

A STATISTICAL REFERENCE CRITERION FOR ADAPTIVE FILTERING

Josep Sala, Gregori Vázquez.

Dpt. of Signal Theory and Communications
Universitat Politècnica de Catalunya
c/ Gran Capità s/n, Mòdul D5
08034 Barcelona, SPAIN.

ABSTRACT †

A cost function for adaptive filtering is proposed, based on the Statistical Reference Criterion. The purpose of this criterion is to enforce a given probability density function at the output of a system. The only aprioristical knowledge utilized is that of the target pdf. This makes it particularly suitable for blind adaptation in the context of arrays in contrast to Temporal or Spatial Reference Techniques. Applications that are foreseen are: blind channel equalization, adaptive beamforming for single or multiple signals and source separation. The main goal is blind signal recovery in unfavourable wanted signal-to-interference power ratios.

1. INTRODUCTION

When trying to recover a signal that has been filtered by an unknown channel or trying to separate independent signals, some kind of statistical matching is usually performed (i.e., enforce constant amplitude, independence between outputs, etc.) [3][4][5]. We intend here not to constrain ourselves to a given feature but instead to perform a complete distribution (pdf) matching [1][2]. This is specially suitable for communications scenarii where we have complete knowledge of the signal statistics, that is, of their constellation and coding schemes. Assuming that interfering signals and noise are present in our scenario, we show that enforcing that the actual distribution at the system output be as close as possible to the target distribution will lead to minimization of output noise power and residual interference. We need only to use an easily implementable functional that measures the distance between two pdf's from a particular realization of the process that will be available to us at the system output. We propose next one such cost function which is closely related to the Kullback-Leibler Distance (KLD) between two pdf's,

$$J(\mathbf{w}) = -E_{A'} \ln E_Z \frac{1}{\pi\sigma_i^2} e^{-|z-a'|^2/\sigma_i^2} \quad (1.1)$$

† This work has been jointly supported by the National Research Plan of SPAIN and the HCM ERBCHRXC930405

where A' and Z denote the RV's corresponding to the target and actual distributions, respectively and $E_{A'}$ and E_Z denote expectation operators with respect to the subscripted RV. $J(\mathbf{w})$ denotes the cost function where \mathbf{w} is the coefficient vector that produces the system output from the data \mathbf{x} as $z = \mathbf{w}^H \mathbf{x}$. This cost function has been proposed for a scalar RV. In a more general formulation where we consider a vector RV, we must enforce the joint distribution by means of,

$$J(\mathbf{W}) = -E_{A'} \ln E_Z \frac{1}{\pi^N \det \mathbf{R}_t} e^{-(z-a)^H \mathbf{R}_t^{-1} (z-a)} \quad (1.2)$$

where now A' and Z denote the vector RV's corresponding to the target and actual joint distributions and \mathbf{R}_t is a noise correlation matrix usually taken to be $\sigma^2 \mathbf{I}$.

Some important points about this criterion are reflected in the following propositions.

2. PROPERTIES

Proposition 1: the cost function can be expressed in the alternative way,

$$J(\mathbf{w}) = -E_{A'} \ln p_{Z+N_t}(a') \quad (2.1)$$

where the argument of the natural logarithm is precisely the pdf of the actual RV Z plus a term of additive independent Gaussian noise, N_t . The variance of this term is precisely the resolution in the estimation of the pdf of Z . The intuitive interpretation of this proposition is straightforward. Let us take the argument of the natural logarithm as it appears in 1.1 It reduces to the expectation of a Gaussian non-linearity centered in the a' value of the random variable A' . This is precisely a measure of how likely it is that the actual random variable Z falls in a neighbourhood of a' , with the 'neighbourhood' defined according to the variance σ_i . That is, the Gaussian non-linearity plays the role of an indicator function. Its exact relationship with the pdf of Z may be construed from the interpretation of this non-linearity as the probability density function of a noise term N_t , independent of Z . Let us note that the pdf of the addition $Z+N_t$ is precisely the convolution of the

respective pdf's, and that the argument of the logarithm is simply this convolution, although it has been expressed through expectations of random variables. In writing, we have that,

$$\begin{aligned} E_Z \frac{1}{\pi\sigma_i^2} e^{-|z-a|^2/\sigma_i^2} &= \int_C p_Z(z) \frac{1}{\pi\sigma_i^2} e^{-|z-a|^2/\sigma_i^2} dz \\ &= \int_C p_Z(z) p_{N_t}(a'-z) dz = p_{Z+N_t}(a') \end{aligned} \quad (2.2)$$

where we have resorted to the symmetry of the distribution of N_t : $p_{N_t}(t) = p_{N_t}(-t)$.

It appears then that we can estimate the pdf of the underlying process Z with any degree of precision, depending only on how small we choose the variance σ_i to be. The limitation is only practical. The smaller this variance is, the larger the number of samples to obtain reliable estimates. In practice, this variance is not chosen very small to guarantee rapid convergence. Simulations show that this is not very critical to obtain a good performance of the algorithm. The relationship enunciated in proposition 1 leads to the following proposition,

Proposition 2: it is easy to prove from the last proposition that for continuous RV, the cost function can be expressed in terms of the KLD and the entropy of the target distribution as,

$$J(\mathbf{w}) = KLD(A', Z + N_t) + H(A') \quad (2.3)$$

such that the only dependence on the coefficient vector is found in the first term, the Kullback-Leibler distance between A' and $Z+N_t$. Therefore, minimization of the cost function is akin to minimization of a KLD. The proof is very simple, we only have to operate as follows,

$$J = -E_{A'} \ln p_{Z+N_t}(a') = E_{A'} \ln \frac{p_{A'}(a')}{p_{Z+N_t}(a')} - E_{A'} \ln p_{A'}(a') \quad (2.4)$$

so that we can establish a one-to-one correspondence between the terms of this expression and those in 2.3. Note that the first term is precisely the KL distance between the actual and the target random variables (except for the noise term N_t resulting from the estimation window). The second term is precisely the entropy of the target random variable A .

Proposition 3: given the following model for the data \mathbf{x}_k ,

$$\mathbf{x}_k = \mathbf{H}\mathbf{a}_k + \mathbf{n}_k, \quad E\mathbf{n}_k\mathbf{n}_k^H = \mathbf{R}_m \quad (2.5)$$

where the signal of interest is one component of the signal vector \mathbf{a}_k , the vector \mathbf{w}_0 that minimizes the cost

function can be expressed in the following way, analogous to the Minimum Variance Criterion,

$$\mathbf{w}_0 = \arg \min_{\mathbf{w}} J(\mathbf{w}) = \mathbf{R}_m^{-1} \mathbf{H}\alpha \quad (2.6)$$

for some vector α .

The value of α reflects just what the contribution from neighbouring symbols is for $\mathbf{w}=\mathbf{w}_0$. As in all blind equalization criteria, there always exists some residual intersymbol interference term, as these kind of criteria do not tend to minimize the power of additive noise. Rather, they minimize the overall power of noise and residual interference. This is also linked to the fact that for that solution, constellation shrinkage occurs. When we consider the single source case, proposition 3 states that the vector that minimizes the cost function is proportional by a scalar term to the Minimum Variance solution. In the multiple source case, it is not necessarily implied that extraneous sources (or neighbouring symbols) are nulled at the system output.

The proof of proposition 3 has not been included in this paper due to its intrinsic complexity. The fact that the coefficient vector is expressed in this way reflects the property that this function has of maximizing SNR at the system output. When the extended cost function for the vector case is used in the context of signal separation, it can also be shown that the minimizing \mathbf{W}_0 is also analogous to the Minimum Variance Criterion in a similar way. That is, we do not only achieve separation of independent signals but also minimization of the noise power and residual interference (in contrast with HOS-based methods which are not sensitive to Gaussian noise power).

3. UPDATE LAW

Up to this point we have considered only the cost function. The coefficient vector \mathbf{w} is updated with the gradient rule. The gradient equations can be easily derived as,

$$\begin{aligned} \nabla_{\mathbf{w}^H} J &= -E_{A'} \frac{E_Z e^{-|z-a|^2/\sigma_i^2} (z-a)^* \mathbf{x}_k}{E_Z e^{-|z-a|^2/\sigma_i^2}} \\ &= -E_{A'} E_Z q(z, a') (z-a)^* \mathbf{x}_k \end{aligned} \quad (3.1)$$

where the q -functions have been defined in the following way,

$$q(z, a') = \frac{e^{-|z-a|^2/\sigma_i^2}}{E_Z e^{-|z-a|^2/\sigma_i^2}} \quad (3.2)$$

Note that these functions depend on the particular distribution of the variable Z . As the statistics of this random variable changes for each iteration, re-evaluation

of these memory non-linearities will be mandatory for each update of \mathbf{w} . It is precisely this dependence on the actual distribution and not on the target distribution, tightly coupled with its memory, what makes this criterion truly sensitive to the distance between pdf's. Usually, in blind algorithm of the Bussgang type, the underlying non-linearity is memoryless and fixed, dependent only on the target distribution. The consequence of this is that for Bussgang-type algorithms, the coefficient vector \mathbf{w} may be captured by some local minimum induced by the presence of interferers. We understand here by local minimum the the target pdf and the actual pdf do not coincide. This happens because the utilized cost function is only sensitive to the maxima of the pdf and not to the frequency of occurrence of its likeliest values. It can be avered that in general convergence of the cost function herein proposed to a desirable minimum is with all likelihood assured, as it has been endorsed by comprehensive set of simulations. We do not purport here that convergence is absolute for all possible input signal distributions. Rather, that the local minimum population is reduced in extreme in comparison to Bussgang-type algorithms, and that convergence is safe for a broad range of situations with unfavourable signal-to-interference power ratios.

Proposition 4: we will prove herein that the solution \mathbf{w} of the cost function is somewhat related to the Wiener criterion,

$$\mathbf{w} = \tilde{\mathbf{R}}_{xx}^{-1} E \hat{a}_k^* \mathbf{x}_k \quad (3.3)$$

Proof: equating the gradient already derived in 3.1 to vector $\mathbf{0}$, we arrive at,

$$E_{A'} E_Z q(z, a') z^* \mathbf{x}_k = E_{A'} E_Z q(z, a') a'^* \mathbf{x}_k \quad (3.4)$$

using the fact that the output z can be expressed as $z = \mathbf{w}^H \mathbf{x}_k$, we have,

$$E_Z (E_{A'} q(z, a')) \mathbf{x}_k \mathbf{x}_k^H \mathbf{w} = E_Z (E_{A'} q(z, a') a')^* \mathbf{x}_k \quad (3.5)$$

naming the term within parantheses on the right-hand side of the equality as \hat{a}_k , we can finally prove the proposition,

$$\tilde{\mathbf{R}}_{xx} \mathbf{w} = E_Z \hat{a}_k^* \mathbf{x}_k \quad (3.6)$$

♦ The non-linear function on the satum z which we have called \hat{a}_k , plays the role of a regeneration function of the data. This function is sensitive to the true actual distribution at the system output. When the output pdf does not coincide with the target pdf, it regenerates a distorted reference. The difference with Bussgang algorithms is that this distortion effects help avoid the equalizer to get captured by false minima.

The update law when applied to array signal processing is somewhat different. We consider in this case that the architecture of the receiver is that of a beamformer plus an equalizer in cascade. Therefore, it is described by two vectors in the following way.

$$y = \mathbf{w}_s^H \mathbf{X} \mathbf{w}_t^* \quad (3.7)$$

This time, the snapshots captured by the array are arranged into matrix \mathbf{X} . The update equations are now derived by taking the joint gradient with respect to the beamformer and the equalizer weights. In some cases (high SIR ratio) convergence may be slow as the quadratic architecture in 3.7 causes that the beamformer may mask the signal from the equalizer if it is not well within focus. This is easily shown by the update equations.

$$\begin{aligned} \mathbf{w}_s(k+1) &= \mathbf{w}_s(k) + \mu_s \varepsilon^*(y_k) \mathbf{X} \mathbf{w}_t^* \\ \mathbf{w}_t(k+1) &= \mathbf{w}_t(k) + \mu_t \varepsilon^*(y_k) \mathbf{X}^T \mathbf{w}_s^* \end{aligned} \quad (3.8)$$

It is clear from the last expression how the algorithm bootstraps itself into convergence: data temporally filter by the equalizer weights are utilized to generate the beamformer updates. Concurrently, data spatially filtered by the present setting of the beamformer weights are used for the equalizer updates.

3. SIMULATIONS

The efficiency of this criterion in recovering a signal of the wanted distribution is proven in the context of array signal processing. For the test scenarii, we assume a given spatial distribution of point sources of varying distributions, where those signals having unwanted statistics are more powerful than the one of interest. Only in this case will we be able to prove the validity of the Statistical Reference Criterion.

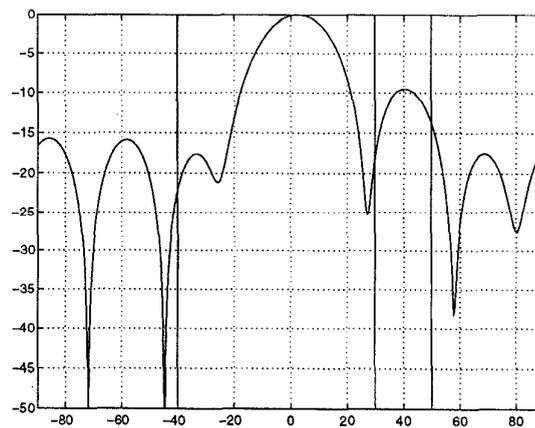


Figure 3.1: Beam pattern obtained in convergence

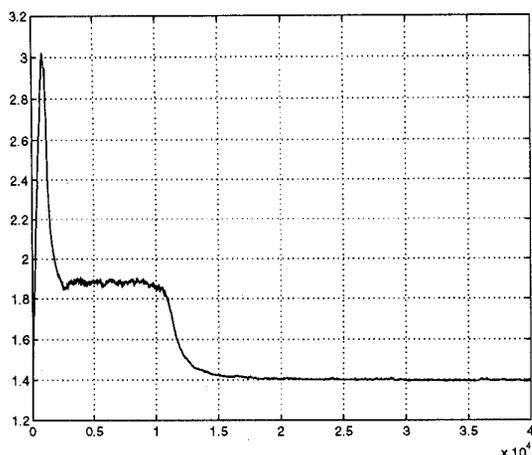


Figure 3.2: time evolution of the cost function

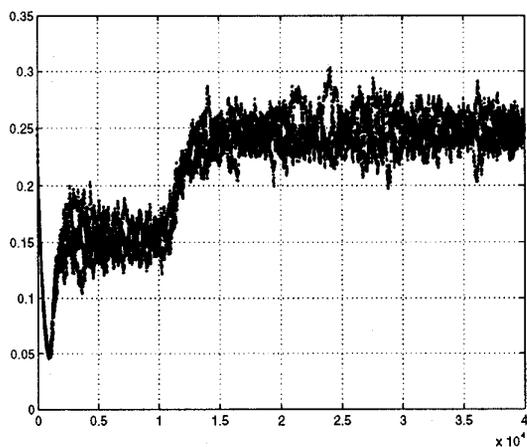


Figure 3.3: time evolution of the estimated probabilities

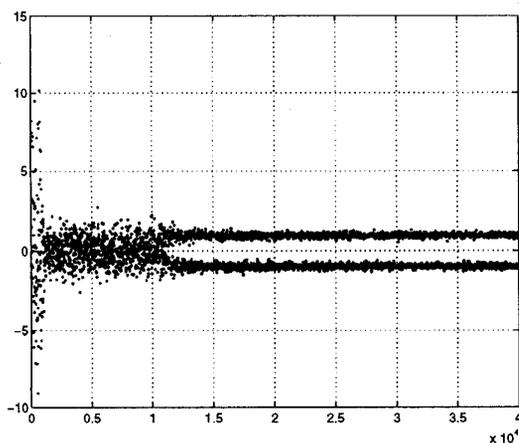


Figure 3.4: In-phase channel. Aperture of the eye versus time.

We have considered a 7-sensor array on which 4 signals impinge: a Gaussian interferer at 50 degrees and 20 dB over the sensor background noise, two 9-QAM interferers at 30 and -40 degrees and with a power level 6.5 dB over that of the signal of interest: a QPSK at 0

degrees. The number of equalizer coefficients is 30. Note how the algorithm is able to distinguish between signals of differing distributions. Even though the interferers are more powerful than the signal of interest, the receiver is able to null out the signals not of interest in the spatial stage. There exist a limit for the robustness of the algorithm to the capture problem. Absence of convergence may be observed if the interferers are in the order of (roughly) 10 dB more powerful than the signal of interest. This could in principle be solved using correspondingly smaller step-sizes. Unfortunately, this increases the convergence time and the relationship of the convergence rate with the step size is not linear.

At the moment of writing this paper, no other algorithm that imposes a given pdf at the output of an adaptive system was known to us. Therefore, comparisons for the same scenarii were made with variations of the constant modulus algorithm. In a way, this algorithm imposes a circular pdf at the system output. Simulations showed that convergence could not be achieved under the conditions in which the cost function herein presented successfully operates. Also, the CMA algorithm is limited to circular distribution and lacks the freedom of choice of the statistical reference philosophy. When applying statistical reference to blind equalization, the aperture of the eye can also be achieved at the expense of longer acquisition time, compared with classical algorithms such as Benveniste-Goursat, Sato, etc. It is only in tough scenarii in terms of interference and noise where our cost function is at its best and other algorithms fail. Work is now being undertaken to extend these results to signal separation, that is, using information of the joint probability density of several signals in the way of statistical reference.

4. REFERENCES

- [1] Sala, Josep. "Information Theoretical Criteria in Adaptive Signal Processing and Applications", PhD. Thesis, Dpt. of Signal Theory and Communications, Universitat Politècnica de Catalunya, June 1995.
- [2] Sala, Josep; Vázquez, Gregori. "A cost function for Blind Signal Recovery based on an Implicit Cumulant Expansion". Proceedings of the IEEE HOS'95 Workshop, Girona, Spain.
- [3] Castedo, Luís; Figueiras-Vidal, Anibal. "Blind equalization and Interference Suppression based on Cyclostationary Signal Properties". Conf. on Adaptive Methods and emergent Techniques for Signal Processing and Communications, Vigo, Spain, 1993.
- [4] Nowlan, Steven J.; Hinton, Geoffrey. "A Soft-Decision Directed LMS Algorithm for Blind Equalization". IEEE Trans. on Comm. VOL-36 No-5, May 1988.
- [5] Haykin, Simon. "Adaptive Filter Theory". Prentice-Hall International Editions. 1991