On Self-Consistency of Cost Functions for Blind Signal Processing Based on Neural Bayesian Estimators

Simone Fiori *

Abstract

In some blind signal processing tasks, such as blind source deconvolution and blind source separation, the optimal signal processing structure is designed adaptively through cost function optimization. A class of cost functions known in the literature is based on pseudo-error defined on the basis of Bayesian estimation of the source signals. The exact Bayesian estimators may rarely be computed, so that their neural approximations are often invoked. The present paper aims at investigating the self-consistency of the cost functions based on such neural Bayesian estimators.

1 Introduction

Blind system deconvolution and blind source separation have become popular research fields in the signal processing and neural network communities over the recent years [3, 4]. They concern the recovery of a time-series distorted by a system with memory and the recovery of a set of random signals from their mixtures induced by a system, respectively. The term 'blind' denotes the partial lack of knowledge of the involved signals and of the features of the distorting systems.

Although motivated by different applications, the two research streams possess several common points and, in fact, may be formulated within common frameworks (see e.g. [1, 3]). One of the common frameworks may be the optimization of a cost function based on a pseudo-error defined on the basis of Bayesian estimation of the source signals. In blind deconvolution, this approach was initially proposed by Bellini [2], while in blind separation such interpretation was proposed by Oja [11].

Because of the blindness of the problem, however, the exact Bayesian estimators required in order to compute the cost functions to be optimized

*Facoltà di Ingegneria dell'Università di Perugia – Polo Scientifico e Didattico del Ternano, Loc. Pentima bassa, 21, I-05100 Terni (Italy). Email: sfr@unipg.it

may be rarely accessed, so that their neural approximations are often invoked. That is, approximated Bayesian estimators are to be learnt from the available data as well as the deconvolution/separation structures' parameters.

In blind deconvolution, the idea of replacing the inaccessible exact estimator by a simple neural system was initially proposed by Haykin in [10] and successively exploited, in an adaptive form, by Fiori in [5, 7]. In blind separation, the same idea, which may be referred to as 'source adaptivity', was exploited by many authors (see, for instance, [6, 8, 13]).

In particular, in our previous contributions [5, 7], we suggested to regard the neural approximator as a parametric estimator, whose free parameters may be adapted on the basis of the same optimization principle that drives the deconvolving/separating signal-processing structure. Namely, we suggested to consider the cost function that drives the signal processing structure as a function of the estimator's parameters, too.

Good results have been obtained with the corresponding algorithms designed in this way [5, 7]. However, a detailed investigation of the consistency of the proposed estimation strategy was never carried out before. In particular, some issues such as existence and uniqueness of optimal estimator parameters did not find appropriate investigation in the previous contributions.

The aim of this paper is to present recent formal results on the mentioned theoretical questions.

2 Generality

In Bussgang filtering, the following framework is considered. In the linear model, the input-output transmission system description writes:

$$y(t) = \mathbf{c}^T \mathbf{s}(t) + \mathcal{N}(t) \ , \tag{1}$$

where $\mathbf{s}(t)$ is a vector containing zero-mean time-shifted input samples and $\mathbf{c}$ is the system's impulse response vector. Also, t indicates discrete time, and $\mathcal{N}(t)$ denotes zero-mean additive channel noise that may originate from many simultaneous effects [12], as cross-talk and sampling errors. Since both source signal and channel noise are unobservable, they cannot be distinguished, thus usually the latter is ignored in the theoretical developments [10]; also, usually the need for blind deconvolution arises from severe intersymbol interference (ISI) and not from additive noise, thus the effect of noise on the algorithms and their performances is usually small [2].

A linear equalizer described by its impulse response $\mathbf{w}$ deconvolves $s(t)$ if it cancels the effects produced by the system on the source signal. Denoting by $\mathbf{y}(t)$ the vector containing time-shifted observed samples $y(t)$, the equation describing the inverse discrete-time filter output $x(t)$ reads:

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

In a noiseless situation, perfect equalization would imply $x(t) = cs(t - \Delta)$, where c is a scaling factor and Δ represents the total group-delay of the channel-filter cascade. During filter adaptation, the non-null quantity $x(t) - cs(t - \Delta)$ may be thought of as an infinitely long linear combination of independent source random variables and, in virtue of central limit theorem of statistics, it may be well-represented as a Gaussian random process $n(t)$ termed *deconvolution noise* [2, 10]. Formally:

$$x(t) = cs(t - \Delta) + n(t) \ . \tag{3}$$

The noise $n(t)$ is zero-mean, incorrelated with the source signal and completely characterized by its (time-dependent) variance [10].

From filter output signal model (3) we can imagine a way to get an estimate of the source sequence $s(t)$ from $x(t)$ by means of a Bayesian estimator. In fact, (3) is a deterministic model but for the deconvolution noise, thus the above question gives rise to a classical estimation problem. Motivated from simplicity and robustness, a memoryless estimator was suggested in [2]. Likely, the estimator will depend upon the inverse filter response through the probability density function (pdf) of the source and on the level of convolutional noise. In symbols we write:

$$\hat{s}(t - \Delta) = b(x(t)) \ . \tag{4}$$

The main question is now how to select an appropriate estimator. The answer comes from Bayesian estimation theory (complete details on this are give in [5, 7]). For a wide noise power range a suitable approximation of the actual Bayesian estimator is the bilateral 'sigmoidal' function:

$$\hat{b}(x) = \frac{a}{c}\tanh(\lambda x) \ , \tag{5}$$

with a and λ being properly chosen parameters [10].

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

$$U(\mathbf{w}) = \frac{1}{2}E_x\left[(cb(x) - x)^2\right] \tag{6}$$

was proposed. The function $b(x)$ provides an estimate of the source signal, so the optimal deconvolving filter $\mathbf{w}_\star$ minimizes U because it assumes its lowest values when $x = \mathbf{w}_\star^T\mathbf{y} \approx cs$.

In [10] a pair of values for a and λ is obtained by fitting the expression (5) with the actual estimator for a given convolutional distortion level. Anyway, it is clear that as an optimal constant value for the convolutional noise variance cannot be found, a suitable pair of *constant* parameters a and λ cannot be determined, too.

In order to overcome this problem, we proposed to adapt a and λ through time by means of a gradient steepest descent algorithm applied to U (thought of as a function of a, λ and x). The optimal values of these parameters find by:

$$\begin{cases} \frac{\partial U}{\partial a} &= \mathbb{E}_x\left[(c\hat{b}(x)-x)\frac{c\hat{b}(x)}{a}\right] = 0 \ , \\ \frac{\partial U}{\partial \lambda} &= \mathbb{E}_x\left[(c\hat{b}(x)-x)\{a^2 - c^2\hat{b}^2(x)\}\frac{x}{a}\right] = 0 \ . \end{cases} \tag{7}$$

Straightforward computations lead to the mean-field equilibrium conditions for the learning equations (7):

$$\begin{cases} \mathbb{E}_s[a^2\tanh^2(\lambda x) - ax\tanh(\lambda x)] &= 0 \ , \\ \mathbb{E}_s[(a\tanh(\lambda x) - x)(1-\tanh^2(\lambda x))x] &= 0 \ , \end{cases} \tag{8}$$

with $x = cs$. Also, the residual Bussgang cost function for $x = cs$, denoted as U_R, reads:

$$U_R(a,\lambda) = \frac{a^2}{2}\mathbb{E}_s[\tanh^2(c\lambda s)] + \frac{c^2}{2}\mathbb{E}_s[s^2] - ac\mathbb{E}_s[s\tanh(c\lambda s)] \ . \tag{9}$$

The residual cost function $U_R(a,\lambda)$ represents the minimal possible deconvolution noise power value attainable by the deconvolution algorithm when the neural approximated Bayesian estimator is endowed with the parameters' values a and λ. Consequently, the residual cost function may be advantageously used in order to select the values of these parameters that grant minimal distortion after filter updating.

The shape of the neural Bayesian estimator depends on the probability density function of the source signals, therefore the analyses which were carried out focus on some density function models.

3 Three cases-study

In the following sections, a uniformly-distributed source, a binary source and a source endowed with Laplacean distributions are examined. The common Gaussian distribution was avoided because it is not considered in blind signal processing because it might violate certain identifiability conditions (see e.g. [1]). The uniform and the binary distributions were considered because they are common in telecommunications, while the Laplacean distribution was considered because it (roughly) represents the distribution of values of speech signals. The self-consistency of cost functions for blind signal processing based on the neural Bayesian estimators in the above-mentioned cases has been investigated. As an interesting result, it turns readily out that the analysis of cost-function optimum can be investigated through two macro-variables that parameterize the equations.

3.1 Case I: Uniformly-distributed source signal

Here we consider a uniformly distributed source signal, namely with:

$$p_s(s) = \frac{H(s+s_0) - H(s-s_0)}{2s_0} \ , \tag{10}$$

where $H(\cdot)$ denotes the Heaviside (unit-step) function and $[-s_0, +s_0]$ denotes the support of the distribution ($s_0 > 0$). The computation of the terms involved in the equations (8) is simplified by the following positions:

$$\Gamma \stackrel{\text{def}}{=} c\lambda s_0 \ , \ T_{h,k}(u) \stackrel{\text{def}}{=} \int_0^u z^h \tanh^k(z) dz \ . \tag{11}$$

The first term involved in equations (8) expresses as:

$$\begin{aligned} \mathbb{E}_s[a^2 \tanh^2(\lambda x)] &= \frac{1}{2s_0} \int_{-s_0}^{+s_0} a^2 \tanh^2(\lambda cs) ds \\ \text{(with the variable change } \sigma \stackrel{\text{def}}{=} \lambda cs) &= a^2 \int_0^\Gamma \tanh^2(\sigma) \frac{d\sigma}{\Gamma} \\ &= \frac{a^2}{\Gamma} T_{0,2}(\Gamma) \ . \end{aligned}$$

The other terms may be computed in a similar way and prove to be:

$$\begin{aligned} \mathbb{E}_s[ax \tanh(\lambda x)] &= \frac{acs_0}{\Gamma^2} T_{1,1}(\Gamma) \ , \\ \mathbb{E}_s[ax \tanh^3(\lambda x)] &= \frac{acs_0}{\Gamma^2} T_{1,3}(\Gamma) \ , \\ \mathbb{E}_s[x^2] &= \frac{c^2 s_0^2}{3} \ , \\ \mathbb{E}_s[x^2 \tanh^2(\lambda x)] &= \frac{c^2 s_0^2}{\Gamma^3} T_{2,2}(\Gamma) \ . \end{aligned}$$

Plugging these expressions into the equations (8) yields:

$$\left\{ \begin{array}{lcl} \frac{a^2}{\Gamma} T_{0,2}(\Gamma) - \frac{acs_0}{\Gamma^2} T_{1,1}(\Gamma) & = & 0 \ , \\ \frac{acs_0}{\Gamma^2} T_{1,1}(\Gamma) - \frac{acs_0}{\Gamma^2} T_{1,3}(\Gamma) - \frac{c^2 s_0^2}{3} + \frac{c^2 s_0^2}{\Gamma^3} T_{2,2}(\Gamma) & = & 0 \ . \end{array} \right. \tag{12}$$

Let us multiply both members of both equations by λ^2 and define $\Lambda \stackrel{\text{def}}{=} a\lambda$. In the hypothesis that $\Lambda \neq 0$ and $\Gamma \neq 0$, the above system of equations may be recast into:

$$\left\{ \begin{array}{lcl} \Lambda T_{0,2}(\Gamma) - T_{1,1}(\Gamma) & = & 0 \ , \\ 3\Lambda(T_{1,1}(\Gamma) - T_{1,3}(\Gamma)) - 3T_{2,2}(\Gamma) + \Gamma^3 & = & 0 \ . \end{array} \right. \tag{13}$$

Let us now discuss the relationships among the functions $T_{h,k}(u)$ that help understanding the calculations indicated in the above system of non-linear equations.

The function $T_{1,1}(u)$ appears to play the role of a basis function for the family $T_{h,k}(u)$. In fact, it cannot be integrated in closed form, while some mathematical work show that:

$$T_{0,2}(u) = u - \tanh(u) \ , \tag{14}$$

$$T_{1,3}(u) = T_{1,1}(u) + \frac{\text{sech}^2(u)(2u - \sinh(2u))}{4} \ , \tag{15}$$

$$T_{2,2}(u) = 2T_{1,1}(u) + \frac{u^3}{3} - u^2 \tanh(u) \ , \tag{16}$$

therefore $T_{1,1}(u)$ only needs to be integrated numerically, while $T_{1,3}(u)$ and $T_{2,2}(u)$ may be evaluated in terms of $T_{1,1}(u)$ and of elementary and hyperbolic functions. A graphical representation of the four considered functions $T_{h,k}(u)$ is given in the Figure 1.

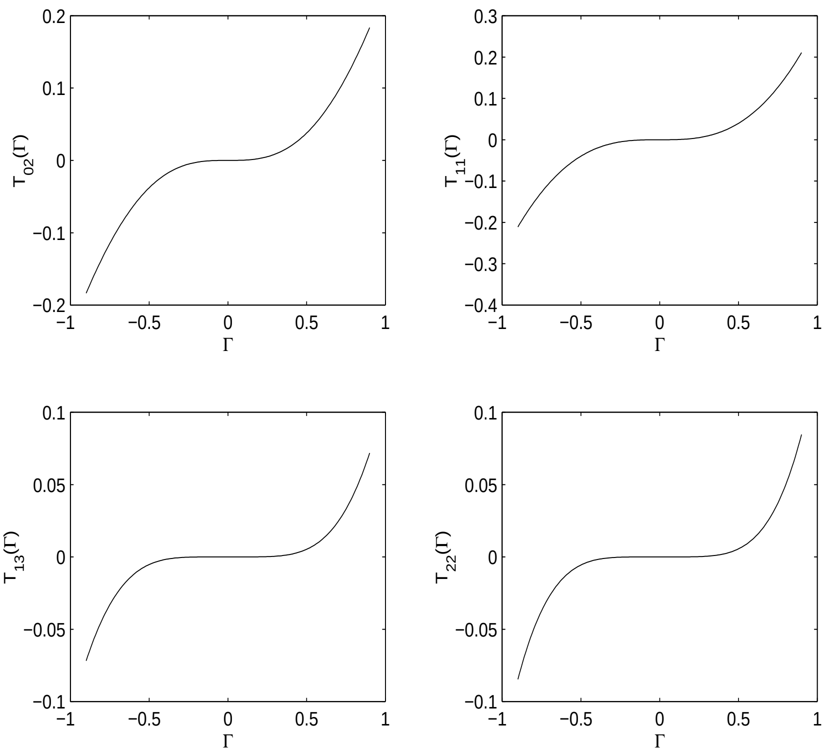

Figure 1: Graphical representation of the four considered functions $T_{h,k}(\cdot)$.

The residual cost function, in this case, has the structure:

$$\lambda^2 U_R = \frac{\Lambda^2}{2\Gamma} T_{0,2}(\Gamma) + \frac{\Gamma^2}{6} - \frac{\Lambda}{\Gamma} T_{1,1}(\Gamma) \ . \tag{17}$$

The (13) represents a system of two non-linear equations in the two unknowns Γ and Λ that may be solved graphically. In particular, what can be done is to make explicit the variable Λ as a function of Γ from the first equation, let us say $\Lambda_1 = \Lambda_1(\Gamma)$ and the same may be done with the second equation, by constructing the solution $\Lambda_2 = \Lambda_2(\Gamma)$. The solutions Λ of the system are the common values $\Lambda_1 = \Lambda_2$ – and the corresponding values of Γ – that may be found numerically.

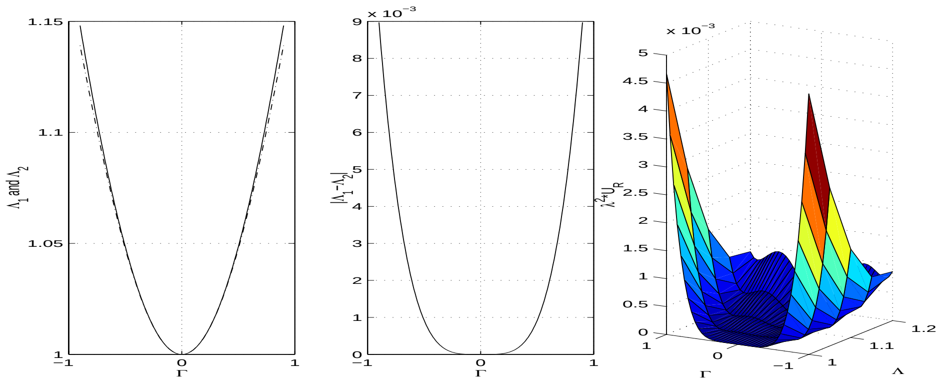

Figure 2: Graphical representation of the functions $\Lambda_1 = \Lambda_1(\Gamma)$ (solid-line) and $\Lambda_2 = \Lambda_2(\Gamma)$ (dashed-line), of their absolute deviation, and of the residual cost as a function of Λ and Γ for the uniformly-distributed source-signal case.

As can be seen from the Figure 2, the only point in which the two curves meet is the origin, that should be excluded, however, the difference-curve $|\Lambda_1 - \Lambda_2|$ is very flat for a large interval of values of Γ, therefore there exist many possible values for the pair Λ and Γ that grant a practically optimal residual cost function, as also testified by the shape of the residual cost function, represented in the Figure 2, that appear to be flat for a relatively large portions of the Λ—Γ plane.

3.2 Case II: Binary source signal

In the present section it is considered a (symmetric) binary source signal, namely with:

$$p_s(s) = \frac{\delta(s+s_0) + \delta(s-s_0)}{2} \ , \tag{18}$$

where the symbol $\delta(\cdot)$ denotes the Dirac's delta and the values $-s_0$ and $+s_0$ denote the only values in the distribution.

The computation of the terms involved in the equations (8) is simplified by the position $\Gamma \stackrel{\text{def}}{=} c\lambda s_0$. The first term involved in equations (8) expresses as:

$$\begin{aligned} \mathbb{E}_s[a^2 \tanh^2(\lambda x)] &= \frac{1}{2}\int_{-\infty}^{\infty} a^2 \tanh^2(\lambda c s)[\delta(s+s_0) + \delta(s-s_0)]ds \\ &= \frac{a^2}{2}[\tanh^2(\lambda c s_0) + \tanh^2(-\lambda c s_0)] \end{aligned}$$

$$= \quad a^2 \tanh^2(\Gamma) \ .$$

The other terms may be computed accordingly and write:

$$\begin{aligned}
\mathbb{E}_s[ax \tanh(\lambda x)] &= acs_0 \tanh(\Gamma) \ , \\
\mathbb{E}_s[ax \tanh^3(\lambda x)] &= acs_0 \tanh^3(\Gamma) \ , \\
\mathbb{E}_s[x^2] &= c^2 s_0^2 \ , \\
\mathbb{E}_s[x^2 \tanh^2(\lambda x)] &= c^2 s_0^2 \tanh^2(\Gamma) \ .
\end{aligned}$$

Plugging these expressions into the equations (8) yields:

$$\left\{ \begin{array}{lcl} \frac{a^2}{\Gamma} T_{0,2}(\Gamma) - \frac{acs_0}{\Gamma^2} T_{1,1}(\Gamma) & = & 0 \ , \\ \frac{acs_0}{\Gamma^2} T_{1,1}(\Gamma) - \frac{acs_0}{\Gamma^2} T_{1,3}(\Gamma) - \frac{c^2 s_0^2}{3} + \frac{c^2 s_0^2}{\Gamma^3} T_{2,2}(\Gamma) & = & 0 \ . \end{array} \right. \tag{19}$$

Let us again multiply both sides of both equations by λ^2 and define $\Lambda \stackrel{\text{def}}{=} a\lambda$. In the hypothesis that $\Lambda \neq 0$ and $\Gamma \neq 0$, the above system of equations may be recast into:

$$\left\{ \begin{array}{lcl} \Lambda^2 \tanh^2(\Gamma) - \Lambda\Gamma \tanh(\Gamma) & = & 0 \ , \\ \Lambda\Gamma \tanh(\Gamma) - \Lambda\Gamma \tanh^3(\Gamma) - \Gamma^2 + \Gamma^2 \tanh^2(\Gamma) & = & 0 \ . \end{array} \right. \tag{20}$$

Straightforward computations show that the above equations are not independent: In fact, they both represent the following, simple, relationship:

$$\Lambda \tanh(\Gamma) = \Gamma \ , \tag{21}$$

The residual cost function, in the present case, is easily proven to be identically equal to zero.

The fact that, for the symmetric binary case, the two non-linear optimal equations reduce to a single relationship means that there exist infinitely many pairs of values of Γ and Λ that satisfy the optimality conditions. Such pairs lie on the curve illustrated in the Fugure 3.

3.3 Case III: Source signal with Laplacean distribution

To end with, we consider a Laplacean distribution, that is useful in modeling human speech signals [9], namely:

$$p_s(s) = \frac{\rho}{2} e^{-\rho|s|} \ , \tag{22}$$

where $\rho > 0$ denotes the Laplacean dispersion parameter. The computation of the terms involved in the equations (8) is simplified by the following positions:

$$\Gamma \stackrel{\text{def}}{=} \frac{c\lambda}{\rho} \ , \ B_{h,k}(u) \stackrel{\text{def}}{=} \int_0^{+\infty} z^h \tanh^k(uz) e^{-z} dz \ . \tag{23}$$

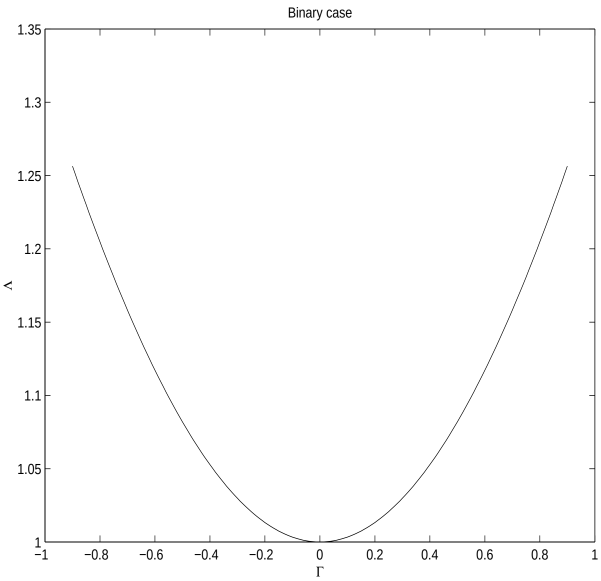

Figure 3: Graphical representation of the function (21).

Thanks to the above positions, the first term involved in equations (8) expresses as:

$$\begin{aligned}
\mathbb{E}_s[a^2\tanh^2(\lambda x)] &= \frac{a^2\rho}{2}\int_{-\infty}^{+\infty}\tanh^2(\lambda c s)e^{-\rho|s|}ds \\
\text{(with the variable change } \sigma \overset{\text{def}}{=} \rho s) &= a^2\int_0^{+\infty}\tanh^2(\Gamma\sigma)e^{-\sigma}d\sigma\ , \\
&= a^2 B_{0,2}(\Gamma)\ .
\end{aligned}$$

The remaining terms may be computed in a similar way and are found to be:

$$\begin{aligned}
\mathbb{E}_s[ax\tanh(\lambda x)] &= \frac{ac}{\rho}B_{1,1}(\Gamma)\ , \\
\mathbb{E}_s[ax\tanh^3(\lambda x)] &= \frac{ac}{\rho}B_{1,3}(\Gamma)\ , \\
\mathbb{E}_s[x^2] &= \frac{2c^2}{\rho^2}\ , \\
\mathbb{E}_s[x^2\tanh^2(\lambda x)] &= \frac{c^2}{\rho^2}B_{2,2}(\Gamma)\ .
\end{aligned}$$

Plugging these expressions into the equations (8) yields:

$$\left\{\begin{array}{lcl}
a^2B_{0,2}(\Gamma) - \frac{ac}{\rho}B_{1,1}(\Gamma) & = & 0\ , \\
\frac{ac}{\rho}B_{1,1}(\Gamma) - \frac{ac}{\rho}B_{1,3}(\Gamma) - \frac{2c^2}{\rho^2} + \frac{2c^2}{\rho^2}B_{2,2}(\Gamma) & = & 0\ .
\end{array}\right. \tag{24}$$

By multiplying every terms of both equations by λ^2, defining $\Lambda \stackrel{\text{def}}{=} a\lambda$ and hypothesizing that $\Gamma \neq 0$ and $\Lambda \neq 0$, the above system of equations may be recast into:

$$\begin{cases} \Lambda B_{0,2}(\Gamma) - \Gamma B_{1,1}(\Gamma) & = & 0 \ , \\ \Lambda(B_{1,1}(\Gamma) - B_{1,3}(\Gamma)) + \Gamma B_{2,2}(\Gamma) - 2\Gamma & = & 0 \ . \end{cases} \tag{25}$$

There appear to be no closed-from expressions for the integral quantities $B_{h,k}(u)$, that need therefore to be evaluated numerically. It is worth noting that the chosen form of the integrand contains the function $z^h e^{-z}$ that decreases to zero rather rapidly for $0 \leq h \leq 2$, and a hyperbolic term that is absolutely bounded by the unity. A graphical representation of the four considered functions $B_{h,k}(u)$ is given in the Figure 4.

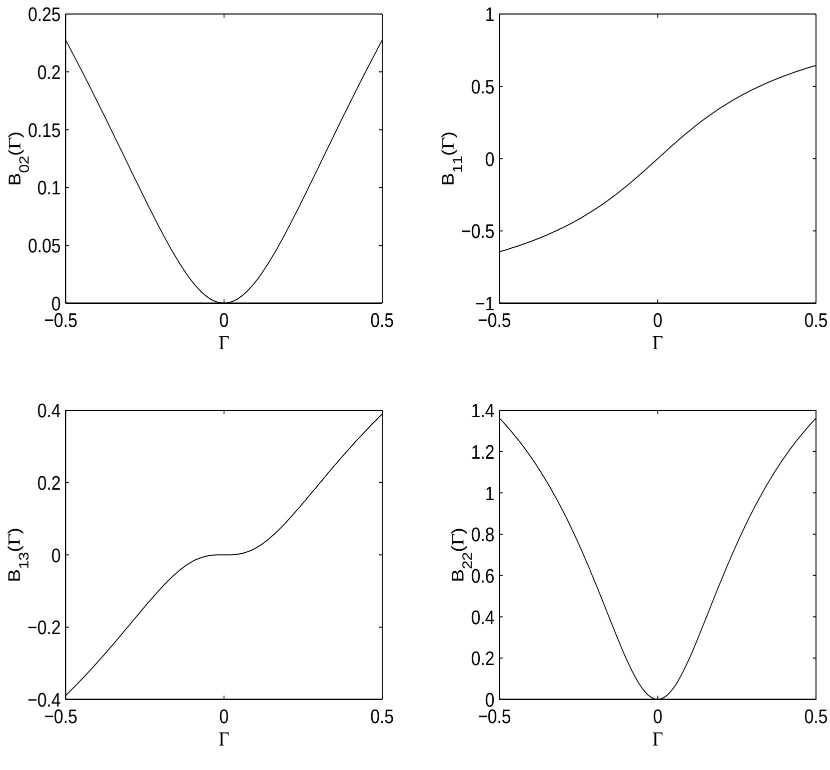

Figure 4: Graphical representation of the four considered functions $B_{h,k}(\cdot)$.

The residual cost function, in this last case, possesses the structure:

$$\lambda^2 U_R = \frac{\Lambda^2}{2} B_{0,2}(\Gamma) + \Gamma^2 - \Lambda\Gamma B_{1,1}(\Gamma) \ . \tag{26}$$

The (25) represents again a system of two non-linear equations in the two unknows Γ and Λ that may be solved numerically. By making again explicit from the first equation the solution $\Lambda_1 = \Lambda_1(\Gamma)$ and from the second equation the solution $\Lambda_2 = \Lambda_2(\Gamma)$, the solutions Λ of the system are the common values $\Lambda_1 = \Lambda_2$ that may be found numerically.

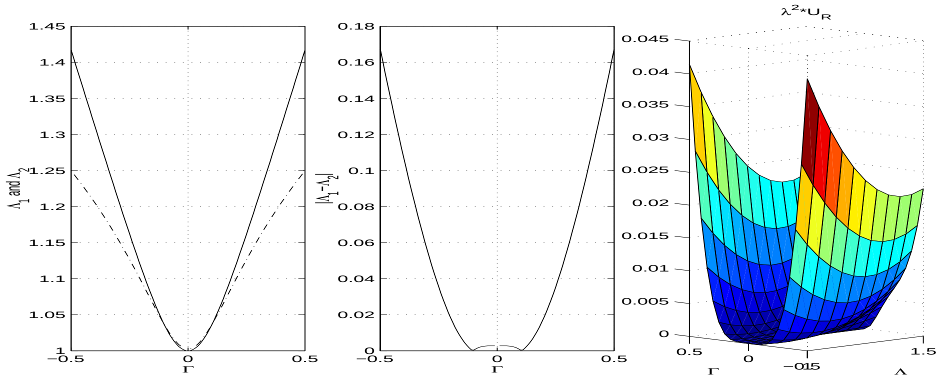

Figure 5: Graphycal representation of the functions $\Lambda_1 = \Lambda_1(\Gamma)$ (solid-line) and $\Lambda_2 = \Lambda_2(\Gamma)$ (dashed-line), of their absolute deviation, and of the residual cost as a function of Λ and Γ for the Laplacean source-signal case.

As can be seen from the Figure 5, there exist two non-null points in which the two curves meet. However, again in practice there exist many possible values for the pair Λ and Γ that grant a practically optimal residual cost function, as testified by the shape of the residual cost function, represented in the Figure 5, that look nearly flat for a relatively large portion of the Λ—Γ plane.

4 Conclusion

The aim of this report was to elucidate the structure of the cost function for the Bussgang technique in blind signal processing, with the aim of optimizing the parameters of the neural approximated Bayesian estimator that the algorithm is equipped with. The analysis is conditioned by the statistical distribution of the source signal values, thus some possible distributions have been considered. The results of the analysis showed that in the binary case it is always possible to find parameters values that theoretically grant null residual interference, while for the uniform distribution and the Laplacean distribution it is, in practice, possible to find parameters values that grant a very low residual interference after filter adaptation.

References

[1] A.J. Bell and T.J. Sejnowski, *An information-maximization approach to blind separation and blind deconvolution*, Neural Computa-

tion, Vol. 7, No. 6, pp. 1129 – 1159, 1996

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

[3] A. Chicocki and S.-i. Amari, *Adaptive Blind Signal and Image Processing*, J. Wiley & Sons, 2002

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

[5] S. Fiori, *A contribution to (neuromorphic) blind deconvolution by flexible approximated Bayesian estimation*, Signal Processing, Vol. 81, No. 10, pp. 2131 – 2153, September 2001

[6] S. Fiori, *Hybrid independent component analysis by adaptive LUT activation function neurons*, Neural Networks, Vol. 15, No. 1, pp. 85 – 94, January 2002

[7] S. Fiori, *Blind deconvolution by simple adaptive activation function neuron*, Neurocomputing, Vol. 48, No. 1-4, pp. 763 – 778, October 2002

[8] S. Fiori, *Notes on Bell-Sejnowski pdf-matching neuron*, Neural Computation, Vol. 14, No. 12, pp. 2847 – 2856, December 2002

[9] S. Fiori, *Neural Independent Component Analysis by 'Maximum-Mismatch' Learning Principle*, Neural Networks. Accepted for publication

[10] S. Haykin, *Adaptive Filtering*, Ed. Prentice-Hall, 1991 (Chapter 20)

[11] E. Oja, *Nonlinear PCA criterion and maximum likelihood in independent component analysis*, Proc. ICA'98, pp. 143 – 148, 1998

[12] 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

[13] N. Vlassis and Y. Motomura, *Efficient source adaptivity in independent component analysis*, IEEE Trans. on Neural Networks, Vol. 12, No. 3, pp. 559 – 566, 2001