjustable parameters (MBA-2). The conclusion of paper [8] was that the 'Bussgang' algorithm endowed with flexible adaptive Bayesian approximated estimator is effective, converges faster than the algorithm with fixed approximated estimator, and is able to attain better numerical results in terms of interference residual.

In [9], the number of adjustable parameters was reduced to one by properly modeling their dependency. Such one-parameter algorithm will be hereafter referred to as MBA-1. Paper [9] showed that this algorithm is effective and exhibits fast convergence and low computational complexity compared to the original 'Bussgang' and their natural-gradient-based versions.

From a critical examination of contributions [8], [9], at least three questions arise:

1. Apparently, the performances of MBA-1 algorithm does depend on the parameter of the model that explains the dependency among the coefficients of the approximated estimator. Is it possible to determine its optimal value ?
2. Both MBA-2 and MBA-1 algorithms are effective and show superior performances in comparison to the original 'Bussgang' algorithm, but how do the performances of the MBA-1 and the older MBA-2 compare ?
3. Can any adaptive version of 'Bussgang' algorithm be envisaged and which merits and demerits would it exhibit with respect to the MBAs ?

The aim of this paper is to provide an investigation of the above questions. In particular, section II briefly summarizes the new algorithms introduced in [8], [9]. Section III is devoted to a theoretical investigation of the steady-state equations for these algorithms, while in section IV some experimental results are discussed in order to numerically assess the soundness of the theoretical findings. Section V concludes the paper.

II. Blind Equalization by Iterative Bayesian Estimation

The aim of the present section is to summarize the basic theory of blind channel equalization by iterative Bayesian estimation and the new algorithms introduced in [8], [9]. Also, a new adaptive 'Bussgang' algorithm is proposed for comparison purpose.

A. Basic blind equalization theory

The linear system to equalize is described by the following model:

$$u(t) = \mathbf{h}^T \mathbf{s}(t) + \mathcal{N}(t) \ , \qquad (1)$$

where $\mathbf{s}(t)$ is the system's input vector-stream at time t, namely $\mathbf{s}(t) \stackrel{\text{def}}{=} [s(t)\ s(t-1)\ s(t-2)\ \ldots\ s(t-L_h+1)]^T$, with L_h being the length of the impulse response $\mathbf{h}$. The signal $\mathcal{N}(t)$ represents a channel disturbance, such as a zero-mean AWGN, uncorrelated with the source signal. Normally, the effect of additive noise is neglected because it is understood that the degradation of the performance of data transmission is dominated by the inter-symbol interference [2] and that, in digital transmissions relying on discrete alphabets, symbol estimation is not affected by small perturbations [3]. Thus in the analytical development we shall neglect the term $\mathcal{N}(t)$, while the effect of additive noise will be considered again in the section devoted to numerical experiments in order to survey the equalization performance degradation due to operation over non-ideal channels.

The following assumptions are considered [14]: The channel's impulse response satisfies $\mathbf{h}^T\mathbf{h} = 1$ and its inverse has finite energy; we further assume the channel is non-minimum phase; the source sequence is IID and has $\mathbb{E}_s[s(t)] = 0$ and $\mathbb{E}_s[s^2(t)] = 1$; also, if $p_s(s)$ denotes the probability density function (PDF) of the input signal, it is supposed that $p_s(s) = p_s(-s)$ and that

The author is with the Facoltà di Ingegneria dell'Universitá di Perugia – Polo Scientifico e Didattico del Ternano, Loc. Pentima bassa, 21, I-05100 Terni (Italy). Email: fiori@unipg.it

it is non-Gaussian. Some comments on these assumptions are in order. The hypothesis on channel's phase-distortion is assumed in order to make the model represent real-world communication channels, such as the telephonic ones. The hypotheses on the channel energy and the source sequence power are related to the use of Automatic Gain Control (AGC) and helps removing the uncertainty in the recovering of source signal amplitude inherent to blind processing. The hypotheses on the PDF of the source stream are related to well-known identifiability conditions [3].

In the noiseless limit, a filter described by the vector impulse response $\mathbf{w}$ is the approximate inverse of system (1) if $\mathbf{w}$ cancels – to the maximum permitted extent – the effects of $\mathbf{h}$ on the source signal. Denoting by $\mathbf{u}(t)$ the vector containing the filter input samples $\mathbf{u}(t) \stackrel{\text{def}}{=} [u(t)\ u(t-1)\ u(t-2)\ \ldots\ u(t-L_w+1)]^T$, with L_w being the number of tap-weights in $\mathbf{w}$, the output of the filter writes:

$$x(t) = \mathbf{w}^T(t)\mathbf{u}(t) \ . \tag{2}$$

Since $\mathbf{h}$ and $s(t)$ are both unknown, the filter $\mathbf{w}_\star$ such that $x(t) \sim s(t)$ has to be *blindly* identified possibly by means of an iterative algorithm. From the basic theory of blind channel equalization it is known that the source signal may be recovered up to arbitrary time-delay [14] while, because of the assumed hypotheses, the amplitude uncertainty reduces to the sign.

As a general remark, only in a high signal-to-noise environment the degradation of the performance of data transmission is dominated by the inter-symbol interference, hence the blind equalization algorithms are sufficient to provide symbol/data recovery. Otherwise, channel coding is necessary.

Even in the noiseless case, the inversion of the channel (1) cannot be performed exactly by means of the FIR filter (2) because the true inverse is IIR; also, during filter adaptation the misadjustment of filter's coefficients makes the filter output signal differ from the source signal. Both phenomena are described by the following filter output signal model:

$$x(t) = cs(t-\delta) + n(t) \ , \tag{3}$$

where $n(t)$ is the so-called *convolutional noise*, δ is a finite delay and $c \in \mathcal{R}$ is an amplitude factor. The statistical properties of convolutional noise have been completely characterized in [14]: It may be described as a zero-mean IID process of time-varying variance, denoted here with $\sigma^2(t)$; also, from equation (3) the noise $n(t)$ appears to be the weighted sum of many statistically independent and identically distributed random variables, thus, provided the channel-filter combined impulse response is sufficiently long, the Central Limit Theorem of statistics applies and makes a Gaussian model for the convolutional noise plausible. In conclusion, the convolutional noise is modeled as a zero-mean AWGN that is statistically independent of the source sequence. It is worth noting that, provided the filter adaptation algorithm converges, the convolutional noise power σ^2 decreases to low values.

From the model (3) of the filter response, a way can be envisaged to get an estimate of the source sequence $s(t)$ by means of a statistical estimator. In fact, the model (3) reveals that the relationship between $x(t)$ and $s(t)$ is deterministic but for the convolutional noise. On the basis of such model and of the mentioned assumptions on the convolutional noise, a memoryless estimator $\hat{s}(t-\delta) = b(x(t))$ was suggested in [2], [13]. An appropriate estimator $b(\cdot)$ is suggested from Bayesian estimation theory, which allows defining a non-linear estimator that guarantees the minimal distortion between the true and the estimated source sequence. Such estimator coincides with the conditional mean function [14]:

$$b(x) = \mathbb{E}_s[s|x] = \frac{\sigma^2}{c}\left[\frac{p'_x(x)}{p_x(x)} + \frac{x}{\sigma^2}\right] \ , \tag{4}$$

where the symbol $\mathbb{E}[\cdot]$ denotes statistical expectation (ensemble average).

The derivation of the above estimator is exact and relies on the filter's output signal PDF $p_x(x)$. Full details about its computation have been supplied e.g. in [8]. Provided that the source statistics is known, the statistics of the channel response might be evaluated and the required estimator be computed, as well. It is worth remarking that the chosen estimator explicitly depends on the convolutional noise power, whose estimation is quite burdensome. Moreover, this variable is normally treated as a constant parameter of the estimator, to be decided during the algorithm's design phase, while it actually varies according to filter adaptation progress.

On the basis of the described estimator, in [2] an error criterion like:

$$\tilde{C}(\mathbf{w}) = \frac{1}{2}\mathbb{E}_x\left[(cb(x) - x)^2\right] \tag{5}$$

had been proposed. The minimization of the cost function $\tilde{C}$ may be achieved also by means of a stochastic gradient steepest descent (SGSD) algorithm. In the present context this rule assumes the following expression:

$$\Delta\mathbf{w} = -\eta[cb'(x) - 1][cb(x) - x]\mathbf{u} \ , \tag{6}$$

where η is a positive adapting step-size and $b'(x)$ denotes the derivative of the function $b(x)$ with respect to x.

B. Previous contribution on 'Bussgang' filtering

A key observation in the design of simplified estimators is that, for uniformly distributed source sequences, a suitable approximation of the actual Bayesian estimator $b(x)$ is the bilateral 'sigmoidal' function [14]:

$$\hat{b}(x) = \frac{a}{c}\tanh(\lambda x) \ , \tag{7}$$

with a and λ being properly selected parameters, where the values of a and λ were obtained by fitting the expression (7) with the actual estimator (4) for a given convolutional noise level. The corresponding expression for the updating rule (6) was given in [8].

In order to allow convolutional noise adaptivity and to avoid its repeated estimation, in the contribution [8] we proposed to adapt a and λ through time by means of a SGSD algorithm applied to $\tilde{C}$, thought of as a function of a, λ and $\mathbf{w}$. The obtained updating equations – constituting the MBA-2 adapting algorithm – are reported below because they are the basis for the following analytical considerations:

$$\Delta a = -\eta_a\left[c\hat{b}(x) - x\right]\frac{c\hat{b}(x)}{a} \ , \tag{8}$$

$$\Delta\lambda = -\eta_\lambda\left[c\hat{b}(x) - x\right]\left[a^2 - c^2\hat{b}^2(x)\right]\frac{x}{a} \ . \tag{9}$$

Here η_a and η_λ are constant positive adapting step-sizes.

From the experimental results reported in the contribution [8], the value of the product $a\lambda$ was noted to asymptotically converge to a constant value around 1. This phenomenon was also analytically investigated by approximately solving the steady-state equations for the adapting system by truncating the MacLaurin expansions of the non-linear functions to the *first-order* terms [8].

Motivated by the observation that the product $a\lambda$ tends toward some constant value Λ, in the manuscript [9] the scaling variable a was replaced with $\frac{\Lambda}{\lambda}$, so that the only adjustable variable remained λ, which adapts by:

$$\Delta\lambda = -\eta(c\hat{b}(x) - x)\left[-\frac{c}{\lambda}\hat{b}(x) + \frac{\Lambda x}{\lambda} - \frac{\lambda x c^2}{\Lambda}\hat{b}^2(x)\right] . \quad (10)$$

Also, the adjusting rule for $\mathbf{w}$, obtained by replacing $\hat{b}$ for b in equation (6), reads:

$$\Delta\mathbf{w} = -\eta(c\hat{b}(x) - x)\left[\Lambda - 1 - \frac{\lambda^2 c^2}{\Lambda}\hat{b}^2(x)\right]\mathbf{u} , \ \mathbf{w} \leftarrow \frac{\kappa\mathbf{w}}{\|\mathbf{w}\|} , \quad (11)$$

where the last operation gives a simple way to implement the AGC, that is to constrain the norm of $\mathbf{w}$ to a constant $|\kappa|$. The constant η is again a positive adapting step-size. The above described equations give the MBA-1 algorithm.

In the experiments carried out in [9], a good value for the constant Λ was set. However, this finding was completely heuristic and was validated through a series of experiments, but its value was not related to the actual behavior of adapting equations at equilibrium.

About the convergence of the above adapting algorithm we refer the Reader to [8], [9].

C. Proposal of an adaptive 'Bussgang' algorithm

To end with, it is worth mentioning the interesting possibility to make the original 'Bussgang' algorithm adaptive by iteratively estimating the convolutional noise level. In fact, on the basis of the model (3) and on the statistical characterization proposed for the convolutional noise, the following equations hold true:

$$\mathbb{E}_x[x^2] = c^2\mathbb{E}_s[s^2] + \mathbb{E}_n[n^2] ,$$
$$\mathbb{E}_x[x^4] = c^4\mathbb{E}_s[s^4] + \mathbb{E}_n[n^4] + 6c^2\mathbb{E}_s[s^2]\mathbb{E}_n[n^2] .$$

By hypothesis $\mathbb{E}_s[s^2] = 1$, due to the Gaussianity of the noise $\mathbb{E}_n[n^2] = \sigma^2$ and $\mathbb{E}_n[n^4] = 3\sigma^4$. Therefore, the above system of equations may be cast as follows:

$$c = \sqrt{\mathbb{E}_x[x^2] - \sigma^2} , \quad (12)$$
$$Z(\sigma) \stackrel{\text{def}}{=} \mathbb{E}_x[x^4] - (\mathbb{E}_x[x^2] - \sigma^2)^2\mathbb{E}_s[s^4] - 9\sigma^2 + 6\mathbb{E}_x[x^2]\sigma^2 . \quad (13)$$

The zero of the function $Z(\sigma)$ represents an estimate of the standard deviation of the convolutional noise[1], while the first equation gives the corresponding value of the scaling factor c. The 'Bussgang' algorithm [2], [14] where the convolutional noise level is iteratively estimated with the above-mentioned method will be referred to, in the following, as adaptive 'Bussgang' algorithm (ABA)[2].

It is worth noting that the use of the above formulas implies the availability of a whole filter-output signal segment as well as the source kurtosis, thus the ABA is inherently a batch-type algorithms. Also, note that on-line high-order moment estimation is also possible if this does not lead to a significant decrease on the algorithm performance.

[1] In the experiments, the 'fzero' zero-crossing procedure provided by MATLAB was employed, which uses a combination of bisection, secant, and inverse quadratic interpolation methods.

[2] To the best of our knowledge, we are not aware of the existence of any similar proposal in the scientific literature (it is interesting to note that in the contribution [2] it is written "Preliminary results show, however, that small improvements (if any) may be expected by [estimator] adaptation.")

III. Steady-State Analysis

As already noted in the previous section II-B, the algorithms MBA-1 and MBA-2 are linked though the relationship among their parameters: The MBA-1 adapting equations are obtained by the empirical observation that there exists a relationship between the two parameters of the MBA-2 algorithm, namely, that $a\lambda = \Lambda$ which is postulated to be a constant of the problem.

The question is how to define a general procedure that allows to find an optimal value of the constant Λ for the equalization problem at hand. An answer to this question comes from the postulate that the optimal Λ should satisfy some general-purpose equations involving the MBA-2's parameters. As general-purpose set of equations we selected the equilibrium conditions for the adapting rules (8)-(9) that describe the situation in which the adaptive-filter parameters stop adjusting (in the mean-field sense). Therefore, we postulate that *the optimal parameter Λ arises from the steady-state state analysis of the MBA-2 adapting equations.*

The proposed analysis begins by considering the mean-field equilibrium conditions for the adapting equations (8) and (9):

$$\mathbb{E}_s[a^2\tanh^2(\lambda x) - ax\tanh(\lambda x)] = 0 , \quad (14)$$
$$\mathbb{E}_s[(a\tanh(\lambda x) - x)(1 - \tanh^2(\lambda x))x] = 0 . \quad (15)$$

As already mentioned, in the present analytical developments we consider noiseless channels. We also take advantage of the fact that the conditions (14) and (15) represent the optimality of the parameters of the estimator, which implies perfect equalization[3], that means $x = cs$. These give rise to two non-linear equations in two unknowns (a and λ) whose solution is looked for (at equilibrium, c is known and equals the AGC gain $|\kappa|$).

It is worth remarking that the above equations imply integration with respect to the source statistics, thus, for example, one of the involved terms is:

$$\mathbb{E}_s[x\tanh(\lambda x)] = \int_{-\infty}^{+\infty} cs\tanh(\lambda cs)p_s(s)ds .$$

Therefore, the study of the solutions of the above steady-state equations relates to the statistical features of the source. In the following we consider two statistical distributions of interest: The uniform distribution and the binary (± 1-valued) one.

A. Analysis of the uniform-distribution case

In order to possess zero mean and unit variance, a uniform distribution should range within $[-\sqrt{3}, +\sqrt{3}]$, namely:

$$p_s(s) = \begin{cases} \frac{1}{2\sqrt{3}}, & s \in [-\sqrt{3}, +\sqrt{3}] , \\ 0, & \text{elsewhere} . \end{cases} \quad (16)$$

With this probability density function, the system of steady-state non-linear equations cannot be solved directly, thus we propose to expand the involved non-linear functions in MacLaurin series in order to be able to integrate term-by-term the involved non-linearities with respect to the source PDF. In this manner, polynomial equations are obtained, which appear easier to solve by the help of a suitable numerical procedure.

The invoked expansions are:

$$\tanh(x) = \sum_{n=1}^{\infty} E_{2n-1}x^{2n-1} , \ \tanh^2(x) = \sum_{n=1}^{\infty} S_{2n}x^{2n} ,$$
$$\tanh^3(x) = \sum_{n=1}^{\infty} Q_{2n+1}x^{2n+1} .$$

[3] Note that, from the point of view of the statistical description of concern in the present analysis, the time-delay between $x(t)$ and $s(t)$ is negligible.

The expansion coefficients E_n are known from calculus, while coefficients S_n and Q_n have been introduced in order to simplify the notation and can be computed on the basis of the fundamental coefficients E_n. The MacLaurin expansion coefficients of the hyperbolic tangent functions are known to be [22]:

$$E_{2n-1} = \frac{(-1)^{n-1}2^{2n}(2^{2n}-1)B_n}{(2n)!} \ , \ E_{2n} = 0 \ ,$$

where B_n denotes the n^{th} Bernoulli number. From basic algebra it is known that the coefficients of a polynomial computed as the product of two polynomials are given by the convolution $\star$ of the coefficients of the polynomial factors [1]. This means that $S_n = E_n \star E_n$ and $Q_n = E_n \star S_n$.

From the above expansions, the terms in equation (14) write:

$$a^2 \tanh^2(\lambda x) = \sum_{n=1}^{\infty} a^2 S_{2n} \lambda^{2n} x^{2n} \ ,$$

$$-ax\tanh(x) = -\sum_{n=1}^{\infty} aE_{2n-1} x^{2n} \lambda^{2n-1} \ .$$

By summing hand-by-hand and multiplying both sides by λ^2 we get:

$$\sum_{n=1}^{\infty} [(a\lambda)^2 S_{2n} - (a\lambda)E_{2n-1}](c\lambda)^{2n} \mathbb{E}_s[s^{2n}] = 0 \ . \qquad (17)$$

In a similar way, by summing the four terms arising from condition (15) and multiplying again both sides by λ^2, we obtain the new condition:

$$\sum_{n=1}^{\infty} [(a\lambda)E_{2n-1}\mathbb{E}_s[s^{2n}](c\lambda)^{2n} - (a\lambda)Q_{2n+1}\mathbb{E}_s[s^{2n+2}](c\lambda)^{2n+2} + S_{2n}\mathbb{E}_s[s^{2n+2}](c\lambda)^{2n+2}] = (c\lambda)^2 \mathbb{E}_s[s^2] \ . \qquad (18)$$

These two equations contain three quantities (a, λ and c), however, remarkably they always appear only within the products $a\lambda$ and $c\lambda$. The following definitions prove useful:

$$\Lambda \stackrel{\text{def}}{=} a\lambda \ , \ \Gamma \stackrel{\text{def}}{=} c\lambda \ , \ \mu_{2n} \stackrel{\text{def}}{=} \mathbb{E}_s[s^{2n}] \ .$$

Note that, by definition, the source PDF is symmetric, thus odd-order moments are zero. The even-order moments of a uniform distribution are $\mu_{2n} = \frac{3^n}{2n+1}$ for $n \geq 1$.

From a practical point of view, the resulting polynomial equations have infinitely many terms and cannot be solved numerically, thus we should resort to truncation. Let us denote by N the number of terms in the approximation of the hyperbolic tangent function; we then make use of the following truncated expansions:

$$\tanh(x) = \sum_{n=1}^{N} E_{2n-1} x^{2n-1} \ , \ \tanh^2(x) = \sum_{n=1}^{2N-1} S_{2n} x^{2n} \ ,$$

$$\tanh^3(x) = \sum_{n=1}^{3N-2} Q_{2n+1} x^{2n+1} \ .$$

In fact, if 'tanh' is approximated with N (nonzero) terms, then its square would have $2N-1$ terms and its cube would exhibit $3N-2$ terms, as noted from the convolutional expression of S_n and Q_n and from the fact that the length of a sequence obtained from the convolution of two finite-length sequences equals the sum of their length minus one.

With the truncated expansions and the new variables Λ and Γ, the conditions (17) and (18) write:

$$\Lambda \sum_{n=1}^{2N-1} S_{2n}\Gamma^{2n}\mu_{2n} = \sum_{n}^{N} E_{2n-1}\Gamma^{2n}\mu_{2n} \ , \qquad (19)$$

$$\Lambda \sum_{n=1}^{N} E_{2n-1}\Gamma^{2n}\mu_{2n} - \Lambda \sum_{n=1}^{3N-2} Q_{2n+1}\mu_{2n+2}\Gamma^{2n+2} + \sum_{n=1}^{2N-1} \Gamma^{2n+2}\mu_{2n+2} = \Gamma^2\mu_2 \ . \qquad (20)$$

In conclusion, by properly arranging the involved terms, the resolving system of polynomial equations becomes:

$$\Lambda = \frac{\sum_{n=1}^{N} E_{2n-1}\Gamma^{2n}\mu_{2n}}{\sum_{n=1}^{2N-1} S_{2n}\Gamma^{2n}\mu_{2n}} \ , \qquad (21)$$

$$(\Lambda-1)\mu_2 + \sum_{n=1}^{N}(\Lambda E_{2n+1} + S_{2n} - \Lambda Q_{2n+1})\mu_{2n+2}\Gamma^{2n} + \sum_{n=N}^{2N-1}(S_{2n} - \Lambda Q_{2n+1})\mu_{2n+2}\Gamma^{2n} - \sum_{n=2N}^{3N-2} \Lambda Q_{2n+1}\mu_{2n+2}\Gamma^{2n} = 0 \ . \qquad (22)$$

It is worth remarking that the equation (21) directly stems from equation (19), while the terms in formula (22) are the terms in the relation (20) which have been re-grouped so that they read as non-overlapping monomials constituting the polynomial equation. It is also important to remark that the polynomial equation (22) holds only for $N \geq 2$. The particular case $N = 1$ must be studied apart, and is interesting just because it confirms the weak analysis results already obtained in the previous work [8]. In this special case, the resolving system writes:

$$\begin{cases} \Lambda^2 S_2 \Gamma \mu_2 = \Lambda \Gamma \mu_2 \ , \\ \Lambda E_1 \Gamma \mu_2 - \Lambda Q_3 \mu_4 \Gamma^4 + S_2 \Gamma^4 \mu_4 = \Gamma^2 \mu_2 \ . \end{cases}$$

As $E_1 = S_2 = Q_3 = \mu_2 = 1$, the above system gives $\Lambda = 1$.

The system of equations (21) and (22) can be solved by a relaxation method. The following sequence of operations may be carried out iteratively: a) Fix the order N and an initial value for Λ; b) Replace the current solution of Λ in equation (22) and find the real-valued positive root of such polynomial; c) Solve equation (21) for Λ; d) return to (b) until the values of Γ and Λ stabilize.

The asymptotic value of the two unknowns may be found numerically through the above numerical procedure, provided its overall consistency. In fact, it is important to note that, for a fixed N, the polynomial equation (22), in the unknown Γ^2, has $3N-2$ roots, therefore, the procedure is consistent only if at most one among such roots is real and positive. Also, we must ensure that during the iteration the chosen root Γ does not break down to zero. We experimentally verified that these conditions are met only for N even. Under these conditions, the procedure quits after 5/6 iterations, at least for small values of N, and never takes more than 10 iterations to stabilize for any value of N in the considered range.

The Figure 1 shows the final result of 10 iterations of the numerical procedure for N ranging from 2 to 20: The last approximated value is $\Lambda \approx 1.185$. This is the result expected from

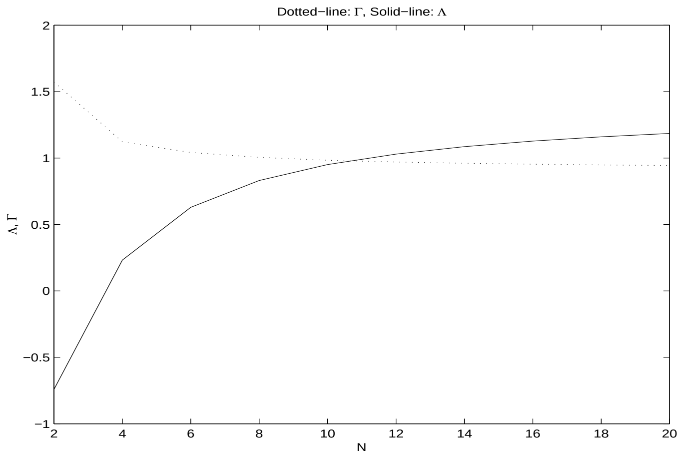

Fig. 1. Approximated values of Λ (solid-line) and Γ (dashed-lines).

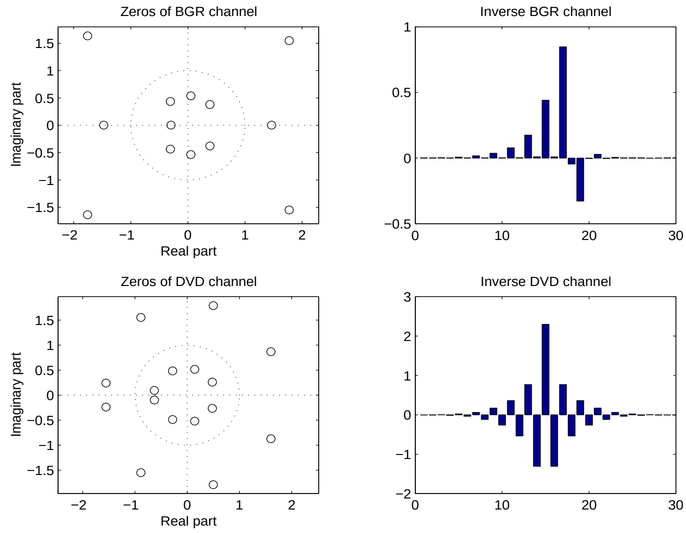

Fig. 2. Sampled telephonic channel and DVD-ROM equivalent channel: Zero-plot and inverses.

the presented steady-state analysis: The above value for the parameter Λ will be used in the following section in order to run the MBAs.

B. Analysis of the binary-distribution case

Analyzing the steady-state behavior of MBA-2, in view of optimizing MBA-1, in presence of a binary source is useful in practical applications and, remarkably, gives rise to tractable mathematics, so that the steady-state equations may be explicitly solved. The distribution may be assumed as:

$$p_s(s) = \frac{1}{2}[\delta(s-1) + \delta(s+1)] \ , \qquad (23)$$

where $\delta(u)$ denotes the Dirac's delta.

In this case, the integrals involved in the equations (14) and (15) may be explicitly computed, so that the steady state equations take on the form:

$$a^2 \tanh^2(\lambda c) - ac\tanh(\lambda c) = 0 \ , \qquad (24)$$

$$ac\tanh(\lambda c) - ac\tanh^3(\lambda c) - c^2 + c^2\tanh^2(\lambda c) = 0 \ . \qquad (25)$$

By multiplying both sides of both equations by λ^2 and by invoking again the definitions of Λ and Γ, we get:

$$\Gamma = \Lambda\tanh(\Gamma) \ , \qquad (26)$$

$$\Gamma\Lambda\tanh(\Gamma) - \Gamma\Lambda\tanh^3(\Gamma) - \Gamma^2 + \Gamma^2\tanh^2(\Gamma) = 0 \ , \qquad (27)$$

where the hypothesis $\Gamma \neq 0$ has been used. It is now easy to see that the first equation replaced in the second one makes it identically satisfied. Therefore, in the binary case, there exist infinitely many values of the constant Λ that satisfy the optimality conditions, due to the trivial structure of the source distribution. However, it is readily seen that if $\Lambda \leq 1$ then the only solution to the equation $\Gamma = \Lambda\tanh(\Gamma)$ in the unknown Γ would be $\Gamma = 0$, while for $\Lambda > 1$ there exists one (and only one) additional nonzero solution.

IV. Experimental Results

The aim of this part is to present computer-based experimental results pertaining to the MBAs, in order to gain an insight into their numerical behavior. The algorithms are tested on the sampled telephonic channel (termed BGR) described in [3] and the DVD-ROM equivalent channel described in [17]. Their features are depicted in Figure 2. In the experiments, the equalization accuracy is measured through the inter-symbol interference (ISI) residual defined as in [20]. Also, in the binary case the accuracy of equalization is measured by the help of the bit-error-rate (BER), defined conventionally.

In the experiments, the initial impulse response of the filter $\mathbf{w}(0)$ was assumed a null sequence, except for the central weight that equals the AGC gain $|\kappa|$. Also, for the MBA-2 it was always chosen $a^{\text{init}} = 1$, $b^{\text{init}} = 1$, $\alpha = 0.001$ and $\beta = 0.01$, while for MBA-1 it was always assumed $\lambda^{\text{init}} = 1$.

As a general remark on the choice of algorithms' parameters such as κ and η, it is worth noting that the value of η substantially controls the convergence speed (the larger η, the speedier the convergence) but it also affects the steady-state precision (the larger η, the larger the oscillations of the filter tap-weights around the optimal solution). About the choice of the value of κ, general considerations have been presented in [8]; in the present paper we basically switch between the values $\kappa = 0.5$ and $\kappa = 1$ that grant convergence.

In the present section, first some single-trial experiments on algorithm MBA-1 are illustrated in order to validate and corroborate the analytical findings of section III. Then, a comparison of MBA-1, MBA-2, adaptive 'Bussgang' algorithm (ABA) and constant-modulus algorithm (CMA, [12]) are presented and discussed.

A. Experiments on MBA-1 algorithm evaluation and validation

The results pertaining to the BGR channel are illustrated in Figure 3; the parameters value are: $L_w = 20$ tap-weights, 100 epochs, 1000 samples/epoch, $\eta = 0.01$, and $\kappa = 0.5$. The results are very good, confirming the good behavior of the MBA-1 algorithm. The results obtained with the equivalent DVD system are displayed in Figure 4; the parameters value were: $L_w = 20$ tap-weights, 200 epochs, 500 samples/epoch, $\eta = 0.15$, and $\kappa = 1$. Even in this single-trial case the results are good. The value of Λ arising from the analysis of section III was adopted.

In the above experiments, the number of filter-taps was set according to the results of a pre-validation process, which shows that the best trade-off between minimal interference residual and filter structure complexity is attained in correspondence

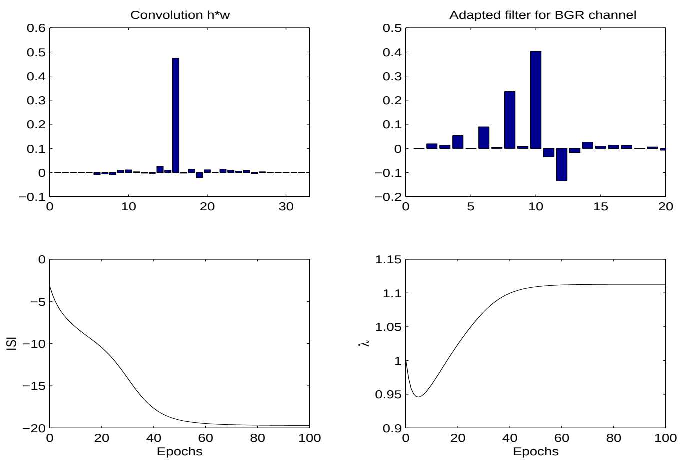

Fig. 3. Telephonic-channel equalization by MBA-1 with uniformly-distributed source signal: Convolution of **h** and **w** after adaptation, adapted inverse filter, ISI and λ during the adatping phase.

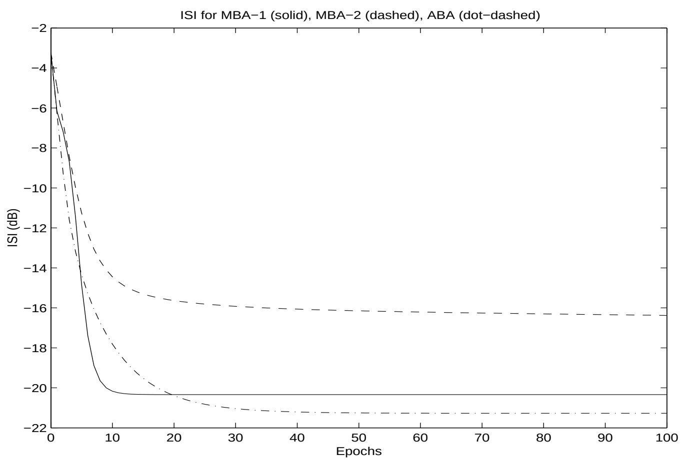

Fig. 5. Comparison of BGR equalization results. ISI of the MBA-1, MBA-2 and ABA algorithms during the adaptation phase.

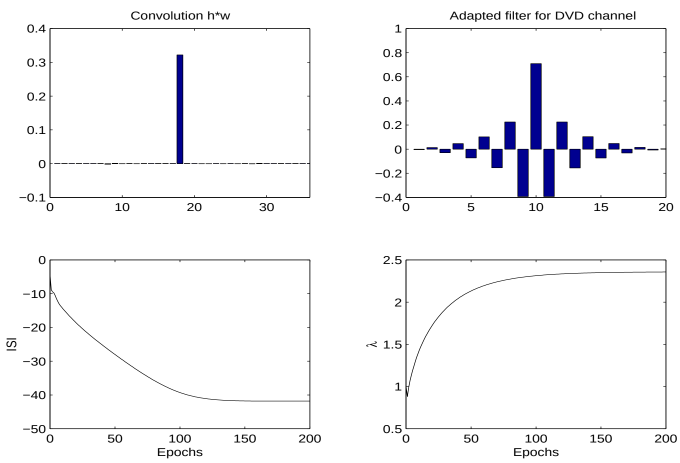

Fig. 4. DVD-ROM equalization by MBA-1 with binary source signal: Convolution of **h** and **w** after adaptation, adapted inverse filter, ISI and λ during the adapting phase.

to 20 tap-weights. These results confirm the suitability of the analytical investigation that led to the optimal value of the parameter Λ. Note that these results are reported for illustration purpose and do not pertain to optimally-chosen values of the parameters κ, η, etc.

B. *Comparative experiments on noiseless channels*

The following experiments aim at comparing the behavior of MBA-1 and MBA-2 with the behavior of adaptive 'Bussgang' algorithm over the BGR channel and with the behavior of Godard's constant modulus algorithm on DVD-ROM. In particular, in the second case, the storage of a textual string and its retrieval, without and with the equalization system, have been simulated in order to obtain some qualitative information on equalization suitability.

About the experiment on noiseless BGR channel, it has been performed by comparing the MBAs and ABA in presence of a uniformly-distributed source signal. In particular, the number of filter taps was set to $L_w = 20$, the number of adapting epochs was set to 100 with 1000 source-samples per epoch. The adapting parameters for the algorithms were: MBA-1: $\eta = 0.1$, $\kappa = 0.5$; MBA-2: $\eta = 0.004$, $\kappa = 0.5$. ABA: $\kappa = 1$ and $\eta = 0.005$. The result of adaptation are illustrated in the Figure 5. As clearly evidenced by the graphs, in this case the MBA-1 algorithm outperforms the MBA-2 one, while the ABA, which relies on a well-tuned exact estimator, outperforms both algorithms about equalization ability. The MBA-1 exhibits the faster convergence, in this experiment.

About the experiment on DVD-ROM equivalent channel, after definition of a proper textual string, it has been back-converted to a binary sequence using standard ASCII encoding, then storage/retrieval over a DVD system has been simulated by passing the binary sequence through the corresponding equivalent channel and by successively passing the distorted sequence through the adapted filter. A simple threshold-based decision algorithm has finally been used in order to obtain again a ± 1 sequence that, after time-tails removal, has been converted to a text-string. The string used in the experiment was 'Signal Processing Task'.

In order to compare the behavior of the three algorithms, in this experiment the adaptation parameters were chosen in order to emphasize their behavior in terms of bit-error-rate. In fact, in the present (binary, noiseless) case it is not necessary to achieve very low residual ISI values in order to vanish the BER. The selected values where: 2 taps-weight, 10 adapting epochs with 50 samples/epoch.

The figure 6 shows the obtained input/output sequences. Before equalization the storage/retrieval BER was equal to about 0.71%. As it is easily recognized, the source string may be perfectly recovered after DVD-ROM equalization, leading to a zero BER. The Figure 7 shows the values of the ISI and BER indices for the three algorithms, as computed at the end of each adapting epoch; it also shows the ISI-BER plane that empirically explains the correspondence of the two indices for the present equalization problem. These results pertain to the following parameter values: MBA-1: $\kappa = 1$ and $\eta = 0.005$; MBA-2: $\eta = 0.004$ and $\kappa = 0.5$; CMA: $\eta = 0.0005$.

Also, two important elements in the comparison are considered: The number of flops and the CPU-computation time. They refer to a MATLAB code implemented on a 500MHz machine with 64MB memory and are reported in Table I. The substantial equivalence about computational complexity of the MBAs may be explained as follows: Though the MBA-2 requires two adapting equations to update the two tunable parameters in the flexible Bayesian estimator, while MBA-1 requires only one of such rules, the MBA-2 does not use a variable adapting stepsize, while in MBA-1 the computational burden due to

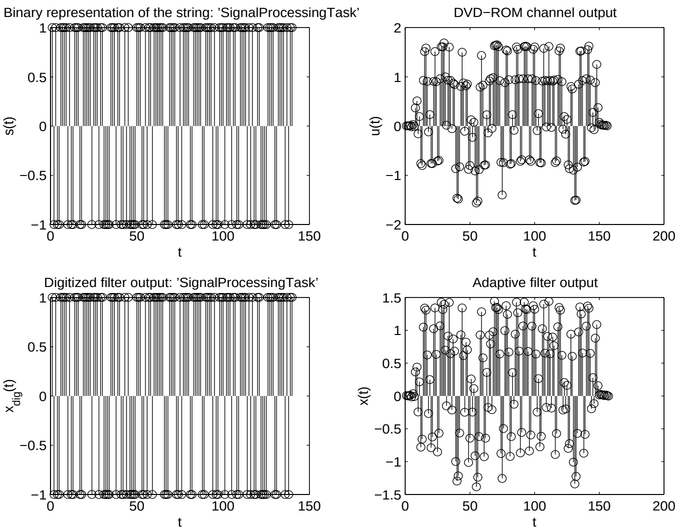

Fig. 6. DVD-ROM inverse filter experiment with text string as input: $s(t)$ denotes the coded string; $u(t)$ arises when the source sequence is passed through the DVD-ROM channel in order to simulate storage/retrieval operation; $x(t)$ is the result of passing the channel output through the equalization filter, while $x_{\rm dig}(t)$ is the result of thresholding.

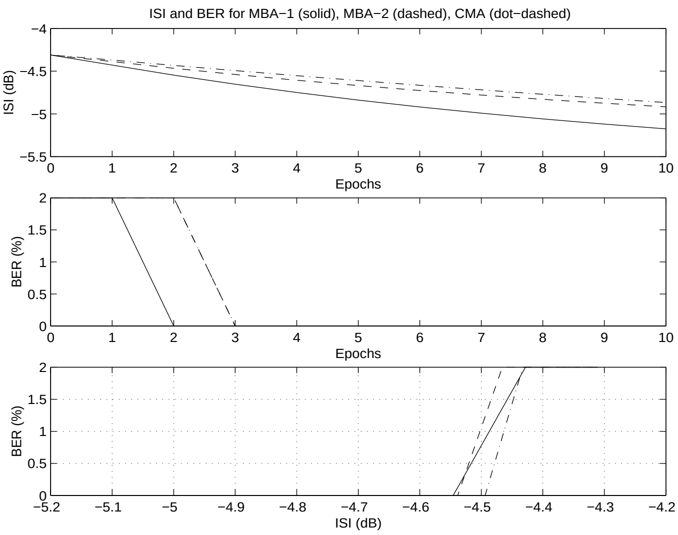

Fig. 7. Comparison of DVD-ROM equalization results. Top: ISI of the MBA-1, MBA-2 and CMA algorithms during the adaptation phase. Middle: BER of the three algorithms during adaptation. Bottom: ISI-BER plane.

Algorithm	Flops	Total time (sec.s)
MBA-2	270.66	101.11
MBA-1	274.23	82.93
CMA	221.68	32.3
ABA	421.24	296.44

TABLE I

Computational complexity (flops per iteration, averaged over 200 epochs of 500 samples each) and total computation time (500×200 iterations) over a 20-taps long transversal filter.

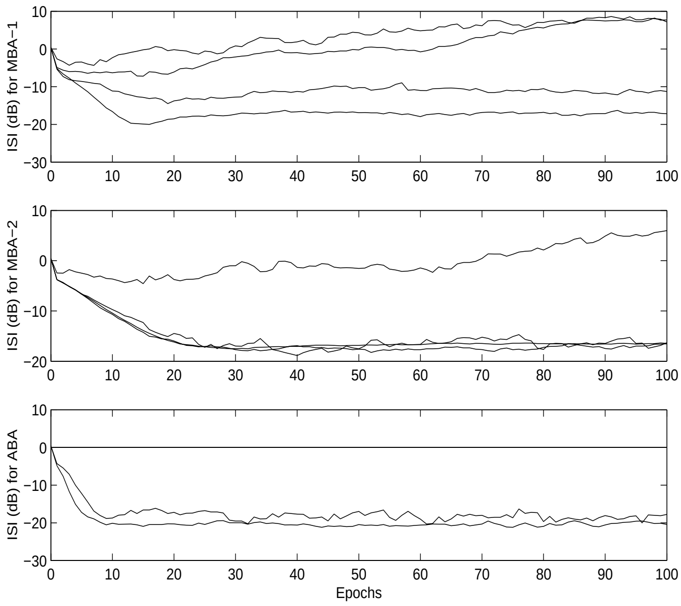

Fig. 8. Comparison of MBA-1, MBA-2 and ABA algorithms when operating over noisy BGR telephonic channel. (MBA-1: The curves, from bottom to top, pertain to the values SNR = 20, 10, 5 and 1 dB respectively. MBA-2: The three curves on bottom, which are almost superimposed, pertain to the values SNR = 20, 10 and 5 dB, while the curve on top pertains to 1 dB. ABA: The lowest curve pertains to SNR = 20, while the higher curve pertains to 10 dB. No convergence was observed for the SNR values 10 and 5 dB.

adaptive adapting step-size cannot be avoided because the algorithm would become unstable (except that for very low fixed step-size values): These differences compensate and make the required computational efforts nearly similar. About the adaptive 'Bussgang' algorithm, it certainly performs very well, however, the numerical procedure required in order to estimate the convolutional noise level make its computational burden much higher than the complexity of the MBAs.

The results of the above comparison are clear: The old MBA-2 algorithm and the new version MBA-1 exhibit nearly the same computational complexity; however, in the considered equalization problem the MBA-1 converges somewhat slower than MBA-2, though their asymptotic performances are equal. They both converge much better than the CMA algorithm.

C. Comparative experiments on noisy channels

An interesting comparison concerns the operation over noisy channels. In particular, the whole AWGN model (1) can be considered, where different signal-to-noise (SNR) ratios may be implemented by adding AWGN with different powers. Because of the hypotheses on the source, in this case SNR$=-10\log_{10}\mathbb{E}_{\mathcal{N}}[\mathcal{N}^2]$ dB.

For the BGR channel, the cases SNR = 20, 10, 5 and 1 dB were considered for the MBA-1, MBA-2 and ABA algorithms. The adapting parameters were: For MBA-2, $\eta = 0.004$ and $\kappa = 0.5$; for MBA-1, $\eta = 0.02$ and $\kappa = 0.5$; for ABA, $\eta = 0.005$ and $\kappa = 1$; common parameters: 20 tap-weights, 1000 samples/epoch for 100 adapting epochs. The results are depicted in the Figure 8. Note that, in the present experiment, ABA did not converge for SNR = 5 and 1 dB.

For the DVD-ROM, a single case was considered that well evidences the behavior of the algorithms in presence of additive noise. The results are reported in the Figure 9 and pertain to the very low signal-to-noise power ratio value of SNR=0 dB,

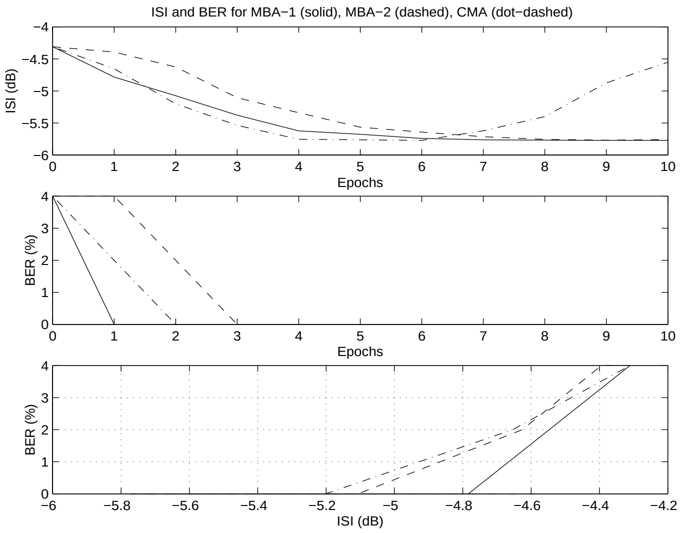

Fig. 9. Comparison of noisy DVD-ROM equalization results with SNR=0 dB.

which evidences some minimal dissimilarities in the behavior of the algorithms. The results illustrated in the Figure ensure that in 1/3 adapting epochs the algorithms are able to cancel completely the effect of the equivalent channel and to drive the bit error to zero (this happens in correspondence of relatively high values of the inter-symbol interference level). In these experiments, the MBAs grant quite good stability in time while CMA looks a little more sensitive to noise.

V. Conclusion

Within the present paper we introduced a theoretical investigation of the adapting equations pertaining to modified 'Bussgang' algorithms through a suitable analysis of their steady-state versions. The older MBA-2 and MBA-1 algorithms have been also tested on a sampled telephonic channel and an equivalent DVD-ROM reading system and on both synthetic and real-world source signals in comparison to adaptive 'Bussgang' and constant modulus algorithms. This analysis allowed us to answer to the three questions presented in the introduction:

1. The performances of MBA-1 depends on the value of the parameter Λ. Is it possible to determine its optimal value ? The answer is positive and the optimal value of the parameter, for the two considered distributions, was computed through the general steady-state analysis introduced in section III.
2. How do the performances of the MBA-1 and the older MBA-2 compare ? From the experimental results illustrated in section IV it can be noticed that the MBA-1 algorithm is a little lighter with respect to MBA-2 about computational complexity and it behaves better in ideal conditions (e.g. on a noiseless channel). MBA-2, however, exhibits a better robustness against additive noise.
3. Can any adaptive version of 'Bussgang' algorithm be envisaged and which are its merits and demerits ? An adaptive version of 'Bussgang' theory was proposed in section II which gives rise to the ABA algorithm. From the experimental results of section IV it appear to be very burdensome about computational complexity and very sensitive to channel disturbances, therefore we deem it does not constitute a valid alternative to the modified 'Bussgang' algorithms.

Also, the numerical results showed that the MBAs converge faster and steadier than the CMA. The present theoretical and numerical investigation concludes the analysis of the class of iterative modified 'Bussgang' algorithms for blind channel equalization based on FIR adaptive filters. A possible line of research in the future may be the replacement of standard stochastic gradient-based optimization algorithms with more efficient optimization algorithms like the fixed-point one. Some preliminary results on this topic are available in the letter [10].

References

[1] E.J. Barbeau, *Polynomials*, New York: Springer-Verlag, 1989

[2] S. Bellini, *Blind Equalization*, Alta Frequenza, Vol. 57, pp. 445 – 450, 1988

[3] A. Benveniste, M. Goursat, and G. Ruget, *Robust Identification of a Non-minimum Phase System: Blind Adjustment of a Linear Equalizer in Data Communication*, IEEE Trans. on Automatic Control, Vol. AC-25, No. 3, pp. 385 – 399, June 1980

[4] T. Bertozzi, D. Le Ruyet, G. Rigal, and H. Vu Thien, *Joint data-channel estimation using the particle filtering on multipath fading channels*, Proceeding of the ICT'03, Tahiti, Feb. 23-28, 2003

[5] C. Cafforio, C. Prati, and F. Rocca, *Full resolution focusing of Seasat SAR images in the frequency-wavenumber domain*, International Journal of Remote Sensing, Vol. 12, pp. 491 – 510, 1991

[6] S. Choi, S. Ong, J. Cho, C. You, and D. Hong, *Performances of Neural Equalizers on Partial Erasure Model*, IEEE Trans. on Magnetics, Vol. 33, No. 5, pp. 2788 – 2790, Sept. 1997

[7] Z. Ding and Y. Li, *Blind Equalization and Identification*, Marcel Dekker, New York, 2001

[8] S. Fiori, *A Contribution to (Neuromorphic) Blind Deconvolution by Flexible Approximated Bayesian Estimation*, Signal Processing, Vol. 81, No. 10, pp. 2131 – 2153, Sept. 2001

[9] S. Fiori, *Blind Deconvolution by Simple Adaptive Activation Function Neuron*, Neurocomputing, Vol. 48, No. 1-4, pp. 763 – 778, Oct. 2002

[10] S. Fiori *A Fast Fixed-Point Neural Blind Deconvolution Algorithm*, IEEE Trans. on Neural Networks. Accepted for publication

[11] M. Ghosh and C.L. Weber, *Blind Deconvolution Using a Maximum Likelihood Channel Estimator*, Proc. of 10^{th} Annual International Phoenix Conference on Computers and Communications, pp. 448 – 452, 1991

[12] D.N. Godard, *Self Recovering Equalization and Carrier Tracking in Two-Dimensional Data Communication System*, IEEE Trans. on Communications, Vol. COM-28, No. 11, pp. 1867 – 1875, Nov. 1980

[13] R. Godfrey and F. Rocca, *Zero Memory Nonlinear Deconvolution*, Geophisical Prospecting, Vol. 29, pp. 189 – 228, Apr. 1981

[14] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, 1991 (Chapter 20: Blind Deconvolution)

[15] J.R. Hopgood and P.J.W. Rayner, *Bayesian single channel blind deconvolution using parametric signal and channel models*, IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, pp. 151 – 154, 1999

[16] G.-K. Lee, S.B. Gelfand, and M.P. Fitz, *Bayesian techniques for blind deconvolution*, IEEE Trans. on Communications, Vol. 44, No. 7, pp. 826 – 835, July 1996

[17] A.B. Marchant, *Optical Engineering: A Technical Overview*, Addison Wesley, 1996

[18] A. Neri, G. Scarano and G. Jacovitti, *Bayesian iterative method for blind deconvolution*, SPIE Proceedings on Adaptive Signal Processing, Vol. 1565 (S. Haykin, ed.), pp. 196 – 208, 1991

[19] Y. Sato, *A Method for Self-Recovering Equalization for Multilevel Amplitude-Modulation Sistem*, IEEE Trans. on Communication, Vol. COM-23, pp. 679 – 682, June 1975

[20] O. Shalvi and E. Weinstein, *New Criteria for Blind Deconvolution of Non-mimimum Phase Systems (Channels)*, IEEE Trans. on Information Theory, Vol. IT-36, No. 2, pp. 312 – 321, March 1990

[21] R.A. Wiggins, *Minimum entropy deconvolution*, Geoexploration, Vol. 16, pp. 21 – 35, 1978

[22] D. Zwillinger (ed.), *CRC Standard Mathematical Tables and Formulae*, 30th ed., CRC Press (Boca Raton, FL), 1996

Simone Fiori was born in Rimini (Italy) in June 1971. He received the Italian Laurea (Dr.Ing.) cum laude in electronic engineering in July 1996 from the University of Ancona (Italy), and the Ph.D. degree in electrical engineering (circuit theory) in March 2000 from the University of Bologna (Italy). In November 2000 he has been appointed assistant professor of circuit theory at the University of Perugia (Italy). His research interests include unsupervised learning theory for artificial neural networks, linear and non-linear adaptive discrete-time filter theory, vision and image processing by neural networks, continuous-time and discrete-time circuits for stochastic information processing. He is author of more than 95 refereed journal and conference papers on these topics and is serving as reviewer for the IEEE Transactions on Neural Networks, the IEEE Transactions on Signal Processing, SIAM Journal on Applied Mathematics and the IEEE Transactions on Antennas and Propagation. Dr. Fiori has been the recipient of the 2001 "E.R. Caianiello Award" for the best Ph.D. thesis in the artificial neural network field.