

Convex Cost Functions in Blind Equalization

Sridhar Vembu, *Student Member, IEEE*, Sergio Verdú, *Fellow, IEEE*, Rodney A. Kennedy, and William Sethares

Abstract—Existing blind adaptive equalizers that use nonconvex cost functions and stochastic gradient descent suffer from lack of global convergence to an equalizer setup that removes sufficient ISI when an FIR equalizer is used. In this paper, we impose convexity on the cost function and anchoring of the equalizer away from the all-zero setup. We establish that there exists a globally convergent blind equalization strategy for 1-D pulse amplitude modulation (PAM) systems with bounded input data (discrete or continuous) even when the equalizer is truncated. The resulting cost function is a constrained l_1 norm of the joint impulse response of the channel and the equalizer. Our results apply to arbitrary linear channels (provided there are no unit circle zeros) and apply regardless of the initial ISI (that is whether the eye is initially open or closed). We also show a globally convergent stochastic gradient scheme based on an implementable approximation of the l_1 cost function.

I. INTRODUCTION

A. Statement of the Problem

A REAL, ZERO mean, stationary sequence is transformed by an unknown linear channel as shown in Fig. 1.

We assume that the sequence is drawn independently and identically from a symmetric distribution with finite support (discrete or continuous). We want to remove the intersymbol interference (ISI) caused by the channel by adapting a linear equalizer so that its coefficients converge to the inverse of the channel. This is the case for baseband pulse amplitude modulation (PAM) channels. In the context of blind equalization, we do not assume that an initial training sequence is sent. Solutions to the conventional equalization problem (where the input to the channel is known) are well known (see, for instance, [16]). So we assume that the receiver has access only to the channel output and the equalizer can only use this and the knowledge of the statistics of the input process.

The motivation for studying this problem is that in many-communications and system identification problems the assumption of a training sequence that *precedes* the data is unrealistic. Examples include multi-point networks and fading

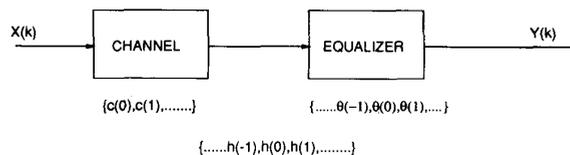


Fig. 1. The channel and the equalizer.

channels. A discussion of such problems is found in [1]. However, as pointed out in [5], there is some interesting work on having a very low duty cycle training sequence *inside* the bitstream which shows promise in decision directed equalization in terrestrial digital radio systems operating in a fading environment¹. See also [6] where the term “probing sequences” is used.

Several comments are in order about the nature of the problem. First of all, if the input process is Gaussian it is well known [11] that we can only identify the magnitude spectrum of the channel. This is because a stationary Gaussian process is completely characterized by its power spectral density which depends only on the magnitude spectrum of the channel. Even if the process is non-Gaussian, if we only use the second order moments of the channel output we will not be able to identify the phase. This is sufficient if the channel is known to be minimum phase but many of the channels of interest such as telephone channels are known to be nonminimum phase. See [16] for other examples.

From the above discussion it is clear that we need to use the higher order statistics of the channel output in order to identify the phase. In fact methods exist in which the channel coefficients are obtained by solving systems of nonlinear equations that involve the higher order cumulants of the channel output [15]. Some recursive algorithms based on cumulants have also been reported and they are computationally intensive (see [19]). In this article we adopt a cost function based approach which is explained in the next section.

One final remark about what we can hope to achieve through a blind equalization algorithm. Due to the stationarity of the input distribution, we can equalize only up to an arbitrary delay. In addition, if the input distribution is symmetric a sign inversion may occur. These are due to the fact that the statistics of the input distribution are invariant with respect to delays and sign changes. In fact, as first suggested in [2], [3], an unknown gain is also acceptable. We can take care of this through automatic gain control. So we will focus on equalization algorithms that introduce an unknown delay and gain. We ignore noise throughout this article. This is justified because in many situations of interest in this article, the ISI

¹ We thank an anonymous reviewer for pointing out this and [6] to us.

Manuscript received October 27, 1991; revised October 11, 1993. S. Vembu and S. Verdú were supported by the National Science Foundation under PYI Grant ECSE-8857689. R. A. Kennedy was supported by the ANU Centre for Information Science Research and the Australian Research Council. The associate editor coordinating the review of this paper and approving it for publication was Dr. H. Fan.

S. Vembu and S. Verdú are with Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA.

R. A. Kennedy is with Department of Systems Engineering, Research School of Physical Sciences and Engineering, and Interdisciplinary Engineering Program, Faculty of Science, Australian National University, Canberra ACT 2601, Australia.

W. Sethares is with the Department of Electrical and Computer Engineering, University of Wisconsin, Madison, WI 53706 USA.

IEEE Log Number 9401918.

caused by the channel is the main impairment. Moreover, globally convergent finitely parametrized blind equalization algorithms are not known even in the absence of noise (except in the case of minimum phase channels). Finally, like all linear equalization schemes, the present one also will not work properly if the channel has deep spectral nulls (zeros on or near the unit circle).

B. Previous Work

We focus on equalization algorithms which use the cost function

$$\mathcal{J}(\underline{\theta}) = Ef(Y_k) \tag{1.1}$$

where $\underline{\theta}$ is the parameter vector of the equalizer, Y_k is the output of the equalizer, $f(\cdot)$ is nonnegative and differentiable, and E denotes expectation. The coefficients of the equalizer are updated according to a stochastic gradient algorithm

$$\underline{\theta}^{k+1} = \underline{\theta}^k - \mu f'(Y_k) \nabla_{\underline{\theta}} Y_k(\underline{\theta}) \tag{1.2}$$

where μ is the step size and $'$ denotes differentiation. Thus we are using the estimate of the gradient to update $\underline{\theta}$ (assuming some regularity conditions on $f(\cdot)$). The advantage of this approach is the obvious simplicity of implementation compared to more computationally intensive approaches. Note that if $f(\cdot)$ is a convex function of its argument, $\mathcal{J}(\cdot)$ is a convex function of $\underline{\theta}$ as well. Obviously, an arbitrary convex function of $\underline{\theta}$ need not have a representation in the form of (1.1). In fact, in this article we will concentrate on convex $\mathcal{J}(\cdot)$ that are not necessarily of the form of (1.1) and derive a necessary condition which turns out to be very restrictive. Then we exhibit a function that satisfies this condition and show how to approximate it arbitrarily closely by functions of the form (1.1).

One of the first algorithms proposed in the literature which is of the form of (1.1) is the Godard algorithm [1]. This uses a nonconvex cost function. Global convergence to a valid equalizer setup is true for doubly infinite parametrizations of the Godard equalizer, as has been shown in [5]. However, when we use truncated FIR implementations of the equalizer (as we have to do in practice) the Godard algorithm might end up converging to an equalizer setup that removes an insufficient amount of ISI. This has been demonstrated in [9]. In other words, there exist simple channels for which the Godard cost function has undesirable local minima which result in insufficient removal of ISI. Even if infinitely parametrized, but causal, transversal Godard equalizers are allowed, convergence is guaranteed only for minimum-phase channels [3]. The same behavior holds for other blind equalization algorithms like Bussgang [20], Sato [22], and Shalvi-Weinstein [8]. As commented upon in [21] in the Sato case, this ill-convergence is not merely the result of under- or over-parametrization of the equalizer but seems to be a fundamental feature of the approach which attempts to restore some property of the channel input sequence (like modulus restoration in the Godard algorithm). It should be noted that the ‘‘tap-centering’’ strategy (that is, to prevent the equalizer taps from drifting to one end or the other) proposed by Foschini [5] is a

practically attractive way of overcoming the ill-convergence of the truncated versions of the equalizer. It works well in practice, though as yet there is no theoretical demonstration of its global convergence.

The work by Verdu, Anderson, and Kennedy [3], [4] considers in part a special case of the setting of this paper. The fundamental insight of [3], [4] (originally commented upon in [2]) is that not trying to identify the gain of the channel could lead to a globally convergent strategy. This is also vindicated by the present work. References [3], [4] use as cost function the average energy of the equalizer output $\mathcal{J}(\underline{\theta}) = EY_k^2$ along with θ_0 anchored at 1. Because of such anchoring, equalization is achieved with a gain which is not predetermined. Global convergence for FIR as well as finite dimensional IIR parametrizations of the equalizer have been demonstrated in [3]. For an arbitrary channel, this algorithm cancels its minimum phase equivalent and hence if the channel is not minimum phase, the channel equalizer combination becomes an allpass transfer function which is then identified using the information provided by the equalizer parameters. The idea of anchoring was the starting point of the present work.

We would like to mention the work of Rupprecht who has also investigated the use of convex cost functions (with a fixed tap) [25]–[27]. It is in these references that one first finds the approach to blind equalization that we study but with limited analysis of the convergence properties. Our work, however, goes beyond these earlier investigations by establishing that among the class of convex cost functions (with or without memory) there is only this single plausible candidate. We show theoretically and generically that the global convergence point is indeed located at an equalizer parameter setting which achieves perfect equalization, which is a central result. Further, we characterize the channels which lead to a (nongeneric) problem of nonuniqueness of the cost minimization and we develop a means to avoid such cases by generalizing the single tap anchoring strategy to a linear constraint on the equalizer parameters. Finally, we demonstrate that truncated (finite dimensional) equalizers preserve arbitrarily closely the ideal global convergence properties.

The following references have been brought to our attention recently² [12]–[14]. In these, an approach to system identification using a prediction error estimate is presented. The cost function turns out to be the sup-norm of the prediction error. It is also proposed to approximate the sup-norm by the ℓ_p -norm for large p . The main difference between this and our approach is that we do not form any error estimate - instead our cost function is the sup-norm of the equalizer output itself, with the equalizer center tap anchored at 1. Also in our work we do not identify the system parameters explicitly; instead we find the inverse system.

II. MAIN RESULTS

A. Overview and Notation

The main question that we address is the following: What is the best we can do if we restrict our attention to cost functions

²We thank an anonymous referee who pointed out these references to us.

that are convex in the equalizer coefficients? This question is important because if convexity is imposed, truncated causal finite dimensional versions of the ideal noncausal infinite dimensional equalizers will also exhibit global convergence. This is due to the following: truncating the equalizer means setting some of the coefficients to zero. Any convex function has the property that if we linearly transform the argument, the function is still convex as a function of the new argument (in our case, the argument is the set of equalizer coefficients). This means that any convex function of the equalizer taps is also a convex function of the combined response coefficients of the channel and the equalizer and any truncated version thereof. Of course, when we use a truncated finite dimensional equalizer, it may leave some residual ISI; but that residual error can be made as small as desired by increasing the number of equalizer taps.

We will now introduce some notation and make some assumptions. Let the impulse response of the channel be parametrized by a causal l_1 sequence, that is, $\underline{c} \triangleq \{c_0, c_1, c_2 \dots\}$ with

$$\sum_{i=0}^{\infty} |c_i| < \infty.$$

We also assume that a stable (possibly noncausal) inverse of the channel exists, i.e., there exists $\underline{d} \triangleq \{\dots d_{-1}, d_0, d_1 \dots\}$ such that

$$\sum_{i=-\infty}^{\infty} |d_i| < \infty$$

where

$$c_n \otimes d_n = \delta_n$$

with \otimes denoting convolution and δ_n the Kronecker delta. These assumptions are satisfied, for instance, if the channel is a stable autoregressive moving average (ARMA) channel with no zeros on the unit circle.

Let the equalizer be parametrized as $\underline{\theta} \triangleq \{\dots \theta_{-1}, \theta_0, \theta_1 \dots\}$. In the algorithms that we consider we will impose a linear constraint on the equalizer to keep it away from the all-zero setup. From the previous discussion, it is clear that such a linear constraint preserves the convexity of the cost function. Let the combined impulse response of the channel and the equalizer be parametrized as $\underline{h} \triangleq \{\dots h_{-1}, h_0, h_1 \dots\}$. In other words

$$\underline{c} \otimes \underline{\theta} = \underline{h}$$

or

$$\underline{\theta} = \underline{d} \otimes \underline{h}.$$

We need that

$$h_n = G\delta_{n-j}$$

as the condition for perfect equalization for some unknown gain G and delay j . Since the cost function $\mathcal{J}(\cdot)$ is assumed to be convex in $\underline{\theta}$ and $\underline{\theta}$ is linear in \underline{h} , $\mathcal{J}(\cdot)$ is convex in \underline{h} as well. From now on we will write $\mathcal{J}(\cdot)$ to denote both the function of $\underline{\theta}$ and the function of \underline{h} .

B. Convex Cost Necessary Conditions

We can also assume that $\mathcal{J}(\cdot)$ is even and delay-insensitive, i.e., $\mathcal{J}(\underline{h}) = \mathcal{J}(-\underline{h})$ and $\mathcal{J}(\underline{h}^1) = \mathcal{J}(\underline{h}^2)$ if \underline{h}^1 and \underline{h}^2 are such that $h_n^1 = h_{n-k}^2$ for all n where k is some delay. These assumptions do not result in any loss of generality because the input process has been assumed to be stationary and its distribution symmetric.

We will state a simple result that holds for any such cost function. Let \underline{h} be a fixed vector. Let

$$g(x) \triangleq \mathcal{J}(x\underline{h}) \text{ for all } x \in \mathbb{R}.$$

Then $g(\cdot)$ is an even, convex, nonnegative function. Furthermore it follows easily by convexity that

$$g(x_1) \geq g(x_2) \text{ if } |x_1| \geq |x_2|. \quad (2.1)$$

It can also be shown that if $g(\cdot)$ is not a constant then

$$\lim_{x \rightarrow \infty} g(x) = \infty. \quad (2.2)$$

It is clear that the all-zero setup of the equalizer achieves the lowest cost. This points to the need for constraining the equalizer away from the all-zero setup. As in [3] we will do this by constraining θ_0 to be 1. In general, we could impose a linear constraint on the equalizer

$$\underline{a}'\underline{\theta} = 1 \quad (2.3)$$

where $\{a_i\}$ is a sequence with finite support. The results below are true for such a linear constraint as well.

We now impose the equalization condition: we want $\mathcal{J}(\cdot)$ to have its global minimum at an equalizer setup that results in $h_n = G\delta_{n-k}$, i.e., a valid equalizer, for a single delay k to be determined (this differs from the nonconvex cost approaches in [22,1] which admit multiple numbers of desirable minima corresponding to different delays). This results in another necessary condition:

Proposition 1: When we fix $\theta_0 = 1$, among all equalizers achieving perfect equalization, the one achieving the lowest cost is

$$\theta_n^* = d_{n-k}/d_{-k} \text{ for all } n$$

where k is such that

$$|d_{-k}| = \max_{i \in \mathbb{Z}} |d_i|.$$

Proof: It is clear such a maximum exists because \underline{d} is absolutely summable. Moreover, any sequence $\underline{\theta}^1$ that results in perfect equalization with $\theta_0^1 = 1$ is such that

$$\theta_n^1 = d_{n-j}/d_{-j} \text{ for some } j.$$

Therefore

$$\mathcal{J}(\underline{\theta}^1) = \mathcal{J}(\underline{d}/d_{-j}) \text{ (delay insensitivity)}$$

whereas

$$\mathcal{J}(\underline{\theta}^*) = \mathcal{J}(\underline{d}/d_{-k}).$$

Using the fact that $|d_{-k}| \geq |d_{-j}|$ and applying (2.1) we get

$$\mathcal{J}(\underline{\theta}^*) \leq \mathcal{J}(\underline{\theta}^1)$$

thus proving the desired result.

We will now derive another necessary condition and then exhibit a valid cost function that satisfies all these necessary conditions. Given arbitrary i and j with $i \neq j$ we can find a channel such that the following nonuniqueness property is true: The inverse of the channel, parametrized by \underline{d} , satisfies the nongeneric condition

$$|d_{-i}| = |d_{-j}| = \max_{n \in \mathcal{Z}} |d_n|.$$

Then for such a channel, the above proposition indicates that, $\underline{\theta}^1$ and $\underline{\theta}^2$ where

$$\theta_n^1 = d_{n-i}/d_{-i} \text{ and } \theta_n^2 = d_{n-j}/d_{-j}$$

achieve the minimum cost and due to convexity any equalizer on the line joining these two also achieves the minimum. We can rephrase this in terms of \underline{h} as follows. Let \underline{h}^1 correspond to $\underline{\theta}^1$ and \underline{h}^2 correspond to $\underline{\theta}^2$, i.e.

$$h_n^1 = \delta_{n-i}/d_{-i} \text{ and } h_n^2 = \delta_{n-j}/d_{-j}.$$

Then, for any $0 \leq a \leq 1$

$$\mathcal{J}(ah^1 + (1-a)h^2) = \mathcal{J}(h^1) = \mathcal{J}(h^2).$$

The right-hand side is independent of a and this shows that $\mathcal{J}(\cdot)$ is only a function of $|h_{-i}| + |h_{-j}|$. This is true for any i and j . Hence, when all the arguments of $\mathcal{J}(\cdot)$ except any two are zero, $\mathcal{J}(\cdot)$ depends on those two arguments only through the sum of their absolute values. The foregoing reasoning can be extended to replace two by any finite positive integer.

These are all necessary conditions resulting from the assumptions of convexity, evenness and delay insensitivity. The only constraining assumption is convexity. The last condition we have derived is fairly restrictive and the simplest cost function satisfying this is the l_1 norm of the combined impulse response. We will show that this is indeed a valid cost function.

Proposition 2: The cost function $\mathcal{J}(\underline{\theta}) = \mathcal{J}(\underline{h}) = \sum_{i=-\infty}^{\infty} |h_i|$ with $\theta_0 \equiv 1$ achieves its global minimum at $\underline{\theta}^*$ such that

$$\theta_n^* = d_{n-k}/d_{-k}$$

where $|d_{-k}| = \max_{i \in \mathcal{Z}} |d_i|$. The global minimum is unique if and only if the maximum of $\{|d_i|\}$ is attained at only one point.

Proof: First of all note that this cost function is convex, even and delay insensitive. Convexity follows from the fact that the absolute value function is convex, i.e., $q(x) = |x|$ is convex and the sum of any number of convex functions is also convex [27]. It is clear that $\theta_0^* = 1$. Let $A \triangleq 1/d_{-k}$. Then, $c_n \otimes \theta_n^* = Ad_{n-k}$. Now consider any $\underline{\theta}$ which satisfies $\theta_0 = 1$. Let $\underline{\mu} \triangleq \underline{\theta} - \underline{\theta}^*$. Note that $\mu_0 = 0$. Let $c_n \otimes \mu_n = \epsilon_n$. Then, from $c_n \otimes d_n = \delta_n$ we have

$$\epsilon_n \otimes \theta_n^* = c_n \otimes \mu_n \otimes Ad_{n-k} = A\mu_{n-k}.$$

Then

$$\begin{aligned} \mathcal{J}(\underline{\theta}) - \mathcal{J}(\underline{\theta}^*) &= \sum_i |c_i \otimes \theta_i| - \sum_i |c_i \otimes \theta_i^*| \\ &= \sum_i |c_i \otimes \mu_i + c_i \otimes \theta_i^*| - \sum_i |c_i \otimes \theta_i^*| \\ &= \sum_{i \neq k} |\epsilon_i| + |\epsilon_k + A| - |A| \\ &\geq \text{sgn}(A) \left[\epsilon_k + \sum_{i \neq k} \epsilon_i \theta_{k-i}^* \right] = \text{sgn}(A) \left[\epsilon_k \otimes \theta_k^* \right] \\ &= |A| \mu_0 = 0. \end{aligned}$$

For the last inequality we have used the following facts

Fact A: $|\epsilon_k + A| - |A| \geq \text{sgn}(A)\epsilon_k$

Fact B: $\sum_{i \neq k} |\epsilon_i| \geq \pm \sum_{i \neq k} \epsilon_i \theta_{k-i}^*$ because $|\theta_i^*| \leq 1$ for all i .

Thus we have $\mathcal{J}(\underline{\theta}) - \mathcal{J}(\underline{\theta}^*) \geq 0$ for any $\underline{\theta}$, thus showing that $\underline{\theta}^*$ is a global minimum. It is clear that this minimum is not unique if the channel is such that there exist k and l with $k \neq l$ such that $|d_k| = |d_l| = \max_{i \in \mathcal{Z}} |d_i|$. The converse is also true, i.e., the minimum is unique if there exists a unique $|d_k|$ achieving the above maximum. This is because, in that case, Fact B holds with strict inequality ($|\theta_i^*| < 1$ for $i \neq 0$). Hence, $\mathcal{J}(\underline{\theta}) - \mathcal{J}(\underline{\theta}^*) > 0$ for any $\underline{\theta} \neq \underline{\theta}^*$, proving that $\underline{\theta}^*$ is the unique global minimum.

B. Problem Channels and Linear Constraint

The significance of the results in Section II-B is the following: We have shown that imposing convexity requires us to constrain the equalizer away from the all-zero setup. However, when we fix one of the coefficients, a necessary condition on the cost function results. In fact the same condition is true for any linear constraint on the equalizer as in (2.3). Then we exhibited a valid cost function satisfying this necessary condition. However, when we implement this cost function with a particular linear constraint on the equalizer, there will exist nongeneric "problem channels," i.e., channels for which the algorithm will converge to a nonequalized setting. To see what we mean by nongeneric consider the case of $\theta_0 = 1$. From the uniqueness part of Proposition 3 it is clear that those channels for which the sequence of inverse coefficients has nonunique maximum in absolute value are precisely the ones for which the cost function will not work with the constraint $\theta_0 = 1$. It can be argued that such channels will be rare in practice. Moreover, it is easy to take care of such problem channels by passing the channel output through a linear filter before equalization. In effect, we are changing the channel. Alternatively, we could impose a linear constraint other than $\theta_0 = 1$, e.g., as in (2.3). In either case we will create a different set of problem channels and this situation can be handled by having a bank of equalizers, each with a different (possibly randomized) linear constraint.

There is a more elegant formulation of the above approach which lends itself to easy implementation. Let us consider the equalizer constrained as in (2.3). To see how to adapt an equalizer $\underline{\theta}$ subject to (2.3), let $\underline{\theta}$ be any solution of (2.3) and

let the columns of matrix V span the space orthogonal to \underline{d} . Then

$$\underline{\theta} = \hat{\underline{\theta}} + V\underline{\alpha} \quad (2.4)$$

with $\underline{\alpha}$ representing the vector of adjustable parameters. The recursive update law for $\underline{\alpha}$ in (2.4) becomes

$$\begin{aligned} \underline{\alpha}^{k+1} &= \underline{\alpha}^k - \mu \nabla_{\alpha} f(Y_k(\alpha)) \\ &= \underline{\alpha}^k - \mu f'(Y_k) \nabla_{\alpha} Y_k(\alpha) \end{aligned}$$

where Y_k is the equalizer output. We can evaluate the gradient vector in this equation by multiplying the vector of channel outputs by the matrix V . The $\underline{\alpha}$ is passed back to (2.4) to determine the actual equalizer parameter setting.

C. Truncated Equalizers

We now address the issue of truncated equalizer parametrization. Through a simple argument using the continuity of the cost function in the equalizer coefficients, we show that if the number of equalizer taps is sufficiently large, the minimum of the cost function occurs close to the minimum obtained in the case when the number of equalizer taps is infinite. We then consider a FIR channel coupled with a FIR equalizer. Coupled with the fact that global convergence is guaranteed due to convexity, this example illustrates that if our channel is such that its inverse can be adequately approximated with some number of coefficients, we need to parametrize the equalizer only with that many taps. This is a very desirable feature, especially in view of the fact that for existing nonconvex cost function approaches global convergence is not guaranteed even when the equalizer is parametrized with no excess nor deficit [9].

Let i_0 be such that

$$|d_{i_0}| = \max_{i \in Z} |d_i|$$

where $\{d_i\}$ is the channel inverse sequence. Also assume that such i_0 is unique. Fix $\epsilon > 0$. Suppose the equalizer is parametrized by $2N + 1$ taps such that

$$\sum_{i \notin \{i_0 - N, \dots, i_0 + N\}} |d_i| \leq \epsilon \sum_{i \in Z} |d_i| = \epsilon D. \quad (2.5)$$

Since by assumption, the channel inverse sequence $\{d_i\}$ is absolutely summable, we can always find such a N . Let us parametrize the equalizer with N anticausal taps and $N + 1$ causal taps with the central tap $\theta_0 = 1$. We will show that, under these assumptions the minimum of the cost function occurs "close" to the absolute minimum possible when the equalizer is doubly infinite (provided ϵ is chosen small). Note that the absolute minimum of the cost corresponds to perfect equalization. Define

$$\theta_i^* \triangleq \frac{d_{i+i_0}}{d_{i_0}} \text{ for all } i$$

which realizes the minimum cost \mathcal{J}^* , and a truncated version

$$\theta_i^N \triangleq \begin{cases} \frac{d_{i+i_0}}{d_{i_0}} & \text{for } i = -N, \dots, N \\ 0 & \text{otherwise.} \end{cases}$$

The cost corresponding to $\underline{\theta}^N$ is

$$\begin{aligned} \mathcal{J}^N &\triangleq \sum_{i \in Z} |c_i \otimes \theta_i^N| = \sum_{i \in Z} |c_i \otimes \theta_i^* - c_i \otimes (\theta_i^* - \theta_i^N)| \\ &\leq \sum_{i \in Z} |c_i \otimes \theta_i^*| + \sum_{k \in Z} \sum_{i \notin \{-N, \dots, N\}} |c_{k-i} \theta_i^*| \\ &\leq \mathcal{J}^* + \sum_{k \in Z} \sum_{i \notin \{-N, \dots, N\}} |c_{k-i} \theta_i^*| \\ &= \mathcal{J}^* + \sum_{i \notin \{-N, N\}} |\theta_i^*| \sum_{k \in Z} |c_{k-i}| \\ &\leq \mathcal{J}^* + \frac{CD\epsilon}{|d_{i_0}|} \text{ (using (2.4))} \end{aligned}$$

where $C = \sum_{j \in Z} |c_j| < \infty$ by assumption. Hence, by decreasing ϵ (corresponding to increasing N) we can force \mathcal{J}^N to be as close to \mathcal{J}^* , thereby achieving near perfect equalization.

We will now illustrate the truncated equalizer case with an example. We choose a 10-tap FIR impulse response

$$[0.04, -0.05, 0.07, -0.21, -0.5, 0.72, 0.36, 0.21, 0.03, 0.07].$$

This example is from Proakis [16] and is typical of the response of a data-quality telephone channel. Note that this channel has no spectral nulls. It introduces substantial ISI and the eye is initially not open. We use a 13-tap equalizer with the center tap anchored to 1. The minimization of the l_1 norm of the joint response under a linear constraint is a linear programming problem (see, for example, [23]). The resulting equalizer is (approximately) (see equation at the bottom of this page). The normalized ISI defined by

$$ISI = \frac{1}{|h_{n_0}|} \sum_{n \neq n_0} |h_n| \quad (2.6)$$

where $|h_{n_0}|$ refers to the maximum in absolute value of the sequence $\{\dots h_{-1} h_0 h_1 \dots\}$, initially is 2.139. The ISI after equalization is 0.0823. As we can see, the ISI is considerably reduced. For comparison we also computed the equalizer through the least squares approach with the center tap fixed at 1. In other words we minimize the l_2 norm of the combined response of the channel and equalizer subject to the center tap constraint on the equalizer. The resulting equalizer is at the bottom of this page. The resulting ISI is 3.17. The

$$[0.1010, 0.0523, 0.2129, 0.1061, 0.4125, 0.2833, 1.000, -0.5879, -0.019, 0.082, -0.069, 0.061, -0.011].$$

reason for this is that this channel is not minimum phase and so the least squares criterion results in converting the channel-equalizer combination into an allpass filter. Later we use the same channel and a 13-tap equalizer to illustrate the stochastic approximation procedure to recursively update the equalizer.

D. Implementation

How do we implement the blind equalizer adaptive law based on the l_1 cost function we have proposed? To answer this, note that we can write $Y_k = h_k \otimes X_k$ and so

$$\sum_{i=-\infty}^{\infty} |h_i| = M^{-1} \sup_{\omega \in \Omega} |Y_k| = M^{-1} \lim_{p \rightarrow \infty} \{E|Y_k|^p\}^{1/p} \quad (2.7)$$

where $M = \max \{X_k\}$. The second equality is true if the input random variables $\{X_k\}$ are bounded and iid. However, (2.7) also holds for certain classes of correlated inputs X_k and this is important in the instances of encoded data. Even though this cost function is not of the form specified in (1.1), it is a limiting form of functions of that type. An approximate cost function which has the desired properties is obtained by taking $\mathcal{J}_p(\theta) = E|Y_k|^p$ for p large. In other words, we are approximating the l_∞ norm of the equalizer output with the l_p norm for high p . Note that $\mathcal{J}_p(\cdot)$ is convex, even and delay insensitive and therefore preserves the global (unimodal) convergence properties. Furthermore, as $p \rightarrow \infty$, the minimum (or minima) of $\mathcal{J}_p(\cdot)$ approaches that of $\mathcal{J}(\cdot)$ for any fixed linear constraint on the equalizer coefficients. In fact, the cost functions $\mathcal{J}_p(\cdot)$ for $p \geq 2$ all have a global minimum at the valid equalizer if we constrain the equalizer such that $h_0 = 1$. (cf. [17]). Moreover, this is of little use since such a constraint needs the knowledge of the channel impulse response which we do not have. This has an interesting geometrical interpretation: Any linear constraint on the equalizer taps defines a manifold in the equalizer tap space. The functions $\mathcal{J}_p(\cdot)$ have their minimum at the valid equalizer if we look at the slice of the function defined on the manifold $h_0 = 1$. But in other manifolds corresponding to other linear constraints such as $\theta_0 = 1$, they do not achieve the minimum at the valid equalizer. In this class of cost functions only the l_∞ overcomes this problem. We would like to add in passing that if we have a fast estimator of h_0 , (either explicitly, or through an estimate of the channel impulse response coefficients) we could combine that with, say, the l_2 cost function and use the $h_0 = 1$ constraint. Such an estimator has to deal with nonminimum phase channels. The l_1 cost function is the same as the ISI criterion function considered by Lucky [17], [18] at the inception of adaptive equalization. He considers fixing $h_0 = 1$. The initial work [17] was based on sending a sequence of test pulses to estimate h_0 . In the second work [18] a decision-directed technique is suggested in order to update

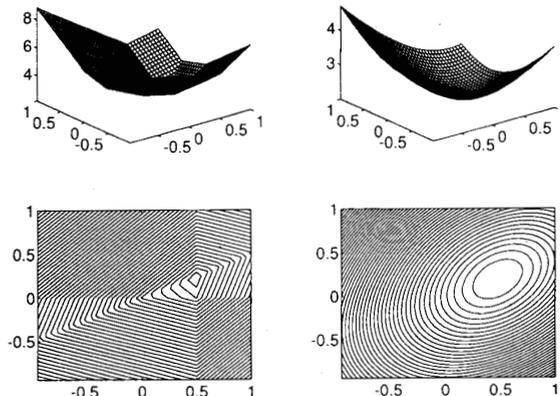


Fig. 2. The surfaces and the contours of the ideal cost function (left) and its l_8 approximation (right).

the equalizer based on such a constraint. Lucky also explicitly considers an FIR equalizer such that

$$\theta_i \equiv 0 \text{ if } i \notin K$$

where K is a finite subset of Z . A sufficient condition is derived in [17] which says that if the channel is such that the eye is initially open, i.e., $|h_0| > \sum_{i=1}^{\infty} |h_i|$, then the distortion criterion has its minimum at that equalizer setup which forces the combined impulse response coefficients to be zero within the set K , i.e., $h_i = 0$ for all $i \in K - \{0\}$. In general $h_i \neq 0$ for values of $i \notin K$. In fact, he illustrates with an example that if the channel is such that the eye is closed, it might happen that the equalizer converges to a configuration that increases the ISI and the problem becomes worse as we increase the number of equalizer taps.

Now we illustrate the approximation of the ideal cost function by the l_8 norm in Fig. 2. On a simple nonminimum channel, we plot the cost surfaces and the contours for the l_∞ cost function and the l_8 cost function. The channel chosen is $1 - 2z^{-1}$. The equalizer is parametrized with three taps of which the third tap is anchored at 1. At the top left is the surface of the ideal cost function and the one below is the contour map. The corresponding plots for the l_8 cost are on the right. From the contour maps we can see that the minima of the two cost functions are close to each other and they correspond to the equalizer setting $0.25z^2 + 0.5z + 1$. The actual stable anticausal inverse of the channel is $1 + 0.5z + 0.25z^2 + 0.125z^3 + \dots$ (left)

Now we carry out a local analysis of the optimum of the approximate cost function. We choose a causal minimum phase channel as an example, since this makes the analysis simpler. For this case, we note that the cost function EY_k^p achieves its global minimum at the exact equalizer. From now on we restrict p to be even for simplicity.

[0.0303, 0.0244, 0.0710, 0.0440, 0.1973, -0.0625, 1.000, -0.0625, 0.1973, 0.0440, 0.0710, 0.0244, 0.0303].

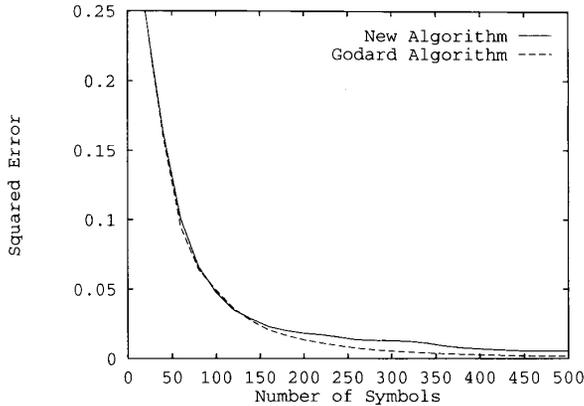


Fig. 3. Two tap channel: Equalizer convergence in the noiseless case.

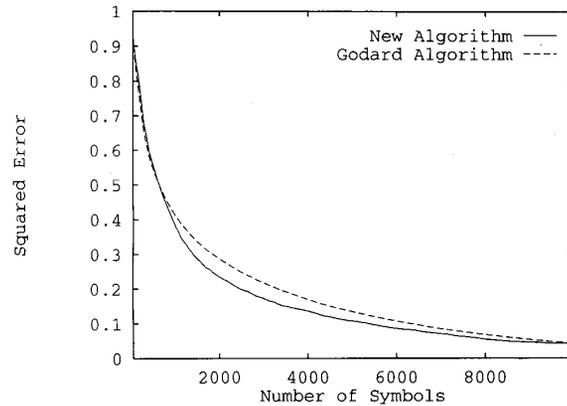


Fig. 4. Telephone channel: Equalizer convergence in the noiseless case.

We consider the equalizer $\hat{\theta}$ to consist of the exact equalizer θ^* plus a small perturbation $\epsilon\tilde{\theta}$, where $\tilde{\theta}$ is parametrized as $\{0, \tilde{\theta}_1, \tilde{\theta}_2, \dots\}$. In other words, we don't perturb the leading coefficient, which is anchored at 1. Now we can write the output of this equalizer Y_k as $X_k + \epsilon\tilde{Y}_k$ where X_k is the channel input data sequence and \tilde{Y}_k is the output of the cascade of the channel and the system parametrized by $\tilde{\theta}$. Under the assumption of iid data sequence, the cost can be expressed as

$$EY_k^p = E(X_k + \tilde{Y}_k)^p = \sum_{i=0; i \text{ even}}^p \binom{p}{i} \epsilon^i E\tilde{Y}_k^i EX_k^{p-i}.$$

For equally likely $\{+1, -1\}$ data sequence $EX_k^{p-i} = 1$, since both p and i are even. When ϵ is very small and p is moderate, the perturbation in the cost is dominated by the ϵ^2 term which has the coefficient $\binom{p}{2} E\tilde{Y}_k^2$. This gives us a sense of the local behavior of the cost function.

When p increases the cost increases, as is evident from the decomposition. In fact for larger p , the cost becomes steeper. This creates problems of stability. However in order to approximate the ideal cost well, we should not choose a value of p that is too small. A sensible strategy which we used in our simulations is to increase p as the adaptation progresses. Thus we prevent the equalizer from being stuck at a nonideal setup - which happens because we are not capturing the ideal cost function well when we use too low a value of p .

Some Monte-Carlo simulations were performed in order to verify the convergence properties. Two channels are used: one is the channel $1 - 2z^{-1}$ and the other is the more realistic telephone channel example from [16] which we used in Section II-D. Note that both channels are nonminimum phase and the eye diagram is closed in the second case. In both cases, we used a iid 8-level PAM data sequence. We use as performance measure the squared error between the ideal, equalized joint response—the delta function—and the joint response of the channel and the adapting equalizer. In other words, the performance measure is

$$\text{Squared Error} = \sum_{i=-\infty}^{\infty} h_i^2 - 1$$

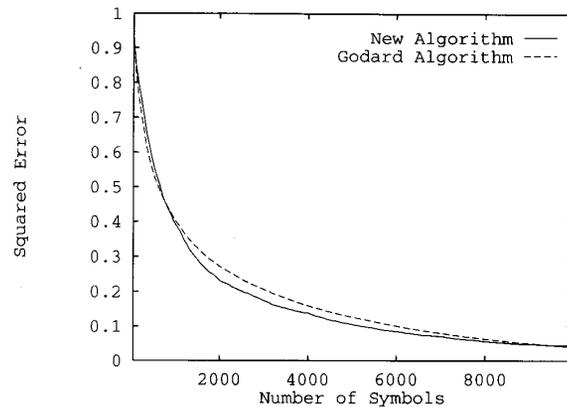


Fig. 5. Telephone channel: Equalizer convergence when SNR = 40 dB.

where we normalize the maximum in absolute value of h_i to be 1. For comparison, we have also plotted the Godard equalizer, though it has the problem of ill-convergence without suitable tap centering procedures.

Three plots are given (Figs. 3–5). The first two illustrate equalizer convergence in the absence of noise. For the telephone channel we also did a simulation with noise—with 40 db SNR. All the plots are averages of 20 Monte-Carlo runs. As the approximation to the ideal cost we use the p th moment EY_k^p . In the two tap channel case we switch from $p = 8$ to $p = 12$ after 300 symbols and stayed there. In the telephone channel case, we switch from $p = 8$ to $p = 12$ after 4000 symbols and then to $p = 14$ after 7000 symbols. For the telephone channel, the eye opens after about 4000 symbols. The plots shown are only for illustration of the convergence issue studied in this paper; no attempt was made to optimize the step sizes.

We also noted that if we use too high a value of p , noise would become a problem because the outliers are raised to a high power. In practice, it is desirable to choose a p that leads to good noise performance and opens the eye quickly. Simulations as well as contour plots indicate that p of the order 8–10 is a good tradeoff. In addition, it appears to be beneficial to combat the effect of noise by a saturation type

nonlinearity before the equalizer, in order to eliminate the outliers responsible for slowing convergence of the equalizer.

III. CONCLUSION

We have derived necessary and sufficient conditions for *global convergence* of the equalizer update algorithm regardless of the initial equalizer setup and also showed how to approximate the convex cost function. We have also treated the FIR equalizer. Convexity guarantees a unique global minimum (if the channel is not a nongeneric "problem channel") at a valid equalizer and hence a sufficiently long equalizer will reduce most of the ISI. We also indicated how to tackle the problem channels. Crucial from the view-point of implementation, we impose the linear constraint on the equalizer and not on the combined response. This obviates the need for decision-directed implementations and hence the need for sufficient conditions on the initial eye pattern to guarantee convergence.

The extension to the QAM case remains an important consideration. The theoretical demonstrations presented here do not extend easily to the complex case although it is straightforward to investigate algorithms based on the appropriate generalization of (2.7) through simulations. These simulations have proven encouraging [24].

REFERENCES

[1] 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.

[2] S. Verdú, "On the selection of memoryless adaptive laws for blind equalization in binary communications," in *Proc. 6th Int. Conf. Anal., Optimization Syst.* (Nice, France), June 1984, pp. 239-249.

[3] S. Verdú, B. D. O. Anderson, and R. A. Kennedy, "Anchored blind equalization," in *Proc. Conf. Inform. Sci., Syst.* (Baltimore, MD), Mar. 1991, pp. 774-779.

[4] S. Verdú, B. D. O. Anderson, and R. A. Kennedy, "Blind equalization without gain identification," *IEEE Trans. Inform. Theory*, pp. 292-297, Jan. 1993.

[5] G. J. Foschini, "Equalization without altering or detecting data," *Bell Syst. Tech. J.*, pp. 1885-1911, Oct. 1985.

[6] N. Amitay and L. Greenstein, Patent no. 4,606,054.

[7] C. R. Johnson Jr., "Admissibility in blind adaptive channel equalization," *IEEE Contr. Syst. Mag.*, Mar. 91.

[8] O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of nonminimum phase systems (channels)," *IEEE Trans. Inform. Theory*, vol. 36, no. 2, pp. 312-321, Mar. 1990.

[9] Z. Ding, C. R. Johnson, and R. A. Kennedy, "On the (non)existence of undesirable equilibria of Godard blind equalizers," *IEEE Trans. Signal Processing*, pp. 2425-2432, Oct. 1992.

[10] Z. Ding, C. R. Johnson Jr., R. A. Kennedy, and B. D. O. Anderson, "On the ill-convergence of Godard blind equalizers in data communications," in *Proc. 23rd Conf. Inform. Sci. Syst.* (Baltimore, MD), Mar. 1989, pp. 538-543.

[11] A. Benveniste, M. Goursat, and G. Ruget, "Robust identification of a nonminimum phase system: Blind adjustment of a linear equalizer in data communications," *IEEE Trans. Automat. Contr.*, vol. AC-25, pp. 385-399, June 1980.

[12] L. Ljung, *System Identification: Theory for the User*. Englewood Cliffs, NJ: 1987, p. 201.

[13] M. Milanese and R. Tempo, "Optimal algorithms theory for robust estimation and prediction," *IEEE Trans. Automat. Contr.*, vol. AC-30, pp. 730-738, 1985.

[14] M. Milanese, R. Tempo, and A. Vicino, "Strongly optimal algorithms and optimal information in estimation problems," *J. Complexity*, vol. 2, pp. 78-94, 1986.

[15] G. B. Giannakis, "On the identifiability of non-Gaussian ARMA models using cumulants," *IEEE Trans. Automat. Contr.*, vol. 35, pp. 18-26, Jan. 1990.

[16] J. G. Proakis, *Digital Communications*, 2nd ed. New York: McGraw-Hill, 1982, pp. 520-590.

[17] R. W. Lucky, "Automatic equalization for digital communication," *Bell Syst. Tech. J.*, pp. 547-588, Apr. 1965.

[18] R. W. Lucky, "Techniques for adaptive equalization of digital communications systems," *Bell Syst. Tech. J.*, pp. 255-286, Feb. 1966.

[19] B. Friedlander and B. Porat, "Adaptive IIR algorithms based on high-order statistics," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no. 4, pp. 485-495, Apr. 1989.

[20] S. Bellini and F. Rocca, "Blind deconvolution: Polyspectra or bussgang techniques?" in *Digital Communications*, E. Bilgieri and G. Prati, Eds. North-Holland: Elsevier, 1986, pp. 251-263.

[21] R. A. Kennedy, B. D. O. Anderson, Z. Ding, and C. R. Johnson Jr., "Local stable minima of the sato recursive identification scheme," in *Proc. 29th Conf. Decision, Contr.* (Honolulu, HI), Dec. 1990, pp. 3194-3199.

[22] Y. Sato, "A Method of self-recovering equalization for multilevel amplitude modulation," *IEEE Trans. Commun.*, vol. COM-29, pp. 679-683, June 1975.

[23] F. S. Hillier and G. J. Lieberman, *Introduction to Operations Research*. San Francisco, CA: Holden-Day, 1980, pp. 172-173.

[24] Z. Ding, "Joint blind equalization and carrier recovery of QAM systems," in *Proc. 25th Conf. Inform. Sci., Syst.* (Baltimore, MD), Mar. 1991, pp. 786-791.

[25] W. Rupprecht, "Adaptive equalization of binary NRZ-signals by means of peak value minimization," in *Proc. 7th European Conf. Circuit Theory, Design* (Prague, Hungary), June 1985, pp. 352-355.

[26] W. Rupprecht, "Analysis of the mean-absolute-deviation-error for adaptive equalizers in data transmission systems," in *Proc. Int. Symp. Appl. Signal Processing, Digital Filtering* (Paris, France), May 1985, pp. 99-102.

[27] W. Rupprecht, *Orthogonalfilter und adaptive Daten-Signalentzerrung*. München, Wien: R. Oldenbourg Verlag, 1987.



Sridhar Vembu (S'90) was born in Umayalpuram, Tamil Nadu, India, on June 28, 1968. He received the B.Tech. in electrical engineering from the Indian Institute of Technology, Madras, in July 1989, and the M.A. in electrical engineering from Princeton University, Princeton, NJ, in October 1990, where he is currently working toward the Ph.D. degree.

His research interests are in communications, signal processing, and information theory.

Mr. Vembu was awarded the 1991 IEEE Communications Society Scholarship and was supported

by an IBM Graduate Fellowship in 1991-1992 and 1992-1993.



Sergio Verdú (S'80-M'84-SM'88-F'93) received the Ph.D. degree in electrical engineering from the University of Illinois at Urbana-Champaign in 1984.

In 1984, he joined the Faculty of Princeton University, Princeton, NJ, where he is a Professor of electrical engineering. His research interests are in information theory and multiuser communication.

Dr. Verdú was a recipient of the NSF Presidential Young Investigator Award and the IEEE Donald Fink Paper Award. He has served as Associate

Editor for the Shannon Theory of the IEEE TRANSACTIONS ON INFORMATION THEORY and is currently a member of the Board of Governors of the IEEE Information Theory Society.

Rodney A. Kennedy was born in Sydney, Australia, on October 23, 1960. He received the B.E.(Hons.) degree in electrical engineering from the University of New South Wales, Australia, in 1982. He received the M.E. degree in digital control theory from the University of Newcastle, Australia, in 1986. He received the Ph.D. degree in December 1988 from the Department of Systems Engineering, The Australian National University, Canberra, in the area of digital communications.

From 1983 to 1986, he was with the CSIRO Division of Radiophysics, Sydney, Australia, working on antenna design and control problems. He has held a QEII Fellowship in the Department of Systems Engineering at The Australian National University between 1990 and 1993. Currently, he is a Senior Lecturer in the Department of Engineering in the Faculty of Engineering and Information Technology at The Australian National University. His research interests are in the fields of digital communications and signal processing.

Dr. Kennedy is currently an Editor of the *International Journal of Adaptive Control and Signal Processing*.



William Sethares received the B.A. degree in mathematics from Brandeis University, Waltham, MA, and the M.S. and Ph.D. degrees in electrical engineering from Cornell University, Ithaca, NY.

He has worked at the Raytheon Company as a Systems Engineer and is currently on the faculty of the Department of Electrical and Computer Engineering at the University of Wisconsin, Madison. His research interests include adaptive systems in signal processing, communications, control, and electronic music.