FIXED WINDOW CONSTANT MODULUS ALGORITHMS

Xiang-yang Zhuang and A. Lee Swindlehurst

Department of Electrical and Computer Engineering
Brigham Young University, Provo, UT 84602
email:{jeffz, swindle}@ee.byu.edu

ABSTRACT

We propose two batch versions of the constant modulus algorithm in which a fixed block of samples is iteratively reused. The convergence rate of the algorithms is shown to be very fast. The delay to which the algorithms converge can be determined if the peak position of the initialized global channel/equalizer response is known. These fixed window CM algorithms are data efficient, computationally inexpensive and no step-size tuning is required. The effect of noise, and the relationship between the converging delay and noise enhancement are analyzed as well.

1. INTRODUCTION

Constant Modulus Algorithms (CMA) have been studied for years since the pioneering work of Godard [3]. Although LMS adaptive implementations of CMA are simple and robust to adverse channels, their drawbacks include convergence to local minima, slow convergence, and unpredictable performance variations due to *ad hoc* initialization. Under the assumption of multiple channels obtained by either oversampling or the use of an antenna array, the CMA approach has been extended to fractionally spaced or spatio-temporal equalization in which an FIR filter is able to achieve perfect equalization. To achieve faster convergence, the use of techniques from classical adaptive filters, such as replacing the training (desired) signal by a nonlinear function of the equalizer output (e.g. a "sgn" function or a tentative decision closest to the constellation), have been found to be effective. Examples of fast CMA implementations are normalized CMA [5], LS-CMA [1], normalized sliding window CMA [2], *etc.* Since the convergence rate and the steady state error associated with each delay to which the equalizer converges may be dramatically different depending on the filter initialization, some re-initialization schemes have also been proposed to seek the globally optimum equalizer [6]. Other recent developments for CMA include its extension to the multiple user case and studies of its behavior with noisy channels.

We propose two variations of CMA here. We call them fixed window CMA (FWCMA) since they re-use a stored fixed-length block of data. Since fourth order cumulants are involved in CMA(2,2), its slow convergence is usually ascribed to the involvement of HOS rather than to the recursive LMS technique employed. The approximation of the gradient by its instantaneous value is very inaccurate, especially when the cost function involves HOS. The algorithms we present demonstrate that if a good approximation of the gradient can be obtained by averaging over a block of data, HOS-based algorithms can be data efficient (30-100 samples are usually enough depending on how good the *i.i.d.* assumption is). The proposed methods iteratively reuse the data block to compute the gradient. Convergence is shown to be quite fast, with only 2-8 iterations required depending on the nature of the channel. The delay to which the algorithm converges can be determined as a valuable by-product if some information about the initialized global system response is available. The iterations require no step size adjustment which is critical to LMS-CMA. We also analyze the behavior for noisy channels, especially the relationship between the noise amplification effect of different delays and the eigenstructure of the channel matrix. Finally, several approaches are proposed for the multi-user case where near-far problems may occur.

2. DATA MODEL

We consider a single user transmitting through multiple channels which are obtained by either oversampling in time or space. The data model can be written as

$$\begin{aligned}\mathbf{x}(n) &= [\mathbf{x}_{MP}(n)\cdots\mathbf{x}_{MP}(n-E+1)]^T \\ &= \begin{bmatrix} \mathbf{h}_0\cdots\mathbf{h}_{L-1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0}\ \mathbf{h}_0 & \cdots & \mathbf{h}_{L-1} & \mathbf{0}\ \ \mathbf{0} \\ \vdots\ \ddots & \ddots & \ddots & \ddots\ \ddots\ \ddots \\ \mathbf{0}\ \ \mathbf{0} & \mathbf{0} & \mathbf{h}_0 & \cdots\mathbf{h}_{L-1} \end{bmatrix} \mathbf{s}(n) + \mathbf{n}(n) \\ &= \mathcal{H}\mathbf{s}(n) + \mathbf{n}(n) \end{aligned} \quad (1)$$

where M is the number of sensors, P is the oversampling factor, $\mathbf{x}_{MP}(n)$ is a vector containing samples from all MP channels at time n, LT is the effective channel length, E is the length of the temporal filter for each subchannel, $\mathbf{s}(n) = [s(n)\cdots s(n-E-L+2)]^T$ and $\mathbf{n}(n) = [\mathbf{n}_{MP}^T(n)\cdots\mathbf{n}_{MP}^T(n-E+1)]^T$.

3. ALGORITHMS

Our problem is to find an equalizer $\mathbf{w}$ that restores the constant modulus property of the signal. Denoting y as the equalizer output $y = \mathbf{w}^H\mathbf{x}$, the CMA(2,2) cost function is:

$$J(\mathbf{w}) = E(1-|y|^2)^2 = E(1-\mathbf{w}^H\mathbf{x}\mathbf{x}^H\mathbf{w})^2\,. \quad (2)$$

We introduce an approximation of the above cost function as follows:

$$J(\mathbf{w}_{k+1}) = E(1-\mathbf{w}_{k+1}^H\mathbf{x}\mathbf{x}^H\mathbf{w}_k)^2 \quad (3)$$

where the subscript k indicates the iteration number and superscript H denotes the complex conjugate transpose. This

approximation will be valid when $\mathbf{w}_k$ is near convergence. Writing $\mathbf{w}_{k+1}$ as $\mathbf{w}_k + \Delta\mathbf{w}_k$, we observe that the minimum of (3) may be achieved by searching for the weight increment $\Delta\mathbf{w}_k$ with minimum norm for which y_{k+1} is constant modulus. This underlying idea is similar to that of the normalized sliding window algorithms.

Denoting $\mathbf{z}_k = \mathbf{x}\mathbf{x}^H\mathbf{w}_k = y_k^*\mathbf{x}$, the Wiener solution to (3) for $\mathbf{w}_{k+1}$ is simply

$$\mathbf{w}_{k+1} = \left[E(\mathbf{z}_k\mathbf{z}_k^H)\right]^{-1} E(\mathbf{z}_k) = R_{zz}^{-1}R_{xx}\mathbf{w}_k \ . \qquad (4)$$

We normalize the composite global system response, $\mathbf{g}_k^H \stackrel{\text{def}}{=} \mathbf{w}_k^H\mathcal{H}$ so that $\|\mathbf{g}_k\|_2^2 = 1$. The reason for this will be clear in the proof of the following theorem. The normalization can be implemented as

$$\mathbf{w}_{k+1} = \mathbf{w}_{k+1}/\sqrt{\mathbf{w}_{k+1}^H\mathcal{H}\mathcal{H}^H\mathbf{w}_{k+1}} \approx \mathbf{w}_{k+1}/\sqrt{\mathbf{w}_{k+1}^H R_{xx}\mathbf{w}_{k+1}} \ . \qquad (5)$$

We use "$\approx$" since R_{xx} asymptotically approaches $\mathcal{H}\mathcal{H}^H$ if the source sequences are assumed to be *i.i.d.*.

Theorem 1: If $\mathcal{H}$ is full column rank and the data sequence is *i.i.d.*, then in the absence of noise the block adaptation rule of (4)-(5) will converge to the global solution in which $\mathbf{g}$ has only one non-zero component. If the i^{th} entry $|g_{0i}|$ of $\mathbf{g}_0$ is the maximum element of the initial global response, then $|g_{0i}|$ will converge to 1 (hence, the equalizer will converge to delay i).

Proof: Note that $R_{zz} = E(\mathbf{x}\mathbf{x}^H|y_k|^2)$, where $y_k = \mathbf{w}_k^H\mathbf{x} = \mathbf{w}_k^H\mathcal{H}\mathbf{s} = \mathbf{g}_k^H\mathbf{s}$. From equation (4), we have

$$\begin{aligned}
E(\mathbf{x}\mathbf{x}^H|y_k|^2)\mathbf{w}_{k+1} &= R_{xx}\mathbf{w}_k \\
\mathcal{H}E(|\mathbf{g}_k^H\mathbf{s}|^2\mathbf{s}\mathbf{s}^H)\mathcal{H}^H\mathbf{w}_{k+1} &= \mathcal{H}\mathcal{H}^H\mathbf{w}_k \\
\mathcal{H}E\{(\sum_{i=1}^{q} g_{ki}^* s_i)(\sum_{i=1}^{q} g_{ki}s_i^*)\mathbf{s}\mathbf{s}^H\}\mathbf{g}_{k+1} &= \mathcal{H}\mathbf{g}_k \\
\mathcal{H}\begin{bmatrix} 1 & g_{k1}g_{k2}^* & \cdots & g_{k1}g_{kq}^* \\ g_{k2}g_{k1}^* & 1 & \cdots & g_{k2}g_{kq}^* \\ \vdots & & \ddots & \vdots \\ \vdots & \cdots & & \vdots \\ g_{kq}g_{k1}^* & \cdots & & 1 \end{bmatrix}\mathbf{g}_{k+1} &= \mathcal{H}\mathbf{g}_k \\
\mathcal{H}R_g\mathbf{g}_{k+1} &= \mathcal{H}\mathbf{g}_k
\end{aligned}$$

where $\mathbf{g}_k = [g_{k1}, \cdots, g_{kq}]^T$ and $q = col(\mathcal{H}) = L + E - 1$ is also the number of delays at which the source can be recovered. We have used the normalization $\sum_{i=1}^{q}|g_{ki}|^2 = 1$, as well as the following HOS property of MPSK (but not BPSK) signals

$$E(s_i s_j^* s_k s_l^*) = \begin{cases} 1 & i = j \text{ and } k = l \\ 0 & \text{otherwise} \ . \end{cases}$$

If $\mathcal{H}$ is full column rank, $\mathbf{g}_{k+1}$ can be uniquely determined as

$$\mathbf{g}_{k+1} = R_g^{-1}\mathbf{g}_k \ ,$$

since

$$R_g = diag(1 - |g_{k1}|^2, \cdots, 1 - |g_{kq}|^2) + \mathbf{g}_k\mathbf{g}_k^H \stackrel{\text{def}}{=} D + \mathbf{g}_k\mathbf{g}_k^H \ .$$

From the matrix inversion lemma, we have

$$R_g^{-1} = D^{-1} - \frac{D^{-1}\mathbf{g}_k\mathbf{g}_k^H D^{-1}}{1 + \mathbf{g}_k^H D^{-1}\mathbf{g}_k}$$

and hence

$$\mathbf{g}_{k+1} = R_g^{-1}\mathbf{g}_k = \frac{D^{-1}\mathbf{g}_k}{1 + \mathbf{g}_k^H D^{-1}\mathbf{g}_k} = c[\frac{g_{k1}}{1 - |g_{k1}|^2}, \cdots, \frac{g_{kq}}{1 - |g_{kq}|^2}]^T$$

where $c = 1/(1 + \mathbf{g}_k^H D^{-1}\mathbf{g}_k)$. The ratio between any pair of the elements of $\mathbf{g}$ is thus

$$\frac{g_{(k+1)i}}{g_{(k+1)j}} = \frac{g_{ki}}{g_{kj}} \cdot \frac{1 - |g_{kj}|^2}{1 - |g_{ki}|^2} \ , \qquad (6)$$

and we observe that after an iteration, the ratios between the peak value of the previous global response and the others increase. If $|g_{0i}|$ is initially the largest entry of $\mathbf{g}_0$, $\mathbf{g}_k$ will converge to the indicator vector $\mathbf{e}_i$ and the convergence speed is super fast. If the initialization $\mathbf{g}_0$ has k exactly equal maximum elements, $\mathbf{g}$ will converge to an undesired equilibrium where those k entries are equal and the rest are zeros. However, these are saddle points and usually will not occur in practice. The saddle point issue coincides with the discussion in [7]. ■

Some observations:

1. In the noiseless case, R_{zz} is rank deficient and the inverse does not exist. However, this does not affect the above proof since R_{zz} is moved to the other side of the equation. In implementing the FWCMA filter in high SNR conditions, the inverse can be computed by an appropriate regularization such as diagonal loading.
2. For real-valued signals (*e.g.*, BPSK), we have

$$E(s_i s_j^* s_k s_l^*) = \begin{cases} 1 & i = j \text{ and } k = l \\ 1 & i = k \text{ and } j = l \\ 0 & \text{otherwise} \ . \end{cases} \qquad (7)$$

It is easy to show that in this case, the off-diagonal entries of R_g are all doubled in value. Therefore, the relationship between the elements of $\mathbf{g}_k$ pre- and post-iteration is

$$\frac{g_{(k+1)i}}{g_{(k+1)j}} = \frac{g_{ki}}{g_{kj}} \cdot \frac{1 - 2|g_{kj}|^2}{1 - 2|g_{ki}|^2} \ ,$$

and negative values appear when $|g_{ki}|^2 > 1/2$. One simple remedy is to modify the normalization to $\|\mathbf{g}_k\| = 2$. Another way will be discussed later.

In the following, we re-derive the above rule from a stochastic gradient viewpoint, and obtain a simpler version of FWCMA. We first look at the gradient of the CMA(2,2) cost function which is

$$\begin{aligned}
\nabla J \stackrel{def}{=} \frac{\partial J}{\partial \mathbf{w}^*} &= E\left\{2(|y|^2 - 1)\frac{\partial(\mathbf{w}^H\mathbf{x}\mathbf{x}^H\mathbf{w})}{\partial\mathbf{w}^*}\right\} \\
&= 4E\{(|y|^2 - 1)\mathbf{x}\mathbf{x}^H\mathbf{w}\} \\
&= 4E(|y|^2\mathbf{x}\mathbf{x}^H)\mathbf{w} - 4E(\mathbf{x}\mathbf{x}^H)\mathbf{w} = 4R_{zz}\mathbf{w} - 4R_{xx}\mathbf{w} \ .
\end{aligned}$$

The stochastic steepest-descent update is of the form

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \frac{1}{2}\mu\nabla J \ . \qquad (8)$$

If μ is a constant (or normalized) scalar step size, we get the (normalized) LMS-CMA. If μ is taken to be $\mu = \frac{1}{2}R_{zz}^{-1}$, we obtain (4). If $\mathbf{z}$ is treated as the desired training signal and the update is adaptively processed symbol-by-symbol, this is actually the RLS algorithm. But ours is a batch

iteration rule. The previous algorithm (we call it FWCMA-1) requires the inverse computation R_{zz}^{-1} at every iteration, but if we replace R_{zz}^{-1} by R_{xx}^{-1}, just one inverse is required. Indeed, R_{xx} is a good approximation to R_{zz} when $\mathbf{w}$ is near convergence. So, taking $\mu = \frac{1}{2}R_{xx}^{-1}$ into (8), we obtain FWCMA-2:

$$\begin{aligned} \mathbf{w}_{k+1} &= 2\mathbf{w}_k - R_{xx}^{-1}R_{zz}\mathbf{w}_k \\ \mathbf{w}_{k+1} &\approx \mathbf{w}_{k+1}/\sqrt{\mathbf{w}_{k+1}^H R_{xx}\mathbf{w}_{k+1}} \end{aligned} \quad (9)$$

Theorem 2: If $\mathcal{H}$ is full column rank and the data sequence is *i.i.d.*, then in the absence of noise the block adaptation rule of (9) will converge to the global solution in which $\mathbf{g}$ has only one non-zero component. If $|g_{0i}|$ is the maximum element of the initial response $\mathbf{g}_0$, this entry will converge to 1 (hence, the equalizer will converge to delay i).

Proof: Taking the conjugate transpose of the above equation and right multiplying by $\mathcal{H}$ on both sides, we have

$$\begin{aligned} l\mathbf{g}_{k+1}^T &= 2\mathbf{g}_k^T - E(|\mathbf{g}_k^T\mathbf{s}|^2\mathbf{g}_k^T\mathbf{s}\mathbf{s}^H)\mathcal{H}^H R_{xx}^{-1}\mathcal{H} \\ &= 2\mathbf{g}_k^T - [g_{k1} + \sum_{i\neq 1} g_{k1}|g_{ki}|^2, \cdots, g_{kq} + \sum_{i\neq 1} g_{kq}|g_{ki}|^2] \\ &= [g_{k1}|g_{k1}|^2, \cdots, g_{kq}|g_{kq}|^2] \end{aligned}$$

The ratio between any pair of the elements of $\mathbf{g}$ is $\frac{g_{(k+1)i}}{g_{(k+1)j}} = \frac{g_{ki}|g_{ki}|^2}{g_{kj}|g_{kj}|^2}$. If $|g_{0i}|$ is the maximum element of the initialized response $\mathbf{g}_0$, this entry $|g_{0i}|$ will converge to 1. As discussed previously, R_{xx} is rank deficient in the noise-free case, but diagonal loading or a pseudo-inverse could be used. In the latter case, when $\mathcal{H}$ is a full column rank matrix, we have $\mathcal{H}^H R_{xx}^{\dagger}\mathcal{H} \approx \mathcal{H}^H(\mathcal{H}\mathcal{H}^H)^{\dagger}\mathcal{H} = I$, which is required in the proof. ■

If we write $\mu = \alpha R_{zz}^{-1}$ (or αR_{xx}^{-1}), it can be shown that the range $0 < \alpha \leq 1$ leads to convergence for both algorithms, while convergence is fastest for $\alpha \to 1$. BPSK signals require $0 < \alpha \leq 1/2$, as per the discussion above.

4. NOISE ANALYSIS

CMA(2,2) minimizes the modulus variation of the output. The convergence of this intuitive cost function is proved in the important work of Shalvi and Weinstein [7], in which CMA(2,2) is shown to be equivalent to maximizing the kurtosis under the constraint of uncorrelated outputs. The convergence proof requires an *i.i.d.* source sequence, so although CMA(2,2) appears to use the modulus property of signal, it is actually attempting to restore the original distribution of source. Indeed, restoring the source distribution is the universal principle of blind equalization. This explains why CMA is also successful in its extension to non-CM signals. The equivalence of CMA(2,2) and Shalvi and Weinstein's criterion suggest that we can use the Shalvi-Weinstein criterion for our analysis. If Gaussian noise is assumed, the noise mainly affects the second order statistical constraint, as follows:

$$\begin{aligned} 1 &= \mathbf{w}^H(\mathcal{H}\mathcal{H}^H + \sigma_n^2 I)\mathbf{w} \\ &= \mathbf{w}^H U diag(\lambda_1^2 + \sigma_n^2, \cdots, \lambda_q^2 + \sigma_n^2, \sigma_n^2, \cdots, \sigma_n^2)U^H\mathbf{w} \\ &= |f_1|^2(\lambda_1^2 + \sigma_n^2) + \cdots + |f_q|^2(\lambda_q^2 + \sigma_n^2) + \sigma_n^2 \sum_{i=q+1}^{m} |f_i|^2 \\ &= |f_1|^2\lambda_1^2 + \cdots + |f_q|^2\lambda_q^2 + \sigma_n^2\|\mathbf{w}\|^2 \end{aligned}$$

where $\mathbf{f}^T = \mathbf{w}^H U$, $\mathcal{H} = U_{m\times m}\Sigma_{m\times q}V_{q\times q}$ is the SVD of $\mathcal{H}$, and $m = row(\mathcal{H}), q = col(\mathcal{H})$. The singular values of $\mathcal{H}$ are $\lambda_1, \cdots, \lambda_q$. The equalizer output is given by

$$\begin{aligned} \mathbf{w}^H\mathbf{x} &= \mathbf{w}^H\mathcal{H}\mathbf{s} + \mathbf{w}^H\mathbf{n} \\ &= \mathbf{w}^H U\Sigma V\mathbf{s} + \mathbf{w}^H\mathbf{n} \\ &= \mathbf{f}^H\Sigma V\mathbf{s} + \mathbf{w}^H\mathbf{n} \\ &= [\lambda_1 f_1, \cdots, \lambda_q f_q]V\mathbf{s} + \mathbf{w}^H\mathbf{n}\,. \end{aligned}$$

First of all, the squared norm of the global response is

$$\begin{aligned} \|\mathbf{g}\|^2 &= \|[\lambda_1 f_1, \cdots, \lambda_q f_q]V\|^2 = \lambda_1^2|f_1|^2 + \cdots + \lambda_q^2|f_q|^2 \\ &= 1 - \sigma_n^2\|\mathbf{w}\|^2 \stackrel{def}{=} \eta^2 \end{aligned} \quad (10)$$

which is less than 1 when noise is present. Secondly, to cancel ISI, we need to compensate for the unitary matrix V which requires $[\lambda_1 f_1, \cdots, \lambda_q f_q] = \eta\mathbf{v}_i^H, (1 \leq i \leq q)$, where $V = [\mathbf{v}_1 \ldots \mathbf{v}_q]$. This step involves HOS and the Gaussian noise effect is negligible. So, in effect, the noise perturbs the global response around the noise-free norm ($\|\mathbf{g}\|^2 = 1$). This perturbation is similar to that shown in the results of [4].

The output SNR for delay i is

$$\begin{aligned} SNR_i &= \frac{E|\mathbf{w}^T H\mathbf{s}|^2}{E|\mathbf{w}^T\mathbf{n}|^2} = \frac{\lambda_1^2|f_1|^2 + \cdots + \lambda_q^2|f_q|^2}{\sigma_n^2\|\mathbf{w}\|^2} = \frac{\eta^2}{\sigma_n^2\sum_{i=1}^{m}|f_i|^2} \\ &= \frac{1}{\sigma_n^2(|\frac{v_{1i}}{\lambda_1}|^2 + \cdots + |\frac{v_{qi}}{\lambda_q}|^2 + \sum_{i=q+1}^{m}|f_i|^2/\eta^2)}. \end{aligned}$$

Obviously, we want η^2 (≤ 1) to be as big as possible, which is equivalent to making $\sigma_n^2\|\mathbf{w}\|^2$ as small as possible. If V is an identity matrix (*i.e.*, $\mathcal{H}$ is a matrix with orthogonal rows), the output SNR can attain its approximate minimum $\frac{\lambda_{min}^2}{\sigma_n^2}$ and maximum $\frac{\lambda_{max}^2}{\sigma_n^2}$. We can also see how the SNR is determined by the column entries of V and the singular values of $\mathcal{H}$.

Since $m\lambda_{max} \geq ||H||_F$, then for a channel oversampled by, for example, a factor of two, $||H||_F^2 = m(\|\vec{h}_1\|^2 + \|\vec{h}_2|^2)/2$, and we have $SNR_{max} \geq SNR_{in}(\|\vec{h}_1\|^2 + \|\vec{h}_2|^2)/2$; *i.e.*, the upper bound is bigger than half of the theoretical maximum SNR which is obtained by two filters ideally matched to the two sub-channels.

5. MULTIPLE USER CASE

With d users, the model becomes

$$\mathbf{x}(n) = [\mathcal{H}_1 \cdots \mathcal{H}_d]\begin{bmatrix} \mathbf{s}_1(n) \\ \vdots \\ \mathbf{s}_d(n) \end{bmatrix} + \mathbf{n}(n) = \mathcal{H}\mathbf{s} + \mathbf{n}(n) \quad (11)$$

where $\mathcal{H} \in \mathcal{C}^{MPE\times d(E+L-1)}$ is assumed to be full rank, and assuming equal channel lengths for each user. From Theorem 1, we notice that the initialization $\mathbf{w}_0 = \mathbf{e}_i$ (*i.e.*, $\mathbf{g}_0 = \mathcal{H}(i,:)$) will recover $s(n-j+1)$, where j is the position of the peak value in the i^{th} row of $\mathcal{H}$. When there are multiple users, it is difficult to control the convergence to the desired user (or delay) from some easy initialization, especially when some users are much stronger than others. Since all amplitude factors are absorbed into the $\mathcal{H}$ matrix, a simple spike initialization will only recover the stronger signals.

We propose three ways to overcome this problem. The first is to use prewhitening, which amounts to transforming

$\mathcal{H}$ to a unitary matrix V (all the users are then of equal "power"). Two drawbacks of this approach are the computational burden incurred by such a step, and the possible increase in convergence time due to the fact that the peak values of V may not be as large in relation to other matrix entries. We also lose control over which user the algorithm may converge to. The second approach is to project $\mathbf{w}$ onto the space orthogonal to that spanned by previously obtained $\mathbf{w}$'s, since the set of $\mathbf{w}$ vectors for different users (or delays) should be linearly independent. To prevent previous errors in $\mathbf{w}$ from affecting the next one, this projection step can be omitted after several iterations. We have occasionally observed an increase in convergence time due to this projection step. The third approach is to successively recover sources beginning with the strongest one, as in the multistage CMA approach. The column of $\mathcal{H}$ corresponding to a given user at some delay can be estimated as

$$\hat{\mathcal{H}}(:,i) = E(X\hat{\mathbf{s}}^*) = E(\mathcal{H}S\hat{\mathbf{s}}^*) \qquad (12)$$

where $\hat{\mathbf{s}}$ is the sequence obtained from some spike initialization . The effect of this strong component can be suppressed by subtracting it from the data: $X - \hat{H}(:,i)\hat{\mathbf{s}}^T$. We then proceed to recover the next strongest user (or delay) from the same initialization until all users are resolved. We can also initialize with an identity matrix and resolve all strong user(s) and delay(s). This option also has the advantage of being able to choose the "best" delay and the processing is completely parallel. We have found that this approach is better than the other two.

6. SIMULATIONS

The single user case is simulated using FWCMA-2 (results for FWCMA-1 are similar) for a three-ray channel whose impulse response is truncated by a window of length 4T. The simulation parameters were $M = 1$ sensor, an oversampling factor of $P = 2$, and a temporal equalizer tap length of $E = 4$. Thus $\mathcal{H} \in \mathcal{C}^{8\times 7}$, and the particular $\mathcal{H}$ used had a condition number of $cond(\mathcal{H}) \cong 22.1$. The modulus of the coefficients of the global response are plotted in Figure 1 for $SNR = 30$dB and 8dB. A block of $N = 100$ samples was used and the results were averaged over 100 Monte-Carlo trials. We initialize with an identity matrix $I_{8\times 8}$, which amounts to initializing the global response with the eight rows of $\mathcal{H}$. We also plot in Figure 2 the averaged modulus error (AME) performance corresponding to these eight initializations (only the $SNR = 30$dB case is shown, the $SNR = 8$dB case is very similar).

We make the following observations. First, the delays to which the algorithm converges correspond to the column positions of the peak values in those rows of $\mathcal{H}$. Second, the solutions for different delays have different steady-state AME. Only "good" delays (2-5 in this case, 1,6,7 are "bad" solutions) are resolved. Third, the convergence speed depends on how the peak value of the initialized global response stands out from the others (the left and right subplots in Figure 2 show convergence for the same delays with different initializations). Finally, the noise perturbation of the global response verifies our theoretical analysis (the peak of the global response is less than 1).

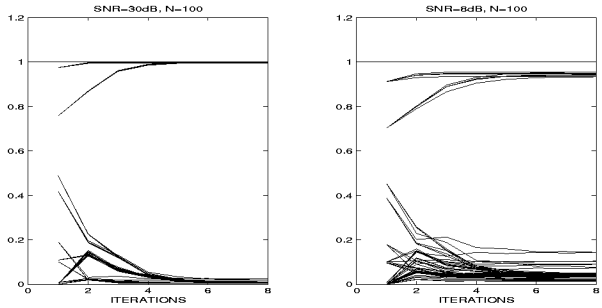

Figure 1: Modulus of the global response coefficients (M=1, P=2,E=4, SNR=30 and 8dB)

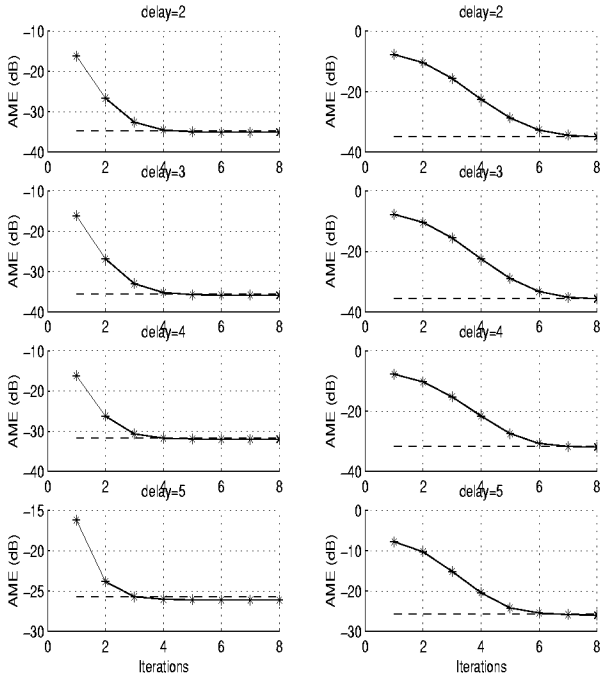

Figure 2: AME performance and converging delays with spike initializations. (Dashed lines represent AME of training-based MMSE filters at corresponding delays)

7. REFERENCES

[1] B.G. Agee. "The Least-Squares CMA: A New Technique for Rapid Correction of Constant Modulus Signals". *IEEE ICASSP*, pages 953–956, April 1986.

[2] C.B. Papadias and A. Paulraj. "A Space-Time Constant Modulus Algorithm for SDMA Systems". *IEEE Signal Processing Letters*, pages 178–181, June 1997.

[3] Dominique N. Godard. "self-recovering equalization and carrier tracking in two-dimensional data communication systems". *IEEE Trans. on Comm.*, pages 1867–1875, November 1980.

[4] I. Fijalkow, A. Touzni, and J.R. Treichler. "Fractionally-spaced Equaliztion by CMA: Robustness to Channel Noise and Lack of Disparity". *IEEE Transaction on signal processing*, pages 56–66, January 1997.

[5] K. Hilal and P. Duhamel. "A Convergence Study of the Constant Modulus Algorithm Leading to a Normalized-CMA and a Block-Normalized-CMA". *Proc. EUSIPCO*, pages 135–138, August 1992.

[6] Lang Tong and Hanks H. Zeng . "Channel-surfing Reintialization of CMA". *Proc. ICASSP*, pages 45–48, April 1997.

[7] Ofir Shalvi and Ehud Weinstein. "New Criteria for Blind Deconvolution of Nonminimum Phase Systems (Channels)". *IEEE Trans. on Information Theory*, pages 312–321, March 1990.