

## CMA FRACTIONALLY SPACED EQUALIZERS: STATIONARY POINTS AND STABILITY UNDER IID AND TEMPORALLY CORRELATED SOURCES

JAMES P. LEBLANC<sup>1,\*</sup>, INBAR FIJALKOW<sup>2</sup> AND C. RICHARD JOHNSON, JR.<sup>3</sup>

<sup>1</sup> *Klipsch School of ECE, New Mexico State University, Las Cruces, NM, U.S.A.*

<sup>2</sup> *ENSEA/ETIS, 6 av. du Ponceau, 95014 Cergy-Pontoise Cdx, France*

<sup>3</sup> *Electrical Engineering Department, Cornell Univ. Ithaca, NY, U.S.A.*

### SUMMARY

A common assumption in blind equalization schemes using the Constant Modulus Algorithm (CMA) is that the source sequence is an independent identically distributed (i.i.d.) sequence with equiprobable symbols. Much of the analysis demonstrating the global convergence of CMA in a noiseless channel to an open-eye setting uses this assumption. This work investigates the effect of source statistics (distributions and correlations) on the location of CMA stationary points in the fractionally sampled equalizer case under the conditions of equalizability. The work identifies the stationary points as the solution set of a system of multivariate polynomial equations with monomial coefficients given by the source moments. © 1998 John Wiley & Sons, Ltd.

Key words: blind equalization; Godard algorithm; CMA; stationary point analysis

### 1. INTRODUCTION

In a digital communication setting, the transmitted signal is often distorted through a dispersive channel which introduces intersymbol interference (ISI). In a high-SNR situation, where ISI is the limiting factor more so than channel noise or interference, a linear equalizer may be used to remove ISI. Successful source estimation from the equalizer output is accomplished by decision device (e.g. a quantizer) if enough of the ISI is removed. A condition on the combined channel–equalizer resulting in source recovery is known as an ‘open-eye’ condition. Due to the fact that the (unknown) channel impulse response may vary over time, an adaptive equalization scheme, with its ability to track a time-varying system is considered.

A standard adaptive filtering scheme (e.g. LMS) adjusts the tap weights based on an error signal which is the difference between the filter (equalizer) output and the desired output (e.g. the transmitted sequence). To differentiate our problem from those that have the desired output available at least intermittently (also known as a training sequence), the term *blind* is used to denote the absence of a known transmitted sequence. The lack of need for such a training signal is an advantage of blind equalization schemes since the use of training signals may become cumbersome in broadcast mode and does not make efficient use of available bandwidth.

\* Correspondence to: James P. LeBlanc, Klipsch School of ECE, New Mexico State University, Las Cruces, NM, U.S.A.  
E-mail: leblanc@nmsu.edu.

A variety of blind equalization algorithms may be used during the initial phase of adaptive equalization. After any of these have converged to appropriately low ISI equalization settings (i.e. open-eye), the equalization scheme may be switched to a specific blind algorithm known as the 'decision-directed' approach to eke out maximum performance of the equalizer.<sup>1</sup> In such schemes, it is crucial that the blind equalizers have convergent behaviour to low ISI channel-equalizer settings. Hence, the study of the convergent behaviour of blind equalization algorithms has been the focus of much research. An excellent tutorial on the subject of adaptive equalization in general may be found in Reference 2.

We focus on one particular, commonly used, blind equalization method, considering an adaptive linear feedforward equalizer with tap weights updated by the Constant Modulus Algorithm (CMA)<sup>3</sup> or Godard<sup>4</sup> which uses a specific memoryless polynomial error function for parameter update. Although we focus specifically on CMA in this work, it is the polynomial nature of the error function that allows application of the described methods. Thus, the formulation and methodology presented here are not limited strictly to CMA, but may be applicable to any Bussgang-style scheme using a memoryless, polynomial error function.

To date, *infinitely* long baud rate equalizers adapted by CMA have been shown to be globally convergent to a perfect equalization setting for an independent source sequence with equiprobable symbols.<sup>5</sup> However, to reflect the situation in which many equalizers are actually used, more recent work has regarded fractionally spaced equalizers (FSEs) in which the received waveform is sampled at a rate higher than the baud rate. FSEs have been shown to be able to 'perfectly equalize' moving average (MA) channels provided certain conditions on *finite* length moving average equalizer and channel are met.<sup>7</sup> Moreover, lacking channel noise and assuming a source drawn from an independent and identically distributed (i.i.d.) and non-Gaussian sequence then fractionally spaced CMA is known to be globally convergent to a perfect equalization setting.<sup>8,9</sup>

To our knowledge little published work<sup>10-15</sup> has appeared addressing how critical the source sequence assumption is for the proper behaviour of blind equalization schemes. There are some source sequence estimation methods (hidden Markov models, direct source sequence estimation, Viterbi-like methods) which are possibly analysable and applicable to source sequences which are correlated. However, as they may be computationally formidable, they are considered to be in a different class of algorithms than the one studied herein. Also, some of these methods require *a priori* knowledge of the source statistics which is not necessarily available. We choose to address the *effect* of source statistics on the popular and computationally simple adaptive blind equalization algorithm, CMA. Source statistics are considered with respect to the source distribution from which source elements are drawn, as well as their temporal correlation. The investigation on distributional aspects is relevant to constellation shaping (shaping gain), while temporally correlated sources may appear in any system lacking data scramblers and/or source encoders. Several situations in which CMA fails to converge are documented in Reference 14.

## 2. PROBLEM SETTING

### *Fractionally spaced equalization*

Much of the analytic work on blind equalization algorithms considered the baud spaced equalization problem. This was despite the fact that many equalizers built today use a technique known as 'fractional sampling'.<sup>16,17</sup>

Consider the adaptive fractionally spaced equalizer in the system block diagram in Figure 1. Here the path from discrete source sequence  $s(k)$  with baud spacing of  $T$  seconds to the received

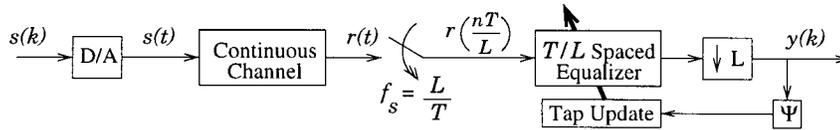


Figure 1. Fractionally-spaced adaptive equalization block diagram

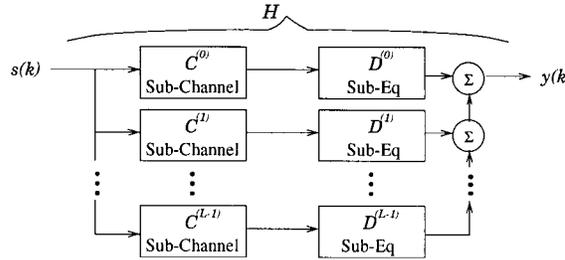


Figure 2. Associated sub-channel model

(analog) signal  $r(t)$  is modelled as the cascade of a discrete-to-continuous device (D/A) and a continuous-time channel impulse response  $c(t)$ . The received waveform is then sampled at intervals of  $T/L$  seconds (where  $L$  is a positive integer) or sampled at rate  $T$  using  $L$  sensors. This sampled sequence is then filtered through a discrete-time equalizer whose output is then downsampled by  $L$  to produce the baud-space output sequence  $y(k)$ .

The fractionally spaced system may be alternatively described by the sub-channel model in Figure 2, in which all sub-blocks are considered as baud-spaced discrete-time systems. This formulation eases introduction of the aforementioned ‘equalizability’ conditions allowing perfect equalization. With the time-span of the equalizer exceeding the time-span of the channel, and sub-channels (as polynomials in  $z$ ) having no common roots, then there exists an equalizer parametrization capable of setting the combined channel–equalizer system to a pure delay with no intersymbol interference.<sup>7</sup> More specifically related to CMA equalizers are the additional conditions that the i.i.d. source sequence be drawn from an equiprobable alphabet and lacking channel noise. Then the fractional equalizer adapted under the CMA algorithm is globally convergent to perfect equalization setting (no ISI).<sup>8,9</sup>

**CMA**

CMA is considered to be the most widely used and tested blind equalization algorithm (Proakis and Nikias in Reference 18). Due to CMA’s popularity, it is chosen as the algorithm for investigation. The blind adaptive equalization algorithm which is the focus of this work was independently developed by Godard<sup>4</sup> and Treichler<sup>3</sup> and is known as the Godard Algorithm, or Constant Modulus Algorithm (CMA).

The CMA cost function is defined as

$$J = \frac{1}{4} E\{(|y|^2 - \gamma)^2\} \tag{1}$$

where  $\gamma$  is a positive real constant. Differentiating  $J$  with respect to  $y$  yields the error function as

$$\Psi = y(|y|^2 - \gamma) \tag{2}$$

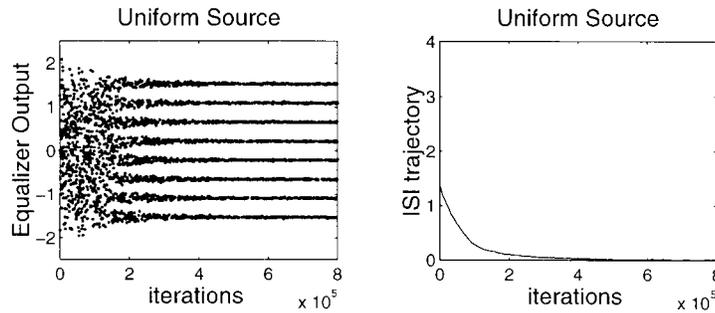


Figure 3. Example A

The update law of the equalizer impulse response ( $D$ ) then follows as

$$D(k+1) = D(k) + \mu \Psi(k) R(k) \quad (3)$$

where  $R(k)$  is the vector containing samples of the received signal (i.e. the equalizer state given by  $R(k) = [r_k \ r_{k-1} \ \dots \ r_{k-n}]$ ).

In order to characterize the performance of a linear equalizer, a quantitative measure of ISI is defined as

$$\varphi = \frac{\sum_i |h_i| - \max_i |h_i|}{\max_i |h_i|} \quad (4)$$

also known as the *peak distortion*.<sup>19</sup> This is a measure of the maximum intersymbol interference present. For an  $m$ -PAM source, when  $\varphi < 1/(m-1)$  then the channel–equalizer combination is considered to be ‘open-eye’, and a decision device will successfully reconstruct the source in the absence of noise.

#### Numerical Examples of CMA

Examples A, B, and C demonstrate source statistics effects on CMA, these examples share the common parameters: over sampling factor  $L=2$ , adaptive stepsize  $\mu = 8 \times 10^{-5}$ , 5-tap (half-baud spaced) channel, 5-tap (half baud spaced) equalizer, 8-PAM unit variance alphabet, ( $\mathcal{A} = \{\pm\Delta, \pm 3\Delta, \pm 5\Delta, \pm 7\Delta\}$  with  $\Delta = 2/\sqrt{84}$ ), channel taps,  $c = [-0.57 \ 1.00 \ -0.28 \ -0.82 \ -0.50]^T$ , equalizer initialization  $D(0) = [0 \ 0 \ 1 \ 0 \ 0]^T$ .

*Example A (Proper CMA behaviour).* The source sequence elements  $s(k)$  are independently drawn from the alphabet with equal probability. Figure 3 plots the equalizer output  $y(k)$ . Initially the equalizer output is smeared over the range from  $-2$  to  $2$ . As tap adaptation takes place the smearing is reduced, appearing finally as very tight clusters about the true alphabet member values. Here, the equalizer has reduced ISI to allow source sequence recovery by the use of a quantizer (nearest-neighbour decision device). Note also from Figure 3,  $\varphi$  approaches zero indicating that only one combined channel–equalizer ( $H$ ) tap is appreciably non-zero.

*Example B (Source distribution).* Here, all parameters are the same as in Example A, except that the probabilities of occurrence of alphabet members are unequal;  $p(\pm\Delta) = 0.75$ ,  $p(\pm 3\Delta) = 0.05$ ,

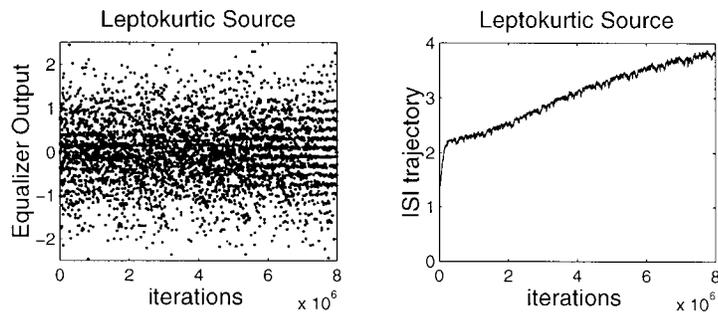


Figure 4. Example B

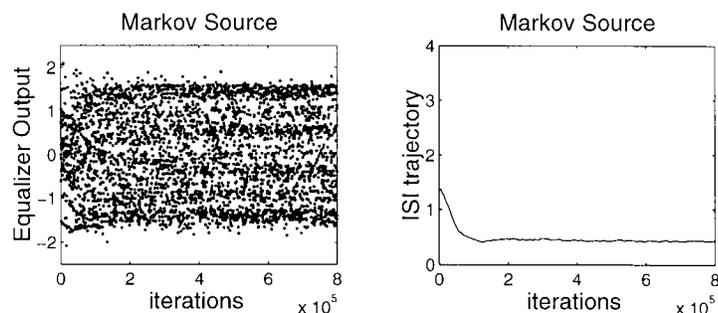


Figure 5. Example C

$p(\pm 5\Delta) = 0.05$ ,  $p(\pm 7\Delta) = 0.15$  (yielding a *leptokurtic* source—a term to be defined later). Figure 4 displays the results below which demonstrate dismal performance in which CMA yields ISI enhancement.

A detailed analysis of the situation is presented later with a discussion on more subtle effects of the source distribution which, while yielding acceptable ISI reduction, reduce error surface curvature resulting in slower equalizer convergence rates. The implication of this will be recognized as a tradeoff between source constellation shaping gain and equalizer convergence rates.

*Example C (Source correlation).* Consider a temporally correlated source sequence generated by a Markov model (see Appendix II) under similar parameters as in Examples A and B. The deleterious effects are shown in Figure 5 wherein the equalizer output never coalesces into bands about the alphabet members, demonstrating significant ISI.

Although such simulations of adaptive systems often provide insight to the convergence properties of an algorithm, it is difficult to draw conclusions about the asymptotic convergent parametrizations from such limited iteration simulations. However, such poorly performing settings will be shown to exemplify true CMA minima under correlated sources. A method which may be used to find *all* stationary points of the CMA error surface given specific correlation statistics is introduced. Stationary point computations are performed for two models of temporally correlated sources (e.g. periodic and Markov sources).

### Multivariate polynomial formulation

We specifically address the use of fractionally spaced CMA under the equalizability conditions. This allows analysis of CMA to take place in the combined channel–equalizer space ( $H$ -space) rather than strictly in the equalizer tap space ( $D$ -space). Thus, our results are general to any channel (meeting equalizability).

Although CMA is often used for two-dimensional signal constellations (i.e. complex alphabet), we limit our discussion to situations using real alphabets (e.g.  $M$ -PAM—Pulse Amplitude Modulation with  $M$  discrete amplitude values). The motivation behind this limitation is to maintain differentiability of the cost function and keep the notation simple. The CMA cost as a function of equalizer output  $y$  with a real alphabet becomes

$$J = \frac{1}{4} \mathbb{E}\{(y^2(k) - \gamma)^2\} \quad (5)$$

where  $\gamma = \mathbb{E}\{a^4\}/\mathbb{E}\{a^2\}$  for  $a \in \{\mathcal{A}\}$  and  $\mathcal{A}$  is a real alphabet. With the combined channel–equalizer tap weights defined as  $H = [h_0 \ \dots \ h_n]^T$  and  $S(k) = [s(k) \ \dots \ s(k-n)]^T$  as the column vector containing source sequence elements, we have  $y(k) = H^T S(k)$ . The combined channel–equalizer response,  $H$  (also assumed real), may be written as the product of the  $L$ -multichannel convolution matrix  $\mathbb{C}$  and the equalizer tap weight vector  $D$  as  $H = \mathbb{C}^T D$  where

$$\mathbb{C} = \begin{bmatrix} c_0^{(0)} & \dots & c_n^{(0)} \\ \vdots & & \vdots \\ c_0^{(L-1)} & \dots & c_n^{(L-1)} \\ & \ddots & & \ddots \\ & & c_0^{(0)} & \dots & c_n^{(0)} \\ & & \vdots & & \vdots \\ & & c_0^{(L-1)} & \dots & c_n^{(L-1)} \end{bmatrix} \quad (6)$$

and  $[c_0^{(i)} \ c_1^{(i)} \ \dots \ c_n^{(i)}]$  is the baud spaced impulse response of sub-channel  $i$ . The CMA cost function in terms of the channel and equalizer parametrizations is

$$J = \frac{1}{4} \mathbb{E}\{((\mathbb{C}^T D)^T S(k))^2 - \gamma\}^2 \quad (7)$$

Invoking the assumptions of equalizability, yields the matrix  $\mathbb{C}$  to be full rank,<sup>20</sup> meaning that all of  $H$ -space is reachable by choice of equalizer  $D$ . The error surface stationary points are given by  $\partial J / \partial D = \mathbf{0}$  where

$$\frac{\partial J}{\partial D} = \mathbb{C} \mathbb{E}\{((H^T S(k))^3 - \gamma H^T S(k)) S(k)\}. \quad (8)$$

With  $\mathbb{C}$  full column rank then  $\partial J / \partial D = \mathbf{0}$  necessarily implies

$$\mathbb{E}\{((H^T S(k))^3 - \gamma H^T S(k)) S(k)\} = \mathbf{0}. \quad (9)$$

There are a variety of reasons to investigate the error surface in terms of  $H$  space. First, it is precisely the *combined* channel–equalizer parametrization which determines the equalizer output. In turn, any meaningful performance metric must be considered in terms of the equalizer output (e.g. bit error rate, ISI level). Also, in certain situations there may be a variety of different equalizer parametrizations that yield identical combined channel–equalizer responses. In Reference 20 it is noted that in cases where the equalizer time support sufficiently exceeds the channel time support, a dense subspace in  $D$  corresponds to a point in  $H$  space. As such, it is difficult to discuss stationary *points* in equalizer space. Expanding (9) for clarity we have

$$E \left\{ \left( (s(k)h_0 + \dots + s(k-n)h_n)^3 - \gamma(s(k)h_0 + \dots + s(k-n)h_n) \right) \begin{bmatrix} s(k) \\ \vdots \\ s(k-n) \end{bmatrix} \right\} = \mathbf{0}$$

Taking the expectation operator with respect to the source regressor elements  $s_i$  leads to  $n+1$  multivariate polynomial equations in  $H$  space with monomials having coefficients given by the fourth and second moments of the source signal. Denote these moments of a symmetrically distributed source as

$$\mathcal{R}_k^i = E\{s(m-i)s(m-j)s(m-k)s(m-\ell)\}, \quad \mathcal{R}_j^i = E\{s(m-i)s(m-j)\}. \quad (10)$$

We are assuming a symmetric distribution wherein  $E\{s\} = E\{s^3\} = 0$  but not temporal independence as in References 8, 9 and 21. Since we are considering correlations of stationary processes, keeping all four (two) lag indices for the fourth (second) moments is unnecessary, but they are kept since they aid in showing the structure. The resulting system of equations  $F$  may then be written as

$$F(H) = \begin{bmatrix} f_0(H) \\ \vdots \\ f_n(H) \end{bmatrix} = \begin{bmatrix} \sum_i \mathcal{R}_i^0 h_i^3 + \sum_{i \neq j} \mathcal{R}_j^0 h_i^2 h_j + \sum_{i \neq j \neq k} \mathcal{R}_k^0 h_i h_j h_k - \gamma \sum_i \mathcal{R}_i^0 h_i \\ \vdots \\ \sum_i \mathcal{R}_i^n h_i^3 + 3 \sum_{i \neq j} \mathcal{R}_j^n h_i^2 h_j + \sum_{i \neq j \neq k} \mathcal{R}_k^n h_i h_j h_k - \gamma \sum_i \mathcal{R}_i^n h_i \end{bmatrix} \quad (11)$$

Values of  $H$  for which  $F(H) = \mathbf{0}$  define the CMA stationary points. For an  $n + 1$  tap system, this yields a set of  $n + 1$  equations in the  $n + 1$  unknowns ( $h_m, 0 \leq m \leq n$ ).

Each of the stationary points (solutions of  $F(H)$ ) may be categorized as a saddle, local minimum, or local maximum of the error surface depending on the eigenvalues of the Hessian matrix evaluated at the stationary point. Calculations similar to those in (9) verify that the sign of the eigenvalues of the cost function Hessians with respect to  $D$  and  $H$  agree. Denote the error surface Hessian by the matrix  $\mathbb{M}$ , with  $\mathbb{M}(i, j) = \partial f_i / \partial h_j$ . For a local minima (maxima),  $\mathbb{M}$  is positive (negative) definite.  $\mathbb{M}$  having both positive and negative eigenvalues denotes a saddle point. In cases where  $\mathbb{M}$  is singular, the stationary points are degenerate (i.e. a dense neighborhood satisfies  $\partial J / \partial H = \mathbf{0}$ ).

### 3. I.I.D. SOURCE CASE: DISTRIBUTIONAL ASPECTS

In this section, we assume that source symbols are drawn independently from some distribution (i.i.d.). The system of equations defining the stationary points is solved and classification of the stationary points by the number of non-zero elements in the tap vector is presented. Categorization

of stationary points into minima, saddles, and maxima is shown to be dependent on *class membership* and the source *kurtosis* (quantities to be introduced). Analytic results are meshed with computer simulations describing the disappearance of error surface curvature (and related equalizer convergence rate issues) as the source kurtosis approaches that of a Gaussian distribution (such as occurs with constellation shaping).

A concept that appears repeatedly in the stability analysis of the CMA stationary points having a profound effect on the error surface is the dimensionless quantity known as *kurtosis*.<sup>\*</sup> Denote the central moments by  $M_n = E\{(x - \mu)^n\}$  where  $\mu = E\{x\}$ . Then kurtosis is defined to be the ratio,  $\kappa_x = M_4/M_2^2$ . The minimum possible kurtosis value is  $\kappa = 1$  which corresponds to a symmetric equally probable distribution of only two values (i.e. BPSK source). The continuous uniform distribution has  $\kappa = 1.8$ . The continuous Gaussian distribution has  $\kappa = 3$  and is regarded as forming the boundary between *platykurtic* (low kurtosis,  $\kappa_s < 3$ ) distributions and *leptokurtic* (high kurtosis,  $\kappa_s > 3$ ) distributions.

The importance of kurtosis in the deconvolution problem has appeared in References 23 and 24, and its relation to the CMA cost function can be seen by rewriting the CMA cost function as  $J = \frac{1}{4}((\kappa_y - 1)E\{y^2\}^2 + (E\{y^2\} - \gamma)^2)$ . For both platykurtic and leptokurtic sources the CMA error surface minima coincide with minimum kurtosis of equalizer output. It will be shown that in the platykurtic source case kurtosis reduction realizes *ISI reduction*, while for leptokurtic sources this yields *ISI enhancement*.

*Stationary points*

Note that for the i.i.d. source all the correlation coefficients of (11) are zero, except  $\mathcal{R}_i^i = M_4$ ,  $\mathcal{R}_j^j = M_2^2$  for  $i \neq j$  and  $\mathcal{R}_i^i = M_2$ . Then each of the  $n + 1$  equations of  $F$  can be rewritten as  $f_m$ , for  $0 \leq m \leq n$ ,

$$f_m = M_2^2 h_m \left( \kappa_s h_m^2 + 3 \sum_{i \neq m} h_i^2 - \kappa_s \right) \tag{12}$$

The CMA stationary points are those values of  $H$  solving each  $f_m = 0$ . Any  $h_m = 0$  trivially solves  $f_m = 0$ . So, only  $f_m$  for which  $h_m \neq 0$  need be further considered. For any  $h_m \neq 0$  we have the condition

$$\kappa_s h_m^2 + 3 \sum_{i \neq m} h_i^2 - \kappa_s = 0 \tag{13}$$

When  $\kappa_s = 3$ , all the  $f_m$  equations are the same, our system is degenerate and we need only satisfy  $\sum_i h_i^2 = 1$ . In this case, the stationary points in  $H$  are not discrete but a dense set corresponding to the unit sphere. Thus, for  $\kappa_s = 3$  CMA adaptation will converge to and then wander about this sphere.

Alternatively, when  $\kappa_s \neq 3$  each non-zero  $h_m$  in a solution set must solve *each* equation  $f_m$ , we then have that the value of  $h_m^2$  for all non-zero  $h_m$  in this solution must be the same

<sup>\*</sup> The reader is cautioned that some texts define kurtosis a bit differently as  $\kappa = (M_4/M_2^2) - 3$ . We shall however follow the definition above as found in Reference 22.

(i.e. all  $h_m$  have the same magnitude). This is a generalization of a similar result in Reference 25 pertaining to a BPSK source with equal probabilities.

For the i.i.d. source case, the *value* of each non-zero  $h_m$  will depend on the total *number* of non-zero elements of  $H$  for any  $F(H)=\mathbf{0}$ . A useful construct is to divide the set of solutions of (11) into classes. Define the class,

$$\mathcal{C}_N = \{H: F(H)=\mathbf{0} \text{ where } H \text{ has exactly } N \text{ non-zero elements}\}. \quad (14)$$

Denote the size of the class (i.e. the number of its members) by  $|\mathcal{C}_N|$ .

Of the total  $n+1$  elements of  $H$ , since each of its  $N$  non-zero elements may have one of two values and there are  $\binom{n+1}{N}$  ways to group  $N$  elements we have,  $|\mathcal{C}_N|=2^N \binom{n+1}{N}$ . The total number of CMA stationary points under i.i.d. source is  $3^{n+1}$ .

Class membership defines the magnitude of the non-zero taps, for  $H \in \mathcal{C}_N$  (13) becomes

$$\kappa_s h_m^2 + 3(N-1)h_i^2 - \kappa_s = 0 \Rightarrow h_m = \begin{cases} 0, \\ \pm \sqrt{\frac{\kappa_s}{\kappa_s - 3 + 3N}}, \end{cases} \quad 0 \leq m \leq n \quad (15)$$

#### Categorization of solutions by Hessian

The Hessian evaluated at stationary point categorizes each as a local minimum ( $\mathbb{M}$  is positive definite), local maximum ( $\mathbb{M}$  is negative definite), or a saddle point ( $\mathbb{M}$  is non-definite). The diagonal elements of the Hessian are

$$\mathbb{M}(m, m) = \frac{\partial f_m}{\partial h_m} = M_2^2 \left( 3\kappa_s h_m^2 + 3 \sum_{i \neq m} h_i^2 - \kappa_s \right) \quad (16)$$

while the off-diagonal elements ( $m \neq n$ ) can be written as,

$$\mathbb{M}(m, n) = \frac{\partial f_m}{\partial h_n} = 6M_2^2 h_n h_m. \quad (17)$$

Evaluating these expressions at a stationary point for the case  $H \in \mathcal{C}_N$  we need to consider the three cases  $\mathbb{M}(m, m)$  with  $h_m \neq 0$ ,  $\mathbb{M}(m, m)$  with  $h_m = 0$ , and  $\mathbb{M}(n, m)$  with  $m \neq n$ .

All the elements of  $\mathbb{M}$  at a stationary point in  $\mathcal{C}_N$  for  $N \geq 1$  contain a common factor of  $M_2^2(\kappa_s/(\kappa_s - 3 + 3N))$  which may be factored out to simplify notation. Introduce  $\widehat{\mathbb{M}}$  as

$$\mathbb{M} = M_2^2 \left( \frac{\kappa_s}{\kappa_s - 3 + 3N} \right) \widehat{\mathbb{M}} \quad (18)$$

Since the common factor is strictly positive for  $N \geq 1$ , the signs of the eigenvalues of  $\mathbb{M}$  and  $\widehat{\mathbb{M}}$  agree for  $H \in \mathcal{C}_N$  with  $N \geq 1$ . The diagonal elements  $\widehat{\mathbb{M}}$  may be described by

$$\widehat{\mathbb{M}}(m, m) = \begin{cases} 3 - \kappa_s, & h_m = 0 \\ 2\kappa_s, & h_m \neq 0 \end{cases} \quad (19)$$

and off-diagonal terms are

$$\widehat{\mathbb{M}}(m, n) = \begin{cases} 6, & h_m = h_n \neq 0 \\ -6, & h_m = -h_n \neq 0 \\ 0, & \text{else} \end{cases} \quad (20)$$

Table I. Summary of stationary point analysis

Class	$ \mathcal{C}_i $	$\kappa_s < 3$			$\kappa_s > 3$			Comment
		No. pos. eigvals	No. neg. eigvals	Type	No. pos. eigvals	No. neg. eigvals	Type	
$H \in \mathcal{C}_0$	1	0	$n + 1$	max	0	$n + 1$	max	Origin
$H \in \mathcal{C}_1$	$2(n + 1)$	$n + 1$	0	min	1	$n$	saddle	Perfect eq.
$H \in \mathcal{C}_N$	$2^N \binom{n+1}{N}$	$n - N$	$N + 1$	saddle	$N + 1$	$n - N$	saddle	For $2 \leq N \leq n$
$H \in \mathcal{C}_{n+1}$	$2^{n+1}$	1	$n$	saddle	$n + 1$	0	min	Max ISI

Since the eigenvalues of  $\widehat{\mathbb{M}}$  are merely scaled versions of the eigenvalues of  $\mathbb{M}$  we investigate the eigenvalues of  $\widehat{\mathbb{M}}$  to ascertain categorization of a stationary point. Such categorization is based upon class membership.

$H \in \mathcal{C}_0$ : For  $H \in \mathcal{C}_0$  (i.e. the origin), we have  $\mathbb{M} = -\kappa_s \mathbb{I}$ , where  $\mathbb{I}$  is the identity matrix. Since  $\kappa_s$  is strictly positive,  $\mathbb{M}$  is negative definite and  $H = \mathbf{0}$  is a local maximum.

$H \in \mathcal{C}_1$ : In the case of a single non-zero tap element,  $\widehat{\mathbb{M}}$  is a diagonal matrix with elements given by (19). The diagonal entry corresponding to the single non-zero  $h_m$  is  $2\kappa_s$  (which is strictly positive). The remaining diagonal entries have value  $3 - \kappa_s$ . The sign of these entries (and hence, the corresponding eigenvalues) are determined by the kurtosis of the source signal,  $\kappa_s$ . The cases are:

$\kappa_s < 3$ : For platykurtic source distributions, all diagonal entries of  $\widehat{\mathbb{M}}$  are positive and these channel-equalizer parametrizations are stable (i.e. local minima).

$\kappa_s > 3$ : For leptokurtic source distributions, there is the one positive diagonal entry of  $2\kappa_s$  and all remaining diagonal entries of  $\widehat{\mathbb{M}}$  are  $3 - \kappa_s$  which are negative and these parametrizations are saddle points (i.e. 1 positive eigenvalue, rest negative).

$H \in \mathcal{C}_N$ ,  $2 \leq N \leq n$ : Stationary points in this class are found to be saddle points, regardless of the distribution of the i.i.d. source. To demonstrate this, we will first define a few terms and note some properties of  $\widehat{\mathbb{M}}$ . Define the ordered index lists,  $\mathcal{J} = \{i : h_i \neq 0\}$  and  $\mathcal{O} = \{i : h_i = 0\}$ . Concatenating these ordered lists, we form  $\mathcal{JO}$ . That is,  $\mathcal{JO}$  lists the ascending indices of the non-zero  $h_m$  followed by the ascending indices of the zero valued  $h_m$ . To evaluate the eigenvalues of  $\widehat{\mathbb{M}}$  we make use of matrix transformations, see Appendix I. The first term  $2\kappa_s + 6(N - 1)$  is always positive and can be shown to correspond to the radial direction eigenvector. The signs of the remaining eigenvalues will depend on the value of  $\kappa_s$ . For

$\kappa_s < 3$ : Here,  $2\kappa_s - 6 < 0$ . So, this block has one positive eigenvalue, and the remaining  $N - 1$  eigenvalues being negative.

$\kappa_s > 3$ : Here, the sub-block has all  $N$  eigenvalues being positive.

In any case, whenever  $\kappa_s \neq 3$ ,  $\widehat{\mathbb{M}}$  has mixed eigenvalues with the diagonal sub-matrix and circulant sub-matrix always containing eigenvalues of opposing signs.

$H \in \mathcal{C}_{n+1}$ : In the case of all non-zero elements of  $H$  we have the entire matrix  $\widehat{\mathbb{M}}$  being circulant, with eigenvalues of  $2\kappa_s + 6(N - 1)$  and  $2\kappa_s - 6$ . For  $\kappa_s < 3$ , this corresponds to  $H$  being a saddle point. However, for  $\kappa_s > 3$ , we have that all  $H$  in this class are minima (with maximal ISI).

A summary of the stationary point stability analysis is given in Table I for  $H \in \mathbf{R}^{n+1}$ : For i.i.d. platykurtic sources ( $\kappa_s < 3$ ) the only stable stationary points are solutions  $H \in \mathcal{C}_1$ . As any

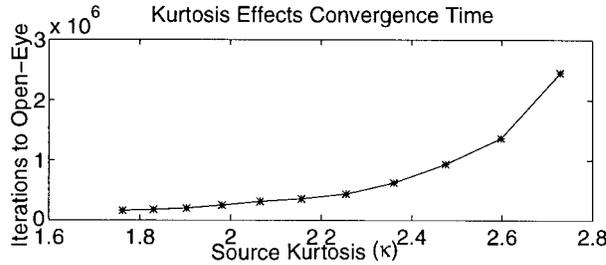


Figure 6. Eye-opening time vs. kurtosis

$H \in \mathcal{C}_1$  is a perfect equalization setting we have that CMA is globally convergent to a perfect equalization setting in this case, duplicating the results of References 8 and 9. However, with leptokurtic sources ( $\kappa_s > 3$ ), CMA is globally convergent to settings in which all taps have equal, non-zero values. Such parametrizations correspond to maximal ISI settings, providing disastrous equalization performance.

A third alternative is the case  $\kappa_s = 3$ . A Gaussian source is known to be a source which causes ill-behaviour of CMA and other blind equalization schemes<sup>6</sup> and we note that a Gaussian source distribution has  $\kappa_s = 3$ . It is shown that in the case of  $\kappa_s = 3$ , CMA will adapt to any parametrization on the sphere which, in general, does not provide adequate equalization. The class of sources for which this behaviour occurs is much broader than the oft-cited Gaussian class. (Strictly, in a digital communication setting where the source is drawn from a finite alphabet it is obvious that a *truly* Gaussianly distributed source is *not* possible.) There are many other distributions which may attain  $\kappa_s = 3$  (e.g. let  $s_i = \pm 1$  with probability  $p(\pm 1) = \frac{79 - \sqrt{97}}{90}$ ,  $s_i = \pm 4$  with probability  $p(\pm 4) = \frac{11 + \sqrt{97}}{90}$ ) which have this extreme effect on CMA. We propose a name '*meso-kurtic*' to denote them.

#### Error surface curvature and convergence

Global convergence to perfect equalization settings for platykurtic sources is shown through the above stability analysis. However, a potentially disturbing quality appears. A vast number of saddle points (of the order of  $3^{n+1}$ ) might lead to slow convergence. This concern is heightened in cases where the source kurtosis increases towards 3. From (26) and (27) it is seen that as  $\kappa_s$  approaches 3, *all but one* (the one in the radial direction) of the eigenvalues of all stationary points approaches zero. The neighbourhood of all saddle points becomes 'flatter', warning of prolonged convergence times. This should be especially noted in cases where CMA equalizers are used in conjunction with constellation shaping which increases the 'Gaussianity' of the source (e.g. the v.34 modem standard). In Reference 26 it is shown that optimal shaping for a large two-dimensional constellation is close to a Gaussian distribution. So, improved source shaping may hinder blind equalization ability and the tradeoff should be considered in the design of a communication system (see Reference 27).

In order to view the effect of kurtosis on convergence time another example is introduced. The equalizer is initialized so that the combined channel-equalizer setting has maximal ISI (i.e.  $H = [1 \ 1 \ 1 \ 1]^T$ ) and we note the number of iterations necessary to achieve an open-eye setting. Using the same channel and stepsize  $\mu$  as in earlier numerical examples Figure 6 reflects the dependence of convergence time on source kurtosis and highlights the necessity of avoiding a source kurtosis too close to  $\kappa_s = 3$ . Thus, we have an apparent tradeoff between source shaping gains (driving  $\kappa_s \rightarrow 3$ ) and blind equalizer convergence (preferring  $\kappa_s$  away from value 3).

## 4. TOPOLOGICAL VIEWPOINT

In this section we look at the characteristics of the CMA error surface which are invariant to source statistics. We recognize topology as the study of deformational invariants, and view correlation as inducing a deformation of the i.i.d. error surface. This section serves as a bridge between the distributional aspects of i.i.d. source considered in the previous section and temporally correlated sources to be discussed in following section.

 $\gamma$  Non-effect

In the case of an unknown constellation or source distribution *a priori* knowledge of the true value of  $\gamma$  is not possible, raising concern about proper selection of  $\gamma$ 's value. However, the choice of  $\gamma$ 's value is somewhat unimportant, in the sense that it only introduces a scaling effect on the achieved system parametrization (and hence equalizer output  $y$ ). The ability to reduce ISI is unaffected. Consider the effect of differing values of  $\gamma$  on the cost function. Here, the original CMA cost function,  $J(H)$  from (5) is compared with a cost function using a version of  $\gamma$  which has been scaled by a positive scalar  $\lambda$ ,  $J_\lambda(H) = \frac{1}{4}E\{((H^T S)^2 - \lambda\gamma)^2\}$ . Notice that,  $J_\lambda(H) = \lambda^2 J(\frac{H}{\sqrt{\lambda}})$ . Thus, the value of  $\gamma$  affects the error surface only by a scaling of the range and a dilation in the domain. No qualities of the error surface (e.g. number of minima, maxima, etc.) are changed. The 'shape' remains the same. Hence, the equalizer output  $y$  is merely scaled, ISI levels are unaffected.

 $\gamma$  matching manifold is  $n$ -sphere

There exists a closed, bounded manifold of the error surface to which CMA converges regardless of initialization. In a general sense, the algorithm's convergent behaviour can be described by characterization of this manifold which will be denoted as  $\mathcal{M}_\gamma$ . To develop this concept, note that along any ray extending from the origin in  $H$  space, there is one and only one minimum. This has been demonstrated for i.i.d. sources with equal probabilities<sup>25, 28</sup> but is extended here to include any distribution and temporal source correlation.

Consider the ray given by  $\lambda\bar{H}$  for some  $\bar{H} \in \mathbf{R}^{n+1}$  with  $\lambda \geq 0$ . Let  $\bar{y}$  be the output resulting from  $\bar{H}$  (i.e.  $\bar{y} = \bar{H}^T S$ ), and consider the output  $y$  resulting from any parametrization along the ray (i.e.  $y = \lambda\bar{y} = \lambda\bar{H}^T S$ ,  $\lambda \geq 0$ ). The CMA error surface in terms of  $\lambda$  may be written,  $J(\lambda) = 1/4E\{((\lambda\bar{y})^2 - \gamma)^2\}$ . Differentiating to find the extrema along this ray yields,  $\partial J(\lambda)/\partial \lambda = \lambda^3 E\{\bar{y}^4\} - \gamma\lambda E\{\bar{y}^2\}$ . There are only two solutions,  $\lambda = \{0, +\sqrt{\gamma E\{\bar{y}^2\}/E\{\bar{y}^4\}}\}$  with corresponding second derivative values of which are positive and negative, respectively. Hence, the origin is a local maximum, and the other stationary point is the only minimum along that ray, in agreement with References 25 and 28. Furthermore, it can be shown the output at any radial minimum has the property that its ratio of fourth to second moments matches  $\gamma$ . The union of all such radial minima is the manifold of interest and denoted as the ' $\gamma$  manifold', or  $\mathcal{M}_\gamma$ . This manifold is topologically equivalent to the sphere  $\mathcal{S}^n$ . The concept is that all stationary points necessarily lie on this manifold, since  $\mathcal{M}_\gamma$  is composed of all points having zero gradient in the radial direction, and stationary points must have zero gradient in *all* directions.

Benveniste<sup>6</sup> uses the notion of a manifold in his analysis of the Sato (and extensions thereof) under i.i.d. source. The manifold under consideration is the unit sphere. Stationary points of the vector field restricted to this unit sphere are identified. In such a formulation however, these are *not* stationary points of the algorithm itself, as there may exist a radial component orthogonal to the manifold. In contrast, herein we use the  $\gamma$  manifold which, while being topologically equivalent

to the sphere has the property that stationary points on this manifold are indeed stationary points of the CMA algorithm.

An important point is that the radial gradient is seen to be positive whenever  $E\{y^4\}/E\{y^2\} < \gamma$ , and negative whenever  $E\{y^4\}/E\{y^2\} > \gamma$ . Then, on average settings residing within the topological sphere have trajectories which increase in norm to approach  $\mathcal{M}_\gamma$ , and vice versa for settings outside this manifold. The convergence to the  $\gamma$ -matching manifold may be quite quick due to the quartic nature of the error surface in the radial direction. After reaching the neighborhood of the manifold it remains there and convergence is governed by the vector field existing on the manifold. It is this vector field that may be greatly influenced by source distributions and temporal correlation. In Reference 25 a presentation of the expanding/contracting radius is found for binary, zero-mean i.i.d. source. The results here hold for arbitrary distributions and correlations.

### *Symmetries*

Since every monomial in  $F(H)$  has an odd total degree, we have  $F(H) = \mathbf{0} \Leftrightarrow F(-H) = \mathbf{0}$ . This is due to the sign ambiguity caused by the  $y^2$  term in the cost function. However, the point is meant to be extended here beyond just a property at the stationary points. The property that  $F(H) = -F(H)$  is true everywhere in  $H$  space. One manifestation of this is that the error surface will always have an even number of minima, an even number of maxima, and an even number of saddle points (discounting the origin).

### *Connection to CMA effort: $\gamma$ manifold*

It has been pointed out that all CMA stationary points lie on the  $\gamma$  manifold. The vector field on the manifold will govern the asymptotic behaviour. In this manner, we consider only this manifold and view the CMA error surface as a vector field on this manifold.

Note that on  $\mathcal{M}_\gamma$ , the CMA gradient has no radial component (by definition of  $\mathcal{M}_\gamma$ ), and hence the CMA gradient is tangent to  $\mathcal{M}_\gamma$  at all points  $H \in \mathcal{M}_\gamma$ . Then, the CMA gradient suits the definition of a vector field on a manifold topologically equivalent to a sphere.

In a vector field, there are large regions which may be deemed 'uninteresting'. That is, uninteresting in the sense that they have little to do with the asymptotic behaviour of the gradient algorithm. For points not near a critical point, the vector field around  $x$  is nearly constant. Contrarily, the interesting parts of the vector field are those in the neighbourhood of a critical point. For this reason we focus on describing the vector field in the neighbourhood of critical points. Near critical points, a variety of behaviours are possible. CMA, being a gradient system (i.e. curl-free), precludes some of these, leaving the only critical points to be saddles, minima (sinks), or maxima (sources).

The manifold places certain restrictions on the relative number of minima, maxima, and saddles. A form of this type of restriction is often first seen in introductory graph theory, taking the form of 'Euler's formula' which relates the number of vertices, edges and faces of a polytope in three space. Morse theory<sup>29</sup> extends this concepts and identifies the constraints between critical points of vector fields on manifolds.

### *Generalized results for IID sources*

By applying similar stationary point Hessian analysis, for any i.i.d. source where  $\kappa_s \neq 3$ , we have that the radial direction is always attractive, and it is always the strongest attractive direction.

For any platykurtic i.i.d. sources, (i) the number of repulsive directions on the manifold equals  $N - 1$  for stationary points  $H \in \mathcal{C}_N$  and (ii) the number of attractive directions on the manifold equals  $n + 1 - N$  for stationary points  $H \in \mathcal{C}_N$ . Since, *all* of the non-radial Hessian eigenvalues for  $\kappa_s < 3$  have opposite signs when  $\kappa_s > 3$  we see that the vector field at critical points (on the manifold) is negative (reversed) for a leptokurtic source. Thus, the minima and maxima are exchanged, as are the number of repulsive and attractive directions for all stationary points.

## 5. TEMPORALLY CORRELATED SOURCES

In Reference 18, Bellini points out the lack of understanding of blind equalization schemes when the source is non-white. The global convergence to perfect equalization settings demonstrates a rather attractive property of CMA-FSEs meeting the length, diversity, and platykurtic i.i.d. source conditions. Next we relax the condition of the source ‘whiteness’ and investigate CMA stationary points under (a class of) temporally correlated sources.

### *No perfect equalization*

Let  $\vec{\mathcal{R}}$  denote the vector containing all such correlation coefficients appearing in (11) and consider the system  $F(H, \vec{\mathcal{R}})$ . Of interest is what happens to the result of global convergence to perfect equalization when a temporally correlated source is introduced. We first turn this result for i.i.d. sources around and ask, ‘for which source correlations do there exist perfect equalization parametrizations which are stationary points?’. Denote such perfect equalizing parametrizations by  $H^{k\star}$  (i.e. the global system parametrization which has all zero entries, except for a 1 (or  $-1$ ) in the  $k$ th position). Note that  $H^{k\star} \in \mathcal{C}_1$ . Evaluating  $F(H^{k\star}, \vec{\mathcal{R}})$  yields the conditions,

$$\mathcal{R}_i^0 - \gamma \mathcal{R}_i^0 = 0 \quad \text{for } 0 < i < n - 1 \quad (21)$$

Recall that for an i.i.d. source all correlation coefficients are zero except  $\mathcal{R}_i^0$  and  $\mathcal{R}_i^0$  when  $i = 0$ , which trivially meets the stated conditions for perfect equalization. However, rewriting (21) for clarity, we recognize how stringent these conditions may be for non-white sources.

$$h_k = \pm \sqrt{\frac{\gamma \mathcal{R}_0^0}{\mathcal{R}_0^0}} = \pm \sqrt{\frac{\gamma \mathcal{R}_1^0}{\mathcal{R}_1^0}} = \dots = \pm \sqrt{\frac{\gamma \mathcal{R}_n^0}{\mathcal{R}_n^0}} \quad (22)$$

This is a very tight constraint and appears to admit only ‘pathological’ source sequences (e.g.  $s(k)$  constant for *all*  $k$ , or  $s(k)$  alternating between some constant and its negative). In general, it may be said that non-white source correlations do not admit a perfect equalization setting to be a stationary point of CMA. Of course, the real question is ‘how far from perfect equalization settings are the resulting stable stationary points?’

### *Metrics of performance and correlation*

A metric of equalization performance and source correlation must be defined to properly describe the results regarding equalization under temporally correlated sources. In addition to the previously introduced performance metric of ISI ( $\varphi$ ) in (4), an appropriate metric of equalization performance

is the mean squared error (MSE) between source and equalizer output at a stable stationary channel–equalizer parametrization defined as

$$\text{MSE} = \min_{\delta} \{E\{(s(k - \delta) + y(k))^2\}, E\{(s(k - \delta) - y(k))^2\}\} \quad (23)$$

wherein the presence of both the sum and difference terms denotes a toleration of the sign ambiguity (i.e. assuming the source is differentially encoded). As we are motivated to show the effect of source correlation on achieved MSE, we define a metric on non-whiteness as

$$\mu_{\text{COR}}(\vec{\mathcal{H}}) = \|\vec{\mathcal{H}} - \vec{\mathcal{H}}_w\|_1 \quad (24)$$

where  $\vec{\mathcal{H}}_w$  is the vector  $\vec{\mathcal{H}}$  evaluated for an i.i.d. source with equally probable symbol values, and  $\|\cdot\|_1$  is the  $\ell_1$  vector norm. The choice of the  $\ell_1$  vector norm as opposed to the perhaps more prevalent  $\ell_2$  norm was somewhat arbitrary. (None of the qualitative results of the experiments to be presented are changed by use of the  $\ell_2$  norm.)

#### Numerical solution method

Direct algebraic solution for the roots of  $F(H, \vec{\mathcal{H}})$  is difficult in the general case, as  $F(H, \vec{\mathcal{H}})$  is a set of multivariate polynomial equations with many monomials. However, the numerical approach of continuation methods<sup>30, 31</sup> (or homotopy methods) may be applied to solve for roots of  $F(H, \vec{\mathcal{H}})$  given a *specific*  $\vec{\mathcal{H}}$ . In this method, one introduces the homotopy

$$\Phi(\lambda) = \lambda F(H, \vec{\mathcal{H}}) + c(1 - \lambda)G \quad (25)$$

consisting of the system of interest (here,  $F(H, \vec{\mathcal{H}})$ ), a system  $G$  with known roots, a random complex constant  $c$ , and a scalar  $0 \leq \lambda \leq 1$ . As  $\Phi(\lambda)$  is a system with one degree of freedom, the solutions breakup into paths almost everywhere. The desired roots of  $F(H, \vec{\mathcal{H}})$  may be found by ‘tracing’, in a predictor–corrector fashion, the roots of  $\Phi$  from the known roots of  $G$  to the roots of  $F(H, \vec{\mathcal{H}})$  by varying  $\lambda$  from 0 to 1.

Mathematically, there are three issues of concern. First, in order to perform such root tracing, we must be guaranteed that there are the same number of roots of systems  $F(H, \vec{\mathcal{H}})$  and  $G$ . We may construct  $G$  such that this is true and follows from Bernstein’s theorem<sup>32</sup> or by recognizing that our system  $F(H, \vec{\mathcal{H}})$  satisfies the Bezout upper bound. Furthermore, nastiness such as bifurcations (i.e. path crossings during roots tracings) must be ruled out. This is accomplished herein by the use of the random constant  $c$  which puts the systems  $F(H, \vec{\mathcal{H}})$  and  $G$  in relatively general position (see Reference 31). Specifically, we choose  $G$  to be the CMA-FSE system under independent source with equally probable symbol value, with the known  $3^{n+1}$  solutions derived earlier.

For all homotopy methods calculations we have used PELICAN.\* Also, to differentiate effects of correlated sources investigated in this section from the effects of source distributions considered earlier, all source sequences used here (periodic and Markov) have an equiprobable symbol occurrence.

\* PELICAN is a software package for polynomial continuation methods under development by Birk Huber at the Dept. of Mathematics, Cornell University. Documentation and release information available at <http://math.cornell.edu/~birk>.

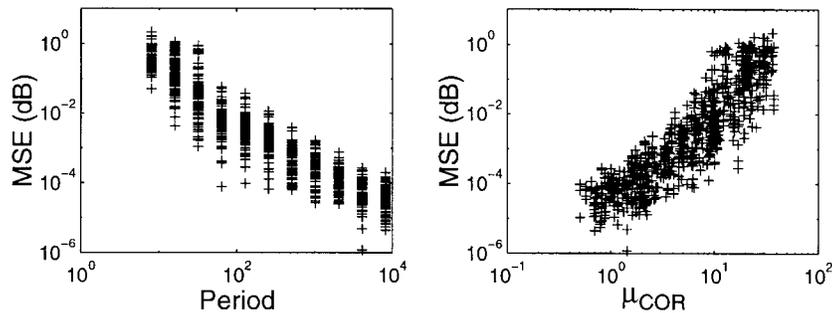
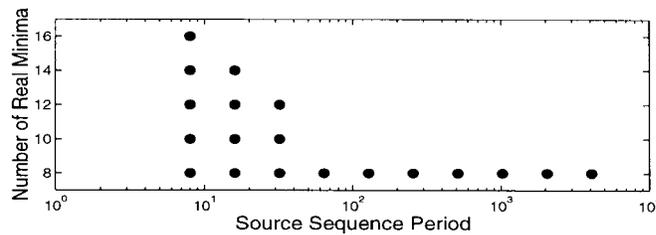


Figure 7. Periodic source: MSE vs. period and MSE vs. correlation

Figure 8. Number of minima under source of period  $P$ 

### Experiment I—Periodic Sources

Using continuation methods, the stationary points for CMA under periodic sources may be computed. To gain some insight into how the MSE at stable stationary points may be effected as the period increases, a Monte Carlo computation was performed. For each period ( $P$ ) of length 32, 64,  $\dots$ , 8192, twenty source sequences with equal symbol occurrence were randomly drawn from a 4-PAM source. For each sequence, *all* real stationary combined channel–equalizer parametrizations corresponding to a 4 tap (half-baud spaced) channel and 4 tap (half-baud spaced) equalizer when  $L=2$  were computed.

The resulting  $\mu_{\text{COR}}$  and MSE for the subset of stable, strictly real, stationary parametrizations resulting from these computations was also computed. Results relating MSE and  $\mu_{\text{COR}}$  are plotted in Figure 7 which exhibits a general trend of MSE with increasing correlation. Also plotted is MSE versus the sequence period ( $P$ ) where as the period increases, MSE tends to decrease.

A wide range of equalization performance (about 20 dB MSE) was observed for a given period (seen from Figure 7). Some sequences of a given period deform the error surface stationary points away from perfect equalization parametrizations more than others. However, there is another more significant point that may be culled from the continuation-method computations. For each of the 220 source sequences considered, the number of (possibly complex) stable stationary solutions to  $F(H, \vec{\mathcal{R}})$  due to a particular sequence of period  $P$  was computed. Restricting our attention to the strictly real (i.e. achievable) parametrizations results in Figure 8. For sequences of short period, the number of minima may be double of that of sequences of long periods (as well as white). In the white source case the number of minima is  $2^{n+1}$ . For  $n+1=3$  we expect eight minima.

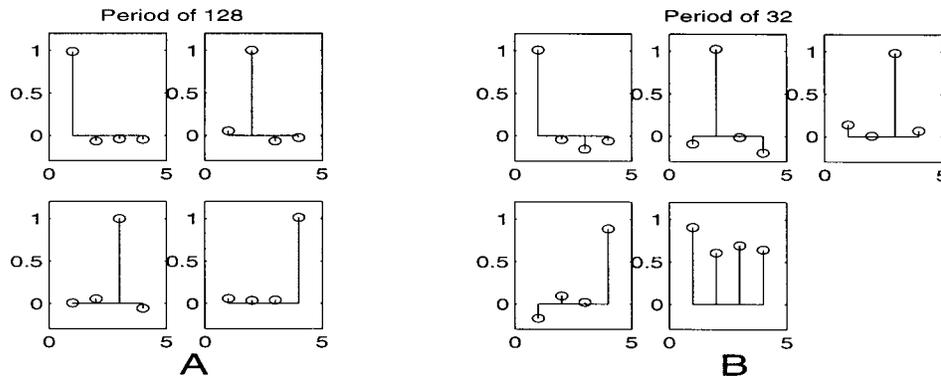


Figure 9. Period of 128, 8 minima with Low ISI

However, under source correlation we see that as many as 16 minima may exist. Such an increase in the number of minima may be alarming. The error surface has *not* merely deformed in such a way that results in only slight perturbations in the locations of minima. To the contrary, *new* minima have been created. The results herein for the 4 tap case shows only increases of number of minima under source correlation, it is also possible for the number of minima to decrease as well.<sup>33</sup>

Consider Figure 9A which shows all eight stable parametrizations depicted for a source sequence period of 128 (four parametrizations are drawn explicitly, the other four are the negatives of those shown). All parametrizations have one significant tap and three taps weights of small magnitude correspond to relatively low ISI settings.

Contrast this to Figure 9B in which a sequence of period 32 yields *ten* stable stationary points. Eight are fairly close to a good equalization setting (may be considered a displacement from perfect equalization by correlation induced deformation), the other two (the one shown and its negative) have high ISI settings (may be considered to be ‘new’ minima created by error surface deformation).

### Experiment II—Markov Sources

Next, a Markov model which allows the modelling of a white source as well as temporally correlated sources is investigated. Since convex combinations of Markov transition matrices are themselves valid transition matrices, we can produce a source model which allows smoothly varying  $\bar{\mathcal{R}}$ . These  $\bar{\mathcal{R}}$  may be calculated directly from the transition matrix  $\Xi$ .

Consider a Markov state transition matrix which generates a white source,  $\Xi_w$ , (in which each state represents an alphabet member) which generates a source with equally probable symbol values (i.e. each matrix element equals  $1/M$  for an  $M$ -ary source). Next, we choose a transition matrix,  $\Xi_c$  which corresponds to a source correlation of interest and redefine  $\Phi$  to be  $\Phi(\lambda) = F(H, \lambda \Xi_c + (1 - \lambda) \Xi_w)$  which yields a system wherein performing continuation methods yields a data point of interest at each evaluation of  $\lambda$ .

An example for the case  $\Xi_c$  as described in Appendix II demonstrates the stationary point movement from  $H = [0 \ 0 \ 1 \ 0 \ 0]^T$  for a combined channel–equalizer having a 5 baud span under increasing source correlation. Figure 10 plots the stationary point trajectory as  $\lambda$  in (10) is

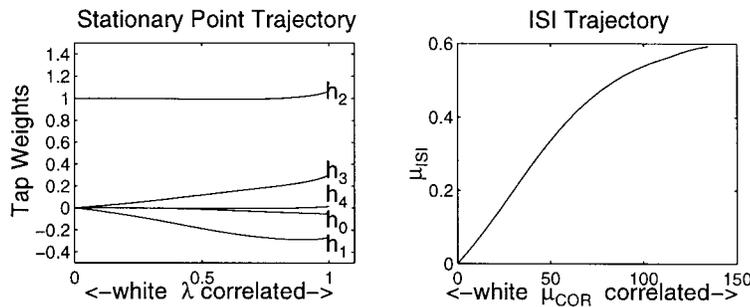


Figure 10. Markov source stationary point and ISI trajectories

varied from 0 to 1 and the associated Hessian of the cost function was checked at each iteration to verify whether the stationary point remained a stable one.

## 6. CONCLUSION

This work has investigated the CMA under the equalizability condition which allows examination in the combined channel–equalizer space using an equalizer of finite order. Behaviour of FSE-CMA is understood through characterization of the CMA error surface. A main point of interest concerning the error surface is the location of stationary points, as the stable stationary points influence the asymptotic behaviour, while the saddles influence the transient behaviour.

While global convergence for an i.i.d. source with samples drawn with equal probability from a finite alphabet is not a new (nor surprising) result, the phrasing of the problem as a system of multivariate polynomial equations leads to powerful extensions. We are able to quantitatively address the effects of various source distributions on the error surface. The *kurtosis* of the source (and its relation to constellation shaping) is identified as the important quantity which influences convergent points (shown analytically) and convergence speed (demonstrated through analysis and numerical examples). Identification of *all* the stationary point locations, as well as their categorization is stretched to include enumeration of the number of positive and negative eigenvalues of their Hessian as well as the associated directions (eigenvectors) describing the relative directional flow of the vector field near the stationary point.

The results of the i.i.d. source case are combined with some simple properties of the error function to identify the concept of the  $\gamma$  manifold. Further geometric structure is introduced relating the vector field stationary points on the  $\gamma$  manifold to Morse theory. The work on i.i.d. sources represents a mixture of refining known results and moderate extensions thereof. The true strength perhaps lies in the unified presentation allowed by the problem formulation.

The more novel contributions are found in the later sections. Herein, the solution to the polynomial system identifying the stationary points is considered for the temporally correlated source. The avenue pursued incorporates a numerical approach. Homotopy methods are introduced as a method of solving for stationary points for *given* source statistics. Monte Carlo computations are run for a class of correlated sources. The results are presented and discussed in terms of convergent ISI and MSE. An idea of the sensitivity to source correlation is gleaned.

It has been demonstrated that correlated sources have the ability to do more than merely perturb error surface minima. They allow the creation of new (poorly performing) minima. This places added emphasis on the source independence assumption.

APPENDIX I

To determine the eigenvalues of  $\widehat{\mathbb{M}}$  note that similarity transformations do not change the eigenvalues of a matrix. Scaling these by the common factor in (18) yields the eigenvalues of  $\mathbb{M}$ . Form the transformation matrix  $\mathbb{P}$  as  $\mathbb{P}(m, \mathcal{S}\mathcal{O}_m) = \text{sign}(h_{\mathcal{S}\mathcal{O}_m})$  and note that  $\mathbb{P}$  is unitary ( $\mathbb{P}\mathbb{P}^T = \mathbb{I}$ ).

A quick example for the case  $H^T = \sqrt{\frac{\kappa_s}{\kappa_s+6}} [0 \ 1 \ 1 \ 0 \ -1 \ 0]$  follows. The index list is  $\mathcal{S}\mathcal{O} = \{1, 2, 4, 0, 3, 5\}$ . We then write,  $\mathbb{B} = \mathbb{P}^{-1}\widehat{\mathbb{M}}\mathbb{P}$  as

$$\begin{bmatrix} 2\kappa_s & 6 & 6 & & & \\ 6 & 2\kappa_s & 6 & & & \\ 6 & 6 & 2\kappa_s & & & \\ & & & 3-\kappa_s & 0 & 0 \\ & & & 0 & 3-\kappa_s & 0 \\ & & & 0 & 0 & 3-\kappa_s \end{bmatrix} = \mathbb{P}^{-1} \begin{bmatrix} 3-\kappa_s & 0 & 0 & 0 & 0 & 0 \\ 0 & 2\kappa_s & 6 & 0 & -6 & 0 \\ 0 & 6 & 2\kappa_s & 0 & -6 & 0 \\ 0 & 0 & 0 & 3-\kappa_s & 0 & 0 \\ 0 & -6 & -6 & 0 & 2\kappa_s & 0 \\ 0 & 0 & 0 & 0 & 0 & 3-\kappa_s \end{bmatrix} \mathbb{P}$$

where

$$\mathbb{P} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{26}$$

For  $H \in \mathcal{C}_N$  the similarity transformation of  $\widehat{\mathbb{M}}$  into  $\mathbb{B}$  yields a block-form matrix, wherein one block is  $(3 - \kappa_s)\mathbb{I}$  of dimension equal to the number of  $h_m$  which are zero. The other block is an  $N \times N$  circulant matrix with strictly positive elements. Such a structure allows easy identification of the eigenvalues. Interestingly, the signs of the eigenvalues of both sub-matrices are determined by the source kurtosis  $\kappa_s$ .

*Diagonal sub-matrix:* The diagonal sub-matrix eigenvalues are  $3 - \kappa_s$ . For  $\kappa_s < 3$ , this sub-matrix is positive definite. However, for  $\kappa_s > 3$  we have a negative definite sub-matrix. For the case  $\kappa_s = 3$ , this sub-matrix becomes singular with all elements zero.

*Circulant sub-matrix:* The circulant block's eigenvalues are identified by the discrete Fourier transform (DFT) of a column (a well-known property of circulant matrices, see Reference 34). Thus,

$$\begin{aligned} \text{Upper block eigenvalues} &= \mathcal{DFT}([2\kappa_s \ 6 \ \dots \ 6]) \\ &= \mathcal{DFT}([2\kappa_s - 6 \ 0 \ \dots \ 0]) + \mathcal{DFT}([6 \ 6 \ \dots \ 6]) \\ &= [2\kappa_s - 6 \ 2\kappa_s - 6 \ \dots \ 2\kappa_s - 6] + [6N \ 0 \ \dots \ 0] \\ &= [2\kappa_s + 6(N - 1) \ 2\kappa_s - 6 \ \dots \ 2\kappa_s - 6] \end{aligned} \tag{27}$$

where  $\mathcal{DFT}(\cdot)$  represents the discrete Fourier operator.

APPENDIX II

The states ( $\mathcal{S}$ ) of the Markov process are identified with a particular alphabet member, and the transition from present state to next state is associated with the Markov transition matrix  $\Xi$ .

### Markov source for Example C

For the sample Markov source model used in Example C, we have the state vector

$$\mathcal{S} = [-7\Delta \quad -5\Delta \quad -3\Delta \quad -1\Delta \quad +1\Delta \quad +3\Delta \quad +5\Delta \quad +7\Delta]^T$$

and  $\Xi$  is the circulant matrix having the first column as  $[0 \ 0.025 \ 0.025 \ 0 \ 0.025 \ 0.3 \ 0.25 \ 0.6]^T$ , with 0.6 becoming the first element of the second column, etc. The  $\Xi(i, j)$  elements give the probabilities of transitioning from state  $i$  to state  $j$ . Notice by the symmetry of  $\Xi$ , we have equiprobability of symbol occurrence. This can be shown formally by recognizing that the vector  $[\frac{1}{8} \ \frac{1}{8} \ \frac{1}{8} \ \frac{1}{8} \ \frac{1}{8} \ \frac{1}{8} \ \frac{1}{8} \ \frac{1}{8}]^T$  is an eigenvector of  $\Xi$  whose elements are strictly positive and sum to one. Hence, it is the valid steady-state probability vector.

### Markov source for continuation methods

In correlated source experiment 2, a Markov source was used with state transition matrix  $\Xi_c$  given by

$$\Xi_c = \begin{bmatrix} 0 & 0.6 & 0.2 & 0.2 \\ 0 & 0 & 0.2 & 0.8 \\ 0.8 & 0.2 & 0 & 0 \\ 0.2 & 0.2 & 0.6 & 0 \end{bmatrix}$$

### ACKNOWLEDGEMENTS

This work was partially supported by NSF Grant MIP-9509011.

### REFERENCES

- Bellini, S., 'Blind equalization', *Alta Frequenza*, **LVII**, 445–450 (1988).
- Qureshi, S. U. H., 'Adaptive equalization', *Proc. IEEE*, **73**, 1349–1387 (1985).
- Treichler, J. R. and M. G. Agee, 'A new approach to multipath correction of constant modulus signals', *IEEE Trans. Acoust. Speech Signal Process.*, **xx**, 459–472 (April 1983).
- Godard, D. N., 'Self-recovering equalization and carrier tracking in two-dimensional data communication systems', *IEEE Trans. Commun.*, **COM-28**, 1867–1875 (1980).
- Foschini, G. J., 'Equalization without altering or detecting data', *AT&T Tech. J.*, **64**, 1885–1911 (1985).
- Benveniste, A., M. Goursat and G. Ruget, 'Robust identification of a nonminimum phase system: blind adjustment of a linear equalizer in data communications', *IEEE Trans. Automat. Control*, **AC-25**, 385–399 (1980).
- Tong, L., G. Xu and T. Kailath, 'Blind identification and equalization based on second-order statistics: a time domain approach', *IEEE Trans. Inform Theory*, 340–349 (1994).
- Fijalkow, I., F. Lopez de Victoria and C. R. Johnson Jr., 'Adaptive fractionally spaced blind CMA equalization', in: *Proc. IEEE DSP Workshop*, 1994, pp. 257–260.
- Li, Y. and Z. Ding, 'Global convergence of fractionally spaced godard adaptive equalizers', *IEEE Trans. Signal Process.*, 818–829 (April 1996).
- Afkhamie, K. H. and Z. Q. Luo, 'Blind equalization using second-order statistics', *Proc. IEEE Conf. on Acoust. Speech Signal Process.*, May 1995, pp. 1053–1056.
- Chan, C. K. and J. J. Shynk, 'Stationary points of the constant modulus algorithm for real Gaussian signals', *IEEE Trans. Acoust. Speech Signal Process.*, **38**, 2176–2180 (1990).
- Hua, Y., H. Yang and W. Qiu, 'Source correlation compensation for blind channel identification based on second-order statistics', *IEEE Signal Process. Lett.*, **1**, 119–120 (1994).
- Johnson, C. R. Jr., J. P. LeBlanc and V. Krishnamurthy, 'Godard blind equalizer misbehavior with correlated sources: two examples', *Moroccan J. Control Comput. Sci. Signal Process.*, **II** (3), 1–39 (1993).

14. Treichler, J. R., V. Wolff and C. R. Johnson Jr., 'Observed misconvergence in the constant modulus adaptive algorithm', *Proc. 25th Asilomar Conf. Signals, Systems and Computers*, 1991, pp. 663–667.
15. Zervas, E., J. G. Proakis and V. Eyuboglu, 'Effects of constellation shaping on blind equalization', *SPIE*, **1565**, 178–187 (1991).
16. Gitlin, R. D. and S. B. Weinstein, 'Fractionally-spaced equalization: an improved digital transversal equalizer', *Bell System Tech. J.*, **60**, 275–296 (1981).
17. Ungerboeck, G., 'Fractional tap-spacing equalizer and consequences for clock recovery in data modems', *IEEE Trans. Commun.*, **COM-24**, 856–864 (1976).
18. Haykin, S., *Blind Deconvolution*, Prentice-Hall, Englewood Cliffs, NJ, 1994.
19. Gitlin, R. D., J. F. Hayes and S. B. Weinstein, *Data communications principles*, Plenum Press, New York, 1992.
20. Moulines, E., P. Duhamel, J. Cardoso and S. Mayrargue, 'Subspace methods for the blind identification of multichannel FIR filters', *IEEE Trans. Signal Process.*, xx, 516–525 (Feb. 1995).
21. Fijalkow, I., A. Touzni and J. R. Treichler, 'Fractionally spaced equalization using CMA: Robustness to channel noise and lack of disparity', *IEEE Trans. Signal Process.*, xx, 56–66 (January 1997).
22. Bulmer, M. G., *Principles of Statistics*, Dover Publications, New York, 1967.
23. Wiggins, R. A., 'Minimum Entropy Deconvolution', *Geoexploration*, **16**, 21–35 (xxxx).
24. Donoho, D. L., 'On minimum entropy deconvolution', in: D. F. Findley (Ed.), *Applied Time Series Analysis*, Academic Press, New York, 1981, pp. 565–608.
25. Johnson, C. R. Jr. and B. D. O. Anderson, 'Godard blind equalizer error surface characteristics: white, zero-mean, binary source case', *Int. J. Adapt. Control Signal Process.*, **9**, 301–324 (1995).
26. Laroia, R., N. Farvardin and S. A. Tretter, 'On optimal shaping of multidimensional constellations', *IEEE Trans. Inform. Theory*, **40**, 1044–1056 (1994).
27. LeBlanc, J. P. and S. W. McLaughlin, 'Non-equiprobable constellation shaping and blind constant modulus algorithm equalization', *Proc. 30th Conf. on Information Sciences and Systems, Princeton, NJ*, March 1996, pp. 901–903.
28. Ding, Z. and R. A. Kennedy, 'On the whereabouts of local minima for blind adaptive equalizers', *IEEE Trans. Circuits and Systems—II: Analog and Digital Signal Processing*, **39**, 119–223 (1992).
29. Milnor, J., *Morse Theory*, Princeton University Press, Princeton, NJ, 1963.
30. Allgower, E. L., *Numerical Continuation Methods*, Springer, Berlin, 1990.
31. Morgan, A. P., *Solving Polynomial Systems using Continuation for Engineering and Scientific Problems*, Prentice-Hall, Englewood Cliffs, NJ, 1987.
32. Bernstein, D. N., 'The number of roots of a system of equations', *Funct. Anal. Appl.* **9**, 1-4XS (1975).
33. LeBlanc, J. P., K. Doğançay, R. A. Kennedy and C. R. Johnson Jr., 'Effects of input data correlation on the convergence of blind adaptive equalizers', *Proc. IEEE Conf. on Acoustics, Speech and Signal Processing*, May 1994, pp. III-313–317.
34. Willsky, A. S., *Digital Signal Processing and Control and Estimation Theory: Points of Tangency, Areas of Intersection, and Parallel Directions*, MIT Press, Cambridge, MA, 1979.