It is easy to check that σ_n is a length function of order n [13, Appendix, Lemma A.3]. Define a partial function $f: \hat{A}^* \times \{0,1,2,\cdots\} \to \{0,1,2,\cdots\}$ by

$$f(y,n) = \sigma_n(y), \qquad \text{for } y \in \hat{A}^n \text{ and } n \geq 2N. \tag{55}$$

Clearly, f is a partial recursive function. By Theorem 3.2 of [4], there exists a partial recursive function $g: B^* \times \{0,1,2,\cdots\} \to \hat{A}^*$ that satisfies for $n \geq 2N$ the following:

a) the domain $D(n)$ of $g(\cdot,n)$ is a prefix code, and $g(\cdot,n)$: $D(n) \to \hat{A}^n$ is one to one and onto,

b) if $g(z,n) = y$, then $y \in \hat{A}^n$ and $l(z) = \sigma_n(y)$,

where $B = \{0,1\}$ and B^* is the set of all finite words from B. By the definition of conditional Chaitin complexity (in old fashion), for $y \in \hat{A}^*$ with $l(y) \geq 2N$,

$$\begin{aligned} C(y|l(y)) &\leq \min\{l(z)|z \in B^*, g(z,l(y)) = y\} + O(1) \\ &= \sigma_{l(y)}(y) + O(1) \\ &= \min_{1\leq j\leq N} \left\{ \sum_{\substack{i\equiv j \pmod N \\ 1\leq i\leq l(y)-N+1}} \sigma\left(y_i^{i+N-1}\right) \right\} + O(1). \end{aligned} \tag{56}$$

ACKNOWLEDGMENT

The authors wish to thank the anonymous reviewers for their helpful comments and suggestions that improved the presentation of this correspondence and strengthened the earlier version of Theorem 3.

REFERENCES

[1] T. Berger, *Rate Distortion Theory.* Englewood Cliffs, NJ: Prentice-Hall, 1971.

[2] A. N. Kolmogorov, "Logical basis for information theory and probability theory," *IEEE Trans. Inform. Theory*, vol. IT-14, pp. 662–664, Sept. 1968.

[3] J. Ziv, "Distortion-rate theory for individual sequences," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 137–143, Mar. 1980.

[4] G. J. Chaitin, "A theory of program size formally identical to information theory," *J. ACM*, vol. 22, pp. 329–340, July 1975.

[5] S. K. Leung-Yan-Cheong and T. M. Cover, "Some equivalences between Shannon entropy and Kolmogorov complexity," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 331–338, 1978.

[6] E. H. Yang, "The proof of Levin's conjecture," *Chinese Sci. Bull.*, vol. 34, pp. 1761–1765, Nov. 1989.

[7] L. K. Zvonkin and L. A. Levin, "Complexity of finite objects and justification of the notion of information and randomness in the framework of the theory of algorithms," *Usp. Mat. Nauk*, vol. 25, pp. 85–127, 1970.

[8] T. M. Cover, P. Gacs, and R. M. Gray, "Kolmogorov's contributions to information theory and algorithmic complexity," *Ann. Probab.*, vol. 17, pp. 840–865, 1989.

[9] J. C. Kieffer, "Extension of source coding theorems for block codes to sliding-block codes," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 679–692, Nov. 1980.

[10] ——, " On the minimum rate for strong universal block coding of a class of ergodic sources," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 693–702, Nov. 1980.

[11] ——, "Sample converses in source coding theory," *IEEE Trans. Inform. Theory*, vol. 37, pp. 263–268, Mar. 1991.

[12] D. S. Ornstein and P. C. Shields, "Universal almost sure data compression," *Ann. Probab.*, vol. 18, pp. 441–452, 1990.

[13] J. C. Kieffer, "Finite-state adaptive block to variable-length noiseless coding of a nonstationary information source," *IEEE Trans. Inform. Theory*, vol. 35, pp. 1259–1263, Nov. 1989.

Blind Equalization Without Gain Identification

Sergio Verdú, *Fellow, IEEE*, Brian D. O. Anderson, *Fellow, IEEE*, and Rodney A. Kennedy, *Member, IEEE*

Abstract— **Blind equalization up to a constant gain of linear time-invariant channels is studied. Dropping the requirement of gain identification allows equalizer anchoring. This results in the elimination of a degree of freedom that causes ill-convergence of conventional blind equalizers, and affords the possibility of using simple update rules based on the stochastic approximation of output energy. Unlike conventional blind equalizers, truncations of the nonrecursive infinite-dimensional realizations of those equalizers inherit the convergence properties of their infinitely parametrized counterparts. A globally convergent blind recursive equalizer for channels without all-pass sections is obtained based on the exact equalization of the minimum-phase part of the channel and the identification of its nonminimum-phase zeros.**

Index Terms—**Blind equalization, deconvolution, ARMA models, adaptive filtering.**

I. Introduction

Finite-dimensional discrete-time linear time-invariant systems are popular models for digital communication channels that introduce intersymbol interference. In many (but not all) situations intersymbol interference is removed prior to data demodulation by means of an equalizer—a linear time-invariant system whose transfer function is equal to the inverse of the channel transfer function. If the receiver does not know the actual transfer function of the channel, the need arises for an adaptive equalizer which is updated using the channel outputs. In addition, classical adaptive equalization methods [12] rely on the input being a training sequence of data which is known by the receiver.

The objective of *blind equalization* is to drop the requirement of a training sequence which in many applications (such as multiuser channels) is too cumbersome to be realistic. Thus, a blind equalizer has access to the output, but not the input, of the channel.

That information is enough to identify the channel (asymptotically) because the input data is a non-Gaussian i.i.d. sequence. (Although the results hold for other modulations, we assume throughout, for the sake of clarity, that the input data is i.i.d. equally likely to be +1 or −1.) Then, the channel coefficients can be obtained by solving systems of equations dependent on higher-order statistics of the channel output sequence (e.g., [7]). Intense research efforts are currently under way in order to make such an indirect solution a viable alternative for on-line equalization. Instead, blind equalization imposes a specific structure on the adaptive scheme so that it can be easily implemented: the equalizer coefficients are updated according to a stochastic approximation scheme governed by a *cost function* that satisfies the following admissibility conditions:

C1) it depends on the input data and the unknown channel only through the channel output;

Manuscript received January 3, 1991; revised May 26, 1992. This work was supported in part by NSF PYI Award ECSE-8857689. This work was presented in part at the Conference on Information Sciences and Systems, Baltimore, MD, March 1991.

S. Verdú is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544.

B. D. O. Anderson and R. A. Kennedy are with the Department of Systems Engineering, Australian National University, Canberra, ACT 2601, Australia.

IEEE Log Number 920318.

0018–9448/93$03.00 © 1993 IEEE

C2) the local minima of its expected value (as a function of the equalizer coefficients) occur at systems which differ from the channel inverse transfer function by at most an arbitrary delay and a change of sign.

Because of the practical importance and conceptual interest of blind equalization, the search for an admissable cost function has attracted the attention of a number of researchers during the last fifteen years. Admissible cost functions have been found for doubly-infinite transversal equalizers [3], [6], [8], [15]. A shortcoming of these solutions is that their *causal* implementations are known to converge globally only for minimum-phase channels. (A restriction that arises from the invertibility of the semi-infinite channel convolution operator discussed in [5].) More importantly, *finite-dimensional* (realizable) approximations of those blind equalizers exhibiting global convergence have not been reported. Even within the domain of relatively simple classes of minimum-phase channels (e.g., first-order autoregressive), implementable blind equalization remains an open problem (see [10] for an up-to-date account of the main efforts and the fundamental issues in this open problem.) As in [4], [5], we study (in addition to infinite-dimensional nonrecursive equalizers) finite-dimensional implementable blind equalizers which achieve the inverse of the channel transfer function exactly. Our starting point is the observation (made previously in [19]) that condition C2) for the admissibility of the cost function is unnecessarily restrictive. Indeed, because of the symmetry of the alphabet $\{+1, -1\}$, it is irrelevant whether the data is recovered exactly or the equalizer introduces an arbitrary constant gain, even if that gain is *a priori* unknown. Note that this remains true even in the presence of noise at the receiver input (as the signal-to-noise ratio is unchanged). Even if the input alphabet is not polarity-symmetric, it is perfectly tolerable to remove the intersymbol interference leaving a residual gain, which can be easily estimated if necessary for demodulation purposes by an automatic gain control subsystem. Therefore, Condition 2 is replaced by the following:

C2′) The local minima of the expected value of the cost function occur at systems which differ from the channel inverse transfer function by at most an arbitrary delay and an arbitrary gain factor. Those systems will be referred to in the sequel as *valid equalizers.*

In Section II, we solve the open problem motivated by the counterexample in [5] of whether there exist finitely-parametrized blind equalizers for which it is possible to prove global convergence as long as the channel can be equalized by an FIR. Using the cost function and the equalizer *anchoring* proposed in Section II, we take two different approaches in Section III in order to deal with arbitrary stable channels. The first one is the traditional approach of double-infinite nonrecursive equalization, for which, unlike the cost functions considered in the past, the convex cost function considered here allows the proof of convergence of truncated versions of the equalizer to approximations of the desired channel inverse. The second approach is based on a recursive blind equalizer that exhibits global convergence to the desired system for any stable channel without all-pass factors.

II. Blind Equalization of Autoregressive Channels

Assume that the channel is described by the difference equation:

$$r_k = \sum_{i=1}^{N} a_i r_{k-i} + G x_k, \tag{1}$$

where $\{x_k\}$ is the data sequence, $\{r_k\}$ is the channel output sequence (Fig. 1), and the receiver knows that the channel is autoregressive and the value of N. If the receiver knew the channel coefficients,

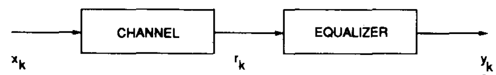

Fig. 1. Channel equalization.

$(a_1, \cdots, a_N)$ (a scaled version of) the input data could be simply recovered with an FIR equalizer:

$$y_k = r_k - \sum_{i=1}^{N} a_i r_{k-i}. \tag{2}$$

In the absence of such knowledge, the output of an infinitely long nonrecursive equalizer,

$$y_k = \sum_{i=0}^{\infty} \lambda_i r_{k-i}, \tag{3}$$

will asymptotically coincide with $\pm\{x_k\}$ if the equalizer is updated according to the Godard algorithm [8] or the Shalvi–Weinstein algorithm [15]. Even though those algorithms assume a doubly-infinite equalizer, in this case a causal equalizer is sufficient because the invertibility condition of [4] is satisfied for an autoregressive channel. No such convergence property has been shown for any realizable (i.e., finitely parametrized) blind equalizer. In particular, it has been shown in [4] that the FIR equalizer

$$y_k = \sum_{i=0}^{N} \lambda_i r_{k-i} \tag{4}$$

may end up converging to local minima instead of the desired solution $\pm \frac{1}{G}[1, -a_1, \cdots, -a_N]$. For example, if $N = 1$, $a_1 = -\alpha$, and $G = 1$, see [5] the Godard equalizer has four local minima located at $\pm[1 - \alpha]$ and $\pm\left[\sqrt{1-\alpha^4}/\sqrt{1+5\alpha^2}\right]$. The location of those points of convergence is continuous in α at $\alpha = 0$, because if $\alpha = 0$, both $\pm[1\ 0]$ and $\pm[0\ 1]$ are valid equalizers. This behavior is a consequence of the overparametrization of the equalizer brought about by the requirement of Condition C2). However, as we argued in Section I, Condition C2′) is all we really need and the natural choice suggested by (2) is to fix the first equalizer coefficient to 1 and set the equalizer structure

$$y_k = r_k - \sum_{i=1}^{N} \lambda_i r_{k-i}. \tag{5}$$

Next, we will propose a simple algorithm for updating $(\lambda_1 \cdots \lambda_N)$ based on the observation of $\{r_k\}$ which converges to $(a_1 \cdots a_N)$ as long as the channel is stable. For now, notice that in contrast to (2), for every $(a_1 \cdots a_N)$ there is only *one* valid equalizer (5). In order to introduce our proposed cost function note that the combined transfer function of the cascade of channel and equalizer is equal to

$$G\frac{1 - \lambda_1 z^{-1} - \cdots - \lambda_N z^{-N}}{1 - a_1 z^{-1} - \cdots - a_N z^{-N}} = G\left(1 + \lambda_1 z^{-1} + \lambda_1 z^{-2} + \cdots\right)$$

and therefore the energy of the output is (since the input is i.i.d.) equal to

$$G^2 + G^2 \sum_{i=1}^{\infty} \lambda_i^2 \geq G^2$$

with equality if and only if $(\lambda_1, \cdots, \lambda_N) = (a_1, \cdots, a_N)$. Therefore, the output energy has a global minimum when the equalizer is equal to the inverse of the channel.

This motivates the choice of the cost function as proportional to the instantaneous value of the output energy: $y_k^2/2$. (The same conclusion would be achieved with any other absolute moment of the output.) Notice that this cost function cannot be used with the classical approach to blind equalization (CMA, Godard, etc.) where all FIR coefficients are degrees of freedom, as the global minimum would correspond to an equalizer with zero transfer function. The gradient of the cost function with respect to the equalizer coefficients: $\lambda = [\lambda_1, \cdots, \lambda_N]^T$ is

$$\nabla y_k^2/2 = -y_k \boldsymbol{r}_{k-1}, \tag{6}$$

where $\boldsymbol{r}_{k-1} = [r_{k-1} r_{k-2} \cdots r_{k-N}]^T$, and the updating of the equalizer coefficients proceeds according to

$$\lambda^{(k+1)} = \lambda^{(k)} + \mu y_k \boldsymbol{r}_{k-1}, \tag{7}$$

where μ is the algorithm step size. It remains to check whether the proposed cost function satisfies Condition C2′). To that end we show that its expected gradient does not vanish except at the global minimum. Denote the parameter-error vector $\boldsymbol{\delta} = [\lambda_1 - a_1, \cdots, \lambda_N - a_N]^T$, then

$$E[y_k \boldsymbol{r}_{k-1}] = E\left[\left(Gx_k + \boldsymbol{r}_{k-1}^T \boldsymbol{\delta}\right)\boldsymbol{r}_{k-1}\right] = E\left[\boldsymbol{r}_{k-1}\boldsymbol{r}_{k-1}^T\right]\boldsymbol{\delta}, \tag{8}$$

where the first equation follows from (1) and (5) and the second equation follows from the causality of the channel and the fact that the input is an independent sequence. Finally, it is sufficient to check that the $N \times N$ correlation matrix $\boldsymbol{R} = E\left[\boldsymbol{r}_{k-1}\boldsymbol{r}_{k-1}^T\right]$ is positive-definite: a straightforward computation shows that the (i, j) entry of $\boldsymbol{R}$ is $\sum_{l=0}^{\infty} h_l h_{l+|i-j|}$, where $\{h_0, h_1 \cdots\}$ is the channel impulse response; therefore, for any N-vector $\boldsymbol{x}$, the quadratic form $\boldsymbol{x}^T \boldsymbol{R} \boldsymbol{x}$ is equal to the energy of the output of the channel when driven by a finite duration sequence $\cdots 0, x_1, \cdots x_N, 0 \cdots$. If the finite length input is not identically zero, i.e., $\boldsymbol{x} \neq (0, \cdots, 0)$ then the output sequence cannot be identically zero because the channel is causal.

The conclusion is that the equalizer in (5) updated according to (7) will converge to the inverse of the autoregressive channel regardless of the initial setting of coefficients (with suitable scheduling of the step size μ). This solves the open problem suggested in [5]. The close connection of the form of the update equation (7) to the more general recursive prediction error method [11] will be examined in Section III-B.

III. Blind Equalization of ARMA Channels

We consider in this section the general case of a finite-dimensional linear time-invariant channel where

$$r_k = \sum_{i=1}^{N} a_i r_{k-i} + G\left[x_k + \sum_{i=1}^{L} b_i x_{k-i}\right]. \tag{9}$$

If $L \geq 1$, the channel cannot be equalized exactly by a finite-dimensional transversal filter. We will explore two approaches for the blind equalization of channels with zeros: 1) infinite-dimensional nonrecursive equalization (as in previous works) and 2) finite-dimensional recursive equalization.

A. Nonrecursive Blind Equalization of ARMA Channels

Let us assess the feasibility of extending the approach taken in Section II to a nonrecursive equalizer with an infinite number of taps. For that purpose, the equalizer in (5) is substituted by

$$y_k = r_k - \sum_{i=1}^{\infty} \lambda_i r_{k-i} \tag{10}$$

along with the updating rule in (7) (the vectors therein are now semiinfinite). The expected output energy is equal to the energy of the impulse response of the cascade of channel and equalizer which can be written as

$$\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \lambda_i \lambda_j R(j-i), \tag{11}$$

where $R(l)$ is the autocorrelation of the channel impulse response and, by convention, $\lambda_0 = -1$. Notice that the cost function is a *convex* function of the equalizer coefficients $(\lambda_1, \lambda_2, \cdots)$ and therefore, any local minimum has to be a global minimum. Moreover, since the leading coefficient in the equalizer impulse response is fixed to 1, the leading coefficient in the combined response is equal to that of the channel for every $(\lambda_1, \lambda_2, \cdots)$. The minimum-energy combined response is achieved, if and only if all the other coefficients are equal to zero, i.e., only if the equalizer is equal to the channel inverse (modulo gain). If the channel is minimum-phase there is one and only one causal stable realization of the channel inverse. Otherwise, the channel does not have a stable and causal inverse and the minimum-energy combined response is not achieved by the channel inverse even if a more general doubly-infinite equalizer with an anchored central tap is used. The reason why we discuss this approach is its behavior with truncated approximations (which in contrast to previous approaches) lends itself to the proof of desirable convergence with realizable equalizers. This approach suggests a modification of the cost function pursued in [18] which is able to deal successfully with non minimum phase channels. The finite-length truncation of doubly-infinite nonrecursive blind equalizers has been the caveat emptor of previous solutions (cf. [10]). In our setting it is easy to prove global convergence to an equalizer which approximates the channel inverse up to any prespecified degree of accuracy in those cases where the infinite-dimensional equalizer algorithm works (i.e., minimum-phase channels and maximum-phase FIR channels). If the nonrecursive equalizer has $K < \infty$ taps, then the cost function still admits the expression in (11) except that the summations range up to K and, thus, the function is a quadratic form of a finite-dimensional positive-definite matrix. When the leading equalizer coefficient is fixed to 1, then the cost function is strictly convex in the remaining equalizer coefficients and thus it has a unique global minimum. That minimum value will approach the minimum achieved by the infinite-dimensional valid stable equalizer as $K \to \infty$. The reason is that since the valid equalizer is stable, its distance from a K-dimensional truncation can be made as small as desired provided K is large enough. Then, the continuity of the cost function in the neighborhood of the infinite-dimensional global minimum guarantees that the cost function achieved by the K-dimensional truncation can be made as close as desired to the global minimum. It is of interest to estimate the difference in the depths of the global minima as a function of K. Clearly, this depends on the unknown channel response. However, in applications, a judiciously conservative choice for K can be made based on prior information on the class of potential channel responses. The difference between the depths of the ideal global minimum and the one resulting from the truncated equalizer is equal to the difference between (11) and the corresponding expression where the summation indices range up to K. Given a channel impulse response, it is straightforward to evaluate that quantity (albeit, closed-form expressions do not exist) by computing the energy (in the time or frequency domains) of the cascade of channel response and its truncated inverse. It can be seen that the error can be made arbitrarily small with sufficiently large K, because the ideal equalizer impulse response belongs to l_1 and the channel autocorrelation sequence belongs to l^{∞}.

B. Recursive (IIR) Blind Equalization of ARMA Channels

We consider here a different approach: the blind equalization of (not necessarily minimum-phase) ARMA channels with a (finitely parametrized) ARMA adaptive filter. The first point to observe is that there exists a causal stable ARMA equalizer, if and only if the channel transfer function $H(z)$ is minimum-phase. If the channel is not minimum-phase then the natural solution within the domain of

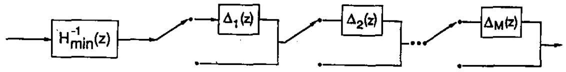

Fig. 2. Switched all-pass sections for identification of minimum-phase zeros.

exact equalization is to 1) equalize the minimum-phase equivalent transfer function $H_{\min}(z)$, and 2) identify which zeros are outside the unit circle. Those zeros can then be taken care of by established techniques in digital communication such as maximum-likelihood sequence detection, decision-feedback equalization or approximate equalization with a finitely parametrized linear system [2]. Denote

$$H(z) = G\,\frac{B_1(z)\cdots B_M(z)}{A(z)},$$

where the numerator has been factored into first- and second-order polynomials with real coefficients. The number of factors $\lceil \frac{L}{2} \rceil \leq M \leq L$ is equal to the number of complex conjugate pairs of zeros of $H(z)$ plus the number of its real-valued zeros. We label $I \subset \{1, \cdots, M\}$ as the set of sections whose zeros belong to the unit circle $\{|z| \leq 1\}$. Define the second-order polynomial (either of whose two coefficients may be zero) $C_i(z) = B_i(z)$ if $i \in I$ and $C_i(z) = B_i(z^{-1})$ otherwise, i.e., $C_i(z)$ is the minimum-phase equivalent of $B_i(z)$. Then,

$$H_{\min}(z) = G\,\frac{C_1(z)\cdots C_M(z)}{A(z)}. \quad (12)$$

Suppose that we were able to somehow obtain $H_{\min}(z)$ in the factored form in (12) but without knowledge of the set I, i.e., without knowing which are the minimum-phase zeros of $H(z)$. Could that missing information be obtained from the factored form of $H(z)$? Define the following all-pass section with an anticausal stable realization

$$\Delta_j(z) = \frac{C_j(z)}{C_j(z^{-1})},$$

which can be approximated (modulo delay) up to any desired degree of accuracy by an FIR filter whose taps are a function of two parameters only: the coefficients of $C_i(z)$. Consider the cascade in Fig. 2, where the inverse of $H_{\min}(z)$ is followed by a series of M sections consisting of (an FIR approximation to) $\Delta_i(z)$ bypassed by a switch. The output is binary for one and only one combination of the M switches.[1] (The output can be binary only if intersymbol interference is eliminated.) A search procedure can be implemented to identify the sought-after combination of switches, which can easily incorporate any a priori knowledge on the maximum-part of the channel, such as an upper bound on its order. Thus, we see that the central problem is the identification of $H_{\min}(z)$ in the factored form in (12) (cf. [17] for a related approach). Once this is accomplished, $H(z)$ can be equalized, using the aforementioned two-step approach. Let us now focus attention of the derivation of an adaptive law which can be shown to converge to the inverse of (12).

We first consider the following adaptive recursive equalizer:

$$y_k = -\sum_{i=1}^{L} \theta_i y_{k-i} + r_k - \sum_{i=1}^{N} \lambda_i r_{k-1}. \quad (13)$$

Note that if $\theta^T = [\theta_1 \cdots \theta_L] = [b_1 \cdots b_L]$ and $\lambda^T = [\lambda_1 \cdots \lambda_N] = [a_1 \cdots a_N]$; then $y_k = Gx_k$ for all k. Furthermore, since the leading coefficient in the equalizer impulse response is equal to 1, that choice of equalizer coefficients achieves the unique global minimum of the output energy. As before, the updating of θ and λ will proceed according to a stochastic approximation of the output energy. In order to obtain the gradient of the output energy with respect to λ and θ, it is convenient to write (13) as the cascade of its regressive and moving average components:

$$y_k = w_k - \sum_{i=1}^{N} \lambda_i w_{k-i}, \quad (14a)$$

$$w_k = -\sum_{i=1}^{L} \theta_i w_{k-i} + r_k. \quad (14b)$$

It readily follows from (14a) that the gradient with respect to the numerator coefficients is

$$\frac{1}{2}\nabla_\lambda y_k^2 = -[w_{k-1} \cdots w_{k-N}]^T y_k. \quad (15)$$

In order to find the gradient with respect to the denominator coefficients let us introduce

$$d_k = -\sum_{i=1}^{L} \theta_i d_{k-i} + y_k;$$

from (14), we get

$$\frac{\partial y_k}{\partial \theta_j} = \frac{\partial w_k}{\partial \theta_j} - \sum_{i=1}^{N} \lambda_i \frac{\partial w_{k-i}}{\partial \theta_j}, \quad (16a)$$

$$\frac{\partial w_k}{\partial \theta_j} = -\sum_{i=1}^{L} \theta_i \frac{\partial w_{k-i}}{\partial \theta_j} - w_{k-j}, \quad (16b)$$

which implies that $\frac{\partial w_k}{\partial \theta_j}$ is the response of the denominator of the equalizer to $-w_{k-j}$ or, equivalently, $\frac{\partial y_k}{\partial \theta_j}$ is equal to $-d_{k-j}$. Thus, the equalizer coefficients are updated according to

$$\lambda_j^{(k+1)} = \lambda_j^{(k)} + \mu y_k w_{k-j}, \qquad j = 1, \cdots, N, \quad (17a)$$

$$\theta_j^{(k+1)} = \theta_j^{(k)} + \mu y_k d_{k-j}, \qquad j = 1, \cdots, L. \quad (17b)$$

Formulae (17a) and (17b) may be interpreted as a special case of the recursive prediction error algorithm (RPEA) [11, pp. 385–387]. Even for more general system models than those that we are considering, the RPEA for ARMA-type parametrizations leads to an expression for the gradient of the form of a filtered regressor vector where the filtering depends on the parameters being estimated [11, pp. 110–114]. Our pending further development, however, will lead us to consider a different parametrization and algorithm motivated by the structure of (12). The simplicity of the cost function we are using allows us to verify that this equalizer will converge to the channel parameters regardless of its initialization (as long as it is stable) by invoking a result from maximum-likelihood identification of Gaussian linear systems due to Astrom and Soderstrom [1], and Stoica and Soderstrom [16]:

Theorem: Consider the cascade of two systems,

$$r_k = \sum_{l=1}^{N} a_i r_{k-i} + e_k + \sum_{i=1}^{L} b_i e_{k-i}, \quad (18)$$

$$y_k = -\sum_{i=1}^{L'} \theta_i y_{k-i} + r_k - \sum_{i=1}^{N'} \lambda_i r_{k-i}. \quad (19)$$

[1] In the presence of moderate noise, the correct switch combination will still be discernible from the output.

Assume that $\{e_k\}$ is a second-order i.i.d. sequence, and $L' \geq L, N' \geq N$. If (18) is a minimum-phase stable system without pole-zero cancellations, then the only stationary points of $E[y_k^2]$ as a function of θ and λ occur at the global minima:

$$1+\theta_1 z^{-1}+\cdots+\theta_{L'} z^{-L'} = \left(1+b_1 z^{-1}+\cdots+b_L z^{-L}\right) \cdot \left(1+l_1 z^{-1}+\cdots+l_{\boldsymbol{m}} z^{-\boldsymbol{m}}\right),$$

$$1+\lambda_1 z^{-1}+\cdots+\lambda_{N'} z^{-N'} = \left(1+a_1 z^{-1}+\cdots+a_N z^{-N}\right) \cdot \left(1+l_1 z^{-1}+\cdots+l_{\boldsymbol{m}} z^{-\boldsymbol{m}}\right),$$

where $m = \min(L'-L, N'-N)$ and $\left(1+l_1 z^{-1}+\cdots+l_{\boldsymbol{m}} z^{-\boldsymbol{m}}\right)$ has all zeros inside the unit circle, but is otherwise arbitrary.

In the case we are considering, (18) and (19) become the minimum-phase equivalent of (9) and (13), respectively. The result of the theorem ensures that the lines of steepest descent of the recursion in (17) converge to the minimum-phase equivalent of the channel (provided that the initialization is stable and the step size is suitably scheduled), because the cost function is transparent to replacing the channel by its minimum-phase part, and the gradients do not depend on the channel impulse response. However, this falls short of our objective since the equalizer in (13) is not in the factored form required in (12). Besides, in any realization of the stochastic approximation in (17), the trajectory followed by $\lambda^{(k)}$ and $\theta^{(k)}$ deviates randomly from its expected trajectory (a line of steepest descent) and in particular may wander outside the set of stable systems, in which case convergence is no longer guaranteed. This is a well-known issue in adaptive IIR filtering [9], and can be remedied by discarding updates that would lead to unstable systems or reflecting poles lying outside the unit circle to their reciprocals. To resolve both shortcomings we will parametrize the recursive part of the filter as a cascade of second-order sections, the stability of which is easy to monitor:

$$\frac{1}{1+\theta_1 z^{-1}+\cdots+\theta_L z^{-L}} = \prod_{m=1}^{M} \frac{1}{\Theta^m(z^{-1})}, \tag{20}$$

where $\Theta^m(z^{-1}) = 1+\theta_1^m z^{-1}+\theta_2^m z^{-2}$. We will denote the output of the mth stage by w_k^m (cf. Fig. 3), in particular, $w_k = w_k^M$. The gradient with respect to the numerator coefficients remains intact (cf. (15)). In order to find $\nabla_m y_k$ (the gradient of the output with respect to the coefficients of the mth section) we will denote $\Lambda(z^{-1}) = 1+\lambda_1 z^{-1}+\cdots+\lambda_L z^{-L}$, and we will use the informal notation

$$w_k^{n-1} = \Theta^n(z^{-1}) w_k^n, \tag{21}$$

$$y_k = \frac{\Lambda(z^{-1})}{\Theta^{n+1}(z^{-1})\cdots\Theta^M(z^{-1})} w_k^n. \tag{22}$$

From (21), it follows that

$$0 = \nabla_n w_k^{n-1} = \begin{bmatrix} z^{-1} \\ z^{-2} \end{bmatrix} w_k^n + \Theta^n(z^{-1}) \nabla_n w_k^n$$

or, equivalently,

$$\nabla_n w_k^n = \frac{1}{\Theta^n(z^{-1})} \begin{bmatrix} -w_{k-1}^n \\ -w_{k-2}^n \end{bmatrix}. \tag{23}$$

Putting (22) and (23) together, we obtain

$$\nabla_n y_k = \frac{\Lambda(z^{-1})}{\Theta^{n+1}(z^{-1})\cdots\Theta^M(z^{-1})} \nabla_n w_k^n$$

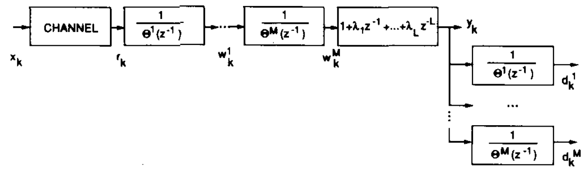

Fig. 3. Recursive blind equalizer with cascaded AR second-order sections.

$$= \frac{\Lambda(z^{-1})}{\Theta^n(z^{-1})\cdots\Theta^M(z^{-1})} \begin{bmatrix} -w_{k-1}^n \\ -w_{k-2}^n \end{bmatrix}$$

$$= -\frac{1}{\Theta^n(z^{-1})} \begin{bmatrix} y_{k-1} \\ y_{k-2} \end{bmatrix} = -\begin{bmatrix} d_{k-1}^n \\ d_{k-2}^n \end{bmatrix}, \tag{24}$$

where d_k^n is the response of the nth recursive stage to the equalizer output (cf. Fig. 3).

With the new parametrization, there is no longer a unique global minimum, as all permutations of second-order sections are equivalent. However, this does not give rise to spurious local minima, i.e., if a given unfactored system is not a local minimum, none of its factored counterparts can be local minima. To see this, fix any unfactored system which is not a local minimum. Then, there is a neighboring system achieving lower cost. The factored versions of both of those systems (nonunique in general) attain the same respective costs as their unfactored counterparts, and are also neighbors due to the continuity of the roots of a polynomial as a function of its coefficients [13, p. 3]. This implies that the corresponding factored system cannot be a local minimum—a conclusion that is in accordance with the general results on mean-square surfaces for cascaded IIR filters in [14].

Regarding the implementation of the algorithm, it should be noted that the assumption that the channel order parameters M and N are known and are exactly matched by the equalizer can be dropped, as long as the equalizer overparametrizes both M and N (see the conditions in the theorem allowing $L' \geq L, N', \geq N$). In the context of conventional nonrecursive blind equalization, related conditions (on the length of the equalizer relative to the channel order) are also necessary.

IV. SUMMARY

Globally convergent blind equalization has been shown in [6, 15] for *doubly-infinite* equalizers of the Godard type. The truncation required to implement those equalizers destroys their global convergence as demonstrated in [4]. Blind equalization without gain identification leads naturally to the study of *anchored* equalizers (with fixed taps) which eliminate some of the excess degrees of freedom that underpin the ill-convergence mechanism reported in [4]. A sensible strategy to the blind equalization of nonminimum-phase channels is to decouple the equalization of the minimum-phase part (which can be done exactly) and the identification of the zeros located outside the unit circle (cf. [17]). We showed a simple system that accomplishes this task and exhibits global convergence, as long as the channel does not have all-pass factors.

In contrast to Godard-type cost functions, the convexity of energy functions used in this correspondence allow that the convergence properties of infinite-dimensional equalizers be inherited by their truncations, and the ideal channel inverse can be approximated up to any prespecified degree of accuracy provided the truncation is long enough. The next step motivated by this correspondence is the study of anchored equalizers using cost functions other than energy. An anchored equalizer whose cost function is not energy (but it is convex) has been identified in [18] where it is shown to achieve essentially the same convergence properties as the doubly-infinite Godard equalizer, with the advantage that those properties are not destroyed by finite truncations.

Comparing the anchored blind equalizers with energy cost functions and the Godard-type blind equalizers, we conclude the following.

a) For doubly-infinite nonrecursive equalization Godard-type cost functions are superior as they achieve global convergence, whereas the anchored equalizers may have inadmissible global minima.

b) For semi-infinite nonrecursive equalization, both equalizers achieve convergence if the channel is minimum-phase, otherwise, both equalizers have spurious local minima. The reason why the Godard equalizer suffers from this problem is the noninvertibility of the semi-infinite channel convolution matrix [5] in the nonminimum-phase case.

c) For finitely parametrized equalizers, the truncation of the Godard-type equalizers leads to ill-convergence even for the simplest channels [4], whereas the anchored equalizers based on convex cost functions inherit the convergence properties of their infinite-dimensional counterparts which can be approximated as accurately as desired. Furthermore, in contrast to the Godard-type equalizers, exact finite-dimensional implementable blind equalization is achievable by anchored blind equalizers.

References

[1] K. J. Astrom and T. Soderstrom, "Uniqueness of the maximum likelihood estimates of the parameters of an ARMA model," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 769–773, Dec. 1974.

[2] S. Benedetto, E. Biglieri, and V. Castellani, *Digital Transmission Theory.* Englewood Cliffs, NJ: Prentice-Hall, 1987.

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

[4] Z. Ding, R. A. Kennedy, B. D. O. Anderson, and C. R. Johnson, "Existence and avoidance of ill-convergence of Godard blind equalizers in data communication systems," *IEEE Trans. Commun.*, vol. 39, pp. 1313–1327, Sept. 1991.

[5] Z. Ding, C. R. Johnson, and R. A. Kennedy, "On the (non)existence of undesirable equilibria of Godard blind equalizers," *IEEE Trans. Signal Processing*, 1992, to appear.

[6] G. J. Foschini, "Equalizing without altering or detecting data," *AT&T Tech. J.*, vol. 64, pp. 1885–1911, Oct. 1985.

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

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

[9] C. R. Johnson, "Adaptive IIR filtering: Current results and open issues," *IEEE Trans. Inform. Theory,* vol. IT-30, pp. 237–250, Mar. 1984.

[10] C. R. Johnson, "Admissibility in blind adaptive equalization," *IEEE Control Syst. Mag.*, vol. 11, pp. 3–15, Jan. 1991.

[11] L. Ljung and T. Soderstrom, *Theory and Practice of Recursive Identification.* Cambridge, MA: MIT Press, 1987.

[12] R. W. Lucky, "Automatic Equalization for Digital Communications," *BSTJ*, vol. 44, pp. 547–588, Apr. 1965.

[13] M. Marden, *Geometry of Polynomials.* Providence, RI: American Mathematical Society, 1966.

[14] N. Nayeri and W. K. Jenkins, "Alternate realizations of adaptive IIR filters and properties of their performance surfaces," *IEEE Trans. Circuits Syst.*, vol. CAS-36, pp. 485–496, Apr. 1989.

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

[16] P. Stoica and T. Soderstrom, "Uniqueness of the maximum likelihood estimates of ARMA model parameters—An elementary proof," *IEEE Trans. Automat. Contr.*, vol. AC-27, pp. 736–738, June 1982.

[17] J. K. Tugnait, "Identification of nonminimum phase linear stochastic systems," *Automatica*, vol. 22, pp. 457–464, July 1986.

[18] S. Vembu, S. Verdú, R. A. Kennedy, and W. Sethares, "Convex cost functions in blind equalization," *Proc. 24th Conf. Inform. Sci. Syst.*, Baltimore, MD, Mar. 1991, pp. 792–797.

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

Convergence of Best ϕ-Entropy Estimates

Marc Teboulle and Igor Vajda, *Senior Member, IEEE*

***Abstract*—Minimization problems involving ϕ-entropy functionals (a generalization of Boltzmann–Shannon entropy) are studied over a given set A and a sequence of sets A_n and the properties of their optimal solutions x_ϕ, x_n. Under certain conditions on the objective functional and the sets A and A_n, it is proven that as n increases to infinity, the optimal solution x_n converges in L_1 norm to the best ϕ-entropy estimate x_ϕ.**

***Index Terms*—Entropy functionals, norm convergence, maximum entropy methods, convex optimization, set-convergence.**

I. Introduction

Let us consider a convex, continuous function $\phi : \mathbb{R}_+ \to \mathbb{R}$ and put

$$\phi(u) = \begin{cases} \lim_{v\downarrow 0} \phi(v), & \text{if } u = 0 \\ +\infty, & \text{if } u < 0. \end{cases}$$

Denote by $\phi'_+(u)$ the (finite) right-hand derivative of ϕ at $u \in \mathbb{R}_+$. It is well known (see, Rockafellar [25]), that for any $v \in \mathbb{R}_+$ it holds

$$\phi(u) \geq \phi(v) + \phi'_+(v)(u - v), \qquad u \in \mathbb{R}. \tag{1}$$

It follows from here

$$\phi(u) \in (-\infty, \infty], \qquad u \in \mathbb{R}.$$

Consider further a finite measure space (S, μ) and the corresponding Banach spaces $L_\alpha(S, \mu)$, $1 \leq \alpha \leq \infty$, with norms $\|\cdot\|_\alpha$. It follows from (1) that the formula

$$I_\phi(x) := \int_S \phi(x(s))d\mu(s) \tag{2}$$

defines a mapping $I_\phi : L_1(S, \mu) \to (-\infty, \infty]$. This mapping is called the ϕ-entropy.[1]

Manuscript received November 5, 1991; revised May 3, 1992. M. Teboulle was supported in part by AFOSR Grant 91008 and NSF Grant DMS-9201297. I. Vajda was supported by CSAS Grant 17503.

M. Teboulle is with the Department of Mathematics and Statistics, University of Maryland, Baltimore County Campus, Baltimore, MD 21228.

I. Vajda is with the Institute for Information Theory and Automation, 18208 Prague, Czechoslovakia.

IEEE Log Number 9203638.

[1] Throughout this correspondence, we use the extended real line arithmetic rules common to integration theory. These rules include e.g., $0 \cdot \infty = 0$, see e.g., Rudin [27, p. 19].