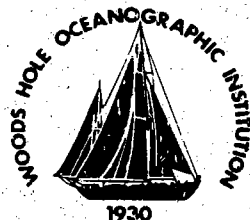


WHOI-90-08

COPY 2

Woods Hole Oceanographic Institution



Universal Criteria for Blind Deconvolution

by

Ofir Shalvi & Ehud Weinstein

February 1990

Technical Report

Funding was provided by the Office of Naval Research
under Grant Number N00014-90-J-1109.

Approved for public release; distribution unlimited.

DOCUMENT
LIBRARY
Woods Hole Oceanographic
Institution

WHOI-90-08

Universal Criteria for Blind Deconvolution

by

Ofir Shalvi and Ehud Weinstein

Woods Hole Oceanographic Institution
Woods Hole, Massachusetts 02543

February 1990

Technical Report



Funding was provided by the Office of Naval Research
under Grant Number N00014-90-J-1109.

Reproduction in whole or in part is permitted for any purpose of the
United States Government. This report should be cited as:
Woods Hole Oceanog. Inst. Tech. Rept., WHOI-90-08.

Approved for publication; distribution unlimited.

Approved for Distribution:

A handwritten signature in cursive script, which appears to read 'William J. Plant', is written over a horizontal line.

A.J. Williams 3rd, Chairman
Department of Applied Ocean Physics
and Engineering

Universal Criteria For Blind Deconvolution

by

Ofir Shalvi and Ehud Weinstein*

Department of Electronic Systems
Faculty of Engineering
Tel-Aviv University
Ramat-Aviv, Tel-Aviv 69978

* Ehud Weinstein is also an adjunct scientist at the Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543.

Abstract

We present necessary and sufficient conditions for blind equalization/deconvolution (without observing the input) of an unknown, possibly non-minimum phase linear time invariant system (channel). Based on that, we propose a family of optimization criteria and prove that their solution correspond to the desired response. These criteria, and the associated gradient-search algorithms, involve the computation of high order cumulants. The proposed criteria are universal in the sense that they do not impose any restrictions on the probability distribution of the input symbols. We also address the problem of additive noise in the system and show that in several important cases, e.g. when the additive noise is Gaussian, the proposed criteria are unaffected.

I. Introduction

Inverse filtering, or blind equalization/deconvolution, is a problem of considerable practical interest in diverse fields including seismology, radio astronomy, underwater acoustic telemetry, and data communication. The problem is illustrated in Figure 1. We observe the output y_n of an unknown possibly non-minimum phase linear time-invariant system H with input a_n being a sequence of independent identically distributed (i.i.d.) random variables with a prespecified probability distribution. We want to recover the input sequence, or equivalently, to identify the inverse H^{-1} of H using a tap-delay line C . This, in turn, requires the identification of both the magnitude and the phase of the unknown system's transfer function. The magnitude can be identified using second order moments of the output signal. However, phase identification requires the calculation of higher order moments/cumulants.

The paper by Sato [1] and Godard [2] approach the problem of blind equalization by introducing new criteria, different from the mean square error (m.s.e.) criterion used for trained equalizers, and then apply gradient-search algorithms to optimize the selected criteria. Sato's method and Godard's method are further analyzed by Benveniste and Goursat [3] and by Foschini [4], respectively, and the conditions necessary to ensure the convergence of the respective algorithms are specified.

The paper by Benveniste, Goursat, and Ruget [5] presents several concepts and results that significantly contributed to the understanding of the problem. First, it has been established that a criterion based on second order statistics, e.g. the m.s.e. criterion, is insufficient for phase identification. For that reason, the problem cannot be solved when the probability distribution of the input symbols is Gaussian since the second order moments completely specify the input-output statistics. Next, it has been proven that a sufficient condition for equalization is that the probability distribution of the output (recovered) symbols z_i be equal the probability distribution of the input symbols a_i . This principle is then used to formulate a general class of criteria that

converge to the desired system response under the assumption that the probability distribution of the input symbols belong to a certain class of continuous-type distributions. This is, however, a rather restrictive condition; for example, in digital communications, the input distribution is inevitably of discrete type.

The paper by Shalvi and Weinstein [6] proves that it is sufficient to equalize the second and fourth order cumulants of the input and the output probability distributions. Based on that, new criteria are presented, that require only partial knowledge of the input distribution. Godard's criterion is shown to be a special case of these criteria. An important feature of the proposed criteria is that their maximization correspond to the desired response, and they do not suffer from unwanted local maxima. Therefore, the associated gradient-search algorithm is expected to converge to the desired response regardless of initialization.

In this paper we extend the results of [6]. First, we present necessary and sufficient conditions for equalization based on high order cumulants of the corresponding probability distributions. We then propose a general class of criteria, prove that their optimization must yield the desired solution, and present the associated gradient-based algorithms. These criteria are universal in the sense that they do not impose any restrictions on the probability distribution of the input symbols. We briefly address the problem of additive noise in the system and show that in some important cases, e.g. when the additive noise is Gaussian, the proposed criteria are unaffected.

II. Mathematical Preliminaries

Let x_1, x_2, \dots, x_n be a set of real/complex random variables with the joint characteristic function

$$\Phi(w_1, w_2, \dots, w_n) = E \left\{ e^{j \sum_{i=1}^n w_i x_i} \right\}$$

where $j = \sqrt{-1}$, and $E\{\cdot\}$ stand for the expectation operation.

The joint cumulants are defined by:

$$\begin{aligned} \text{cum}(x_1 : p_1; x_2 : p_2; \dots; x_n : p_n) &= \\ &= (-j)^p \frac{\partial^p \ln \Phi(w_1, w_2, \dots, w_n)}{\partial w_1^{p_1} \partial w_2^{p_2} \dots \partial w_n^{p_n}} \Big|_{\underline{\omega}=\underline{0}} \end{aligned}$$

where p_i are non-negative integers, and $p = \sum_{i=1}^n p_i$.

For notational convenience, if $p_i = 1$ we do not write it in, that is:

$$\text{cum}(\dots; x_i : 1; \dots) \equiv \text{cum}(\dots; x_i; \dots)$$

The following properties can be verified:

(p.1)

$$\begin{aligned} \text{cum}(x_1 : p_1; x_2 : p_2; \dots; x_n : p_n) &= \\ &= \text{cum} \left(\underbrace{x_1; x_1; \dots; x_1}_{p_1 \text{ times}}; \underbrace{x_2; x_2; \dots; x_2}_{p_2 \text{ times}}; \dots; \underbrace{x_n; x_n; \dots; x_n}_{p_n \text{ times}} \right) \end{aligned}$$

(p.2) If

$$y_i = \sum_{k_i=1}^n b_{k_i} x_{k_i} \quad i = 1, 2, \dots, m$$

Then

$$\text{cum}(y_1; y_2; \dots; y_m) = \sum_{k_1=1}^n \sum_{k_2=1}^n \dots \sum_{k_m=1}^n b_{k_1} b_{k_2} \dots b_{k_m} \text{cum}(x_{k_1}; x_{k_2}; \dots; x_{k_m})$$

(p.3) If x_i and x_j are statistically independent and $p_i, p_j > 0$, then :

$$\text{cum}(x_1 : p_1; x_2 : p_2; \dots; x_n : p_n) = 0$$

(p.4) If x_1, x_2, \dots, x_n are jointly Gaussian, then:

$$\text{cum}(x_1 : p_1; x_2 : p_2; \dots; x_n : p_n) = 0$$

whenever $p = \sum_{i=1}^n p_i > 2$.

III. Development of the Criteria

The basic system of interest is illustrated in Figure 1. We shall make the following assumptions:

- (i) The input sequence a_i consists of real/complex i.i.d. random variables with a non-Gaussian but otherwise completely arbitrary continuous/discrete probability distribution. We shall assume the existence of certain cumulants of a_i .
- (ii) The unknown system (channel) $H = \{h_i\}$ is a possibly non-minimum phase linear time invariant filter whose transfer function has no zeros on the unit circle, that is:

$$H(\omega) = \sum_i h_i e^{j\omega i} \neq 0 \quad 0 < \omega \leq 2\pi \quad (1)$$

- (iii) The equalizer $C = \{c_i\}$ is a tap-delay line of sufficient length so that truncation effects can be ignored.

Let $S = \{s_i\}$ denote the combined channel-equalizer response, that is the convolution of H with C (see Figure 1). Thus

$$s_i = h_i \circ c_i = \sum_l c_l h_{i-l} \quad (2)$$

and

$$z_i = a_i \circ s_i = \sum_l s_l a_{i-l} \quad (3)$$

Invoking properties (p.1) and (p.2)

$$\begin{aligned} \text{cum}(z_i : p; z_i^* : q) &= \text{cum} \left(\underbrace{z_i; z_i; \dots; z_i}_{p \text{ times}}; \underbrace{z_i^*; z_i^*; \dots; z_i^*}_{q \text{ times}} \right) \\ &= \text{cum} \left(\sum_{l_1} s_{l_1} a_{i-l_1}; \sum_{l_2} s_{l_2} a_{i-l_2}; \dots; \sum_{l_p} s_{l_p} a_{i-l_p}; \right. \\ &\quad \left. \sum_{k_1} s_{k_1}^* a_{i-k_1}^*; \sum_{k_2} s_{k_2}^* a_{i-k_2}^*; \dots; \sum_{k_q} s_{k_q}^* a_{i-k_q}^* \right) \\ &= \sum_{l_1} \sum_{l_2} \dots \sum_{l_p} \sum_{k_1} \sum_{k_2} \dots \sum_{k_q} s_{l_1} s_{l_2} \dots s_{l_p} s_{k_1}^* s_{k_2}^* \dots s_{k_q}^* \text{cum} (a_{i-l_1}; a_{i-l_2}; \dots \\ &\quad \dots; a_{i-l_p}; a_{i-k_1}^*; a_{i-k_2}^*; \dots; a_{i-k_q}^*) \end{aligned} \quad (4)$$

Invoking properties (p.3) and (p.1)

$$\begin{aligned} & \text{cum} \left(a_{i-l_1}; a_{i-l_2}; \dots; a_{i-l_p}; a_{i-k_1}^*; a_{i-k_2}^*; \dots; a_{i-k_q}^* \right) = \\ & = \begin{cases} \text{cum}(a_i : p; a_i^* : q) & l_1 = l_2 = \dots = l_p = k_1 = k_2 = \dots = k_q \\ 0 & \text{otherwise} \end{cases} \end{aligned} \quad (5)$$

Substituting (5) into (4)

$$\text{cum}(z_i : p; z_i^* : q) = \left(\sum_l s_l^p s_l^{*q} \right) \text{cum}(a_i : p; a_i^* : q) \quad (6)$$

This equation, relating the cumulants of the output symbols z_i to the cumulants of the input symbols a_i , form the basis to the subsequent development. We shall distinguish between the complex case in which a_i and/or s_i are complex valued sequences and the real case in which both sequences are real valued.

Theorem 1. (complex case)

Let a_i be a sequence of real/complex i.i.d. random variables with some given probability distribution. Suppose that $\text{cum}(a_i; a_i^*) > 0$ and $\text{cum}(a_i : p; a_i^* : q) \neq 0$, where p and q are some non-negative integers such that $p + q > 2$, exist. Let s_i be a deterministic (possibly infinite) real/complex valued sequence. Let the random variables z_i be specified by (3).

If

$$\text{cum}(z_i; z_i^*) = \text{cum}(a_i; a_i^*) \quad (7)$$

Then:

$$|\text{cum}(z_i : p; z_i^* : q)| \leq |\text{cum}(a_i : p; a_i^* : q)| \quad (8)$$

where equality holds if and only if

$$s_l = e^{j\phi} \delta_{l-k} = \begin{cases} e^{j\phi} & l = k \\ 0 & l \neq k \end{cases} \quad (9)$$

for some fixed integer k and $\phi \in \mathcal{R}^1$.

Proof: Invoking (6) for $p = q = 1$, if

$$\sum_l |s_l|^2 = \frac{\text{cum}(z_i; z_i^*)}{\text{cum}(a_i; a_i^*)} = 1$$

Then

$$|s_l|^2 \leq 1 \quad \forall l$$

Thus

$$|s_l| < 1 \quad \forall l$$

Hence, for $p + q > 2$,

$$|s_l|^{p+q} \leq |s_l|^2 \quad \forall l$$

Thus, invoking (6),

$$\begin{aligned} \frac{|\text{cum}(z_i; p; z_i^*; q)|}{|\text{cum}(a_i; p; a_i^*; q)|} &= \left| \sum_l s_l^p s_l^{*q} \right| \\ &\leq \sum_l |s_l|^{p+q} \leq \sum_l |s_l|^2 = 1 \end{aligned}$$

where equality holds if and only if s_l satisfies (9) for some fixed integer k and $\phi \in \mathcal{R}^1$. Q.E.D.

Note that the condition $\text{cum}(a_i; a_i^*) = E\{|a_i - E\{a_i\}|^2\} > 0$ implies that a_i is a non-trivial random variable with a non-zero variance. The assumption $\text{cum}(a_i; p; a_i^*; q) \neq 0$ where $p + q > 2$ implies that the probability distribution of the input symbols is non-Gaussian (see property (p.4)).

Corollary 1.1

Under the assumptions of theorem 1.1,

$$z_i = e^{j\phi} a_{i-k} \quad (\text{w.p.1.}) \quad (10)$$

for some fixed integer k and $\phi \in \mathcal{R}^1$ if and only if

$$\text{cum}(z_i; z_i^*) = \text{cum}(a_i; a_i^*) \quad (11)$$

and

$$|\text{cum}(z_i : p; z_i^* : q)| = |\text{cum}(a_i : p; a_i^* : q)|, \quad (12)$$

for any non-negative integers p and q such that $p + q > 2$.

Proof: By theorem 1, if (11) and (12) are satisfied then

$$s_l = e^{j\phi} \delta_{l-k} \quad (13)$$

for some fixed integer k and $\phi \in \mathcal{R}^1$. Substituting (13) into (3) immediately yields (10).

On the other hand, if (10) is satisfied, then by properties (p.1) and (p.2), for any non-negative integers p and q ,

$$\begin{aligned} \text{cum}(z_i : p; z_i^* : q) &= \text{cum}(\underbrace{z_i; z_i; \dots; z_i}_{p \text{ terms}}; \underbrace{z_i^*; z_i^*; \dots; z_i^*}_{q \text{ terms}}) \\ &= \text{cum}(\underbrace{e^{j\phi} a_{i-k}; e^{j\phi} a_{i-k}; \dots; e^{j\phi} a_{i-k}}_{p \text{ terms}}; \\ &\quad \underbrace{e^{-j\phi} a_{i-k}^*; e^{-j\phi} a_{i-k}^*; \dots; e^{-j\phi} a_{i-k}^*}_{q \text{ terms}}) \\ &= e^{j(p-q)\phi} \text{cum}(\underbrace{a_{i-k}; a_{i-k}; \dots; a_{i-k}}_{p \text{ times}}; \underbrace{a_{i-k}^*; a_{i-k}^*; \dots; a_{i-k}^*}_{q \text{ times}}) \\ &= e^{j(p-q)\phi} \text{cum}(\underbrace{a_i; a_i; \dots; a_i}_{p \text{ terms}}; \underbrace{a_i^*; a_i^*; \dots; a_i^*}_{q \text{ terms}}) \\ &= e^{j(p-q)\phi} \text{cum}(a_i : p; a_i^* : q) \end{aligned} \quad (14)$$

which implies (11) and (12). Q.E.D.

Corollary 1.1 asserts that the output (recovered) sequence is identical to the input sequence up to a constant delay and possibly a constant phase shift if and only if the variance of the individual z_i and the magnitude of any of the non-zero cumulant of order $p + q > 2$ are equal to that of a_i . the constant delay is unavoidable because of the stationarity of the input sequence. However, the phase shift can be identified in some cases, as indicated below.

Corollary 1.2

Under the assumptions of theorem 1,

$$z_i = e^{j\phi} a_{i-k} \quad \text{w.p.1} \quad (15)$$

where k is some fixed integer and

$$(p - q)\phi = 2\pi N, \quad N - \text{integer} \quad (16)$$

if and only if

$$\text{cum}(z_i; z_i^*) = \text{cum}(a_i; a_i^*) \quad (17)$$

and

$$\text{cum}(z_i : p; z_i^* : q) = \text{cum}(a_i : p; a_i^* : q) \quad (18)$$

for some non-negative integers p and q such that $p + q > 2$.

Proof: If z_i and a_i are related by (15), then the corresponding cumulants are related by (14), which immediately imply (17) and (18).

On the other hand, if (17) is satisfied, then by theorem 1

$$|\text{cum}(z_i : p; z_i^* : q)| = |\text{cum}(a_i : p; a_i^* : q)|$$

if and only if

$$s_l = e^{j\phi} \delta_{l-k} = \begin{cases} e^{j\phi} & l = k \\ 0 & l \neq k \end{cases} \quad (19)$$

for some fixed integer k and $\phi \in \mathcal{R}^1$, in which case

$$\frac{\text{cum}(z_i : p; z_i^* : q)}{\text{cum}(a_i : p; a_i^* : q)} = \sum_l s_l^p s_l^{*q} = e^{j(p-q)\phi}$$

which is equal to 1, if and only if (16) is satisfied. Substituting (19) into (3) immediately yields (15). Q.E.D.

Corollary 1.2 asserts that if $\text{cum}(a_i : p; a_i^* : q) \neq 0$ where $p \neq q$, then the indicated phase shift can be identified up to an ambiguity of order $|p - q|$. Thus, if $|p - q| = 1$, then the phase shift between the input and recovered symbols can be completely eliminated. However, we note that in most signal constellations used for data communications, the input distribution is symmetric under rotation, in which case all cumulants of order $p \neq q$ are equal to zero.

Theorem 2 (real-case)

Let a_i be a sequence of real i.i.d. random variables with some given probability distribution. Suppose that $\text{cum}(a_i : 2) > 0$ and $\text{cum}(a_i : p) \neq 0$, where $p > 2$, exist. Let s_i be a deterministic (possibly infinite) real-valued sequence. Let the random variables z_i be specified by (3).

If

$$\text{cum}(z_i : 2) = \text{cum}(a_i : 2) \quad (20)$$

Then:

$$\frac{\text{cum}(z_i : p)}{\text{cum}(a_i : p)} \leq 1 \quad (21)$$

where equality holds if and only if

$$s_i = \rho \delta_{i-k} \quad (22)$$

where k is some fixed integer and

$$\rho = \begin{cases} 1 & p \text{ odd} \\ \pm 1 & p \text{ even} \end{cases} \quad (23)$$

Proof: Invoking (6) for $p = 2$ and $q = 0$, if

$$\sum s_l^2 = \frac{\text{cum}(z_i : 2)}{\text{cum}(a_i : 2)} = 1$$

Then

$$s_l^2 \leq 1 \quad \forall l$$

Thus

$$s_l \leq 1 \quad \forall l$$

Hence, for $p > 2$,

$$s_l^p \leq s_l^2 \quad \forall l$$

Thus, invoking (6) with $q = 0$,

$$\frac{\text{cum}(z_i : p)}{\text{cum}(a_i : p)} = \sum_l s_l^p \leq \sum_l s_l^2 = 1$$

where equality holds if and only if (22) is satisfied. Q.E.D.

The condition $\text{cum}(a_i : 2) = E\{(a_i - E\{a_i\})^2\} > 0$ implies that a_i is a non-trivial random variable (with non-zero variance). The condition $\text{cum}(a_i : p) \neq 0$ where $p > 2$ implies that the probability distribution of a_i is non-Gaussian (see property (p.4)).

Corollary 2.1

Under the assumptions of theorem 2,

$$z_i = \rho a_{i-k} \quad (\text{w.p.1}) \quad (24)$$

where ρ is specified by (23), if and only if

$$\text{cum}(z_i : 2) = \text{cum}(a_i : 2) \quad (25)$$

and

$$\text{cum}(z_i : p) = \text{cum}(a_i : p), \quad (26)$$

for any $p > 2$.

Proof: By theorem 1, if (25) and (26) are satisfied, then

$$s_l = \rho \delta_{l-k} \quad (27)$$

where ρ is specified by (23). Substituting (27) into (3) immediately yields (24).

On the other hand, if (24) is satisfied, then following the same development as in (14)

$$\text{cum}(z_i : p) = \rho^p \text{cum}(a_i : p) \quad (28)$$

where by (23) $\rho^p = 1$. Q.E.D.

Recall [5] theorem 2.2, a sufficient condition for the output (recovered) sequence to be identical to the input sequence, up to a delay and a sign, is that the probability distribution of the individual z_i be equal the probability distribution of a_i . According to corollary 2.1, it is necessary and sufficient to equalize the second order cumulant and any other non-zero cumulant of order $p > 2$. Furthermore, if the higher order cumulant being used is odd, the sign ambiguity can be resolved.

Theorem 1 suggests the following family of equalization criteria, indexed by p and q :

$$\begin{aligned} & \text{Max} |\text{cum}(z_i : p; z_i^* : q)| \\ & \text{Subject to : } \text{cum}(z_i; z_i^*) = \text{cum}(a_i; a_i^*) \end{aligned} \quad (29)$$

We note that the choice $p = q = 2$ yields the optimization criterion proposed in [6]. In [6] it is assumed that the input symbols are zero-mean (i.e., $E\{a_i\} = 0$), in which case we obtain the average power constraint $E\{|z_i|^2\} = E\{|a_i|^2\}$.

In the real case, theorem 2 suggests the following family of criteria, indexed by p :

$$\begin{aligned} & \text{Max } \{\text{sign} [\text{cum}(a_i : p)] \text{cum}(z_i : p)\} \\ & \text{Subject to : } \text{cum}(z_i : 2) = \text{cum}(a_i : 2) \end{aligned} \quad (30)$$

where

$$\text{sign}(x) = \begin{cases} -1 & x < 0 \\ +1 & x > 0 \end{cases} \quad (31)$$

Recall (6), if p is even and s_l are real valued,

$$\text{sign}(\text{cum}(z_i : p)) = \text{sign}(\text{cum}(a_i : p)) \quad (32)$$

Thus, for even p , (30) reduces to:

$$\begin{aligned} & \text{Max} |\text{cum}(z_i : p)| \\ & \text{Subject to : } \text{cum}(z_i : 2) = \text{cum}(a_i : 2) \end{aligned} \quad (33)$$

By the theorems, the set of solutions of these constrained optimization criteria correspond to the desired response. We now want to make sure that the criteria function do not have spurious local maxima.

Consider first the real case. By (6), the constrained maximization in (30) is equivalent to:

$$\begin{aligned} & \text{Max} \sum_l s_l^p \\ & \text{Subject to : } \sum_l s_l^2 = 1 \end{aligned} \quad (34)$$

By the constraint

$$s_0 = \left(1 - \sum_{l \neq 0} s_l^2 \right)^{1/2} \quad (35)$$

By the constraint, there is at least one non-zero s_l . Thus, we suppose, without any loss of generality, that $0 < |s_0| \leq 1$. Substituting (35) into (34) we obtain:

$$f(\underline{s}) = \sum_{l \neq 0} s_l^p + \left(1 - \sum_{l \neq 0} s_l^2 \right)^{p/2} \quad (36)$$

Clearly, the maximization of $f(\underline{s})$ is equivalent to solving (34). We note that $f(\underline{s})$ is well defined for $\sum_{l \neq 0} s_l^2 = 1 - s_0^2 < 1$, which is the domain of interest. Differentiating (36),

$$\frac{\partial f(\underline{s})}{\partial s_i} = p s_i \left[s_i^{p-2} - \left(1 - \sum_{l \neq 0} s_l^2 \right)^{p/2-1} \right] \quad (37)$$

The roots of (37) are, for even p :

$$s_i = 0 \quad \text{or} \quad s_i = \pm \left(1 - \sum_{l \neq 0} s_l^2 \right)^{1/2} = \pm s_0$$

and for odd p :

$$s_i = 0 \quad \text{or} \quad s_i = + \left(1 - \sum_{l \neq 0} s_l^2 \right)^{1/2} = s_0$$

vectors \underline{s} having only one non-zero component of magnitude 1 must be local maxima. By theorem 2, we already know that these are also the global maxima. For odd p , all the eigenvalues are either negative if $\alpha = 1$, or positive if $\alpha = -1$. Thus, for odd p , all vectors \underline{s} having one non-zero component that is equal to 1 are local maxima, by theorem 2 they are global maxima, and all vectors \underline{s} having one non-zero component that is equal to -1 are local minima. In fact, they are global minima. This indicates once again the ability to resolve sign ambiguity in case p is odd.

If $M = 2$, one eigenvalue of the Hessian equals to 2η , and all other eigenvalues equal to μ . If $M > 2$, it can easily be shown that one eigenvalue equals to $(M - 2)\eta$, $(M - 2)$ eigenvalues are equal to 2η , and all other eigenvalues are equal to μ . Since μ and η have opposite signs, it indicates that all other stationary points specified by $M = 2, 3, \dots$ are unstable equilibria (saddle points). We can therefore state the following lemma:

Lemma 2.1 (real-case)

Under the assumptions of theorem 2, the only local (hence, global) maxima of the constrained optimization in (34), or equivalently (30), correspond to the desired solution $s_i = \rho\delta_{i-k}$, or equivalently $z_i = \rho a_{i-k}$, where k is an arbitrary constant delay, and ρ is specified by (23).

The proof of the lemma is self-evident.

We now turn to the complex case. If $p = q$, then the constrained maximization in (29) is equivalent to:

$$\begin{aligned} & \text{Max } \sum_l |s_l|^{2p} \\ & \text{Subject to : } \sum_l |s_l|^2 = 1 \end{aligned} \quad (38)$$

The optimization in (38) is identical to the optimization in (34), only that instead of s_l we have the variables $|s_l|$, and instead of p we have $2p$. Thus, in complete analogy, the set of stationary points of the constrained optimization in (38) are all vectors \underline{s} having M non-zero components of magnitude $1/\sqrt{M}$. For $M = 1$ we obtain the set of local maxima which, by theorem 2, are also the

global maxima. All other stationary points for $M = 2, 3, \dots$ are unstable equilibria. We therefore conclude the following:

Lemma 1.1 (complex case)

Under the assumptions of theorem 1, the only local (hence, global) maxima of the constrained optimization in (38), or equivalently (29) with $p = q$, correspond to the desired solution $s_i = e^{j\phi} \delta_{i-k}$, or equivalently $z_i = e^{j\phi} a_{i-k}$, where k is some fixed delay and ϕ is an arbitrary phase.

Comment: The analysis of the stationary points in the complex case with $p \neq q$ is more complicated and will therefore not be presented here. As already pointed out, in most signal constellations used for data communications, the probability distribution of the input symbols is symmetric under rotation, in which case only $p = q$ is of interest.

ALGORITHM IMPLEMENTATION

The constrained optimizations in (34) and in (38) give rise to the following gradient-based algorithm:

$$s'_i = s_i + \delta \frac{\partial F}{\partial s_i} \quad (39)$$

$$s''_i = \frac{1}{\sqrt{\sum_l |s'_l|^2}} s'_i \quad (40)$$

where F is the objective function to be maximized, $\delta > 0$ is the step-size, s_i are the unit sample response coefficients of the combined channel-equalizer prior to the iteration, and s''_i are the unit sample response coefficients after the iteration. The normalization operation in (40) is necessary to satisfy the indicated constraint. In the case of (34), $F = \sum_l s_l^p$, and in the case of (38) $F = \sum_l |s_l|^{2p}$. Since in both cases it has been proved that F has no spurious local maxima, the algorithm is expected to converge to the desired solution regardless of initialization.

A common measure of equalization performance is the inter-symbol-interference (ISI) defined by:

$$\text{ISI}(\underline{s}) = \frac{\sum_l |s_l|^2 - |s_{max}|^2}{|s_{max}|^2} \quad (41)$$

where $|s_{max}|$ is the component of \underline{s} having the maximum absolute value. Clearly, $\text{ISI} = 0$ if and only if \underline{s} has only one non-zero component - that is the desired response, and small ISI indicates the proximity to the desired solution.

To analyze the rate of convergence of the ISI using the gradient-search algorithm in (39), (40) we suppose, without any loss of generality, that s_0 is the component of \underline{s} having the largest absolute value. Thus,

$$\text{ISI}(\underline{s}) = \sum_{l \neq 0} |s_l|^2 / |s_0|^2 \quad (42)$$

Now, if $F = \sum_l |s_l|^{2p}$ (complex case), then $\partial F / \partial s_i = 2p |s_i|^{2p-2} s_i$, in which case

$$\frac{|s''_i|}{|s''_0|} = \frac{|s'_i|}{|s'_0|} = \frac{1 + 2p\delta |s_i|^{2p-2}}{1 + 2p\delta |s_0|^{2p-2}} \cdot \frac{|s_i|}{|s_0|} \quad (43)$$

Since $\sum_l |s_l|^2 = 1$ and since $|s_i| \leq |s_0| \leq 1$, then:

$$1 \leq 1 + 2p\delta |s_i|^{2p-2} < 1 + 2p\delta, \quad i \neq 0 \quad (44)$$

Using (44) in (43),

$$\frac{1}{1 + 2p\delta} \frac{|s_i|}{|s_0|} \leq \frac{|s_i''|}{|s_0''|} < \frac{|s_i|}{|s_0|} \quad (45)$$

Therefore, using (45) in (42),

$$\frac{1}{(1 + 2p\delta)^2} \text{ISI}(\underline{s}) \leq \text{ISI}(\underline{s}'') < \text{ISI}(\underline{s}) \quad (46)$$

The right inequality asserts that the ISI monotonically decreases from iteration to iteration, where the left inequality sets an upper bound on the rate of convergence of the ISI. It asserts that the factor of improvement in the ISI from iteration to iteration is bounded by $1/(1 + 2p\delta)^2$. Near the point of convergence, $|s_0| \approx 1$ and $|s_i| \approx 0$ for $i \neq 0$, in which case the left inequality is tight indicating that the rate of convergence is approximately $1/(1 + 2p\delta)^2$. It therefore suggests to use a large step-size, where the left inequality in (46) ensures monotonic convergence for any choice of δ .

If we choose a very large δ , the term $\delta \frac{\partial F}{\partial s_i}$ appearing on the right hand side of (39) becomes dominant, and in the limit we approach the following algorithm:

$$s_i' = \frac{\partial F}{\partial s_i} \quad (47)$$

$$s_i'' = \frac{1}{\sqrt{\sum_l |s_l'|^2}} s_i' \quad (48)$$

where we note that δ can be factored out of (47) because of the normalization operation in (48).

In this setting, the iterated \underline{s} is determined solely by the direction of the gradient. In this case,

$$\frac{|s_i''|}{|s_0''|} = \frac{|s_i'|}{|s_0'|} = \frac{|s_i|^{2p-1}}{|s_0|^{2p-1}} \quad (49)$$

since

$$\sum_i |s_i|^{2(2p-1)} \leq \left(\sum_i |s_i|^2 \right)^{2p-1} \quad (50)$$

then

$$\begin{aligned} \text{ISI}(\underline{s}'') &= \sum_{i \neq 0} |s_i''|^2 / |s_0''|^2 = \sum_{i \neq 0} |s_i|^{2(2p-1)} / |s_0|^{2(2p-1)} \\ &\leq \left(\sum_{i \neq 0} |s_i|^2 / |s_0|^2 \right)^{2p-1} = [\text{ISI}(\underline{s})]^{2p-1} \end{aligned} \quad (51)$$

indicating a very fast, at least exponential, convergence of the ISI.

Similar results can be obtained in the real case where $F = \sum_l s_l^p$.

The algorithms in (39), (40) and in (47), (48) are explicit since they are presented in terms of the unit sample response coefficients s_i of the combined system S . We want to express these algorithms in terms of the tapes c_i of the equalizer C . To do that, we use the convolutional relation between s_i and c_i :

$$s_i = h_i \circ c_i = \sum_l c_l h_{i-l} \quad (52)$$

where h_i is the unit sample response of the unknown system (channel) H . Invoking the chain rule for differentiation,

$$\frac{\partial F}{\partial c_i} = \sum_l \frac{\partial F}{\partial s_l} h_{l-i}^* = h_{-i}^* \circ \frac{\partial F}{\partial s_i} \quad (53)$$

where we note that in the complex case $\partial F / \partial c_i = \partial F / \partial R_e(c_i) + j \partial F / \partial I_m(c_i)$.

By assumption (ii), the Fourier transform $H(\omega)$ of h_i contains no zeros on the unit circle. Therefore, its inverse $H^{-1}(\omega)$ exists. We denote by h_i^{-1} the inverse Fourier transform of $H^{-1}(\omega)$. Thus, $h_i^{-1} \circ h_i = \delta_i$. Convoluting both sides of (53) by $(h_{-i}^{-1})^*$

$$\frac{\partial F}{\partial s_i} = (h_{-i}^{-1})^* \circ \frac{\partial F}{\partial c_i} \quad (54)$$

Substituting (52) and (54) into (39),

$$h_i \circ c_i' = h_i \circ c_i + \delta (h_{-i}^{-1})^* \circ \frac{\partial F}{\partial c_i} \quad (55)$$

convolving both sides of (55) by h_i^{-1} ,

$$c'_i = c_i + \delta h_i^{-1} \circ (h_i^{-1})^* \circ \frac{\partial F}{\partial c_i} \quad (56)$$

The Fourier transform of $h_i^{-1} \circ (h_i^{-1})^*$ is $1/|H(\omega)|^2$. Thus, if $H(\omega)$ is spectrally white, that is:

$$|H(\omega)| = 1 \quad 0 \leq \omega < 2\pi \quad (57)$$

then $h_i^{-1} \circ (h_i^{-1})^* = \delta_i$, and (56) reduces to:

$$c'_i = c_i + \delta \frac{\partial F}{\partial c_i} \quad (58)$$

Also,

$$\begin{aligned} \sum_l |s_l|^2 &= \frac{1}{2\pi} \int_{-\pi}^{+\pi} |S(\omega)|^2 d\omega = \frac{1}{2\pi} \int_{-\pi}^{+\pi} |H(\omega)C(\omega)|^2 d\omega \\ &= \frac{1}{2\pi} \int_{-\pi}^{+\pi} |C(\omega)|^2 d\omega = \sum_l |c_l|^2 \end{aligned} \quad (59)$$

Therefore, normalizing the s_l is the same as normalizing the c_l :

$$c''_i = \frac{1}{\sqrt{\sum_l |c'_l|^2}} c'_i \quad (60)$$

The algorithm specified by (58), (60) is identical to (39), (40) if H is spectrally white, implying a spectral pre-whitening operation. Only in this case, a gradient step in the S domain is equivalent to a gradient step in the C domain.

Similarly, under the spectral pre-whitening operation, the algorithm is (47) (48) is equivalent to:

$$c'_i = \frac{\partial F}{\partial c_i} \quad (61)$$

$$c''_i = \frac{1}{\sqrt{\sum_l |c'_l|^2}} c'_i \quad (62)$$

Next, we want to develop an explicit expression for the partial derivatives $\partial F/\partial c_m$. In the complex case with $p = q$,

$$\begin{aligned} F &= |\text{cum}(z_i : p; z_i^* : p)| = \\ &= |\text{cum}(a_i : p; a_i^* : p)| \sum_l |s_l|^{2p} \end{aligned} \quad (63)$$

Therefore,

$$\frac{\partial F}{\partial s_m} = 2p |\text{cum}(a_i : p; a_i^* : p)| s_m^p (s_m^*)^{p-1} \quad (64)$$

Substituting (64) into (53),

$$\frac{\partial F}{\partial c_m} = 2p |\text{cum}(a_i : p; a_i^* : p)| \sum_l s_l^p (s_l^*)^{p-1} h_{l-m} \quad (65)$$

Now,

$$\begin{aligned} \text{cum}(z_i : p; z_i^* : p-1; y_{i-m}^*) &= \text{cum} \left(\sum_{l_1} s_{l_1} a_{i-l_1}; \dots; \sum_{l_p} s_{l_p} a_{i-l_p}; \sum_{k_1} s_{k_1}^* a_{i-k_1}^* \right. \\ &\quad \left. \dots; \sum_{k_{p-1}} s_{k_{p-1}}^* a_{i-k_{p-1}}^*; \sum_{k_p} h_{k_p}^* a_{i-m-k_p}^* \right) \\ &= \sum_{l_1} \dots \sum_{l_p} \sum_{k_1} \dots \sum_{k_p} s_{l_1} \dots s_{l_p} s_{k_1}^* \dots s_{k_{p-1}}^* h_{k_p}^* \cdot \\ &\quad \cdot \text{cum} \left(a_{i-l_1}; \dots; a_{i-l_p}; a_{i-k_p}^*; \dots; a_{i-k_{p-1}}^*; a_{i-m-k_p}^* \right) \\ &= \text{cum}(a_i : p; a_i^* : p) \sum_l s_l^p (s_l^*)^{p-1} h_{l-m} \end{aligned} \quad (66)$$

Combining (65) with (66)

$$\begin{aligned} \frac{\partial F}{\partial c_m} &= 2p \frac{|\text{cum}(a_i : p; a_i^* : p)|}{\text{cum}(a_i : p; a_i^* : p)} \text{cum}(z_i^* : p; z_i^* : p-1; y_{i-m}^*) \\ &= 2p \cdot \text{sign}[\text{cum}(a_i : p; a_i^* : p)] \text{cum}(z_i : p; z_i^* : p-1; y_{i-m}^*) \end{aligned} \quad (67)$$

In the real case,

$$\begin{aligned} F &= \text{sign}[\text{cum}(a_i : p)] \text{cum}(z_i : p) \\ &= |\text{cum}(a_i : p)| \sum_l s_l^p \end{aligned} \quad (68)$$

Following the same considerations leading from (63) to (67),

$$\frac{\partial F}{\partial c_m} = p \operatorname{sign}[\operatorname{cum}(a_i : p)] \operatorname{cum}(z_i : p - 1; y_{i-m}) \quad (69)$$

The computation of the gradient using either (67) or (69) requires knowledge of the sign of the corresponding cumulant of a_i , and this is the only prior information concerning the input distribution that is required. We also need to know the joint cumulant of the input y_i and the output z_i of the equalizer. These cumulants can be estimated using the available data.

To demonstrate that, consider the real case with $p = 3$. Assuming that a_i is zero mean, y_i and z_i are also zero mean, in which case (69) reduces to:

$$\frac{\partial F}{\partial c_m} = \operatorname{sign}[E\{a_i^3\}] E\{z_i^2 y_{i-m}\} \quad (70)$$

The expectation $E\{z_i^2 y_{i-m}\}$ may be approximated using the current realization $z_i^2 y_{i-m}$, or it may be estimated by the cumulative average $\frac{1}{N} \sum_i z_i^2 y_{i-m}$, where N corresponds to the number of terms in the sum. These approximations result recursive/sequential algorithms (the cumulative averaging can be evaluated sequentially) in which we perform one iteration per symbol.

Alternatively, using the input-output relation:

$$z_i = y_i \circ c_i = \sum_l c_l y_{i-l} \quad (71)$$

we obtain:

$$\begin{aligned} E\{z_i^2 y_{i-m}\} &= E \left\{ \left[\sum_l c_l y_{i-l} \right]^2 y_{i-m} \right\} \\ &= \sum_{k_1} \sum_{k_2} c_{k_1} c_{k_2} E \{ y_{i-l} y_{i-k_1} y_{i-k_2} \} \end{aligned} \quad (72)$$

This equation suggests a non-recursive iterative algorithm in which we first estimate the expectation $E\{y_{i-l} y_{i-k_1} y_{i-k_2}\}$ by the empirical averaging $\frac{1}{N} \sum_i y_{i-l} y_{i-k_1} y_{i-k_2}$ over the observed block of data, and then we iterate using the proposed algorithms until convergence is accomplished.

As another case, consider (69) with the choice $p = 4$. Assuming the a_i is zero mean, the components of the gradient (subject to the indicated constraint) are:

$$\frac{\partial F}{\partial c_m} = 4 \text{sign}[\text{cum}(a_i : 4)] E\{z_i^3 y_{i-m}\} \quad (73)$$

where we note that $\text{cum}(a_i : 4) = E\{a_i^4\} - 3E\{a_i^3\}$. The term $E\{z_i^3 y_{i-m}\}$ can similarly be approximated by replacing the expectation by the current realization, or by performing empirical averaging.

Consider now the complex case with $p = q = 2$. Once again we assume that a_i is zero mean. For simplicity, we further assume that $E\{a_i^2\} = 0$ (e.g., the real and complex components of a_i are statistically uncorrelated with equal variance - a condition that is satisfied for most signal constellations used for data communications). Under these assumptions, (67) reduces to:

$$\frac{\partial F}{\partial c_m} = 4 \text{sign}[\text{cum}(a_i : 2; a_i^* : 2)] E\{|z_i|^2 z_i y_{i-m}^*\} \quad (74)$$

This formula coincides with the result developed in [6]. The expectation in (74) can be estimated by the current realization (in which case we obtain the stochastic gradient algorithm presented in [6]), or by cumulative averaging operation.

EFFECT OF ADDITIVE NOISE

So far, we have completely ignored the presence of additive noise in the system. To study this effect, let the noise (error) signal e_i be modeled as the output of a linear time-invariant filter $T = \{t_i\}$ driven by a white (i.i.d.) noise process v_i as illustrated in Figure 2. We suppose that a_i and v_i are statistically independent. Note that if $T = C$ then v_i represents the additive noise generated at the input to the equalizer. If $T = H \circ C$, the v_i represents the additive noise generated at the channel input. The extension of the model to include several noise sources is straightforward.

Invoking the statistical independence between a_i and v_i ,

$$\begin{aligned} \text{cum}(z_i : p; z_i^* : q) &= \text{cum}(a_i : p; a_i^* : q) \sum_l s_l^p s_l^{*q} \\ &+ \text{cum}(v_i : p; v_i^* : q) \sum_l t_l^p t_l^{*q} \end{aligned} \quad (75)$$

Thus, if the cumulants of $e_i = v_i \circ t_i$ are known, or independently measured, their effect can be removed by a simple subtraction.

Perhaps the most interesting observation is that if v_i is Gaussian, then $\text{cum}(v_i : p; v_i^* : q) = 0$ for $p + q > 2$. Therefore, the criteria functions in (29) and (30) are unaffected by the presence of additive Gaussian noise. Substituting $p = q = 1$ in (75),

$$\text{cum}(z_i; z_i^*) = \text{cum}(a_i; a_i^*) \sum_l |s_l|^2 + \text{cum}(v_i; v_i^*) \sum_l |t_l|^2 \quad (76)$$

Thus, the additive Gaussian noise only affect the constraint. Consequently, it can be verified that the optimization criteria still converge to the set of desired solutions up to a constant gain factor.

References

- [1] Y. Sato, "A method for self recovering equalization," IEEE Trans. on Commun., Vol. COM-23, No. 6, pp. 679-682, June 1985.
- [2] D. Godard, "Self recovering equalization and carrier tracking in two-dimensional data communication systems," IEEE Trans. on Commun., Vol. COM-28, No. 11, pp. 1867-1875, November 1980.
- [3] A. Benveniste and M. Goursat, "Blind Equalizer," IEEE Trans. on Commun., Vol. COM-32, No. 8, pp. 871-883, August 1984.
- [4] G.J. Foschini, "Equalization without altering or detecting data," Bell Syst. Tech. J., Vol. 64, No. 8, pp. 1885-1911, October 1985.
- [5] A. Benveniste, M. Goursat, and G. Ruget, "Robust identification of a non-minimum phase system: blind adjustment of a linear equalizer in data communication," IEEE Trans. on Auto. Contr., Vol. AC-25, No. 3, pp. 385-399, June 1980.
- [6] O. Shalvi and E. Weinstein, "New Criteria for blind deconvolution of non-minimum phase systems (channels)," to appear in the March issue of the IEEE Trans. on Information Theory.

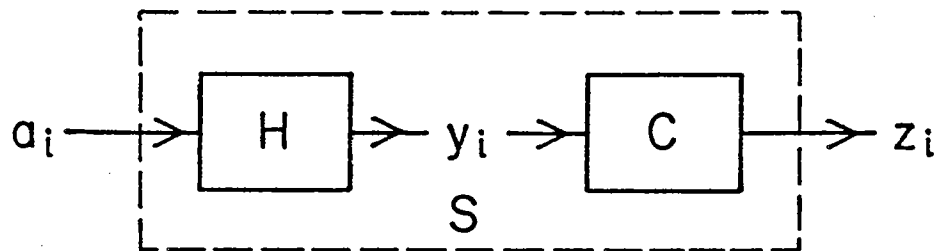


Figure 1

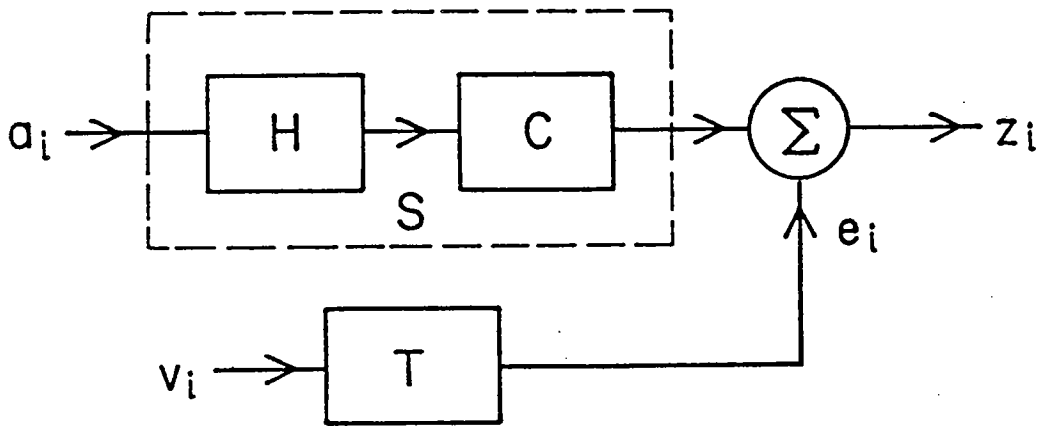


Figure 2

DOCUMENT LIBRARY

January 17, 1990

Distribution List for Technical Report Exchange

Attn: Stella Sanchez-Wade
Documents Section
Scripps Institution of Oceanography
Library, Mail Code C-075C
La Jolla, CA 92093

Hancock Library of Biology &
Oceanography
Alan Hancock Laboratory
University of Southern California
University Park
Los Angeles, CA 90089-0371

Gifts & Exchanges
Library
Bedford Institute of Oceanography
P.O. Box 1006
Dartmouth, NS, B2Y 4A2, CANADA

Office of the International
Ice Patrol
c/o Coast Guard R & D Center
Avery Point
Groton, CT 06340

NOAA/EDIS Miami Library Center
4301 Rickenbacker Causeway
Miami, FL 33149

Library
Skidaway Institute of Oceanography
P.O. Box 13687
Savannah, GA 31416

Institute of Geophysics
University of Hawaii
Library Room 252
2525 Correa Road
Honolulu, HI 96822

Marine Resources Information Center
Building E38-320
MIT
Cambridge, MA 02139

Library
Lamont-Doherty Geological
Observatory
Columbia University
Palisades, NY 10964

Library
Serials Department
Oregon State University
Corvallis, OR 97331

Pell Marine Science Library
University of Rhode Island
Narragansett Bay Campus
Narragansett, RI 02882

Working Collection
Texas A&M University
Dept. of Oceanography
College Station, TX 77843

Library
Virginia Institute of Marine Science
Gloucester Point, VA 23062

Fisheries-Oceanography Library
151 Oceanography Teaching Bldg.
University of Washington
Seattle, WA 98195

Library
R.S.M.A.S.
University of Miami
4600 Rickenbacker Causeway
Miami, FL 33149

Maury Oceanographic Library
Naval Oceanographic Office
Bay St. Louis
NSTL, MS 39522-5001

Marine Sciences Collection
Mayaguez Campus Library
University of Puerto Rico
Mayaguez, Puerto Rico 00708

Library
Institute of Oceanographic Sciences
Deacon Laboratory
Wormley, Godalming
Surrey GU8 5UB
UNITED KINGDOM

The Librarian
CSIRO Marine Laboratories
G.P.O. Box 1538
Hobart, Tasmania
AUSTRALIA 7001

Library
Proudman Oceanographic Laboratory
Bidston Observatory
Birkenhead
Merseyside L43 7 RA
UNITED KINGDOM

REPORT DOCUMENTATION PAGE	1. REPORT NO. WHOI-90-08	2.	3. Recipient's Accession No.
4. Title and Subtitle Universal Criteria for Blind Deconvolution		5. Report Date February, 1990	
7. Author(s) Ofir Shalvi and Ehud Weinstein		6.	
9. Performing Organization Name and Address The Woods Hole Oceanographic Institution Woods Hole, Massachusetts 02543		8. Performing Organization Rept. No. WHOI 90-08	
12. Sponsoring Organization Name and Address Funding was provided by the Office of Naval Research.		10. Project/Task/Work Unit No.	
15. Supplementary Notes This report should be cited as: Woods Hole Oceanog. Inst. Tech. Rept., WHOI-90-08.		11. Contract(C) or Grant(G) No. (C) (G) N00014-90-J-1109	
16. Abstract (Limit: 200 words) We present necessary and sufficient conditions for blind equalization/deconvolution (without observing the input) of an unknown, possible non-minimum phase linear time invariant system (channel). Based on that, we propose a family of optimization criteria and prove that their solution correspond to the desired response. These criteria, and the associated gradient-search algorithms, involve the computation of high order cumulants. The proposed criteria are universal in the sense that they do not impose any restrictions on the probability distribution of the input symbols. We also address the problem of additive noise in the system and show that in several important cases, e.g. when the additive noise is Gaussian, the proposed criteria are unaffected.		13. Type of Report & Period Covered Technical Report	
17. Document Analysis a. Descriptors 1. blind deconvolution 2. channel equalization 3. high order cumulants b. Identifiers/Open-Ended Terms c. COSATI Field/Group		14.	
18. Availability Statement Approved for publication; distribution unlimited.	19. Security Class (This Report) UNCLASSIFIED	21. No. of Pages 28	
	20. Security Class (This Page)	22. Price	