D3.21

Approaches to Blind Equalization of Signals with Multiple Modulus

William A. Sethares, Gonzalo A. Rey*, C. Richard Johnson, Jr.*

Dept. of Electrical and Computer Engineering, University of Wisconsin, Madison, WI, 53706
*School of Electrical Engineering, Cornell University, Ithaca NY, 14853

Abstract

The Constant Modulus Algorithm (CMA) and Decision Directed (DD) equalizer are two ways to approach blind equalization of signals which are known to lie on a circle of fixed radius, but where specific values at any given time are unknown. In M-ary Quadrature Amplitude Modulation, the signals lie on n circles of known radius. This paper presents two possible approaches to the n-modulus problem, both in the spirit of "feature reconstruction" algorithms. The Multiple Modulus Algorithm uses a straightforward generalization of the CMA cost function to derive its update, while the Decision Adjusted Modulus Algorithm is a hybrid of the CMA and the DD approaches. The algorithms are analyzed and compared in a simple problem setting.

1. Introduction

Blind equalization is a notoriously difficult task. Consider the scenario of figure 1, where a signal y is sent across a channel and a corrupted version x is received. Suppose that the channel can be modeled by the linear autoregression $1/W(q^{-1})$. The task of the blind equalizer is to build the inverse of the channel. *If* $\hat{W}(q^{-1})$ could be made equal to $kq^{-d}W(q^{-1})$, then the source could be adequately reconstructed. But how can such $\hat{W}(q^{-1})$ be built?

If nothing is known about the character of the transmitted signal, then blind equalization is patently impossible. Often, however, the source will have some known property which can be exploited to help determine how the received signal has been corrupted. In the baud synchronous binary phase shift keying problem, for instance, the source consists of just two possible values, +1 and -1. In this case there are two popular methods used to adjust the coefficients of the $\hat{W}(q^{-1})$ polynomial.

The first, the Decision Directed (DD) equalizer (see [2] and [6]) compares the reconstructed signal $\hat{y}$ to the nearest value +1 or -1, and uses this difference to drive the adaptive algorithm. Thus the "decision" to compare $\hat{y}$ to +1 or to -1 "directs" the subsequent evolution of the adaptive filter parameters. The second approach, the Constant Modulus Algorithm (CMA) [1] attempts to minimize the difference between unity and the modulus of the reconstructed signal. The cost function $J_{CMA}(k) = (\hat{y}^2(k) - 1)^2$ is used to define the gradient descent algorithm

$$\hat{W}(k+1) = \hat{W}(k) - \mu \hat{y}(k)(\hat{y}^2(k) - 1) X(k) \qquad (1)$$

where $\hat{W}(k)$ is a vector of the weights of the adaptive filter at time k, $X^T(k)=(x(k), x(k-1),..., x(k-n))$ is a regressor vector of past received signals, and μ is a small positive stepsize. Thus the error between $\hat{y}^2$ and 1 drives the adaptive update. Analysis of the local stability of real CMA can be found in [3]. Throughout this paper we restrict discussion to the 1-D real case instead of the 2-D complex case, with the understanding that the same philosophy applies in the more general setting.

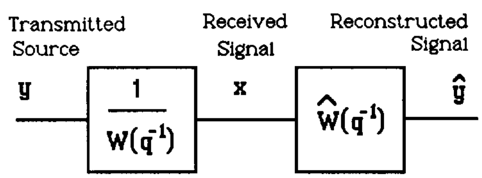

Figure 1 : Blind Equalization

Sometimes the transmitted signal does not have a single constant modulus as in the above example. In M-ary QAM (Quadrature Amplitude Modulation [5]), for instance, the transmitted source can assume a whole constellation of possible values which can generally be considered to lie on several circles of known modulus. The DD strategy can be employed, but as the number of points increases, finer distinctions must be made, and the probability of erroneous decisions increases. CMA can also be applied, but it can only adapt to match a single constant modulus. Straightforward application of the algorithm to the n-modulus problem will cause the parameters to converge to some average location which is a tradeoff between the various moduli.

* supported by NSF grant MIP8608787

Yet the idea of "feature reconstruction" that underlies CMA is enticing. If the source lies on several known circles, why not reconstruct this feature by generalizing the cost function to one which attempts to minimize the error between the reconstructed signal and each of the known moduli? This leads to the Multiple Modulus Algorithm (MMA) of section 3. Alternatively, why not reconstruct this feature by minimizing the error between the reconstructed signal and the nearest of the known moduli - essentially "deciding" at each step which "constant" to use. This leads to the Decision Adjusted Modulus Algorithm (DAMA) of section 4, which combines many of the strengths of CMA with the DD approach.

2. CMA in a Multiple Modulus Setting

Consider a simple test case where the transmitted source y is known to consist of the four real values +1, -1, +3, and -3. The channel model is the simple first order autoregression $\frac{1}{w_1 - w_2 q^{-1}}$ with $w_1=1$, $w_2=-.9$, and the adaptive equalizer contains two adjustable parameters $\hat{w}_1$ and $\hat{w}_2$. The reconstructed signal is calculated as

$$\hat{y}(k) = \hat{w}_1(k)\ x(k) + \hat{w}_2(k)\ x(k-1). \quad (2)$$

Note that if $w_1 = \hat{w}_1$ and $w_2 = \hat{w}_2$, then $\hat{y}(k+1)$ exactly reconstructs y(k). If the ratios $\hat{w}_2/\hat{w}_1$ and w_2/w_1 are equal, this is an admissible solution, since then $\hat{W}(q^{-1}) = kq^{-d}W(q^{-1})$ for some k and d. The job of CMA is to adapt the $\hat{w}_i$, based only on measurements of the x and knowledge that y consists of +/-1 and +/-3 (but without knowing the actual value of y at any particular time).

Suppose that the four y values are chosen independently and uniformly, and apply the CMA update (1). Then the parameter estimates $(\hat{w}_1, \hat{w}_2)$ tend to converge to (0.3489, -0.3149) or its negative, depending on the initialization. The ratio $\hat{w}_2/\hat{w}_1$ is -0.9025, which is the admissible value -0.9 to within experimental accuracy. Suppose, however, that y is chosen in a heavily dependent way, say by the Markov chain detailed in figure 2. Then the CMA updates tend to converge to (0.2988, -0.2544) or its negative, which are not admissible solutions since the value of the ratio is -0.85.

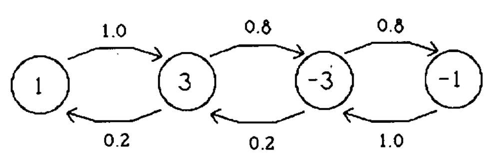

Figure 2: Transition probabilities for example

Such behavior appears to be generic... when the transmitted source is white and uniform, then CMA can converge to an admissible solution. When the source is highly correlated and nonuniform, CMA does not attain an admissible solution. What we desire, then, is an algorithm for which an admissible solution is always a point of local stability, irrespective of the correlation and uniformity of the source. We propose two candidate algorithms.

3. The Multiple Modulus Algorithm

Suppose that the transmitted source lies on n circles of known modulus, where $M_1, \ldots, M_n$ designate the squares of the moduli. The feature reconstruction idea suggests that it might be desirable to try to minimize the cost function

$$J_{MMA}(k)=(\hat{y}^2(k)-M_1)^2(\hat{y}^2(k)-M_2)^2\cdots(\hat{y}^2(k)-M_n)^2. \quad (3)$$

One way to proceed with this minimization is to pursue a gradient descent strategy, updating the parameter estimates $\hat{W}$ in the direction opposite the gradient at each timestep. The MMA adaptive algorithm is then

$$\hat{W}(k+1) = \hat{W}(k) - \mu \frac{\partial J_{MMA}(k)}{\partial \hat{W}(k)}. \quad (4)$$

Since $\hat{y}(k) = X^T(k)\ \hat{W}(k)$, this derivative can be evaluated directly as

$$\frac{\partial J_{MMA}}{\partial \hat{W}} = \hat{y}\ (\hat{y}^2-M_1)\ (\hat{y}^2-M_2) \cdots$$

$$\cdots\ (\hat{y}^2-M_n)\left(\sum_{i=1}^{n} \prod_{j=1\ j\neq i}^{n} (\hat{y}^2-M_j)\right). \quad (5)$$

Due to its form as a gradient minimization of the cost function J_{MMA}, analysis of (4) can proceed similarly to the analysis of CMA in [3] and [4]. The algorithm (4) will stop updating, on average, whenever

$$\mathrm{avg}\left(\frac{\partial J_{MMA}(k)}{\partial \hat{W}(k)}\right)\Bigg|_{\hat{W}(k)=W^*} = 0 \quad (6)$$

where "avg" represents an appropriate time sample average and W* indicates an averaged equilibrium. Moreover, in order for this stationary point W* to be locally exponentially stable we will need

$$\mathrm{avg}\left(\frac{\partial^2 J_{MMA}(k)}{\partial \hat{W}^2(k)}\right)\Bigg|_{\hat{W}(k)=W^*} > \alpha > 0, \quad (7)$$

and a sufficiently small stepsize.

Once local stability is established for a given average equilibrium point, the key question concerns the size of the region of attraction. Since (5) contains high powers of $\hat{y}$, (4n-1, to be exact), one might anticipate numerical problems. Indeed, typical simulations require stepsizes on the order of 10^{-8}. A practical modification, then, is to normalize the stepsize by the largest even power of $\hat{y}$ found in the update term. Such normalization is common in LMS to take into account variations in the size of signals.

For the example of the previous section, n=2, $M_1=1$, and $M_2=9$. Substituting (5) into the gradient algorithm form (4) and normalizing yields the update form

$$\hat{W}(k+1)=\hat{W}(k)-\mu_k\hat{y}(k)(\hat{y}^2(k)-M_1)(\hat{y}^2(k)-M_2)\cdots (2\hat{y}^2(k)-(M_1+M_2))X(k) \qquad (8)$$

where $\mu_k = \mu/(\hat{y}^6(k)+1)$ is the normalized stepsize.

There are several equilibria for (8), which correspond to $\hat{W}$ values that cause $\hat{y}=0$, $\hat{y}^2= M_1$, $\hat{y}^2= M_2$, and $\hat{y}^2= (M_1+M_2)/2$.To examine the stability of these equilibria, we can construct the matrix of partial derivatives as in (7). This gives

$$\begin{aligned}\frac{\partial^2 J_{MMA}}{\partial \hat{W}^2} = &(\hat{y}^2-M_1)\ (\hat{y}^2-M_2)\ (2\hat{y}^2-(M_1+M_2))\ X\,X^T\\ &+ 2\,\hat{y}^2\,(\hat{y}^2-M_2)\ (2\hat{y}^2-(M_1+M_2))\ X\,X^T\\ &+ 2\,\hat{y}^2\,(\hat{y}^2-M_1)\ (2\hat{y}^2-(M_1+M_2))\ X\,X^T\\ &+ 4\hat{y}^2\ (\hat{y}^2-M_1)\ (\hat{y}^2-M_2)\ X\,X^T\end{aligned} \qquad (9)$$

where the time index k has been dropped for conciseness.

Each term in (9) consists of a scalar, dependent on $\hat{y}$, M_1 and M_2 multiplied by the vector outer product XX^T. Since XX^T can itself never be positive definite, it is necessary to consider the average of (9). In particular, if avg(XX^T) > α > 0, (often called the persistence of excitation condition) and if the sum of the scalar multiples in (9) are positive for particular $\hat{y}$, M_1 and M_2, then (7) will hold and the algorithm will be locally exponentially stable about that averaged equilibrium value. For the equilibrium at $\hat{y}=0$ (which corresponds to $\hat{W}=0$), the second, third and fourth scalars of (9) are zero, while the first is $-M_1M_2(M_1+M_2)$. Thus this equilibrium point is unstable, since M_1 and M_2 are positive. The equilibria at $\hat{y}^2=M_1$ corresponds to the "correct" answer at $\hat{W}=W^*$ and to its negative at $\hat{W}=-W^*$. The scalar is $2M_1(M_1-M_2)(2M_1-(M_1+M_2))$, which equals the positive constant $2M_1(M_1-M_2)^2$, indicating stability for both of these answers. A symmetric argument shows that the equilibria at $\hat{y}^2=M_2$ are also stable. The last pair of equilibria, at $\hat{y}^2=(M_1+M_2)/2$ has the scalar

$$4\frac{(M_1+M_2)}{2}\left(\frac{(M_1+M_2)}{2} - M_1\right)\left(\frac{(M_1+M_2)}{2} - M_2\right). \qquad (10)$$

Since exactly one of the last two factors must be negative, these are unstable equilibria.

Using the simple test case outlined above and the MMA algorithm (8), we simulated the adaptation of the coefficients from a variety of different initial locations. In the simulations, the stepsize was μ=.0002, and the four possible values of y were chosen independently from a uniform distribution. This is adequate to ensure that the persistence of excitation condition is fulfilled, and the simulations verify that if the $\hat{W}$ parameters are initialized near the stable equilibrium values, then they tend to converge to the equilibrium, while if they are initialized near one of the unstable equilibria, they tend to move elsewhere. Not only do the parameters converge to (1.0,-0.9) or its negative, but they may also converge to (0.35,-0.32) or its negative. This corresponds to when ($\hat{y}^2$ - M_1) is on average zero, even though it never *equals* zero. All these solutions are admissible.

The second method used to examine the behavior of the algorithm in this example is to plot the "arrow" diagrams as studied in [7]. These are essentially the averaged, discrete-time equivalent of continuous-time flow lines. The circles of figure 3 represent the bases, or starting points, and the lines represent the average direction of parameter motion from the base. Each line is averaged over 3,000 inputs. One may think of figure 3 as approximating the "error surface" over which the MMA algorithm moves. It is not a quadratic surface.

Due to the intricacy of this error surface, and the corresponding possibility of spurious equilibria, we consider an alternative algorithm, which is essentially a hybrid of the DD and CMA approaches.

4. The Decision Adjusted Modulus Algorithm

Once again, suppose that the transmitted source is known to lie on n circles of known radius, and let M_1, ... , M_n designate the squares

$$J_i(k) = (\hat{y}^2 - M_i)^2 \qquad i=1,\ldots,n, \qquad (11)$$

which correspond to the n known radii. The update at each time instant is then

$$\hat{W}(k+1)=\hat{W}(k) - \mu\,\hat{y}\,\min_i(\hat{y}^2 - M_i)^2\,X(k) \qquad (12)$$

where the "min" function compares the reconstructed signal to the nearest M_i at each time k.

In essence, DAMA acts like n CMAs, one for each of the known radii. While CMA is incapable of matching more than one circle, DAMA (like MMA) holds the promise of adapting to all n. Unlike MMA, DAMA does not seem to suffer from a bizarrely shaped error surface. It makes a decision at each timestep and adapts the coefficients towards the nearest circle. Due to the nondifferentiability of the minimum function, it is not trivial to apply the averaging style results to the study of DAMA. On the other hand, the application of averaging requires only a Lipshitz continuity in the algorithm update term, and one can envision an averaging approach based on the conical properties of the derivative of the "min" function.

Figure 4 shows the arrow diagram for the same equalization setup as in the previous section, utilizing the DAMA algorithm in place of MMA. Both algorithms converge to the same correct answer ($\hat{W}$=W*), the same negative answer ($\hat{W}$=-W*), and the same pair of "other" admissible values at (0.35, -0.32) and (-0.35, 0.32), depending on the initialization.

5. Conclusions

This paper provides a starting point from which a discussion of blind equalization of multiple modulus signals can begin. We have presented two possible approaches, one based solely on the feature reconstruction idea as in CMA, and the other a hybrid of CMA and DD. By drawing "arrow" diagrams, we compared the performance and region of attractions of the two algorithms in one simple problem setting. After convergence, the performance of the two algorithms appears to be roughly comparable, but the size of the region of attraction appears to be somewhat larger for DAMA than for MMA. Clearly, a lot of work remains to be done before such adaptive schemes are ready for application.

6. References

[1] J. R. Treichler and B. G. Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Trans. on Acoustics,, Speech and Signal Proc.*, vol. ASSP-31, pp.459-472, April 1983.

[2] J. G. Proakis, "Advances in equalization for intersymbol interference," in *Advances in Communications Systems*, vol. 4, Academic Press, New York, 1975.

[3] C. Richard Johnson, Jr., S. Dasgupta, and W. A. Sethares, "Averaging analysis of local stability of a real constant modulus algorithm adaptive filter," *IEEE Trans. On Acoustics, Speech and Signal Proc.*, vol. ASSP-36, pp. 900-910, June 1988.

[4] Gonzalo Rey, "Local stability analysis of the constant modulus algorithm for adaptive equalization," Thesis, Cornell University, January 1989.

[5] S. Haykin, *Digital Communications*, John Wiley & Sons, Inc. New York, 1988.

[6] O. Macchi and E. Eweda, "Convergence analysis of self adaptive equalizers," *IEEE Trans. on Information Theory*, vol. IT-30, no. 2, March 1984.

[7] C. R. Elevitch, W. A. Sethares, and C. R. Johnson, Jr., "Quiver Diagrams and Signed Adaptive Algorithms," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, Feb. 1989.

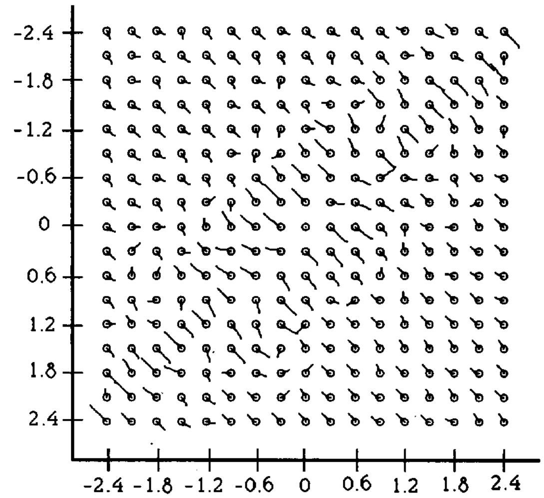

Figure 3: Arrow diagrams for MMA

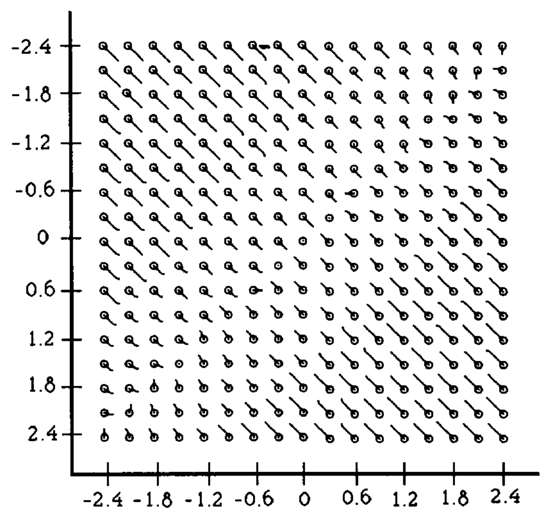

Figure 4: Arrow diagram for DAMA