ON THE SELECTION OF MONEITLESS ADAPTIVE LAWS FOR BLIND EQUALIZATION IN KINNAY COMMUNICATIONS

Sergio Bress
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), unequilized impulse responses with a 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

Notation: For observations, \( y = \{y_n \in \mathbb{Z} \} \), the output of an unknown discrete-time linear time-invariant system driven by a Bernoulli sequence \( \xi = \{\xi_n \in \mathbb{Z} \} \) are used. In order to adjust the impulse response of an (IR) linear equalizer \( \theta = \{\theta_n \in \mathbb{Z} \} \) such that when driven by \( \xi \), its output is the original Bernoulli sequence. A possible strategy is to use a nonempty adaptive law of the type

\[
\theta[n+1] = \theta[n] + \tau \pi^T[n] \xi[n] \xi[n] \theta[n]
\]

where \( \pi[n] \) is a sequence of small positive scalars, \( \pi[n] \) is the current output of the equalizer, and \( \pi[n] \) is a real-valued function to be specified.

The scheme (1) for the blind equalization problem has been made in the past in [1]-[3], and in based on the fact that (1) is a stochastic approximation [4] for the minimization of the risk \( R(\theta) = E[\xi^T \xi | \theta] \), where \( \xi^T \xi \) is a function of the original input.

As popularized by [5], the analysis of the convergence of (1) can be carried out, if \( \tau \) is smooth enough, by first studying the steepest descent lines of

\[
R(\theta) = E[\xi^T \xi | \theta] = \theta^T \pi^T[n] \xi[n] \xi[n] \theta[n] \quad \text{and then analyzing the behavior of (1) with respect to that of the integral curve of}
\]

\[
\frac{d\theta}{dt} = -\gamma R(\theta).
\]

Ideally, the designer's goal would be to select an adaptive law \( \pi(n) \) such that all the steepest descent lines of \( R(\theta) \) converge to \( \cdots, 0, 1, 0, \cdots \). However, to date, no
function $\varphi(\cdot)$ has been found to satisfy such a requirement. In fact, since the existence of steepest descent lines that converge to unstable equilibria (i.e., tangent 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**

$\varphi$ is admissible if the steepest descent lines of $H(\varphi)$ converge and its local minima belong to the set $L = \{x \in \mathbb{R}, \text{sgn}(x) = \{\ldots, 0, 0, 0, \ldots\} \} = \delta$.

In their seminal work [1], Heunen, Vantieghem, Courant, and Huet showed that if the distribution (1) of the i.i.d. input sequence is sub-Gaussian rather than Bernoulli and $\varphi$ satisfies:

(a) $\varphi(x) = -\text{sgn}(x) + \varphi(x)$
(b) $\int \varphi(x)(\text{d}x) = 0$ (i.e., $0 \in \delta$)
(c) $\varphi$ is an odd, twice differentiable and convex on $\delta$

Then $\varphi$ is an admissible adaptive law, and the only local minima are $\{\ldots, 0, 0, 0, \ldots\}$. 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 $\varphi(\cdot)$ such that an adaptive scheme with the above desirable properties - admissibility and existence of descent 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 $\overline{\varphi}$ 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 $\varphi(x) = \lim_{x \to 0^+} \varphi(x)$ $h(x)$
(c) $\psi(x) = \lim_{x \to 0^-} \varphi(x-h)$ $h(x)$
(d) $\psi$ is odd. (Note that since $\varphi$ has symmetric distribution, any function $\overline{\varphi}$ results in the same risk as the even function $\psi(x) = \psi(-x).$)

Conditions (a)-(d) ensure the existence of the left-hand and right-hand derivatives $\overline{\psi}(x)$, which is necessary for the analysis of the steepest descent minimization of $H(\varphi)$. Let the (unnormalized) Cramer directional derivative of $H(\cdot)$ at point $\xi$ in the direction $\overline{\psi}(\xi)$ be defined

**Definition 2**

$\xi$ is a point of co-variation (a) for all $\xi$, $\overline{\psi}(\xi)$ (b) for every neighbor $\overline{\psi}(\xi)$.

Our first result will be a fulfillment of condition (a) for a problem when $\xi$ will be the zero terms $\xi$.

**Proposition 1**

If for every $\xi$ such that $\xi_0$

then (a) holds for every direction $\xi$.

**Proof**

For an arbitrary $\xi_0$, we have

$\overline{\psi}(\xi_0) = \lim_{h \to 0^+} \overline{\psi}(\xi_0)$

$\lim_{h \to 0^+} \overline{\psi}(\xi_0) = \lim_{h \to 0^+} \overline{\psi}(\xi_0)$

where the last equation follows for any nonnegative scalar $\lim_{h \to 0^+} \overline{\psi}(\xi_0) = \overline{\psi}(\xi_0)$

if $\overline{\psi}(\xi_0) \geq 0$ for every $\xi_0$ at $\xi_0 = \xi_0$ if $\xi_0 \neq 0$, and $\xi_0$

$\lim_{h \to 0^+} \overline{\psi}(\xi_0) = \overline{\psi}(\xi_0)$

$\overline{\psi}(\xi_0) (1 + \xi_0)$
the direction \( \hat{d} \) be defined by
\[
\hat{d}(t) = \lim_{h \to 0} \frac{1}{h} \left[ R(t + h\hat{d}) - R(t) \right].
\] (1)

The steepest descent algorithm [7] for the minimization of \( R(\hat{d}) \) selects at each point the direction at which \( R(t) \) for some norm \( \| \bullet \| \) is minimum. If such a minimum exists and is negative then a new approximate to the sought-after global minimum of \( R(\hat{d}) \) is obtained. Hence the three of the steepest descent lines have the following properties:

Definition 2

\( t \) is a point of convergence if

(a) for all \( \delta \), \( R(t + \delta \hat{d}) > 0 \)

(b) for every neighborhood \( N(\hat{d}) \) there exists \( t^* \in N(\hat{d}) \) such that \( R(t^*) < R(t) \).

Our first result will be invoked again and again in the sequel and reduces the fulfillment of condition (a) to 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 \( t \).

Proposition 1

If for every \( t \) such that \( t_1 = 0 \), \( \delta = \hat{d} \), we have
\[
R(t_1 + \delta t) > 0,
\]
then (4) holds for every direction \( \hat{d} \) such that the directional derivative \( R(t_1; \hat{d}) \) exists.

Proof

For an arbitrary \( \delta \), we have
\[
R(t + \delta \hat{d}) = \lim_{h \to 0} \frac{1}{h} \left[ R(t + h\hat{d}) - R(t) \right]
\]

\[
= \lim_{h \to 0} \frac{1}{h} \left[ E\left(\tilde{t} + h\hat{d} \right) - E(\tilde{t}) \right]
\]

\[
= \lim_{h \to 0} \frac{1}{h} \left[ E\left(\tilde{t} + h\hat{d} \right) - E(\tilde{t}) \right]
\]

where the last equation follows because \( \hat{d} \) is even. Therefore it is enough to show that for any positive scalar \( \epsilon \),
\[
E(\tilde{t} + \epsilon \hat{d} \tilde{t}) = E(\tilde{t} \epsilon \hat{d} \tilde{t} \hat{d}) - E(\tilde{t}) \epsilon > 0.
\] (5)

If \( R(t \hat{d}) > 0 \) for every \( \varepsilon \) such that \( t_1 = 0 \) unless \( t_1 \tilde{t} \neq 0 \). Let \( \tilde{d} \) be such that
\[
\tilde{d}_1 = \frac{\tilde{d}_1}{\| \tilde{d} \|} \text{ if } t_1 \tilde{t} \neq 0, \text{ and } \tilde{d}_1 = 0 \text{ if } t_1 \tilde{t} = 0.
\]

\[
E(\tilde{t} + \epsilon \hat{d} \tilde{t}) = E(\tilde{t} \epsilon \hat{d} \tilde{t} \hat{d}) - E(\tilde{t}) \epsilon > 0.
\]
\[ M(\alpha), M(\beta), \alpha, \beta \in \mathbb{R}, -\mathbb{R} \]
Right hand side of the last line of the text is not clear. It seems to be an equation or a definition, but the exact meaning is not discernible.

The text discusses the conditions under which a system is stable or unstable, based on the signs of certain expressions. It mentions that if \( \psi \) is admissible and there exists a local minimum of \( \psi(x) \in (K, 0) \) at some point, then the system is stable. The proof involves showing that the derivative of \( \psi \) is negative at this point, which implies stability.

The proof concludes with a statement about the derivative of \( \psi \) being negative, which is a condition for the system to be in a stable equilibrium.
Since all the steepest descent times converge and there are no local minima of \( H(x) \) other than \( H(x) = 0 \), it is necessary that

\[
\psi'(x) > 0 \quad x \in [0,x_0]
\]

(9)

Now consider the case:

(a) \( \psi'(x) + \psi(x) > 0 \) and \( \psi'(x) + \psi(x) < 0 \).

These imply that there exists a strict local minimum of \( \psi(x) \) and an upper bound of \( \psi(x) > \psi(-t) \). On the one hand this implies that \( \psi(x) > \psi(t) \) and \( \psi(x) > \psi(0) \) for all \( x \) in the interval \( [0,x] \). On the other hand, we have that \( \psi'(x) > 0 \) for all \( x \) in the interval \( [0,x_0] \). Therefore \( \psi(x) = \psi(x') + \psi(t) \) and \( \psi(x) > \psi(-t) \). (Note that condition (a) in the statement of Proposition 4 follows from (a) and (b).) The alternative condition for avoiding points of convergence of \( \psi(x') = \psi(x) > \psi(-t) \).

Proportion 3 shows that the discontinuity of the adaptive law is necessary for the existence of points of convergence at \( \psi(x') = \psi(x) > \psi(-t) \). This implies that even in the even that there exists an adaptive law without points of convergence at unequal time systems, in practice, the adaptive scheme will lack the necessary robustness for ensuring 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 control 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 attention to adaptive laws whose derivative $\psi'(\epsilon)$ exists for $\epsilon \in (0,\beta)$.

Proposition 4
If there exists $\epsilon \in (0,\beta)$ such that

1. $\psi'(\epsilon) + \psi(3\epsilon) < 0$
2. $\psi'(\epsilon) + 3\psi'(3\epsilon) > 0$
3. $\psi'(\epsilon) > 0$

and at least one inequality is strict, then $\psi$ is not admissible.

Proof
Under the above conditions there exists a local minimum at $(...,0,1,1,0,...)$. In order to show this, we prove that

$$R_i \bigg| \epsilon = (...,0,1,1,0,...)$$

if and only if (1) holds (Equations (6)-(7) reduce to (6) if $\psi$ is continuous), and that

$$H_i \left[ 3\psi'(\epsilon) \psi(3\epsilon) \right] > 0, \quad i = (...,1,1,0,...)$$

and $H \neq 0$ if and only if (11) and (111) are true. Notice that this is enough because of Proposition 1 and $\psi'(\epsilon) > 0$ at $\epsilon = (...,0,1,1,0,...)$ if $\epsilon > 0$. In order to show the nonnegativity of the Hessian, we have that

$$\sum_{1 \leq i, j \leq 3} \left| H_{ij}(\epsilon) \right| > 0, \quad i = (...,0,1,1,0,...)$$

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

$$\psi'(3\epsilon) + 3\psi'(\epsilon) > 0$$

and

$$\psi'(3\epsilon) + 3\psi'(\epsilon) > 0$$

which is equivalent to

$$\psi'(\epsilon) + \psi'(3\epsilon) > 0$$

and

$\psi'(\epsilon) > 0$.

Corollary
If $\psi'(\epsilon) > 0$ for every $\epsilon \in (0,\beta)$ then $\psi$ is not admissible.

Proof
Suppose that $\psi$ is admissible. Then since it is strictly increasing there exists a point $k$ such that $\psi(k) < 0 < \psi(y)$ for $0 < k < x < y$. Now, using the fact that $\psi$ is
This corollary shows that a continuously differentiable convex set $(0,\infty)$-convex function $f(\cdot)$ is admissible. This is particularly interesting since the admissible 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 admissible from the viewpoint of global convergence. Furthermore, a much larger class of laws (including the continuous functions) has been shown to result in points of convergence at unequalized systems. All this has been accomplished by restricting attention to the behavior of the cost function and the system whose impulse response have no more than three nonzero coefficients and hence avoiding the need to specify the underlying space of the variables. 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 $f$, for example) could be investigated. It is straightforward to show that if a real-magnitude nonzero coefficients are allowed the gradient of the $\mathcal{V}$-function is zero for all directions at \(\cdots, 0, t, 0, \cdots\) if and only if

\[\mathcal{V}(0) + \frac{n-3}{2} \mathcal{V}(n-2) + \frac{1}{2} \sum_{i=1}^{n-2} \mathcal{V}(n-2i-1) = 0,\]

it is tedious to find conditions for the nonnegativity of the function for generic $\mathcal{V}$. On the other hand, when $\mathcal{V}$ has more than three or four nonzero coefficients it is generally tedious to obtain necessary and sufficient conditions for $f(\cdot, \delta) > 0$ for all $\delta$ (cf. (0)-(6) for $n = 3$). However, it is possible to systematize the derivations of such conditions by using the following result.

**Proposition 4.**

Let $1$ be the set of all vertices of the simplex defined by subsets of the following hyperplanes in $\mathbb{R}^n$:

- $x_i > 0, \quad i = 1, \ldots, n$
- $x_1 \leq 1$
- $x_i \leq 1, \quad i = 1, \ldots, n$
- $\mathcal{V}(0) + \frac{n-3}{2} \mathcal{V}(n-2) + \frac{1}{2} \sum_{i=1}^{n-2} \mathcal{V}(n-2i-1) = 0.$

If $f(\mathcal{V}) > 0$ for all $\mathcal{V} \in V$, then $f(\mathcal{V}(\delta)) > 0$ for all $\delta \in \mathbb{R}^n$.

**Proof.**

Select any $\delta \in \mathbb{R}^n - (0)$ and let $V_\delta = \mathcal{V} \in (-1, 1)^n, s.t. f(\mathcal{V}) > 0$, then $1 + \frac{1}{\mathcal{V}(\delta)}$ belongs to the simplex (with vertices $V_\delta \in V$) defined by

- $x_i < 1, \quad i = 1, \ldots, n$
- $\mathcal{V}(0) = 1$
- $\mathcal{V}(n-2i-1) > 0, \quad i = 1, \ldots, n$,

and therefore $\mathcal{V}$ can be put as a convex combination of the vertices in $V_\delta = \{V_1, \ldots, V_k\}$.

The proof is completed by showing that

\[f(\mathcal{V}(\delta)) = \frac{1}{\mathcal{V}(\delta)} \mathcal{V}(\mathcal{V}(\delta)) = \frac{1}{\mathcal{V}(\delta)} \mathcal{V}(\mathcal{V}(\delta)) > 0,\]

To that end we have

\[(\mathcal{V}(\delta))/\mathcal{V}(\delta) = \mathcal{V}(\mathcal{V}(\delta))/\mathcal{V}(\mathcal{V}(\delta)) > 0,\]

where we have used $\mathcal{V}(\mathcal{V}(\delta)) = \mathcal{V}(\mathcal{V}(\delta))$, similarly $K_{\delta}$.

The import of Proposition 3 is that (recall Prop. 1) it is enough to restrict the given by vertices of polytopes in the hypervolumes $\mathcal{G} \cap \mathcal{G}$ would result in the systematical. Once the conditions generated they can be reduced to a set.

Figure 1 illustrates the case 4 and (8). It turns out that these equations and $\mathcal{V}$ and $\mathcal{V}$ respectively. In particular, because the same archetypes $\mathcal{V} = y \in V$, where $y \in V$, and $y$ result in the this approach and its connections with obtaining an algorithm that generates for the nonnegativity of the directness.

Overall, it appears that the approach showing the existence of unequalized of classes of adaptive laws. Several approach is needed to prove or dig out the reasonably general space of impulse the system and its inverse have this. Another point is that although the more realistic than the more restricted to an equalized system with a larger freedom makes it difficult to
\( \phi(t) + \psi(t) = 0. \)

The proof is completed by showing that

\[
\phi(t) = \sum_{i=1}^{K} a_i \phi_i(t), \quad \psi(t) = \sum_{i=1}^{K} a_i \psi_i(t).
\]

So that we have

\[
(t, \xi) \mid \phi(t) = \phi(t) \mid \sum_{i=1}^{K} a_i \phi_i(t) \sum_{i=1}^{K} a_i \phi_i(t) = \sum_{i=1}^{K} a_i \phi_i(t) \sum_{i=1}^{K} a_i \phi_i(t) = \sum_{i=1}^{K} a_i \phi_i(t) \sum_{i=1}^{K} a_i \phi_i(t).
\]

where we have used \( \phi(\xi) = \psi(\xi) = \phi(t) = \psi(t) = 0 \), and the fact that both \( \phi(t) \) and \( \psi(t) \) belong to the simplex \( S_3 \).

The result of Proposition 4 is that when \( \alpha \) 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 hypotheses e.g. \( ||.|| \) 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 \( \phi_1, \phi_2, \phi_3 \) and \( \phi_4 \) respectively. In particular, note that not every vertex must be investigated because if \( \phi \in \mathcal{V} \), then \( \phi \in \mathcal{V} \) where \( \phi_1, \phi_2, \phi_3 \) and \( \phi_4 \) is any subjective function of \( (\phi_1, \phi_2, \phi_3) \) and \( \phi_4 \) 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 insufficiency 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 relatively general space of impulse responses is allowed (note that requiring that the system and its inputs have finite energy \( ||.|| \) — say by 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 Hermitian known gain, it introduces a further degree of freedom which makes it difficult to obtain similar results to those obtained with

\( \phi(t) = \sum_{i=1}^{K} a_i \phi_i(t), \quad \psi(t) = \sum_{i=1}^{K} a_i \psi_i(t). \)
Figure 1. Simplices in unit cube for generating necessary and sufficient conditions for nonemptiness of the directional derivative of the risk.

the narrower scope definition (e.g., Proposition 3). More finally than the class of nonredundant adaptive laws considered here only takes advantage of the one-dimensional distribution of the output of the non-idealized 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 Predoctoral Fellowship and by the U. S. Office of Naval Research under Contract N00014-81-K-0014.

REFERENCES


[5] G. D'Angelo, "Comments on 'S


