Does Fractionally–Spaced CMA Converge Faster Than LMS ?

A. Touzni, I. Fijalkow

ENSEA / ETIS, 95014 Cergy-Pontoise Cedex, France

fax: (33-1) 30 73 66 27

e-mail:touzni,fijalkow@ensea.fr

ABSTRACT

This paper addresses the convergence rate study of the Fractionally-Spaced Equalizer updated by Constant Modulus Algorithm (FSE-CMA). By analyzing the average algorithm behavior we compare the FSE-CMA to the FSE-LMS. Although the FSE-CMA is based on a fourth order statistics criterion, we will show for constant modulus input that the algorithm has the amazing property to converge locally twice as fast as FSE-LMS (which requires a training sequence). Furthermore, we will show that the global FSE-CMA transient behavior convergence is accomplished in two steps.

Keywords: Fractionally spaced equalization, Constant Modulus Algorithm, Ordinary Differential Equation.

1. INTRODUCTION

In digital communications, the Constant Modulus Algorithm (**CMA**) is one of the blind adaptive algorithms most commonly used in digital communications to remove the intersymbol interference (**ISI**) induced by channel transmission.

In recent contributions, it has been shown that channel diversity (due to oversampling the received analog signal and / or using a sensors array) allows, under some conditions, the fractionally-spaced CMA equalizers (**FSE-CMA**) to achieve perfect equalization ([1]). Moreover, on the contrary of the recent approaches based on second order statistics methods ([2], [3], [4]), FSE-CMA shows some robustness when the conditions required for perfect equalization are not met (see [5], [6]).

Nevertheless, since the FSE-CMA cost-function is a highly non-linear form, the convergence speed is known to be very dependent of the algorithm initialization, inducing sometimes trouble for practical on line implementation. In order to analyze the convergence rate, we address herein the averaged FSE-CMA behavior by using the Ordinary Differential Equation (**ODE**) technic. The result is compared then to that of FSE-LMS. It yields the asymptotic convergence behavior in the vicinity of the convergence point of the algorithm and the mean transient behavior.

2. FSE-CMA

The spatio-temporal channel equalization problem consists in choosing the $L \times 1$ Finite Impulse Response (**FIR**) equalizer transfer function $\vec{e}(z) = (\mathrm{e}_1(z), ..., \mathrm{e}_L(z))^\top$, with $e_k(z) = \sum_{i=0}^{N} e_{k,i}\, z^{-i}$ so that its output $y(n)$ achieves a "good" estimate of the i.i.d. input sequence $s(n-\nu)$ (with ν an integer delay), as displayed on Figure 1. The L-dimensional FIR channel transfer function is denoted $\vec{c}(z) = (\mathrm{c}_1(z), ..., \mathrm{c}_L(z))^\top$, where each $\mathrm{c}_k(z)$ writes as $c_k(z) = \sum_{i=0}^{Q} c_{k,i}\, z^{-i}$.

The additive noise is described by the L-dimensional $(w_1(n), ..., w_L(n))^\top$ vector, and is neglected in our study.

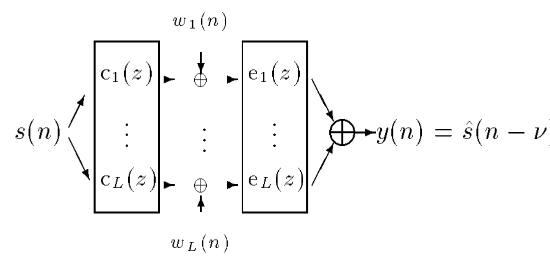

Figure 1: Spatio-Temporal Equalization Scheme

From the previous propagation model, the equalizer $\vec{e}(z)$ must satisfy the Bezout equation:

$$\vec{e}^\top(z)\vec{c}(z) = e_1 c_1 + ... + e_L c_L = z^{-\nu} \qquad (1)$$

where $h(z) = z^{-\nu}$ represents the global (channel + equalizer) desired transfer function.

This problem formulation is turned on choosing the NL $(N-1 \geq Q)$ equalizer impulse response $\vec{e}$ (with entries the taps of $\vec{e}(z)$), such that $\vec{e}$ minimizes the FSE-CMA cost-function $\mathcal{J}(\vec{e}) = E[(r_2 - y(n)^2)^2]$, where $r_2 = \frac{E[s^4]}{E[s^2]}$ is the a priori statistical information of the input signal, called dispersion constant. Note that, for $L = 1$, this algorithm is known as the Godard algorithm ([7]) or CMA ([8]).

The FSE-CMA is the gradient descent algorithm that minimizes $\mathcal{J}(\vec{e})$. It can be written as,

$$\vec{e}(n+1) = \vec{e}(n) + \mu\, y(n)(r_2 - y^2(n))\vec{R}(n) \qquad (2)$$

where $y(n) = \vec{e}^{\top}(n)\vec{R}(n)$ is the equalizer output and $\vec{R}(n) = \mathcal{T}(\mathbf{c})S(n)$ is a regressor vector of the N last L-dimensional observations. $S(n)$ contains the input sequence at $n, n-1, ..., n-N-Q+1$. $\mathcal{T}(\mathbf{c})$ is the $NL \times (N+Q)$ Sylvester channel convolution matrix containing the taps of $\vec{c}(z)$ (see [4]). Note that, $\mathcal{T}(\mathbf{c})$ is a full column-rank matrix, under the fundamental hypothesis of channel identifiability, i.e., when there is no common zeros between all components $c_k(z)$ k=1,...,L.

The FSE-CMA will be compared to the well known **FSE-LMS** algorithm:

$$\vec{e}(n+1) = \vec{e}(n) + \mu\,(s(n-\nu) - y(n))\vec{R}(n) \qquad (3)$$

where the input sequence $s(n)$ must be known to allow the receiver to adapt the linear equalizers taps.

3. AVERAGED FSE-CMA

We introduce, in this section, the so-called Ordinary Differential Equation (**ODE**) associated to the adaptive algorithm (2). The ODE describes the stochastic adaptive FSE-CMA by a continuous differential equation which leads to analyze the FSE-CMA mean behavior more easily and provides the equilibria setting of the cost-function $\mathcal{J}(\vec{e})$.

However, before expressing the ODE, we need to take some caution. In fact, we propose to express the ODE in function of the impulse response h corresponding to the channel/equalizer global scalar system. So, we introduce $h = \mathcal{T}(\mathbf{c})^{\top}\vec{e}$, where h is a $(N+Q)-$length vector, with entries the taps of $h(z)$. Because $\mathcal{T}(\mathbf{c})$ is full column-rank, for a given $h \in \mathcal{R}^{N+Q}$ there is infinitely many possible equalizers settings $\vec{e}$ spanning a dense subspace of dimension $NL - (Q+N)$.

One of the main interest of this parametrization is to be transparent with respect to a possible leakage effect in the equalizers space $\vec{e}$.

In term of h, the ODE is:

$$\frac{1}{4}\,\frac{\partial h}{\partial t} = E\big[y(n)(r_2 - (h^{\top}S(n))^2)S(n)\big] = E[s^2]^2\, D_h \quad (4)$$

where $E[.]$ stands for the mean operator expectation. In [1] it was shown that $D_h = \Delta(h)\, h$, with

$$\Delta(h) = (3\,\|h\|^2 - \rho)I - (3-\rho)\mathrm{diag}(hh^{\top})$$

$\rho = \frac{E[s^4]}{E[s^2]^2}$ is the input signal kurtosis, and $\mathrm{diag}(A)$ is the operator defined as the matrix extracted from A with the same diagonal entries and 0 elsewhere.

The possible convergence settings of the FSE-CMA (2) satisfy the equation $\Delta(h)\, h = 0$. The stability of each stationary point is then determined by the sign-definiteness of the Hessian matrix:

$$\Psi(h) = (3\|h\|^2 - \rho)I + 6hh^{\top} - 3(3-\rho)\mathrm{diag}(hh^{\top}) \quad (5)$$

Lemma 1 *The stationary points of (4) can be classified as (see [1]):*
- *maximum for* $h = 0$,
- *global minima for* $h = h_\nu = (0...1...0)^{\top}$, *where the non-zero entry is the* $(\nu+1)^{th}$ *component.*
- *saddle points. The* M *non-zero components of* h *are all equal to* $\kappa_M = \sqrt{\rho/(3(M-1)+\rho)}$.

The important result is that when $L > 1$ all global minima achieve perfect equalization, (on the contrary of the case $L = 1$). More accurately, for each h_ν, there is an infinity of FIR equalizers satisfying (1). Note that, when the noise can not be neglected the global minima are slightly perturbed and may sometimes induce local minima (see [6] for more details).

4. LOCAL MEAN CONVERGENCE RATE

In the vicinity of a stationary point h_*, a straightforward Taylor development of the gradient D_h, leads to,

$$D_h = D_{h_*} + \nabla D_{h_*}(h - h_*) + o(\|h - h_*\|) \qquad (6)$$

So that, for a stationary stable point h_ν, we get:

$$\frac{\partial(h - h_\nu)}{\partial t} = -\Psi(h_\nu)\,(h - h_\nu) \qquad (7)$$

With $\Psi(h_\nu)$ a diagonal matrix with entry $(3-\rho)$ when $i \neq \nu+1$, and 2ρ when $i = \nu+1$.
In particular, when $\rho = 1$ (i.e., $s(n)$ is a constant modulus signal), we have $\Psi(h_\nu) = 2\,I$. In this case, one can see that there is an "hyper convergence" to h_ν. Indeed, locally we obtain the expression,

$$h(t) \simeq h_\nu + e^{-2\,t}\,(h(0) - h_\nu) \qquad (8)$$

when $h(0)$ is an initial value in a neighborhood of h_ν. Accordingly, for the FSE-LMS, the ODE writes as $\frac{\partial h}{\partial t} = h - h\nu$, where the delay ν is fixed by (3). In this case, the convergence is locally $h(t) \simeq h_\nu + e^{-\,t}(h(0) - h_\nu)$. Consequently, the averaged FSE-CMA converges twice as fast as FSE-LMS ! This result is all the more interesting (and surprising) that although FSE-CMA is a fourth-order statistics blind based algorithm, the FSE-CMA convergence is locally faster than the second-order statistics based FSE-LMS algorithm, which needs a training sequence. Note that a similar result has been given by Treichler in the non fractional scheme, for the specific case of sinusoidal inputs ([9]).

When $\rho \neq 1$ (and $\rho < 3$) the FSE-CMA behavior is slightly different. In this case, all components of $\Psi(h_\nu)$, but one, are equal to $(3-\rho)$. So, the convergence speed depends on the kurtosis ρ of the input signal (note that the input signal must be a non-gaussian distributed sequence). Moreover, for a PAM or QAM input sequences $\rho < 1$, thus $3 - \rho \geq 2$. So that the convergence speed is even higher, for almost all taps, than in the constant modulus signal case !

As we use a global impulse response parametrization, we may notice that the convolution channel matrix $\mathcal{T}(\mathbf{c})$ does not appear explicitly in (8). In fact, in term of $\vec{c}$, the convergence speed for the FSE-LMS as well as for the FSE-CMA, is of course dependent of the condition number of the observed regressor covariance matrix $\mathcal{R}(c) = E[s^2]\,\mathcal{T}(\mathbf{c})\mathcal{T}(\mathbf{c})^\top$ and more accurately to the distance between the zeros of $\vec{c}(z)$.

Note that, since (8) is only valid locally, there are no real practical consequences in term of algorithm initialization, for instance. In fact, the interest is mostly due to the fact that this result is highly non intuitive.

Simulations: The simulations were computed with a 2-dimensional multichannel vector $\vec{c}(z)$ with zeros -1.4 and 0.6 for $c_1(z)$, 1.1 and -0.4 for $c_2(z)$. The input signal is a BPSK (+1/−1) sequence ($\rho = 1$). FSE-LMS (Figure 1) and FSE-CMA (Figure 2) use a step-size $\mu = 0.04$. Figure 1 shows the convergence of the global impulse response h to the stationary canonical vector setting h_ν in terms of the Minimum Mean Square Error (MMSE). One can see that convergence is achieved within about 100 iterations for FSE-CMA initialized close to h.

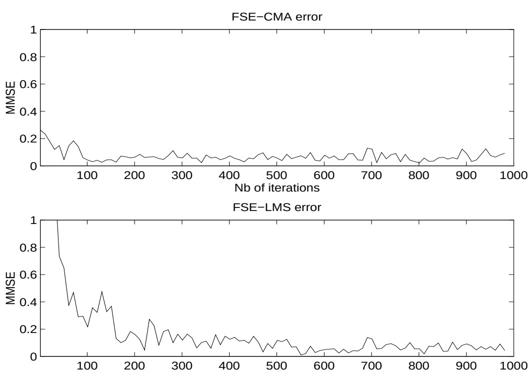

Figure 2: MMSE

5. GLOBAL BEHAVIOR

Our goal in this section is to bring out in relief two transient different FSE-CMA behaviors whether h is close or not to a region including the equilibria settings. Note that, we do not intend to give an exhaustive mathematical study but only some insights for a better understanding of FSE-CMA behavior.

To do so, we need first to introduce the following corollary of Lemma 1.

Corollary 1 *The set of minima and saddle points of the FSE-CMA cost-function is located in the region:*

$$\mathcal{B} = \{\ (N+Q)\,\kappa_{N+Q}^2 \leq \|h\|^2 \leq 1\} \subset \mathcal{R}^{N+Q}$$

where, $\kappa_{N+Q}^2 = \rho/(3(N+Q-1)+\rho)$, $\rho < 1$.

Proof: The upper boundary is defined by the set of the $(N+Q)$ global minima h_ν. When $\rho < 1$, the lower boundary is $min\{M\,\kappa_M^2\}_{M=2,\ldots,N+Q} = (N+Q)\,\kappa_{N+Q}^2$. □

We propose to decompose the mean gradient vector D_h of the ODE (4) in a projection on the vector h, denoted $D_h^{//}$ (corresponding to the radial contribution) and on its orthogonal complement $D_h^\perp$, (the tangential contribution). Hence, we have:

$$D_h = \underbrace{\frac{h^\top D_h}{\|h\|^2}\,h}_{D_h^{//}} + \underbrace{\left[D_h - \frac{h^\top D_h}{\|h\|^2}\,h\right]}_{D_h^\perp}$$

Based on this decomposition, we further derive the following properties.

Lemma 2 *If h has P "large" enough almost equal components and $N+Q-P$ "small" components a close-form equation of ODE is given by the expression,*

$$\frac{\partial h}{\partial t} \simeq D_h^{//} \qquad (9)$$

Proof: We set for convenience $A = 3(P-1)+\rho$ and $M = N+Q-P$. We suppose that h has P non-zero components almost equal to h_0 (up to ϵ) and M components ϵ such as $h_0 \gg \epsilon$ and $\epsilon h_0 \ll 1$. A straightforward calculus leads to the expression, $\|D_h^\perp\|^2 = (3-\rho)^2 M\epsilon^2 h_0^4 + o(\epsilon^2 h_0^4)$ and $\|D_h^{//}\|^2 = P h_0^2 (h_0^2 A - \rho)^2 - 2MA\rho\epsilon^2 h_0^2(1+1/P) + \epsilon^2 h_0^4 A^2 M(1+1/P) + o(\epsilon^2 h_0^4)$. Thus, as soon as $\epsilon^2 h_0^2 \to 0$, the tangential contribution can be neglected behind the radial contribution. □

Lemma 2 characterizes the regions of $\mathcal{R}^{N+Q}$ where the FSE-CMA trajectories "go straight" to the region of interest $\mathcal{B}$. Indeed, as the tangential contribution is negligible, the first step transient trajectories fit (approximatively) the trajectories corresponding to the radial projection of the initialization algorithm setting $h(0)$ on the boundary of $\mathcal{B}$ (denoted $h_\mathcal{B}$). When the hypothesis of Lemma 2 does not hold the tangential contribution can not be neglected and some correction must be given to find a more accurate close-form expression of the ODE.

Nevertheless, in both cases the convergence speed is mainly driven by the norm of the gradient $\|D_h\|$ which is all the more important that h is "far" from the region $\mathcal{B}$ (i.e., when $\|h\| \gg 1$), see Figure 3.

When h is in the neighborhood of the region $\mathcal{B}$ the norm $\|D_h\|$ is slow. So, the convergence is much slower. Consequently, the path followed by the trajectories from $h(0)$ to a stationary stable point is much greater than in the previous case.

To simplify the analysis we distinguish the two limit cases:

Case (a): $h(0)$ has one non-zero component h_0. Invoking lemma 2, the main gradient contribution is $\|D_h^{//}\|^2 = \rho^2 h_0^2(h_0^2-1)^2$, and the radial projection of $h(0)$ on $\mathcal{B}$ leads to $h_{\mathcal{B}} = (0...1...0)^\top = h_\nu$ corresponding exactly to a global stationary setting of $\mathcal{J}(\vec{e})$.

Case (b): $h(0)$ has P non-zero exactly equal components (with $P>2$). The main contribution of the mean gradient vector is $\|D_h^{//}\|^2 = P h_0^2\,(h_0^2(3(P-1)+\rho)-\rho)^2$. Thus, the projection of $h(0)$ on $\mathcal{B}$ corresponds to a saddle setting, i.e., $h_0^2\,(3(P-1)+\rho)=\rho$.

According to the cases (a) and (b), we may easily understand that the "worst" initialization corresponds to $h(0)$ first attracted by a saddle point. Unfortunately, when $N+Q$ increases, the number of saddle points $\sum_{k=2}^{N+Q} \frac{(N+Q)!}{(N+Q-k)!\,k!}$ increases too, and is much greater than $N+Q$ the number of global minima h_ν. Thus, when initialization is arbitrary, it is all the more likely to be at first attracted by a saddle setting that the global impulse response is large. In this case, the penalization is mostly due to the last step of convergence, i.e., from the vicinity of a saddle point to a stable setting. Worse still, we may notice that even if a saddle point is close to a global stationary setting, the damage in terms of ISI is, in most cases, non negligible.

Simulations: We simulate the trajectories of the ODE for different initialization settings versus the norm of D_h. The simulation were computed with $\rho = 1$ and a h having 2 entries. We represent 2 global minima $(0,1)$, $(1,0)$, and a saddle point $(0.5, 0.5)$. The limit cases (a) and (b) are displayed in dotted line. When the initialization respects the condition of Lemma 2 (curves 1-2-3), we can see that the convergence is "straight" to a neighborhood of stationary settings. Furthermore, the convergence is all the more faster that $\|h\| \gg 1$.

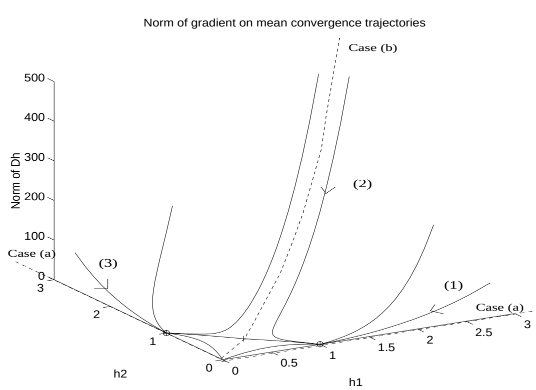

Figure 3: Trajectories of ODE

Concerning the comparison with the FSE-LMS, we must remark that on the contrary of FSE-CMA, the ODE of the FSE-LMS $\frac{\partial h}{\partial t} = E[s^2](h-h_\nu)$ has only one global minimum as equilibria setting. Thus, independently of the initialization $h(0)$, the FSE-LMS trajectories go straight to the convergence point h_ν. In order to compare the convergence speed, we set in Table-I, the magnitude of the gradient vector D_h for the specific cases (a) and (b). It appears that when $h_0 \gg 1$ the convergence speed of FSE-CMA is greater than that of the FSE-LMS. The behaviors are opposite when $h_0 < 1$. Consequently, when h is in a neighborhood of $\mathcal{B}$ FSE-LMS is globally faster, even if in the vicinity of h_ν, FSE-CMA is better (see previous section).

Table-I

	FSE-LMS	**FSE-CMA**
D_h	$\|h-h_\nu\|^2$	$\|D_h^{//}\|^2$
(a)	$(h_0-1)^2$	$\rho^2 h_0^2(h_0^2-1)^2$
(b)	$Mh_0^2+1-2\sum_j h_j h_{j,\nu}$	$Mh_0^2(h_0^2(3(M-1)+\rho)-\rho)^2$

6. CONCLUSION

In this paper we have given some insights on the FSE-CMA convergence speed behavior. We have shown that in the neighborhood of the global minima FSE-CMA converges twice as fast as FSE-LMS, but that when initialization is arbitrary the two steps transient behavior of convergence may be much slower.

References

[1] I. Fijalkow, C. Manlove, C.R. Johnson Jr., *Adaptive Fractionally Spaced Blind Equalization*, submitted to IEEE Tr. on SP, January 1995. Partially published in Proc DSP workshop, 1994.

[2] D.T.M. Slock *Blind Fractionally-Spaced Equalization, Perfect Reconstruction Filter Banks and Multichannel Linear Prediction*, in Proc ICASSP, vol 4, pp. 585-588, 1994.

[3] L. Tong, G. Xu, and T. Kailath, *Blind identification and based on second-order statistics: a time domain approach*, IEEE Tr. on IT, vol. 40, n. 2, pp. 340-349, 1994.

[4] E. Moulines, P. Duhamel, J.-F. Cardoso, and S. Mayrargue, *Subspace methods for the blind identification of multichannel FIR filters*, IEEE Tr. on SP, January 1995.

[5] I. Fijalkow, J.R. Treichler, C.R. Johnson Jr., *Fractionally Spaced Blind Equalization: Loss of Channel Disparity*, in Proc. ICASSP' 95, vol. 4, pp. 1988-1991, 1995.

[6] A. Touzni, I. Fijalkow and J.R. Treichler *Robustness of Fractionally-Spaced Equalization by CMA to Lack of Channel Disparity and Noise*, in Proc SSAP'96, 1996.

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

[8] J.R. Treichler, and B.G. Agee, *A new approach to multipath correction of constant modulus signals*, IEEE Tr. on ASSP, vol. 31, n. 2, pp. 459-472, 1983.

[9] J.R. Treichler and M.G. Larimore, *Convergence rate for the constant modulus algorithm with sinusoidal inputs*, 1984.