CONSTANT MODULUS PERFORMANCE SEARCH USING LMS METHOD

Hamadi Jamali, Member, IEEE and Tokunbo Ogunfunmi, Senior Member, IEEE

Santa Clara University, EE Dept.

ABSTRACT

The solution obtained by the CMA algorithm depends on both initial conditions and signal realizations. This paper uses the LMS method to seek the global minimum of the CM performance measure only. The resulting increase in computation is not prohibitive.

1. INTRODUCTION

The Constant Modulus (CM) minimization is the estimation of an unobservable sequence $\{s_t\}$, given n samples of a correlated sequence $\{x_t\}$. The estimate $\{y_t\}$ is chosen as:

$$y_t = W^* X_t \tag{1}$$

where W and X_t are the $n \times 1$ parameter and sample vectors, and * is the complex transpose. The vector W is selected by minimizing the CM criterion $\mathcal{J}$ [1] given as:

$$\mathcal{J} = \frac{1}{4} E[(\mid y_t \mid^2 - \gamma)^2] \tag{2}$$

where $\mid . \mid$ and $E[.]$ are the modulus and the expectation operators and γ is a known constant.

In practice, this problem is solved in real time using the Constant Modulus Adaptive (CMA) algorithm given by:

$$W_t = W_{t-1} - \mu \varepsilon_{t-1} y_{t-1} X_{t-1} \tag{3}$$

where μ is a constant, W_t tends to a minimum of $\mathcal{J}$ as t tends toward infinity, and $\varepsilon_t = \mid y_t \mid^2 - \gamma$.

A large body of research exists in this topic as described in [2]. This paper falls in a subcategory that uses the kronecker product to improve the performance of the CMA algorithm. In particular, [3] demonstrates the benefits of this notation in visualizing the topology of the CM function. These results are used in [4] to cast the problem of blind beamforming as a constrained matrix factorization. Similarly, [5] observes that some of the minima of the CMA algorithm in the context of equalization can be obtained by solving a linear system of equations where the unknowns are nonlinear functions of the equalizer parameters. Both [4] and [5] are, however, application specific and provide off-line solutions only. The first of these limitations is overcome by [6], which, formulates the CM as a generic minimization problem irrespective of the application where it is used. The second limitation of [4] and [5] is overcome by [7], which, solves the minimization of [6] using Newton's method. However, the approach of [7] requires calculating a matrix inverse and higher order statistics. This contribution proposes an LMS algorithm that avoids these limitations. Section 2 presents the new algorithm, its convergence properties and its implementation. Section 3 contrasts the behavior of the new algorithm against that of the CMA through an example and highlights its limitations. Section 4 offers concluding remarks.

2. LMS BASED CMA

This section introduces an LMS search of the CM function.

Theorem 1: Given an initial value, $W(-1)$, of an unknown $n \times 1$ vector, an $n \times 1$ sample vector $X(k)$, and a constant γ, an LMS search of the CM function is given as:

$$\theta(-1) = W(-1) \otimes \overline{W}(-1)$$

$$\textit{For} \quad k = 0 \quad \textit{to} \quad k = N-1$$

$$\varphi(k) = X(k) \otimes \overline{X}(k) \tag{4}$$

$$\varepsilon(k) = \theta(k-1)^* \varphi(k) - \gamma \tag{5}$$

$$\nabla(k) = \varepsilon(k)\varphi(k) \tag{6}$$

$$\theta(k) = \theta(k-1) - \mu\nabla(k) \tag{7}$$

$$M_\theta(k) = \textit{matrix}(\theta(k), n) \tag{8}$$

$$[U, S, V](k) = \textit{svd}(M_\theta) \tag{9}$$

$$W_*(k) = \sigma_1 u_1 \tag{10}$$

$$W_\otimes(k) = \frac{W_*(k)}{W_*(k)[1]} \tag{11}$$

where N is the number of iterations needed to converge, n is the data length, $\otimes$ is the kronecker product, $\overline{(.)}$ is the complex conjugate, $svd(.)$ is the function that computes the singular value decomposition of a given matrix, σ_1 is the largest singular value of the matrix M_θ, u_1 is the left singular vector corresponding to σ_1, $W_*(k)[1]$ is the first component of the vector $W_*(k)$, and $matrix(\theta, n)$ is the operator that converts the entries of a vector θ, n at time, to corresponding columns of a matrix M_θ.

Proof: Equation (2) may be expressed as [6]:

$$\mathcal{J} = \frac{1}{4}\theta^* R_{\varphi\varphi}\theta - \frac{\gamma}{2}\theta^* P_\varphi + \frac{\gamma^2}{4} \tag{12}$$

with the different variables defined as:

$$\begin{aligned} P_\varphi &= E[\varphi_t] \\ R_{\varphi\varphi} &= E[\varphi_t \varphi_t^*] \\ \theta &= W \otimes \overline{W} = [w_0 W^* \, w_1 W^* \, \ldots \, w_{n-1} W^*]' \\ \varphi_t &= X_t \otimes \overline{X}_t = [x_t X_t^* \, x_{t-1} X_t^* \, \ldots \, x_{t-n+1} X_t^*]' \end{aligned}$$

where the $l \times l$ matrix $R_{\varphi\varphi}$ is positive definite, with $l = n^2$.

Now, consider a function $\mathcal{K}(\theta)$ defined as in equation (12), but where the vector θ is not necessarily a kronecker product. The gradient $\nabla_\theta \mathcal{K}$ of $\mathcal{K}(\theta)$, with respect to θ, is:

$$\nabla_\theta J = R_{\varphi\varphi}\theta - P_\varphi \tag{13}$$

The global minimum θ_{kp} of $\mathcal{K}(\theta)$ is:

$$\theta_{kp} = R_{\varphi\varphi}^{-1} P_\varphi \tag{14}$$

To proceed, notice that equation (6) and the term $\nabla(k)$ in equation (7) are both instantaneous estimates of equation (13). Hence, equations (4) through (7) combined form an LMS search of the function $\mathcal{K}(\theta)$.

Equations (8) through (11) are output equations only and do not affect the iterative process since they do not enter into the feedback loop. Hence, the behavior of equations (4) through (7) may be considered separately. In particular, since $\mathcal{K}(\theta)$ is convex, the LMS algorithm of equations (4) through (7), if it converges, leads to a solution θ_{opt} in the vicinity of the absolute minimum θ_{kp} in equation (14).

Equations (8) through (10) compute a vector $W_*(k)$ that is the best rank 1 approximation to the vector $\theta(k)$. To see this, recall that the problem of approximating a given $m \times n$ matrix M_θ by a nearest rank 1 matrix [8] is stated as:

$$\min \quad || M_\theta - bc^* || \tag{15}$$

where b and c are $m \times 1$ and $n \times 1$ vectors, respectively.

The solution of this problem is given by:

$$b_{opt} = \sigma_1 u_1 \tag{16}$$

$$c_{opt} = v_1 \tag{17}$$

The vectors u_1 and v_1 are the first columns of the orthogonal matrices U and V in the decomposition:

$$M_\theta = U \Sigma V^* \tag{18}$$

Letting M_θ be the $n \times n$ matrix defined in equation (8), we see that equation (10) is just the expression of b_{opt} in equation (16). Also, using the properties of the kronecker product, equation (13) can be written as:

$$R_2 M_\theta R_1' = R_{xx} \tag{19}$$

where the matrices R_1 and R_2 are related as:

$$R_{\varphi\varphi} = R_1 \otimes R_2 \tag{20}$$

That is, the matrix M_θ is also positive semidefinite. Hence, the singular value decomposition of equation (18) becomes:

$$M_\theta = U \Sigma U^* \tag{21}$$

In this case, the vector c_{opt} is:

$$c_{opt} = \overline{u}_1 \tag{22}$$

Consequently, equation (10) is simply the solution of the particular nearest rank 1 matrix approximation defined as:

$$\min \quad || M_\theta - W_\otimes W_\otimes^* || \tag{23}$$

The normalization of equation (11) is used to ensure uniqueness of the vector $W_\otimes$ only, since any vector $W_\alpha = \alpha W_\otimes$ is also solution to the problem of equation (23).

Finally, since $\mathcal{K}(\theta)$ is convex, the vector W_{opt} extracted from θ_{opt} is close to the absolute minimum vector W_{cm} of the CM function $\mathcal{J}$. To quantify, the difference between these two solutions, notice that if θ_{kp} is a kronecker product form, then θ_{opt} is equal to θ_{kp}. Hence, $W_{opt} = W_{cm}$. However, if θ_{kp} is not a kronecker product form, then the vector θ_{opt} solves the system of equations:

$$R_{\varphi\varphi}\theta_{opt} = \gamma P_{opt} \tag{24}$$

P_{opt} exists since $R_{\varphi\varphi}$ is nonsingular. On the other hand, the vector $\theta_{cm} = W_{cm} \otimes \overline{W}_{cm}$, verifies the system:

$$R_{\varphi\varphi}\theta_{cm} = \gamma P_{cm} \tag{25}$$

Again, P_{cm} exists since $R_{\varphi\varphi}$ is nonsingular.

Using perturbation theory [8], the norm of the difference between the two solutions is bounded as:

$$\frac{|| \theta_{opt} - \theta_{cm} ||}{|| \theta_{opt} ||} \leq \varepsilon \varpi(R_{\xi\xi}) \frac{|| P_{cm} - P_{opt} ||}{|| P_{opt} ||} \tag{26}$$

where ε is arbitrarily small and $\varpi(R_{\varphi\varphi}) = \frac{\lambda_{max}(R_{\varphi\varphi})}{\lambda_{min}(R_{\varphi\varphi})}$, with $\lambda_{max}(R_{\varphi\varphi})$ and $\lambda_{min}(R_{\varphi\varphi})$ being the maximum and the minimum eigenvalues of the matrix $R_{\varphi\varphi}$, respectively. This difference is small when the signal model is adequate and the noise is not excessive. This ends the proof of theorem 1.

Theorem 1 seeks the CM absolute minimum only regardless of initial conditions. Unlike the case of the CMA algorithm, the null vector can be used as an initial value since the update term of the new algorithm is not equal to zero for a null vector. Equations (8) through (11) may be computed using the explicit singular value decomposition if small matrices are involved. However, if the matrix M_θ is large, it may be more economical to use the Lanczos method instead. The method of [7] may also be used.

We should emphasize that theorem 1 searches $\mathcal{K}(\theta)$ and not the CM function. In general, these two functions are not equal except in the special case when the augmented vector

θ can be decomposed into a kronecker product form. In fact, there is a gap between the minimum values of these two expressions. This gap goes to zero only when the estimated signal $\{y_t\}$ matches perfectly the unknown signal $\{s_t\}$.

The new algorithm differs from the conventional LMS method in three ways. First, it performs the iterations on an augmented vector θ instead of the parameter vector W. Second, it uses $\varepsilon(k)$ and $\varphi(k)$ instead of $e(k)$ and $X(k)$ needed for Wiener filter. Finally, the new algorithm features an output block $C(.)$ for computing the values of the parameter vector $W(k)$ from those of the augmented parameter vector $\theta(k)$. This output block may be viewed as a nonlinear measurement block analogous to that used in Kalman filtering.

The algorithm of theorem 1 may also be implemented as a block optimization procedure. In this case, there is a preamble period of duration N_1 needed for the iterations of equations (4) through (7) to converge to θ_{opt} with the time $t \leq N_1$. Once θ_{opt} is reached, we proceed with equations (8) through (11) to estimate W_{opt} and y_t with $t \geq N_1$. Note that no estimation of W_{opt} or of y_t is performed prior to time N_1. This block optimization is most effective when implemented in two separate chips. The first chip hosts equations (4) through (7) and runs at a faster rate than the communications link. The second chip performs the operations in equations (8) through (11) and runs at the same rate as the communications link. The rate of the first chip is such that θ_{opt} is reached before the next cycle of the communications link comes around.

Performing its update on an $l = n^2$ vector, the computational complexity of the algorithm of theorem 1 is of order $O(n^2)$, if W is computed as in [7]. If W is computed by the standard explicit singular value decomposition algorithm, the order of complexity increases to $O(n^3)$. Since only the largest singular value and the corresponding left singular vector are needed, more efficient algorithms requiring an order $O(n^2)$ only, are however, available [8]. This is still more expensive than the standard LMS method. But, it is in the same order as that of the commonly used Least Squares (LS) approach. As such, it is not too prohibitive. This complexity may be reduced further by using the properties of the kronecker product or parallel architectures.

Theorem 2: A necessary and sufficient condition for the LMS type algorithm of theorem 1 to converge to an approximate solution of the CM absolute minimum is:

$$0 < \mu < \frac{2}{\lambda_{max}} \tag{27}$$

Proof: Some of the details of this proof are omitted due to space limitations. Subtracting θ_{kp} from both sides of equation (7), using equation (6), replacing $\varepsilon(k)$ in terms of $\theta(k-1)$ and $\varphi(k)$, and rearranging, we get:

$$V(k) = (I - \mu\varphi(k)\varphi^*(k))V(k-1) - \mu\varepsilon_0(k)\varphi(k) \tag{28}$$

where $V(k) = \theta(k) - \theta_{kp}$ and $\varepsilon_0(k) = \theta^*_{kp}\varphi(k) - \gamma$. Taking the expectation of both sides of equation (28) and invoking the independence assumption, we obtain:

$$E[V(k)] = (I - \mu R_{\varphi\varphi})E[V(k-1)] \tag{29}$$

Converting $R_{\varphi\varphi}$ to a diagonal form, equation (29) leads to a set of n homogeneous difference equations of the first order whose solution is stable if and only if:

$$-1 < 1 - \mu\lambda_i(R_{\varphi\varphi}) < 1, \qquad \forall i \tag{30}$$

In this case, we have:

$$\lim_{k\to\infty} E[W(k)] = W_{opt} \tag{31}$$

Since the eigenvalues of $R_{\varphi\varphi}$ are all real and positive, it therefore follows that condition (30) implies condition (27).

To prove convergence in the mean square, observe that computing the square of the norm of equation (28), taking the expectation of the result and rearranging, we obtain:

$$\begin{aligned} E[|\, V(k)\,|^2] = {} & E[V^*(k-1)A(k)V(k-1)] \\ & + E[U_1(k)] + E[U_2(k)] + E[U_3(k)] \end{aligned}$$

where the expressions for $A(k)$, $Z_{\varphi\varphi}$,$U_1(k)$, $U_2(k)$ and $U_3(k)$ are easily obtained by expanding $E[|\, V(k)\,|^2]$.
The sum $E[U_1(k)] + E[U_2(k)] + E[U_3(k)]$ is bounded as:

$$E[U_1(k)] + E[U_2(k)] + E[U_3(k)] \leq a\mu^2, \;\; a > 0 \tag{32}$$

As a result, we have:

$$E[|\, V(k)\,|^2] \leq [1 - \rho(R_{\varphi\varphi}, Z_{\varphi\varphi})]E[V(k-1)] + a\mu^2 \tag{33}$$

where $\rho(R_{\varphi\varphi}, Z_{\varphi\varphi})$ is:

$$2\mu\lambda_{min}(R_{\varphi\varphi}) - \mu^2\lambda_{max}(Z_{\varphi\varphi}) \tag{34}$$

Equation (33) is stable if and only if:

$$|\, 1 - 2\mu\lambda_{min}(R_{\varphi\varphi}) + \mu^2\lambda_{max}(Z_{\varphi\varphi})\, | < 1 \tag{35}$$

or equivalently,

$$0 < \mu < \frac{2\lambda_{min}(R_{\varphi\varphi})}{\lambda_{max}(Z_{\varphi\varphi})} \tag{36}$$

Since $Z_{\varphi\varphi}$ is positive semidefinite, we deduce:

$$\lambda^2_{min}(R_{\varphi\varphi}) \leq \lambda^2_{max}(R_{\varphi\varphi}) \leq \lambda^2_{max}(Z_{\varphi\varphi}) \tag{37}$$

Hence, condition (36) implies condition (27). Consequently,

$$\lim_{k\to\infty} E[|\, W(k) - W_{opt}\,|^2] = 0 \tag{38}$$

This ends the proof of theorem 2.

Theorem 2 provides a necessary and sufficient condition for both convergence of the mean and in the mean square as the number of iterations approaches infinity. This condition is remarkably similar to that needed for the conventional LMS algorithm to converge. However, keep in mind here that λ_{max} is the largest eigenvalue of the fourth order moment matrix $R_{\varphi\varphi}$ and not that of the correlation matrix R_{xx} as is the case in the standard LMS setting.

3. EXPERIMENTAL RESULTS

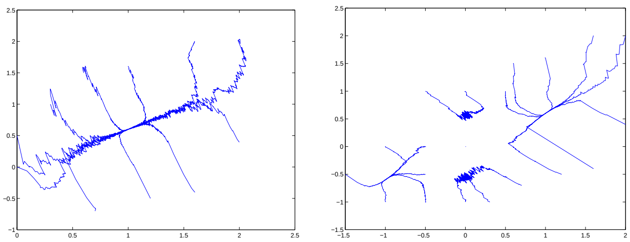
To illustrate the performance of the new algorithm, consider two real valued sequences $\{x_t\}$ and $\{s_t\}$ related as:

$$s_t = h_0 x_t + h_1 x_{t-1} \tag{39}$$

where h_0 and h_1 are unknown scalars.

The sequence $\{x_t\}$ is used to drive a filter with two real valued taps w_0 and w_1. The output $\{y_t\}$ of the filter is the desired estimate of the unknown sequence $\{s_t\}$. At the same time, the parameters w_0 and w_1 are the estimates of the unknown parameters h_0 and h_1, respectively.

In our simulations, we used $h_0 = 1$, $h_1 = 0.6$, $\gamma = 1$, $\mu = 0.005$, $N = 2500$ and a number of initial conditions. The new algorithm converges to a unique solution for all initial conditions as shown in figure 1 (a). In contrast, the CMA algorithm exhibits multiple minima depending on the initial conditions as seen in figure 1 (b). More elaborate examples are given in [6] and show similar behavior.



(a) New Algorithm (b) CMA Algorithm

Fig. 1. Example showing the new algorithm with a single solution when the CMA exhibits multiple minima.

In addition to its computational complexity discussed earlier, the new algorithm also requires a large number of samples to converge. As expected, our simulations showed that the eigenvalue spread of the matrix $R_{\varphi\varphi}$ plays a major role in the speed of convergence of the new algorithm. For low eigenvalue spread, the number of samples required by the new algorithm is the same as that needed by the original CMA. This number is however, larger than that used by the conventional LMS method since both the CMA and the new algorithm are implicitly based on the fourth order statistical moments.

However, we observed also that for the same value of μ, the new algorithm needed as much as 4 times the number of samples used by the original CMA algorithm when the eigenvalue spread is high. Moreover, our simulations revealed also that the type of digital signals typically encountered in the CM minimization setting often lead to large eigenvalue spreads. This observation highlights the importance of data conditioning in practice and may explain, in part, why several researchers have experienced slow convergence when using the CMA algorithm.

4. CONCLUDING REMARKS

This paper has introduced a new LMS algorithm for minimizing the CM performance measure. Unlike the original CMA algorithm, the new algorithm has the advantage of converging to a desired solution only regardless of initial conditions. In addition, the resulting increase in computational complexity is not too excessive. However, data conditioning practices need to be incorporated with the new algorithm in order to improve its speed of convergence.

5. REFERENCES

[1] J.R. Treichler and B.G. Agee, “A New Approach to Multipath Correction of Constant Modulus Signals,” *IEEE Trans. ASSP*, Vol. ASSP-31, No. 2, Apr. 1983.

[2] C. R. Johnson Jr., et al, “Blind Equalization Using the Constant Modulus Criterion: A Review”, *Proc. IEEE*, Special Issue, Oct. 1998.

[3] H. Jamali and S.L. Wood, “Error Surface Analysis for the Complex Constant Modulus Adaptive Algorithm”, *Proc. ACSSC*, Pacific Grove, CA., Nov. 1990.

[4] A. J. Van Der Veen and A. Paulraj “An Analytical Constant Modulus Algorithm”, *IEEE Trans. on Sig. Proc.*, Vol. 44, No. 5, May. 1996.

[5] K. Dogancay and R.A. Kennedy, A globally admissible off-line modulus restoral algorithm for low-order adaptive channel equalisers, in Proc. IEEE ICASSP, (Adelaide), Apr. 1994.

[6] H. Jamali and T. Ogunfunmi, “Stationary Points of the Finite Length Constant Modulus Optimization”, *Signal Processing*, Apr. 2002.

[7] T. Ogunfunmi and H. Jamali, “Constant Modulus Performance Search Using Newton’s Method”, *Proc. ICASSP*, Salt Lake City, UT., Apr. 2001.

[8] G. H. Golub and C. F. Van Loan, “Matrix Computations,” 3^{rd} edition, Baltimore, John Hopkins University Press, 1996.