Global CMA Error Surface Characteristics, Source Statistic Effects: Polytopes and Manifolds

James P. LeBlanc \textsuperscript{*} \\
leblanc@nmsu.edu \\
Klipsch School of ECE \\
New Mexico State University, Las Cruces, NM USA \\
C. Richard Johnson, Jr. \textsuperscript{*} \\
johnson@ee.cornell.edu \\
School of Electrical Engineering \\
Cornell University, Ithaca, NY USA

Abstract: This work focuses on gaining insight to CMA behavior through identification of global properties of CMA stationary point locations due to source statistics (distributions and temporal correlation). The CMA error function under source correlation effects is viewed as a deformation of the iid CMA error surface. As deformations are the realm of topology, we look at some of the topological aspects connected with CMA source correlation. Some general characteristics of CMA under iid source are presented, noting relations to Morse Theory, polytopes, and Euler Characteristics.

1. Introduction

Blind equalization is a form of signal estimation used in digital communications where the transmitted signal has been corrupted by inter-symbol interference (ISI) through an unknown channel. The notation used throughout and reflected in Figure 1 uses a baseband model. Here s(k) denotes the elements of the zero mean source sequence at time k. The samples of the received waveform are denoted as r(k), and the equalizer output as y(k). The channel impulse response is denoted by C, the adaptive equalizer parametrization at time k by D(k), and the combined channel-equalizer response is denoted by H(k).

Figure 1. Blind Equalization Block Diagram

Let the (n+1) \times 1 column vectors R(k) and D(k) represent the equalizer regressor and equalizer tap weights at time k respectively,

\[ R(k) = [r(k) \ r(k-1) \ \ldots \ r(k-n)]^T \]

\[ D(k) = [d_0(k) \ d_1(k) \ \ldots \ d_n(k)]^T. \]

The equalizer output at time k is then given by y(k) = R^T(k)D(k).

In blind equalization schemes (as opposed to LMS), the “error” signal \( e(k) = y(k) - s(k) \) is not available. Hence, a signal which is strictly a function of the equalizer output must be used to act as an error signal. This is typically represented by \( \Psi(y) \) and called the error function. The tap update for such blind equalizers is analogous to the LMS update replacing \( e(k) \) with \( \Psi(y(k)) \) yielding,

\[ D(k+1) = D(k) + \mu \Psi(y(k))R(k) \]

where \( \mu \) is the adaptation step-size. We limit our discussion here to the Constant Modulus Algorithm (CMA) [1] wherein,

\[ \Psi_{\text{CMA}} = \gamma |y|^2 - \gamma \]

where \( \gamma = E\{|s|^2\}/E\{|s|^4\} \) is a positive real constant. This algorithm may be viewed as a gradient descent algorithm in which an instantaneous estimate of the gradient of some cost function is used in place of the actual gradient. The CMA cost function associated with \( \Psi_{\text{CMA}} \) is

\[ J_{\text{CMA}} = \frac{1}{4} E \left\{ |(y|^2 - \gamma)^2 \right\}. \]

The modulus operator in \( J_{\text{CMA}} \) allows any phase shift of y to have the an identical cost. This phase ambiguity may be overcome by the use of a differential encoding scheme.

Many equalizers built today use “fractional spacing” which may be described by a sub-channel model, in which all sub-blocks are baud-spaced discrete-time systems. Such fractional sampling in a digital communications setting has the ability to “perfectly equalize” a moving average channel with a finite order equalizer, under the conditions: the time-span of equalizer exceeds the time-span of the channel, and not all the sub-channels (as polynomials in z) share a common root [2].

With the additional conditions of independent, uniformly distributed source, and lack of channel noise, then the fractional equalizer adapted under CMA is known to be globally convergent to a no ISI setting, [3]. Recent work may be found regarding CMA’s behavior when not meeting the noiseless condition [4], and sub-channel disparity [5].

Herein we investigate the effect of relaxing the assumptions on distributions and temporal correlations of source sequence on CMA. The interest in distributional effects of iid sources is related to often used constellation shaping which can have a remarked affect on the convergence of CMA [6]. The motivation to study source correlation comes from documented evidence of CMA failure when the source sequence is drawn from a variety of cyclostationary, periodic, and non-white inputs. [7].

2. Mathematical Framework

Behavior of CMA can be understood through a study of the error surface. Such analysis of adaptive algorithms and their convergent properties is often pursued through the investigation of stationary points and their Hessian evaluation. Although CMA is often used for two-dimensional signal constellations, we limit our discussion to the use of strictly real alphabets (e.g. M-PAM) to maintain differentiability of the cost function and keep notation simple. (Strictly speaking the absolute value operator in (5) does not meet the Cauchy-Riemann criterion and hence is not differentiable.)

The system of equations which defines CMA’s stationary points is described in [8] and denoted as \( F(H) = [f_0(H) \ \ldots \ f_n(H)]^T = [0 \ \ldots \ 0]^T \) where \( H = [h_0 \ \ldots \ h_n]^T \) is the \( n^\text{th} \) order baud-space combined channel-equalizer response, and each \( f_i(H) \)

\[ \sum_i \sum_{j \neq i} R^i_j h^j_i + 3 \sum_i \sum_{j \neq i} R^i_j h^j_i h^j_i + \sum_i \sum_{j \neq k \neq i} R^i_j h^j_i h^j_k - \gamma \sum_i R^i_i h^i_i \]

where the moments of the symmetrically distributed source \( E\{s\} = E\{s^2\} = 0 \) are denoted as

\[ R^i_i = E\{s(m-i)\} = E\{s(m-j)\} = E\{s(m-k)\} \]

(7)

For an \( n+1 \) tap system, this yields a set of \( n+1 \) equations in the \( n+1 \) unknowns \( \{h_m, 0 \leq m \leq n\} \) of each of the stationary points (solutions of \( F(H) \)) may then be categorized depending on the eigenvalues of the Hessian matrix \( M(\{s\}) = \partial^2 f/\partial h_j \) evaluated at a stationary point. For a local minima (maxima), \( M \) having both positive and negative eigenvalues denotes a saddle point. Since every nonmonial has odd total degree, we have \( F(-H) = -F(H) \). Thus, the error surface will always have an even number of minima, an even number of saddle points and an even number of maxima (discounting the origin).
3. TOPOLOGICAL VIEWPOINT

“If your head were a doughnut, you could comb your hair without leaving a bald spot.” [9]

Before actually solving \( F(H) = 0 \) (which is a function of the source statistics), we make some observations on the CMA error surface which are source independent.

3.1. \( \gamma \) Non-Effect

In the case of an unknown constellation, or source distribution, there may be some concern as to the improper selection of this constant. However, \( \gamma \)'s value is somewhat unimportant, in that it only introduces a scaling effect on the achieved system parametrization (and hence equalizer output \( y \)), the ability to reduce ISI is unaffected.

Compare the original CMA cost function, \( J_{\text{cma}}(H) \) with one using a \( \gamma \) value which has been scaled by a positive scalar \( \beta \),

\[
J_{\beta}(y) = \frac{1}{4} E \left\{ (y^2 - \beta \gamma)^2 \right\}.
\]

Notice that, \( J_{\beta}(y) = \beta^2 J(y/\sqrt{\beta}) \). Changing \( \gamma \) changes 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, ISI levels are unaffected.

3.2. \( \gamma \) Matching Manifold is equivalent to \( \alpha \)-Sphere

There exists a closed, bounded manifold of the error surface to which CMA converges regardless of initialization or source statistics. Note that along any ray extending from the origin in \( H \) space, there is one and only one minimum. This has been demonstrated for uniformly iid sources ([10]) but is extended here to allow any distribution and temporal source correlation.

Consider the manifold given by \( \alpha \tilde{H} \) for some \( \tilde{H} \in \mathbb{R}^{n+1} \) with \( \alpha \geq 0 \). Let \( \bar{y} \) be the output resulting from \( \tilde{H} \), and consider the output \( y \) resulting from any parametrization along the ray (i.e. \( y = \alpha \bar{y}, \alpha \geq 0 \)). The CMA error surface in terms of \( \alpha \) may be written,

\[
J(\alpha) = \frac{1}{4} (\alpha^4 E \left\{ \bar{y}^2 \right\} - 2 \alpha^2 E \left\{ \bar{y}^2 \right\}^2 + \gamma^2).
\]

Differentiating along ray yields to find the extrema only two solutions, \( \alpha = 0, +\sqrt{\gamma E \left\{ \bar{y}^2 \right\}/E \left\{ \bar{y}^2 \right\}} \) with corresponding second derivative values of, \(-\gamma E \left\{ \bar{y}^2 \right\} + 2 \gamma^2 E \left\{ \bar{y}^2 \right\}\). The origin is a local maximum, and the other stationary point is the only minimum along that ray.

It can be shown that the equalizer output at this radial minimum has the property that its ratio of fourth to second moments match \( \gamma \). The union of all such radial minima is the manifold of interest and denoted the “\( \gamma \) matching manifold”, or \( \mathcal{M}_\gamma \). This manifold is topologically equivalent to the sphere \( S^4 \in \mathbb{R}^{n+1} \) (loosely speaking, it is a distorted sphere). All stationary points (except the one at the origin) necessarily lie on this manifold having zero gradient in the radial direction. That is, since stationary points have zero gradient in all directions, they must also be zero in the radial direction.

Note that the radial gradient is positive whenever \( E \left\{ \bar{y}^4 \right\}/E \left\{ \bar{y}^2 \right\} < \gamma \). So, the average update term has a radially expanding component. Initial settings residing within the topological sphere have trajectories which increase in norm to approach \( \mathcal{M}_\gamma \), and vice versa for settings outside \( \mathcal{M}_\gamma \). The convergence to \( \mathcal{M}_\gamma \) 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 gradient (or vector field) existing on \( \mathcal{M}_\gamma \). In this manner, we consider only this manifold and view this portion of the CMA error surface as a vector field. Since the CMA gradient has no radial component on \( \mathcal{M}_\gamma \), the gradient is tangent to \( \mathcal{M}_\gamma \) at all points \( H \in \mathcal{M}_\gamma \), suiting the definition of a vector field on a manifold topologically equivalent to a sphere.

Benveniste ([11]) uses the notion of a manifold and considers the unit sphere. Stationary points of the vector field restricted to this unit sphere are identified. Note, these are not stationary points of the algorithm itself (they may have a radial component). In contrast we have \( \mathcal{M}_\gamma \) which, while being topologically equivalent to the sphere, includes actual CMA stationary points.

3.3. Manifold Constraints

The existence of a manifold on which all non-zero stationary points must lie, coupled with the fact that CMA is a gradient algorithm, places constraints on the relation of stationary points due to the topology of the manifold. The seemingly unrelated quote about doughnut shaped heads has relevance here. The connection to CMA may be realized by:

- Relating “hair combing of doughnut head” to assigning vector field on torus.
- Recognize that the structure of the torus allows a vector field with no critical points (Figure 2a).
- Note that the structure of the sphere disallows a vector field with no critical points (Figure 2b).

![Figure 2. Differing Fields on Torus and Sphere](image)

This demonstrates a qualitative difference between the torus and the sphere. Moreover this difference remains even if these objects are bent or twisted (it is the “structure” of the object, not the “shape”). Topology is the study and identification of such qualities which remain unchanged under deformations.

All non-zero critical points of CMA lie on \( \mathcal{M}_\gamma \) whose topological structure places constraints on CMA behavior by limiting the combinations of types of stationary points (in the same manner that the torus and sphere have differing constraints). Source correlation can be viewed as imparting a deformation of the error surface by changing the correlation coefficients of (6). In this manner, topology yields some description of what may or may not happen under source correlation.

In a vector field, there are large regions which may be deemed “uninteresting” in the sense that have little to do with the asymptotic behavior of the gradient algorithm. Away from critical points, the field is nearly constant as in Figure 3a. Contrarily, the interesting parts of the vector field are those in the neighborhood of a critical point where a variety of behaviors are possible. CMA, being a gradient system (i.e. curl-free), precludes some of these. The only critical points are saddles, minima (sinks), or maxima (sources) as in Figure 3b (adapted from [9]).

![Figure 3. Vector Field Away From and Near Critical Points](image)

3.4. \( \mathcal{M}_\gamma \) for IID Sources

The system of equations (6) defining the stationary points is solved along with a stationary point analysis for iid sources of varying source distributions in [12]. A useful quantity that appears repeatedly in this analysis, is the dimensionless quantity known as kurtosis defined as...
\( \kappa_x = E\{x^4\}/E\{x^2\}^2 \) for zero mean \( x^1 \). The Gaussian distribution has \( \kappa = 3 \) and is regarded as forming the boundary between platykurtic (\( \kappa_x < 3 \)) distributions and leptokurtic (\( \kappa_x > 3 \)) distributions.

When \( \kappa_x \neq 3 \) each non-zero element in \( H \) for \( F(H) = 0 \) (call this element \( h_m \), has the same magnitude. Specifically,

\[
h_m = \left\{ \begin{array}{ll}
0 & 0 \leq m \leq n \quad (10)
\end{array} \right.
\]

where \( N \) is the number of non-zero elements of \( H \).

For any platykurtic iid sources,

- The number of repulsive directions on the manifold is \( N - 1 \), for stationary point \( H \) having \( N \) non-zero elements.
- The number of attractive directions on the manifold is \( n + 1 - N \), for stationary point \( H \) having \( N \) non-zero elements.

Since, all of the non-radial Hessian eigenvalues for \( \kappa_x < 3 \) have opposite signs when \( \kappa_x > 3 \) we see that the vector field at critical points (on the manifold) is the negative (reversed) for a leptokurtic source. Thus, the minima and maxima are exchanged.

For both platykurtic and leptokurtic sources the CMA error surface minima coincide with minimum kurtosis of equalizer output. In the platykurtic source case kurtosis reduction realizes ISI reduction, while for leptokurtic sources this yields ISI enhancement.

For example, the results of CMA stationary point analysis for \( H \in \mathbb{R}^n \) under a platykurtic source are shown in Figure 4a. The minima are denoted by a “+”. The saddle points with two non-zero elements are denoted by a “0”, and those with three non-zero elements are denoted by a “−”.

![Figure 4. Manifold of CMA Error Surface](image)

Arrows have been painted on the “sphere” in Figure 4b to represent the CMA flow lines (i.e. vector field). In typical blind equalization settings the combined channel-equalizer order greatly exceeds three and the \( \gamma \) manifold may be of high dimensions which makes visualization difficult.

### 5. Euler Characteristic

An important topological invariant on compact manifolds is the Euler Characteristic. This is perhaps the most elementary of topological invariants and may be defined in a number of ways (some rather abstract). However, most of us are introduced to a simpler form of this concept in the setting of polyhedra in 3 space by the name of the Euler Equation written as \( 2 = V - E + F \). Where \( V \) is the \# of vertices, \( E \) is the \# of edges, \( F \) is the \# of faces of a polyhedron. But let’s expand on this, allowing a bit more abstraction.

Polyhedra in three space may be deformed to lie on a sphere (\( S^2 \)). Furthermore, the “2” in the Euler equation may actually regarded as referring to the Euler Characteristic of the space on which these polyhedra lie (i.e. \( S^2 \)). Furthermore, the vertices \( V \) may be viewed as faces of dimension 0, edges \( E \) are then faces of dimension 1, and the faces \( F \) become faces of dimension 2. With this slightly more abstract form framework, the Euler Characteristic for “polyhedra” in \( n + 1 \) space is defined as the alternating sum of the faces of various dimension,

\[
\chi(S^n) = \sum_{i=0}^{n} (-1)^i |F_i|.
\]

where \( |F_i| \) denotes the number of faces of dimension \( i \). Interestingly, \( \chi \) is a function of the manifold, not the polyhedron. The Euler Characteristic of the sphere \( S^n \) is found in most topology texts as:

\[
\chi(S^n) = \left\{ \begin{array}{ll}
0 & n \text{ odd} \\
2 & n \text{ even}
\end{array} \right.
\]

Admittedly, these polyhedra and their Euler Characteristic seem rather far removed from our interest in the vector field on \( M_n \). We address this in the next section which bridges these concepts of to vector fields on manifolds.

#### 3.6. Morse Theory

Morse Theory deals with smooth mappings on manifolds detailing how functions with no degenerate critical points can yield insight into the topology of the manifold of their domain and vice versa. A basic tenet of Morse theory is that near a nondegenerate critical point, a function can be “completely” determined by the Hessian [9]:

**Morse Lemma:** Suppose the point \( a \in \mathbb{R}^k \) is a nondegenerate critical point of the function \( f \), and

\[
(H_{ij}) = \frac{\partial^2 f}{\partial x_i \partial x_j}(a)
\]

is the Hessian of \( f \) at \( a \). Then there exists a local coordinate system \((x_1, \ldots, x_k)\) around \( a \) such that

\[
f = f(a) + \sum h_{ij} x_i x_j
\]

So, all neighborhoods of nondegenerate critical points are quadratic in nature. The connection to polytopes is found in [13]:

**Theorem 3.5:** If \( f \) is a differentiable function on a manifold \( M \) with no degenerate critical points, and if each \( M^c \) is compact, then \( M \) has the homotopy type of a CW-complex, with one cell of dimension \( \lambda \) for each critical point of index \( \lambda \).

A bit of explanation, \( M^c \) denotes a level set. In the CMA error surface all level sets have finite bound, the compactness is met. The “index” of a critical point is the number of negative eigenvalues of its Hessian. Although the term “CW-complex” is fairly technical in a general topology sense, in our setting here (a finite dimensional compact, manifold, with a finite number of stationary points) we may take it to simply mean a polytope. Thus we finally have a connection between local behavior of the vector field at the critical points and the global structure of the manifold. As this is a rather abstract connection, we continue with our example of the \( M_n \) for a three tap channel-equalizer.

#### 3.6.1. CW-Complex of \( M_n \) for Platykurtic IID Source

For platykurtic sources, we had from Figure 4a “+” = minima, “0” = saddles, “−” = maxima. On \( S^2 \), this means that “+” has a Morse index of 0, “0” has a Morse index of 1, “−” has a Morse index of 2, so construction of the homotopic equivalent CW-complex means that points associated with “+” become a face with dimension 0 (vertex), “0” become a face with dimension 1 (edge), “−” become a face with dimension 2 (face). The resulting CW-complex forms the octahedron in 3 space (more generally the \( n \)-octahedron or cross-polytope). As shown in Figure 5a.

The relation to Euler Characteristic is seen to be,

\[
\chi(S^2) = \chi(M_n) = \sum_{i=0}^{n} (-1)^i |F_i| = 6 - 12 + 8 = 2
\]

So, we have agreement between the Euler Characteristic of \( M_n \) and the relation of the numbers of types of stationary points.
3.6.2. CW-Complex of \( M_s \) for Leptokurtic IID Source

Interestingly, performing the same for a leptokurtic source a cube is formed (more generally an \( n \)-cube) as in Figure 5b which is the dual polytope (i.e. exchanging faces and vertices) of the octahedron.

Figure 5. CW-Complex of for Platykurtic & Leptokurtic Sources

Up to this point, we have considered only the CW-complexes generated by iid cases because we know the stationary points locations and their Hessian. However, for the correlated case we do not know this information in general.

4. TEMPORALLY CORRELATED SOURCES

Next we relax the condition of the source “whiteness” and investigate CMA under (a class of) temporally correlated sources. Direct algebraic solution for the roots of \( F(H) \) is quite difficult in this case. However, the numerical approach of continuation methods [14] may be applied to solve for roots of \( F(H, \tilde{R}) \) given a specific \( \tilde{R} \) (see [8]). Choosing temporal correlations which have equiprobable symbols (hence \( \kappa = 3 \)) we determined the stable, stationary points (and their equalization performance) for a four tap system \((n + 1 = 4)\) for 220 randomly chosen periodic sequences of varying period. Curiously, for any given source, some minima performed better than others. At times, the performance difference was quite large!

The number of stable stationary points due to a particular sequence of period \( P \) was computed and shown in Figure 6. For sequences of short period, the number of minima may be double of that of sequences of long periods. Under iid sources, for a four tap system we expect 8 minima. However, under source correlation there are as many as 16 minima.

Figure 6. Number of Minima under vs. Period \( P \).

Such an increase in the number of minima may be alarming meaning that the error surface has not merely deformed (shifting slightly the minima) but deformed to the point that new minima have been created.

Figure 7 plots 5 of the 10 (instead of 8) minima due to a source of period 32 (the other 5 minima are the negatives of those shown). Eight are fairly close to a good equalization performance. However, under correlation there are as many as 16 minima.

For this 4 tap experiment we have

\[
\chi(M_s) = \chi(S^g) = 0 = \sum_{i=0}^{3} (-1)^i |\mathcal{F}_i|
\]  

(14)

where \(|\mathcal{F}_i|\) is the number of stationary points with \( i \) negative eigenvalues of its Hessian on \( M_s \). From Figure 7 we have \( I_0 = 10 \) (minima), \( I_2 = 14 \), \( I_4 = 10 \), \( I_6 = 6 \) (maxima). By (14) we have, \( I_0 - I_2 + I_4 - I_6 = 10 - 14 - 10 - 6 = 0 \).

5. CONCLUSION

We observe behavior of CMA under varying source distributions and correlation. Under correlated sources CMA exhibits some rather extreme behavior. Numbers of minima and saddle points change. Thus, not only is asymptotic behavior changed (minima), but also the transient behavior due to the saddles. Identification of the CMA error surface through classification of polytopes is presented.

REFERENCES