V. CONCLUSIONS

In this correspondence, relationships of the transfer function polynomials of the direct-form and two lattice realizations were introduced. It was shown that these equations yield computationally efficient implementation of lattice-based algorithms for adaptive IIR filters, including members of the equation error family of algorithms, requiring $O(N)$ multiplications per iteration, where N is the filter order. The proposed lattice-based implementation led to a set of parameters that realize identical transfer functions to the ones obtained by corresponding direct-form algorithms.

ACKNOWLEDGMENT

The authors would like to thank the reviewers for all their helpful comments.

REFERENCES

[1] I. L. Ayala, "On a new adaptive lattice algorithm for recursive filters," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-30, pp. 316–319, Apr. 1982.
[2] A. H. Gray, Jr. and J. D. Markel, "Digital lattice and ladder filter synthesis," *IEEE Trans. Audio Electroacoust.,* vol. 21, pp. 492–500, Dec. 1973.
[3] ——, "A normalized digital filter structure," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-23, pp. 268–277, June 1975.
[4] S. Horvath, Jr., "A new adaptive recursive LMS filter," *Digital Signal Processing,* V. Cappellini and A. G. Constantinides, Eds. New York: Academic, 1980, pp. 21–26.
[5] ——, "Lattice form adaptive recursive digital filters: Algorithms and applications," in *Proc. Int. Symp. Circuits Syst.,* Houston, TX, Apr. 1980, pp. 128–133.
[6] F. Itakura and S. Saito, "Digital filtering techniques for speech analysis and synthesis," in *Proc. 7th Int. Cong. Acoustics,* Budapest, Hungary, 1971, pp. 261–264.
[7] J.-N. Lin and R. Unbehauen, "Bias-remedy least mean square equation error algorithm for IIR parameter recursive estimation," *IEEE Trans. Signal Processing,* vol. 40, pp. 62–69, Jan. 1992.
[8] K. X. Miao, H. Fan, and M. Doroslovački, "Cascade lattice IIR adaptive filters," *IEEE Trans. Signal Processing,* vol. 42, pp. 721–742, Apr. 1994.
[9] S. L. Netto and P. S. R. Diniz, "Composite algorithms for adaptive IIR filtering," *IEE Electron. Lett.,* vol. 28, no. 9, pp. 886–888, Apr. 1992.
[10] S. L. Netto and P. Agathoklis, "A new composite adaptive IIR algorithm," in *Proc. 28th Asilomar Conf. Signals, Systems, Computers,* Pacific Grove, CA, Oct./Nov. 1994, pp. 1506–1510.
[11] D. Parikh, N. Ahmed, and S. D. Stearns, "An adaptive lattice algorithm for recursive filters," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-28, pp. 110–112, Feb. 1980.
[12] P. A. Regalia, "Stable and efficient lattice algorithms for adaptive IIR filtering," *IEEE Trans. Signal Processing,* vol. 40, pp. 375–388, Feb. 1992.
[13] J. A. Rodríguez-Fonollosa and E. Masgrau, "Simplified gradient calculation in adaptive IIR lattice filters," *IEEE Trans. Signal Processing,* vol. 39, pp. 1702–1705, July 1991.
[14] J. J. Shynk, "On lattice-form algorithms for adaptive IIR filtering," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing,* New York, Apr. 1988, pp. 1554–1557.

Adaptive Fractionally Spaced Blind CMA Equalization: Excess MSE

I. Fijalkow, C. E. Manlove, and C. R. Johnson, Jr.

***Abstract*— The performance of the constant modulus algorithm (CMA)—a reference algorithm for adaptive blind equalization—is studied in terms of the excess mean square error (EMSE) due to the nonvanishing step size of the gradient descent algorithm. An analytical approximation of EMSE is provided, emphasizing the effect of the constellation size and resulting in design guidelines.**

***Index Terms*—Adaptive blind equalization, excess mean square error, multichannel equalization.**

I. INTRODUCTION

Digital communication is subject to intersymbol interference (ISI) due to nonideal pulse shaping, multipath propagation, and residual clock or carrier phase error. ISI is more severe when the channel dispersion time (or channel time span) cannot be neglected with respect to the input signal symbol time duration, thereby making its removal all the more crucial. Traditionally, channel equalization (i.e., input sequence extraction directly from the received signal) and identification (with input sequence recovery, e.g., by Wiener filtering of the observed signal using the channel estimate) are performed using a training sequence. However, in many applications, either the training sequence is unavailable, or the bandwidth occupied by the training sequence is to be spared for input carrying information. Consequently, one pursues *blind equalization*, i.e., without training nor any *a priori* knowledge of the channel dynamics. Due to its potential benefits, blind equalization has become an important topic in digital communications. Blind methods use the received signal sequence and some *a priori* knowledge of the input sequence statistics. Nonminimum phase channel equalization was performed using methods based on high-order statistics or other nonlinearities that are effective only with non-Gaussianly distributed input sequences [4]. Contrast-based methods (see [13], for example) or adaptive Bussgang algorithms (see [1]) have been proposed and studied over the last ten years. In this correspondence, we study the most popular adaptive blind equalization Bussgang algorithm (the Godard algorithm [7]) or constant modulus algorithm (CMA) [15] in the context of nonconstant modulus data with spatio-temporal diversity.

The constant modulus (CM) criterion minima have been proven to achieve zero-forcing fractionally spaced equalization (see [5] and [11] for a simple algebraic proof) under various ideal conditions, including the absence of noise. Even under the presence of a small amount of additive channel noise, CM minima were proved to

Manuscript received January 27, 1995; revised June 17, 1997. This work was supported by the French Ministére des Affaires Etrangéres (bourse Lavoisier), Applied Signal Technology, Inc., and NSF Grant MIP-9509011. The associate editor coordinating the review of this paper and approving it for publication was Dr. Pierre Duhammel.

I. Fijalkow was with the School of Electrical Engineering, Cornell University, Ithaca, NY 14853 USA. She is now with ENSEA de Cergy, Groupe ETIS, Cergy-Pontoise, France (e-mail: fijalkow@ensea.fr).
C. E. Manlove was with School of Electrical Engineering, Cornell University, Ithaca, NY 14853 USA. She is now with the Applied Physics Laboratory, Johns Hopkins University, Baltimore, MD 21218 USA.
C. R. Johnson is with the School of Elecricial Engineering, Cornell University, Ithaca, NY 14853 USA (e-mail: johnson@anise.ee.cornell.edu).

Publisher Item Identifier S 1053-587X(98)00505-4.

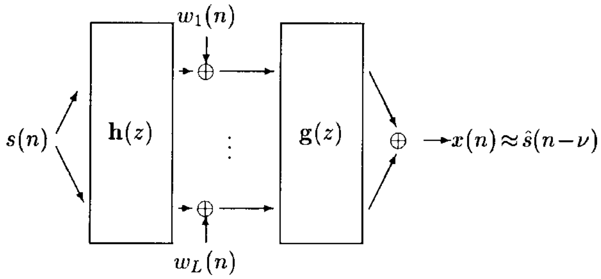


Fig. 1. Noisy fractionally-space equalization scheme.

achieve similar performance as the trained Wiener optimal solution [6], [18]. These results concern the solution reached on average by an equalizer updated using CMA at steady state. However, when a stochastic gradient descent algorithm such as CMA is adapted, the solution at steady state jitters around the mean solution. The amplitude of the jittering is critical to the algorithm performances since the jittering area of the equalized points must be smaller than the distance between the points of the input constellation in order to "open the eye." This jitter is usually quantified by the excess mean square error (EMSE) between the delayed input and output of the system.

Following the approach of Benvéniste *et al.* in [2], in order to study the performance of a stochastic gradient descent algorithm at steady state, in this correspondence, we express analytically the EMSE for CMA, and we note the relative importance of channel, equalizer length, and input constellation on the EMSE. Analytical results are then checked by numerical simulations. We also introduce the notion of (re)selecting step size across changes in constellation size to maintain the percentage of distance between adjacent constellation points closed by the EMSE.

II. Spatio-Temporal Equalization

Using spatio-temporal diversity (either by oversampling the received data with respect to the input data rate or by observing the received data through a sensor array), the communication system can be viewed as a single input/multiple output (SIMO) transfer function $\mathbf{h}(z)$ with a zero-mean input data stream $s(n)$ (see [10]). We assume herein that the L-variate channel $\mathbf{h}(z)$ encompasses the effects of the transmission filter, channel response, reception filter, and any other linear dispersive effect encountered between the transmitter's symbol generation and the receiver's equalizer. The system is corrupted by additive noise $\mathbf{w}(n) = [w_1(n), \cdots, w_L(n)]^\top$ (as displayed in Fig. 1).

The observed sequence $\mathbf{y}(n)$ is described as

$$\mathbf{y}(n) = [\mathbf{h}(z)]\, s(n) + \mathbf{w}(n) = \sum_{k=0}^{Q} \mathbf{h}_k s(n-k) + \mathbf{w}(n) \quad (1)$$

where $\mathbf{h}(z)$ is assumed to be a causal polynomial with degree Q. $\mathbf{y}(n)$ and $\mathbf{w}(n)$ are L-variate signals, where L denotes either the temporal oversampling factor or the number of sensors.

The equalizer $\mathbf{g}$ has an NL-long impulse response chosen so that its output

$$x(n) = \mathbf{g}^\top \mathbf{Y}_N(n) = \underbrace{\mathbf{g}^\top \mathcal{T}(\mathbf{h})}_{f^\top} S_{N+Q}(n) + \mathbf{g}^\top \mathbf{W}_N(n)$$

is a good estimate of the delayed input $s(n-\nu)$ (up to some multiplicative scalar constant). $\mathcal{T}(\mathbf{h})$ is the $NL \times (N+Q)$ block-Toeplitz channel convolution matrix. The regressor vector $\mathbf{Y}_N(n)$ is the concatenation of the observed signal $\mathbf{y}(n)$ at instants $n, n-1, \cdots, n-N+1$. f is the global system (i.e., channel and equalizer convolved combination) impulse response. The hypotheses on the system are as follows.

TABLE I
M-PAM: Distance d Between Points of Constellation Versus M

M	2	4	8	16	32
d	2	0.8944	0.4364	0.2169	0.1083

- H1) Each component of $\mathbf{h}(z)$ is assumed to be causal with finite degree less or equal to Q, $\mathbf{h}(z) \neq 0,\ \forall z$.
- H2) The equalizer time span N is larger than the channel time span Q.
- H3) The noise $\mathbf{w}(n)$ is a zero-mean, circular (i.e., each $E[w_k^2] = 0$ for complex valued signals), Gaussian, temporally, and spatially white ($E[\mathbf{w}(n)\mathbf{w}^{*\top}(m)] = E[|w|^2]I$ if $m = n$, and 0 otherwise. I stands for the $L \times L$ identity matrix.)
- H4) $s(n)$ is an i.i.d. (i.e., independent and identically distributed) sequence, zero-mean, circular with variance $E[|s|^2] = 1$ and sub-Gaussian (i.e., $E[|s|^4] - 2E[|s|^2]^2 - |E[s^2]|^2 < 0$).
- H5) $\mathbf{w}(n)$ is independent from the input $s(n)$.

Under H1) and H2), $\mathcal{T}(\mathbf{h})$ is proved in [3] to be a full-column rank block-Toeplitz matrix. H2) is greater than the minimal value required to have more rows than columns in the matrix $\mathcal{T}(\mathbf{h})$ when $L > 3$. However, it ensures an equal integer number of time delays considered on each of L subchannels. The sub-Gaussian assumption in H4) is not too restrictive since it is satisfied for virtually all digital communications modulations (including PAM, PSK, and QAM). Note that the whiteness and Gaussian assumptions in H3) are not crucial but simplify the expressions in the sequel.

III. CM Criterion—CMA

The CM cost function, as a function of a generic equalizer $\mathbf{g}$, is defined as $J(\mathbf{g}) = E[(r_2 - |x(n)|^2)^2]$, where $x(n) = \mathbf{g}^\top \mathbf{Y}_N(n)$, and $r_2 = E[|s|^4]/E[|s|^2]$. It was introduced in [7] and [15]. Correction proportional to the instantaneous gradient over $\mathbf{g}$ yields the CMA update of the equalizer impulse response as

$$\mathbf{g}(n+1) = \mathbf{g}(n) + \mu\, x(n)[r_2 - |x(n)|^2]\mathbf{Y}_N(n)^* \quad (2)$$

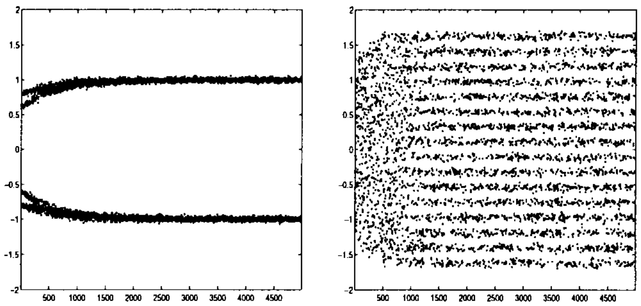
with $x(n) = \mathbf{g}(n)^\top \mathbf{Y}_N(n)$. Here, * denotes conjugation.

In [17], CMA was presented in a fractionally spaced fashion, i.e., the received signal is oversampled with respect to the input rate. In most digital communications systems, such an oversampling is used (see [16]) so that considering the channel SIMO model is the realistic assumption considered in the following.

A. Simulations

For CMA with a nonconstant modulus input sequence, EMSE at steady state is a crucial factor for the performance of the algorithm. Here, we briefly illustrate this phenomenon on M-pulse amplitude modulation (M-PAM) input sequences with a normalized power $1 = E[s^2] = 2\sum_{k=0}^{M/2-1} (2k+1)^2 = d^2[M(M^2-1)/12]$, where d is the distance between two successive values of the constellation (see Table I).

The channel chosen for simulations is described by its two-component z transform $(1 - 0.4z^{-1})(1 - 1.2z^{-1})$ for the first component of $\mathbf{h}(z)$ and $(1 - 0.7z^{-1})(1 + 1.5z^{-1})$ for its second



(a) (b)

Fig. 2. Equalizer output versus number of iterations.

component. CMA is initialized with a center spike equalizer value of unity and is updated with a step size equal to $\mu = 10^{-4}$. Next, we display the equalizer outputs for a binary (2-PAM) [Fig. 2(a)] and a 16-PAM [Fig. 2(b)] input sequences; in both cases, the signal to noise ratio (SNR) is 40 dB. We assume in the sequel that the error due to noise can be neglected with respect to the error due to a nonzero step size and nonconstant modulus sources.

In order to "open the eye" (i.e., to be able to decide of the input value from the equalizer output), the jitter of the equalizer output around constellation values must be smaller than d. The only factor that the user can then tune to reduce the misadjustment is the algorithm step size μ. From Fig. 2, we observe that the same step size should not be used for different constellation sizes. Unfortunately, a reduction in step size reduces the algorithm convergence rate. Thus, a tradeoff must be made by the user between the constellation size (and, subsequently, bit transmission rate) and the step size. The following analysis should assist the user in performing this trade-off.

B. Cost-Function Characterization

In order to characterize the jitter around the CM criterion minima, we recall the description of these points given by averaging theory [2]. We consider hereafter the noise-free case, but the study can be extended to the case of additive channel noise. When there is no additive channel noise, $J(\mathbf{g})$ depends only on the combined channel-equalizer impulse response f and on the input signal statistics. Minima characterization was first introduced in [5], [9], and [11]. Results from averaging theory concerning the behavior of such stochastic adaptive gradient algorithms [2] and the work in [8] in the single-input case for a BPSK input are the basis for [5], which studies the CM cost function extrema in the noise-free case.

Lemma 1: In [5], we have CMA stationary points zero on the ordinary differential equation (ODE)

$$\begin{aligned}\frac{\partial \mathbf{g}}{\partial t} &= E\{x(n)[r_2 - |x(n)|^2]\mathbf{Y}_N(n)^*\}\\ &\qquad \text{where } x(n) = \mathbf{g}^\top \mathbf{Y}_N(n)\\ &= E[s^2]^2 \mathcal{T}(\mathbf{h})\,\Delta(f)\,f\\ &\qquad \text{where } \Delta(f) = (3f^\top f - \rho)I - (3-\rho)\mathrm{diag}(ff^\top) \end{aligned} \quad (3)$$

with $\rho = E[s^4]/E[s^2]^2$, and we have $\mathrm{diag}(A)$ defined as the matrix extracted from A with the same diagonal entries and 0 elsewhere. The stationary points can be classified by their stability. The sign definition of the Hessian

$$\begin{aligned}\mathcal{H}(g) &= -E[s^2]^2 \mathcal{T}(\mathbf{h})\,\Psi(f)\,\mathcal{T}(\mathbf{h})^\top\\ &\qquad \text{where } \Psi(f) = (3f^\top f - \rho)I + 6ff^\top\\ &\qquad - 3(3-\rho)\mathrm{diag}(ff^\top).\end{aligned}$$

When $\mathcal{T}(\mathbf{h})$ is full column rank [i.e., under H1) and H2)], the stationary points are defined by $\Delta(f)f = 0$, and their stability is described by $\Psi(f)$. They are also classified in terms of f as

- $f = 0$, maximum;
- $f = \delta_\nu = (0 \cdots 010 \cdots 0)^\top$ with $\nu = 0, \cdots, N+Q-1$ minima;
- f having P nonzero components equal to $\{\rho/[3(P-1)+\rho]\}^{1/2}$ saddle points.

It should be noticed that each setting described in terms of the combined channel-equalizer impulse response f corresponds to a dense subspace of settings in the equalizer tap space $\mathbf{g}$. All settings corresponding to the same combined impulse response are equivalent in terms of equalizer input–output performance in the noise-free context. However, the existence of some channel additive noise regularizes the cost function and forces the equalizer toward minimal norm solutions, [6].

IV. Excess MSE: Stochastic Jitter at Steady State

From the previous section, we know that the equalizer converges asymptotically in mean to a value such that the corresponding noise-free output satisfies $\mathbf{g}^\top \mathbf{Y}_N(n) = f^\top S(n) = s(n-\nu)$. The minimum mean square error (MMSE) $E\{[\mathbf{g}^\top \mathbf{Y}_N(n) - s(n-\nu)]^2\}$ is thus 0. However, due to the random character of the input signal, the instantaneous error $\mathbf{g}(n)^\top \mathbf{Y}_N(n) - s(n-\nu)$ is not zero, even when there is no channel noise. The jitter around the mean solution is quantified by the EMSE, $E\{[(\mathbf{g}(n)^\top \mathbf{Y}_N(n) - s(n-\nu)]^2\}$. Using results of averaging theory, an analytical approximation of the EMSE is provided for small values of the step size μ in a noise-free context.

Proposal 1: Under conditions H1), H2), and H4) and in the absence of noise, the EMSE of CMA can be approximated by

$$\mu N \frac{\dfrac{E[s^6(n)]}{E[s^2(n)]^3} - \rho^2}{2(3-\rho)} \{E[s^2(n)]\}^2 E[\mathbf{y}(n)^\top \mathbf{y}(n)] \quad (4)$$

where ρ is the kurtosis of the input signal $E[s^4(n)]/E[s^2(n)]^2$.

Proof: $x(n) = f(n)^\top S(n)$ with $f(n) = \mathcal{T}(\mathbf{h})^\top \mathbf{g}(n)$

$$f(n+1) = f(n) - \mu\,\mathcal{T}(\mathbf{h})^\top \mathcal{T}(\mathbf{h})[x^2(n) - r_2]x(n)S(n). \quad (5)$$

Let us denote δ_ν as the attractor, in terms of f, around which there is jittering for a small enough μ.

$$\begin{aligned}E\{[(x(n) - s(n-\nu)]^2\} &= E(\{[f(n)-\delta_\nu]^\top S(n)\}^2)\\ &= \mathrm{Trace}(E\{[f(n)-\delta_\nu][f(n)-\delta_\nu]^\top\\ &\qquad \cdot S(n)S(n)^\top\}).\end{aligned}$$

With the usual independence hypothesis between the global channel-equalizer system and the input signal at steady state, the EMSE becomes

$$\begin{aligned}E\{[(x(n) - s(n-\nu)]^2\} &\approx \mathrm{Trace}(E\{[f(n)-\delta_\nu][f(n)-\delta_\nu]^\top\}\\ &\qquad \cdot E[S(n)S(n)^\top])\\ &= E[s(n)^2]\mathrm{Trace}(E\{[f(n)-\delta_\nu]\\ &\qquad \cdot [f(n)-\delta_\nu]^\top\}).\end{aligned}$$

An approximation of $E\{[f(n)-\delta_\nu][f(n)-\delta_\nu]^\top\}$ for an algorithm such as (5) can be drawn from averaging theory [2] using the following result.

Lemma 2: (See [2, Th. 1, pp. 107].) Consider a stochastic approximation algorithm of the type

$$f(n+1) = f(n) + \mu\, H[f(n),\, S(n)] \tag{6}$$

where H is a "smooth" function, and $[f(n)]$ "converges" in mean to the attractor f_* (here, δ_ν). For a "small enough" value of the fixed step size μ, the covariance matrix $E\{[f(n)-f_*][f(n)-f_*]^\top\}$ can be approximated by μP_*, where P_* is the unique positive solution of the Lyapunov equation

$$\mathcal{H}(f_*)P_* + P_*\mathcal{H}(f_*)^\top + \mathcal{R}(f_*) = 0 \tag{7}$$

where $\mathcal{H}(f_*)$ is the Hessian corresponding to the ODE associated with (5), taken for $f = f_*$, and

$$\mathcal{R}(f_*) = \sum_{n=-\infty}^{+\infty} E(H[f_*,\, S(n)]\{H[f_*,\, S(0)]\}^\top).$$

It should be mentioned that the formal proof of this result requires f_* to be a global attractor.

In the case of CMA, the Hessian corresponding to (5) is $\mathcal{H}(f) = -\{E[s(n)^2]\}^2\mathcal{T}(\mathbf{h})^\top\mathcal{T}(\mathbf{h})\Psi(f)$. At the convergence point, $\Psi(\delta_\nu) = (3-\rho)I - 3(1-\rho)\delta_\nu\delta_\nu^\top$. $\mathcal{R}(\delta_\nu) = \{E[s(n)^2]\}^2\mathcal{T}(\mathbf{h})^\top\mathcal{T}(\mathbf{h})\mathcal{R}_0\mathcal{T}(\mathbf{h})^\top\mathcal{T}(\mathbf{h})$, where $\mathcal{R}_0$ is diagonal with entry $i,\, i$ equals $E\{s^2(n)[s^2(n)-r_2]^2\}/E[s^2(n)]$ when $i \neq \nu+1$, and $E\{s^4(n)[s^2(n)-r_2]^2\}/\{E[s^2(n)]\}^2$ when $i = \nu+1$. For a constant modulus signal, $s^2(n) = r_2$ so that $\mathcal{R}(\delta_\nu) = 0$ but not $\mathcal{H}(f_*)$; thus, the EMSE is zero in a noisy-free situation. For a nonconstant modulus input constellation, the Lyapounov equation to be solved becomes $\Psi(\delta_\nu)\mathcal{T}(\mathbf{h})^\top\mathcal{T}(\mathbf{h})P_0 + P_0\mathcal{T}(\mathbf{h})^\top\mathcal{T}(\mathbf{h})\Psi(\delta_\nu) = \mathcal{R}_0$, where we are looking for $P_* = \mathcal{T}(\mathbf{h})^\top\mathcal{T}(\mathbf{h})P_0\mathcal{T}(\mathbf{h})^\top\mathcal{T}(\mathbf{h})$. The exact solution of this Lyapounov equation is difficult to express analytically. For a large enough number of equalizer taps, we neglect the difference between the term corresponding to $i = \nu+1$ and the others so that $\mathcal{T}(\mathbf{h})^\top\mathcal{T}(\mathbf{h})P_0 + P_0\mathcal{T}(\mathbf{h})^\top\mathcal{T}(\mathbf{h}) = \alpha I$ with $\alpha = E\{s^2(n)[s^2(n)-r_2]^2\}/\{E[s^2(n)](3-\rho)\}$ is to be solved. $P_0 = (\alpha/2)(\mathcal{T}(\mathbf{h})^\top\mathcal{T}(\mathbf{h}))^{-1}$, and thus, $P_* = (\alpha/2)\mathcal{T}(\mathbf{h})^\top\mathcal{T}(\mathbf{h})$. Finally, noticing that $\mathrm{Trace}\{\mathcal{T}(\mathbf{h})^\top\mathcal{T}(\mathbf{h})\} = NE[\mathbf{y}^\top(n)\mathbf{y}(n)]/E[s^2(n)]$ yields (4). △

Proposal 1 is to be interpreted with respect to previous asymptotic error results about nonfractional LMS and CMA; see [12] and [14], for example. Similarly to these studies, the CMA MSE is linearly proportional to the step size μ, the equalizer time-span N, and the received signal power $E[\mathbf{y}^\top(n)\mathbf{y}(n)]$. The new term appearing in this study is related to the input constellation shape and size M by

$$\mathrm{m}(M) = E[s^2(n)]^2\{E[s^6(n)]/E[s^2(n)]^3 - \rho^2\}/(3-\rho).$$

For sake of simplicity, we study, in the sequel, the effect of this term for a M-PAM sequence. When the input signal is a binary sequence (i.e., it has a constant modulus), $m(2) = 0$ so that the MSE depends only on the channel additive noise, which is assumed here to be very low. As M is chosen greater than 2, the jitter increases significantly with the constellation order; see Table II. This explains the behavior noticed on simulations (Fig. 2). However, if one looks more carefully, $\mathrm{m}(M)$ tends to some upper-bound value. This means that the EMSE is bounded despite increases in M. In order to compare the EMSE to $(d/2)^2$, which is the square of half the distance between points of the input constellation, we display the EMSE normalized by the distance between points of constellation—it will be called "distance normalized EMSE." (See Table II.) As expected, it increases strongly when M increases. When the EMSE becomes constant, the distance normalized EMSE should be proportional to $M(M^2-1)$.

TABLE II
M-PAM: EFFECT OF M ON THE DISTANCE NORMALIZED $m(M)$ AND A POSSIBLE CHOICE OF STEP-SIZE VERSUS M

M	2	4	8	16	32
$\mathrm{m}(M)/(d/2)^2$ in EMSE	0	0.2118	2.1758	10.380	43.288
$\mu(M)$	any	$4.72\mu_0$	$0.4596\mu_0$	$0.0963\,\mu_0$	$0.0231\,\mu_0$

TABLE III
M-PAM: EFFECT OF M ON EXPERIMENTAL EMSE USING THE STEP-SIZE SUGGESTED IN TABLE II

M	4	8	16	32
$\mu(M)$	$4.7\ 10^{-3}$	$4.6\ 10^{-4}$	$9.6\ 10^{-5}$	$2.3\ 10^{-5}$
exper. EMSE	$8.521\ 10^{-3}$	$1.833\ 10^{-3}$	$4.40\ 10^{-4}$	$1.38\ 10^{-4}$
exper. EMSE $/(d/2)^2$	0.043	0.039	0.037	0.047

We choose to design a step size adjusted to the constellation size $[\mu(M)]$ in order to achieve a given value of distance normalized EMSE. For example, the step size could be determined for $M = 8, 16, \cdots$ from that tuned for $M = 4$ by keeping $\mu(M)\mathrm{m}(M)/d^2(M)$ equal to some fixed value. In the last row of Table II, we show how the step-sizes for $M = 8, 16$, and 32 are obtained by scaling the step-size μ_0 chosen for $M = 4$. Indeed, the resulting $\mu(M)$ is reduced as M increases.

In order to check the validity of (4) and the distance-normalized EMSE consistancy of our step-size scaling guideline, we consider the step-size values suggested in Table II. If (4) is a realistic approximation, the normalized EMSE found for different constellation sizes should be equivalent. In Table III, we display the EMSE observed in the case $L = 2$ with the transfer function of the channel described in Section III-A and 45-dB SNR. The step-size values are these suggested in Table II with $\mu_0 = 10^{-3}$. Each result corresponds to CMA simulation over 50 000 iterations to assure the attainment of steady state. EMSE is averaged over the last 1000 values.

The bottom row of Table III presents the expected result with the distance-normalized EMSE's in close agreement.

V. CONCLUSION

We derive an analytical approximation (verified by simulation) of the CMA EMSE around the mean solutions corresponding to the minima of the noise-free CM cost function. The importance of the distribution of nonconstant modulus constellations is shown in terms of its normalized fourth- and sixth-order moments. Finally, we provide the reader with a design guideline for the selection of step size in order to preserve EMSE normalized by the distance between constellation points across changes in constellation size.

REFERENCES

[1] S. Bellini, "Bussgang techniques for blind deconvolution and equalization," *Blind Deconvolution.* Englewood Cliffs, NJ, Prentice-Hall, 1994.

[2] A. Benveniste, M. Métivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximations.* New York: Springer-Verlag, 1990.

[3] R. R. Bitmead, S.-Y. Kung, B. D. O. Anderson, and T. Kailath, "Greatest common divisors via generalized Sylvester and Bezout matrices," *IEEE Trans. Automat. Contr.,* vol. AC-23, pp. 1043–1047, 1978.

[4] D. Donoho, "On minimum entropy deconvolution," *Applied Time-Series Analysis II.* New York: Academic, 1981, pp. 565–609.

[5] I. Fijalkow, F. Lopez de Victoria, and C. R. Johnson, Jr., "Adaptive, fractionally spaced blind equalization," in *Proc. IEEE Digital Signal Process. Workshop,* Yosemite, CA, 1994, pp. 257–260.

[6] 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.

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

[8] C. R. Johnson, Jr., and B. D. O. Anderson, "Godard blind equalizer error surface characteristics: White, zero-mean, binary source case," *Int. J. Adaptive Contr. Signal Process.,* vol. 9, pp. 301–324, 1995.

[9] S. Mayrargue, "A blind spatio-temporal equalizer for radio-mobile channel using the constant modulus algorithm (CMA)," in *Proc. ICASSP,* 1993, pp. 344–347.

[10] 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.

[11] Y. Li and Z. Ding, "Global convergence of fractionally spaced Godard adaptive equalizers," in *Proc. Asilomar Conf.,* 1994.

[12] S. U. H. Quershi, "Adaptive equalization," *Proc. IEEE,* vol. 73, pp. 1349–1387, 1985.

[13] 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.

[14] J. J. Shynk, R. P. Gooch, G. Krisnamurthy, and C. K. Chan, "A comparative performance study of several blind equalization algorithms," *Proc. SPIE, Adv. Signal Process.,* vol. 1565, 1991.

[15] 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.

[16] J. R. Treichler, I. Fijalkow, and C. R. Johnson, Jr., "Fractionally spaced equalizers: How long should they be?," *IEEE Signal Processing Mag.,* vol. 13, pp. 65–81, May 1996.

[17] V. Wolf, R. Gooch, and J. Treichler, "Specification and development of an equalizer/demodulator for wideband digital microwave radio signals," in *WESCON Conf. Record,* 1987, paper 6.3.

[18] H. Zeng, L. Tong, and C. R. Johnson, Jr., "Mean square error performance of CMA receivers," in *Proc. Asilomar Conf.,* Oct. 1996.

Nonlinear Location and Scale Estimators of Fuzzy Numbers

Vassilios Chatzis and Ioannis Pitas

***Abstract*— In this correspondence, the extension principle is used in order to fuzzify location and scale estimators when used on fuzzy numbers. First, fuzzy nonlinear means are defined as extensions of the corresponding crisp means. Fuzzy L location and scale estimators, which are based on fuzzy-order statistics, are defined as extensions of the crisp L location and scale estimators. The most widely used scale estimator, which is known as the sample standard deviation, is also extended to fuzzy numbers through the extension principle. Equivalent relations that can be used to calculate the fuzzy estimators by using crisp arithmetic are also given for each one of the proposed fuzzy estimators.**

***Index Terms*—Extension principle, fuzzy estimastors, fuzzy numbers, fuzzy ranking.**

I. Introduction

The *fuzzy set theory* was first introduced by Zadeh [8]. He used this word to generalize the mathematical concept of the *set* to one of the *fuzzy set.* Suppose that the available information is such that the uncertain value can be located inside a closed interval $I \subset R$, which we call *interval of confidence*. Then, a *membership function* that maps each element of the interval of confidence to a value in the interval [0, 1] is defined. The concept of a fuzzy number X is presented either by its membership function $X = \{(x, \mu_X(x)), x \in I_X\}$ or by the union of its α *cuts* $X = \bigcup_\alpha \alpha \cdot [x_l^{(\alpha)}, x_r^{(\alpha)}]$, where $\alpha \in [0, 1]$. In order to extend mathematical laws of crisp numbers in fuzzy theory, we can use the *extension principle* [2], [3], [6], which provides the theoretical warranty that fuzzifies the parameters or arguments of a function, resulting in computable fuzzy sets. The extensions of the basic operations are presented in Table I.

The novel contribution of this correspondence is the use of fuzziness concepts in estimation theory. The observation data will be considered to have uncertain values (fuzzy numbers). For example, fuzzy numbers can describe knowledge regarding the conditions of the observation by changing fuzziness when observation conditions change. Fuzzy numbers can also be considered to be the output of fuzzy inference mechanisms. When fuzzy numbers have to be combined in a nonlinear way, fuzzy nonlinear estimators (e.g., fuzzy median) have to be defined. In this correspondence, the extension principle is used to fuzzify location and scale estimators that are useful in estimation theory as well as in applications, e.g., in digital signal and image processing.

The paper has the following structure. In Section II, the fuzzy nonlinear means and the fuzzy L location estimators based on fuzzy-order statistics are defined. In Section III, the fuzzy scale estimators, such as fuzzy L scale estimators and the fuzzy sample standard deviation, are also defined. Equivalent relations are also given for every estimator, which can be used to calculate the corresponding fuzzy estimators by using crisp arithmetic. Conclusions are drawn in Section IV.

Manuscript received January 8, 1996; revised June 10, 1997. The associate editor coordinating the review of this paper and approving it for publication was Prof. Peter C. Doerschuk.

The authors are with the Department of Informatics, Aristotle University of Thessaloniki, Thessaloniki, Greece (e-mail: pitas@zeus.csd.auth.gr).

Publisher Item Identifier S 1053-587X(98)00521-2.