Cost Function Adaptivity in Bussgang Filtering

Simone Fiori

Abstract

Classical Bussgang algorithms requires hypothesizing deconvolution noise level, while flexible-estimator based algorithms are endowed with learnable parameters that accumulate knowledge on source and channel characteristics. Remarkably, parameters adaptivity induces cost function adaptivity and makes the algorithm 'more blind'. The aim of the present Letter is to investigate on cost function adaptivity phenomenon in Bussgang filtering for blind channel equalization.

Keywords

Adaptive 'Bussgang' blind deconvolution; Bayesian estimation; cost function adaptivity.

I. Introduction

The aim of blind deconvolution/equalization is to recover a source signal distorted by the linear mean that it propagates within [5], [6]. Its most known applications are blind image restoration [10], optical memory-support storage and retrieval enhancement [4], remote sensing [3], [11], [14] and telecommunications [1], [2].

The most known class of blind deconvolution algorithm is perhaps the 'Bussgang' one, which relies on iterative Bayesian estimation of the source sequence [1]. Some modifications to the basic 'Bussgang' algorithm have recently been proposed by the present author in the contributions [6], [7].

Classical Bussgang algorithms requires hypothesizing deconvolution noise level, while flexible-estimator based algorithms are endowed with tunable parameters that are able to accumulate knowledge on source statistics and channel features. Remarkably, parameters adaptivity induces cost function adaptivity. The aim of this Letter is to investigate on cost-function adaptivity arising by the theory of

Faculty of Engineering, Perugia University, Loc. Pentima bassa, 21, I-05100 Terni (Italy). Email: sfr@unipg.it

flexible Bayesian estimation in Bussgang filtering. In particular, a two-tap filtering structure is examined in details.

II. Blind Deconvolution by Iterative Bayesian Estimation

Let us recall the essentials of blind deconvolution from the literature. In vector notation, the linear system to deconvolve is described by the following input/output model:

$$u(t) = \mathbf{h}^T \mathbf{s}(t) \ , \tag{1}$$

where $\mathbf{s}(t) \stackrel{\text{def}}{=} [s(t)\ s(t-1)\ s(t-2)\ \ldots\ s(t-L_h+1)]^T$ is the system's input vector-stream at time t, with L_h denoting the length of the impulse response $\mathbf{h}$. The following minimal hypotheses are usually made about the channel and data stream: The channel's impulse response satisfies $\mathbf{h}^T\mathbf{h} = 1$ and its inverse has finite energy, $\mathbb{E}_s[s(t)] = 0$, $\mathbb{E}_s[s^2(t)] = 1$; 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 it is non-Gaussian.

A filter described by the vector impulse response $\mathbf{w}$ is the inverse of system (1) if $\mathbf{w}$ cancels the effects of $\mathbf{h}$ on the source signal. Denoting by $\mathbf{u}(t) \stackrel{\text{def}}{=} [u(t)\ u(t-1)\ u(t-2)\ \ldots\ u(t-L_w+1)]^T$ the vector containing the filter input samples, where L_w is 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 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 deconvolution it is known that the source signal may be recovered up to arbitrary amplitude scaling and time-delay [9].

When $\mathbf{h}$ represents a non-minimum phase system, its inversion cannot be performed exactly by means of the FIR filter (2); also, during filter adaptation the misadjustment of filter's coefficients make filter output different 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 *deconvolution noise*, $c \in \mathcal{R}$ is an amplitude factor and δ is a finite delay. A suitable representation of $n(t)$ is a zero-mean Gaussian random process [9] with variance σ^2.

From filter output signal model (3), a way can be envisaged to get an estimate of the source sequence $s(t)$, knowing $x(t)$: The model (3) reveals that the relationship between $x(t)$ and $s(t)$ is deterministic but for the convolutional noise, thus Bayesian estimation theory suggests the existence of a causal estimator [1]:

$$\hat{s}(t-\delta) = B(x(t)) \ , \ B(x) \stackrel{\text{def}}{=} \mathbb{E}_s[s|x] = \int_{\mathcal{R}} s p_{s|x}(s|x) ds \ , \tag{4}$$

where $p_{s|x}(s|x)$ is the pdf of s conditioned to the knowledge of x. Provided that the source statistics be known, channel output statistics might be evaluated and the required estimator be computed, as well. As expected, the chosen estimator depends on deconvolution noise power and on the inverse filter response through $p_x(x)$.

On the basis of the available estimator, in [1] it had been proposed an error criterion like:

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

As expected, the function $B(x)$ is dependent upon the deconvolution noise power σ^2 [1], [9]. A suitable estimation for this parameter is made difficult by the fact that an optimal constant value for σ^2 does not exist since actually it changes through time according to adaptation progress.

Despite this, for a wide noise power range a suitable approximation of the actual Bayesian estimator $B(x)$ for uniformly distributed source sequences is [7]:

$$\hat{B}(x) \stackrel{\text{def}}{=} \frac{\Lambda}{c\lambda} \tanh(\lambda x) \ , \tag{6}$$

with Λ being a properly chosen constant; the learnable parameter $\lambda(t)$ adapts by the gradient-descent rule:

$$\Delta\lambda = -\eta(c\hat{B}(x) - x)\left(-\frac{c}{\lambda}\hat{B}(x) + \frac{\Lambda x}{\lambda} - \frac{\lambda x c^2}{\Lambda}\hat{B}^2(x)\right) \ . \tag{7}$$

Also, the gradient-descent adapting rule for $\mathbf{w}$ reads:

$$\Delta\mathbf{w} = -\eta(c\hat{B}(x) - x)\left(\Lambda - 1 - \frac{\lambda^2 c^2}{\Lambda}\hat{B}^2(x)\right)\mathbf{u}\ ,\ \mathbf{w} \leftarrow \frac{\kappa\mathbf{w}}{\|\mathbf{w}\|}\ . \tag{8}$$

The constant η denotes a positive learning step-size and the stochastic gradient approximation is invoked, and the iterative normalization of weight-vector implements automatic gain control (AGC).

The above described algorithm for the channel equalizer will hereafter be referred to as modified Bussgang algorithm with one-learnable-parameter estimator (MBA-1).

About the value of the constant Λ, an extensive theoretical analysis [8] based on minimum-distortion steady-state equalization requirements, led to the value $\Lambda = 1.185$. This is assumed as optimal parameter value in this Letter.

III. Cost Function Adaptivity

The adaptivity of the Bayesian estimator makes it matched to the knowledge on the source statistics and channel characteristics accumulated by the equalizer and represented by the state-variable $\lambda(t)$. Because of the dependence (5) of the cost function on the Bayesian estimator, estimator adaptivity induces cost function adaptivity.

In order to investigate the behavior of the self-shaping cost function $\tilde{C}(\cdot)$, an experiment with a 2-tap inverse filter is illustrated in the following. The experiment deals with a simple IIR channel with impulse response $h[n] = 0.8^n u[n]$. It has a only pole in $p = 0.8$, therefore the exact inverse filter has coefficients $\mathbf{w}_\star = [1\ -0.8]^T$. This experiment is interesting because, under AGC constraint, the inverse filter has only one free parameter, thus it is possible to represent graphically the cost function $\tilde{C}$ at any learning epoch. Within each epoch the enseble-average approximation of expectation operator is computed.

The actual accuracy degree of the deconvolution can be measured by means of the residual inter-symbol interference (ISI) [13] defined as $ISI \stackrel{\text{def}}{=} (\mathbf{T}^T\mathbf{T} - T_{\max}^2)/T_{\max}^2$, where $\mathbf{T}$ denotes the convolution between the channel impulse response and the inverse filter impulse response, and $T_{\max}$ denotes the component of $\mathbf{T}$ having the maximal absolute value.

First, we represent $\mathbf{w} = [w_0 \; w_1]^T$. Let us hypothesize that the AGC keeps $\|\mathbf{w}\| = 1$: As we already know that, at convergence, $w_0 > 1$ and $w_1 < 0$, we may study the cost function at the end of any learning epoch by considering $w_1 \in [-1, 0]$ and $w_0 = \sqrt{1 - w_1^2}$. It is expected that, after any learning epoch, the shape of the cost function changes because it is a function of λ, which varies with time. More formally, by the decomposition $\mathbf{u} = [u_0 \; u_1]^T$, we have:

$$\tilde{C}(w_1) = \frac{1}{2}\mathbb{E}_{\mathbf{u}}\left[\left(\frac{\Lambda}{\lambda}\tanh\left(\lambda\sqrt{1-w_1^2}u_0 + \lambda w_1 u_1\right) - \sqrt{1-w_1^2}u_0 - w_1 u_1\right)^2\right] . \tag{9}$$

Figure 1 shows the behavior of the algorithm. Figure 2 shows instead the shape of the cost function $\tilde{C}(w_1)$ after any learning epoch. These results have been obtained with 100 epochs, 500 samples/epoch, $\eta = 0.003$ and $\kappa = 1$. The curves show that the value of λ changes appreciably during iteration and its adaptation slows down as soon as it reaches the optimal value for the estimator slope.

The results show that the algorithm is able to deconvolve the IIR channel attaining a residual ISI of about -30dB. As a result of investigation, the surface of the cost function evolution shows that the adaptive estimator tends to work better and better by making the minimum of the cost function deeper and sharper in the correct position (around $w_1 = -0.8$.)

In conclusion, Bayesian estimator adaptation has been proven to induce useful cost function tuning in Bussgang adaptive filtering. This has the effect of making the blind cost function suited to the channel equalization task without requiring the user to hypothesize channel deconvolution noise level and thus, ultimately, making the algorithm 'more blind'.

References

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

[2] A. Benveniste and M. Goursat, *Blind Equalizers*, IEEE Trans. on Communications, Vol. COM-32, No. 8, pp. 871 – 883, Aug. 1984

[3] 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

[4] 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

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

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

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

[8] S. Fiori, *Analysis of Bayesian Channel Equalization Algorithms*, Unpublished technical report.

[9] S. Haykin, (Ed.) *Blind Deconvolution*, Ed. Prentice-Hall, 1994

[10] D. Kundur and D. Hatzinakos, *Blind Image Deconvolution*, IEEE Signal Processing Magazine, Vol. 13 , No. 3, pp. 43 – 64, May 1996

[11] F.K. Li, D. Held, J. Curlander, and C. Wu, *Doppler parameter estimation for spaceborne synthetic aperture radar*, IEEE Trans. on Geoscience and Remote Sensing, Vol. GE-23, pp. 47 – 51, 1985

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

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

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

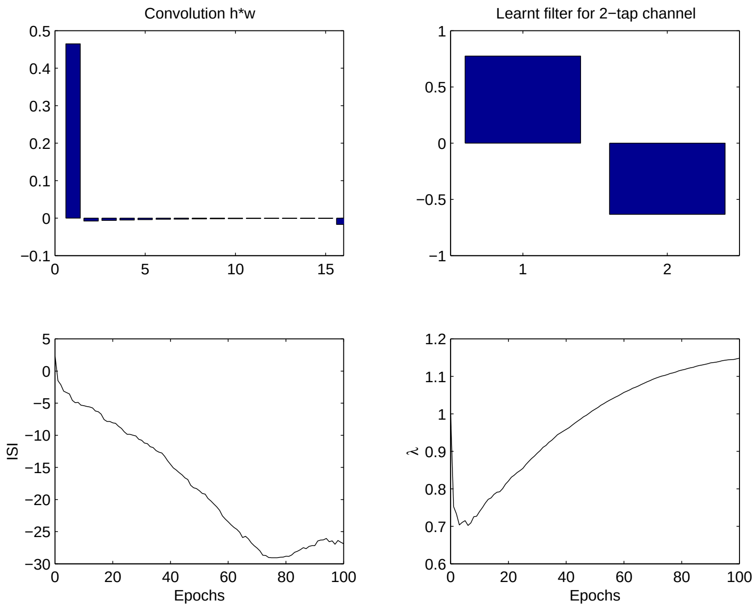

Fig. 1. Two-tap inverse filter experiment: Convolution of **h** and **w** after learning, the learnt inverse filter, the course of ISI and of variable λ during learning.

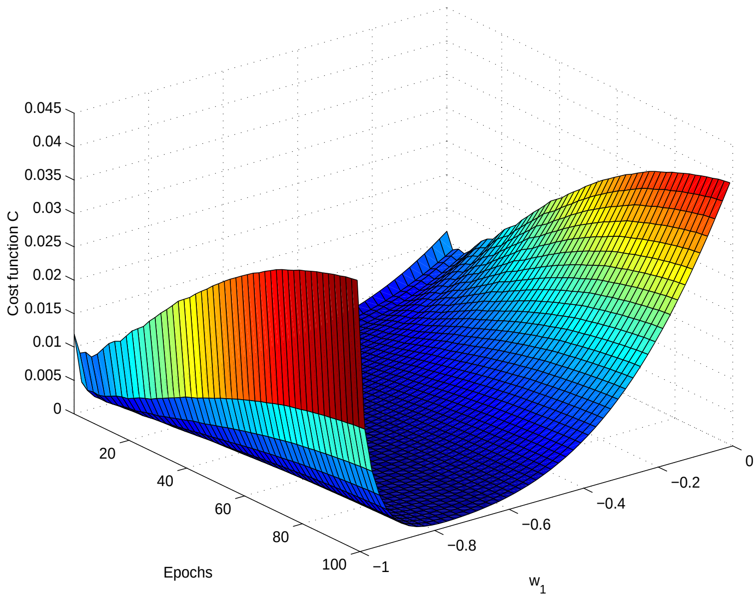

Fig. 2. Shape of the cost function $\tilde{C}(w_1)$ after any learning epoch.