Statistical Reference Criteria for Adaptive Signal Processing in Digital Communications

Josep Sala-Alvarez, *Member, IEEE,* and Gregori Vázquez-Grau, *Member, IEEE*

Abstract—A general criterion for the design of adaptive systems in digital communications called the statistical reference criterion is proposed. The criterion is based on imposition of the probability density function of the signal of interest at the output of the adaptive system, with its application to the scenario of highly powerful interferers being the main focus of this paper. The knowledge of the pdf of the wanted signal is used as a discriminator between signals so that interferers with differing distributions are rejected by the algorithm. Its performance is studied over a range of scenarios. Equations for gradient-based coefficient updates are derived, and the relationship with other existing algorithms like the minimum variance and the Wiener criterion are examined.

I. INTRODUCTION

ADAPTIVE algorithms are usually applied in the context of signal recovery whose characteristics have been modified in passing through an unknown system. Training sequences may be applied in a MMSE criterion (Wiener solution) to configure the system coefficients. It is intended in blind adaptation that the optimum system configuration be achieved in the absence of a reference. The design of a suitable cost function then becomes the primary goal, which constitutes the main point of this paper. We assume throughout that *a priori* knowledge of the intended signal statistics is available. That is, this cost function will be based on availability of a statistical reference rather than on a temporal (realizational) or spatial (structural) reference. Previous work by the authors may be found in [1]–[3]. The main goal will be therefore to impose a given probability density at the output of the adaptive system. The driving purpose behind this paper is the establishment of a framework for blind signal recovery that is robust in the presence of interference from strong unwanted signals, as is the case, for example, in array signal processing. Therefore, some measure of the "distance" between two density functions should be used in the formulation of the adaptive algorithm: The Kullback–Leibler Distance (KLD) [4] between two density functions provides the starting point from which this criterion can be realized. Other possible measures such as Itakura–Saito might be of interest but are not considered in this paper. The KLD is given by the following expression in terms of the density functions of the processes

$$D(X\|Y) = \int_C p_X(z) \ln \frac{p_X(z)}{p_Y(z)}\, dz. \tag{1}$$

Although we will be able to provide cost functions that implement this criterion, due to their extremely nonlinear nature, it is difficult to prove convergence most of the time. We do not aim to provide this theoretical study here. Rather, convergence has been validated through simulation for a wide range of conditions. It is important to remark, though, that these cost functions are, in general, multimodal. Two cost functions will be presented, which we have called the first and second cost function, respectively. The expression of the two cost functions will be derived, justified, and validated by simulation. The range of applicability of both cost functions is different. The first cost function (FCF) is proven to work in not very complex scenarios (small alphabet modulations and high SINR ratio) and is linked to a soft decision or Bussgang philosophy, and hence, it inherits its limitations. Departing from this prior formulation, a second cost function (SCF) is proposed that truly implements the KLD philosophy. It is important to remark that the FCF will serve the purpose of a stepping stone toward a more elaborate and robust criterion impersonated by the SCF. The SCF instead overcomes the limitations of the FCF and converges under very tough conditions that are evaluated both in complexity of the target pdf and of the interfering signals distribution. The population of local minima is greatly reduced as compared with the FCF, hence, solving the capture problem when interferers stronger than the signal of interest are present. A theoretical study is also carried out to establish the relationship of both cost functions with the Wiener and the minimum variance criteria [6] and [8]. Finally, extensive results are presented for applications that are common in communications signal processing, and conclusions are drawn.

II. STATEMENT: FIRST COST FUNCTION

We propose in this section a cost function based on the statistical reference philosophy we have already introduced. In a real environment, the *a priori* knowledge of the channel behavior and of the noise statistics is practically null. In order to guarantee a minimum robustness of the algorithms, we intend for the adaptive system to be capable of capturing the signal of interest from side information of the source statistics. We should be reminded here that not all algorithms utilize all possible information in its full extent. This is not necessarily

a drawback if it leads to simpler, although not so robust, algorithms. We will show how this is related to soft decision algorithms via the formulation of the first cost function, and how, from this definition, another one is proposed, where usual convergence problems associated with soft decision schemes are solved.

A. Notation

By convention, uppercase letters will be used to refer to a random variable (RV) and lower-case to the values it takes. In addition, it will be necessary in this paper to distinguish between expectations taken over different RV's. Therefore, we will denote with the subscript Z_i that a multivariate nonlinear function is averaged with respect to Z_i

$$E_{Z_i} \Psi(z_1, \cdots, z_L) = \int_C p_{Z_i}(z_i) \Psi(z_1, \cdots, z_L) \, dz_i. \quad (2)$$

We will assume all along that we are dealing with complex random variables. Therefore, the integration support in the last equation is the complex plane C. Expectation over two (or more) RV's Z and Z' is denoted by $E_{Z, Z'}$. We will also at some points consider the expectation with respect to a RV Z that is a function of other RV's $\mathbf{X} = [X_1, X_2, \cdots, X_N]$: $Z = Z(\mathbf{X})$, where, as an example, $\mathbf{Z}$ might be the output of an adaptive filter, and $\mathbf{X}$ is its input at the taps: $z = z(\mathbf{x}) = \mathbf{w}^H \mathbf{x}$. We will understand, therefore, that

$$E_Z \Psi(z) = E_X \Psi[z(\mathbf{x})]. \quad (3)$$

In some forthcoming derivations and for ease of notation, we will sometimes use the terms E_Z and E_X interchangeably, which are equivalent if $z = \mathbf{w}^H \mathbf{x}$ (for fixed $\mathbf{w}$). We have a useful property for deriving some theoretical conclusions: The expectation over a RV $A + B$ of a nonlinearity is examined in Proposition 1.

B. Signal Model

In this section, we will introduce the signal model that will be assumed in the rest of the paper. Two possible cases will be considered: that of a single source and that of multiple sources. The distinction between both models will prove useful in the theoretical analyses that will be carried out further on. A general analysis of the behavior of the cost function is not possible as it greatly depends on the statistics of the input signals. Nonetheless, we will be able to show implicit relationships independent of the statistics of the input signals. Let us define the data vector $\mathbf{x}_k$ in the general multiple source case as

$$\mathbf{x}_k = \mathbf{H} \mathbf{a}_k + \mathbf{n}_k \quad (4)$$

where the subscript k is the time index. $\mathbf{H}$ is the matrix that transforms the information conveyed by each user from vector $\mathbf{a}_k$ to vector $\mathbf{x}_k$, plus a noise term $\mathbf{n}_k$. In the single source case, we have the simpler expression

$$\mathbf{x}_k = a_k \mathbf{h} + \mathbf{n}_k. \quad (5)$$

In an array processing context, $\mathbf{H}$ and $\mathbf{h}$ would be the steering matrix and the steering vector, respectively.

C. FCF Derivation

This cost function has been already utilized in [5] for the equalization of BPSK signals. See also [12] and references thereof for the utilization of the KLD concept. We will show here how it relates to the Kullback–Leibler distance measure. A new formulation will be derived to express it in terms of an error measure for blind algorithms. The derivation starts from the KLD. We will investigate an alternative formulation in terms of the conditioned probability density functions, and we will see how it can also be related to a *temporal reference* algorithm. We will express the cost function by means of a nonlinear averaging of the error norm between the RV at the output of the adaptive system Z (the actual RV) and the RV we would like to have A_o (the *target* RV). Later, in Section II-C1, we will examine a gradient algorithm and see how the concept of a *regeneration* function can be derived from it. This will lead us to the definition of a generalized error as a process that controls the coefficient updates of the adaptive filter. The matching of probability density functions is posed in general terms, independent from the continuous or discrete nature of the density function. For illustrative purposes, in Section II-C1, we provide an algorithm for the recovery of QPSK signals for a discrete pdf, as well as a constant modulus algorithm for a continuous pdf. Let us now examine the expression of the KLD in (1) and pose in the following equation the KLD between the output RV Z and some RV B to be defined later on. We can see that the first term is minus the entropy of the actual RV Z, which at this stage is not possible to evaluate. The FCF $J(\mathbf{w})$ is defined as the second term of the KLD in (6). Information is doubtlessly left out in the process of keeping a single term, but we will show how we can recover from this in the SCF. We prove that the second term is important in the sense that if it is taken alone (expectation of the log-likelihood as in [5]), it leads to soft decision schemes (they do not have information of the differential entropy of the random process). The first term, or the entropy, is also important as it keeps information on the true actual random variable distribution, and thus, it is used implicitly in the SCF. Therefore, we define the FCF $J(\mathbf{w})$ as

$$D(Z\|B) = \int_C p_Z(z) \ln p_Z(z) \, dz - \int_C p_Z(z) \ln p_B(z) \, dz$$
$$= E_Z \ln p_Z(z) - E_Z \ln p_B(z)$$
$$J(\mathbf{w}) \overset{\text{def}}{=} -E_Z \ln p_B(z) \Rightarrow$$
$$D(Z\|B) = -H(Z) + J(\mathbf{w}) \quad (6)$$

where the dependence of the cost function on the coefficient vector $\mathbf{w}$ is implicit in the expectation operator over Z. B is a random variable that we take as the addition of two other RV's: that of the (noiseless) target distribution A_o plus a term of Gaussian noise N_t : $B = A_o + N_t$. The variance of the Gassing noise term, which we call σ_t, models the noise at the output of the adaptive filter and is set to a fixed tentative value. We will refer to it henceforth as the tentative variance parameter. Given that it will not be possible to make the joint power of residual ISI and noise go to zero, we pose instead the

KLD between the actual random variable and a noisy version of the target variable, using the tentative variance for the latter. It has been proven in the simulations that the performance of the adaptive algorithm does not depend critically on the choice of this parameter. For the FCF, it regulates the "softness" of the decision. A formulation of (6) that has been further elaborated upon will provide further insight into the nature of the cost function. A fundamental condition is that the terms that make up B, A_o, and N_t be independent RV's. The pdf of B can then be expressed as follows, where * is the convolution operator:

$$
\begin{aligned}
p_B(z) &= p_{A_o + N_t}(z) \Rightarrow \\
p_B(z) &= p_{A_o}(z) * p_{N_t}(z) \\
&= \int_C p_{A_o}(\alpha) p_{N_t}(z - \alpha) \, d\alpha \\
&= E_{A_o} p_{N_t}(z - a_o).
\end{aligned}
\tag{7}
$$

Resorting to the expression of the pdf of a complex Gassing random variable, we can finally express the FCF in the more insightful way:

$$
J(\mathbf{w}) = -E_Z \ln E_{A_o} \frac{1}{\pi \sigma_t^2} e^{-|z - a_o|^2 / \sigma_t^2}
\tag{8}
$$

where the cost function is expressed as a function of the difference between the two random variables Z and A_o. Equations (6) and (8) as they stand are applicable when a temporal reference is not available. Let us examine how J is modified when we take into consideration that the symbols transmitted by a given user are known (temporal reference). We are constrained in this case to using the output pdf conditioned on the *a priori* known reference d_k or RV D. The modified cost function is now expressed as

$$
J(\mathbf{w}) = -E_Z \ln p_{B|D}(z)
\tag{9}
$$

where $D = \{d_k\}$ constitute the sequence of known values of the random variable A_o. Therefore, the cost function becomes

$$
\begin{aligned}
J(\mathbf{w}) &= -E_Z \ln \frac{1}{\pi \sigma_t^2} e^{-|z_k - d_k|^2 / \sigma_t^2} \\
&= E_Z \frac{1}{\sigma_t^2} |z_k - d_k|^2 + \ln \pi \sigma_t^2
\end{aligned}
\tag{10}
$$

which is coherent as we arrive at the MMSE criterion. Nevertheless, when the temporal reference D is unknown, the cost function can be expressed as in (8). Let us now go on to derive the generalized error function for the FCF to be used in the adaptation of the coefficients.

1) Expression of the Error: All along, we will perform the coefficient updates with the gradient rule. Let us now consider the gradient of the cost function with respect to the Hermitian of the coefficient vector $\mathbf{w}^H$. It is important to remember that we are working in the complex domain and that when taking derivatives, the involved functions are not necessarily analytical (note that although so far we have referred to J as $J(\mathbf{w})$, it is actually $J(\mathbf{w}, \mathbf{w}^*)$). Therefore, derivatives must always be taken with respect to the considered variables and

their conjugates as set out next in (12), but first, let us note that

$$
\begin{aligned}
\nabla_{\mathbf{w}^H} J &= -\nabla_{\mathbf{w}^H} E_Z \ln E_A \frac{1}{\pi \sigma_t^2} e^{-|z - a|^2 / \sigma_t^2} \\
&= -\nabla_{\mathbf{w}^H} E_{\mathbf{X}} \ln E_A \frac{1}{\pi \sigma_t^2} e^{-|\mathbf{w}^H \mathbf{x} - a|^2 / \sigma_t^2} \\
&= -E_{\mathbf{X}} \nabla_{\mathbf{w}^H} \ln E_A \frac{1}{\pi \sigma_t^2} e^{-|\mathbf{w}^H \mathbf{x} - a|^2 / \sigma_t^2} \\
&= -E_Z \nabla_{\mathbf{w}^H} \ln E_A \frac{1}{\pi \sigma_t^2} e^{-|z - a|^2 / \sigma_t^2}
\end{aligned}
\tag{11}
$$

where the expectation operator is now taken with respect to the input RV's and not the output RV's to justify the interchangeability of the expectation and the gradient. This was, in principle, not clear as the expectation operator implicitly contains dependence on vector $\mathbf{w}$. Elaborating on the gradient, we have

$$
\begin{aligned}
\nabla_{\mathbf{w}^H} J &= -E_{\mathbf{X}} \left[\left(\frac{\partial}{\partial z} \Psi \right) \nabla_{\mathbf{w}^H} z + \left(\frac{\partial}{\partial z^*} \Psi \right) \nabla_{\mathbf{w}^H} z^* \right], \\
\Psi &= \ln E_A \frac{1}{\pi \sigma_t^2} e^{-|z - a|^2 / \sigma_t^2}
\end{aligned}
\tag{12}
$$

but the output does not depend on $\mathbf{w}$ but on $\mathbf{w}^*$ for this architecture. Therefore, the gradient is expressed in the shorter form

$$
\nabla_{\mathbf{w}^H} J = -E_{\mathbf{X}} \left(\frac{\partial}{\partial z} \Psi \right) \mathbf{x}_k.
\tag{13}
$$

The derivative of the nonlinearity has an important significance in the interpretation of the algorithm

$$
\begin{aligned}
\frac{\partial}{\partial z} \Psi &= -\frac{1}{\sigma_t^2} \frac{E_A[e^{-|z - a|^2 / \sigma_t^2}(z - a)^*]}{E_{A'} e^{-|z - a'|^2 / \sigma_t^2}} \\
&= -\frac{1}{\sigma_t^2} E_A q(z, a)(z - a)^*
\end{aligned}
\tag{14}
$$

whereby we define the quality functions $q(z, a)$

$$
q(z, a) \stackrel{\text{def}}{=} \frac{e^{-|z - a|^2 / \sigma_t^2}}{E_{A'} e^{-|z - a'|^2 / \sigma_t^2}}.
\tag{15}
$$

Intuitively, these functions quantify how likely it is that the sample z has been produced by the symbol a. Let us now denote the derivative of the cost function as a generalized error. The gradient finally takes the form

$$
\nabla_{\mathbf{w}^H} J = -\frac{1}{\sigma_t^2} E_{\mathbf{X}} E_{A'} q(z, a')(z - a')^* \mathbf{x}_k.
\tag{16}
$$

The q functions have the following property:

$$
E_{A'} q(z, a') = 1.
\tag{17}
$$

Therefore, elaborating on (16), the gradient is straightforwardly expressed as an error as follows:

$$
\begin{aligned}
\nabla_{\mathbf{w}^H} J &= -\frac{1}{\sigma_t^2} E_{\mathbf{X}}[z - E_{A'} a' q(z, a')]^* \mathbf{x}_k \\
&= -\frac{1}{\sigma_t^2} E_{\mathbf{X}} \varepsilon_k^* \mathbf{x}_k
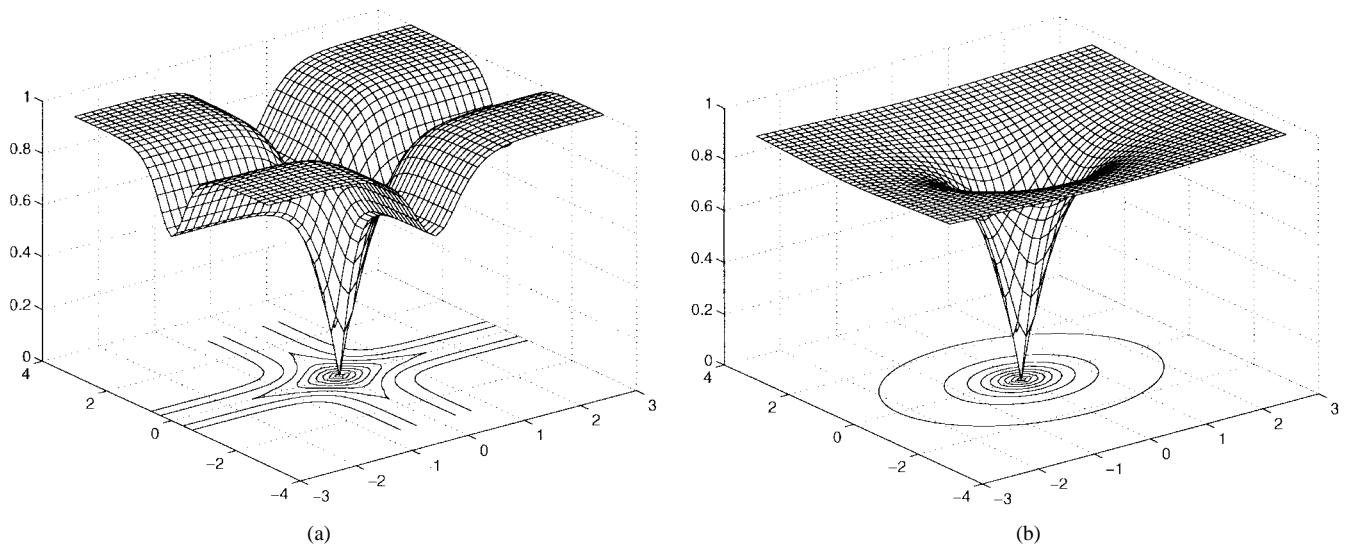\end{aligned}
\tag{18}
$$

Fig. 1. Depiction of regeneration functions (vertical axis) in the complex plane (horizontal plane) for (a) a QPSK signal and (b) constant amplitude signals. Note how in (a) the complex plane is divided into four sectors whose "centers" are the four phases. All values at the frontiers are regenerated as zeros, yielding high values of the generalized error.

where there appears a nonlinear operator on the datum z that plays the role of a regeneration of the data (soft decision) when a reference is not available.

$$\hat{a}_k \overset{\text{def}}{=} E_{A'} a' q(z, a'). \tag{19}$$

We can see that we are led to a Bussgang-type algorithm [6], where the data is regenerated by a memoryless nonlinearity. We will see later on when we introduce the SCF function that the corresponding equivalent regeneration function is indeed a memory nonlinearity (its operation is slightly different, however). We will also see that only memory nonlinearities can retain information on the true pdf at the system output. We now have an example for the case of a discrete distribution: the regeneration function for an M-PSK signal of equiprobable phases, which is immediate

$$p_A(a) = \frac{1}{M} \sum_{i=0}^{M-1} \delta(a - e^{j2\pi i/M}) \Rightarrow$$

$$\hat{a}_k = \sum_{i=0}^{M-1} \frac{e^{-|z-e^{j2\pi i/M}|^2/\sigma_t^2}}{\sum_{i'=0}^{M-1} e^{-|z-e^{j2\pi i'/M}|^2/\sigma_t^2}} e^{j2\pi i/M} \tag{20}$$

and is depicted in Fig. 1(a) for the case of QPSK. The regeneration function for a constant amplitude signal would, in turn, be given by

$$p_A(a) = \frac{1}{2\pi A} \delta(|a| - A) \Rightarrow$$

$$\hat{a}_k = \frac{I_1\left(\frac{2A}{\sigma_t^2}|z|\right)}{I_0\left(\frac{2A}{\sigma_t^2}|z|\right)} A \frac{z}{|z|} \tag{21}$$

with A the amplitude of the wanted distribution and $I_i(x)$ the modified Bessel function of the first kind and order i. See the Appendix and [13] for a derivation. This regeneration function appears in in Fig. 1(b).

III. PROPERTIES

In the last section, we identified the basic concept of a statistical reference algorithm based on the FCF. The objective of this section is to present theoretical results at a more fundamental level. We will center ourselves on different ways to express the coefficient vector that minimizes the criterion. It will be useful to use the following proposition.

Proposition 1: Given two independent random variables A and B, it holds that

$$E_{Z=A+B}\Psi(z) = E_A(\Psi * p_{-B})(a). \tag{22}$$

Proof: Let us expand the left-hand side of Proposition 1:

$$E_{Z=A+B}\Psi(z) = \int_C p_{A+B}(z)\Psi(z)\,dz$$

$$= \int_C \int_C p_A(a)p_B(z-a)\,da\,\Psi(z)\,dz$$

$$= \int_C p_A(a) \int_C \Psi(z)p_B(z-a)\,dz\,da$$

$$= \int_C p_A(a) \int_C \Psi(z)p_{-B}(a-z)\,dz\,da$$

$$= \int_C p_A(a)(\Psi * p_{-B})(a)\,da$$

$$= E_A(\Psi * p_{-B})(a). \tag{23}$$

♦

In this first derivation, we will try to relate the minima of the FCF algorithm to the Wiener and the minimum variance[1] solutions. In an array signal processing context, the former is related to temporal reference and the latter to spatial reference.

[1] For the single source model defined in (5), the Wiener and minimum variance solutions [8] are given by the following equations:

$$\mathbf{w}_{\text{WIENER}} = \mathbf{R}_{xx}^{-1}\mathbf{h},$$

$$\mathbf{w}_{\text{MV}} = \mathbf{R}_{nn}^{-1}\mathbf{h}(\mathbf{h}^H \mathbf{R}_{nn}^{-1}\mathbf{h})^{-1}.$$

For the multiple source case, just substitute $\mathbf{h}$ with $\mathbf{H}$, taking into account that $\mathbf{w}$ is then a matrix.

In particular, we will show how the optimum coefficient vector for the FCF can be expressed in terms of the data correlation matrix. We will proceed first with the single source case.

Proposition 2: All local minima of the cost function are proportional to the minimum variance solution in the single source case

$$\mathbf{w} = \alpha \mathbf{R}_{nn}^{-1} \mathbf{h}. \tag{24}$$

Proof: We will assume for this proposition that the cost function is of the following form:

$$J(\mathbf{w}) = -E_Z \Psi(z) \tag{25}$$

where the log-likelihood nonlinearity is substituted by a more general nonlinearity Ψ. This will simplify notation and provide more general results. Whenever necessary, particularization to the log-likelihood will be resumed. Let us now elaborate on a different formulation of the cost function using Proposition 1, splitting the expectation operator E_Z into its signal and noise terms $E_{A,\mathbf{N}}$

$$
\begin{aligned}
J &= -E_Z \Psi(z) \\
&= -E_{A,\mathbf{N}} \Psi[\mathbf{w}^H(a_k \mathbf{h} + \mathbf{n}_k)] \\
&= -E_A E_{\mathbf{N}} \Psi(a_k \mathbf{w}^H \mathbf{h} + n_{o,k}) \\
&= -E_A (\Psi * p_{-N_o})|_{a_k \mathbf{w}^H \mathbf{h}}
\end{aligned} \tag{26}
$$

where $n_{o,k}$ is the output noise at time k. Note now that dependence on the coefficient vector $\mathbf{w}$ is explicit in the argument of the nonlinearity $\Psi * p_{-N_o}$ and in the pdf of the output noise p_{-N_o}. This noise pdf can, therefore, be expressed as

$$p_{N_o}(n_o) = \frac{1}{\pi \sigma_o^2} e^{-|n_o|^2/\sigma_o^2}\Big|_{\sigma_o^2 = \mathbf{w}^H \mathbf{R}_{nn} \mathbf{w}}. \tag{27}$$

The gradient can now be easily expressed after some algebra as

$$
\begin{aligned}
\nabla_{\mathbf{w}^H} J &= -E_A \nabla_{\mathbf{w}^H} (\Psi * p_{-N_o})|_{a_k \mathbf{w}^H \mathbf{h}} \\
&= -E_A \nabla_{\mathbf{w}^H} \int_C \Psi(z') p_{-N_o}(a_k \mathbf{w}^H \mathbf{h} - z')\, dz' \\
&= -E_A \int_C \Psi(z') \nabla_{\mathbf{w}^H} p_{N_o}(a_k \mathbf{w}^H \mathbf{h} - z')\, dz'
\end{aligned} \tag{28}
$$

where we have taken into account that the distribution of the noise is an even function. When taking the gradient of the pdf of the output noise, we must remember that its variance is also dependent on the coefficient vector. Therefore, the application of the chain rule leads to

$$
\begin{aligned}
\nabla_{\mathbf{w}^H} J = -E_A &\int_C \Psi(z') \\
&\cdot \left\{ \frac{\partial}{\partial n_o} p_{N_o}(n_o)\Big|_{n_o = a_k \mathbf{w}^H \mathbf{h} - z'} \nabla_{\mathbf{w}^H} a_k \mathbf{w}^H \mathbf{h} \right\} dz' \\
-E_A &\int_C \Psi(z') \\
&\cdot \left\{ \frac{\partial}{\partial \sigma_o^2} p_{N_o}(n_o)\Big|_{n_o = a_k \mathbf{w}^H \mathbf{h} - z'} \nabla_{\mathbf{w}^H} \sigma_o^2 \right\} dz'.
\end{aligned} \tag{29}
$$

The particularization of each summand in terms of the coefficient vector $\mathbf{w}$ yields

$$
\begin{aligned}
\nabla_{\mathbf{w}^H} J = -E_A &\int_C \Psi(z') \\
&\cdot \left\{ \frac{\partial}{\partial n_o} p_{N_o}(n_o)\Big|_{n_o = a_k \mathbf{w}^H \mathbf{h} - z'} a_k \mathbf{h} \right\} dz' \\
-E_A &\int_C \Psi(z') \\
&\cdot \left\{ \frac{\partial}{\partial \sigma_o^2} p_{N_o}(n_o)\Big|_{n_o = a_k \mathbf{w}^H \mathbf{h} - z'} \mathbf{R}_{nn} \mathbf{w} \right\} dz'
\end{aligned} \tag{30}
$$

where now, the optimum $\mathbf{w}$ is obtained after equation to the zero vector $\mathbf{0}$ of the last expression

$$
\begin{aligned}
\mathbf{w} = -&\frac{E_A \int_C \Psi(z') \left\{ \frac{\partial}{\partial n_o} p_{N_o}(n_o)\Big|_{n_o = a_k \mathbf{w}^H \mathbf{h} - z'} a_k \right\} dz'}{E_A \int_C \Psi(z') \left\{ \frac{\partial}{\partial \sigma_o^2} p_{N_o}(n_o)\Big|_{n_o = a_k \mathbf{w}^H \mathbf{h} - z'} \right\} dz'} \\
&\cdot \mathbf{R}_{nn}^{-1} \mathbf{h} \\
= &\,\alpha \mathbf{R}_{nn}^{-1} \mathbf{h}.
\end{aligned} \tag{31}
$$

Finally, we prove that the optimum coefficient vector is proportional to the minimum variance solution. It is interesting to remark that the scaling term of this solution α does coincide with that of the minimum variance solution $1/\mathbf{h}^H \mathbf{R}_{nn}^{-1} \mathbf{h}$ and is dependent on the signal-to-noise ratio. ◆

Note that the last derivation has been carried out for a general nonlinearity Ψ. It appears that the minimizing coefficient vector will always be proportional to the minimum variance solution, regardless of the shape of Ψ. Nonetheless, this nonlinearity must be properly chosen to guarantee a good behavior of the adaptive algorithm.

Proposition 3: All local minima of the FCF in the multiple source case are generated by the following equation, which is analogous to Proposition 2,

$$\mathbf{w} = \mathbf{R}_{nn}^{-1} \mathbf{H} \mathbf{b} \tag{32}$$

for some vector $\mathbf{b}$.

Proof: We proceed in the same way as in Proposition 2, where now, the decomposition of the cost function is as follows:

$$
\begin{aligned}
J &= -E_{\mathbf{A},\mathbf{N}} \Psi(z) \\
&= -E_{\mathbf{A}} E_{\mathbf{N}} \Psi(\mathbf{w}^H \mathbf{H} \mathbf{a}_k + n_{o,k}) \\
&= -E_{\mathbf{A}} (\Psi * p_{-N_o})|_{\mathbf{w}^H \mathbf{H} \mathbf{a}_k}
\end{aligned} \tag{33}
$$

and finally, we obtain

$$
\begin{aligned}
\mathbf{w} = -&\mathbf{R}_{nn}^{-1} \mathbf{H} \\
&\cdot \frac{E_{\mathbf{A}} \int_C \Psi(z') \left\{ \frac{\partial}{\partial n_o} p_{N_o}(n_o)\Big|_{n_o = \mathbf{w}^H \mathbf{H} \mathbf{a}_k - z'} \mathbf{a}_k \right\} dz'}{E_{\mathbf{A}} \int_C \Psi(z') \left\{ \frac{\partial}{\partial \sigma_o^2} p_{N_o}(n_o)\Big|_{n_o = \mathbf{w}^H \mathbf{H} \mathbf{a}_k - z'} \right\} dz'} \\
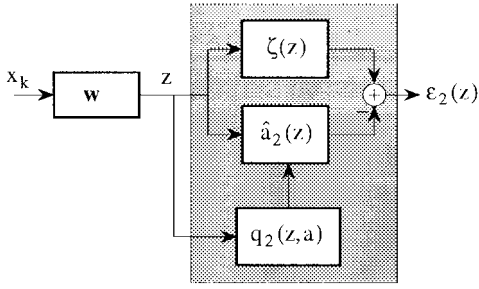= &\,\mathbf{R}_{nn}^{-1} \mathbf{H} \mathbf{b}
\end{aligned} \tag{34}
$$

◆

Fig. 2. Evaluation of the generalized error for the SCF. The actual probabilities of z are estimated in the q_2 block and used to derive the SCF quality function and its generalized error. Correlation with the data vector $\mathbf{x}$ is then used to update the coefficients in $\mathbf{w}$.

We can see that the coefficient vector in the multiple source case is such that all present sources contribute to the filter output. It does not happen that a given source is exactly filtered out. This effect is known in temporal reference algorithms as power inversion. The Wiener criterion is such that it minimizes the joint power of interference and noise. Self interference also appears, as the scaling of the signal of interest is smaller than that of the reference. The error then contains residual signal terms.

Proposition 4: The optimum coefficient vector can be expressed in terms of the regeneration function in the following way, which is analogous to the Wiener solution in a temporal reference setting:

$$\mathbf{w} = \mathbf{R}_{xx}^{-1} E_{\mathbf{X}} \hat{a}_k^* \mathbf{x}_k. \tag{35}$$

Proof: Let us reproduce again the expression for the gradient in (15). We have at the minimizing $\mathbf{w}$ that

$$E_{\mathbf{X}} E_{A'} q(z, a')(z - a')^* \mathbf{x}_k = \mathbf{0}. \tag{36}$$

Let us now split this equality, and let us be reminded that $z = \mathbf{w}^H \mathbf{x}_k$

$$E_{\mathbf{X}} E_{A'} q(z, a') \mathbf{x}_k z^* = E_{\mathbf{X}} E_{A'} a'^* q(z, a') \mathbf{x}_k \Rightarrow$$
$$E_{\mathbf{X}} E_{A'} q(z, a') \mathbf{x}_k \mathbf{x}_k^H \mathbf{w} = E_{\mathbf{X}} \hat{a}^* \mathbf{x}_k. \tag{37}$$

Finally, we have that

$$E_{A'} q(z, a') = 1 \Rightarrow$$
$$\mathbf{w} = \mathbf{R}_{xx}^{-1} E_{\mathbf{X}} \hat{a}^* \mathbf{x}_k \tag{38}$$

where $\mathbf{R}_{xx}$ is the covariance matrix of vector $\mathbf{x}_k$. ♦

The interpretation of this last result is very insightful as it establishes an analogy with the Wiener criterion. Now, $\mathbf{w}$ is expressed as the cross-correlation of the regenerated reference $\hat{a}$ with the data vector $\mathbf{x}_k$, premultiplied by the inverse covariance matrix of the data. We want to impose the regeneration function that contains the information on the target distribution at the output of the adaptive system.

Let us now give a qualitative explanation of how the first cost function works. It tries to make the peaks of the actual pdf coincide with the troughs of the nonlinearity (the log-likelihood we want) to achieve minimization. Therefore, although Proposition 2 has been proven for a general non-linearity, if the troughs of the target pdf do not coincide with those of Ψ, excessive noise may be introduced by

the algorithm. That is, the regeneration function does not work properly. Some problems that arise in the presence of interference are, for example, that the peaks of a BPSK interferer may fall on those of the target distribution of a QPSK signal. Then, the cost function is fooled and captures the interferer. Therefore, this cost function should only be used in scenarios that are not very hostile in terms of signal-to-interference ratios and for simple target distributions. Some results will be presented later in the simulations. The trouble is that for the moment, as the cost function has been formulated, no information on how frequent each peak of the actual pdf is can be obtained from the data. This is a problem associated with all soft-decision type algorithms and, in this case, is caused by dropping the entropy term in (6). True sensitivity of the actual pdf can be obtained in the formulation of the SCF.

IV. STATEMENT: SECOND COST FUNCTION

We have seen the drawbacks of the first cost function when it is used for elaborate distributions of the intended signal and in the presence of interference. This motivated the study of a SCF, which truly preserves the essence of the KLD as a foil to the FCF. Its complexity is also higher than for the FCF, but its multimodal character is much improved, which has been endorsed by direct experience.

We carry out only a minor modification consisting of an interchange of expectations. This is not a shot in the dark but an educated guess instilled by some prior intuition we had of the structure a new cost function that is pdf-sensitive should have. We will justify this a little further on when we speak of the estimation of pdf's. We set out next the two expressions for the FCF and the SCF:

$$J(\mathbf{w}) = -E_Z \ln E_{A_o} \frac{1}{\pi \sigma_t^2} e^{-|z - a_o|^2 / \sigma_t^2} \Leftrightarrow$$
$$J_2(\mathbf{w}) = -E_{A_o} \ln E_Z \frac{1}{\pi \sigma_t^2} e^{-|z - a_o|^2 / \sigma_t^2} \tag{39}$$

where in the SCF, the outer expectation operator is on the target RV, and the inner expectation operator is on the actual RV. We can see in this formulation that the argument of the natural logarithm can be considered to be an estimation of the pdf of the actual RV Z, where the Gassing window plays the role of an indicator function (a measure of the average closeness between z and a_o, whose expectation can be used to estimate the probability of occurrence of z in a neighborhood of a_o). We can introduce, therefore, the pdf estimate $\hat{p}$ of Z in the following way (we will also see its relationship with the actual pdf of Z) in Proposition 5:

$$\hat{p}_Z(a_o) = E_Z \frac{1}{\pi \sigma_t^2} e^{-|z - a_o|^2 / \sigma_t^2}. \tag{40}$$

Proposition 5: The pdf estimate of Z is related to the true pdf of Z by

$$\hat{p}_Z(z) = p_{Z+N_t}(z),$$
$$p_{N_t}(n_t) = \frac{1}{\pi \sigma_t^2} e^{-|n_t|^2 / \sigma_t^2} \tag{41}$$

where N_t is a zero-mean Gassing RV of variance σ_t that is independent of the RV Z.

amplitude function

regeneration function
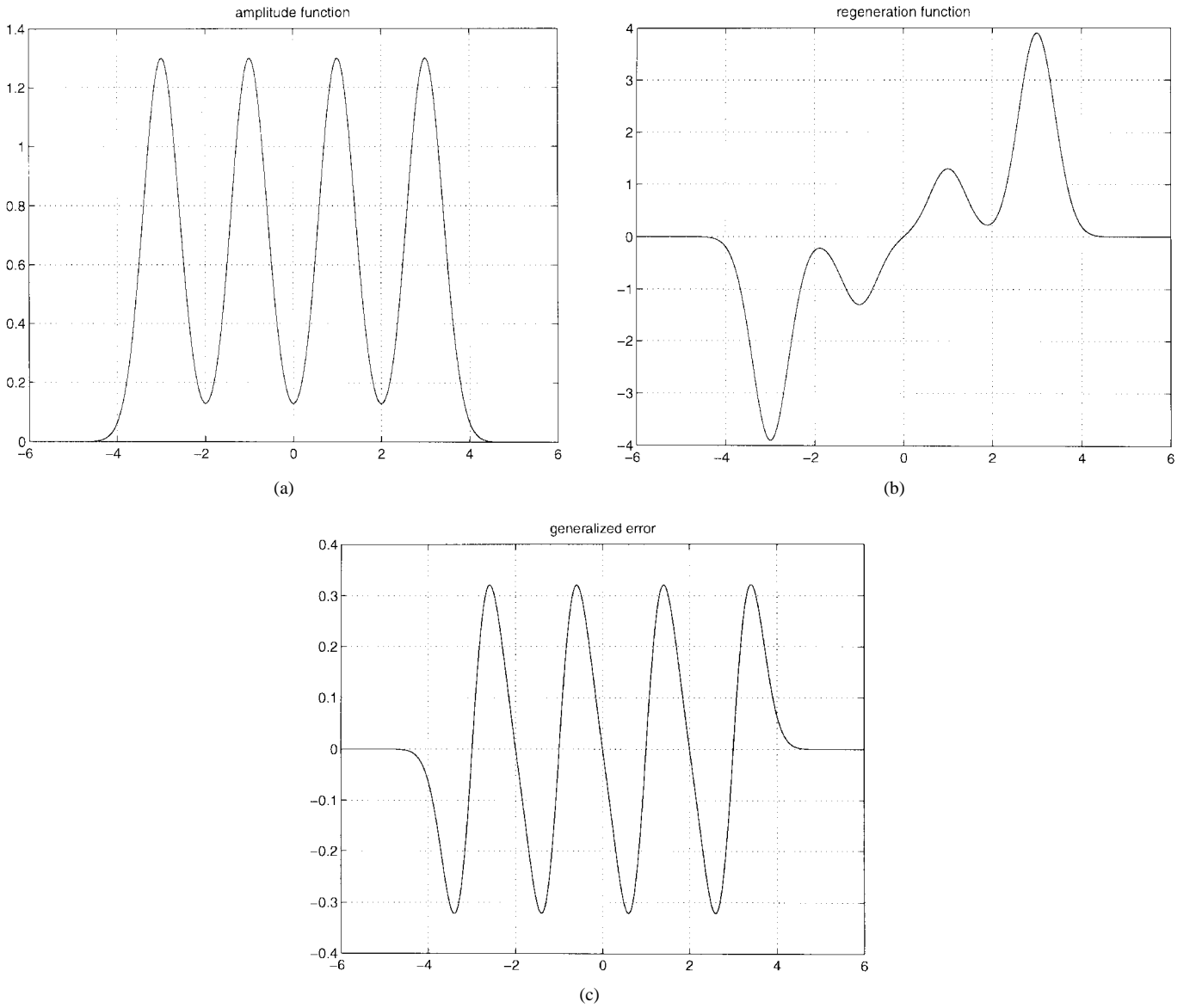
(a)

(b)

generalized error

(c)

Fig. 3. Target and actual pdf's are 4-ASK and 4-ASK, respectively, with alphabet $= [-3, -1, 1, 3]$. (a) Amplitude function. (b) Regeneration function. (c) Generalized error nonlinearities.

Proof: To prove this assertion, we resort to the following alternative relationship for the pdf of the addition of two independent RV's A and B:

$$
\begin{aligned}
p_{A+B}(z) &= \int_C p_A(\tau)p_B(z-\tau)\, d\tau \\
&= E_A p_B(z-a) \\
&= E_B p_A(z-b).
\end{aligned} \tag{42}
$$

Hence, notice in (40) that the Gaussian nonlinearity working as an indicator function is precisely the pdf of a random variable whose variance controls the aperture of this nonlinearity. Hence, Proposition 5 is proven.   ♦

In this sense, we can estimate a pdf with some uncertainty related to the aperture $(1/\sigma_t^2)$ of the indicator function that we are using. In the end, what we have is a blurred version of the original pdf.

The following property provides a mathematical justification of the validity of the second cost function for implementing the

statistical reference criterion. Its relationship with the KLD is established.

Proposition 6: The SCF can be expressed in terms of the KLD between the pdf's of A and $Z + N_t$ as

$$
J_2 = D[A\|(Z+N_t)] + H(A) \tag{43}
$$

with $H(A)$ the differential entropy of the RV A and $D(X\|Y)$ the KLD between random variables X and Y.

Proof: The proof is very simple. We only have to operate as follows:

$$
\begin{aligned}
J_2 &= -E_A \ln p_{Z+N_t}(a) \\
&= E_A \ln \frac{p_A(a)}{p_{Z+N_t}(a)} - E_A \ln p_A(a) \tag{44}
\end{aligned}
$$

which completes the proof as the second summand is precisely the entropy $H(A)$.   ♦

We will also prove some important result relating to the minimum variance solution as in Proposition 3. We will see
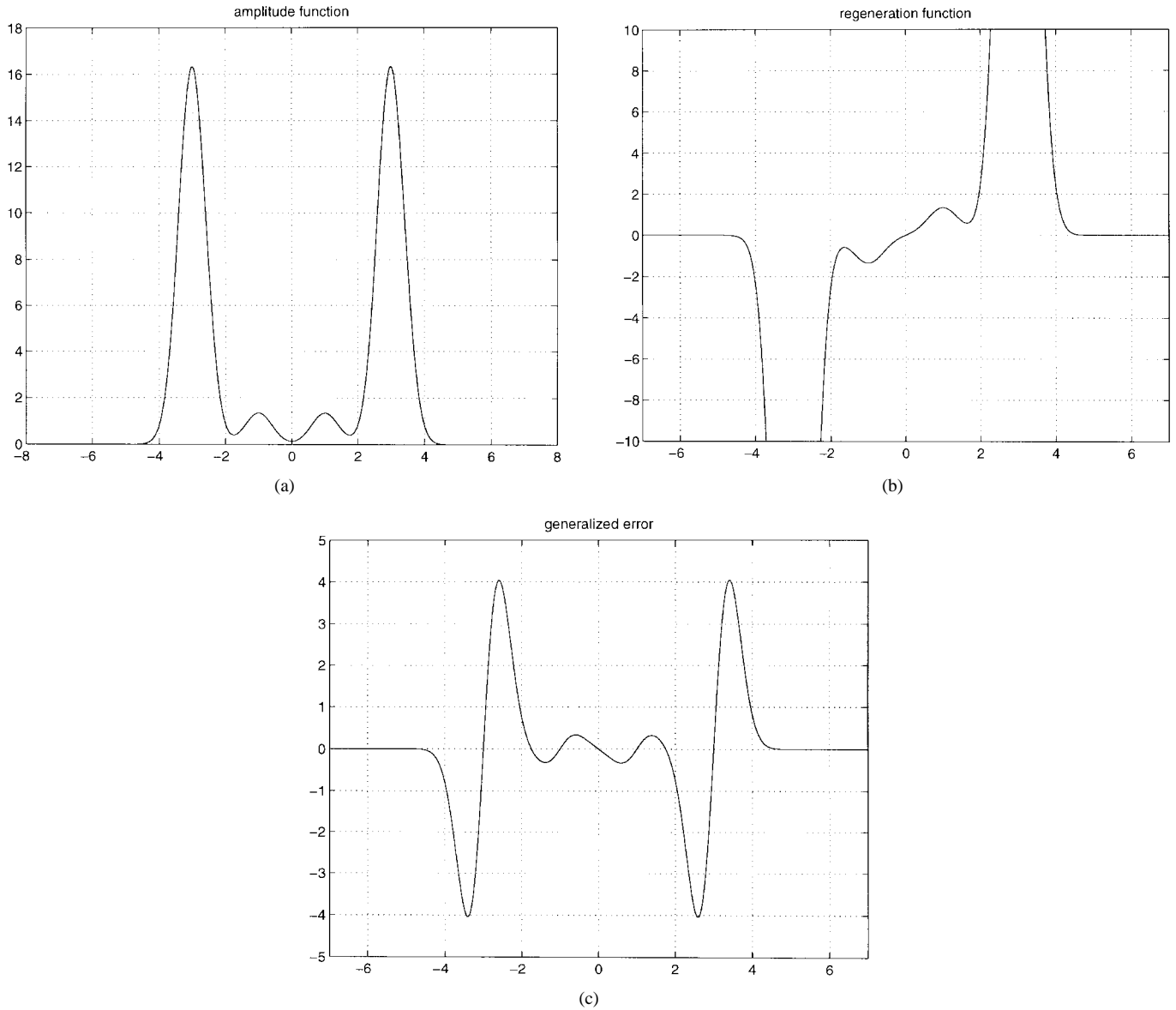
Fig. 4. Target and actual pdf's are 4-ASK and 5-ASK, respectively, with alphabets $= [-3, -1, 1, 3]$ and $[-2, -1, 0, 1, 2]$. (a) Amplitude function, (b) regeneration function, and (c) generalized error.

that results already arrived at for the FCF also apply to the SCF.

Proposition 7: The minimizing $\mathbf{w}$ for the SCF has the following structure for the multiple source case:

$$\mathbf{w} = \mathbf{R}_{nn}^{-1} \mathbf{H} \mathbf{b} \tag{45}$$

for some vector $\mathbf{b}$.

Proof: This proof is rather more elaborate. We can express the SCF as

$$J_2 = -E_{A'} \ln E_Z \frac{1}{\pi \sigma_t^2} e^{-|z-a'|^2/\sigma_t^2}$$
$$= -E_{A'} \ln E_{\mathbf{A}, \mathbf{N}} \frac{1}{\pi \sigma_t^2} e^{-|\mathbf{w}^H \mathbf{H} \mathbf{a}_k + \mathbf{w}^H \mathbf{n}_k - a'|^2/\sigma_t^2} \tag{46}$$

where A' is the target random variable. Expressing Z in terms of the RV's $\mathbf{A}$ and $\mathbf{N}$, where $\mathbf{A}$ includes all signals of interest, $\mathbf{N}$ is the noise vector RV. Using Proposition 1, we

can eliminate the expectation over $\mathbf{N}$ through modification of the nonlinearity

$$J_2 = -E_{A'} \ln E_{\mathbf{A}}$$
$$\left. \left[\frac{1}{\pi \sigma_t^2} e^{-|x-a'|^2/\sigma_t^2} * p_{-N_o}(x) \right] \right|_{x=\mathbf{w}^H \mathbf{H} \mathbf{a}_k},$$
$$p_{N_o}(n_o) = \left. \frac{1}{\pi \sigma_o^2} e^{-|n_o|^2/\sigma_o^2} \right|_{\sigma_o^2 = \mathbf{w}^H \mathbf{R}_{nn} \mathbf{w}.} \tag{47}$$

However, the convolution above is equivalent to the pdf of the addition of two Gaussian RV's. Therefore, we can conclude that the SCF equals

$$J_2 = -E_{A'} \ln E_{\mathbf{A}} \frac{1}{\pi(\sigma_t^2 + \sigma_o^2)}$$
$$\left. \cdot e^{-|x-a'|^2/(\sigma_t^2+\sigma_o^2)} \right|_{\substack{x=\mathbf{w}^H \mathbf{H} \mathbf{a}_k \\ \sigma_o^2 = \mathbf{w}^H \mathbf{R}_{nn} \mathbf{w}}}$$
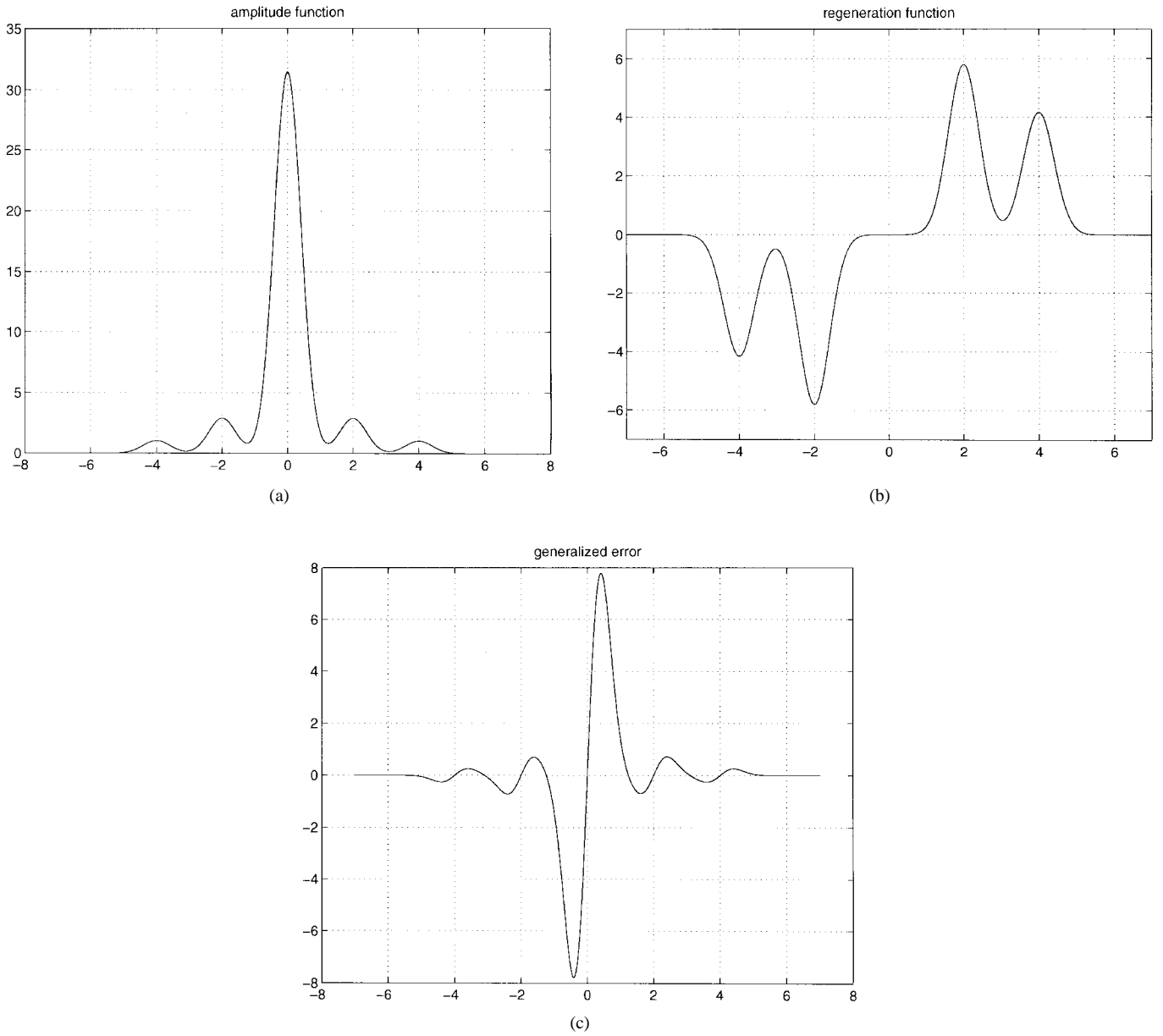$$= -E_{A'} \ln E_{\mathbf{A}} p_{N_t + N_o}(x - a') \tag{48}$$

Fig. 5. Target and actual pdf's are 5-ASK and 4-ASK, respectively, with alphabets $= [-4, -2, 0, 2, 4]$ and $[-4, -4/3, 4/3, 4]$. (a) Amplitude function. (b) Regeneration function. (c) Generalized error.

where the Gaussian nonlinearity has been expressed in terms of a pdf. Let the log-likelihood nonlinearity be expressed in the following terms:

$$\Psi = \ln E_{\mathbf{A}} p_{N_t + N_o}(x - a')$$
$$= \ln E_{\mathbf{A}} \Gamma(x - a'; \sigma_t^2 + \sigma_o^2) \tag{49}$$

where $\Gamma(t; \sigma^2)$ is a Gaussian in t of the specified variance. The gradient is then evaluated as

$$\nabla_{\mathbf{w}^H} J_2 = -E_{A'} \left\{ \left(\frac{\partial}{\partial \sigma_o^2} \Psi \right) \nabla_{\mathbf{w}^H} \sigma_o^2 \right.$$
$$\left. + \frac{E_{\mathbf{A}}[\Gamma \cdot (x - a')^* \nabla_{\mathbf{w}^H} x]}{E_{\mathbf{A}} \Gamma} \right\} \Bigg|_{\substack{x = \mathbf{w}^H \mathbf{H} a_k \\ \sigma_o^2 = \mathbf{w}^H \mathbf{R}_{nn} \mathbf{w}.}} \tag{50}$$

Therefore

$$\nabla_{\mathbf{w}^H} J_2 = -E_{A'} \left\{ \left(\frac{\partial}{\partial \sigma_o^2} \Psi \right) \nabla_{\mathbf{w}^H} \sigma_o^2 \right.$$
$$\left. - E_{\mathbf{A}}[e(x - a') \nabla_{\mathbf{w}^H} x] \right\} \Bigg|_{\substack{x = \mathbf{w}^H \mathbf{H} a_k \\ \sigma_o^2 = \mathbf{w}^H \mathbf{R}_{nn} \mathbf{w}}},$$
$$e(x - a') \stackrel{\text{def}}{=} \frac{\Gamma}{E_{\mathbf{A}} \Gamma} \cdot (x - a')^* \tag{51}$$

and solving for $\mathbf{w}$, we obtain

$$\mathbf{w} = \mathbf{R}_{nn}^{-1} \mathbf{H} \frac{E_{A'} E_A[a_k e(x - a')]}{E_{A'} \frac{\partial}{\partial \sigma_o^2} \Psi}$$
$$= \mathbf{R}_{nn}^{-1} \mathbf{H} \mathbf{b}. \tag{52}$$

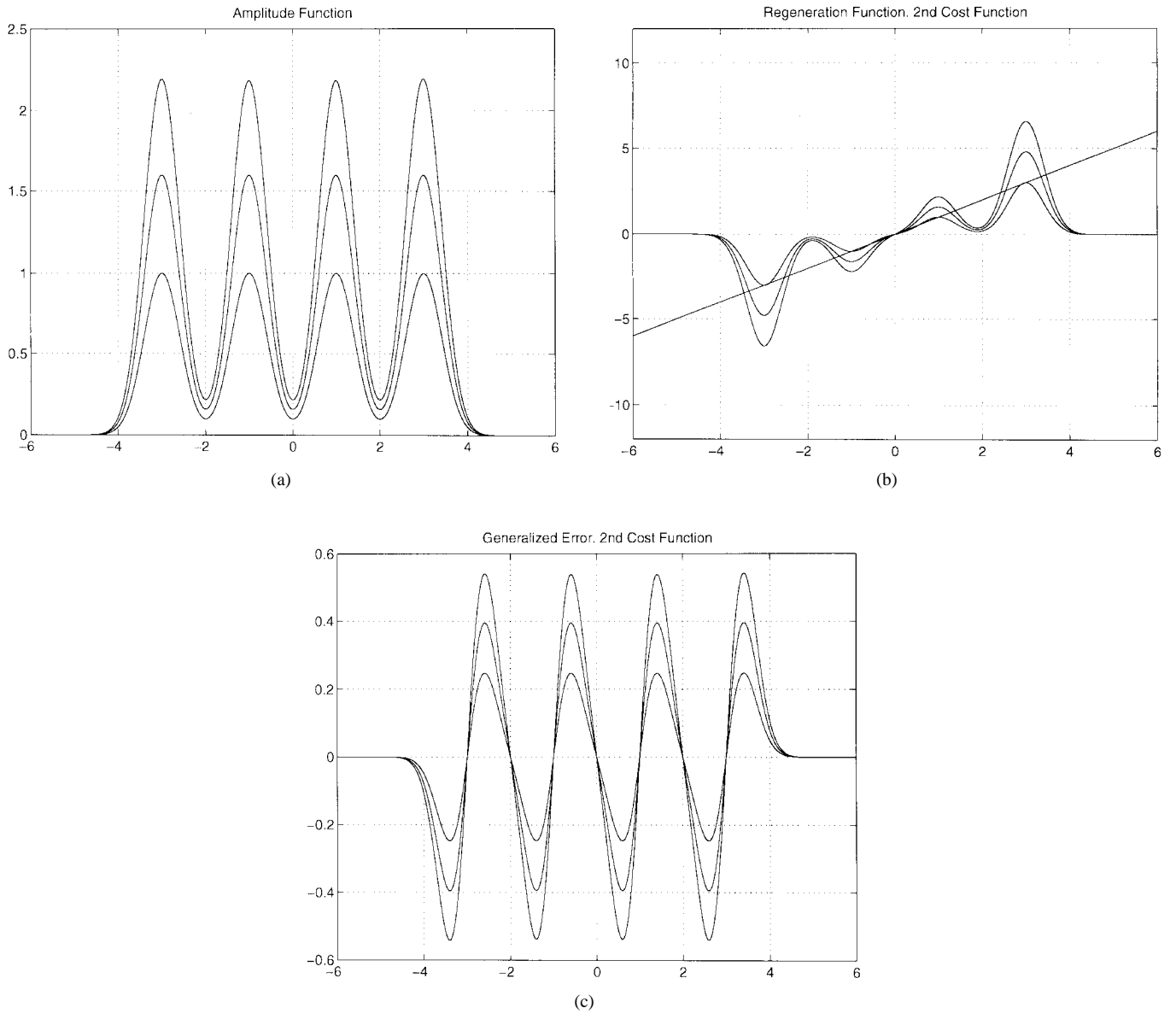which completes the proof.   ◆

Fig. 6. Comparison of the same nonlinearities for several values of the input signal to noise power ratio. Target and actual pdf's are 4-ASK and 4-ASK, respectively, with alphabet $= [-3, -1, 1, 3]$. (a) Amplitude function. (b) Regeneration function. (c) Generalized error nonlinearities.

An expression for the gradient analogous to that of the FCF can also be found, as well as an analogous set of quality functions. The significance of these and of the equivalent regeneration function is rather different for the SCF, as we will see shortly. First, let us note that the gradient of the SCF can be expressed as

$$\nabla_{\mathbf{w}^H} J_2 = E_{\mathbf{X}} E_{A'} q_2(z, a')(z - a')^* \mathbf{x}_k,$$

$$q_2(z, a') \overset{\text{def}}{=} \frac{e^{-|z-a'|^2/\sigma_t^2}}{E_Z e^{-|z-a'|^2/\sigma_t^2}} \qquad (53)$$

where now, the expression for the q functions differs in the denominator, where now, the expectation is taken with respect to the actual distribution Z in lieu of the target distribution A'. The consequence of this is very important as it introduces memory in the generalized error function. This makes the generalized error truly sensitive to the actual pdf of the filter

output and helps avoid many local minima present in the FCF. We can decompose the generalized error in the following way:

$$\varepsilon_2 = E_{A'} q_2(z, a')(z - a')$$
$$= \zeta(z) z - \hat{a}_2,$$
$$\zeta(z) \overset{\text{def}}{=} E_{A'} q_2(z, a'),$$
$$\hat{a}_2 \overset{\text{def}}{=} E_{A'} a' q_2(z, a') \qquad (54)$$

where we have introduced the amplitude function $\zeta(z)$ and the regeneration function for the SCF $\hat{a}_2$. Note that the amplitude function did not appear in the FCF. This is due to the fact that the regeneration function in this case provides us with a distorted reference whose degree of distortion depends on the power of input additive noise. The role of the amplitude function is then to distort the output Z accordingly to yield a sensible generalized error. We will see the working of these nonlinearities in the examples of Figs. 2 to 6.
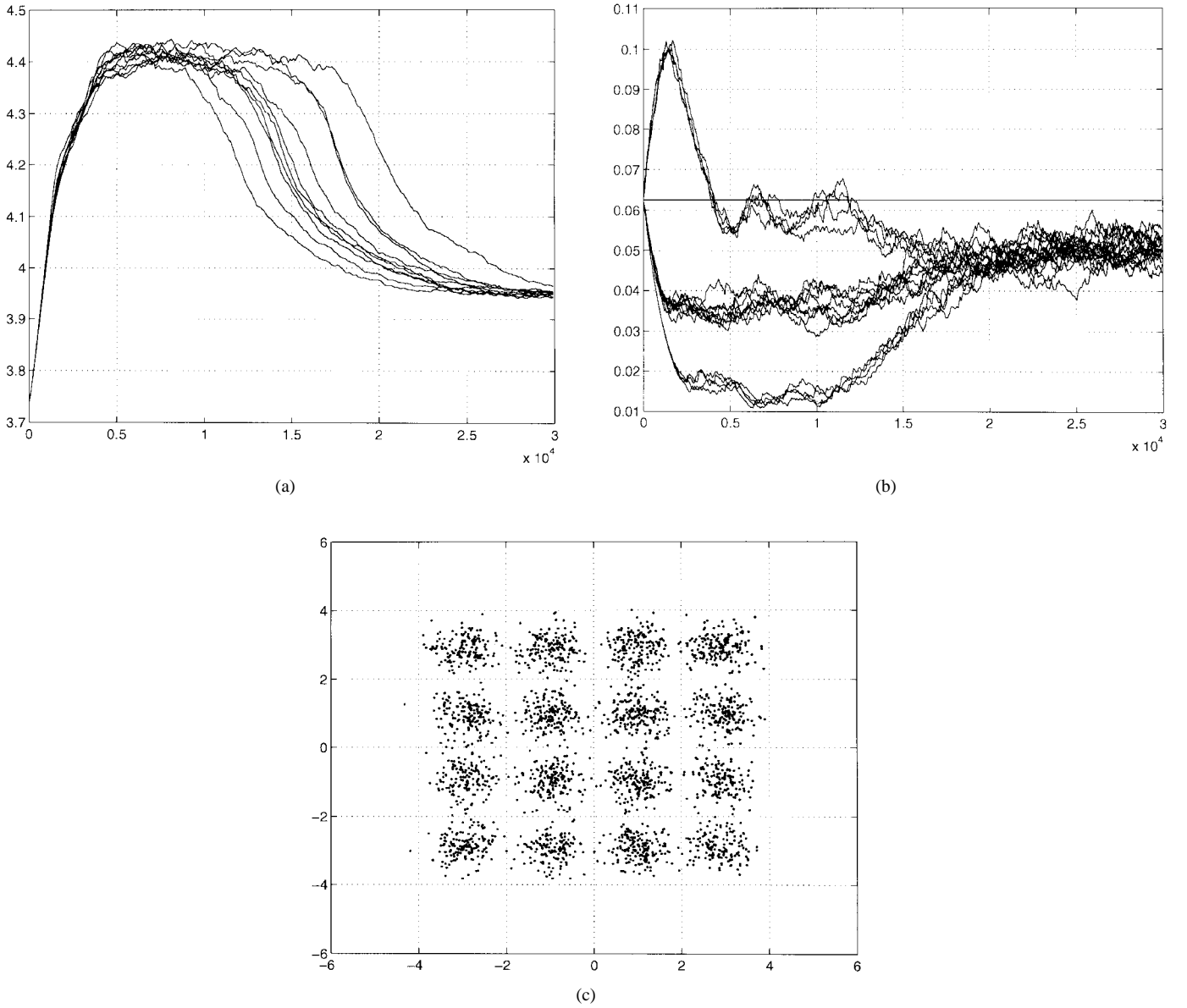
Fig. 7. (a) Time evolution of the cost function over 10 realizations. Note that the time to convergence may display important variations depending on the generated data. The initial rise is due to the time needed for the cost function to fill its memory with reliable probability estimates. (b) Time evolution of the estimated probabilities for a particular realization. The target probability density of the symbols has been represented as a straight line centered at 1/16. The estimated probabilities are smaller than 1/16 due to the aperture of the Gaussian window used for estimation, which is controlled with the σ_t parameter. (c) constellation obtained in convergence.

Proposition 8: The minimizing $\mathbf{w}$ for the SCF can be alternatively expressed as

$$\mathbf{w} = \tilde{\mathbf{R}}_{xx}^{-1} E_{\mathbf{X}} \hat{a}_2^* \mathbf{x}_k,$$
$$\tilde{\mathbf{R}}_{xx} \stackrel{\text{def}}{=} E_{\mathbf{X}} \zeta(z) \mathbf{x}_k \mathbf{x}_k^H \quad (55)$$

where a pseudo-covariance matrix of the data $\mathbf{x}_k$, $\tilde{\mathbf{R}}_{xx}$ has been defined in terms of the amplitude function ζ.

Proof: This can be proven in the same way as Proposition 4. Let us equate the gradient to zero,

$$E_{\mathbf{X}} E_{A'} q_2(z, a')(z - a')^* \mathbf{x}_k = \mathbf{0}.$$

Splitting this equality and using $z = \mathbf{w}^H \mathbf{x}_k$ leads to

$$E_{\mathbf{X}} E_{A'} q_2(z, a') \mathbf{x}_k z^* = E_{\mathbf{X}} E_{A'} q_2(z, a') \mathbf{x}_k a'^*$$

$$E_{\mathbf{X}} E_{A'} q_2(z, a') \mathbf{x}_k \mathbf{x}_k^H \mathbf{w} = E_{\mathbf{X}} \hat{a}_2^* \mathbf{x}_k$$
$$[E_{\mathbf{X}} \zeta(z) \mathbf{x}_k \mathbf{x}_k^H] \mathbf{w} = E_{\mathbf{X}} \hat{a}_2^* \mathbf{x}_k$$

which concludes the proof.   ♦

It is important to note that in the last proposition, the pseudo-covariance, in a way, compensates for the amplitude distortion incurred by the regeneration function of the SCF via the amplitude function present in the denominator. It also bears some resemblance to the Wiener solution, although in a more indirect way as it happened for the FCF.

A. Regeneration and Amplitude Functions

We will compare in this section the behavior of the regeneration function for the SCF with that of the FCF. We will see that there exist some major differences. The regeneration function
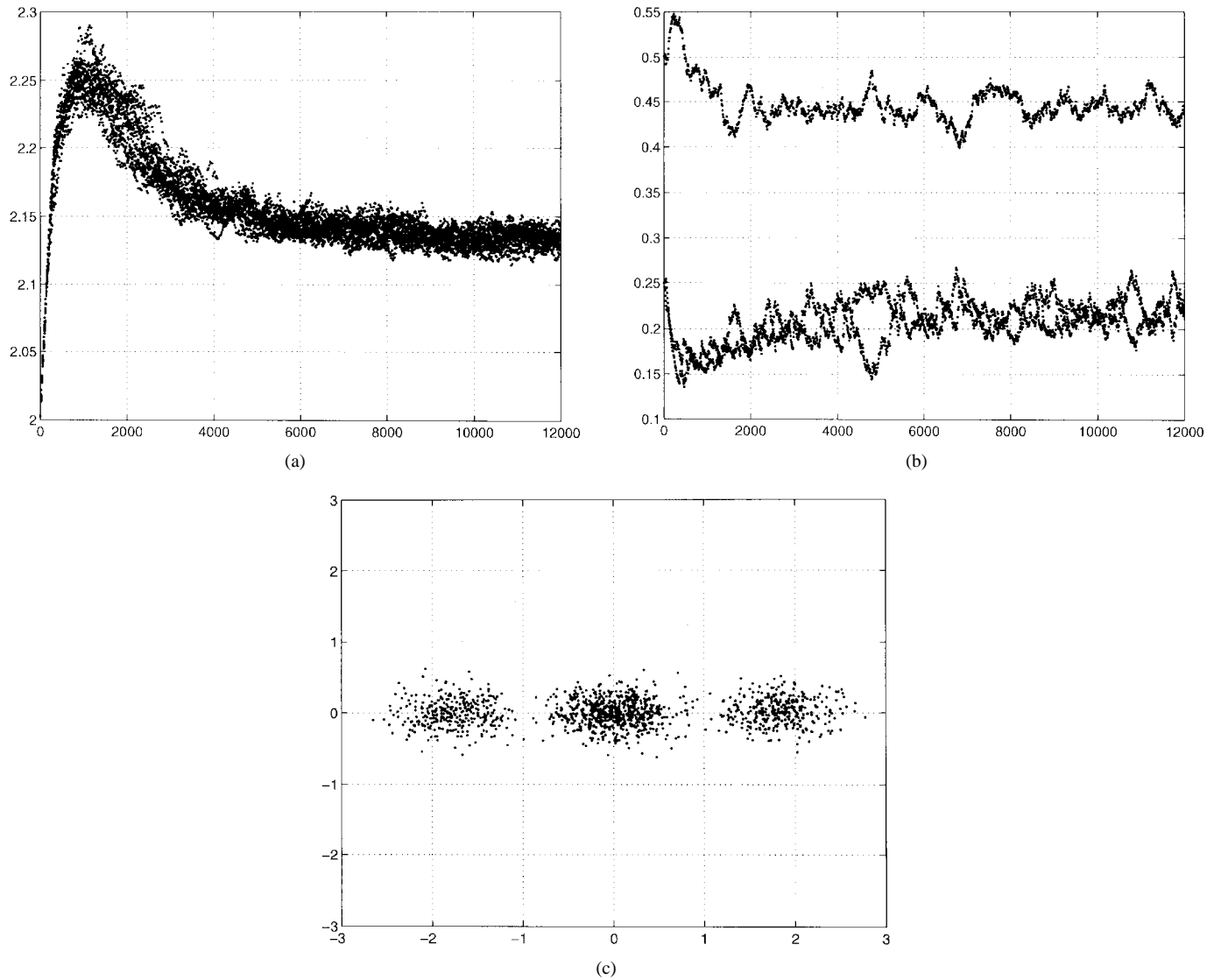
Fig. 8. (a) Time evolution of the cost function over 10 realizations. (b) Time evolution of the estimated probabilities. (c) Constellation obtained in convergence.

for the SCF is truly sensitive to the pdf of the actual RV. The contrary happened for the FCF. It was dependent only on the pdf of the target RV. In addition, some amplitude distortion is introduced by the second regeneration function (SRF). The role of the amplitude function is then to compensate for this distortion in the expression of the error. We will see in the expressions for the minimizing $\mathbf{w}$ that this amplitude function explicitly appears. In this section, we will show some illustrative examples of the working of the SRF and of the amplitude function. We will also see how this is expressed in terms of the generalized error. We will consider three cases involving M-ASK signals in terms of the actual and target RV as they appear in the following figures:

- The target pdf and the actual pdf coincide (Fig. 3).
- The target pdf and the actual pdf do not coincide (Figs. 4 and 5).
- The target pdf and the actual pdf do coincide and several values of input SNR are considered (Fig. 6).

Note that in the last set of figures, the maxima of the amplitude function when the target and the actual pdf's coincide

are always found to be above 1. This amplitude distortion will be compensated for by the regeneration function (observe that for those values coinciding with the alphabet $\{c_i\}$, the regeneration function is also larger: $|\hat{a}_2(c_i)| \geq |c_i|$). We can now see in the generalized error that its zeros with positive slope also coincide with the values of the target alphabet. On the contrary, the features that we have enumerated here do not hold when the target and actual pdf's differ, as shown in the Fig. 4.

In this new set, we have provided the nonlinearities for 4-ASK and 5-ASK as target and actual pdf's. Note that now, large distortions can be observed in all figures due to the difference between both pdf's. In particular, note that those values that are common to both distributions, that is, -1 and 1 do not show a large distortion. On the contrary, those that are not common have undergone a huge amplitude distortion. A situation like this could easily have resulted in a local minimum for a Bussgang-type algorithm. This is not the case for the second cost function as when noncoinciding values of the actual distribution occur at the output of the adaptive system; the present setting of the coefficients will change as
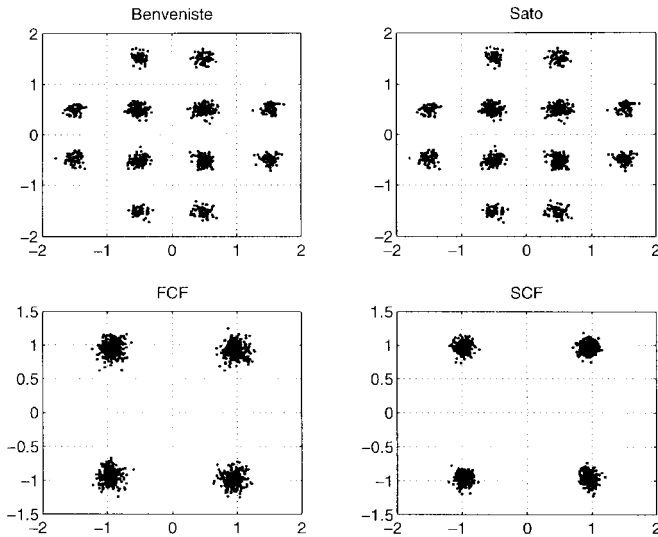
Fig. 9. Constellations obtained in convergence when the input signal is QPSK for (upper left) Benveniste–Goursat, (upper right) Sato, (lower left) FCF, and (lower right) SCF.

the gradient of the SCF is sensitive to this situation. Note that for noncoinciding values, the generalized error shows a negative slope that makes unstable that particular setting of the coefficient vector $\mathbf{w}$.

We have shown in Fig. 5 the reverse of what was happening in Fig. 4. The same conclusions can be obtained.

In Fig. 6, we show the influence of the input additive noise variance on the algorithm nonlinearities when the target and actual pdf's coincide. It is worthwhile to note that as this variance goes to zero, the induced distortion of the amplitude and regeneration functions disappears. See Fig. 6(a), where the maxima of the amplitude function goes toward one, and Fig. 6(b), where $\hat{a}_2(c_i)$ goes to c_i accordingly.

B. Extended Cost Function

Some remarks concerning the second cost function are in order. The second cost function has been shown to recover a signal of the wanted statistics. Nevertheless, we can imagine a scenario (i.e., an array) where more than one signal of the same statistics is being received. There is no way of discerning between the two unless some extra information is provided to the algorithm. There is one way, however. A suitable extension of the cost function is defined to impose at the output of the adaptive system the joint pdf of several random variables. In this way, we can impose that a vector signal whose marginal distributions are equal can be recovered. Due to extension constraints, this is the subject of a forthcoming paper.

C. Comparison between the FCF and the SCF

We have seen that the first cost function leads to soft-decision strategies in terms of the regeneration function. This regeneration is more robust than hard decision and achieves better performance, as has already been proved in other papers. The second cost functions instead differ in nature from the first, although their mathematical expression is very similar. The regeneration function is indeed fully sensitive to the actual probability density function at the output of the adaptive
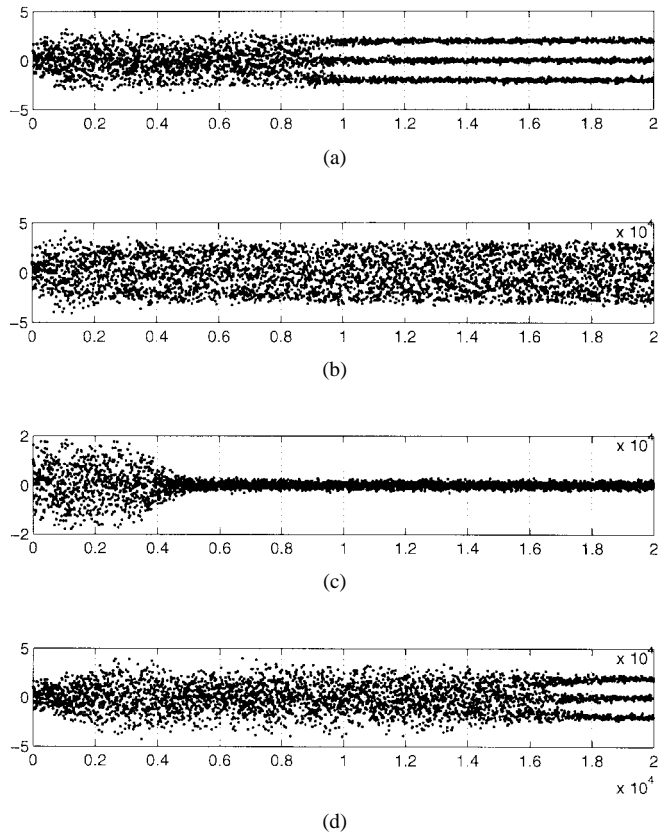

Fig. 10. Time evolution of the in-phase channel when the input signal is 9-QAM for: (a) Benveniste–Goursat. (b) Sato. (c) FCF. (d) SCF. Time is in units of $10\,000$.

system. Let us consider two different kinds of cost functions in the following terms:

$$\Im_1(\mathbf{w}) = E_Z \Phi[E_{A'} \Psi(z - a')]$$
$$\Im_2(\mathbf{w}) = E_{A'} \Phi[E_Z \Psi(z - a')]. \qquad (56)$$

We readily see that the FCF falls under the first category and is therefore soft-decision related (it leads to some kind of signal regeneration), whereas the SCF is of the latter type and, therefore, truly pdf sensitive. The functions Φ and Ψ are monotone decreasing and indicator functions, respectively. The single difference is whether the outer and inner expectation operators are referred to the actual or the target distribution. To really enforce a pdf and fully implement the statistical reference criterion, we should always place it outside the expectation operator on the target pdf. Otherwise, our algorithm may be prone to some capture problems. In this way, Φ is evaluated on quantities that are attributable to estimated probabilities (hence, the pdf sensitivity), and the nonlinear operation of Φ helps in identifying those false solutions where the target and actual pdf's are far from coinciding. It appears then that if we always want to apply statistical reference, we must resort to a memory nonlinearity.

V. RESULTS

In this section, we will validate the theoretical conclusions at which we have arrived in the previous sections. We will consider two possible applications of the statistical reference

criterion, viz., blind equalization and beamforming. We will show also that the SCF is very robust in terms of intended signal to inference powers beause it is able to excise jammers much more powerful than the signal itself. The coefficient update equations are provided for each case.

A. SCF Implementation

Up until now, both cost functions have been formulated using the statistical expectation operator E. In a real setting, some approximation has to be introduced as an infinite data record will not be available. From the expression of the gradient of the SCF in (54), we propose the following approximation to the gradient-based update equations for the coefficients:

$$\hat{\nabla}_{\mathbf{w}^H} J_2 = E_{A'} \frac{\mathbf{d}_k(a')}{\nu_k(a')}$$

$$\begin{cases} \mathbf{d}_k(a') = (1 - \mu_p)\mathbf{d}_{k-1}(a') \\ \qquad + \mu_p e^{-|z_k - a'|^2/\sigma_t^2}(z_k - a')^* x_k \\ \nu_k(a') = (1 - \mu_p)\nu_{k-1}(a') \\ \qquad + \mu_p e^{-|z_k - a'|^2/\sigma_t^2} \end{cases}$$

$$(57)$$

where the statistical expectation present in (54) in both the numerator and the denominator are estimated with a step-size μ_p. The estimates are called $\mathbf{d}_k(a')$ and $\nu_k(a')$, respectively. Note that the term in the denominator is related precisely to the estimated actual probability of the value a' under the output distribution. If the value of μ_p is small enough, we can reasonably surmise that the evolution of the output distribution is more or less stationary within a time interval in the order of $1/\mu_p$. It remains only to choose a step size such that the estimated probabilities are a reasonable approximation to the actual ones. That is, if the output distribution is made up of, say, M centroids, each of probability $1/M$, a proper value for μ_p could be $0.01/M$ such that on average each centroid occurs 100 times within an interval of length $1/\mu_p$, and the estimated probability is close to the true one. The update equations for the gradient are then

$$\mathbf{w}_k = \mathbf{w}_{k-1} - \mu \hat{\nabla}_{\mathbf{w}^H} J_2$$
$$= \mathbf{w}_{k-1} + \mu E_{A'} \frac{\mathbf{d}_k(a')}{\nu_k(a')}. \qquad (58)$$

B. Equalization

The performance of the first and second cost functions has been compared with the Benveniste–Goursat and Sato algorithms. The comparison between both sets of algorithms depends on five factors: type of constellation, parameters of the adaptive algorithm (step-size and other parameters), channel models, signal-to-noise ratio, and presence/absence of interfering signals. Many examples may be found where the behavior of the four algorithms (Benveniste–Goursat, Sato, FCF, and SCF) differ. For example, for a given channel and the same input signal, a change in the step size of the coefficients may mean that Benveniste–Goursat as compared with SCF is relatively slower or faster or that it converges

to the true minimum or to a false minimum. On average, the simulations that we have carried out show that in the absence of interfering signals, Benveniste and Sato appear to be faster than statistical reference algorithms (the reason for this is that statistical reference algorithms need some time to estimate the probabilities of the actual distribution evaluated at the constellation symbols, and this depends on the number of symbols of the target constellation).

Nevertheless, when interfering signals are present, the Benveniste–Goursat and Sato algorithms fail to recover the signal of interest. On the contrary, statistical reference algorithms succeed in delivering the true signal at the output, as shown in the beamforming simulations.

Simulation 1: We now focus on the equalization of a 16-QAM constellation over 30 000 symbols using the second cost function. The shaping pulse used for modulation is the square-root raised cosine of roll-off 0.4 and truncated to three symbols. The channel signal-to-noise ratio is 10 dB. The channel has been modeled as a filter of coefficients: $\mathbf{h} = [-0.1274, 0.5542, -1.0973, -0.7313, 1.4047, -0.6202, 0.2371, -1.5868, -0.4015, -0.7707]$. A fractionally spaced equalizer at four samples per symbol and 40 coefficients in length has been used. The step size for the coefficient updates is 0.004, and the data are normalized to unity power. The step size for the estimated probability updates has been chosen as 0.01th of the symbol probability: 1/1600. The parameter $1/\sigma_t^2$ has been chosen to be 1.2.

Simulation 2: This simulation focusses on equalization of a duobinary signal over 12 000 symbols using the second cost function (see Figs. 7 and 8). The shaping pulse used for modulation is the square-root raised cosine of roll-off 0.4 and truncated to three symbols. The channel signal-to-noise ratio is 10 dB. The channel has been modeled as a filter of coefficients: $\mathbf{h} = [0.5, 0, 0, 0, 1, 0, 0, 0, 1]$. A fractionally spaced equalizer at four samples per symbol and 40 coefficients in length has been used. The step size for the coefficient updates is 0.004, and the data are normalized to unity power. The step size for the estimated probability updates has been chosen as 0.003. The parameter $1/\sigma_t^2$ has been chosen to be 1.2.

Simulation 3: This simulation is a comparison of Benveniste–Goursat, Sato, FCF, and SCF with an equalization of a QPSK signal (Fig. 9) and of a 9-QAM signal (Fig. 10) over 20 000 symbols. The shaping pulse used for modulation is the square-root raised cosine of roll-off 0.4 and truncated to three symbols. The channel signal-to-noise ratio is 20 dB. The channel has been modeled as a filter of coefficients: $\mathbf{h} = [0.1274, 0.5542, 1.0973, 0.7313, 1.4047, -0.6202, -0.2371, -1.5868, -0.4015, 0.7707]$. A fractionally spaced equalizer at four samples per symbol and 30 coefficients in length has been used. The step size for the coefficient updates is 0.0009, and the data are normalized to unity power. The step size for the estimated probability updates has been chosen as 0.0025 for the QPSK case and 0.0011 for the 9-QAM case. The parameter $1/\sigma_t^2$ has been chosen to be 1.1 (see Figs. 11 and 12).

Both Benveniste–Goursat and Sato converge to a false minimum for the QPSK case in Fig. 9. The FCF and the SCF

converge instead to the true minimum. In the second case, only the modulation and the step size for the estimated probability updates have been changed. Benveniste–Goursat and the SCF converge, whereas Sato and the FCF do not. Note also that the SCF takes longer as it needs more time to estimate the probabilities. For the FCF, the coefficient vector $\mathbf{w}$ converges to zero as it interprets the output to be the symbol $0+j0$ of the 9-QAM constellation. For those modulations that contain the $0+j0$ symbol, it is always better to use the SCF. Simulations with interfering signals are shown next in the beamforming section.

C. Beamforming

The structure that has been considered in the beamforming simulations is that of an equalizer in cascade with a beamformer. This has a masking effect in the acquisition regime, with the consequence that until the equalizer does not "see" a sufficiently well focused signal, convergence does not start. To this moment, no other algorithms based on statistical reference are known to us. For the sake of comparison, we have chosen the CMA algorithm that, in a way, enforces a constant amplitude distribution at the system output, although not in the same way that we are considering here. For example, the CMA algorithm would not be able to distinguish between a QPSK and a BPSK signal, both of the same amplitude, whereas our algorithm is able to do so. Several scenarios have been modeled in the experiments, with interferers that are more powerful than the signal of interest.

The correlation between the beamforming coefficients and the normalized steering vector of the interferers is used to represent the time evolution of the beam pattern: $\mathbf{w}_s^H \mathbf{s}_i / \mathbf{s}_i^H \mathbf{s}_i$, with $\mathbf{w}_s$ the beamforming coefficients and $\mathbf{s}_i$ the steering vector of the ith interfering source. For the source of interest, this correlation tends to 1.

Simulation 1: This simulation focusses on combined beamforming and equalization of a 9-QAM signal at $0°$ in the presence of two interferers: a 16-QAM at $50°$ and a duobinary signal at $-40°$. Their power relative to the signal of interest is 1.05 and -5.70 dB, respectively. A Gaussian point source has also been placed at $25°$. The number of symbols is 40 000. The parameter $1/\sigma_t^2$ has been chosen to be 1.1. The step size for the adaptation of the estimated probabilities is 0.0015, and that for the adaptation of the coefficients of the equalizer and beamformer has a exponential variation to guarantee faster convergence: $\mu(n) = 0.0004 + 0.0008e^{-n/10\,000}$. An equalizer 30 coefficients in length and a seven-coefficient beamformer have been chosen. The signal-to-noise ratio is 10 dB.

Simulation 2: This simulation focusses on combined beamforming and equalization of a BPSK signal at $0°$ in the presence of two interferers: a 16-QAM at $50°$ and a duobinary signal at $40°$. Their power relative to the signal of interest is 8.32 and 7.53 dB, respectively. A Gaussian point source has also been placed at $25°$. The number of symbols is 40 000. The parameter $1/\sigma_t^2$ has been chosen to be 1.4. The step size for the adaptation of the estimated probabilities is 0.005, and that for the adaptation of the coefficients of the equalizer and beamformer has been chosen to be 0.004. An equalizer 30

coefficients in length and a seven-coefficient beamformer have been chosen. The signal-to-noise ratio is 10 dB.

The CMA algorithm has been applied to the same scenario in Simulation 2 because the BPSK signal has constant amplitude. For the same settings of the step sizes and a wide range about those values, it has failed to converge due to the very tough conditions in terms of the power of the interferers. An alternative CMA algorithm, where the amplitude error is normalized by $1 + |y|^2$ to guarantee better convergence behavior, has also failed to recover the BPSK signal in this scenario.

VI. CONCLUSION

In this paper, we have introduced a new criterion in the context of adaptive filtering—the statistical reference criterion—that provides a baseline for designing blind algorithms. The universality of this criterion has been proved in successfully dealing with a number of important applications as equalization and beamforming. Robust behavior has been shown in the simulations where, for the beamforming problem, the algorithm can lock on the signal of the intended distribution in very unfavorable SINR ratios. The algorithm is sensitive to the pdf of the actual distribution and is thus more robust in the presence of interference than are blind algorithms of the Bussgang-type, although the cost function is still not convex.

APPENDIX
CONSTANT MODULUS ALGORITHM (FCF)

In this Appendix, we will derive the regeneration function for a constant modulus algorithm based on the FCF. The expression of the pdf of a constant-amplitude RV appears depicted in (21). In the following, we will use A to denote the target amplitude and A_o to denote the constant-amplitude random variable of pdf $p_{A_o}(a_o) = (1/2\pi A)\,\delta(|a_o| - A)$. The argument of the logarithm of the FCF, which appears at the denominator of the regeneration function, is expressed now as

$$E_{A_o} e^{-|z-a_o|^2/\sigma_t^2} = e^{-\sigma_t^{-2}(|z|^2 + A^2)} E_{A_o} e^{2\sigma_t^{-2}\,\mathrm{Re}\,\{z^* a_o\}}$$
$$= e^{-\sigma_t^{-2}(|z|^2 + A^2)} B(z). \tag{59}$$

From now on, we will use the variable $x = 2A|z|/\sigma_t^2 = 2A\rho/\sigma_t^2$ with $a_o = Ae^{j\theta}$ and $z = \rho e^{j\alpha}$ expressed in polar coordinates. $B(z)$ can then be expressed as

$$B(z) = E_{A_o} e^{2\sigma_t^{-2}\,\mathrm{Re}\,\{z^* a_o\}}$$
$$= \int_0^\infty \int_{-\pi}^{+\pi} \frac{1}{2\pi A} \delta(r - A) e^{2\sigma_t^{-2}\rho r \cos(\theta - \alpha)} r \, dr \, d\theta$$
$$= \frac{1}{2\pi} \int_{-\pi}^{+\pi} e^{x \cos(\theta - \alpha)} \, d\theta$$
$$= \frac{1}{2\pi} \int_{-\pi}^{+\pi} e^{x \cos\theta} \, d\theta$$
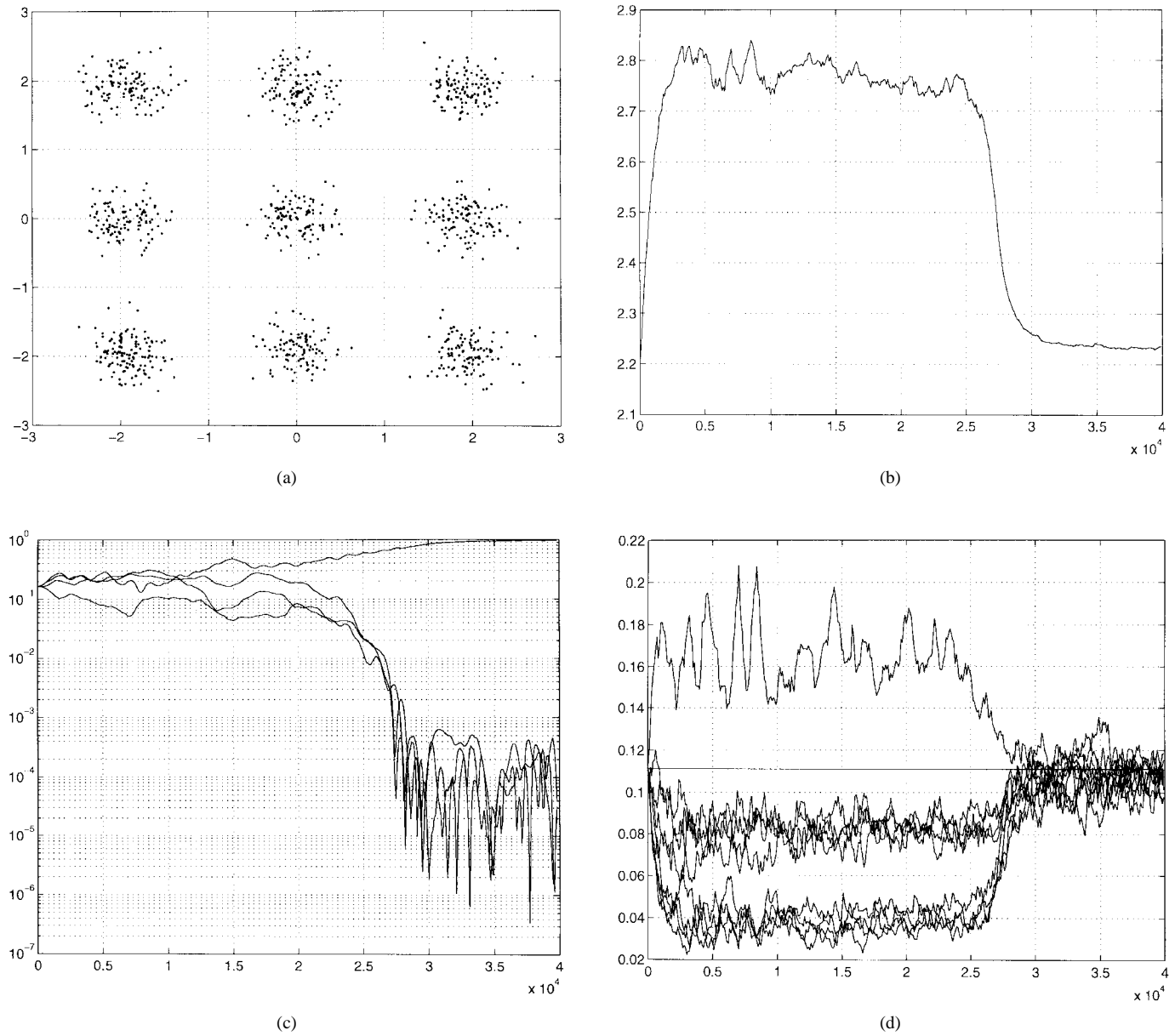$$= \frac{1}{2\pi} B'(x). \tag{60}$$

Fig. 11.   (a) Constellation obtained in convergence. (b) Time evolution of the cost function. (c) Decorrelation obtained between the coefficient vector of the beamformer and the steering vectors of the signal of interest and of the interferers. In this way, we represent the time evolution of a few characteristic samples of the beam response. Observe that for the signal of interest the curve converges to some value close to 0 dB. (d) Time evolution of the estimated probabilities; we can see how they converge to 1/9.

Differentiating with respect to x, the first and second derivatives may be expressed as

$$\dot{B}'(x) = \int_{-\pi}^{+\pi} e^{x \cos \theta} \cos \theta \, d\theta,$$

$$\ddot{B}'(x) = \int_{-\pi}^{+\pi} e^{x \cos \theta} \cos^2 \theta \, d\theta \qquad (61)$$

carrying out the integral in the first derivative by parts, we have

$$\dot{B}'(x) = e^{x \cos \theta} \sin \theta|_{-\pi}^{+\pi} + x \int_{-\pi}^{+\pi} e^{x \cos \theta} \sin^2 \theta \, d\theta$$

$$= x \int_{-\pi}^{+\pi} e^{x \cos \theta} \sin^2 \theta \, d\theta. \qquad (62)$$

Therefore, $B'(x)$ fulfills the following differential equation:

$$\ddot{B}'(x) + \left(\frac{1}{x}\right) \dot{B}'(x) = \int_{-\pi}^{+\pi} e^{x \cos \theta} (\sin^2 \theta + \cos^2 \theta) \, d\theta$$

$$= B'(x). \qquad (63)$$

If we compare this equation with the modified Bessel equation, we get

$$x^2 \ddot{y} + x \dot{y} - (x^2 + n^2) y = 0 \rightarrow$$

$$y = c_1 I_n(x) + c_2 K_n(x) \qquad (64)$$

where I_n and $K_n(x)$ are the Bessel modified functions of the first and second kind of order n, respectively. The modified Bessel equation can be matched to (63) by setting $n = 0$ and dividing by x^2 on both sides. The constants of the linear combination must be such that $B'(0) = 2\pi$ (which is
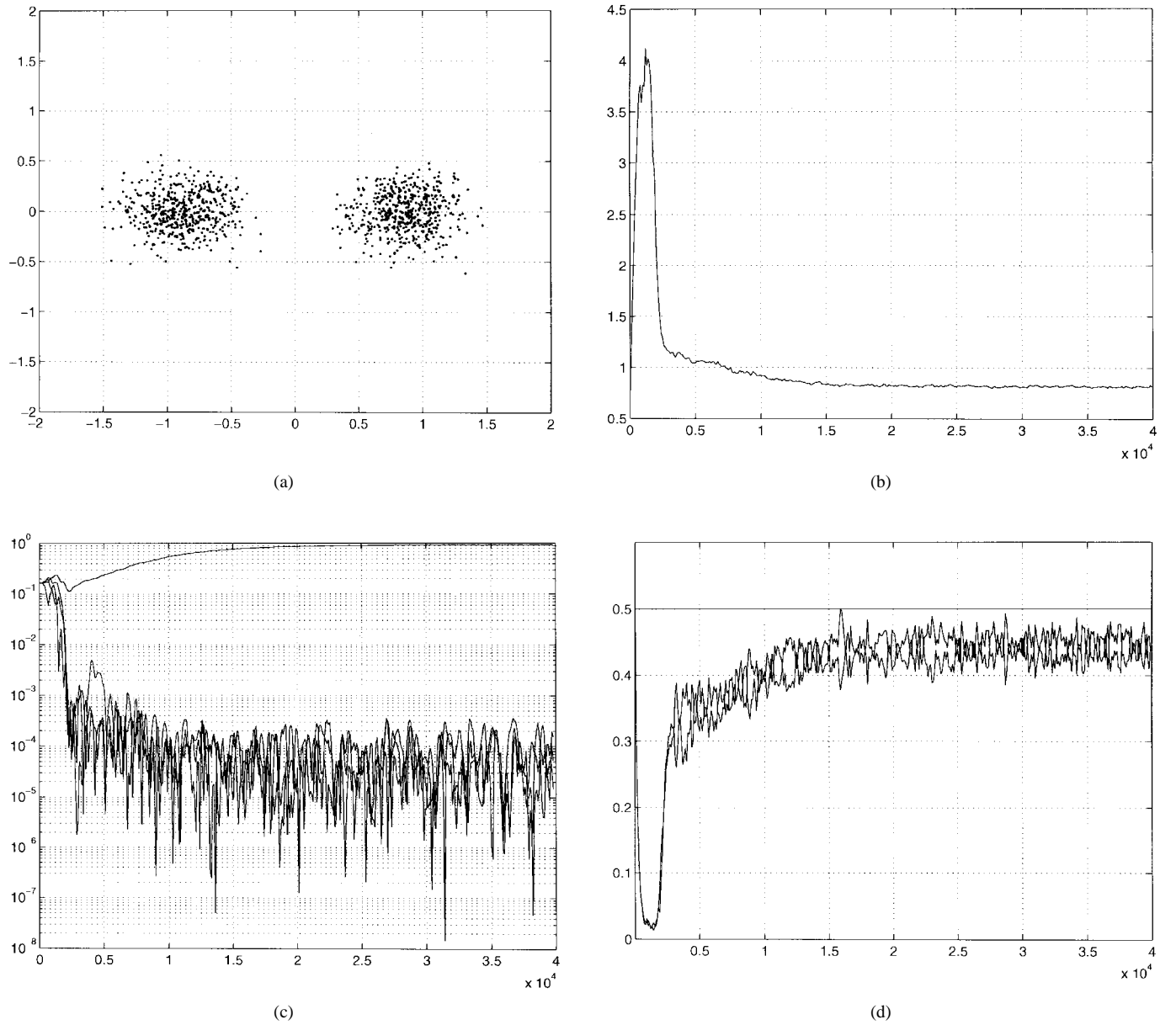
(a)

(b)

(c)

(d)

Fig. 12.   (a) Constellation obtained in convergence. (b) Time evolution of the cost function. (c) Decorrelation obtained between the coefficient vector of the beamformer and the steering vectors of the signal of interest and of the interferers. Observe that for the signal of interest, the curve converges to some value close to 0 dB. (d) Time evolution of the estimated probabilities; we can see how they converge to 1/9.

immediate from the expression of $B'(x)$). As $K_0(0)$ tends to infinity and $I_0(0) = 1$, we must have that $c_1 = 2\pi$ and $c_2 = 0$. Therefore,

$$B'(x) = 2\pi I_0(x)$$
$$= 2\pi I_0\left(\frac{2A}{\sigma_t^2}|z|\right). \tag{65}$$

The regeneration function is expressed now as

$$\hat{a}_k = E_{A_o} a_o \frac{e^{-|z-a_o|^2/\sigma_t^2}}{E_{A'_o} e^{-|z-a'_o|^2/\sigma_t^2}}$$
$$= E_{A_o} a_o \frac{e^{2\sigma_t^{-2} \operatorname{Re}\{z^* a_o\}}}{B(z)}$$
$$= \frac{1}{B(z)} E_{A_o} a_o e^{2\sigma_t^{-2} \operatorname{Re}\{z^* a_o\}}. \tag{66}$$

Expressing the expectation over A_o in terms of an integral in polar coordinates, we have

$$C(z) \overset{\mathrm{def}}{=} E_{A_o} a_o e^{2\sigma_t^{-2} \operatorname{Re}\{z^* a_o\}}$$
$$= \int_0^\infty \int_{-\pi}^{+\pi} p_{A_o}(a_o) a_o e^{2\sigma_t^{-2} \operatorname{Re}\{z^* a_o\}} r\, dr\, d\theta \tag{67}$$

substitution of the expressions of z, a_o, and the pdf of A_o in the above equation leads to

$$C(z) = \int_0^\infty \int_{-\pi}^{+\pi} \frac{1}{2\pi A} \delta(r - A) r e^{j\theta}$$
$$\cdot e^{2\sigma_t^{-2} \rho r \cos(\alpha - \theta)} r\, dr\, d\theta$$
$$= \frac{A}{2\pi} \int_{-\pi}^{+\pi} e^{j\theta} e^{2\sigma_t^{-2} A\rho \cos(\alpha - \theta)}\, d\theta. \tag{68}$$

Letting $\alpha - \theta = \beta$, we have

$$
\begin{aligned}
C(z) &= \frac{A}{2\pi} e^{j\alpha} \int_{-\pi}^{+\pi} e^{-j\beta} e^{2\sigma_t^{-2} A\rho \cos\beta} \, d\beta \\
&= \frac{A}{2\pi} e^{j\alpha} \int_{-\pi}^{+\pi} e^{2\sigma_t^{-2} A\rho \cos\beta} \cos\beta \, d\beta \\
&= \frac{A}{2\pi} e^{j\alpha} \dot{B}'(2\sigma_t^{-2} A\rho)
\end{aligned}
\tag{69}
$$

with $C(z)$ expressed in terms of B'. We have seen that B' is $2\pi I_0$, and we also know that I_1 is the derivative of I_0; therefore, the regeneration function is finally expressed as

$$
\begin{aligned}
\hat{a}_k &= \frac{I_1(2\sigma_t^{-2} A\rho)}{I_0(2\sigma_t^{-2} A\rho)} A e^{j\alpha} \\
&= \frac{I_1(2\sigma_t^{-2} A|z|)}{I_0(2\sigma_t^{-2} A|z|)} A \frac{z}{|z|}.
\end{aligned}
\tag{70}
$$

This completes the proof. ♦

REFERENCES

[1] J. Sala and G. Vàzquez, "Adaptive blind equalization and demodulation without channel and signal parameter extraction," in *Proc. EUSIPCO' 94,* Edinburgh, Scotland, Sept. 13–16, 1994, vol. II, pp. 720–723.

[2] ———, "A cost function for blind signal recovery based on an implicit cumulant expansion," in *Proc. IEEE Signal Processing Athos Workshop on Higher Order Statistics,* Begur, June 1995, pp. 435–438.

[3] ———, "A statistical reference criterion for adaptive filtering," in *Proc. ICASSP' 96,* Atlanta, GA, vol. III, 1996, pp. 1661–1664.

[4] R. E. Blahut, *Principles and Practice of Information Theory.* Reading, MA: Addison-Wesley, 1987.

[5] S. J. Nowlan and G. Hinton, "A soft-decision directed LMS algorithm for blind equalization," *IEEE Trans. Commun.,* vol. 41, no. 2, Feb. 1992.

[6] S. Haykin, *Adaptive Filter Theory.* Englewood Cliffs, NJ: Prentice-Hall, 1991.

[7] J. Sala, "Criterios de teoría de la información en procesado adaptativo de la señal y aplicaciones," Doctoral thesis, Dept. of Signal Theory Commun., Universitat Politécnica de Catalunya, Spain, June 1995.

[8] L. L. Scharf, *Statistical Signal Processing Detection, Estimation and Time Series Analysis.* Reading, MA: Addison-Wesley.

[9] D. N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. Commun.,* vol. COM-28, pp. 1867–1875, Nov. 1980.

[10] Y. Sato, "Two extensional applications of the zero-forcing equalization method," *IEEE Trans. Commun.,* vol. COM-23, pp. 684–687, 1975.

[11] A. Benveniste and M. Goursat, "Blind equalizers," *IEEE Trans. Commun.,* vol. COM-32, pp. 871–883, 1984.

[12] T. Adali and M. K. Sönmez, "Channel equalization with perceptrons: An information-theoretic approach," in *Proc. ICASSP'94,* vol. 3, pp. 297–300.

[13] J. Sala, "A cost function for the equalization of constant amplitude signals based on a statistical reference," in *Proc. EUSIPCO'96,* Sept. 1996, to be published.

Josep Sala-Alvarez (M'95) was born in Calls, Spain, in 1967. He received the M.Sc. and Ph.D. degrees in telecommunications engineering from the Polytechnic University of Catalonia (UPC), Barcelona, Spain, in 1991 and 1995, respectively.

During 1992, he was working at the European Space Operations Centre (ESOC) of the European Space Agency (ESA) in Darmstadt, Germany, in the area of software engineering for telemetry processing. From 1993 to late 1994, he held a grant from the Generalitat de Catalunya for the achievement of his Ph.D. degree at the Department of Signal Theory and Communications (UPC). Since late 1994, he has been assistant professor in the same department. His current research interests are in the area of signal processing and communications, with particular emphasis on array processing, adaptive filtering, and demodulation. He is also involved in the field of ASIC design for communication and signal processing applications.

Gregori Vázquez-Grau (M'88) was born in Barcelona, Spain, in 1961. He received the M.S. and Ph.D. degrees in telecommunications engineering from the Polytechnic University of Catalonia (UPC), Barcelona, Spain, in 1984 and 1988, respectively.

Since 1989, he has been Associate Professor at the Department of Signal Theory and Communications, Polytechnic University of Catalonia. His general interests are signal processing and digital communications. Specifically, he conducts research activities in adaptive filtering and equalization, array processing in mobile and wireless communications, multiuser detection, and synchronization techniques. He has been involved in many signal processing research projects for mobile, wireless, and satellite digital communications in the framework of the RACE and ACTS programs of the European Union, the European Space Agency research programs, and European industries.