SIGNAL PROCESSING VII: Theories and Applications
M. Holt, C. Cowan, P. Grant, W. Sandham (Eds.)

PDF erzeugen

Adaptive Blind Equalisation and Demodulation Without Channel and Signal Parameter Extraction†

Josep SALA-ALVAREZ, Gregori VÁZQUEZ-GRAU

Dpt. of Signal Theory and Communications
Universitat Politècnica de Catalunya
Apt 30002, 08071 Barcelona, Spain
Tel: +34-3-4016440
Fax: +34-3-4016447
E-mail: alvarez@tsc.upc.es / gregori@tsc.upc.es

Abstract. The problem of filtering a source of known statistics in noise and/or interference of unknown statistics is addressed in this paper. In a general framework, the optimum estimation (detection) of the desired signal requires perfect knowledge of the probability density functions of the interfering signal and usually leads to complex non-linear algorithms. Herein we have restricted ourselves to linear estimators and optimize performance within this subset according to a criterion derived from information-theoretic considerations. For validation purposes, the particular application of blind data demodulation has been considered. Synchronisation, carrier phase rotation and equalization are performed with the same algorithm to force a source of a known pdf at the output of the adaptive system. Performance at low SNR is derived through simulations.

1.Introduction·

The motivation of this paper stems from an ample variety of communications problems where a receiving station has to retrieve the transmitted data from a channel-distorted versions of the original signal.. The purpose has been to derive an adaptive algorithm that can ensure blind demodulation, without a training signal. All available statistical knowledge of the clean modulated signal may be employed. In this sense, acquisition is performed on availability of a statistical reference rather than on a time reference (training sequence). The algorithm herein developed is specially suited for those applications where at least partial knowledge of the signal statistics is available and very little is known about the characteristics of the noise or interfering signals. Provision against effects such as coloured noise, unknown channel impulse response, unbalanced I-Q channels, can be achieved by the same algorithm and a suitable architecture [1]. We have striven for its robustness and simplicity.. As a difference to cumulant-based algorithms [2], that estimate the channel from statistics of the incoming signal, this algorithm can be classified as of the Bussgang type. A suitable cost function is derived whose minimization ensures correct demodulation.

2.Signal Model

The signal model of a linearly modulated signal and related notation are presented in this section. Let us assume that a sequence of symbols a_k is modulated on a Nyquist pulse p(t) and transmitted through a channel of impulse response $h_c(t)$. The received signal is expressed by the following equality,

$$r(t)=\sum_{k=-\infty}^{k=+\infty} a_k\left(p*h_c\right)(t-kT)+n_r(t) \qquad [1]$$

where $n_r(t)$ stands for an additive Gaussian noise process. Synchronisation with the symbol sequence as well as carrier phase offset are implicit in the channel response $h_c(t)$. The demodulation architecture is defined in terms of a linear filter $h_a(t)$ working at the symbol rate on the incoming signal, such that the estimates of the symbol sequence are given by,

$$\hat{a}_k=\left(r*h_a\right)(t-kT) \qquad [2]$$

Note that a linear operation on the incoming data is sufficient to resolve stationary effects as symbol time offset or phase rotation (provided that the architecture of the linear filter $h_a(t)$ is suitably defined). Frequency Doppler can also be compensated for without inclusion of an NCO (Number Controlled Oscillator) if the Doppler uncertainty range is much smaller than the signal bandwidtth. Synchronisation can be corrected for with a simple time shift of $h_a(t)$. In a digital implementation we will assume that samples are taken at the rate $1/T_s$, such that the available signal is discrete with $r_k=r(kT_s)$. The filter h_a will be considered as a FIR of N_c coefficients. Hence, if successive snapshots of the signal are taken at the rate of once per symbol, $\mathbf{r}_k = [r(k),r(k-1),\ldots,r(k-N_c+1)]$, the received signal can be expressed as the convolution of the symbol sequence a_k with a vector sequence $\mathbf{h}_r$ of length N_c that contains the discrete snapshots of the impulse response $r*h_c$ in the format already defined for r,

· †This work has been supported by the National Research Plan of Spain CICYT, Grant Number TIC92-0800-C05-05

$$\mathbf{r}_k = a*\mathbf{h}_r + \mathbf{n}_r(k) = \sum_{i=-\infty}^{i=+\infty} a_i \mathbf{h}_r(k-i) + \mathbf{n}_r(k) \quad [3]$$

The symbol estimates are then obtained via multiplication with the coefficient vector w (of h_a),

$$\hat{a}_k = \mathbf{w}^H \mathbf{r}_k = a_k \mathbf{w}^H \mathbf{h}_r(0) + \sum_{i \neq 0} a_i \mathbf{w}^H \mathbf{h}_r(k-i) + n_a(k) \quad [4]$$

The objective of the filter w is to minimise the power of the interfering terms, Gaussian and convolutive noise, to minimise the bit error rate on a_k. It would be desired that **w** were orthogonal to all vectors $\{\mathbf{h}_r(i), i \neq 0\}$ although this seldom happens as $\mathbf{h}_r$ is usually infinite response. The coefficients of w are updated to minimise a suitable cost function to be derived in the next section

3.Derivation of the Cost Function

Assuming perfect knowledge of all relevant model parameters, we seek to derive the optimal estimator of a_k within the subset of all possible *linear* estimators. For the model presented in the introduction, the estimator should (approximately) converge into the subspace orthogonal to the span of $\{\mathbf{h}_r(i), i \neq 0\}$. As the optimization criterion, we have chosen the minimization of a cost function akin to the discrimination between the pdf of a_k and that of its estimate $\hat{a}_k$. Let the Kullback-Leibler information measure (or minus cross-entropy) [3] between two probability density functions, p_X and p_Y, be expressed as,

$$D_{X|Y} = \int p_X(x) \ln \frac{p_X(x)}{p_Y(x)} dx \quad \geq \quad 0 \quad [5]$$

We wish that the statistics of the estimate be as close as possible to the prior pdf of a_k . Let A denote the random variable at the output of the estimator for the optimal case when all parameters of interest (**a**, pdf of $\mathbf{n}_k$) are available. Also let $\hat{A}$ denote the output of the estimating filter w. Therefore, minimization of the discrimination between the real pdf of $\hat{A}$ and the desired pdf of A is enforced.

$$\mathbf{w} = \arg\min\left\{E_{\hat{a}} \ln p_{\hat{A}}(\hat{a}) - E_{\hat{a}} \ln p_A(\hat{a})\right\},\ \hat{a} = \mathbf{w}^H \mathbf{r} \quad [6]$$

The stumbling block of this approach is that the pdf of $\hat{A}$ appears explicitly in the first term of the discrimination, which equals the minus entropy of the symbol estimates. Its estimation is not straightforward (some procedures are referenced in [2] based on high order statistics of the process and hence difficult to obtain). The minimum attainable output noise variance is also unknown. Therefore we propose instead minimisation of the following cost function, obtained as the second summand in eq.6,

$$\mathbf{w}, \sigma_a = \arg\min\left\{-E_{\hat{a}} \ln p_{A_o+N(\sigma_a)}(\hat{a})\right\} \quad [7]$$

where now minimisation has been also extended to the parameters characterising unknown additive noise affecting the output smbols. The adaptive algorithm implementing this criterion steers the output statistics through **w** and controls the shape of the non-linearity through σ. The output pdf is modeled as that of $A_o+N(\sigma)$, with A_o a random variable wiht the same pdf of the symbols and N(σ) additive noise parametrised by its variance σ (or maybe some other paramer for non-Gaussian distributions, see Section 4.1). Intuitively, when the distribution of â is such that its most likely values coincide with the maxima of the pdf of $A_o+N(\sigma)$, the expectation of the non-linearity in eq.7 reaches its minimum value. Although the parametrization of the objective pdf may not be exact (due to interference or convolutional noise), this effect renders a robust cost function for an adaptive algorithm. Application of this cost function also minimises noise variance at the output of the adaptive filter. Additive noise on a random process X(n) manifests itself as a broadening of the peaks of its pdf. If the maxima of p_X coincide with the (convex) peaks of the non-linearity, this broadening should be minimised so as to gather most samples of X(n) around the peaks of the objective pdf. Thus, the expectation in eq.7 is minimised.

In general, the proposed cost function will be multimodal depending on the involved statistics. The criterion outlined in eq.7 should be understood as one of absolute minimization. When the joint minimisation in eq.7 is carried out, it does not necessarily imply that the distribution of $\hat{A}$ equals that of A although it is similar. In practice we have observed this distributional bias in high noise situations as constellation shrinkage. A rigorous proof falls out of the scope of this article. Nevertheless, if the pdf of $\hat{A}$ is kept fixed, partial minimisation with respect to σ can cause that $p_{\hat{A}} \rightarrow p_A$ (provided that the parametrisation of A is exact enough), as endorsed by the cross-entropy (eq.5).

3.1 Relationship with DFE

We will show in this section that the proposed cost function is equivalent to DFE algorithms in the medium to high range of signal-to-noise ratios and in the tracking regime. To that purpose let us consider the distribution of a discrete constant envelope modulation as given in eq.12 with the constraint that $|a_i|^2=1$ for all symbols. Let us also assume that SNR is high: $|n_k|^2 \ll 1$, with n_k addtive noise and that we are already in tracking regime. Then, when symbol a_i is present, all Gaussian exponentials in eq.12 except that corresponding to the actual symbol are vanishingly small (depending on the reliability assigned by the tentative variance σ). The criterion can then be approximated with most likelihood to a least squares of the type,

$$-E_{\hat{a}} \ln p_{A_o+N(\sigma)}(\hat{a}) \approx E_{\hat{a}} \frac{1}{\sigma^2}\left|\hat{a} - a_i(\hat{a})\right|^2 \quad [8]$$

save additive constants, with $a_i(\hat{a})$ the closest symbol to the estimate â. It has been thus proven that the proposed criterion is equivalent to a Decision Directed (DD) algorithm for high SNR. For low SNR, other symbols than the closest also influence the cost function, for the reliability of the decision is not so high. After some algebra, the criterion in eq.7 can also be shown to equal the CMA [4] algorithm for constant amplitude modulations.

4.Adaptive Algorithm

The coefficients of the adaptive algorithm will be updated in terms of the gradient of the cost function defined in Section 3. Note here that the gradient will also have to be calculated with respect to the parametrisation of the noise (see eq.7). Nevertheless, as the expectation operator cannot be realised in practice, we will take a finite sample estimate, such that the non-linearity evaluated at the output of the matched filters is given by the following expression,

[9]

$$J = -N_{av}^{-1}\sum_{k=1}^{k=N_{av}} \ln p_{A_o+N(\sigma_a)}\left(\mathbf{w}^H\mathbf{r}_k\right) = N_{av}^{-1}\sum_k J_k$$

When $N_J>1$, J turns out to be a memory non-linearity. When the gradient with respect to the coefficients is calculated we have that,

[10]

$$\nabla_{\mathbf{w}^H} J = -N_{av}^{-1}\sum_k e^{J_k}\dot{p}_{A_o+N(\sigma_a)}\left(\hat{a}_k\right)\mathbf{r}_k$$
$$= -N_{av}^{-1}\sum_k \varepsilon\left(\hat{a}_k\right)\mathbf{r}_k$$

Where a generalized error $\varepsilon(\hat{a}_k)$ is introduced in terms of the target distribution. Note though that this is not the actual error with respect to the transmitted symbols. From this expression, we can see that each of the terms in the summation is weighted by the exponential of J_k. Therefore, when the output value $\hat{a}_k$ has a low likelihood with respect to the tentative pdf, the exponential term is very high. In this way, when we are far from convergence, the gradient tends to accelerate. As we approach convergence such that the cost function decreases, the weighting terms become less and less important, stabilizing the algorithm. The gradient with respect to the noise parameter is given by,

[11]

$$\nabla_\sigma J = -N_{av}^{-1}\sum_k e^{J_k}\frac{d}{d\sigma}p_{A_o+N(\sigma_a)}\left(\hat{a}_k\right)$$

For the particular case of linearly modulated signals, the objective pdf is given by the convolution of the discrete constellation pdf and the model of the underlying noise distribution. Assuming a Gaussian distribution we are led to,

[12]

$$p_{A_o+N(\sigma_a)}\left(\hat{a}_k\right) = \left(\pi\sigma^2 M_0\right)^{-1}\sum_{i=1}^{i=M_0} e^{-\left|\hat{a}_k-a_i\right|^2/\sigma^2}$$

with M_0 the number of modulation levels. The error $\varepsilon(\hat{a}_k)$ with respect to the coefficients will then be expressed in terms of the following non-linearities,

[13]

$$q_i\left(\hat{a}_k\right) = e^{-\left|\hat{a}_k-a_i\right|^2/\sigma^2}\Big/\sum_{i'=1}^{i'=M_0} e^{-\left|\hat{a}_k-a_{i'}\right|^2/\sigma^2}$$

as,

[14]

$$\varepsilon\left(\hat{a}_k\right) = \sum_{i=1}^{i=M_0} q_i\left(\hat{a}_k\right)\left(\hat{a}_k - a_i\right)$$

Note that the set of non-linearities $\{q_i, 1\leq i\leq M_0\}$ play the role of a measure of the likelihood that $\hat{a}_k$ contains the symbol a_i. They are estimates of the conditional probabilities $p(a_i|\hat{a}_k)$ parametrised by the tentative variance σ. If σ is taken to be the actual noise variance, we obtain by application of the Bayes conditional probability rule that $q_i(\hat{a}_k)=p(a_i|\hat{a}_k)$. Moreover, it holds that $\Sigma_{i=1..M_0}\, q_i(\hat{a}_k)=1$ as was to be expected. Thus, the error defined in [4.6] appears as a weighted average of the error between the output sample $\hat{a}_k$ and all symbols in the constellation. Symbols farther apart than the closest neighbours will have a lesser influence on $\varepsilon(\hat{a}_k)$ as parametrised by σ.

In its turn, the derivative with respect to the tentative variance is expressed as,

[15]

$$\frac{d}{d\sigma^2}J_k = -\frac{1}{\sigma^2}\left(1-\frac{1}{\sigma^2}\sum_{i=1}^{i=M_0} q_i\left(\hat{a}_k\right)\left|\hat{a}_k-a_i\right|^2\right)$$
$$= -\frac{1}{\sigma^2}\left(1-\frac{\hat{\sigma}_n^2}{\sigma^2}\right)$$

This gradient can also be associated to a measurement of the error between the tentative variance σ and the actual noise variance σ_n. Just note that a weighted average is performed of the squared errors of the symbol estimate $\hat{a}k$ with respect to the constellation symbols.

A common pitfall when adaptively changing the cost function itself (as we are doing through adaptation of the tentative variance) is encountered when the filter coefficients are not updated with an equivalent step-size normalized to the curvature of the new cost function. The consequence of this is higher misadjustment noise and possibly divergence as a higher curvature must be offset with a smaller step-size. Then the gradient with respect to the coefficients should be normalised by $1/\sigma^2$, which is approximately the curvature of J evaluated at the peaks. The tentative variance has been bounded from below: $\sigma^2 = \sigma_0^2 + \Delta^2$, in order to preclude that the cost function migrates to an approximately Gaussian distribution during adaptation. This would recover noise instead of data at the equalizer output. Δ becomes then the adaptive parameter. The coefficient update equation results,

[16]

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \mu_w\sigma_k^2\nabla_{\mathbf{w}^H}J \quad , \quad \sigma_k^2 = \sigma_0^2 + \Delta_k^2$$

4.1 Inclusion of Other Distributions

Up till now a Gaussian model has been assumed for the distribution of the additive noise $N(\sigma)$. Nevertheless, other models can also be utilised. In fact, the algorithm is quite robust as concerns the choice of the noise pdf provided that the model pdf and the actual one display similar shapes. We put forward here utilisation of a Laplacian noise distribution to take into account possible interference not accurately modeled by a Gaussian (longer tails). The real Laplacian distribution for the noise $N(\lambda)$ is given by,

[17]

$$p_{N(\lambda)}(n) = (2\lambda)^{-1}e^{-|n|/\lambda}$$

In this way, the criterion in high noise is found equivalent to minimisation of the absolute error between the output sample $\hat{a}_k$ and the closest symbol $a_i(\hat{a})$,

$$-E_{\hat{a}} \ln p_{A_o+N(\lambda)}(\hat{a}) \approx E_{\hat{a}} \frac{1}{\lambda}\{|\mathrm{Re}\,\delta|+|\mathrm{Im}\,\delta|\} \ , \ \delta = \hat{a} - a_i(\hat{a}) \qquad [18]$$

save additive constants. The expressions for the gradient are calculated as in eq.10 and eq.11.

5.Simulations

Blind demodulation of a QPSK signal filtered by a mixed-phase channel has been considered for the simulations. Performance of the algorithm averaged over a set of realizations in a wide range of SNR is depicted in figure.1. (the term N_{av} in eq.9 has been chosen as 1). It displays the inverse of the squared tentative variance parameter (a Gaussian model has been assumed for hte noise) versus time which can also be interpreted as the evolution of the EbNo before the decision device. It appears that the range of acquisition falls in the order of one to two thousand symbols depending on the input SNR. The step size for the adaptation of the coefficients has been chosen as 0.005 while that of the tentative variance is 0.05. In general the tentative variance accepts quite high adaptation rates without danger of divergence. The acquisition curve of the tentative variance is usually very steep during convergence. The smoothed behaviour observed in figure 1 is due rather to statistical averaging of several realizations (the elbow does not always occur at the same instant) than to the actual behaviour of a single realization. A realization of the coefficients error with respect to the optimum equalizer is depicted in figure 3.

6.Conclusions

An algorithm capable of forcing a signal of an a priori known pdf at the output of an adaptive system has been presented. It has also been shown that it is able to minimise (unknown) noise variance at the output. The rate of blind acquisition is rather fast as compared with other methods. Some research is currently being done to solve the capture problem of the algorithm to distributions that are similar to the a priori template.

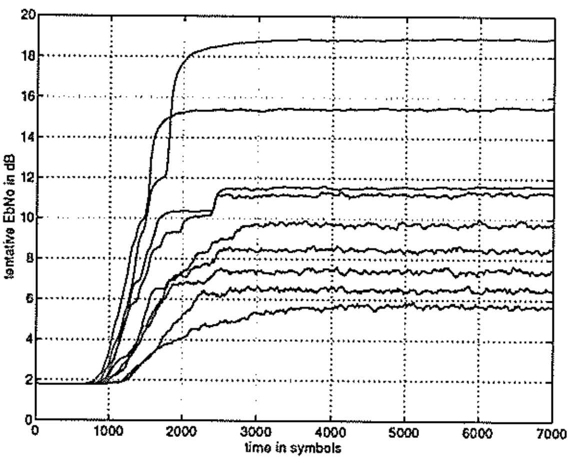

Figure 1. Evolution of the inverse of the squared tentative noise variance before detection or estimated EbNo for several input SNR's (dB): 16.90, 13.37, 10.88, 8.94, 7.35, 6.02, 4.86, 3.83, 2.92.

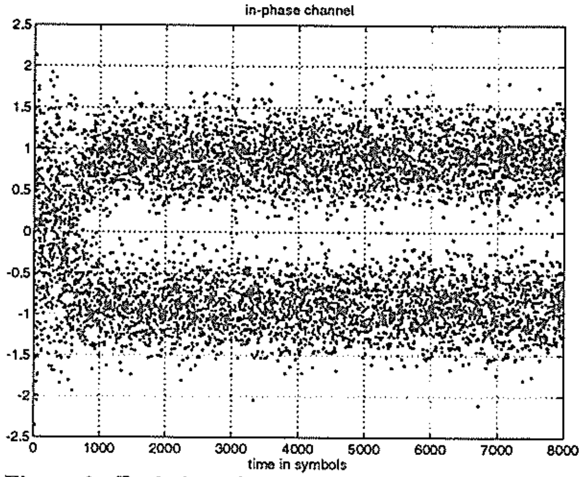

Figure 2.: Evolution of the in-phase channel for an input SNR of 4.86 dB (see figure 1 for output EbNo).

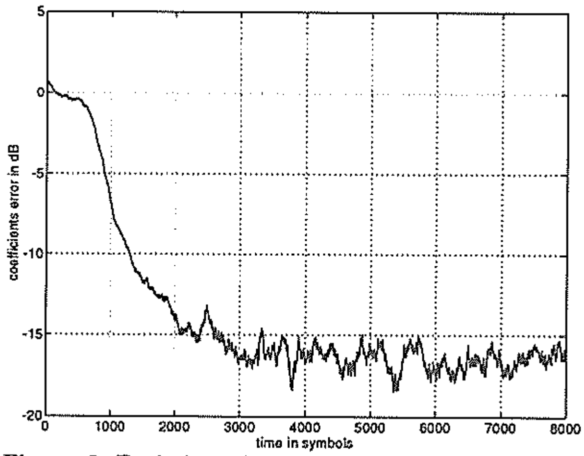

Figure 3. Evolution of the equalizer coefficients error with respect to the optimum filter at an input SNR of 4.86 dB (see figure 1 for output EbNo).

References

[1] Josep Sala, Gregori Vázquez, "Global Beamforming and Demodulation for Linearly Modulated Signals", Seventh SP Workshop on Statistical Signal and Array Processing. Quebec, 1994.

[2] Pierre Common, Jean-Louis Lacoume, "Statistiques d'ordres supérieurs pour le traitement du signal". Fascicule III, Les Houches, école pré-doctorale de physique. August 28th-September 10th, 1993.

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

[4] Simon Haykin, "Adaptive Filter Theory", Prentice-Hall International Editions. 1991