# A Cost Function for Constant Amplitude Signals based on Statistical Reference††

*Josep Sala-Alvarez*
Department of Signal Theory and Communications (GPS)
Universitat Politècnica de Catalunya
c/ Gran Capità s/n, Mòdul D5
08034 Barcelona, Spain

Tel: +34-3-401 64 40 ; Fax: +34-3-401 64 47
E-mail: alvarez@gps.tsc.upc.es

**ABSTRACT**

The equalization of constant amplitude signals is considered in the scope of this paper. A criterion based on the probability density function (pdf) of the signal of interest is proposed. The objective is to derive a suitable soft-decision scheme, more robust than the classical CMA algorithm that ensures recoverability of the signal.

## 1 Introduction

Constant amplitude signals are widely used in digital communications for the equalization of phase-modulated signals. In CMA algorithms one tries to make use of the constant amplitude property of the signals of interest. The most straightforward way is to utilize a cost function based on amplitude errors. Here we propose one based on the probability density function of the signal of interest. In this way we use all aprioristical knowledge we have of the wanted distribution. We will show in terms of the error function of the adaptive algorithm that this is more robust than classical CMA.

## 2 Statement

The pdf of a constant amplitude signal is given in the following terms in the complex plane $C$,

$$p_{A_o}(x) = \frac{1}{2\pi A}\delta(|x| - A) \tag{1}$$

It is chosen to maximize the following criterion,

$$J(\mathbf{w}) = -E_Z \ln p_{A_o+N_t}(\mathbf{w}^{\mathrm{H}}\mathbf{x}) \tag{2}$$

where some uncertainty is modeled via the AWGN term $N_t = N(0, \sigma_t^2)$. We will see that this term is important for the performance of the adaptive algorithm. The subscripted expectation operator describes with respect to which random variable the expectation of the nonlinearity is realized. The final expression is given by,

$$J(\mathbf{w}) = -E_Z \ln E_{A_o} \frac{1}{\pi\sigma_t^2} e^{-1/\sigma_t^2 |z-a_o|^2} \tag{3}$$

where we see that the cost function can be expressed in terms of a non-linear average of the error. The utilization of this cost function always leads to soft-decision type schemes. An alternative cost function as described in [3] can be used. This cost function, which we reproduce here below, always leads to minimization of the Kullback-Leibler information measure,

$$J(\mathbf{w}) = -E_{A_o} \ln E_Z \frac{1}{\pi\sigma_t^2} e^{-1/\sigma_t^2 |z-a_o|^2} \tag{4}$$

Its utilization is precluded here because the target constant amplitude distribution is not discrete but continuous. The evaluation of the outer expectation, $E_{A_o}$ is thus difficult to attain as it involves the evaluation of integrals dependent on the data. For the case of discrete distributions, this can be easily undertaken as the integrals reduce to summations. Back to (3) and using the constant amplitude property of this cost function, we arrive at the final expression,

$$J(\mathbf{w}) = E_Z \left\{ \frac{1}{\sigma_t^2} \left(|z|^2 + A^2\right) - \ln \mathrm{I}_0 \left( \frac{2A}{\sigma_t^2}|z| \right) \right\} \tag{5}$$

where $\mathrm{I}_0(\cdot)$ stands for the modified Bessel function of the first kind and order zero. We can see from the power series development of $\mathbf{I}_0(\cdot)$ that the behaviour of $J$ for $|z|$ close to $A$ is similar to the following error norm,

$$J(\mathbf{w}) \approx \frac{1}{\sigma_t^2} E_Z \left(|z| - A\right)^2 \tag{6}$$

The conventional CMA cost function is denoted instead as,

$$J_{cma}(\mathbf{w}) = E_Z \left||z|^2 - A^2\right|^2 \tag{7}$$

The robustness of this algorithm can be justified when we observe its behaviour for large values of the input data modulus, $|z|$. The classical CMA function is biquadratic. Hence, its gradient behaves as the third power of the data. This may easily lead the algorithm to divergence if the coefficient vector is far from its optimum setting. When the gradient of the classical CMA algorithm is normalized, we get an error function very close in shape to that obtained from statistical reference.

## 3 Derivation

The coefficients are updated according to the gradient rule. The gradient of the cost function is calculated with respect to the Hermitian of the coefficient vector,

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

$$= -E_Z E_{A_o} \frac{e^{-1/\sigma_t^2|z-a_o|^2}(z-a_o)^* \mathbf{x}}{E_{A'_o} e^{-1/\sigma_t^2|z-a'_o|^2}} \quad (9)$$

In the following, it will be understood that by $E_Z$ we mean that the expectation is carried not only with respect to the random variable $z$, but with respect to all random variables depending on the data vector $\mathbf{x}$, such as $z = \mathbf{w}^H\mathbf{x}$. This last equation can be re-formulated in the following fashion, where the quality functions $q(\cdot,\cdot)$ are defined,

$$q(z, a_o) \triangleq \frac{e^{-1/\sigma_t^2|z-a_o|^2}}{E_{A'_o} e^{-1/\sigma_t^2|z-a'_o|^2}} \quad (10)$$

These functions give an indication of how close one value of the input sequence is to the wanted distribution. The gradient is then expressed in terms of these functions as,

$$\nabla_{\mathbf{w}^H} J = 1/\sigma_t^2 E_Z E_{A_o} q(z, a_o)(z-a_o)^* \mathbf{x} \quad (11)$$

If we define a generalized error function from the quality functions, the gradient can be expressed as follows,

$$\nabla_{\mathbf{w}^H} J = 1/\sigma_t^2 E_Z \epsilon^*(z)\mathbf{x} \quad (12)$$

The appearance of the generalized error for the CMA case appears depicted in figure (3). Using the following property, we can express (11) in a more intuitive way.

$$E_{A_o} q(z, a_o) = 1 \quad (13)$$

Then, the gradient becomes,

$$\nabla_{\mathbf{w}^H} J = 1/\sigma_t^2 E_Z \left(z - E_{A_o} a_o q(z, a_o)\right)^* \mathbf{x} \quad (14)$$

$$= 1/\sigma_t^2 E_Z \left(z - \hat{a}_o\right)^* \mathbf{x} \quad (15)$$

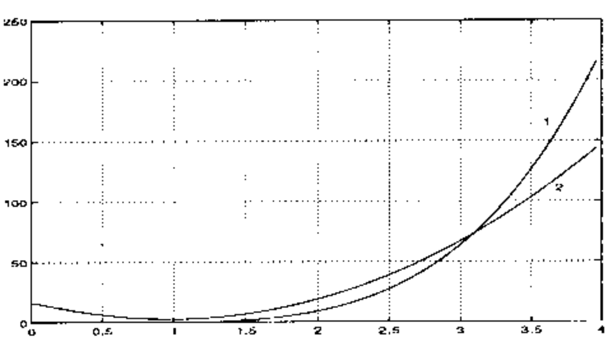


Figure 1: CMA cost functions.

where we have defined a regeneration function as,

$$\hat{a}_o = E_{A_o} a_o q(z, a_o) \quad (16)$$

For the case of constant amplitude signals, this regeneration function is expressed as,

$$\hat{a}_o = \frac{\mathrm{I}_1\left(\frac{2A}{\sigma_t^2}|z|\right)}{\mathrm{I}_0\left(\frac{2A}{\sigma_t^2}|z|\right)} A \frac{z}{|z|} \quad (17)$$

with $\mathrm{I}_1(\cdot)$ the modified Bessel function of the first kind and order 1. Note that the regeneration function only depends on the modulus of the input data, $|z|$, which stems from the symmetry of the wanted distribution. The behaviour of this cost function is such as depicted in figure (1). It preserves the phase of the incoming vector, modifying only its modulus. Small values of $|z|$ with respect to A lead the regeneration function close to zero. Saturation is observed for those values larger than the reference amplitude $A$. It is useful now to compare the behaviour of the regeneration function with respect to that of the generalized error. There exists one value of the amplitude of $z$ for which the generalized error goes to zero with positive slope. Around this point $\alpha_o < A$, the adaptive algorithm is at equilibrium. Note that the slope of the generalized error is *linear* in the modulus of $z$, while that of classical CMA is *cubic*. This causes that classical CMA be more sensitive. A smaller step-size must be used to keepclassical CMA from diverging, which is reflected in a slower convergence rate. A behaviour similar to that of the generalized error can be obtained when the step-size is normalized to make the modified gradient linear with the modulus of the input signal.

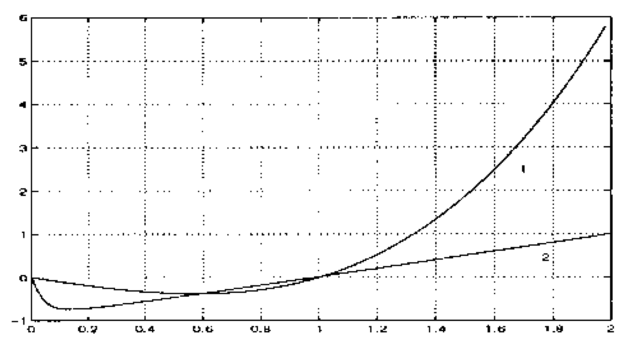


Figure 2: Generalized error functions.

## 4 Adaptive Algorithm

The adaptive algorithm is based on gradient techniques. The expectation operator $E_Z$ is dropped so that the averaging is done implicitly in the coefficient update equations, if the step-size is small enough. The upgrade equations are defined as,

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \mu \epsilon^*(z)\mathbf{x} \quad (18)$$

$$= \mathbf{w}_k + \mu \mathbf{x} \left( |z| - \frac{\mathrm{I}_1\left(\frac{2A}{\sigma_t^2}|z|\right)}{\mathrm{I}_0\left(\frac{2A}{\sigma_t^2}|z|\right)} \right) \frac{z^*}{|z|} \quad (19)$$

The choice of the appropriate $\sigma_t^2$ parameter has important consequences on the convergence rate. To guarantee acquisition, the tentative variance parameter must not be chosen too small, so that the gradient does not deliver zero values when the values of the data amplitudes are far from the wanted amplitude $A$. It is important that the value of the tentative variance be approximately matched to the noise variance at the system output when convergence has been reached. It is also possible to use different variances for acquisition and tracking, but in general, we have found that reasonable choices of $\sigma_t^2$ already guarantee fast convergence and good performance in tracking. The generalized error functions corresponding to several values of the tentative variance have been represented in figure (3). Note that the slope of the error function deviates progressively from linearity in the lower amplitude range as the value of $\sigma_t^2$ increases. Also, the zero cross-point is shifted leftwards. This last effect is known as constellation shrinkage. The more linear the error function is, the more reliability we place on the data. In the statement of the classical CMA cost function, its associated error is not in a linear relationship with the amplitude error. This causes the usual problems of divergence for dissimilarities between the actual and the target amplitude. Modifications of the error function to make it more linear will always improve the behaviour of the algorithm.

## 5 Annex. Derivation of the cost function

In this annex, we will calculate the closed expression for the cost function in the case of constant amplitude signals. Let us consider the argument of the natural logarithm (the pdf) and operate with the exponent of the Gaussian.

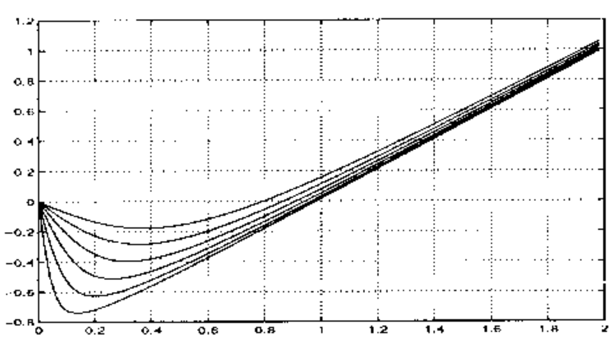


Figure 3: Generalized error functions for several $\sigma_t^2$'s.

$$E_{A_o} \frac{1}{\pi\sigma_t^2} e^{-1/\sigma_t^2|z-a_o|^2} = \frac{1}{\pi\sigma_t^2} e^{-1/\sigma_t^2\left(|z|^2+A^2\right)} \, . \quad (20)$$

$$E_{A_o} e^{2/\sigma_t^2 \mathrm{Re}(z^* a_o)} = \frac{1}{\pi\sigma_t^2} e^{-1/\sigma_t^2\left(|z|^2+A^2\right)} A(z) \quad (21)$$

From now on we will define the variable $x = 2A/\sigma_t^2$. Placing the expectation operator in terms of the pdf of $A_o$, we get,

$$A(x) = \int_{-\pi}^{\pi} e^{x\cos\theta} d\theta \quad (22)$$

whose first and second derivatives are expressed as,

$$\dot{A}(x) = \int_{-\pi}^{\pi} \cos\theta e^{x\cos\theta} d\theta, \ddot{A}(x) = \int_{-\pi}^{\pi} \cos^2\theta e^{x\cos\theta} d\theta, \quad (23)$$

Integrating the first derivative by parts, it is easy to show that,

$$\dot{A}(x) = x \int_{-\pi}^{\pi} \sin^2\theta e^{x\cos\theta} d\theta \quad (24)$$

Therefore, we get that $A$ fulfils the following differential equation,

$$\ddot{A}(x) + 1/x\dot{A}(x) = \int_{-\pi}^{\pi} \left(\sin^2\theta + \cos^2\theta\right) e^{x\cos\theta} d\theta = A(x) \quad (25)$$

Compared with the Bessel modified differential equation,

$$x^2\ddot{y} + x\dot{y} - (x^2 + n^2)y = 0 \quad (26)$$

$$y = c_1 I_n(x) + c_2 K_n(x) \quad (27)$$

where $I_n(x)$ and $K_n(x)$ stand for the Bessel modified functions of the first and second kind, respectively. If we particularize for $n = 0$ and divide by $x^2$ on both sides, we obtain (25). It only remains now to calculate what the constants of the linear combination are. We know that $A(0) = 2\pi$. Hence, given that $K_0(0)$ tends to $\infty$ and that $\mathrm{I}_0(0) = 1$, we must have that $c_1 = 2\pi$ and that $c_2 = 0$. Substituting the value of x, the cost function is equal to that in equation (5).

## 6 Results

In the following figures we compare the constellation obtained with the CMA algorithm and the Statistical Reference algorithm. The classical CMA algorithm usually shows unstability associated with it when the amplitude at the output of the equalizer, in the initial stages of acquisition is very dissimilar from the target amplitude. Very small step-size must be used to guarantee that the algorithm will not diverge. This is due to the fact that the gradient of the cost function is not linear with the true amplitude error. It behaves instead as the third power of the amplitude. Although this makes the CMA algorithm stable, its convergence rate is comparatively very large. In the Statistical Reference algorithm instead, comparatively larger step sizes can be used without making the algorithm stable. The convergence rate is thus much faster. The gradient provides here 'true' undistorted estimates of the amplitude error. In many of the simulations we have carried out, it has not been possible to make the convergence rate of classical CMA equal that of statistical reference for large amplitude dissimilarities with respect to the target amplitude. Simulations with the Statistical Reference Algorithm are shown in figures (4) and (5). The following values have been chosen for the simulations: SNR $= 14dB$, $\mu = 0.0007$, number of coefficients of the fractionally-spaced equalizer $N = 30$ (around 8 symbols), channel response =[-0.1274,0.5542,-1.0973,-0.7313,1.4047,-0.6202,0.2371,-1.5868,-0.4015,-0.7707].

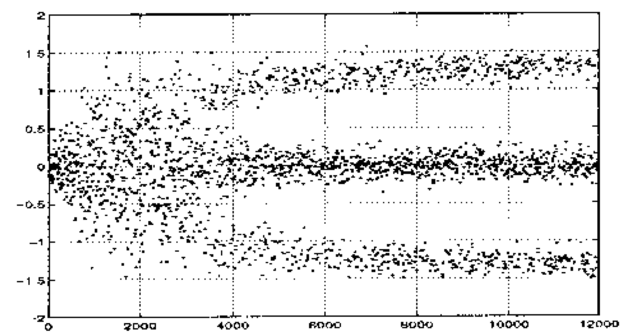

Figure 4: Statistical Reference. Evolution of the in-phase channel

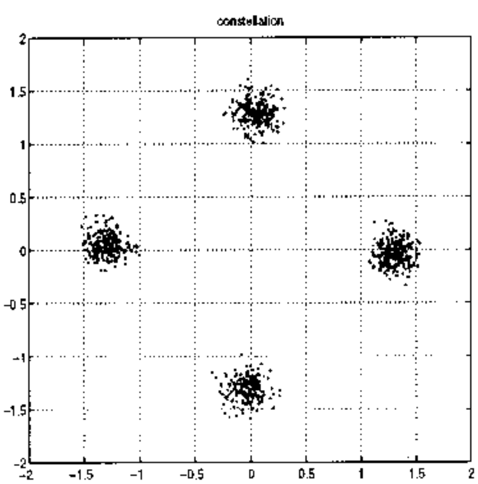


Figure 5: Statistical Reference. Output constellation.

The convergence region of CMA is limited. If the amplitude of the signal is very dissimilar from the target amplitude, power normalization must be carried out to ensure convergence. Due to the special characteristic of statistical reference CMA, this is not necessary. If actual amplitude do not differ much and are close to the target amplitude, performance of both algorithm is also similar. A slightly higher misadjustment can be observed in CMA due to the non-linear characteristic of the error under the same test conditions for both algorithms.

## References

[1] Sala, J. 'Criterios de Teoria de la Informacion en Procesado Adaptativo de la Senal y Aplicaciones'. PhD. Thesis. Dpt. of Signal Theory and Communications. Polytechnic University of Catalonia. June 1995.

[2] Sala, J. Vazquez, G.. 'Adaptive Blind equalization and Demodulation without Channel and Signal Parameter Extraction'. Proceedings of EUSIPCO'94. Edinburgh, 13-16 September 1994.

[3] Sala, J. Vazquez, G.. 'A Statistical Reference Criterion for Adaptive Filtering'. Proceedings of the ICASSP'96, Atlanta, USA. 1996.

[4] Nowlan, Steven J.; Hinton, Geoffrey. 'A Soft-Decision Directed LMS algorithm for Blind Equalization'. IEEE Transactions on Communications. VOL. 41, No. 2, February 1992.

[5] Sala, J. Vazquez, G.'A Cost Funcion for Blind Signal Recovery based on an Implicit Cumulant Expansion'. Proceedings of the Joint IEEE Workshop on Higher Order Statistics. Begur, 1995. Spain.

[6] Blahut, Richard E. 'Principles and Practice of Information Theory'. Addison-Wesley 1987.

[7] Haykin, S.. 'Adaptive Filter Theory'. Prentice-Hall Inc. 1991.