ELSEVIER

Signal Processing 59 (1997) 335-340

SIGNAL PROCESSING

Fast communication

Attraction of saddles and slow convergence in CMA adaptation

S. Lambotharan [a,*], J. Chambers [a,1], C.R. Johnson Jr [b,2]

[a] *Signal Processing Section, Department of Electrical and Electronic Engineering, Imperial College, Exhibition Road, London SW7 2BT, United Kingdom*

[b] *School of Electrical Engineering, Cornell University, Ithaca, NY 14853, USA*

Received 13 February 1997

Abstract

We show that the most widely used blind equalization algorithm, the constant modulus algorithm, CMA, can be attracted during one convergence trajectory to the vicinity of more than one of the saddles in its error performance surface where it exhibits very slow convergence. We also establish bounds on the attraction and escape rates at a saddle and show that the saddles associated with lower energy levels have slower escape rates than the saddles with higher energy levels. These results highlight the need for intelligent initialization schemes for the CMA algorithm. We suggest a step normalisation technique to improve convergence speed in the vicinity of a saddle. © 1997 Elsevier Science B.V.

Keywords: Adaptive signal processing; Blind equalization; Constant modulus algorithms; Convergence and saddle points

1. Introduction

Equalization of a communication channel in the absence of a training sequence is termed *blind* equalization. A blind equalizer increases bandwidth efficiency and is important in a situation where a training sequence is impractical or costly, for example in the proposed HDTV system. Among all blind equalization algorithms, the CMA that was originally proposed in [2] and developed independently in [5], has been shown to be robust to channel under-modelling, channel noise perturbation and loss of channel disparity.

* Corresponding author. E-mail: lambo@ee.ic.ac.uk.

[1] Supported by the Engineering and Physical Science Research Council of the United Kingdom.

[2] Supported by NSF Grant MIP-9509011 and Applied Signal Technology.

2. Attraction of saddles

For certain initializations, the Constant Modulus Algorithm (CMA) can be observed to have flat portions in the parameter error trajectory, apparently due to the attraction of saddles in the error performance surface. The question whether there will be only one such attraction prior to convergence is curious, because if the answer is yes, one might tolerate this temporary attraction, hoping that the next stationary point would be an acceptable minimum. In this section, we demonstrate through a selected example that CMA can be attracted to the vicinity of more than one saddle prior to convergence, even with a centre tap initialization, [2]. Hence, a more judicious initialization strategy or adaptation scheme is required to avoid slow convergence.

PII S0165-1684(97)00103-5

A fractionally spaced (T/2) channel is chosen as $0.4000 + 1.16650z^{-1} - 0.4400z^{-2} + 0.6268z^{-3} + 0.4000z^{-4} + 0.0751z^{-5}$ whose zeros are $z = 0.86\exp(\pm 58)$, $z = 0.27\exp(\pm 149)$ and $z = -3.3579$. The length of the fractionally spaced equalizer (FSE) is 4, i.e. the length of each sub channel is 2, and off baud sampling is assumed. The length of the combined channel + equalizer after decimation is, therefore, 4. The equalizer is initialized at [0 1 0 0], and adapted using an exact gradient descent method (i.e. explicitly calculating the CMA gradient at each iteration). The cost and the evolution of the combined channel + equalizer impulse response (h) are respectively depicted in Figs. 1 and 2. The equalizer is first attracted to a saddle that is formed by three non-zero coefficients $h = [0.378\ 0.378\ 0.378\ 0]$, and then attracted to a second saddle that is formed by two non-zero coefficients $h = [0\ 0.5\ 0.5\ 0]$, before converging to a minimum $h = [0\ 0\ 1\ 0]$.

This shows that the CMA algorithm can be attracted to more than one saddle, around which it exhibits very slow convergence. Such a possibility is indicated in the topological studies of [4].

3. Attraction and escape rates

In this section, we will show that the convergence behaviour of CMA is quite different for saddles with unequal energy levels, by looking at the attraction and escape rates. Consider a fractionally spaced (T/2) channel and an equalizer of order Q and $Q-1$. Each sub channel is assumed to have no common zeros so that the associated channel convolution matrix Δ is full rank. The transmitted sequence is assumed to be BPSK (for example a binary sequence of +1 and −1).

The CMA cost, its gradient and Hessian are respectively written as $E\{(y^2(k)-1)^2\}$, $\nabla_e(J) = 4\Delta^{\mathrm{T}}\Lambda h$ and $\nabla_e^2(J) = 4\Delta^{\mathrm{T}}\Psi\Delta$. (see [3] and [1]). Where $\Lambda = \mathrm{diag}[\Lambda_0\ \Lambda_1\ \ldots\ \Lambda_P]$ and $\Lambda_j = 3\sum_{l=0}^{P} h_l^2 - 1 - 2h_j^2$,

$$\Psi = \begin{bmatrix} 3\sum_{i=0}^{p} h_i^2 - 1 & 6h_0h_1 & \cdots & 6h_0h_p \\ \vdots & & & \vdots \\ 6h_ph_0 & \cdots & 6h_ph_{p-1} & 3\sum_{i=0}^{p} h_i^2 - 1 \end{bmatrix}$$

and Δ is a $(2Q \times 2Q)$ channel convolution matrix

$$\Delta = \begin{bmatrix} c_0^e & \cdots & c_Q^e & 0 & \cdots & 0 \\ c_0^o & \cdots & c_Q^o & 0 & \cdots & 0 \\ 0 & c_0^e & \cdots & c_Q^e & 0 & \cdots \\ 0 & c_0^o & \cdots & c_Q^o & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & c_0^e & \cdots & c_Q^e \\ 0 & \cdots & 0 & c_0^o & \cdots & c_Q^o \end{bmatrix}^{\mathrm{T}}.$$

$y(k)$ is the equalizer output, e is the equalizer parameter vector, $h = \Delta e$ is the combined channel + equalizer impulse response of length $p = 2Q-1$, and $\{c_i^e\}$ and $\{c_i^o\}$ are respectively the even and odd sub channel coefficients. The average behaviour of CMA adaptation can be written as

$$e_{k+1} = e_k - \mu\nabla_{e_k}(J) = e_k - 4\mu\Delta^{\mathrm{T}}\Lambda\Delta e_k, \qquad (1)$$

where μ is the "small" step size. Using a Taylor series expansion, the function $\Delta^{\mathrm{T}}\Lambda\Delta e_k$ can be expanded around any arbitrary $\bar{e}$ as follows:

$$\Delta^{\mathrm{T}}\Lambda\Delta e_k \simeq \nabla_{\bar{e}}(J) + \nabla_{\bar{e}}^2(J)(e_k - \bar{e}). \qquad (2)$$

At a stationary point $\bar{e}$, $\nabla_{\bar{e}}(J)$ is zero. Hence in the vicinity of a stationary point, (1) can be approximated as

$$\tilde{e}_{k+1} \simeq (I - 4\mu\nabla_{\bar{e}}^2(J))\tilde{e}_k = (I - 4\mu\Delta^{\mathrm{T}}\Psi\Delta)\tilde{e}_k, \quad (3)$$

where $\tilde{e}_k = (e_k - \bar{e})$. At a saddle point, it has been shown that the combined channel + equalizer impulse response will have v non-zero h_i's with equal magnitude, $1/\sqrt{3v-2}$, and the associated Hessian, $\nabla_{\bar{e}}^2(J)$, will have both positive and negative eigenvalues [3]. A positive eigenvalue means convergence towards a saddle and a negative eigenvalue means divergence away from a saddle. In order to analyse the attraction/escape rate at a saddle, the following theorem of [4] is important.

Theorem 1. *There are three distinct eigenvalues for Ψ at saddles. They are*

(i) $\lambda_0 = 2$,

(ii) $\lambda_i = \dfrac{2}{3v-2}$, $i = 1, \ldots, p-v+1$,

(iii) $\lambda_i = \dfrac{-4}{3v-2}$, $i = p-v+2, \ldots, p$.

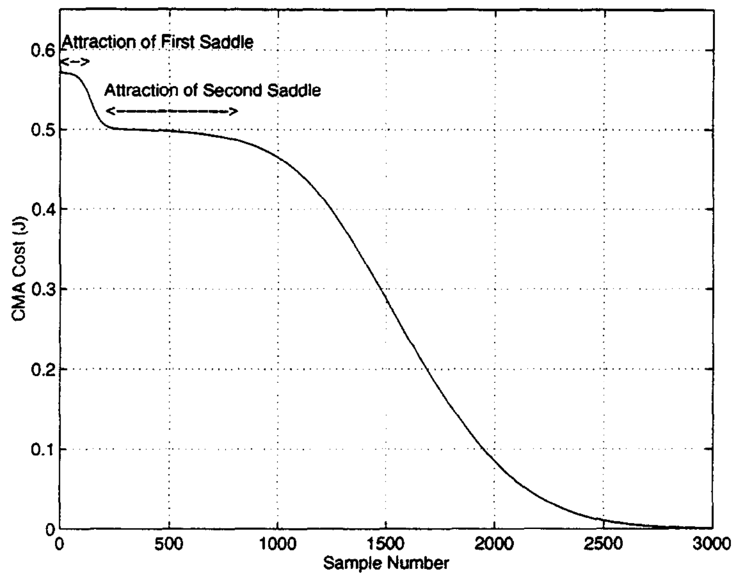

Fig. 1. The evolution of the cost associated with CMA.

The following corollary establishes an upper and lower bound on the escape rate for an FSE-CMA in the vicinity of a saddle.

Corollary 2. *The time constant at which the equalizer setting deviates from a saddle is bounded between*

$$\frac{3\upsilon-2}{16\mu\lambda_{\max}(\Delta^{\mathrm{T}}\Delta))} \leqslant \tau \leqslant \frac{3\upsilon-2}{16\mu\lambda_{\min}(\Delta^{\mathrm{T}}\Delta)}. \qquad (4)$$

See Appendix A for a proof.

Similarly, by writing Eq. (3) as $\tilde{e}_k^{\mathrm{T}}\tilde{e}_k \approx \tilde{e}_k^{\mathrm{T}}(I - 8\mu\Delta^{\mathrm{T}}\Psi\Delta)\tilde{e}_k$ and considering the negative eigenvalues of $\Delta^{\mathrm{T}}\Psi\Delta$, the rate at which the square of the Euclidean distance between the saddle and the equalizer parameter vector grows can be bounded between

$$\frac{3\upsilon-2}{32\mu\lambda_{\max}(\Delta^{\mathrm{T}}\Delta))} \leqslant \tau \leqslant \frac{3\upsilon-2}{32\mu\lambda_{\min}(\Delta^{\mathrm{T}}\Delta)}.$$

For the 5th order channel given in the first example, the equalizer was initialised randomly close to the saddle, $h = [0.378\ 0.378\ 0.3780\ 0]$, and adapted using the exact CMA gradient method. The square of the Euclidean distance between the equalizer parameter vector and the saddle point as a function of the adaptation number is depicted in Fig. 3. The theoretical bounds on the escape rate are also shown as solid lines.

Similar to Corollary 2, the time constant at which the equalizer setting approaches a saddle can be bounded between

$$\frac{1}{8\mu\lambda_{\max}(\Delta^{\mathrm{T}}\Delta))} \leqslant \tau \leqslant \frac{3\upsilon-2}{8\mu\lambda_{\min}(\Delta^{\mathrm{T}}\Delta)}.$$

Though the theoretical bounds above are not very tight, these bounds merely indicate that the escape and approach rates at a saddle point do not depend on the channel convolution matrix alone, but also on the type of the saddle point. As υ increases, the lower and the upper bounds are increased, and the equalizer output power, $\upsilon/(3\upsilon-2)$, is decreased. In other words, the rate at which the equalizer parameter deviates away from a saddle increases as the distance of the saddle from the origin increases.

4. Improving the convergence speed

A technique to improve the convergence speed at a saddle is to apply a normalised step size CMA algorithm.

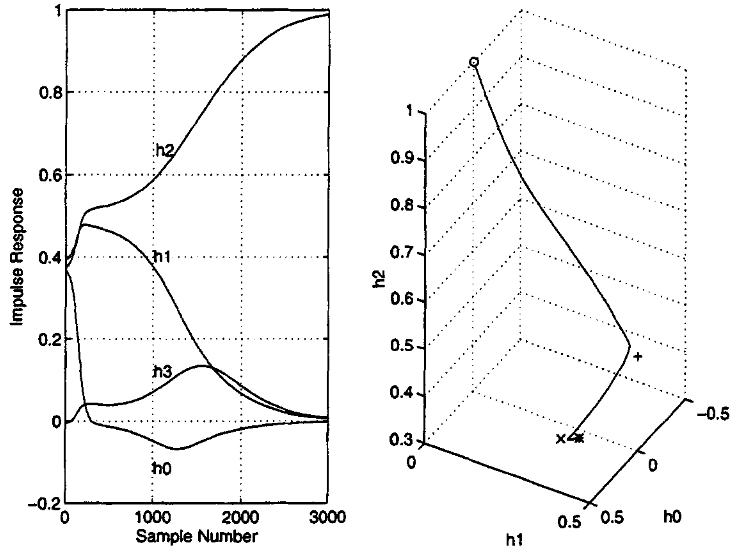

Fig. 2. The evolution of the total channel + equalizer impulse response coefficients: "+" position of saddle, "o" position of minimum and "*" initialization.

$$e_{k+1} = e_k + 4\frac{\mu}{E\{y^2(k)\}}X_k y(k)(1 - y^2(k)), \quad (5)$$

where X_k is the equalizer regressor vector. At a global minimum, $E\{y^2(k)\}$ is unity, hence there is no effect on the step size. At a saddle, however, $E\{y^2(k)\} = v/(3v-2)$, hence, the step size would be greater than μ (by a factor of 2 to 3), and will have an improvement in the convergence speed. The value of $E\{y^2(k)\}$ can be estimated at each sample as $\hat{E}\{y^2(k)\} = (1-\gamma)\hat{E}\{y^2(k-1)\} + \gamma y^2(k)$, where γ is a positive small step size (e.g. 0.02). In order to compare the proposal, a number of channels were randomly chosen and the equalizer was initialized at [0 1 0 0]. The number of samples taken to reach 98% of the final output power (of the equalizer) was averaged over 25 Monte-Carlo experiments. The SNR was 30 dB. Table 1 summaries the performance. A modest improvement is noted with normalisation.

Table 1
The number of samples taken for convergence

CHANNEL	CMA	VS-CMA
{0.67 0.06 0.09 −0.59 0.43 0.07}	435	349
{−0.18 −0.27 0.25 0.08 −0.84 0.34}	359	321
{0.35 0.39 −0.45 0.05 0.70 −0.17}	757	590
{0.75 −0.19 0.37 0.06 −0.39 −0.32}	561	442
{−0.59 0.45 −0.04 −0.41 −0.21 0.48}	733	641

5. Conclusion

We showed that the CMA algorithm can be attracted to the vicinity of more than one of the saddles in its error performance surface, and exhibits slow convergence. One method to avoid this slow convergence is a better initialization scheme. Another is the pragmatic output power normalisation of the equalizer adaptation step size (for which simple tests have demonstrated modest improvements). Yet another could be recognition of the diminished value of $E\{y^2(k)\}$ at a saddle and prescription of some clever equalizer parameter relocation.

Appendix A. Proof of Corollary 2

Write the Hessian at a saddle as

$$H = 4\Delta^{\mathrm{T}}\Psi\Delta = 4(\theta^{\mathrm{T}}\Delta)^{\mathrm{T}}D(\theta^{\mathrm{T}}\Delta) = 4q^{\mathrm{T}}Dq, \quad (6)$$

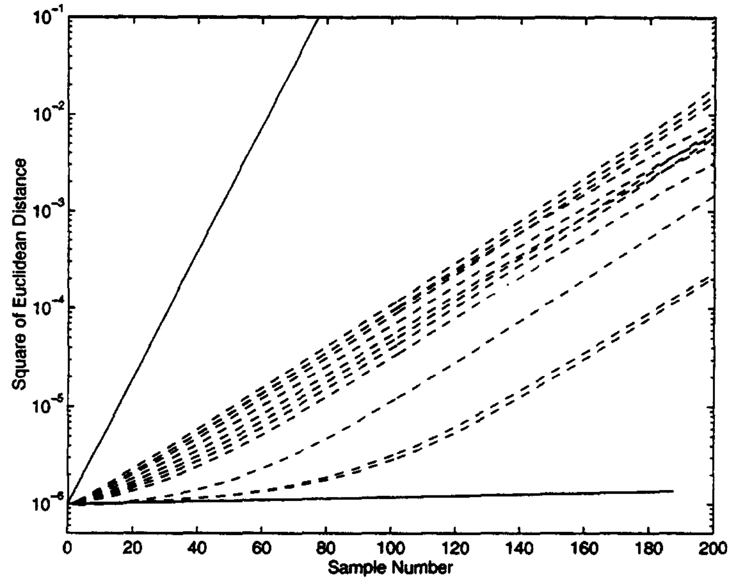

Fig. 3. The square Euclidean distance from the saddle point. Solid: theoretical upper and lower bounds; dashed: simulated results.

where D is a diagonal matrix. The diagonal elements of D and the columns of θ are respectively eigenvalues and corresponding eigenvectors of Ψ. Since $(D - \lambda_{\min}(D)I)$ is non-negative definite,

$$H = 4q^{\mathrm{T}}(\lambda_{\min}(D)I)q + 4q^{\mathrm{T}}(D - \lambda_{\min}(D)I)q \geqslant 4q^{\mathrm{T}}(\lambda_{\min}(D)I)q \tag{7}$$

Therefore $\lambda_{\min}(H) \geqslant 4\lambda_{\min}(D)\lambda_{\max}(\Delta^{\mathrm{T}}\Delta)$. It follows from Theorem 1 that $\lambda_{\min}(D) = -4/(3v-2)$. We introduce the notation $\lambda^n(H)$ to denote the negative eigenvalue of H. Hence

$$|\lambda^n(H)|_{\max} \leqslant 4|\lambda_{\min}(D)|\lambda_{\max}(\Delta^{\mathrm{T}}\Delta) = \frac{16}{3v-2}\lambda_{\max}(\Delta^{\mathrm{T}}\Delta). \tag{8}$$

This gives a lower bound for the time constant. To find an upper bound, write H^{-1} as follows:

$$\begin{aligned} H^{-1} &= 4q^{-1}D^{-1}(q^{\mathrm{T}})^{-1} \\ &= 4q^{-1}(\lambda_{\min}(D^{-1})I)(q^{\mathrm{T}})^{-1} \\ &\quad + 4q^{-1}(D^{-1} - \lambda_{\min}(D^{-1})I)(q^{\mathrm{T}})^{-1} \\ &\geqslant 4q^{-1}(\lambda_{\min}(D^{-1})I)(q^{\mathrm{T}})^{-1}. \end{aligned} \tag{9}$$

Therefore $\lambda^n_{\min}(H^{-1}) \geqslant 4\lambda^n_{\min}(D^{-1})\lambda_{\max}((\Delta^{\mathrm{T}}\Delta)^{-1})$. Notice that $\lambda^n_{\min}(H^{-1}) = 1/\lambda^n_{\max}(H)$ for any square matrix H. Thus,

$$|\lambda^n(H)|_{\min} \geqslant 4|\lambda^n_{\max}(D)|\lambda_{\min}(\Delta^{\mathrm{T}}\Delta) = \frac{16}{3v-2}\lambda_{\min}(\Delta^{\mathrm{T}}\Delta). \tag{10}$$

Therefore, the time constant at which the equalizer deviates from a saddle is bounded as indicated in (4). □

Acknowledgements

The authors would like to thank all members of Cornell University Blind Equalization Research Group.

References

[1] I. Fijalkow, F. Lopez de Victoria and C.R. Johnson Jr, Adaptive fractionally spaced equalization, IEEE Workshop on Digital Signal Processing, 1994, pp. 257–260.
[2] D.N. Godard, Self-recovering equalization and carrier tracking in two-dimensional data communication systems, IEEE Trans. Commun. 28 (11) (1980) 1867–1875.

[3] C.R. Johnson Jr and B.D.O. Anderson, Godard blind equalizer error surface characteristics: White, zero-mean, binary source case, Internat. J. Adaptive Control and Signal Process. 9 (1995) 301–324.
[4] J. LeBlanc, Effects of source distributions and correlation on fractionally spaced blind constant modulus algorithm equalizers, Ph.D. Thesis, Cornell University, 1995.
[5] J.R. Treichler and B.G. Agee, A new approach to multipath correction of constant modulus signals, IEEE Trans. Acoust. Speech Signal Process. 31 (1983) 459–472.