ON THE SELECTION OF MEMORYLESS ADAPTIVE LAWS FOR BLIND EQUALIZATION IN BINARY COMMUNICATIONS

Sergio Verdú
Coordinated Science Laboratory
University of Illinois at Urbana-Champaign
Urbana, IL 61801 USA

ABSTRACT

We consider the adaptive equalization of an unknown linear time-invariant channel without observations of the input sequence, by updating the impulse response coefficients of the equalizer with the output of the channel times a memoryless non-linear function of the equalizer output. To date, no such function is known to result in global convergence to the inverse of the channel when the input consists of binary data. The effect of the selection of the memoryless nonlinearity in the convergence properties of the adaptive scheme is studied, and it is shown that for a large class of laws (including the continuous functions), unequalized impulse responses with few nonzero coefficients are points of convergence, and that there exist undesired local minima for a subset of functions that includes those previously proposed.

I. PRELIMINARIES

Noiseless observations, $\underline{x} = \{x_t, t \in \mathbf{Z}\}$, of the output of an unknown discrete-time linear time-invariant system driven by a Bernoulli sequence $\underline{u} = \{u_t, t \in \mathbf{Z}\}$ are used in order to adjust the impulse response of an (IIR) linear equalizer $\theta = \{\theta_t, t \in \mathbf{Z}\}$ such that when driven by $\underline{x}$ its output is the original Bernoulli sequence. A possible strategy is to use a memoryless adaptive law of the type

$$\underline{\theta}^{n+1} = \underline{\theta}^{n} - \tau_n \psi(c_n) \underline{x}^n \tag{1}$$

where τ_n is a sequence of small positive scalars, c_n is the current output of the equalizer, $\underline{x}^n$ is an n-delayed version of $\underline{x}$ and $\psi(\cdot)$ is a real-valued function to be specified.

Use of the scheme (1) for the blind equalization problem has been made in the past in [1]-[3], and it is based on the fact that (1) is a stochastic approximation [8] for the minimization of the risk $R(\underline{\theta}) = E[\Omega(c)]$, where $\Omega(x) = \Omega(0) + \int_0^x \psi(t)dt$.

As popularized by Ljung [9], the analysis of the convergence of (1) can be carried out, if ψ is smooth enough, by first studying the steepest descent lines of $R(\underline{t}) = E[\Omega(\underline{u}\,\underline{t})] = \sum_{i=-\infty}^{\infty} t_i u_{-i}$, and then analyzing the behavior of (1) with respect to that of the integral curves of

$$\frac{d\underline{t}}{dx} = -\operatorname{grad} R(\underline{t}). \tag{2}$$

Ideally, the designer's goal would be to select an adaptive law $\psi(\cdot)$ such that all the steepest descent lines of $R(\underline{t})$ converge to $(\ldots,0,\pm 1,0,\ldots)$ - a sign uncertainty is inevitable since the input distribution is symmetric. However, to date, no function $\psi(\cdot)$ has been found to satisfy such a requirement. In fact, since the existence of steepest descent lines that converge to unstable equilibria (crest lines) may be a source of loss of efficiency of the algorithm but may not destroy the qualitative convergence properties of (1), we could require an adaptive law with less stringent properties:

Definition 1

ψ is *admissible* if the steepest descent lines of $R(\underline{t})$ converge and its local minima belong to the set $\{\underline{t} \in \mathbb{R}^{\mathbb{Z}}, \exists\, k\neq 0,\ \underline{t} = (\ldots,0,K,0,\ldots)\} \neq \phi$.

In their seminal work [1], Benveniste, Goursat and Ruget showed that if the distribution (λ) of the i.i.d. input sequence is sub-Gaussian rather than Bernoulli and ψ satisfies:

(a) $\psi(x) = -\gamma\,\mathrm{sgn}(x) + \tilde{\psi}(x)$

(b) $\int x\tilde{\psi}(x)\lambda(dx) = \gamma \int |x|\lambda(dx)$

(c) $\tilde{\psi}$ is odd, twice differentiable and convex on $(0,\infty)$.

Then ψ is an admissible adaptive law, and the only local minima are $(\ldots,0,\pm 1,0,\ldots)$. However, the existence of an admissible adaptive law when the input is Bernoulli - the most important case in practice - remains an open problem. As a step in that direction, the goal of this work is to impose conditions on $\psi(\cdot)$ such that an adaptive scheme with the above desirable properties - admissibility (and inexistence of crest lines) - can be obtained. The nature of the input distribution allows us to work in a simple algebraic framework, and to rule out large classes of functions by studying the behavior of the risk around systems with few nonzero coefficients. On the other hand, only restricted types of infinite impulse responses can be dealt with since a general characterization of the distribution of $\underline{u}\,\underline{t}$ in the Bernoulli case is not known [6].

II. DEVELOPMENT

Without significant loss of generality we restrict our attention to laws that satisfy:

(a) Integrable in any finite interval

(b) Left-hand and right-hand limits

$$\psi^{+}(x) = \lim_{h\downarrow 0} \psi(x+h)$$

$$\psi^{-}(x) = \lim_{h\downarrow 0} \psi(x-h)$$

exist at every point (i.e., only discontinuities of the first kind are allowed).

(c) ψ is odd. (Note that since u_t has symmetric distribution, any function Ω results in the same risk as the even function $\Omega(x) + \Omega(-x)$.)

Conditions (a)-(b) ensure the existence of the left-hand and right-hand derivatives of $\Omega(z)$, which is necessary for the analysis of the steepest descent minimization of $R(\underline{t})$. Let the (unnormalized) Gateaux directional derivative of $R(\cdot)$ at point $\underline{t}$ in the direction $\underline{\delta}$ be defined by

$$\rho(\underline{t},\underline{\delta}) = \lim_{h\downarrow 0} \frac{1}{h} [R(\underline{t} + h\underline{\delta}) - R(\underline{t})] \quad . \tag{3}$$

The steepest descent algorithm [7] for the minimization of $R(\underline{t})$ selects at each point the direction at which $\rho(\underline{t},\underline{\delta})/\|\delta\|$ - for some norm $\| \ \|$ - is minimum. If such a minimum exists and is negative then a new approximation to the sought-after global minimum of $R(\underline{t})$ is obtained. Hence the sinks of the steepest descent lines have the following properties:

Definition 2

$\underline{t}$ is a *point of convergence* if

(a) for all $\underline{\delta}$, $\rho(\underline{t},\underline{\delta}) \geqslant 0$

(b) for every neighborhood $N(\underline{t})$ there exists $\underline{t}' \in N(\underline{t})$ such that $R(\underline{t}) < R(\underline{t}')$.

Our first result will be invoked again and again in the sequel and reduces the fulfillment of condition (a) in the last definition to a purely finite dimensional problem when - as will be the case later on - there is only a finite number of non-zero terms in $\underline{t}$.

Proposition 1

If for every $\underline{\delta}$ such that $t_i = 0 \Rightarrow \delta_i = 0$ we have

$$\rho(\underline{t},\underline{\delta}) \geqslant 0 \quad , \tag{4}$$

then (4) holds for every direction $\underline{\delta}$ such that the directional derivative $\rho(\underline{t},\underline{\delta})$ exists.

Proof

For an arbitrary $\underline{\delta}$, we have

$$\begin{aligned}\rho(\underline{t},\underline{\delta}) &= \lim_{h\downarrow 0} \frac{1}{h} E[\Omega(\underline{u}(\underline{t}+h\underline{\delta})) - \Omega(\underline{u}\,\underline{t})] \\ &= \lim_{h\downarrow 0} \frac{1}{h} E[E[\Omega(\sum_{\substack{i\\ t_i=0}} h\delta_i u_i + \sum_{\substack{i\\ t_i\neq 0}} (t_i+h\delta_i)u_i) - \Omega(\underline{u}\,\underline{t}) \,|\, \{u_i, \text{s.t. } t_i=0\}]] \\ &= \lim_{h\downarrow 0} \frac{1}{h} E[E[\Omega(-\sum_{\substack{i\\ t_i=0}} h\delta_i u_i + \sum_{\substack{i\\ t_i\neq 0}} (t_i+h\delta_i)u_i) - \Omega(\underline{u}\,\underline{t}) \,|\, \{u_i, \text{s.t. } t_i=0\}]],\end{aligned}$$

where the last equation follows because Ω is even. Therefore it is enough to show that for any positive scalar ε,

$$\lim_{h\downarrow 0} \frac{1}{h} E[\Omega(h\varepsilon + \sum_{\substack{i\\ t_i\neq 0}} (t_i+h\delta_i)u_i) - \Omega(\underline{u}\,\underline{t})] \geqslant 0 \tag{5}$$

if $\rho(\underline{t},\underline{\delta}) \geqslant 0$ for every $\underline{\delta}$ such that $\delta_i = 0$ unless $t_i \neq 0$. Let $\tilde{\delta}$ be such that $\tilde{\delta}_i = \delta_i/\varepsilon$ if $t_i \neq 0$, and $\tilde{\delta}_i = 0$ if $t_i = 0$.

$$\lim_{h\downarrow 0} \frac{1}{\varepsilon h} E[\Omega(h\varepsilon + \sum_{\substack{i\\ t_i=0}} (t_i+h\delta_i)u_i) - \Omega(\underline{u}\,\underline{t})] = E[\psi^{+}(\underline{u}\,\underline{t})(1+\underline{u}\,\tilde{\underline{\delta}}) \,|\, -1 \leqslant \underline{u}\,\tilde{\underline{\delta}}] P[-1 \leqslant \underline{u}\,\tilde{\underline{\delta}}]$$

$$+ E[\psi^{-}(\underline{u}\,\underline{t})(1 + \underline{u}\,\tilde{\underline{\delta}}) \,|\, \underline{u}\,\tilde{\underline{\delta}} < -1] P[\underline{u}\,\tilde{\underline{\delta}} < -1] =$$

$$= E[\psi^+(\underline{u}\,\underline{t})\,(1+\underline{u}\,\tilde{\underline{\delta}})\,|-1 \leqslant \underline{u}\,\tilde{\underline{\delta}} \leqslant 1]P[-1\leqslant \underline{u}\,\tilde{\underline{\delta}} \leqslant 1]$$

$$+ E[\psi^+(\underline{u}\,\underline{t})\,(2\underline{u}\,\tilde{\underline{\delta}})\,|\,1 < \underline{u}\,\tilde{\underline{\delta}}]\,P[1< \underline{u}\,\tilde{\underline{\delta}}]$$

where we used $\psi^+(x) = -\psi^-(-x)$. Now we show that the right hand side of the last equation is $\frac{1}{2}\,[\rho(\underline{t},\tilde{\underline{\delta}}^+) + (\underline{t},\tilde{\underline{\delta}}^-)]$ where $\tilde{\delta}_i^{\pm} = \tilde{\delta}_i$ for $i \neq 0$ and $\tilde{\delta}_0^{\pm} = \tilde{\delta}_0 \pm 1$ (without loss of generality we may assume $\tilde{\delta}_0 \neq 0$).

$$\begin{aligned}\rho(\underline{t},\tilde{\underline{\delta}}^+) &= E[\psi^+(\underline{u}\,\underline{t})\,(\underline{u}\,\tilde{\underline{\delta}}+u_0)\,|\underline{u}\,\tilde{\underline{\delta}} >1]P[\underline{u}\,\tilde{\underline{\delta}}> 1]\\ &+ P[-1\leqslant\underline{u}\,\tilde{\underline{\delta}} \leqslant 1,u_0 = 1]E[\psi^+(\underline{u}\,\underline{t})(\underline{u}\,\tilde{\underline{\delta}} + 1)\,|-1\leqslant\underline{u}\,\tilde{\underline{\delta}} \leqslant 1,u_0 = 1]\\ &+ P[-1\leqslant\underline{u}\,\tilde{\underline{\delta}} \leqslant 1,u_0 =-1]E[\psi^-(\underline{u}\,\underline{t})(\underline{u}\,\tilde{\underline{\delta}} - 1)\,|-1\leqslant\underline{u}\,\tilde{\underline{\delta}} \leqslant 1,u_0 =-1]\\ &+ P[\underline{u}\,\tilde{\underline{\delta}} <1]E[\psi^-(\underline{u}\,\underline{t})(\underline{u}\,\tilde{\underline{\delta}} +u_0)\,|\,\underline{u}\,\tilde{\underline{\delta}}< 1]\\ &= 2\, E[\psi^+(\underline{u}\,\underline{t})(\underline{u}\,\tilde{\underline{\delta}} + u_0)\,|\underline{u}\,\tilde{\underline{\delta}} >1]P[\underline{u}\,\tilde{\underline{\delta}}< 1]\\ &+2\, E[\psi^+(\underline{u}\,\underline{t})(\underline{u}\,\tilde{\underline{\delta}} +1)\,|\,-1\leqslant\underline{u}\,\tilde{\underline{\delta}} \leqslant1,u_0 =1]P[-1\leqslant\underline{u}\,\tilde{\underline{\delta}} \leqslant1,u_0 =1] \quad .\end{aligned}$$

Analogously, we obtain

$$\begin{aligned}\rho(\underline{t},\tilde{\underline{\delta}}^-) &=2\, E[\psi^+(\underline{u}\,\underline{t})(\underline{u}\,\tilde{\underline{\delta}} - u_0)\,|\underline{u}\,\tilde{\underline{\delta}} >1]P[\underline{u}\,\tilde{\underline{\delta}}< 1]\\ &+ 2\, E[\psi^+(\underline{u}\,\underline{t})(\underline{u}\,\tilde{\underline{\delta}} + 1)\,|-1\leqslant\underline{u}\,\tilde{\underline{\delta}} \leqslant 1,u_0 = -1]P[-1\leqslant\underline{u}\,\tilde{\underline{\delta}} \leqslant 1,u_0 = -1].\end{aligned}$$

Therefore,

$$\begin{aligned}&E[\psi^+(\underline{u}\,\underline{t})\,(1 + \underline{u}\,\tilde{\underline{\delta}})\,|\,-1\leqslant\underline{u}\,\tilde{\underline{\delta}} \leqslant1]P[-1\leqslant\underline{u}\,\tilde{\underline{\delta}} \leqslant1]\\ &\quad+ E[\psi^+(\underline{u}\,\underline{t})(2\underline{u}\,\tilde{\underline{\delta}})\,|1<\underline{u}\,\tilde{\underline{\delta}}\,]P[1<\underline{u}\,\tilde{\underline{\delta}}\,]\\ &\quad= \tfrac{1}{2}\,[\rho(\underline{t},\tilde{\underline{\delta}}^+) + \rho(\underline{t},\tilde{\underline{\delta}}^-)]\geqslant0\end{aligned}$$

where the inequality follows from the assumption of the proposition since $\tilde{\delta}_i^+ = \tilde{\delta}_i^- = 0$ unless $t_i \neq 0$.

Next, our emphasis is on the search for necessary conditions on the adaptive law such that points of convergence that do not correspond to equalized systems are avoided. To that end, we focus attention on the behavior of the risk around systems with two or three nonzero coefficients with equal magnitude. (Recall that the risk is invariant under permutation or change of sign of the impulse response coefficients.)

Proposition 2

If ψ is admissible and continuous at the origin then there exists $c \neq 0$ such that $(\ldots,0,c,c,0,\ldots)$ is a point of convergence.

Proof

Let d be any nonzero scalar. Proposition 1 implies that in order to show

$$\rho((\ldots,0,d,d,0,\ldots),\underline{\delta}) \geqslant 0$$

for every $\underline{\delta}$ it is enough to restrict attention to $\underline{\delta} = (\ldots,0,\delta_1,\delta_2,0,\ldots)$. Hence

$$\rho((\ldots,0,d,d,0,\ldots),(\ldots,0,\delta_1,\delta_2,0,\ldots)) = \begin{cases} 2\psi^+(2d)(\delta_1 + \delta_2) \text{ if } \delta_1 + \delta_2 \geqslant 0 \\ 2\psi^-(2d)(\delta_1 + \delta_2) \text{ if } \delta_1 + \delta_2 \leqslant 0 \end{cases}$$

where the continuity at the origin has been used. Since ψ is admissible, there exists $c \neq 0$, such that $(\ldots,0,2c,0,\ldots)$ is a local minimum of $R(\underline{t})$ and therefore $\psi^-(2c) \leqslant 0 \leqslant \psi^+(2c)$. In order to show that $(\ldots,0,c,c,0,\ldots)$ is not a local maximum, it is enough to consider

$R((\dots,0,c+\varepsilon,c+\varepsilon,0,\dots)) - R((\dots,0,c,c,0,\dots)) = 2[\Omega(2c+2\varepsilon) - \Omega(2c)]$,
which is strictly positive for some ε in every neighborhood of 0, because 2c is a local minimum of $\Omega(\cdot)$. □

The requirement of having a discontinuity at the origin (in order to avoid the existence of points of convergence at unequalized systems) is fulfilled by all the adaptive laws proposed in [1]-[3]. However, as the following result shows, other discontinuities are necessary when the location of the stable points of convergence is restricted to $(\dots,0,\pm1,0,\dots)$.

Proposition 3

Suppose that ψ is admissible and that there exists $K > 0$ such that $(\dots,0,c,0,\dots)$ is not a local minimum of $R(\underline{t})$ if $|c| \neq K$. Then, there exists no $t \in [K/3,K]$ such that $(\dots,0,t,t,t,0,\dots)$ is a point of convergence if and only if $\psi^-(3K) + \psi^-(K) > 0$ and at least one of the following is satisfied.

(a) $\psi^+(K) + \psi^+(K/3) < 0$ and there exists $z \in (K/3,K)$ such that
 (a1) $\psi^+(3z) > \psi^-(3z)$
 (a2) $\psi^+(z) < \psi^-(z)$
 (a3) z is a strict local minimum of $\Omega(3\cdot) + 3\Omega(\cdot)$

(b) $\psi^+(K/3) < \psi^-(K/3)$ and for every $t \in (K/3,K)$ at least one of the following is true:
 (b1) $\psi^+(3t) + \psi^+(t) < 0$
 (b2) $\psi^-(3t) + \psi^-(t) > 0$
 (b3) $\psi^+(t) < \psi^-(t)$
 (b4) $(\dots,0,t,t,t,0,\dots)$ is a local maximum.

Proof

First we obtain necessary and sufficient conditions for

$$\rho((\dots,0,t,t,t,0,\dots),\underline{\delta}) \geq 0 \text{ for all } \underline{\delta}.$$

Invoking Proposition 1, we only need to take into account $\underline{\delta} = (\dots,0,\delta_1,\delta_2,\delta_3,0,\dots)$. Using the definition of ρ and taking expectation with respect to the input sequence, we obtain

$$4\rho((\dots,0,t,t,t,0,\dots),(\dots,0,\delta_1,\delta_2,\delta_3,0,\dots)) =$$

$$= \begin{cases} [\psi^+(3t)+\psi^+(t)](\delta_1+\delta_2+\delta_3), & \text{if } 0 \leq \delta_1,\delta_2,\delta_3 \\ [\psi^-(3t)+\psi^-(t)](\delta_1+\delta_2+\delta_3), & \text{if } 0 \geq \delta_1,\delta_2,\delta_3 \\ [\psi^+(3t)+\psi^+(t)](\delta_1+\delta_2+\delta_3) + [\psi^+(t)-\psi^-(t)](\delta_i-\delta_j-\delta_k), & \text{if } \delta_i \geq 0, \delta_i \geq |\delta_j+\delta_k| \\ [\psi^-(3t)+\psi^-(t)](\delta_1+\delta_2+\delta_3) + 2\delta_i[\psi^+(t)-\psi^-(t)] & \text{if } 0 \leq \delta_i \leq -\delta_j-\delta_k \\ [\psi^+(3t)+\psi^+(t)](\delta_1+\delta_2+\delta_3) - 2\delta_i[\psi^+(t)-\psi^-(t)] & \text{if } -\delta_j-\delta_k \leq \delta_i \leq 0 \\ [\psi^-(3t)+\psi^-(t)](\delta_1+\delta_2+\delta_3) + [\psi^+(t)-\psi^-(t)](-\delta_i+\delta_j+\delta_k), & \text{if } \delta_i \leq 0, \delta_i \leq |\delta_j+\delta_k| \end{cases}$$

From the last expression it is straightforward to check that for the directional derivative to be nonnegative in all directions it is necessary and sufficient that

$$\text{(i)} \quad \psi^+(3t) + \psi^+(t) \geq 0 \tag{6}$$

$$\text{(ii)} \quad \psi^-(3t) + \psi^-(t) \leq 0 \tag{7}$$

(iii) $\psi^+(t) - \psi^-(t) \geqslant 0$ (8)

Since all the steepest descent lines converge and there are no local minima of $\Omega(\cdot)$ other that $\pm K$, it is necessary that

(i) $\psi^+(x) \geqslant 0 \quad x \in [K,\infty)$ (9)

(ii) $\psi^-(x) \leqslant 0 \quad x \in [0,K]$ (10)

If $\psi^-(3K) + \psi^-(K) \leqslant 0$ then particularizing (9) at K and 3K and (10) at K, we obtain that conditions (6)-(8) are satisfied and that $(\ldots,0,K,K,K,0,\ldots)$ is not a local maximum. Hence, we obtain that $\psi^-(3K) + \psi^-(K) > 0$ is necessary and sufficient for the inexistence of points of convergence at $(\ldots,0,K,K,K,0,\ldots)$. Analogously, using (10) we have that $(\ldots,0,K/3,K/3,K/3,0,\ldots)$ is a point of convergence if and only if $\psi^+(K) + \psi^+(K/3) \geqslant 0$ and $\psi^+(K/3) \geqslant \psi^-(K/3)$. (It is easy to check that $(\ldots,0,K/3,K/3,K/3,0,\ldots)$ is not a local maximum, e.g., $R((\ldots,0,K/3,K/3,K/3,0,\ldots)) < R((\ldots,0,K/3-\varepsilon,K/3-\varepsilon,K/3-\varepsilon,0,\ldots))$ for sufficiently small $\varepsilon > 0$.)

In order to find out the conditions for existence of points of convergence at $(\ldots,0,t,t,t,0,\ldots)$ for $t \in (K/3,K)$ when they do not exist at $t = K/3,K$, we first consider the case:

(a) $\psi^-(3K) + \psi^-(K) > 0$ and $\psi^+(K) + \psi^+(K/3) < 0$.

These imply that there exists a strict local minimum (t) of $\Omega(3\cdot) + 3\Omega(\cdot)$ in the interval (K/3,K)-t is an upcrossing of $\psi(3\cdot) + \psi(\cdot)$-. On the one hand this implies that $\Omega(3t + 3\varepsilon) + 3\Omega(t+\varepsilon) > \Omega(3t) + 3\Omega(t)$ for either $\varepsilon > 0$ or $\varepsilon < 0$ sufficiently small. Therefore $(\ldots,0,t+\varepsilon,t+\varepsilon,t+\varepsilon,0,\ldots)$ has strictly higher risk than $(\ldots,0,t,t,t,0,\ldots)$. On the other hand, we have that $\psi^-(3t) + \psi^-(t) \leqslant 0 \leqslant \psi^+(3t) + \psi^+(t)$; so $(\ldots,t,t,t,0,\ldots)$ is a point of convergence if and only if $\psi^+(t) \geqslant \psi^-(t)$. (Note that condition (a1) in the statement of Proposition 3 follows from (a2) + (a3).) The alternative condition for avoiding points of convergence at $(\ldots,0,K/3,K/3,K/3,0,\ldots)$ is

(b) $\psi^-(3K) + \psi^-(K) > 0$ and $\psi^+(K/3) < \psi^-(K/3)$

Conditions (b1)-(b4) are necessary and sufficient for the inexistence of a point of convergence at $(\ldots,0,t,t,t,0,\ldots)$. It can be shown that for $(\ldots,0,t,t,t,0,\ldots)$ to be a local maximum such that (b1)-(b3) are not satisfied it is necessary that $\psi(\cdot)$ be continuous at t and 3t and that t be a downcrossing of $\psi(\cdot) + \psi(3\cdot)$. If, furthermore, $\psi(\cdot)$ is differentiable at 3t and t, then it is necessary and sufficient that $\psi'(t) = 0$ and $\psi'(3t) < 0$. □

Proposition 3 shows that the discontinuity of the adaptive law is necessary for the inexistence of points of convergence at systems $(\ldots,0,t,t,t,0,\ldots)$. This implies that even in the event that there exists an adaptive law without points of convergence at unequalized systems, in practice, the adaptive scheme will lack the necessary robustness for assuring a given speed of convergence. Note also that the above points of convergence occur at systems that are not far from the memoryless ones, and therefore central limit arguments for guaranteeing the behavior of the risk function cannot be used (cf. [1,Remark 4]).

A more crucial point than the inexistence of unequalized points of convergence is the admissibility of the adaptive law. Analogously to the last result, we find a large class of inadmissible functions by examining the systems with three equal-magnitude nonzero coefficients. Here, in order to simplify matters we restrict our attantion to adaptive laws whose derivative $\psi'(x)$ exists for $x \in (0,\infty)$.

Proposition 4

If there exists $t \in (0,\infty)$ such that

(i) $\psi(t) + \psi(3t) = 0$

(ii) $\psi'(t) + \psi'(3t) \geqslant 0$

(iii) $\psi'(t) \geqslant 0$

and at least one inequality is strict, then ψ is not admissible.

Proof

Under the above conditions there exists a local minimum at $(\ldots,0,t,t,t,0,\ldots)$. In order to show this, we prove that

$$\frac{\partial}{\partial t_i} R(\underline{t}) \Bigg|_{\underline{t} = (\ldots,0,t,t,t,0,\ldots)} = 0 \quad i = 1,2,3$$

if and only if (i) holds (Equations (6)-(8) reduce to (i) if ψ is continuous), and that

$$H = \left[\left(\frac{\partial^2 R(\underline{t})}{\partial t_i \partial t_j}\right)_{ij}\right] \geqslant 0, \quad \underline{t} = (\ldots,t,t,t,0,\ldots)$$

and $H \neq \underline{0}$ if and only if (ii) and (iii) are true. Notice that this is enough because of Proposition 1 and $\frac{\partial^2 R(\underline{t})}{\partial t_i \partial t_j} = 0$ at $\underline{t} = (\ldots,0,t,t,t,0,\ldots)$ if $t_i = 0$. In order to show the nonnegativity of the Hessian, we have that

$$4\,\frac{\partial^2 R(\underline{t})}{\partial t_i t_j}\Bigg|_{\underline{t} = (\ldots,0,t,t,t,0,\ldots)} = \begin{cases} \psi'(3t) + 3\psi'(t) & i = j \\ \psi'(3t) - \psi'(t) & i \neq j \end{cases}$$

and therefore ($|H| \geqslant 0$ if the first two principal minors are nonnegative) $H \geqslant 0$ if and only if

$$\psi'(3t) + 3\psi'(t) \geqslant 0$$

and

$$(\psi'(3t) + 3\psi'(t))^2 \geqslant (\psi'(3t) - \psi'(t))^2$$

which is equivalent to

$$\psi'(3t) + \psi'(t) \geqslant 0$$

and

$$\psi'(t) \geqslant 0. \qquad \square$$

Corollary

If $\psi'(t) > 0$ for every $t \in (0,\infty)$ then ψ is not admissible.

Proof

Suppose that ψ is admissible. Then since it is strictly increasing there exists a point K such that $\psi(x) < 0 < \psi(y)$ for $0 < x < K < y$. Now, using the fact that ψ is continuous, there exists a point $t \in (K/3,K)$ such that $\psi(3t) + \psi(t) = 0$. □

This corollary shows that no continuously differentiable convex - in $(0,\infty)$ - cost function $\Omega(\cdot)$ is admissible. This is particularly interesting since the adaptive laws considered by Godard [3] and Sato [2] (see also [1, Sect. VI]) are ruled out.

III. EXTENSIONS AND CONCLUDING REMARKS

A class of functions (including those adaptive laws proposed in the communications literature) has been shown to be inadmissible from the viewpoint of global convergence. Furthermore, a much larger class of laws (including the continuous functions) has shown to result in points of convergence at unequalized systems. All this has been accomplished by restricting attention to the behavior of the cost around systems whose impulse response have no more than three nonzero coefficients and hence avoiding the need to specify the underlying space of systems. It is plausible that by studying more complex systems much larger classes of functions can be excluded, and hopefully more light can be shed into the problem of the existence of an admissible memoryless adaptive law. Following the approach taken in Section II, the sufficient conditions for the existence of local minima (assuming differentiability of ψ, for example) could be investigated. While it is straightforward to show that if n equal-magnitude nonzero coefficients are allowed the gradient of the risk is zero for all directions at $(\ldots,0,t,\ldots,t,0,\ldots)$ if and only if

$$\psi(nt) + (n-2)\psi((n-2)t) + \sum_{i=2}^{\lceil \frac{n}{2}-1 \rceil} \left[\binom{n-2}{i}-\binom{n-2}{i-2}\right]\psi((n-2i)t) = 0 \quad ,$$

it is tedious to find conditions for the nonnegativity of the Hessian for generic n. On the other hand, when $\underline{t}$ has more than three or four nonzero coefficients it is equally tedious to obtain necessary and sufficient conditions for $\rho(\underline{t},\underline{\delta}) \geq 0$ for all $\underline{\delta}$ (cf. (6)-(8) for n = 3). However, it is possible to systematize the derivations of such conditions by using the following result.

Proposition 4

Let V be the set of all vertices of the simplices defined by subsets of the following hyperplanes in $\mathbb{R}^n$:

$$x_i \geq -1 \qquad i = 1,\ldots,n$$
$$x_i \leq 1 \qquad i = 1,\ldots,n$$
$$\underline{u}\,\underline{x} \geq 0 \qquad \underline{u} \in \{-1,1\}^n \quad .$$

If $\rho(\underline{t},\underline{v}) \geq 0$ for all $v \in V$, then $\rho(\underline{t},\underline{\delta}) \geq 0$ for all $\underline{\delta} \in \mathbb{R}^n$.

Proof

Select any $\underline{\delta} \in \mathbb{R}^n - \{\underline{0}\}$ and let $U_\delta = \underline{u} \in \{-1,1\}^n$, s.t. $\underline{u}\,\underline{\delta} \geq 0\}$, then $\tilde{\underline{\delta}} = \underline{\delta}/\|\underline{\delta}\|_\infty$ belongs to the simplex (with vertices $V_\delta \subset V$) defined by

$$S_\delta = \{\underline{x} \in \mathbb{R}^n, \|\underline{x}\|_\infty \leq 1\} \cap_{\underline{u} \in U_\delta} \{\underline{x},\underline{u}\,\underline{x} \geq 0\} \ ,$$

and therefore $\tilde{\underline{\delta}}$ can be put as a convex combination of the vertices in $V_\delta = \{\underline{v}^1,\ldots,\underline{v}^k\}$:

$$\underline{\tilde{\delta}} = \sum_{i=1}^{K} a_i \underline{v}^i, a_i \geq 0, \sum_{i=1}^{K} a_i = 1.$$

The proof is completed by showing that

$$\rho(\underline{t},\underline{\delta}) = \|\underline{\delta}\|_\infty \sum_{i=1}^{K} a_i \rho(\underline{t},\underline{v}^i).$$

To that end we have

$$(\underline{t},\underline{\delta})/\|\delta\|_\infty = \rho(\underline{t},\underline{\tilde{\delta}}) = 2\, E[\psi^+(\underline{u}\,\underline{t})\, \underline{u}\, \underline{\tilde{\delta}} \mid \underline{u}\, \underline{\delta} \geq 0] P[\underline{u}\, \underline{\delta} \geq 0]$$

$$= 2 \sum_{i=1}^{K} a_i E[\psi^+(\underline{u}\,\underline{t})\, \underline{u}\, \underline{v}^i | \underline{u}\, \underline{\delta} \geq 0] P[\underline{u}\, \underline{\delta} \geq 0]$$

$$= 2 \sum_{i=1}^{K} a_i E[\psi^+(\underline{u}\,\underline{t})\, \underline{u}\, \underline{v}^i | \underline{u}\, \underline{v}^i \geq 0] P[\underline{u}\, \underline{v}^i \geq 0]$$

$$= \sum_{i=1}^{K} a_i \rho(\underline{t},\underline{v}^i)$$

where we have used $\psi^+(x) = -\psi^-(-x)$, and the fact that both $\underline{v}^i$ and $\underline{\tilde{\delta}}$ belong to the simplex S_δ. □

The import of Proposition 4 is that when $\underline{t}$ has a finite number of nonzero components (recall Prop. 1) it is enough to restrict attention to a finite set of directions given by vertices of polytopes in the unit cube (note that any other norm defined by hyperplanes e.g. $\| \|_1$ would result in analogous conclusions) which can be generated systematically. Once the conditions corresponding to every vertex have been generated they can be reduced to a smaller linearly independent set.

Figure 1 illustrates the case n = 3 for which we derived conditions (6), (7), and (8). It turns out that these equations are generated by the vertices $\underline{v}^1$, $-\underline{v}^1$ and $\underline{v}^2$ respectively. In particular, note that not every vertex must be investigated because if $\underline{v} \in V$, then $\underline{w} \in V$ where $w_i = v_{p(i)}$ and p(i) is any bijective function on {1,...,n}, and $\underline{v}$ and $\underline{w}$ result in the same condition. The geometrical insight of this approach and its connections with linear programming could be exploited for obtaining an algorithm that generates efficiently necessary and sufficient conditions for the nonnegativity of the directional derivative of the risk.

Overall, it appears that the approach taken in this work is effective for showing the existence of unequalized points of convergence and the inadmissibility of classes of adaptive laws. Nevertheless, it seems that a substantially different approach is needed to prove or disprove the existence of admissible laws if a reasonably general space of impulse responses is allowed (note that requiring that the system and its inverse have finite energy - cf. [1] - may be too restrictive). Another point is that although the concept of admissibility used here is perhaps more realistic than the more restrictive one used in [1] (that requires convergence to an equalized system with a priori known gain), it introduces a further degree of freedom which makes it difficult to obtain similar results to those obtained with

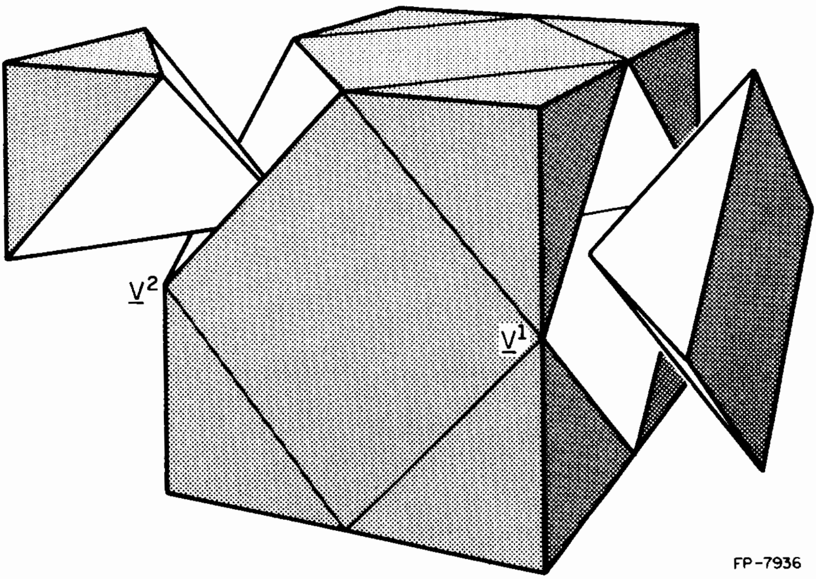

Figure 1. Simplices in unit cube for generating necessary and sufficient conditions for nonnegativity of the directional derivative of the risk.

the narrower sense definition (e.g. Proposition 3). Note finally that the class of memoryless adaptive laws considered here only takes advantage of the one-dimensional distribution of the output of the unequalized linear system; it is plausible that the consideration of dynamic laws which take into account the stochastic dependence of the equalizer outputs, can result in notably improved convergence properties.

ACKNOWLEDGMENT

This work was supported in part by an IBM Pre-doctoral Fellowship and by the U. S. Office of Naval Research under Contract N0014-81-K-0014.

REFERENCES

[1] 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. Automatic Control*, vol. AC-25, no. 3, pp. 385-399, June 1980.

[2] Y. Sato, "A method of self-recovering equalization for multilevel amplitude modulation," *IEEE Trans. Communications*, pp. 679-682, June 1975.

[3] D. N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. Communications*, vol. COM-28, no. 11, pp. 1867-1875, Nov. 1980.

[4] A. Benveniste, M. Goursat and G. Ruget, "Analysis of stochastic approximation schemes with discontinuous and dependent forcing terms with applications to data communication algorithms," *IEEE Trans. Automatic Control*, vol. AC-25, no. 6. pp. 1042-1058, Dec. 1980.

[5] G. Ungerboeck, "Comments on 'Self-recovering equalization and carrier tracking in two-dimensional data communication systems'," *IEEE Trans. Communications*, vol. COM-30, no. 3, pp. 557, Mar. 1982.

[6] F. S. Hill, Jr. and M. A. Blanco, "Random geometric series and intersymbol interference," *IEEE Trans. Information Theory*, vol. IT-19, no. 3, pp. 326-335, May 1973.

[7] L. V. Kantorovich and G. P. Akilov, *Functional Analysis*, Second Edition, Oxford: Pergamon Press, 1982.

[8] L. Ljung, "Analysis of a general recursive prediction error identification algorithm," *Automatica*, vol. 17, no. 1, Jan. 1981.

[9] L. Ljung, "Analysis of recursive stochastic algorithms," *IEEE Trans. Automatic Control*, vol. AC-22, no. 4, Aug. 1977.