Undermodeled Equalization: A Characterization of Stationary Points for a Family of Blind Criteria

Phillip A. Regalia, *Senior Member, IEEE*, and Mamadou Mboup

***Abstract*— We attack specific problems related to equalizer performance in undermodeled cases in which assumptions of perfect equalizability are dismissed in favor of a more realistic situation in which no equalizer setting may achieve perfect channel equalization. We derive a characterization of candidate convergent points for a family of blind criteria which appeal, tacitly or wittingly, to maximizing the ratio of different sequence norms of the combined channel-equalizer impulse response. This may be accomplished in a practical implementation by using equalizer output cumulants of different orders. The popular Godard and Shalvi–Weinstein schemes are accommodated at one extreme of the family of criteria. We also show that each maximum at the other extreme of the family, involving progressively higher order output cumulants, yields, precisely, a Wiener response. This suggests that blind algorithms using progressively higher order statistics may converge more closely to a Wiener response than those using more modest order statistics. We show, moreover, that the superexponential family of algorithms is also included and establish a convergence proof for undermodeled cases that appeals to no approximation. Finally, some apparently novel bounds on attainable open-eye measures in undermodeled cases are also derived.**

The grass is greener on the other side of the fence.
—common proverb

I. Introduction

BLIND equalization methods for digital communication systems are an attractive alternative to their training sequence based counterparts as blind methods do not squander precious channel capacity with the periodic transmission of training sequences. Fractionally spaced equalization methods, in particular, have met with renewed interest in recent years, owing to recent results [1]–[4] showing that perfect channel equalization is attainable, provided the channel impulse response is strictly finite in duration, the polyphase components of the oversampled channel have no common zero (the so-called channel disparity condition), and the equalizer impulse response length is chosen correctly. Some popular blind adaptation algorithms can also claim (e.g., [4]–[6]) to be free from suboptimal convergent points in such cases, i.e., although multiple convergent points generically exist, each achieves perfect channel equalization; the convergent points are distinguished essentially by the residual delay provided by each.

Although this basic result is quite encouraging, knowledge of the channel impulse response length is required to choose the correct equalizer length, and deducing this length is not necessarily a well-conditioned problem. For this task, channel order estimation procedures may certainly be envisaged, but no estimator can claim perfect accuracy, and the true channel length may be slowly varying (compared with the baud rate) in time, reaching, on occasion, a value higher than any estimator may venture to determine. Performance degradations due to insufficient length equalizers are thus inevitable.

Here, we attack specific problems related to equalizer performance in undermodeled cases in which assumptions of perfect equalizability are dismissed in favor of a more realistic situation where no equalizer setting may achieve perfect channel equalization. This occurs, for example, if the equalizer length is insufficient and/or the channel disparity condition is violated, even in the absence of channel noise. The idealized scenario, featuring multiple convergent points with each achieving perfect channel equalization, is deformed into a more realistic and daunting scenario in which multiple convergent points are still present but now offering disparate performance levels.

If undermodeling refers to an equalizer's inability to restore the transmitted sequence perfectly, then contributing factors may be recognized in background noise, multiple sources due to spectral leakage and/or imperfect demultiplexing, and insufficient equalizer lengths. We shall focus our attention on the latter factor, in a single-source, noise-free setting. While it is certainly a simpler setting to treat, it is not, as such, elementary. It does, however, allow us to offer an algebraic definition of "undermodeled" (Definition 1 in Section II) that proves relevant for the present paper, and the insights gained may hopefully motivate starting points for tackling the more challenging multisource, noisy channel setting.

We address, in particular, the characterization of candidate convergent points for a family of blind criteria that appeal, tacitly or wittingly, to maximizing the ratio of different norms of the combined channel-equalizer impulse response. With $\mathbf{s}$ denoting the vector built from this combined response (see Fig. 1), we study a characterization of all stationary points and maxima of $D_{2p}(\mathbf{s}) \triangleq \|\mathbf{s}\|_{2p}/\|\mathbf{s}\|_2$ using the standard ℓ_q sequence norm $\|\cdot\|_q$ as $\mathbf{s}$ varies over the set of attainable combined responses. The problem becomes more challenging once this set excludes "ideal" combined responses, as will be the case once practical undermodeling considerations are taken

Manuscript received January 16, 1998; revised July 28, 1998. The associate editor coordinating the review of this paper and approving it for publication was Prof. Antonio Cantoni.

P. A. Regalia is with the Département Signal et Image, Institut National des Télécommunications, Evry, France (e-mail: regalia@int-evry.fr).

M. Mboup is with the UFR Mathématiques et Informatique, Université René Descartes, Paris, France (e-mail: mboup@math-info.univ-paris5.fr).

Publisher Item Identifier S 1053-587X(99)01339-2.

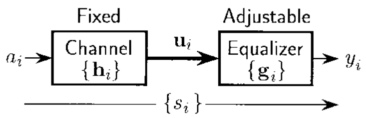

Fig. 1. Channel-equalizer cascade.

seriously. The family of criteria $D_{2p}(\mathbf{s})$ is relevant because the Godard [7] and Shalvi–Weinstein [8] schemes are known to have convergent points at the maxima of $D_4(\mathbf{s})$, as reviewed in Section II, and we show also that the superexponential family of algorithms [5] in fact seek the maxima of $D_{2p}(\mathbf{s})$, where p is a user-chosen integer no smaller than two.

Section II reviews the problem setting and mathematical preliminaries for our study. Section III shows, for undermodeled cases, that each maximum of $\|\mathbf{s}\|_{2p}/\|\mathbf{s}\|_2$ as $p \to \infty$ yields precisely a Wiener response. This suggests that blind equalization methods using progressively higher order statistics may converge more closely to a Wiener response than those using more modest order statistics, although this result must be tempered in practice by the higher variances typically exhibited by higher order statistical estimates.

Section IV presents a necessary and sufficient condition for a candidate attainable combined response to be a stationary point of D_{2p} for finite p. The characterization is nonlinear, as perhaps expected, but tractable. Section V introduces a nonlinear map whose fixed points are precisely the stationary points so sought. A simple iteration based on this nonlinear map allows us to recover the superexponential algorithm of Shalvi and Weinstein [5], who argued for convergence to a solution resembling a Wiener filter, provided the undermodeling effects are not too severe. We give a much stronger result in Section V, namely, that the algorithm approaches a maximum of $D_{2p}(\mathbf{s})$ (where p is user chosen); since our result appeals to no approximation, it applies even if the undermodeling effects are quite severe. As the proofs in Sections IV and V are somewhat technical, the reader wishing to skip them the first time through may do so with little loss of continuity.

Section VI derives some apparently novel bounds on open-eye measures for undermodeled cases, which are useful in simulation studies and prototype verification. Concluding remarks, and directions for further work, are synthesized in Section VII.

II. Problem Setting and Preliminaries

In this section, we review some basic concepts and considerations for the channel-equalizer cascade depicted in Fig. 1.

A. Attainable Responses

With $\{a_i\}$ denoting the scalar-valued source sequence to be transmitted, the equalizer input is obtained after demodulation and sampling of the receiver output, giving the N-element vector

$$\mathbf{u}_i = \sum_k \mathbf{h}_k \, a_{i-k} \quad (N \times 1).$$

Here, $\{\mathbf{h}_i\}$ is the single-input N-output (real or complex) channel impulse response, each term of which is a column vector of N elements. We assume the channel is stable ($\sum_i \|\mathbf{h}_i\| < \infty$) and causal ($\mathbf{h}_i = \mathbf{0}$ for $i < 0$). For the baud-rate, single-sensor case, we take $N = 1$, whereas the fractionally spaced and/or multisensor case corresponds to $N \geq 2$. The equalizer output becomes

$$y_i = \sum_{k=0}^{L} \mathbf{g}_k \, \mathbf{u}_{i-k}$$

in terms of the equalizer impulse response $\{\mathbf{g}_i\}$ of length $L+1$; each term $\mathbf{g}_i$ is a row vector of N elements so that $\{y_i\}$ is a scalar-valued sequence.

From the cascade structure of Fig. 1, the combined impulse response $\{s_i\}$ mapping the source sequence $\{a_i\}$ to the equalizer output sequence $\{y_i\}$, as in

$$y_i = \sum_k s_k \, a_{i-k}$$

is the convolution of $\{\mathbf{h}_i\}$ and $\{\mathbf{g}_i\}$

$$s_i = \sum_{k=0}^{L} \mathbf{g}_k \, \mathbf{h}_{i-k}.$$

If the channel and equalizer are both bounded-input bounded-output stable, then the combined response will be absolutely summable. This implies square summability as ($\sum_i |s_i|^2 < \infty$), meaning $\mathbf{s}$ will belong to the sequence space ℓ_2.

In matrix form, the convolution between $\{\mathbf{h}_i\}$ and $\{\mathbf{g}_i\}$ can be written as

$$\underbrace{\begin{bmatrix} s_0 \\ s_1 \\ \vdots \\ s_L \\ s_{L+1} \\ \vdots \end{bmatrix}}_{\mathbf{s}} = \underbrace{\begin{bmatrix} \mathbf{h}_0^T & \mathbf{0}^T & \cdots & \mathbf{0}^T \\ \mathbf{h}_1^T & \mathbf{h}_0^T & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0}^T \\ \mathbf{h}_L^T & \cdots & \mathbf{h}_1^T & \mathbf{h}_0^T \\ \mathbf{h}_{L+1}^T & \mathbf{h}_L^T & \cdots & \mathbf{h}_1^T \\ \vdots & \ddots & \ddots & \vdots \end{bmatrix}}_{\mathbf{H}} \begin{bmatrix} \mathbf{g}_0^T \\ \mathbf{g}_1^T \\ \vdots \\ \mathbf{g}_L^T \end{bmatrix}.$$

It is clear that the vector $\mathbf{s}$ is restricted to the range space (in ℓ_2) of the matrix $\mathbf{H}$; following Yi and Ding [4], we call this space the set of attainable combined responses, which we denote $\mathcal{S}_A$

$$\mathcal{S}_A = \{\mathbf{s} \in \ell_2 : \{s_i\} = \{\mathbf{g}_i\} * \{\mathbf{h}_i\} \text{ for some equalizer setting} \{\mathbf{g}_i\}.\}$$

This is clearly a linear subspace of ℓ_2. We denote by $\mathcal{P}_A$ the orthogonal projection operator onto $\mathcal{S}_A$; this may be written as $\mathbf{H}(\mathbf{H}^\dagger\mathbf{H})^\sharp\mathbf{H}^\dagger$, where superscript $\dagger$ denotes Hermitian transpose, and superscript $\sharp$ denotes (pseudo-) inversion.

We distinguish next "sufficient order" from "undermodeled" equalizers.[1]

Definition 1: The equalizer $\{\mathbf{g}_i\}$ will be termed *sufficient order* if $\mathcal{P}_A = \mathbf{I}$ (the identity) and *undermodeled* if $\mathcal{P}_A \neq \mathbf{I}$. We illustrate this distinction with two examples:

[1] The definition used here is for convenience in this work. The undermodeled case does not necessarily exclude an equalizer setting that attains an ideal combined response, and accordingly, more restrictive definitions of "undermodeled" may prove more appropriate in other contexts.

Example 1: Suppose the channel impulse response is of finite duration (M, say), and let

$$\mathbf{h}(z) = \sum_{k=0}^{M} \mathbf{h}_k \, z^{-k}$$

be the channel transfer function vector. The combined response $\{s_i\}$ will be a vector in the finite-dimensional complex space $\mathbb{C}^{M+L+1}$ (or its real counterpart $\mathbb{R}^{M+L+1}$ if $\{\mathbf{h}_i\}$ and $\{\mathbf{g}_i\}$ are both real). If $N \geq 2$ (fractionally spaced and/or multisensor case) and $\mathbf{h}(z) \neq \mathbf{0}$ for all z (the channel disparity condition), then choosing the equalizer order L no smaller than $(M-N+1)/(N-1)$ gives the entire space $\mathbb{C}^{M+L+1}$ (or $\mathbb{R}^{M+L+1}$) as the range space of $\mathbf{H}$ (e.g., [4]) so that $\mathcal{P}_A = \mathbf{I}$. An arbitrary vector in $\mathbb{C}^{M+L+1}$ (or $\mathbb{R}^{M+L+1}$) can then be reached by an appropriate setting of the equalizer coefficients $\{\mathbf{g}_i\}$. If, on the other hand, the equalizer order L is chosen smaller than $(M-N+1)/(N-1)$, and/or if the channel disparity condition is violated, and/or $N=1$ (baud rate, single-sensor case), then only a proper subspace of $\mathbb{C}^{M+L+1}$ (or $\mathbb{R}^{M+L+1}$) can be reached by varying the equalizer coefficients, giving $\mathcal{P}_A \neq \mathbf{I}$.

Example 2: Suppose the channel impulse response is infinite in duration ($M \to \infty$). If the channel and equalizer are BIBO stable, the combined response $\{s_i\}$ will still be an element of the infinite sequence space ℓ_2. If the equalizer order L is finite, then only a finite-dimensional (and, hence, insufficient) subspace of ℓ_2 can be reached by varying the equalizer coefficients $\{\mathbf{g}_i\}_{i=0}^{L}$, giving $\mathcal{P}_A \neq \mathbf{I}$. If we let $L \to \infty$, then we will obtain $\mathcal{P}_A = \mathbf{I}$ if and only if $\mathbf{h}(z)$ is bounded and nonzero for all $|z| \geq 1$ (see the Appendix).

B. Criteria to be Maximized

A popular criterion for equalization, originating in Donoho [9], involves normalized cumulants of the form $\text{cum}_{2p}(y_i)/[\text{cum}_2(y_i)]^p$, in which $\text{cum}_{2p}(\cdot)$ is the cumulant of order $2p$ of a random variable (more on cumulants in system theory can be found in [11]). If the source $\{a_i\}$ is an independent, identically distributed (i.i.d.) random sequence and circular if complex, then [8], [9]

$$\frac{\text{cum}_{2p}(y_i)}{[\text{cum}_2(y_i)]^p} = \underbrace{\left(\sum_k |s_k|^{2p} \Big/ \left(\sum_k |s_k|^2 \right)^p \right)}_{[D_{2p}(\mathbf{s})]^{2p}} \frac{\text{cum}_{2p}(a_i)}{[\text{cum}_2(a_i)]^p}.$$

The positive $2p$th root of the parenthetic term is the ratio of different sequence norms, viz

$$D_{2p}(\mathbf{s}) = \frac{\|\mathbf{s}\|_{2p}}{\|\mathbf{s}\|_2}$$

where

$$\|\mathbf{s}\|_{2p} = \begin{cases} \left(\sum_k |s_k|^{2p}\right)^{1/2p}, & 2p < \infty \\ \sup_k |s_k|, & p = \infty. \end{cases}$$

Adaptive algorithms seeking a maximum of the function $|\text{cum}_{2p}(y_i)/[\text{cum}_2(y_i)]^p|$ met with rekindled interest following Shalvi and Weinstein [8], who treated the case $2p = 4$. The Godard algorithm [7], [10] has since been recognized to likewise seek a maximum of $D_4(\mathbf{s})$, provided the source sequence $\{a_i\}$ has a negative fourth-order cumulant [4]; the same may be shown for sources having a positive fourth-order cumulant by using a reparametrization of the equalizer [12].

Deducing candidate convergent points for these procedures thus leads us to the following problem.

Problem 2: Characterize all stationary points (including maxima) of $D_{2p}(\mathbf{s})$ for each $p = 2, 3, \ldots, \infty$ as $\mathbf{s}$ varies over the set of attainable responses $\mathcal{S}_A$.

It is known that $0 \leq D_{2p}(\mathbf{s}) \leq 1$ for any square-summable sequence $\{s_i\}$ and that the upper bound of unity is attained if and only if the sequence $\{s_i\}$ has a sole nonzero term (e.g., [8], [9]). Since $D_{2p}(\alpha\mathbf{s}) = D_{2p}(\mathbf{s})$ for all (nonzero) complex scalars α, two responses ($\mathbf{r}$ and $\mathbf{s}$, say) are equivalent with respect to this criterion if $\mathbf{r} - \alpha\mathbf{s} = \mathbf{0}$ for some scalar $\alpha \neq 0$. Accordingly, we shall say that $D_{2p}(\mathbf{s})$ admits a "unique" maximum in some subset of $\mathcal{S}_A$ if all vectors in the subset attaining the maximum belong to a one-dimensional (1-D) subspace of the form $\alpha\mathbf{s}_*$ for some vector $\mathbf{s}_*$ in the subset.

We denote by

$$\mathbf{e}_n = [\ldots, 0, \underbrace{1}_{n}, 0, \ldots]^T$$

the nth unit vector. In the sufficient order case ($\mathcal{P}_A = \mathbf{I}$), a local maximum of $D_{2p}(\mathbf{s})$, for any $p \geq 2$, is attained if and only if $\mathbf{s} = \alpha\,\mathbf{e}_n$ for any attainable delay n and for an arbitrary nonzero scalar α. In the undermodeled case ($\mathcal{P}_A \neq \mathbf{I}$), the structure of the set of all maxima of $D_{2p}(\mathbf{s})$ is less clear and dependent on p but will be exposed starting in Section III.

C. Dominant Cones

Following [4], the nth dominant cone, denoted $\mathcal{S}_n$, is comprised of all sequences in ℓ_2 (attainable or not) whose nth term is dominant, real, and positive

$$\mathcal{S}_n = \{\{s_i\} \in \ell_2 : s_n > |s_i| \text{ for all } i \neq n\}.$$

This is an open set, and its boundary becomes

$$\mathcal{B}_n = \{\{s_i\} \in \ell_2 : s_n \geq |s_i| \text{ for all } i, \text{ with equality for some } i \neq n\}.$$

The closure of $\mathcal{S}_n$ is denoted $\bar{\mathcal{S}}_n = \mathcal{S}_n \cup \mathcal{B}_n$.

If $|s_n| > |s_i|$ for all $i \neq n$, but s_n is complex or negative real, then $\alpha\mathbf{s}$ will lie in $\mathcal{S}_n$ upon choosing $\alpha = s_n^*/|s_n|$; since $D_{2p}(\alpha\mathbf{s}) = D_{2p}(\mathbf{s})$, such "phase normalization" may be assumed where convenient without loss of generality. Note that if $\mathbf{r}$ and $\mathbf{s}$ are both in $\bar{\mathcal{S}}_n$, so is $\alpha\mathbf{r} + \beta\mathbf{s}$ for all $\alpha, \beta \geq 0$ (so that $\mathcal{S}_n$ is a convex cone), and we have $\|\alpha\mathbf{r} + \beta\mathbf{s}\|_\infty = \alpha\|\mathbf{r}\|_\infty + \beta\|\mathbf{s}\|_\infty$.

Definition 3: The nth cone $\mathcal{S}_n$ is termed *penetrable* if $\mathcal{S}_A \cap \mathcal{S}_n \neq \emptyset$.

Penetrability of the nth cone thus means that some equalizer setting $\{\mathbf{g}_k\}$ exists, which renders the nth term of the combined response dominant.

D. Wiener Combined Responses

The nth Wiener combined response, denoted $\mathbf{w}^{(n)}$, is the best ℓ_2 approximation in $\mathcal{S}_A$ to a pure delay of n samples, i.e.,

$$\mathbf{w}^{(n)} \triangleq \arg\min_{\mathbf{s}\in\mathcal{S}_A} \|\mathbf{e}_n - \mathbf{s}\|_2$$

or $\mathbf{w}^{(n)} = \mathcal{P}_A(\mathbf{e}_n)$. By the projection theorem (e.g., [13]), the nth Wiener response $\mathbf{w}^{(n)}$ is unique, and the approximation error $\mathbf{e}_n - \mathbf{w}^{(n)}$ is orthogonal to $\mathcal{S}_A$. Using the standard inner product in ℓ_2, viz

$$\langle \mathbf{r}, \mathbf{s}\rangle = \sum_k r_k^* s_k$$

we thus have, for any $\mathbf{s}\in\mathcal{S}_A$, the relation $\langle \mathbf{e}_n - \mathbf{w}^{(n)}, \mathbf{s}\rangle = 0$, or

$$s_n = \langle \mathbf{w}^{(n)}, \mathbf{s}\rangle, \quad \text{for all } \mathbf{s}\in\mathcal{S}_A. \tag{1}$$

This shows that each Wiener response behaves as a reproducing kernel over $\mathcal{S}_A$. Since this relation applies to all $\mathbf{s}\in\mathcal{S}_A$, it applies to the particular choice $\mathbf{s} = \mathbf{w}^{(n)}$, giving

$$\mathbf{w}_n^{(n)} = \langle \mathbf{w}^{(n)}, \mathbf{w}^{(n)}\rangle = \left\|\mathbf{w}^{(n)}\right\|_2^2. \tag{2}$$

This shows that the nth term of the nth Wiener response is always real and non-negative. From this, we may readily show that the approximation error of the nth Wiener response is

$$0 \le \left\|\mathbf{e}_n - \mathbf{w}^{(n)}\right\|_2^2 = 1 - \mathbf{w}_n^{(n)} \tag{3}$$

so that $0 \le \mathbf{w}_n^{(n)} \le 1$.

III. Maxima of D_∞ and Wiener Responses

For the sufficient order case ($\mathcal{P}_A = \mathbf{I}$), each maximum of $D_{2p}(\mathbf{s})$ yields $\mathbf{s} = \alpha\mathbf{e}_n$ for any $p \ge 2$. For the undermodeled case, by contrast, the set of maxima varies with p. If maximizing $D_{2p}(\mathbf{s})$ can be accomplished by using output cumulants of orders 2 and $2p$, then $D_\infty(\mathbf{s})$ would represent a limiting criterion as progressively higher order cumulants are used. The following result shows that each maximum of $D_\infty(\mathbf{s})$ in $\mathcal{S}_A$ yields (to within a scale factor) a Wiener combined response:

Theorem 4: Suppose $\mathcal{S}_n$ is penetrable ($\mathcal{S}_A \cap \mathcal{S}_n \ne \emptyset$), and let $\mathbf{w}^{(n)}$ be the nth Wiener response.

1) For all $\mathbf{s}\in\mathcal{S}_A \cap \bar{\mathcal{S}}_n$

$$D_\infty(\mathbf{s}) = \frac{s_n}{\|\mathbf{s}\|_2} \le \frac{\mathbf{w}_n^{(n)}}{\left\|\mathbf{w}^{(n)}\right\|_2}.$$

2) If $\mathbf{w}^{(n)} \in \mathcal{S}_n$, then it attains the unique maximum of D_∞ in $\mathcal{S}_A \cap \bar{\mathcal{S}}_n$.

3) If $\mathbf{w}^{(n)} \notin \mathcal{S}_n$, then any maxima of D_∞ in $\mathcal{S}_A \cap \bar{\mathcal{S}}_n$ must occur along the boundary $\mathcal{B}_n$.

We remark that if a maximum of $D_\infty(\mathbf{s})$ in $\mathcal{S}_A \cap \bar{\mathcal{S}}_n$ is encountered on the boundary $\mathcal{B}_n$ (part 3), then enlarging the search space for $\mathbf{s}$ with an appropriate adjacent dominant cone will generally allow an increase in the function $D_\infty(\mathbf{s})$.

Proof: We exploit the fact that $D_\infty(\mathbf{s}) = D_\infty(\alpha\mathbf{s})$ for any scalar α. We shall choose the scale factor such that the ℓ_2 norm of each $\mathbf{s}$ equals $D_\infty(\mathbf{s})$ within $\bar{\mathcal{S}}_n$ and then deduce the shape of the surface that results.

To this end, suppose each $\mathbf{s}\in\mathcal{S}_A$ is scaled such that

$$s_n = \|\mathbf{s}\|_2^2. \tag{4}$$

(If $\mathbf{s}$ is not so scaled, then $\mathbf{r} = \alpha\mathbf{s}$ will be, upon choosing $\alpha = s_n^*/\|\mathbf{s}\|_2^2$; since $D_\infty(\mathbf{r}) = D_\infty(\mathbf{s})$, we assume that (4) holds without loss of generality.) We then observe that

$$\begin{aligned}\langle \mathbf{w}^{(n)} - \mathbf{s}, \mathbf{s}\rangle &= \langle \mathbf{w}^{(n)}, \mathbf{s}\rangle - \langle \mathbf{s}, \mathbf{s}\rangle\\ &= s_n - s_n = 0\end{aligned}$$

in which we recall that the Wiener response $\mathbf{w}^{(n)}$ is a reproducing kernel [cf. (1)] and that $\langle \mathbf{s}, \mathbf{s}\rangle = \|\mathbf{s}\|_2^2 = s_n$ in view of (4). This shows that $\mathbf{s}$ and $\mathbf{w}^{(n)} - \mathbf{s}$ are orthogonal, and hence, the Pythagorean theorem gives the relation

$$\|\mathbf{s}\|_2^2 + \left\|\mathbf{w}^{(n)} - \mathbf{s}\right\|_2^2 = \left\|\mathbf{w}^{(n)}\right\|_2^2$$

for all $\mathbf{s}\in\mathcal{S}_A$ that are scaled according to (4). Since this relation is equivalent to $\|\mathbf{s} - \frac{1}{2}\mathbf{w}^{(n)}\|_2 = \frac{1}{2}\|\mathbf{w}^{(n)}\|_2$, we recognize a sphere in $\mathcal{S}_A$ with center $\frac{1}{2}\mathbf{w}^{(n)}$ and radius $\frac{1}{2}\|\mathbf{w}^{(n)}\|_2$; the choices $\mathbf{s} = \mathbf{0}$ and $\mathbf{s} = \mathbf{w}^{(n)}$ yield antipodal points on this sphere. Since $s_n = \|\mathbf{s}\|_2^2$, the ℓ_2 distance from the origin at each point of this sphere is $\|\mathbf{s}\|_2 = s_n/\|\mathbf{s}\|_2$, which becomes $\|\mathbf{s}\|_\infty/\|\mathbf{s}\|_2 = D_\infty(\mathbf{s})$ for that part of the sphere that passes through $\bar{\mathcal{S}}_n$.

Now, since the sphere includes the origin, the unique farthest point from the origin on the sphere is its antipode $\mathbf{w}^{(n)}$, which thus attains the unique local maximum of $s_n/\|\mathbf{s}\|_2$. If we restrict our attention to that part of the sphere that passes through $\bar{\mathcal{S}}_n$, we obtain part 1. If $\mathbf{w}^{(n)} \in \mathcal{S}_n$, then it becomes the unique maximum of $D_\infty(\mathbf{s})$ in $\mathcal{S}_A \cap \bar{\mathcal{S}}_n$, giving part 2. If, on the other hand, $\mathbf{w}^{(n)} \notin \mathcal{S}_n$, then seeking a farthest point from the origin on that part of the sphere that passes through $\bar{\mathcal{S}}_n$ must push us to the boundary $\mathcal{B}_n$, to give part 3. □

Example 3: Consider a baud-rate equalizer ($N = 1$) using an FIR channel

$$\begin{aligned}&[\mathbf{h}_0 \quad \mathbf{h}_1 \quad \mathbf{h}_2 \quad \mathbf{h}_3 \quad \mathbf{h}_4]\\ &\quad = [0.06 \quad -0.15 \quad 0.54 \quad 0.13 \quad 0.43]\end{aligned}$$

and a two-coefficient equalizer. The space $\mathcal{S}_A$ is then spanned by the columns of

$$\mathbf{Q} = \begin{bmatrix} -0.0832 & 0.0058\\ 0.2081 & -0.0979\\ -0.7492 & 0.2608\\ -0.1804 & -0.7385\\ -0.5966 & -0.1392\\ 0 & -0.5980\end{bmatrix}$$

obtained from the QR decomposition of $\mathbf{H}$. We then have $\mathcal{P}_A = \mathbf{Q}\mathbf{Q}^T$, and each $\mathbf{s}\in\mathcal{S}_A$ may be written as

$$\mathbf{s} = \mathbf{Q}\bar{\mathbf{g}} = \mathbf{Q}\begin{bmatrix} r\cos\theta\\ r\sin\theta\end{bmatrix}$$

using a polar parametrization of the two-coefficient equalizer $\bar{\mathbf{g}}$.

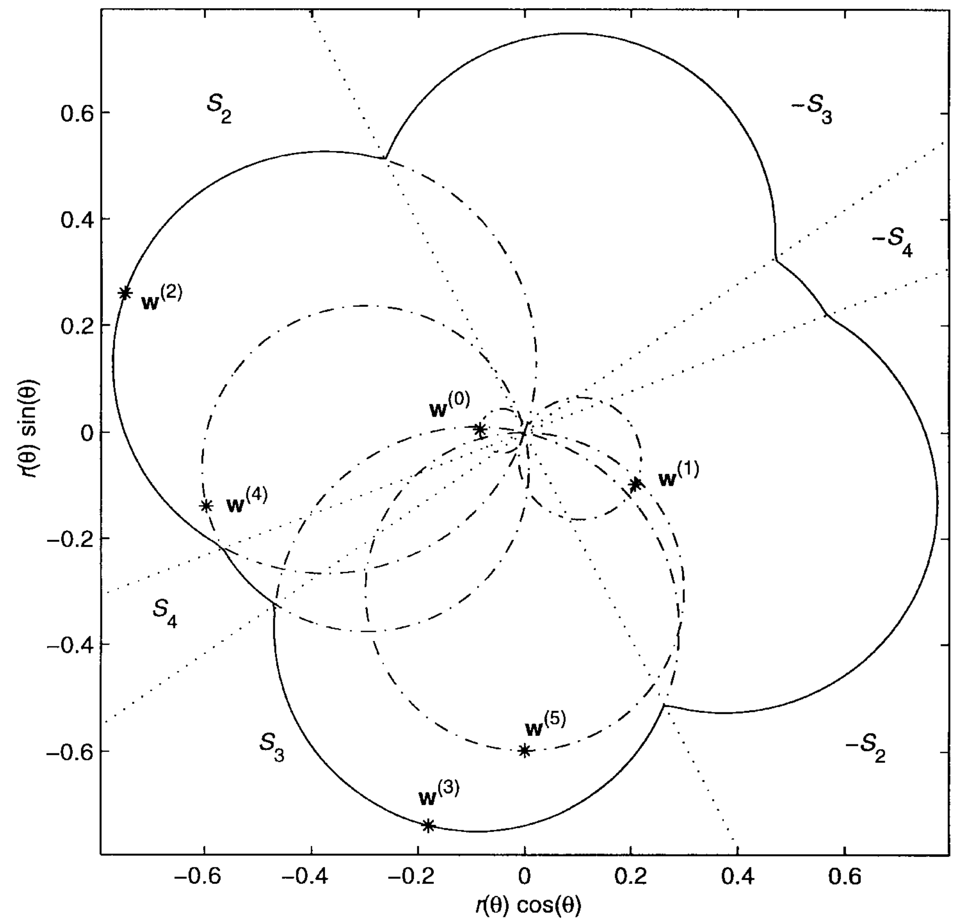

Fig. 2. Two-coefficient equalizer in polar form $[r\cos\theta, r\sin\theta]$. Each circle contains the origin and a Wiener response ("$*$") as antipodal points, whereas dotted lines separate dominant cones.

Since the columns of $\mathbf{Q}$ are orthonormal, we have $||\mathbf{s}||_2 = ||\bar{\mathbf{g}}||_2 = |r|$. A sphere in the combined response space will, by ℓ_2 norm preservation, appear as a sphere in the equalizer coefficient space $[r\cos\theta, r\sin\theta]$. We choose r such that $||\mathbf{s}||_2 = D_\infty(\mathbf{s})$.

Fig. 2 plots $r = r(\theta)$ subject to the constraint that $r = ||\mathbf{s}||_2 = D_\infty(\mathbf{s})$ (solid line). Each dashed circle contains the origin and the equalizer setting attaining a Wiener response as antipodes; the plot of $D_\infty(\theta)$ versus θ is comprised of the outermost part of the union of these circles plus their antipodal images. The dotted lines separate dominant cones. For this example, cones $\mathcal{S}_2$, $\mathcal{S}_3$, and $\mathcal{S}_4$ are penetrable (as are their antipodes $-\mathcal{S}_2$, $-\mathcal{S}_3$, and $-\mathcal{S}_4$). The Wiener responses $\mathbf{w}^{(2)}$ and $\mathbf{w}^{(3)}$ lie in their respective target cones $\mathcal{S}_2$ and $\mathcal{S}_3$ and attain the unique maxima of D_∞ within their respective cones. The Wiener response $\mathbf{w}^{(4)}$ does not lie in $\mathcal{S}_4$—despite this cone being penetrable—and the maximum of D_∞ in $\bar{\mathcal{S}}_4$ occurs at the boundary separating $\mathcal{S}_2$ and $\mathcal{S}_4$.

IV. Stationary Points of $D_{2p}(\mathbf{s})$ for Finite p

We now direct our attention to characterizing stationary points of $D_{2p}(\mathbf{s})$ for finite p as $\mathbf{s}$ varies over the set of attainable responses $\mathcal{S}_A$. We introduce first some notation that will simplify further expressions.

Given two vectors $\mathbf{r}$ and $\mathbf{s}$ (attainable or not), their Hadamard (or componentwise) product will be noted

$$\mathbf{r}\odot\mathbf{s} \quad \text{with } k\text{th component} \quad [\mathbf{r}\odot\mathbf{s}]_k = r_k\, s_k.$$

The mth Hadamard power, denoted $\mathbf{s}^{\odot m}$, follows similarly as

$$\mathbf{s}^{\odot m} = \underbrace{\mathbf{s}\odot\cdots\odot\mathbf{s}}_{m\,\text{terms}}$$

whose kth component is s_k^m. If $\mathbf{s}$ is in ℓ_2, so is $\mathbf{s}^{\odot m}$ for any finite m. The superscript asterisk $*$ will denote componentwise complex conjugation: $[\mathbf{s}^*]_k = s_k^*$. We will frequently encounter expressions of the form $\mathbf{s}^{\odot p}\odot(\mathbf{s}^*)^{\odot(p-1)}$; as this expression is admittedly cumbersome, we choose to introduce a more aesthetic abbreviation.

Definition 5: The notation $\mathbf{s}^{\odot\langle 2p-1\rangle}$ is an abbreviation for the vector $\mathbf{s}^{\odot p}\odot(\mathbf{s}^*)^{\odot(p-1)}$, whose kth component is

$$\left[\mathbf{s}^{\odot\langle 2p-1\rangle}\right]_k = |s_k|^{2p-2} s_k.$$

We now present the main result of this section.

Theorem 6: Let $\mathcal{P}_A$ be the orthogonal projection operator onto $\mathcal{S}_A$. A candidate $\mathbf{s}\in\mathcal{S}_A$ is a stationary point of $D_{2p}(\mathbf{s})$ if and only if

$$\mathcal{P}_A\left(\mathbf{s}^{\odot\langle 2p-1\rangle}\right) = \alpha\mathbf{s}, \quad \text{for some scalar } \alpha.$$

If $\mathbf{s}$ is scaled to unit ℓ_2 norm, then $\sqrt[2p]{\alpha} = D_{2p}(\mathbf{s})$, which is the value obtained at the stationary point.

Proof: If $\mathbf{s}$ is in $\mathcal{S}_A$, and t is any real scalar, the vector

$$\mathbf{q}(t) = \mathbf{s} + t\mathbf{r}$$

will span $\mathcal{S}_A$ as $\mathbf{r}$ spans $\mathcal{S}_A$, by linearity of this subspace. We introduce the moments of orders 2 and $2p$ of $\mathbf{q}(t)$ as

$$m_2(t) = \sum_{k=0}^{\infty} |q_k(t)|^2, \quad m_{2p}(t) = \sum_{k=0}^{\infty} |q_k(t)|^{2p}$$

and we suppose that $\mathbf{s}$ is scaled to unit ℓ_2 norm for simplicity, giving $m_2(0) = 1$.

We now introduce the function

$$F_{2p}(\mathbf{q}(t)) = \big[D_{2p}(\mathbf{q}(t))\big]^{2p} = \frac{m_{2p}(t)}{[m_2(t)]^p}.$$

The directional derivative of the function F_{2p} at $\mathbf{s}$, with respect to a directional vector $\mathbf{r}$, is defined as [13, chap. 7], [14, pt. V]

$$\lim_{t\to 0} \frac{F_{2p}(\mathbf{s}+t\mathbf{r}) - F_{2p}(\mathbf{s})}{t} = \frac{dF_{2p}(\mathbf{q}(t))}{dt}\bigg|_{t=0}.$$

A candidate $\mathbf{s} \in \mathcal{S}_A$ is a stationary point of the function F_{2p} over $\mathcal{S}_A$ if and only if the directional derivative vanishes at $\mathbf{s}$ for all directional vectors $\mathbf{r}$ in $\mathcal{S}_A$

$$\frac{dF_{2p}(\mathbf{q}(t))}{dt}\bigg|_{t=0} = 0, \quad \text{for all } \mathbf{r} \in \mathcal{S}_A.$$

By a direct calculation, we have

$$\begin{aligned}\frac{dF_{2p}}{dt}\bigg|_{t=0} &= \frac{m_2^p(0)\, m_{2p}'(0) - p\, m_{2p}(0)\, m_2^{p-1}(0)\, m_2'(0)}{m_2^{2p}(0)} \\ &= m_{2p}'(0) - p\, m_{2p}(0)\, m_2'(0), \quad \text{since } m_2(0) = 1 \\ &= 2p\,\mathrm{Re}\left(\sum_{k=0}^{\infty} (s_k^*)^p s_k^{p-1}(r_k - \langle \mathbf{s},\mathbf{r}\rangle s_k)\right) \\ &= 2p\,\mathrm{Re}\big\langle \mathbf{s}^{\odot\langle 2p-1\rangle}, \mathbf{r} - \langle \mathbf{s},\mathbf{r}\rangle \mathbf{s}\big\rangle. \qquad (5)\end{aligned}$$

Suppose now we evaluate this latter inner product for some $\mathbf{r} \in \mathcal{S}_A$, i.e.,

$$\big\langle \mathbf{s}^{\odot\langle 2p-1\rangle}, \mathbf{r} - \langle \mathbf{s},\mathbf{r}\rangle \mathbf{s}\big\rangle = x + j\,y.$$

If we replace $\mathbf{r}$ by $j\mathbf{r}$—which also lies in $\mathcal{S}_A$—we get

$$\big\langle \mathbf{s}^{\odot\langle 2p-1\rangle}, j\mathbf{r} - \langle \mathbf{s},j\mathbf{r}\rangle \mathbf{s}\big\rangle = -y + j\,x$$

because the inner product is a linear functional of $\mathbf{r}$. From this, we deduce that the constraint from (5), viz

$$\mathrm{Re}\big\langle \mathbf{s}^{\odot\langle 2p-1\rangle}, \mathbf{r} - \langle \mathbf{s},\mathbf{r}\rangle \mathbf{s}\big\rangle = 0, \quad \text{for all } \mathbf{r} \in \mathcal{S}_A$$

is equivalent to

$$\big\langle \mathbf{s}^{\odot\langle 2p-1\rangle}, \mathbf{r} - \langle \mathbf{s},\mathbf{r}\rangle \mathbf{s}\big\rangle = 0, \quad \text{for all } \mathbf{r} \in \mathcal{S}_A$$

i.e., consideration of the real part by itself is redundant. This final form can be written as

$$\mathbf{s}^{\odot\langle 2p-1\rangle} \perp \mathbf{r} - \langle \mathbf{s},\mathbf{r}\rangle \mathbf{s}, \quad \text{for all } \mathbf{r} \in \mathcal{S}_A. \qquad (6)$$

We now introduce the orthogonal complement to the set of attainable responses, denoted $\mathcal{S}_A^\perp$, and likewise orthogonally decompose $\mathcal{S}_A$ into a 1-D subspace colinear with a candidate $\mathbf{s} \in \mathcal{S}_A$ and its resulting orthogonal complement

$$\begin{aligned}\mathcal{S}_{A\mathbf{s}} &\triangleq \{\mathbf{x} \in \mathcal{S}_A : \mathbf{x} = \alpha\mathbf{s} \text{ for some scalar } \alpha\} \\ \mathcal{S}_{A\mathbf{s}}^\perp &\triangleq \{\mathbf{x} \in \mathcal{S}_A : \mathbf{x} \perp \mathbf{s}\}.\end{aligned}$$

The entire space ℓ_2 then admits the orthogonal decomposition $\mathcal{S}_{A\mathbf{s}} \oplus \mathcal{S}_{A\mathbf{s}}^\perp \oplus \mathcal{S}_A^\perp$.

Now, for all $\mathbf{r} \in \mathcal{S}_A$, the term $\mathbf{r} - \langle \mathbf{s},\mathbf{r}\rangle\mathbf{s}$ is in $\mathcal{S}_A$ and orthogonal to $\mathbf{s}$ because $\mathbf{s}$ is scaled to unit ℓ_2 norm; this gives $\mathbf{r} - \langle \mathbf{s},\mathbf{r}\rangle\mathbf{s} \in \mathcal{S}_{A\mathbf{s}}^\perp$ for all $\mathbf{r} \in \mathcal{S}_A$. The orthogonality condition (6) then holds if and only if the projection of $\mathbf{s}^{\odot\langle 2p-1\rangle}$ along $\mathcal{S}_{A\mathbf{s}}^\perp$ vanishes. This amounts to saying that the projection of $\mathbf{s}^{\odot\langle 2p-1\rangle}$ along $\mathcal{S}_A$ reduces to its projection along $\mathcal{S}_{A\mathbf{s}}$, i.e.,

$$\mathcal{P}_A\big(\mathbf{s}^{\odot\langle 2p-1\rangle}\big) = \alpha\mathbf{s}, \quad \text{for some scalar } \alpha.$$

With $\mathbf{s}$ scaled to unit ℓ_2 norm we finally see that

$$\begin{aligned}\alpha &= \langle \mathbf{s}, \alpha\mathbf{s}\rangle \\ &= \big\langle \mathbf{s}, \mathcal{P}_A\big(\mathbf{s}^{\odot\langle 2p-1\rangle}\big)\big\rangle \\ &= \big\langle \mathbf{s}, \mathbf{s}^{\odot\langle 2p-1\rangle}\big\rangle \quad \text{because } \mathbf{s} \in \mathcal{S}_A \\ &= \sum_k |s_k|^{2p} = [D_{2p}(\mathbf{s})]^{2p}\end{aligned}$$

to complete the proof. □

Remark 1: Note that a scale-factor constraint on $\mathbf{s}$ is arbitrary, as expected; if $\mathbf{s}$ satisfies Theorem 6, so does $\beta\mathbf{s}$ upon adjusting the value of the scalar α.

Remark 2: If we consider the sufficient order case ($\mathcal{P}_A = \mathbf{I}$), then the statement of Theorem 6 reduces to $\mathbf{s}^{\odot\langle 2p-1\rangle} - \alpha\,\mathbf{s} = \mathbf{0}$, which reads componentwise as

$$s_k(|s_k|^{2p-2} - \alpha) = 0, \quad \text{for all } k.$$

This says that all nonzero terms have the same amplitude ($= \sqrt[2p-2]{\alpha}$), as first deduced by Godard [7], but using a different proof that does not readily accommodate the undermodeled case. If two or more terms are nonzero, then the stationary point is a saddle point or minimum (see Yi and Ding [4] for the case $2p = 4$; that proof extends readily to the case $2p > 4$ as well).

For the undermodeled case ($\mathcal{P}_A \neq \mathbf{I}$), we obtain $\mathbf{s}^{\odot\langle 2p-1\rangle} - \alpha\mathbf{s} = \mathbf{b}$ for some $\mathbf{b} \in \mathcal{S}_A^\perp$. As $\mathbf{b}$ is not known *a priori*, this relation does not reveal the form of $\mathbf{s}$. Since a closed-form solution of all stationary points does not appear at hand for undermodeled cases, the next section develops an iterative procedure that converges to a local maximum of $D_{2p}(\mathbf{s})$.

V. Iterative Construction of Maxima of $D_{2p}(\mathbf{s})$

The stationary points fulfilling Theorem 6 may be identified as the fixed points of a nonlinear map $\mathbf{q} = T_{2p}(\mathbf{s})$, which sends a unit sphere of $\mathcal{S}_A$ to itself, which is defined as

1) Take $\mathbf{s} \in \mathcal{S}_A$, scaled to unit norm $||\mathbf{s}|| = 1$ (the choice of norm is arbitrary for now).
2) Project its Hadamard power onto $\mathcal{S}_A$

$$\mathbf{v} = \mathcal{P}_A\big(\mathbf{s}^{\odot\langle 2p-1\rangle}\big).$$

3) Scale the result to unit norm $\mathbf{q} = \mathbf{v}/\|\mathbf{v}\|$ (using the same norm as in step 1).

It is straightforward to check that the fixed points of this map (i.e., those unit-norm $\mathbf{s}$ in $\mathcal{S}_A$ for which $\mathbf{q} = T_{2p}(\mathbf{s}) = \mathbf{s}$) are precisely the stationary points fulfilling Theorem 6. The following inequality applies whenever $\mathbf{s}$ is not a fixed point.

Theorem 7: Let $\mathbf{s} \in \mathcal{S}_A$ be scaled to unit norm. If $\mathbf{q} = T_{2p}(\mathbf{s}) \neq \mathbf{s}$, then $D_{2p}(\mathbf{q}) > D_{2p}(\mathbf{s})$.

We offer some remarks prior to the proof.

Remark 3: It follows readily that the iterative procedure

$$\mathbf{s}_{(i+1)} = T_{2p}\big(\mathbf{s}_{(i)}\big) \tag{7}$$

in which the subscript (i) denotes the iteration number, will approach a local maximum of $D_{2p}(\mathbf{s})$, save for an exceptional set of initial conditions on $\mathbf{s}_{(0)}$.[2] Which maximum is approached depends on the initialization.

Remark 4: If we consider the sufficient-order case ($\mathcal{P}_A = \mathbf{I}$), then the iteration (7) reduces to the super-exponential algorithm proposed by Shalvi and Weinstein [5]; if $\mathbf{s}_{(0)} \in \mathcal{S}_n$, then successive iterates $\mathbf{s}_{(i)}$ converge to $\mathbf{e}_n$ at a superexponential rate. (If we consider the ℓ_∞ norm in steps 1 and 3, then $\|\mathbf{e}_n - \mathbf{s}_{(i+1)}\|_\infty = \|\mathbf{e}_n - \mathbf{s}_{(i)}\|_\infty^{2p-1}$; by induction, we obtain $\|\mathbf{e}_n - \mathbf{s}_{(i)}\|_\infty = (\|\mathbf{e}_n - \mathbf{s}_{(0)}\|_\infty)^{(2p-1)^i}$. This is "superexponentially" contractive because if $\mathbf{s}_0 \in \mathcal{S}_n$ with $\|\mathbf{s}_{(0)}\|_\infty = 1$, then $\|\mathbf{e}_n - \mathbf{s}_{(0)}\|_\infty < 1$). For the undermodeled case ($\mathcal{P}_A \neq \mathbf{I}$), the iteration (7) may be found in [5, (22)–(24)], proposed therein as a realizable approximation to the superexponential algorithm; see also [6]. Shalvi and Weinstein argue that if $\mathcal{P}_A(\mathbf{s}_{(i)}^{\odot\langle 2p-1\rangle}) \approx \mathbf{s}_{(i)}^{\odot\langle 2p-1\rangle}$ for all i, then the trajectory in question should still converge to a solution near a Wiener response. It is not clear from [5], though, how the approximation $\mathcal{P}_A(\mathbf{s}_{(i)}^{\odot\langle 2p-1\rangle}) \approx \mathbf{s}_{(i)}^{\odot\langle 2p-1\rangle}$ should be quantified to ensure convergence. The proof of Theorem 7 to follow, and, hence, the convergence inference of the previous remark, appeals to no approximation.

Proof: For convenience, we assume ℓ_2 normalization in the algorithm: $\|\mathbf{s}\|_2 = \|\mathbf{q}\|_2 = 1$. The inequality $D_{2p}(\mathbf{q}) > D_{2p}(\mathbf{s})$ will then follow by showing that

$$\sum_k |q_k|^{2p} > \sum_k |s_k|^{2p}, \quad \text{whenever } \mathbf{q} = T_{2p}(\mathbf{s}) \neq \mathbf{s}.$$

We begin with the identity

$$\begin{aligned}\sum_k |q_k|^{2p} - \sum_k |s_k|^{2p} &= \big\langle \mathbf{q}^{\odot\langle 2p-1\rangle}, \mathbf{q}\big\rangle - \big\langle \mathbf{s}^{\odot\langle 2p-1\rangle}, \mathbf{s}\big\rangle \\ &= \mathrm{Re}\big\langle \mathbf{q}^{\odot\langle 2p-1\rangle}, \mathbf{q}\big\rangle - \mathrm{Re}\big\langle \mathbf{s}^{\odot\langle 2p-1\rangle}, \mathbf{s}\big\rangle \\ &= \mathrm{Re}\big\langle \mathbf{q}^{\odot\langle 2p-1\rangle} - \mathbf{s}^{\odot\langle 2p-1\rangle}, \mathbf{q}\big\rangle + \mathrm{Re}\big\langle \mathbf{s}^{\odot\langle 2p-1\rangle}, \mathbf{q} - \mathbf{s}\big\rangle.\end{aligned} \tag{8}$$

The proof proceeds in two steps.

1) We show first that the inequality

$$\mathrm{Re}\big\langle \mathbf{q}^{\odot\langle 2p-1\rangle} - \mathbf{s}^{\odot\langle 2p-1\rangle}, \mathbf{q}\big\rangle \geq (2p-1)\,\mathrm{Re}\big\langle \mathbf{s}^{\odot\langle 2p-1\rangle}, \mathbf{q} - \mathbf{s}\big\rangle$$

is valid for any two vectors $\mathbf{q}$ and $\mathbf{s}$ in ℓ_2, with strict inequality whenever $\mathbf{q} \neq \mathbf{s}$. This will give, with respect to (8)

$$\sum_k |q_k|^{2p} - \sum_k |s_k|^{2p} > 2p\,\mathrm{Re}\big\langle \mathbf{s}^{\odot\langle 2p-1\rangle}, \mathbf{q} - \mathbf{s}\big\rangle.$$

2) We then recognize that

$$\mathrm{Re}\big\langle \mathbf{s}^{\odot\langle 2p-1\rangle}, \mathbf{q} - \mathbf{s}\big\rangle > 0$$

whenever $\mathbf{q} = T_{2p}(\mathbf{s}) \neq \mathbf{s}$ by virtue of $\mathbf{q}$ being a scaled projection of $\mathbf{s}^{\odot\langle 2p-1\rangle}$.

For the first part, let $\mathbf{x} = \mathbf{q} - \mathbf{s}$; a direct calculation shows that

$$\begin{aligned}&\mathrm{Re}\big\langle \mathbf{q}^{\odot\langle 2p-1\rangle} - \mathbf{s}^{\odot\langle 2p-1\rangle}, \mathbf{q}\big\rangle - (2p-1)\,\mathrm{Re}\big\langle \mathbf{s}^{\odot\langle 2p-1\rangle}, \mathbf{q} - \mathbf{s}\big\rangle \\ &\quad = \sum_k \big(|s_k + x_k|^{2p} - |s_k|^{2p} - 2p|s_k|^{2p-2}\,\mathrm{Re}(s_k^* x_k)\big) \\ &\quad = \sum_k F(s_k, x_k)\end{aligned} \tag{9}$$

in terms of the two-variable function

$$F(s, x) = |s + x|^{2p} - |s|^{2p} - 2p|s|^{2p-2}\,\mathrm{Re}(s^* x).$$

We show now that $F(s, x) \geq 0$ for all complex s and x. To this end, let $x = r\,e^{j\phi}$, with $r = |x|$. We may expand the terms of $F(s, x)$ involving x as

$$\begin{aligned}\mathrm{Re}(s^* x) &= r\,\mathrm{Re}(s^* e^{j\phi}) \\ |s + x|^{2p} &= (|s|^2 + 2r\,\mathrm{Re}(s^* e^{j\phi}) + r^2)^p.\end{aligned}$$

This gives $F(s, re^{j\phi})$ as a differentiable function of r, whose partial derivative is

$$\begin{aligned}f(s, r, \phi) &\triangleq \frac{\partial F(s, re^{j\phi})}{\partial r} \\ &= 2p\{r\,|s + re^{j\phi}|^{2p-2} \\ &\quad + \mathrm{Re}(s^* e^{j\phi})(|s + re^{j\phi}|^{2p-2} - |s|^{2p-2})\}.\end{aligned}$$

If $r = 0$, then $f(s, 0, \phi) = 0$. If $r > 0$, we can use the identity

$$\mathrm{Re}(s^* e^{j\phi}) = \frac{|s + re^{j\phi}|^2 - |s|^2}{2r} - \frac{r}{2}$$

to obtain, for $r > 0$

$$\begin{aligned}f(s, r, \phi) &= pr(|s + x|^{2p-2} + |s|^{2p-2}) \\ &\quad + \frac{p}{r}(|s + x|^2 - |s|^2)(|s + x|^{2p-2} - |s|^{2p-2}) \\ &> 0.\end{aligned}$$

Now, since $F(s, 0) = 0$ for all s, we may write

$$F(s, re^{j\phi}) = \int_0^r f(s, \rho, \phi)\, d\rho \geq 0$$

which is clearly non-negative because the integrand is non-negative and indeed positive whenever $r > 0$. The sum (9) is thus comprised of non-negative terms. If $\mathbf{q} \neq \mathbf{s}$, then for at least one index k, we have $r_k = |x_k| = |q_k - s_k| > 0$ so that the sum (9) is positive to give the inequality of the first part of the proof.

[2] The exceptional set will include, e.g., all saddle points and all crest lines leading to such saddle points.

For the second part, the projection $\mathbf{v} = \mathcal{P}_A(\mathbf{s}^{\odot\langle 2p-1\rangle})$ is the best ℓ_2 approximation to $\mathbf{s}^{\odot\langle 2p-1\rangle}$ in $\mathcal{S}_A$, i.e.,

$$\begin{aligned}\mathbf{v} &= \arg\min_{\mathbf{r}\in\mathcal{S}_A} \left\|\mathbf{s}^{\odot\langle 2p-1\rangle} - \mathbf{r}\right\|_2^2 \\ &= \arg\min_{\mathbf{r}\in\mathcal{S}_A} \left(\left\|\mathbf{s}^{\odot\langle 2p-1\rangle}\right\|_2^2 + \|\mathbf{r}\|_2^2 - 2\,\mathrm{Re}\langle\mathbf{s}^{\odot\langle 2p-1\rangle}, \mathbf{r}\rangle\right).\end{aligned}$$

If we set $\Delta = \|\mathbf{v}\|_2$, then $\mathbf{v}$ remains the optimal solution to the constrained problem

$$\begin{aligned}\mathbf{v} &= \arg\min_{\substack{\mathbf{r}\in\mathcal{S}_A\\ \|\mathbf{r}\|_2=\Delta}} \left\|\mathbf{s}^{\odot\langle 2p-1\rangle} - \mathbf{r}\right\|_2^2 \\ &= \arg\min_{\substack{\mathbf{r}\in\mathcal{S}_A\\ \|\mathbf{r}\|_2=\Delta}} \left(\left\|\mathbf{s}^{\odot\langle 2p-1\rangle}\right\|_2^2 + \Delta^2 - 2\,\mathrm{Re}\langle\mathbf{s}^{\odot\langle 2p-1\rangle}, \mathbf{r}\rangle\right).\end{aligned}$$

From this, we deduce that if $\mathbf{r} \in \mathcal{S}_A$, with $\|\mathbf{r}\|_2 = \Delta$ and $\mathbf{r} \neq \mathbf{v}$, then $\mathrm{Re}\langle\mathbf{s}^{\odot\langle 2p-1\rangle}, \mathbf{v}\rangle > \mathrm{Re}\langle\mathbf{s}^{\odot\langle 2p-1\rangle}, \mathbf{r}\rangle$; dividing through by the constraint $\Delta = \|\mathbf{v}\|_2 = \|\mathbf{r}\|_2$ gives

$$\mathrm{Re}\left\langle\mathbf{s}^{\odot\langle 2p-1\rangle}, \frac{\mathbf{v}}{\|\mathbf{v}\|_2}\right\rangle > \mathrm{Re}\left\langle\mathbf{s}^{\odot\langle 2p-1\rangle}, \frac{\mathbf{r}}{\|\mathbf{r}\|_2}\right\rangle$$

for any $\mathbf{r}$ in $\mathcal{S}_A$ different from $\beta\mathbf{v}$ with $\beta > 0$. Since $\mathbf{q}$ is precisely the scaled projection $\mathbf{v}/\|\mathbf{v}\|_2$, whereas $\mathbf{s} \neq \mathbf{q}$ is not, the second part now follows, to complete the proof. □

Example 4: Here, we consider a fractionally spaced equalizer $(N = 2)$ with channel impulse response

$$\begin{bmatrix}\mathbf{h}_0^T\\ \mathbf{h}_1^T\\ \vdots\\ \mathbf{h}_{14}^T\end{bmatrix} = \begin{bmatrix} 0.5098 & 0.7379\\ 0.0399 & -0.0191\\ 0.1175 & 0.6870\\ 0.4521 & -0.0497\\ 0.4813 & 0.1858\\ 0.1877 & 0.2997\\ -0.0525 & 0.3872\\ 0.6456 & 0.3901\\ 0.7554 & 0.3561\\ -0.0518 & 0.7833\\ 0.2088 & -0.0582\\ 0.3649 & 0.0521\\ 0.2885 & 0.2640\\ 0.7611 & -0.0740\\ -0.0002 & 0.1192\end{bmatrix}$$

using an equalizer of order $L = 5$. We run the algorithm $\mathbf{s}_{(i+1)} = T_4(\mathbf{s}_{(i)})$, with $\mathbf{s}_{(0)}$ initialized at $\mathbf{w}^{(5)}$ (but scaled to unit ℓ_∞ norm, as with successive iterates) as in Fig. 3(a); this initial point is the maximum of $D_\infty(\mathbf{s})$ in $\mathcal{S}_A \cap \mathcal{S}_5$ (Theorem 4). Fig. 3(b) shows the combined response $\mathbf{s}$ at iteration nine; s_{14} begins to overtake s_5 as the dominant term. Fig. 3(c) shows the combined response at interation 20, which has converged to a maximum of $D_4(\mathbf{s})$ in the penetrable part of cone $\mathcal{S}_{14}$. Fig. 3(d) shows the value of $D_4(\mathbf{s}_{(i)})$ versus the iteration number (i), verifying the monotonic rise to its final value. The curve also illustrates how the convergence rate need not be superexponential in undermodeled cases.

For this example, $D_4(\mathbf{s})$ does not contain a maximum in $\mathcal{S}_A \cap \mathcal{S}_5$, although the criteria $D_{2p}(\mathbf{s})$, for all $p \geq 3$, do contain a maximum there. We have observed in many examples that the number of maxima of $D_{2p}(\mathbf{s})$ is upper bounded by the number of maxima of $D_{2(p+1)}(\mathbf{s})$, but whether some intrinsic ordering relation between the number of maxima of $D_{2p}(\mathbf{s})$ and $D_{2(p+1)}(\mathbf{s})$ can be established remains open.

Example 5: A theme in many works [8], [15], [16] is to gauge the proximity between a convergent point of the Godard algorithm [corresponding to maxima of $D_4(\mathbf{s})$] and a nearby Wiener response. Suppose $\mathbf{s}$ lies at a maximum of $D_{2p}(\mathbf{s})$ and that $\mathbf{s}$ is scaled to unit ℓ_∞ norm. If $\mathbf{s} \in \mathcal{S}_n$, we then have

$$\|\mathbf{e}_n - \mathbf{s}\|_\infty = \sup_{k\neq n} |s_k| < 1$$

and

$$\left\|\mathbf{e}_n - \mathbf{s}^{\odot\langle 2p-1\rangle}\right\|_\infty = \sup_{k\neq n} |s_k|^{2p-1} = \|\mathbf{e}_n - \mathbf{s}\|_\infty^{2p-1}.$$

Since $\alpha\mathbf{s} = \mathcal{P}_A\mathbf{s}^{\odot\langle 2p-1\rangle}$ and $\mathbf{w}^{(n)} = \mathcal{P}_A\mathbf{e}_n$, we see that

$$\begin{aligned}\left\|\mathbf{w}^{(n)} - \alpha\mathbf{s}\right\|_\infty &= \left\|\mathcal{P}_A\left(\mathbf{e}_n - \mathbf{s}^{\odot\langle 2p-1\rangle}\right)\right\|_\infty \\ &\leq \|\mathcal{P}_A\|_\infty \cdot \left\|\mathbf{e}_n - \mathbf{s}^{\odot\langle 2p-1\rangle}\right\|_\infty \\ &= \|\mathcal{P}_A\|_\infty \cdot \sup_{k\neq n} |s_k|^{2p-1}.\end{aligned}$$

The right-hand side can be forced smaller by increasing p, which suggests that the maxima of $D_{2p}(\mathbf{s})$ should lie increasingly closer to a Wiener response as p is increased.

$2p$	Sine of Subspace Angle
4	2.763e−02
6	3.182e−03
8	3.738e−04
10	4.343e−05
12	5.005e−06
14	5.745e−07
∞	0

The table above lists the sine of the subspace angle between the Wiener response $\mathbf{w}^{(14)}$ and the vector maximizing $D_{2p}(\mathbf{s})$ within $\mathcal{S}_A \cap \mathcal{S}_{14}$, for successive values of p using the same channel and equalizer length of the previous example. As p increases, the subspace angle becomes vanishly small, indicating increasing proximity to the Wiener response in the cone. Whether such a monotonic decrease versus p is intrinsic, and what implications this may hold concerning the choice of p, requires further study.

VI. Open Eye Measures

We turn our attention now to bounding ℓ_1 and ℓ_2 open eye measures in undermodeled cases; these measures gauge the degree of dominance of a given term in the combined response. If $\mathbf{s} \in \bar{\mathcal{S}}_n$ (so that component s_n is dominant), the two measures appear as

$$\mathrm{OEM}_1(\mathbf{s}) = \left(\sum_{k\neq n} |s_k|\right) \Big/ |s_n|$$

$$\mathrm{OEM}_2(\mathbf{s}) = \left(\sum_{k\neq n} |s_k|^2\right) \Big/ |s_n|^2.$$

These are also called the maximum distortion and intersymbol interference measures, respectively [17]. Either measure vanishes if and only if $\mathbf{s} = \alpha\mathbf{e}_n$ for some nonzero scalar α. If

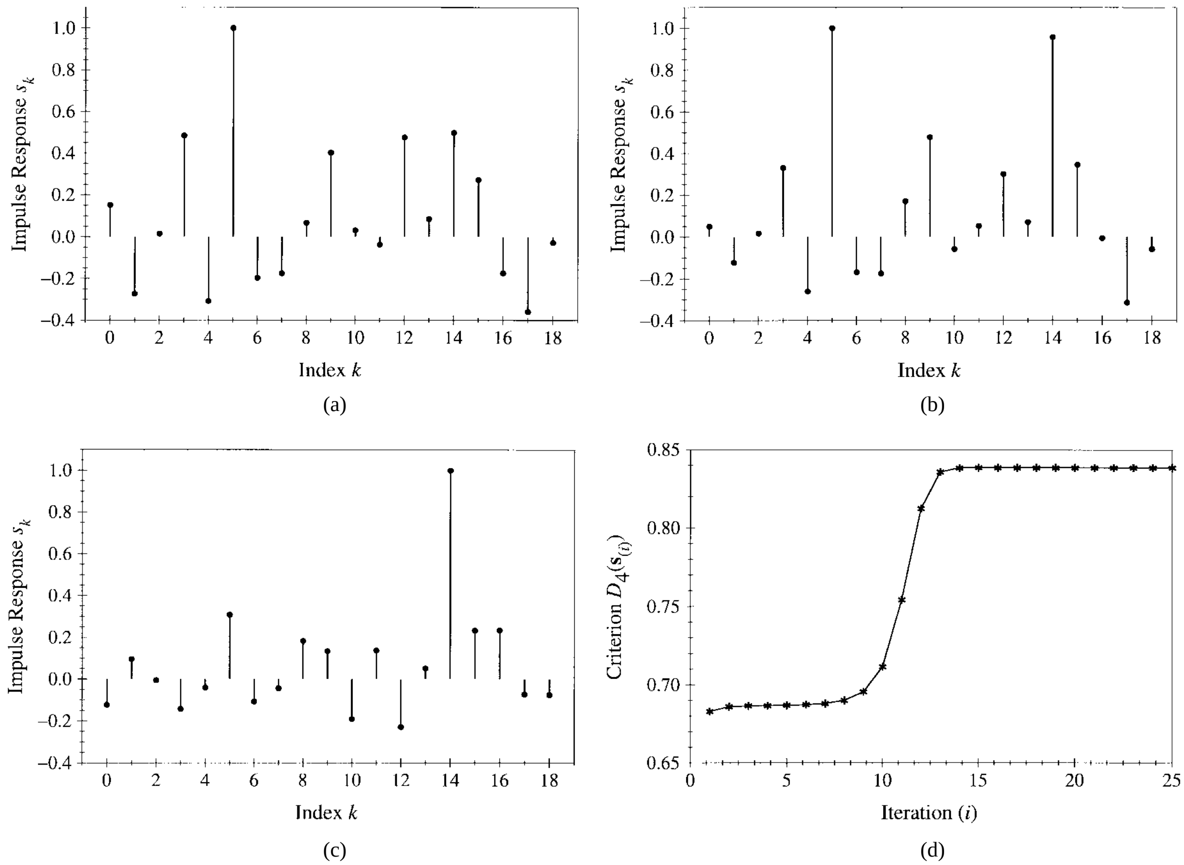

Fig. 3. Illustrating iterative algorithm $\mathbf{s}_{(i+1)} = T_4(\mathbf{s}_{(i)})$. (a) Initialization $\mathbf{s}_0 = \mathbf{w}^{(5)}/\|\mathbf{w}^{(5)}\|_\infty$ in cone $\mathcal{S}_5$. (b) At iteration nine. (c) After convergence, now in cone $\mathcal{S}_{14}$. (d) Monotonic increase of $D_4(\mathbf{s}_{(i)})$.

$\mathbf{e}_n \notin \mathcal{S}_A \cap \mathcal{S}_n$, then an $\mathbf{s} \in \mathcal{S}_A \cap \bar{\mathcal{S}}_n$, which minimizes one measure, need not minimize the other.

The measure $\text{OEM}_2(\mathbf{s})$ is more amenable to analysis; we may verify that

$$\text{OEM}_2(\mathbf{s}) = \frac{1}{[D_\infty(\mathbf{s})]^2} - 1, \quad \text{for all } \mathbf{s} \in \mathcal{S}_A \cap \bar{\mathcal{S}}_n$$

from which follows

$$\frac{d\,\text{OEM}_2(\mathbf{s})}{dD_\infty(\mathbf{s})} = \frac{-2}{[D_\infty(\mathbf{s})]^3} < 0.$$

This shows that $\text{OEM}_2(\mathbf{s})$ and $D_\infty(\mathbf{s})$ are related by a negative monotonic deformation within a given cone so that any minimum of $\text{OEM}_2(\mathbf{s})$ in $\mathcal{S}_A \cap \bar{\mathcal{S}}_n$ must correspond to a maximum of $D_\infty(\mathbf{s})$ in $\mathcal{S}_A \cap \bar{\mathcal{S}}_n$, and vice versa. As a corollary to Theorem 4, if $\mathbf{w}^{(n)} \in \mathcal{S}_n$, then it attains the unique minimum of $\text{OEM}_2(\mathbf{s})$ in $\mathcal{S}_A \cap \bar{\mathcal{S}}_n$; if $\mathbf{w}^{(n)} \notin \mathcal{S}_n$, then any minimum of $\text{OEM}_2(\mathbf{s})$ in $\mathcal{S}_A \cap \bar{\mathcal{S}}_n$ must occur along the boundary $\mathcal{B}_n$.

The ℓ_1 measure $\text{OEM}_1(\mathbf{s})$, however, is more relevant for equalization purposes because it bounds the maximum distortion. For each source constellation, there exists a threshold γ such that, if $\text{OEM}_1(\mathbf{s}) < \gamma$, the intersymbol interference is less than half the distance separating adjacent symbols of the source constellation (in the absence of noise). All residual intersymbol interference may then be removed by passing the equalizer output through a quantizer, which replaces the equalizer output by its nearest neighbor in the source constellation.

We may show that each minimum of $\text{OEM}_1(\mathbf{s})$ corresponds to a maximum of $\|\mathbf{s}\|_\infty/\|\mathbf{s}\|_1$. (The proof mimicks that linking $\text{OEM}_2(\mathbf{s})$ with $\|\mathbf{s}\|_\infty/\|\mathbf{s}\|_2$ above). Deducing the precise minimum value of $\text{OEM}_1(\mathbf{s})$ within a given cone is difficult, however, because $\text{OEM}_1(\mathbf{s})$ is not differentiable in the conventional sense at a minimum. The following result provides at least upper and lower bounds on $\min_{\mathbf{s}} \text{OEM}_1(\mathbf{s})$ within penetrable cones, in terms of Wiener responses.

Theorem 8: Suppose $\mathcal{S}_n$ is penetrable ($\mathcal{S}_A \cap \mathcal{S}_n \neq \emptyset$). For all $\mathbf{s} \in \mathcal{S}_A \cap \bar{\mathcal{S}}_n$

$$\text{OEM}_1(\mathbf{s}) \geq \frac{1 - \mathbf{w}_n^{(n)}}{\sup_{k \neq n} \left|\mathbf{w}_k^{(n)}\right|}.$$

If $\mathbf{w}^{(n)} \in \mathcal{S}_n$, then

$$\min_{\mathbf{s} \in \mathcal{S}_A \cap \mathcal{S}_n} \text{OEM}_1(\mathbf{s}) \leq \text{OEM}_1\big(\mathbf{w}^{(n)}\big).$$

Proof: The second part is trivial since the Wiener response $\mathbf{w}^{(n)}$ is not necessarily optimal with respect to $\text{OEM}_1(\mathbf{s})$.

For the first part, if $\mathbf{s} \in \mathcal{S}_A \cap \bar{\mathcal{S}}_n$, then s_n is positive and dominant. As $\mathbf{w}^{(n)}$ is a reproducing kernel over $\mathcal{S}_A$ [cf. (1)], we may write

$$s_n = \big\langle \mathbf{w}^{(n)}, \mathbf{s} \big\rangle = \sum_k \big[\mathbf{w}_k^{(n)}\big]^* s_k.$$

Since $\mathbf{w}_n^{(n)}$ is real, we can rearrange the above expression as

$$s_n\left(1-\mathbf{w}_n^{(n)}\right)=\sum_{k\neq n}\left[\mathbf{w}_k^{(n)}\right]^* s_k$$

or, since $1-\mathbf{w}_n^{(n)}=\|\mathbf{e}_n-\mathbf{w}^{(n)}\|_2^2\geq 0$ [cf. (3)]

$$1-\mathbf{w}_n^{(n)}=\sum_{k\neq n}\left[\mathbf{w}_k^{(n)}\right]^*\frac{s_k}{s_n}\leq\left(\sup_{k\neq n}\left|\mathbf{w}_k^{(n)}\right|\right)\underbrace{\sum_{k\neq n}\left|\frac{s_k}{s_n}\right|}_{\text{OEM}_1(\mathbf{s})}$$

in which the underbraced term is indeed $\text{OEM}_1(\mathbf{s})$ whenever $\mathbf{s}\in\bar{\mathcal{S}}_n$. Rearrangement of this final expression completes the proof. □

VII. Concluding Remarks

We have derived an analytic characterization of stationary points for a family of blind equalization criteria in undermodeled cases in the single-source, noise-free setting. The family includes at one extreme (D_4), which is the function whose maxima are sought by the Godard and Shalvi–Weinstein schemes, whereas the other extreme (D_∞) was shown to have maxima at Wiener responses. The superexponential algorithms are also incorporated, as we have shown that they seek maxima of D_{2p}, where p is a user-chosen integer.

A key concern in any adaptive filtering scheme admitting multiple convergent points is how many extrema the cost function admits. This number has not been deduced in this work, except for the limiting case $D_\infty(\mathbf{s})$; Theorem 4 shows that the number of maxima of $D_\infty(\mathbf{s})$ is the number of Wiener responses that lie in their target cones. How to explicitly quantify this latter number without inspecting all Wiener responses, however, remains open.

We have also observed that a given cone may be penetrable without a convergent point existing there, contrary to what is suggested by the supporting arguments of [18, Theorem 3.1]. Example 4 presents a case in which a given cone is penetrable and contains a maximum of D_{2p} for all $p\geq 3$ but does not contain a maximum for $p=2$. An algorithm seeking a maximum of D_4 is thus obliged to migrate to another cone containing a better solution, i.e., to where the "grass is greener." This behavior suggests that the Godard or Shalvi–Weinstein schemes may favor convergence to proximities of better Wiener filters over worse ones.

The influence of channel noise and/or multiple sources has not been incorporated in this work. These artefacts, like insufficient equalizer lengths, can prohibit perfect equalization, and extensions of our results to this more general setting are currently under study.

Appendix
Infinite Length versus Sufficient Order

Here, we provide a short verification that if the channel and equalizer impulse response lengths are both infinite, a sufficient order case $(\mathcal{P}_A=\mathbf{I})$ is obtained if and only if the transfer function vector $\mathbf{h}(z)$ of dimensions $N\times 1$ fulfills $\mathbf{h}(z)\neq\mathbf{0}$ for all $|z|\geq 1$.

We note first that the constraint $\mathbf{h}(z)\neq\mathbf{0}$ for all $|z|\geq 1$ is a minimum-phase condition (see, e.g., [19] for connections to spectral factorization); this is equivalent to the existence of a stable and causal equalizer transfer function vector of dimensions $1\times N$ (call it $\mathbf{g}_0(z)$) for which

$$\mathbf{g}_0(z)\,\mathbf{h}(z)=1,\quad \text{for all } z.$$

An arbitrary causal combined transfer function $S(z)=\sum_{k\geq 0}s_k\,z^{-k}$ can then be attained using the causal (and possibly infinite length) equalizer $\mathbf{g}(z)=S(z)\,\mathbf{g}_0(z)$. This shows that all causal combined responses in ℓ_2 are attainable so that $\mathcal{P}_A=\mathbf{I}$.

Conversely, if $\mathbf{h}(z_0)=\mathbf{0}$ for some $|z_0|\geq 1$, then each scalar transfer function in the vector $\mathbf{h}(z)$ must contain a nonminimum phase factor $(1-z_0z^{-1})$. The combined transfer function $S(z)=\mathbf{g}(z)\,\mathbf{h}(z)$ must also contain this nonminimum phase zero since a stable and causal equalizer $\mathbf{g}(z)$ must be devoid of poles in $|z|\geq 1$. If $|z_0|>1$, then $S(z)$ is orthogonal to the stable and causal function $z_0^*/(z_0^*-z^{-1})=\sum_{k\geq 0}(1/z_0^*)^k z^{-k}$ as an exercise will verify. This shows that $\mathcal{S}_A$ has a nontrivial orthogonal complement in ℓ_2 containing the vector

$$[1,1/z_0^*,(1/z_0^*)^2,\ldots,(1/z_0^*)^k,\ldots]^T$$

so that $\mathcal{P}_A\neq\mathbf{I}$. The case $|z_0|=1$ may be handled by a limiting argument.

References

[1] L. Tong, G. Xu, and T. Kailath, "Blind identification based on second-order statistics: A time domain approach," *IEEE Trans. Inform. Theory*, vol. 40, pp. 340–349, Mar. 1994.

[2] E. Moulines, P. Duhamel, J.-F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Processing*, vol. 43, pp. 516–525, Feb. 1995.

[3] D. T. M. Slock, "Blind fractionally-spaced equalization, perfect reconstruction filter banks, and multichannel prediction," in *Proc. ICASSP*, Adelaide, Australia, Apr. 1994, vol. 4, pp. 585–588.

[4] Y. Li and Z. Ding, "Global convergence of fractionally spaced Godard (CMA) adaptive equalizers," *IEEE Trans. Signal Processing*, vol. 44, pp. 818–826, Apr. 1996.

[5] O. Shalvi and E. Weinstein, "Super-exponential methods for blind deconvolution," *IEEE Trans. Inform. Theory*, vol. 39, pp. 504–519, Mar. 1993.

[6] Z. Ding, "On convergence analysis of fractionally spaced adaptive blind equalizers," *IEEE Trans. Signal Processing*, vol. 45, pp. 650–657, Mar. 1997.

[7] D. N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. Commun.*, vol. COMM-28, pp. 1867–1875, Nov. 1980.

[8] O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of nonminimum phase systems (channels)," *IEEE Trans. Inform. Theory*, vol. 36, pp. 312–321, Mar. 1990.

[9] D. Donoho, "On minimum entropy deconvolution," in *Applied Time Series Analysis II*, D. F. Findley, Ed. New York: Academic, 1981.

[10] J. R. Treichler and B. G. Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, pp. 459–472, Apr. 1983.

[11] J. M. Mendel, "Tutorial on higher order statistics (spectra) in signal processing and system theory: Theoretical results and some applications," *Proc. IEEE*, vol. 79, pp. 278–305, Mar. 1991.

[12] P. A. Regalia, "On the equivalence between the Godard and Shalvi-Weinstein schemes of blind equalization," accepted for publication in *Signal Process.*

[13] D. G. Luenberger, *Optimization by Vector Space Methods*. New York: Wiley, 1969.

[14] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton Univ. Press, 1970.

[15] I. Fijalkow, A. Touzni, and J. R. Treichler, "Fractionally spaced equalization using CMA: Robustness to channel noise and lack of disparity," *IEEE Trans. Signal Processing*, vol. 45, pp. 56–66, Jan. 1997.
[16] H. Zeng and L. Tong, "The MSE performance of constant modulus receivers," in *Proc. ICASSP*, Munich, Germany, Apr. 1997, vol. 5, pp. 3577–3580.
[17] J. G. Proakis, *Digital Communications*. New York: McGraw-Hill, 1983.
[18] Y. Li, K. J. R. Liu, and Z. Ding, "Length- and cost-dependent local minima of unconstrained blind channel equalizers," *IEEE Trans. Signal Processing*, vol. 44, pp. 2726–2735, Nov. 1996.
[19] I. Fijalkow and Ph. Loubaton, "Identification of rank one rational spectral densities from noisy observations," *Syst. Contr. Lett.*, vol. 24, pp. 201–205, 1995.

Phillip A. Regalia (SM'96) was born in Walnut Creek, CA, in 1962. He received the B.Sc. (with highest honors), M.Sc., and Ph.D. degrees in electrical and computer engineering from the University of California, Santa Barbara, in 1985, 1987, and 1988, respectively, and the Habilitation à Diriger des Recherches from the University of Paris, Orsay, France, in 1994.

He is presently Professor at the Institut National des Télecommunications, Evry, France, with research interests focused on adaptive signal processing and algorithm design. He is the author of the book *Adaptive IIR Filtering in Signal Processing and Control* (New York: Marcel Dekker, 1995), a member of the editorial board of *Signal Processing*, an editor for the *International Journal of Adaptive Control and Signal Processing*, and an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING.

Mamadou Mboup was born in Kaolack, Sénégal, in 1962. He received the Diplôme d'Ingénieur from the Ecole Nationale d'Ingénieurs de Tunis (ENIT), Tunis, Tunisia, in 1988 and the Doctorat en Sciences from the University of Paris Sud, Orsay, France, in 1992.

From 1989 to 1996, he was with the Laboratoire des Signaux et Systèmes, National Scientific Research Council (CNRS) of France, as an Associate Researcher. Since 1993, he has been with the Department of Mathematics and Computer Science, René Descartes University—Paris V, where he is presently an Assistant Professor. His research interests include adaptive filters, infinite impulse response system identification, rational approximation including matrix-valued rational functions, and blind channel equalization.