A Cost Function Level Analysis of Autocorrelation Minimization Based Blind Adaptive Channel Shorteners

(Invited Paper)

Ciira wa Maina
Dept. of Elec. and Comp. Eng.
Drexel University
Philadelphia, PA 19104
Email: cm527@drexel.edu

John MacLaren Walsh
Dept. of Elec. and Comp. Eng.
Drexel University
Philadelphia, PA 19104
Email: jwalsh@coe.drexel.edu

Abstract— This paper considers a cost function level analysis of the Sum-squared Autocorrelation Minimization (SAM) channel shortening algorithm. We point out that the actual cost the blind adaptive stochastic gradient descent algorithm is minimizing is only indirectly related to the sum squared autocorrelation. We study the asymptotic regimes under which the actual cost yields a reliable surrogate for the sum squared autocorrelation. We investigate the relationship between the minima of the actual cost and sum squared autocorrelation. We also study the upper bound of the approximate cost as a function of the window size used in the approximate autocorrelation calculation.

I. INTRODUCTION

Channel shortening is important in both single and multi-carrier communication systems. In the former it reduces the complexity of maximum likelihood sequence detection [1] while in the latter it reduces inter-carrier and inter-symbol interference (ICI and ISI) [2].

In multicarrier systems, a cyclic prefix (CP) is prepended to each data block before transmission to combat the delay spread of the channel. If the length of the cyclic prefix is ν and the length of the channel is $\leq \nu + 1$, then no ISI or ICI occurs.

To combat the effects of ISI and/or ICI, a channel shortener can be employed at the front-end of the receiver to ensure the effective channel length $\leq \nu + 1$.

II. SYSTEM MODEL

The input signal $x(n)$ is transmitted over a channel modeled as an FIR filter $\mathbf{h} = [h_0, \ldots, h_{L_h}]^T$. The received signal is $r(n) = x(n) * h(n) + u(n)$ where $u(n)$ is additive noise. $r(n)$ is passed through a channel shortener $\mathbf{w} = [w_0, \ldots, w_{L_w}]^T$ to yield the output $y(n) = r(n) * w(n)$. The taps of the channel shortener are adapted to ensure that the significant nonzero taps of the effective channel are within a contiguous block of length $\nu + 1$.

We make the following assumptions:

A1 The input signal $x(n)$ is zero mean, i.i.d and with unit variance.

A2 The noise $u(n)$ and input signal are statistically independent.

A3 The noise is zero mean, i.i.d with variance σ_u^2.

III. SUM-SQUARED AUTOCORRELATION MINIMIZATION

The Sum-squared Autocorrelation Minimization (SAM) algorithm is a blind adaptive channel shortening algorithm that shortens the channel by minimizing the autocorrelation of the output signal outside a desired window [3]. Let $\mathbf{c} = \mathbf{h} * \mathbf{w} = [c_0, \ldots, c_{L_c}]^T$ be the effective channel response. The autocorrelation of $\mathbf{c}$ is given by

$$R_{cc}(l) = \sum_{k=0}^{L_c} c(k)c(k+l). \tag{1}$$

If the channel coefficients are zero outside a window of length $\nu + 1$, then

$$R_{cc}(l) = 0 \quad \text{for} \quad |l| \geq \nu + 1 \tag{2}$$

This motivates the choice of the SAM cost function as

$$J_{SAM} = \sum_{l \geq \nu+1} |R_{cc}(l)|^2. \tag{3}$$

If J_{SAM} is minimized to zero, then the significant non-zero coefficients of the channel response are within a contiguous block of length $\nu + 1$.

Since $\mathbf{h}$ is unknown, we can not use (3) directly. It can be shown that if $y(n)$ is the shortener output,

$$R_{yy}(l) = R_{cc}(l) + \sigma_u^2 R_{ww}(l). \tag{4}$$

Where σ_u^2 is the variance of the channel noise and $R_{ww}(l)$ is the autocorrelation of the shortener $\mathbf{w}$ at lag l [3]. In the absence of channel noise we have $R_{yy}(l) = R_{cc}(l)$. Even in the presence of noise, we can use the following approximate cost function.

$$\hat{J}_{SAM} = \sum_{l \geq \nu+1} |R_{yy}(l)|^2. \tag{5}$$

The adaptive algorithm finds the shortener which minimizes (5) by stochastic gradient descent over the cost surface. The shortener coefficients are updated according to

$$\mathbf{w}^{new} = \mathbf{w}^{old} - \mu \nabla_{\mathbf{w}}[\hat{J}_{SAM}]. \tag{6}$$

To prevent the trivial $\mathbf{w} = \mathbf{0}$ solution, we impose a unit norm constraint on the shortener or the effective channel, that is $||\mathbf{w}||_2 = 1$ or $||\mathbf{c}||_2 = 1$. The unit norm shortener constraint is imposed by dividing the updated shortener by its norm after every iteration. If the source is assumed white, then $||\mathbf{c}||_2^2 \approx \mathbb{E}[y^2(n)]$ and we can impose the $||\mathbf{c}||_2^2 = 1$ constraint by monitoring the energy of the output [3].

In implementation, the expectation is approximated by a moving average over a window of length N as follows

$$\hat{J}_{SAM}^{(k)} = \sum_{l \geq \nu+1} \left\{ \sum_{n=kN}^{(k+1)N-1} \frac{y(n)y(n+l)}{N} \right\}^2. \tag{7}$$

The update equation is given by

$$\mathbf{w}^{k+1} = \mathbf{w}^k - \mu \nabla_{\mathbf{w}}[\hat{J}_{SAM}^{(k)}] \tag{8}$$

which can be considered to be a stochastic gradient descent on $\mathbb{E}[\hat{J}_{SAM}^{(k)}]$.

IV. Cost Function Analysis

Since N is a design parameter, it is important to quantify the performance of the algorithm as a function of N. In particular, it is important to investigate how the difference between $\mathbb{E}[\hat{J}_{SAM}^{(k)}]$ (the actual cost over which the gradient descent is performed) and $\sum_{l \geq \nu+1} |R_{cc}(l)|^2$ is related to the window size. This will allow users of SAM to determine how large N must be for good performance. Also, since $\hat{J}_{SAM}^{(k)}$ is obtained by first approximating $R_{cc}(l)$ by $R_{yy}(l)$ and then approximating the computation of $R_{yy}(l)$, it is important to study the validity of this approximation.

Towards this end we investigate the quantity $\mathbb{E}[\hat{J}_{SAM}^{(k)}]$. We have

$$\mathbb{E}[\hat{J}_{SAM}^{(k)}] = \frac{1}{N^2}\mathbb{E}\left\{ \sum_{l \geq \nu+1} \sum_{n=kN}^{(k+1)N-1} \sum_{m=kN}^{(k+1)N-1} [y(n)y(m) \right.$$
$$\left. \times y(n+l)y(m+l)] \right\}$$
$$= \frac{1}{N^2}\left\{ \sum_{l \geq \nu+1} \sum_{n=kN}^{(k+1)N-1} \sum_{m=kN}^{(k+1)N-1} \mathbb{E}[y(n)y(m) \right.$$
$$\left. \times y(n+l)y(m+l)] \right\} \tag{9}$$

Based on the system model, the ouput $y(n)$ can be written as the sum of two terms, one due to the input and the other due to noise. We have

$$y(n) = x(n) * c(n) + u(n) * w(n) = y_x(n) + y_u(n)$$

where $c(n)$ is the effective channel impulse response and $w(n)$ is the impulse response of the shortener.

We have

$$\mathbb{E}[y(k)y(l)y(m)y(n)] = R_{yyyy}(k,l,m,n)$$
$$= \mathbb{E}[(y_x(k) + y_u(k))(y_x(l) + y_u(l))$$
$$\times (y_x(m) + y_u(m))(y_x(n) + y_u(n))] \tag{10}$$

Which simplifies to (we denote $R_{yyyy}(k,l,m,n)$ by $R_y^4(k,l,m,n)$)

$$R_y^4(k,l,m,n) = R_{y_x}^4(k,l,m,n)$$
$$+R_{y_xy_x}(k,l)R_{y_uy_u}(m,n) + R_{y_xy_x}(k,n)R_{y_uy_u}(l,m)$$
$$+R_{y_xy_x}(l,m)R_{y_uy_u}(k,n) + R_{y_xy_x}(k,m)R_{y_uy_u}(l,n)$$
$$+R_{y_xy_x}(l,n)R_{y_uy_u}(k,m) + R_{y_xy_x}(m,n)R_{y_uy_u}(k,l)$$
$$+R_{y_u}^4(k,l,m,n) \tag{11}$$

where we have made use of assumptions A1, A2 and A3.

The moments appearing in (11) can be determined from the corresponding moments of the input and the noise [4, p. 405]. We have

$$R_{y_x}^4(k,l,m,n) = \sum_{p=0}^{L_c}\sum_{q=0}^{L_c}\sum_{r=0}^{L_c}\sum_{s=0}^{L_c}\left\{ c(p)c(q) \right.$$
$$\left. \times c(r)c(s)R_x^4(k-p,l-q,m-r,n-s) \right\} \tag{12}$$

and

$$R_{y_u}^4(k,l,m,n) = \sum_{p=0}^{L_w}\sum_{q=0}^{L_w}\sum_{r=0}^{L_w}\sum_{s=0}^{L_w}\left\{ w(p)w(q) \right.$$
$$\left. \times w(r)w(s)R_u^4(k-p,l-q,m-r,n-s) \right\}. \tag{13}$$

The second order moments in (11) take the form

$$R_{y_xy_x}(k,l) = \sum_{p=0}^{L_c}\sum_{q=0}^{L_c} c(p)c(q)R_{xx}(k-p,l-q) \tag{14}$$

and

$$R_{y_uy_u}(k,l) = \sum_{p=0}^{L_w}\sum_{q=0}^{L_w} w(p)w(q)R_{uu}(k-p,l-q). \tag{15}$$

A. White and wide sense stationary input

If the input is white and wide sense stationary (WSS) then

$$R_x^4 = \begin{cases} m_4 & k=l=m=n \\ \sigma_x^4 & k=l, m=n \text{ or } k=m, l=n \text{ or } k=n, \\ & l=m \text{ and } \delta[k-l]\delta[l-m]\delta[m-n] \neq 1 \end{cases}$$

where $m_4 = \mathbb{E}[x^4(k)]$.

From (12) we get

$$R^4_{y_x}(k,l,m,n) = (m_4 - 3\sigma_x^4)\times$$

$$\left\{ \sum_{s=0}^{L_c} c(s-\tau_3)c(\tau_1-\tau_3+s)c(\tau_2-\tau_3+s)c(s) \right\}$$

$$+\sigma_x^4\left\{ \sum_{p=0}^{L_c}\sum_{r=0}^{L_c} c(p)c(r)c(p+\tau_1)c(r+\tau_3-\tau_2) \right.$$

$$+\sum_{p=0}^{L_c}\sum_{q=0}^{L_c} c(p)c(q)c(p+\tau_2)c(q+\tau_3-\tau_1)$$

$$\left. +\sum_{p=0}^{L_c}\sum_{q=0}^{L_c} c(p)c(q)c(p+\tau_3)c(q+\tau_2-\tau_1) \right\} \quad (16)$$

where $\tau_1 = l-k, \tau_2 = m-k, \tau_3 = n-k$.

Also

$$R_{y_x y_x}(k,l) = \sigma_x^2 \sum_{p=0}^{L_w} c(p)c(p+\tau) \quad \tau = l-k \quad (17)$$

B. White and Gaussian noise

If the noise is white and Gaussian, we have

$$R_{y_u y_u}(k,l) = \sigma_u^2 \sum_{p=0}^{L_w} w(p)w(p+\tau) \quad \tau = l-k \quad (18)$$

and

$$R^4_{y_u}(k,l,m,n) =$$

$$\sigma_u^4\left\{ \sum_{p=0}^{L_w}\sum_{r=0}^{L_w} w(p)w(r)w(p+\tau_1)w(r+\tau_3-\tau_2) \right.$$

$$+\sum_{p=0}^{L_w}\sum_{q=0}^{L_w} w(p)w(q)w(p+\tau_2)w(q+\tau_3-\tau_1)$$

$$\left. +\sum_{p=0}^{L_w}\sum_{q=0}^{L_w} w(p)w(q)w(p+\tau_3)w(q+\tau_2-\tau_1) \right\} \quad (19)$$

where $\tau_1 = l-k, \tau_2 = m-k, \tau_3 = n-k$.

Returning to the expression for $\mathbb{E}[\hat{J}^{(k)}_{SAM}]$ and assuming that there is no noise and the input is Gaussian with variance 1 we can write

$$\mathbb{E}[\hat{J}^{(k)}_{SAM}] = \frac{1}{N^2} \sum_{l\geq\nu+1} \sum_{n=kN}^{(k+1)N-1} \sum_{m=kN}^{(k+1)N-1}$$

$$\left\{ \sum_{p=0}^{L_c}\sum_{r=0}^{L_c} c(p)c(r)c(p+\lambda_1)c(r+\lambda_1) \right.$$

$$+\sum_{p=0}^{L_c}\sum_{q=0}^{L_c} c(p)c(q)c(p+l)c(q+l)$$

$$\left. +\sum_{p=0}^{L_c}\sum_{q=0}^{L_c} c(p)c(q)c(p+\lambda_1+l)c(q-\lambda_1+l) \right\} \quad (20)$$

where $\lambda_1 = m-n$. (20) can be simplified to

$$\mathbb{E}[\hat{J}^{(k)}_{SAM}] = \sum_{l\geq\nu+1} |R_{cc}(l)|^2$$

$$+\frac{1}{N^2} \sum_{l\geq\nu+1} \left\{ \sum_{n=kN}^{(k+1)N-1} \sum_{m=kN}^{(k+1)N-1} |R_{cc}(m-n)|^2 \right.$$

$$\left. + \sum_{n=kN}^{(k+1)N-1} \sum_{m=kN}^{(k+1)N-1} R_{cc}(l+\lambda_1)R_{cc}(l-\lambda_1) \right\}. \quad (21)$$

Let $\mathbf{r} = [R_{cc}(0), R_{cc}(1), \ldots, R_{cc}(L_c)]^T$ and assume $N \geq L_c + 1$. Then

$$\frac{1}{N^2} \sum_{l\geq\nu+1} \left\{ \sum_{n=kN}^{(k+1)N-1} \sum_{m=kN}^{(k+1)N-1} |R_{cc}(m-n)|^2 \right\}$$

$$= \mathbf{r}^T \mathbf{A r}.$$

where $\mathbf{A} = \text{diag}\left(\frac{(L_c-\nu)}{N}, \frac{2(L_c-\nu)(N-1)}{N^2}, \ldots, \frac{2(L_c-\nu)(N-L_c)}{N^2}\right)$.

Also

$$\frac{1}{N^2} \sum_{l\geq\nu+1} \left\{ \sum_{n=kN}^{(k+1)N-1} \sum_{m=kN}^{(k+1)N-1} R_{cc}(l+\lambda_1)R_{cc}(l-\lambda_1) \right\}$$

$$= \mathbf{r}^T \mathbf{B r}$$

where

$$\mathbf{B} = \sum_{l=\nu+1}^{L_c} \mathbf{B}_l$$

with $\mathbf{B}_l$ being the $L_c + 1$ by $L_c + 1$ matrix whose entries are all zero except[1]

$$b_{l+1,l+1} = \frac{1}{N}$$

$$b_{l+i+1,|l-i|+1} = \frac{2(N-i)}{N^2} \quad i = 1, 2, \ldots, L_c - l$$

Therefore we can write

$$\mathbb{E}[\hat{J}^{(k)}_{SAM}] = J_{SAM} + \mathbf{r}^T \mathbf{C r} \quad (22)$$

where $\mathbf{C} = \mathbf{A} + \mathbf{B}$ and the first term on the right hand side of (21) is recognized as $J_{SAM}(3)$.

C. Illustrative example of the cost surface

In order to investigate the minima of (21) as a function of N relative to the minima of J_{SAM}, we plot the cost surface for various values of N. The channel in this example[2] is given by $\mathbf{h} = [1, 0.3, 0.2]^T$ which we desire to shorten to two taps using a three tap shortener $\mathbf{w} = [w_0, w_1, w_2]^T$. If we impose the constraint $||\mathbf{w}|| = 1$, we can represent the shortener taps in spherical coordinates as $w_0 = \sin(\phi)\cos(\theta)$, $w_1 = \cos(\phi)$, and $w_2 = \sin(\phi)\sin(\theta)$.

Contour plots of $\mathbb{E}[\hat{J}^{(k)}_{SAM}]$ are shown in Figs. 1(a)-1(c) for N=1,100, and 1000. The SAM cost surface is shown in fig. 1(d) for comparison. In the figures, minima of J_{SAM} are indicated by '*' while the local minima of $\mathbb{E}[\hat{J}^{(k)}_{SAM}]$ are

---

[1]The rows and columns of matrices in this paper are indexed from 1.
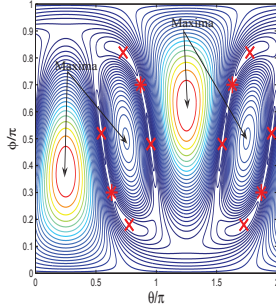[2]This example is taken from [3].

indicated by 'x'. The sam cost is invariant to the reversal of the elements of the shortener and to the negation of shortener coefficients [3]. Thus the local minima occur in quadruples as shown in the plots. Code from [5] was modified to generate these plots.
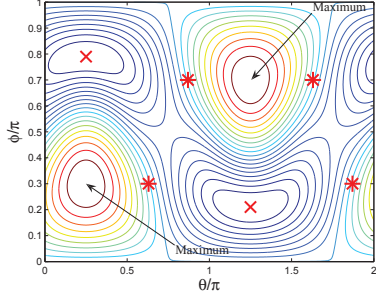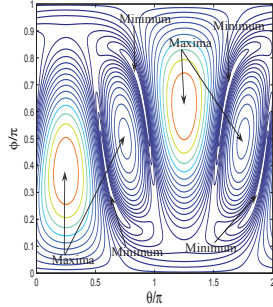

(a) $\mathbb{E}\{\hat{J}_{SAM}^{(k)}\}$ Cost surface for N=1
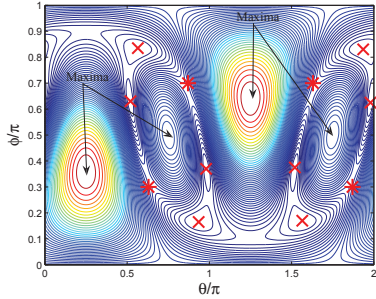
(b) $\mathbb{E}\{\hat{J}_{SAM}^{(k)}\}$ Cost surface for N=100

(c) $\mathbb{E}\{\hat{J}_{SAM}^{(k)}\}$ Cost surface for N=1000

(d) SAM Cost surface

Fig. 1.

From these plots it is seen that the nature of the cost surface $\mathbb{E}[\hat{J}_{SAM}^{(k)}]$ changes significantly as N varies. The position and number of the minima of $\mathbb{E}[\hat{J}_{SAM}^{(k)}]$ also depends on N and these minima do not coincide with the minima of J_{SAM}.

The minima of $\mathbb{E}[\hat{J}_{SAM}^{(k)}]$ appear to be moving towards those of J_{SAM} as N increases. Indeed, the two costs become exactly equal to one another in the limit as $N \to \infty$ due to the weak law of large numbers. However, significant differences remain between the two costs even for the relatively large value $N = 1000$ in this low dimensional example.

V. COST FUNCTION ANALYSIS UNDER UNIT NORM EFFECTIVE CHANNEL CONSTRAINT

If we impose the condition $||\mathbf{c}||_2 = 1$ during implementation, we can investigate the relationship between $\mathbb{E}[\hat{J}_{SAM}^{(k)}]$ and J_{SAM} using (22).

From the system model we have $\mathbf{c} = \mathbf{H}^T\mathbf{w}$ where $\mathbf{H}$ is the convolution matrix of the channel. Therefore, $\mathbf{c}$ is constrained to lie in the intersection of the row space of $\mathbf{H}$ and the unit sphere. However, by employing fractionally-spaced equalizers, we can achieve any point $\mathbf{c}$ on the unit sphere provided the conditions for strong perfect equalization are satisfied [6].

Let $\mathcal{R}$ be the set of possible autocorrelation vectors subject to the constraint $||\mathbf{c}||_2 = 1$. In order to find the upper bound of $\mathbb{E}[\hat{J}_{SAM}^{(k)}]$ for a given J_{SAM} for various values of N, we must solve the program

$$\text{maximize } \mathbb{E}[\hat{J}_{SAM}^{(k)}]$$
$$\text{subject to } J_{SAM} \leq k$$
$$\mathbf{r} \in \mathcal{R}. \quad (23)$$

If we can find a polytope $\mathcal{P}$ such that $\mathcal{R} \subset \mathcal{P}$ we can relax the problem and replace the constraint $\mathbf{r} \in \mathcal{R}$ by a set of linear equality and inequality constraints.

A. Solution to optimization problem

From (1) we can write $R_{cc}(l) = \sum_{k=0}^{L_c} c(k)c(k + l) = \mathbf{c}^T\mathbf{U}^l\mathbf{c}$ where $\mathbf{U}$ is the shift matrix. Thus $\max_{\mathbf{c}^T\mathbf{c}=1} R_{cc}(l) = \frac{1}{2}\lambda_{max}(\{\mathbf{U}^l\}^T + \mathbf{U}^l)$ where $\lambda_{max}(.)$ is the maximum eigenvalue of the matrix argument. Similarly $\min_{\mathbf{c}^T\mathbf{c}=1} R_{cc}(l) = \frac{1}{2}\lambda_{min}(\{\mathbf{U}^l\}^T + \mathbf{U}^l)$. We can therefore form a vector $\mathbf{b}$ with these upper and lower bounds and a matrix $\mathbf{P}$ such that $\mathcal{P} = \{\mathbf{r}|\mathbf{Pr} \leq \mathbf{b}\}$. Thus program (23) can be bounded by

$$\text{maximize } \mathbb{E}[\hat{J}_{SAM}^{(k)}] = \mathbf{r}^T\mathbf{Ar} + \mathbf{r}^T\mathbf{Cr} = \mathbf{r}^T\widetilde{\mathbf{C}}\mathbf{r}$$
$$\text{subject to } J_{SAM} = \mathbf{r}^T\mathbf{Ar} \leq k$$
$$\mathbf{Pr} \leq \mathbf{b}$$
$$R_{cc}(0) = 1. \quad (24)$$

If $\nu \geq 0$ then $\mathbf{Ae}_1 = \mathbf{0}$ where $\mathbf{e}_1$ is the first column of the identity matrix. Also, as a result of the constraint $R_{cc}(0) = 1$ we can rewrite (24) as

$$\text{minimize } -(\tilde{c}_{11} + 2\mathbf{r}^{*T}\mathbf{c}_1^* + \mathbf{r}^{*T}\widetilde{\widetilde{\mathbf{C}}}\mathbf{r}^*)$$
$$\text{subject to } \mathbf{r}^{*T}\widetilde{\mathbf{A}}\mathbf{r}^* \leq k$$
$$\widetilde{\mathbf{P}}\mathbf{r}^* \leq \mathbf{b} \quad (25)$$

where $\tilde{c}_{11} = [\widetilde{\mathbf{C}}]_{11}$, $\mathbf{r}^* = [R_{cc}(1), \ldots, R_{cc}(L_c)]^T$, $\widetilde{\widetilde{\mathbf{C}}}$ and $\widetilde{\mathbf{A}}$ are the submatrices formed from deleting the first row and column of $\widetilde{\mathbf{C}}$ and $\mathbf{A}$ respectively, $\widetilde{\mathbf{P}}$ is obtained by deleting the first column of $\mathbf{P}$ and $\mathbf{c}_1^*$ is obtained by deleting the first entry of the first column of $\widetilde{\mathbf{C}}$.

The Lagrangian associated with problem (25) is

$$\mathcal{L}(\mathbf{r}^*, \lambda, \mu) = -(\tilde{c}_{11} + 2\mathbf{r}^{*T}\mathbf{c}_1^* + \mathbf{r}^{*T}\widetilde{\widetilde{\mathbf{C}}}\mathbf{r}^*)$$
$$+ \lambda(\mathbf{r}^{*T}\widetilde{\mathbf{A}}\mathbf{r}^* - k) + \mu^T(\widetilde{\mathbf{P}}\mathbf{r}^* - \mathbf{b}) \quad (26)$$

and

$$\nabla_{\mathbf{r}^*}\mathcal{L} = -2(\mathbf{c}_1^* + \widetilde{\widetilde{\mathbf{C}}}\mathbf{r}^*) + 2\lambda\widetilde{\mathbf{A}}\mathbf{r}^* + \sum_{i=1}^{n}\mu_i\mathbf{p}_i \qquad (27)$$

with $\widetilde{\mathbf{P}}^T = [\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n]$, $\mu_i \geq 0$ and $\lambda \geq 0$. If $\lambda \neq 0$ then $\mathbf{r}^{*T}\widetilde{\mathbf{A}}\mathbf{r}^* = k$ and if $\mu_i \neq 0$ then $\mathbf{p}_i^T\mathbf{r}^* = b_i$.

Let $\mathcal{S} = \{i | \mathbf{p}_i^T\mathbf{r}^* = b_i\}$ then $\mathcal{S} \subset \{1, 2, \ldots, n\}$. Let $\mathbf{P}_{\mathcal{S}}$ be the matrix formed from the rows of $\mathbf{P}$ contained in $\mathcal{S}$ and $\mathbf{b}_{\mathcal{S}}$ the vector formed from the elements of $\mathbf{b}$ contained in $\mathcal{S}$. Then $\mathbf{P}_{\mathcal{S}}\mathbf{r}^* = \mathbf{b}_{\mathcal{S}}$.

Any minima of the objective function in (25) must satisfy $\nabla_{\mathbf{r}^*}\mathcal{L} = \mathbf{0}$. This can be compactly represented as

$$\underbrace{\begin{bmatrix} -2\widetilde{\widetilde{\mathbf{C}}} + 2\lambda\widetilde{\mathbf{A}} & \mathbf{P}_{\mathcal{S}}^T \\ \mathbf{P}_{\mathcal{S}} & \mathbf{0} \end{bmatrix}}_{\mathbf{M}} \begin{bmatrix} \mathbf{r}^* \\ \mu_{\mathcal{S}} \end{bmatrix} = \begin{bmatrix} 2\mathbf{c}_1^* \\ \mathbf{b}_{\mathcal{S}} \end{bmatrix}. \qquad (28)$$

Assume $\mathbf{M}$ is full rank and $\mathcal{S}$ corresponds to a set of simultaneously satisfiable constraints. If we can find a solution for $\mathbf{r}^*$ which satisfies (28), $\mathbf{r}^{*T}\widetilde{\mathbf{A}}\mathbf{r}^* = k$ and $\widetilde{\mathbf{P}}\mathbf{r}^* \leq \mathbf{b}$ then the corresponding value of $\mathbb{E}[\hat{J}_{SAM}^{(k)}]$ is a candidate for a maximum.

If λ is a generalized eigenvalue of $\mathbf{M}$ then a solution to (28) exists if $\begin{bmatrix} 2\mathbf{c}_1^* \\ \mathbf{b}_{\mathcal{S}} \end{bmatrix}$ is in the range of $\mathbf{M}$. The solution $\mathbf{r}_{sol}^*$ is a sum of a particular solution $\mathbf{r}_{part}$ and a homogeneous solution which is a scalar multiple of the eigenvector coresponding to λ. The scalar is determined by ensuring $\mathbf{r}^{*T}\widetilde{\mathbf{A}}\mathbf{r}^* = k$. The corresponding value of $\mathbb{E}[\hat{J}_{SAM}^{(k)}]$ is a candidate for a maximum if $\widetilde{\mathbf{P}}\mathbf{r}_{sol}^* \leq \mathbf{b}$.

We now present the algorithm used to solve (25):

1) For a set $\mathcal{S} \subset \{1, 2, \ldots, n\}$ corresponding to simultaneously satisfiable constraints form $\mathbf{M}$.

2) Solve (28) for $\mathbf{r}^*$ in terms of λ and solve for λ to ensure $\mathbf{r}^{*T}\widetilde{\mathbf{A}}\mathbf{r}^* = k$. For those λ that are not generalized eigenvalues of $\mathbf{M}$ determine the corresponding value of $\mathbf{r}^*$ and compute $\mathbb{E}[\hat{J}_{SAM}^{(k)}]$. This is a candidate for a maximum provided $\widetilde{\mathbf{P}}\mathbf{r}^* \leq \mathbf{b}$.

3) For those λ that are generalized eigenvalues of $\mathbf{M}$ check that $\begin{bmatrix} 2\mathbf{c}_1^* \\ \mathbf{b}_{\mathcal{S}} \end{bmatrix}$ is in the range of $\mathbf{M}$. That is if

$$\mathbf{M}\mathbf{M}^{\dagger}\begin{bmatrix} 2\mathbf{c}_1^* \\ \mathbf{b}_{\mathcal{S}} \end{bmatrix} = \begin{bmatrix} 2\mathbf{c}_1^* \\ \mathbf{b}_{\mathcal{S}} \end{bmatrix}$$

where $\mathbf{M}^{\dagger}$ is the pseudoinverse of $\mathbf{M}$. The solution $\mathbf{r}_{sol}$ is a sum of the particular solution $\mathbf{r}_{part} = \mathbf{M}^{\dagger}\begin{bmatrix} 2\mathbf{c}_1^* \\ \mathbf{b}_{\mathcal{S}} \end{bmatrix}$ and a homogeneous solution which is a scalar multiple of the eigenvector corresponding to λ. That is $\mathbf{r}_{sol} = \mathbf{r}_{part} + \gamma\mathbf{v}$. Determine γ such that $\mathbf{r}_{sol}\widetilde{\mathbf{A}}\mathbf{r}_{sol} = k$. The corresponding value of $\mathbb{E}[\hat{J}_{SAM}^{(k)}]$ is a candidate for a maximum provided $\widetilde{\mathbf{P}}\mathbf{r}^* \leq \mathbf{b}$.

4) Form another subset and go to 1. If all subsets have been considered go to 5.

5) Select the maximum value among all the candidates.

Using this algorithm, the behaviour of $\max \mathbb{E}[\hat{J}_{SAM}^{(k)}]$ as a function of N was investigated for various values of J_{SAM} for $L_c = 4$ and $\nu = 1$. The results are illustrated in figure 2. The $\mathbb{E}[\hat{J}_{SAM}^{(k)}] = J_{SAM}$ surface is also shown and it is seen that $\max \mathbb{E}[\hat{J}_{SAM}^{(k)}]$ approaches J_{SAM} as N grows.
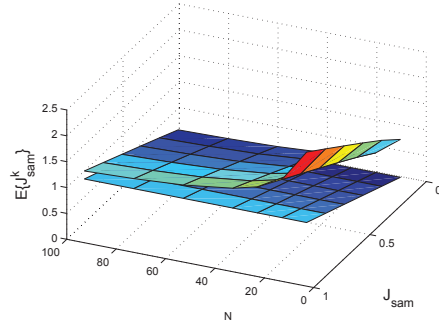

Fig. 2. $\max \mathbb{E}[\hat{J}_{SAM}^{(k)}]$ as a function of N for different values of J_{SAM}

VI. CONCLUSION

In this paper we have studied the relationship between the actual cost the blind adaptive stochastic gradient descent SAM algorithm is minimizing and the sum squared autocorrelation. We have shown that as the window size N used in the approximate autocorrelation calculation becomes large the distance between the minima of the two costs decreases. We have also studied the upper bound of $\mathbb{E}[\hat{J}_{SAM}^{(k)}]$ assuming the input is Gaussian and shown that even for finite N this value is close to the true value of J_{SAM}. Results obtained in [7] bounding the output SIR in terms of J_{SAM} can be used to extend the results of this paper to relate $\mathbb{E}[\hat{J}_{SAM}^{(k)}]$ and the SIR.

Using our algorithm for computation of the upper bound, a user can determine an appropriate value of N for use in a given implementation.

REFERENCES

[1] D. D. Falconer and F. R. Magee, "Adaptive channel memory truncation for maximum likelihood sequence estimation," *Bell Syst. Tech. J.*, pp. 1541–1562, Nov. 1973.

[2] P. Melsa, R. Younce, and C. Rohrs, "Impulse response shortening for discrete multitone transceivers," *IEEE Transactions on Communications*, vol. 44, no. 12, pp. 1662–1672, 1996.

[3] J. Balakrishnan, R. K. Martin, and C. R. Johnson, Jr., "Blind, adaptive channel shortening by sum-squared auto-correlation minimization (SAM)," *IEEE Trans. Signal Processing*, vol. 51, pp. 3086–3093, 2003.

[4] A. Papoulis and S. U. Pillai , *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, 2002.

[5] R. Martin, matlab Code for Papers by R. K. Martin. [Online]. Available: http://bard.ece.cornell.edu/matlab/martin/index.htm.

[6] C. R. Johnson Jr. et al., "The Core of FSE-CMA Behavior Theory," in *Unsupervised Adaptive Filtering, vol. II: Blind Deconvolution*, S. Haykin, Ed. Wiley, 2000, pp. 13–112.

[7] J. M. Walsh, R. K. Martin, and C. R. Johnson, Jr., "Convergence and performance issues for autocorrelation based adaptive channel shortening," in *Asilomar Conference on Signals, Systems, and Computers*, Nov, 2006.