# Gradient–Based Blind Deconvolution with Flexible Approximated Bayesian Estimator

Simone Fiori, Aurelio Uncini, and Francesco Piazza *
Dept. Electronics and Automatics – University of Ancona (Italy)
E-mail: simone@eealab.unian.it

## Abstract

*In this paper a new blind deconvolution algorithm as modification of the Bellini's 'Bussgang' is presented. Firstly, a novel version based on stochastic Gradient Steepest Descent error minimization technique is proposed. Then the Bayesian estimator used by Bellini is approximated with a flexible 'sigmoid' parameterized with adjustable amplitude and slope, and a gradient-based technique is proposed to adapt such parameters in order to avoid their unsuitable choices. Experimental results are shown to assess the usefulness of the new equalization method.*

## 1. Introduction

Blind deconvolution [3, 8, 11, 12, 14, 16] concerns the problem of recovering a source signal $s(t)$ distorted by a linear channel with impulse response $\vec{h}$, from observations of the channel output $x(t)$, without knowledge about $\vec{h}$ nor the statistics and the source's temporal features. In the linear model:

$$x(t) = \vec{h}^T \vec{s}(t) \ , \tag{1}$$

where $\vec{s}(t)$ is a vector containing the input samples:

$$s(t),\ s(t-1),\ s(t-2),\ \ldots,\ s(t-\ell+1) \ ,$$

with $\ell$ being the number of entries in $\vec{h}$.

A transversal filter described by its impulse response $\vec{w}$ is a channel equalizer if $\vec{w}$ cancels the effects of $\vec{h}$ on the source signal. Said $\vec{x}(t)$ a vector containing the samples:

$$x(t),\ x(t-1),\ x(t-2),\ \ldots,\ x(t-m+1) \ ,$$

where $m$ is the number of tap-weights in $\vec{w}$, the output of the filter is:

$$z(t) = \vec{w}^T(t)\vec{x}(t) \ . \tag{2}$$

*This research was supported by the Italian MURST. Please send comments and suggestions to the first author.

Since $\vec{h}$ and $s(t)$ are unknown, the equalizer $\vec{w}_\star$ such that $z(t) \sim s(t)$ has to be *blindly* found usually by means of an iterative algorithm [1, 2, 9].

When $\vec{h}$ is a non-minimum phase system, its inversion cannot be performed by means of an FIR filter, therefore every time an FIR equalizer is used an approximation error occurs [3, 4, 9]. Formally:

$$z(t) = As(t-\delta) + n(t) \ , \tag{3}$$

where $n(t)$ is the so-called *deconvolution noise*, $A$ is a possible amplitude factor and $\delta$ is a finite delay. A suitable representation of $n(t)$ is a gaussian random process [1, 9] with variance denoted here with $\sigma^2$ (called 'deconvolution noise power'). Notice that the same model for $z(t)$ takes into account the 'learning error' due to the fact that during the whole adaptation phase $\vec{w} \neq \vec{w}_\star$.

Following the pioneering work of Sato [13, 17], several blind equalization algorithms have been proposed through years. One of the most known is that of Bellini [1, 2, 7, 9], based on a memoryless Bayesian estimation $\hat{s} = g(z)$ of $s$ by the knowledge of $z$ and a pseudo-LMS adjustment of $\vec{w}$ with the quantity $e = \hat{s} - z$ as error, with the hypothesis of i.i.d. source sequence. Since $g(z)$ depends on $\sigma^2$, a problem relative to that algorithm is to estimate the deconvolution noise power in the best way.

In this paper we propose a self-tuning procedure that allows to automatically determine optimal parameters of a flexible approximated estimator $g(z)$, in connection with a different error-minimization algorithm based on Gradient Steepest Descent technique. Such a self-tuning behavior allows to overcome the problem of finding a suitable value of $\sigma^2$. Moreover, since parameters are continuously refined through time, suitable values are used during any learning phase. This avoids also the common problem of finding appropriate learning stepsizes.

Finally, we show through simulations that such gradient-based self-tuning algorithm is effective, fast and accurate.

0-7803-4859-1/98 $10.00©1998 IEEE

## 2. Gradient–based deconvolution with flexible estimator

In [1, 2, 10] an error criterion like:

$$\tilde{U} = \frac{1}{2}E[(g(z) - z)^2] \quad (4)$$

is proposed. The function $g(z)$ provides an estimation of the source signal based on the Bayesian technique, so the deconvolving filter $\vec{w}_\star$ minimizes $\tilde{U}$. As a method to find iteratively the optimal filter given observations of the channel output $x$, the pseudo-LMS criterion is used [1, 9]. By interpreting the difference $g(z) - z$ in (4) as an 'error', the structure of the algorithm proposed by Bellini is:

$$\Delta\vec{w} = \mu[g(z) - z]\vec{x} \text{ with } z = \vec{w}^T\vec{x} \ , \quad (5)$$

where $\mu$ is a positive learning stepsize.

As already pointed out in [5, 6], there are no theoretical reasons to use the LMS procedure rather than other ones. So, the minimization of the cost function $U$, that is the instantaneous stochastic approximation of $\tilde{U}$, can be attained also by means of a stochastic Gradient Steepest Descent (GSD) algorithm described by $\Delta\vec{w} \propto -\frac{\partial U}{\partial \vec{w}}$. In the present context this rule assumes the following expression:

$$\Delta\vec{w} = -\eta[g'(z) - 1][g(z) - z]\vec{x} \ , \quad (6)$$

where $\eta$ is a positive learning rate and $g'(z)$ denotes the derivative of the function $g(z)$ with respect to $z$.

By comparing equations (5) and (6) one gathers they coincide if in (5) the variable stepsize:

$$\mu(z) = -\eta[g'(z) - 1]$$

is used. The meaning of a variable stepsize is that of a *self-controlled adaptation rate* which assumes large values at the beginning of learning, and takes more and more smaller values as learning goes on, but that can take again large values when contour conditions change, e.g. when a channel commutation happens [8, 16]. Its usefulness has been clearly explained recently in [5, 6] and experimentally proved.

The Bellini's expression for $g(z)$ is dependent upon the deconvolution noise power $\sigma^2$ [1, 2, 9]. The choice of a suitable estimation for this parameter is quite difficult; moreover, an optimal value for $\sigma^2$ probably does not exist since it should be changed through time accordingly with the adaptation progress. Despite this, for a wide noise power spectrum a suitable approximation of the Bellini's $g(z)$ seems to be [9] the bilateral sigmoid:

$$g(z) = a\tanh(bz) \ , \quad (7)$$

with $a$ and $b$ properly chosen parameters.

If the above approximating expression is used, the gradient of $U$ becomes:

$$\frac{\partial U}{\partial \vec{w}} = [ab - \frac{b}{a}g^2(z) - 1][g(z) - z]\vec{x} \ . \quad (8)$$

In [9] a pair of values for $a$ and $b$ is obtained by fitting the expression (7) with the actual Bellini's function. Anyway, it is clear that as an optimal constant value for $\sigma^2$ cannot be found, a suitable pair of *constant* parameters $a$ and $b$ cannot be fixed, too.

In order to get rid of this drawback, we propose to adapt through time their values by means of a GSD algorithm applied to $U$ (thought as a function of $a$, $b$ and $z$). In formulas we get:

$$\Delta a = -\alpha\frac{\partial U}{\partial a} = -\alpha(g - z)\frac{g}{a} \ , \quad (9)$$

$$\Delta b = -\beta\frac{\partial U}{\partial b} = -\beta(g - z)(a^2 - g^2)\frac{z}{a} \ , \quad (10)$$

where $\alpha$ and $\beta$ are constant positive learning stepsizes.

## 3. Refined criterion

In our algorithm all parameters $a$, $b$ and $\vec{w}$ are changed at the same time by means of equations (6), (9) and (10). Because of the structure of $U$ due to the expression (7), this fact implies that now the problem of minimizing $U$ is ill-posed, because a simple way to minimize $U$ is to vanish $||\vec{w}||$. To prevent such a behavior, it is possible to embed a simple constraint on the norm of $\vec{w}$, that is:

$$\vec{w}^T\vec{w} - \kappa^2 = 0 \ , \quad (11)$$

where $\kappa^2$ is an arbitrarily chosen non-null constant that provides an amplification of the filter output signal with a factor $|\kappa|$. This condition can be taken into account by defining a new criterion $\tilde{J}$ as:

$$\tilde{J} = \tilde{U} + \lambda(\vec{w}^T\vec{w} - \kappa^2) \ , \quad (12)$$

where $\lambda$ is called a *Lagrange multiplier*. Using again stochastic approximated cost functions, namely $U$ instead of $\tilde{U}$ and $J$ instead of $\tilde{J}$, the optimum $\vec{w}$ may be found by:

$$\frac{\partial J}{\partial \vec{w}} = \frac{\partial U}{\partial \vec{w}} + 2\lambda\vec{w} = 0 \ ,$$

with $\vec{w}^T\vec{w} = \kappa^2$, therefore:

$$\vec{w}^T\left(\frac{\partial U}{\partial \vec{w}}\right) + 2\lambda\vec{w}^T\vec{w} = \vec{w}^T\left(\frac{\partial U}{\partial \vec{w}}\right) + 2\lambda\kappa^2 = 0 \ .$$

Hence the corresponding optimum $\lambda$ is:

$$\lambda = -\frac{1}{2\kappa^2}\vec{w}^T\left(\frac{\partial U}{\partial \vec{w}}\right) = -(ab - \frac{b}{a}g^2 - 1)(g - z)\frac{z}{2\kappa^2} \ , \quad (13)$$

where equation (8) has been used.

If we still use the stochastic GSD algorithm $\Delta\vec{w} = -\eta\frac{\partial J}{\partial\vec{w}}$ to search for the minimum of $J$, we can replace the unconstrained rule (6) with:

$$\Delta\vec{w} = -\eta(ab - \frac{b}{a}g^2 - 1)(g - z)(\vec{x} - \frac{z}{\kappa^2}\vec{w}) . \quad (14)$$

Equations (14), (9) and (10) give a new gradient-based blind equalization method with the flexible estimator (7).

## 4. Experimental results

In order to check the validity and performance of the new equalization algorithm, it has been simulated under the following conditions:

- as vector $\vec{h}$ we take the sampled impulse response of a typical non-minimum phase telephonic channel with $\ell = 14$ used in [4]; the bar-graph of $\vec{h}$ is shown in Figure 1;
- as source signal a random process uniformly distributed within $[-\sqrt{3}, \sqrt{3}]$ has been taken, like that described in [1] to develop the Bellini's theory;
- as deconvolving structure a transversal filter with $m = 21$ taps as in [4] is used;
- algorithm starts with $\vec{w}(0)$ which has all entries null except that the $10^{\text{th}}$ one equal to 1, and with $a = b = 1$.

Running the new algorithm needs three 'nominal' learning stepsizes $\eta$, $\alpha$ and $\beta$. We chose three values that experimentally prevent instability and provide fast convergence, namely $\eta = 0.08$, $\alpha = 0.08$ and $\beta = 0.08$. As a filter output amplitude gain we chose $\kappa = 2$.

During the learning phase the behavior of the algorithm has been monitored by computing an error measure $E$ obtained averaging function $U$ over a small batch of 200 samples of the input signal $x(t)$ every epoch. In Figure 2 the error $E$ expressed in dB is shown. Clearly the major part of learning happens within the first 250 epochs, then the refinement continues slowly.

Figure 3 shows the bar graph of the learnt filter $\vec{w}$ after 500 epochs, normalized so that $||\vec{w}||^2 = 1$. It can be directly compared with the 'exact' inverse filter reported in [4].

After 500 epochs, sigmoid's amplitude and slope have values:

$$a = 1.1996 \, , \; b = 0.8399 \, .$$

Any deconvolution problem needs a suitable approximated estimator to behave well.

To appreciate the true equalization performance, in Figure 4 the convolution $\vec{\nu}$ of the channel response $\vec{h}$ and the

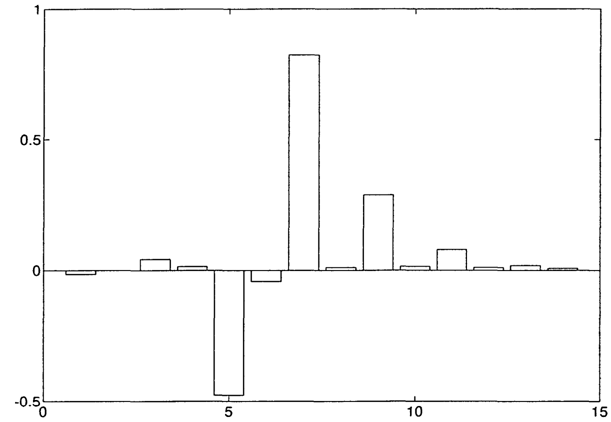


**Figure 1. Sampled telephonic channel response $\vec{h}$.**

learnt deconvolving filter $\vec{w}$ is presented. Ideally a unique central bar there should appear, but the interference residuals have a very low relative intensities, therefore equalization has been attained with a high degree of accuracy. The actual accuracy degree of the deconvolution can be quantitatively measured by means of the residual ISI defined as in [15]:

$$ISI = \frac{\sum_i \nu_i^2 - \nu_{\max}^2}{\nu_{\max}^2} \, , \quad (15)$$

where $\nu_{\max}$ is the component of $\vec{\nu}$ having the maximal absolute value. In Figure 5 the evolution of $ISI$ through learning is depicted. The value of $ISI$ after 500 epochs is $0.0063$.

## 5. Discussion

Several simulations with the new algorithm has been performed keeping fixed the conditions stated at the beginning of Section 4. Indeed, different values of learning stepsizes $\alpha$, $\beta$, $\eta$ and filter output amplitude gain $\kappa$ have been taken, in order to investigate on its stability properties.

Firstly, a criterion without the normalization $\vec{w}^T\vec{w} = const.$ has been considered, thus the adaptation rule with the gradient (8) has been used. As already mentioned above, such an algorithm reaches very low values of $U$ due to the vanishing of $||\vec{w}||$: As the adaptation goes on, the learning terms in (8) become more and more smaller, therefore the correct $\vec{w}$ cannot be achieved in a reasonable number of epochs. Coherently, the obtained algorithm is stable even for high learning stepsizes as, for instance, $10^{-1}$.

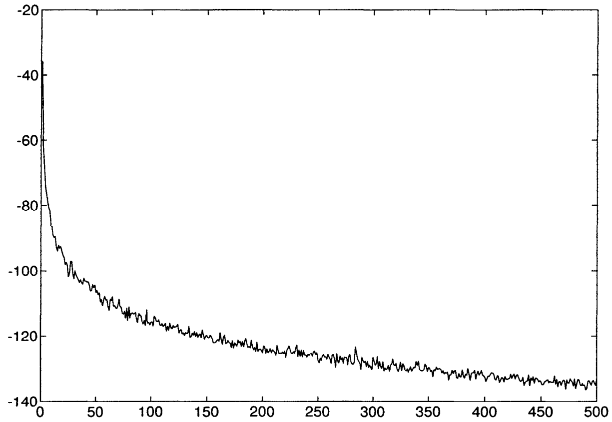


**Figure 2. Averaged criterion $E$ in dB.**

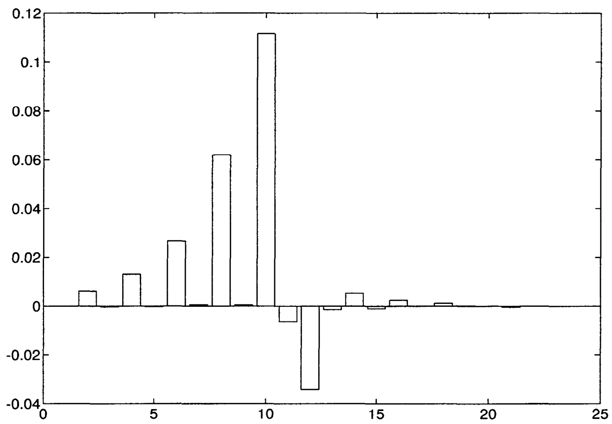


**Figure 3. Learnt filter $\vec{w}_\star$ after 500 epochs.**

To get rid of this drawback the normalization term $\lambda(||\vec{w}||^2 - \kappa^2)$ has been embedded in $U$. Due to its presence a new term appears on (8), namely the quantity $-(z/\kappa^2)\vec{w}$. Through simulations we found such term makes the system unstable for $\kappa = 1$, except when very small stepsizes are used, for instance $10^{-4}$ or smaller. Naturally, this choice makes the algorithm too much slow in converging to the expected equalizer $\vec{w}_\star$. Clearly the simplest solution to this problem is weighting $-z\vec{w}$ with a suitable small number. The choice of $\kappa$ equal to some unit is very good, as can be seen in the previous Section, indeed it allows the system to be stable even when high stepsizes are used, as $10^{-1}$.

About the cost function used in this paper, it is worth noting that it has a very unusual structure. Indeed, expanding

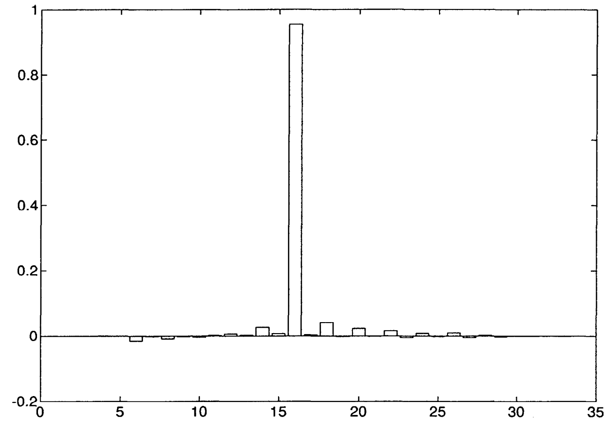


**Figure 4. Convolution between $\vec{h}$ and $\vec{w}_\star$.**

function $g(z)$ up to the third order gives:

$$g(z) \cong abz - \frac{1}{3}ab^3z^3 \,.$$

Since near the equilibrium $ab \cong 1$, $g(z)$ satisfies:

$$g(z) - z \cong -\frac{1}{3}b^2z^3$$

therefore:

$$\tilde{U} \propto b^4 E[z^6] \,, \qquad (16)$$

that means near the equilibrium algorithm tries to minimize the sixth momentum of $z$ under the constraint $\vec{w}^T\vec{w} = \kappa^2$. This behavior provides a little improvement with respect to blind deconvolution algorithms which try to maximize fourth momenta only, like that approximately does (5) with estimator (7), that can be expressed as:

$$\Delta\vec{w} = \mu\frac{\partial}{\partial\vec{w}}\int_0^z [g(\zeta) - \zeta]d\zeta \,. \qquad (17)$$

Another interesting observation is that in [5] an exact Bayesian estimator is established for binary sources, and its structure (see [5, eq. (18)]) looks like (7) where output $z$ appears *squared.* This tells that a function like (7) is not a 'universal' approximating one, and suggests that a better flexible function, as for instance:

$$g(z) = a\tanh(\sum_n b_n z^n) \qquad (18)$$

would be more effective and useful, understanding that coefficients $a$ and $b_n$ should be adaptively changed.

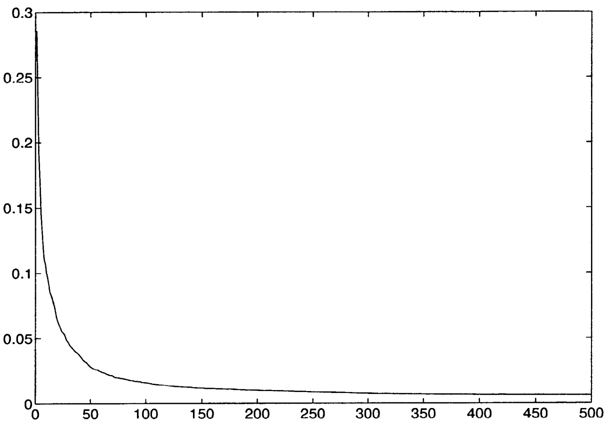


**Figure 5. Inter-Symbol Interference (ISI) residual.**

## 6. Conclusion

In this paper a new blind equalization method based on the Bellini's Bayesian estimation technique is presented. We proposed to overcome the problem of the correct deconvolution noise power assumption, by introducing self-tuned parameters in a suitable approximation of the original estimator. Finally, we showed through simulations that the use of a gradient-based technique instead of the original pseudo-LMS together with self-tuning mechanism give a fast and accurate blind equalization algorithm.

## References


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

[2] S. BELLINI, *Bussgang Techniques for Blind Equalization*, in IEEE Global Telecommunication Conf. Rec., pp. 1634 – 1640, Dec. 1986

[3] A. BENVENISTE AND M. GOURSAT, *Blind Equalizers*, IEEE Trans. on Communications, Vol. COM-32, No. 8, Aug. 1984

[4] A. BENVENISTE, M. GOURSAT AND G. RUGET, *Robust Identification of a Nonminimum Phase System: Blind Adjustment of a Linear Equalizer in Data Communication*, IEEE Trans. on Automatic Control, Vol. AC-25, NO. 3, June 1980

[5] J.B. DESTRO FILHO, G. FAVIER AND J.M. TRAVASSON ROMANO, *New Bussgang Methods for Blind Equalization*, Proc. on International Conference on Acoustics, Speech and Signal Processing, pp. 2269 – 2272, 1997

[6] S. FIORI, *Unsupervised Neural Artificial Models for Blind Signal Processing*, Dr. Eng. (M.Sc.) Dissertation, July 1996 (in Italian)

[7] R. GODFREY AND F. ROCCA, *Zero Memory Nonlinear Deconvolution*, Geophysical Prospecting, vol. 29, pp. 189 – 228, Apr. 1981

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

[9] S. HAYKIN, *Adaptive Filtering*, Ed. Prentice-Hall, 1991 (Chap. 20: Blind Deconvolution)

[10] S. HAYKIN, Editor, *Blind Deconvolutiom*, Prentice-Hall, 1994 (Chap. 2: *Bussgang Techniques for Blind Deconvolution and Equalization*)

[11] O. MACCHI AND E. EWEDA, *Convergence Analysis of Self-Adaptive Equalizers*, IEEE Trans. on Information Theory, Vol. IT-30, No. 2, pp. 161 – 176, March 1984

[12] R. MAKOWSKI, *A Blind Deconvolution Method*, Proc. of Signal Processing V: Theories and Applications, Vol. III, pp. 1959 – 1962, 1990

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

[14] K. TORKKOLA, *Blind Deconvolution, Information Maximization and Recursive Filters*, Proc. of Int. Conf. Acoustics, Speech and Signal Processing, pp. 3301 – 3304, 1997

[15] 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, March 1990

[16] H.H. YANG AND S.-I. AMARI, *Blind Equalization of Switching Channels by ICA and Learning of Learning Rate*, Proc. of International Conference on Acoustics, Speech and Signal Processing, pp. 1849 – 1852, 1997

[17] V. WEERACKODY, S.A. KASSAM, AND K.R. LAKER, *Convergence Analysis of an Algorithm for Blind Equalization*, IEEE Trans. on Communications, Vol. COM-39, No. 6, pp. 856 – 865, June 1991