A Stochastic Gradient Algorithm for Transmit Antenna Weight Adaptation with Feedback

Brian C. Baniater
Department of Electrical and Computer Engineering,
University of California at San Diego, La Jolla, CA 92093
LSI Logic Corporation, San Diego CA 92121

James R. Zeidler
Department of Electrical and Computer Engineering,
University of California at San Diego, La Jolla, CA 92093
Space and Naval Warfare Systems Center, San Diego

ABSTRACT
A simple algorithm for adaptation of the complex baseband weights of a transmit antenna array utilizing feedback from the receiving unit is presented. A dedicated pilot signal is transmitted from the transmitting unit with a pair of perturbed test weights, feedback bit is transmitted from the receiving unit indicating which weight vector provided the larger power; the transmitting unit adjusts a base weight to this selection and continues the process. This approach is shown to be a coarse stochastic gradient algorithm. The settling time and steady state characteristics of the algorithm are investigated analytically and verified by simulation, and the algorithm is shown to converge to those weights which maximize the power delivered to the receiver.

1. INTRODUCTION
It is generally accepted that the downlink of next generation cellular systems will require greater capacity than the uplink, due to the asymmetry of data traffic patterns. The use of transmit adaptive antenna arrays at the base station is a promising area for downlink capacity improvement. This paper describes a gradient algorithm utilizing mobile to base feedback in order to achieve some of those possible gains.

In many systems Frequency Division Duplexing (FDD) is used, causing difficulties in transmit array beam forming since the downlink and uplink channels in a multipath environment are in general not the same. Blind algorithms make use of long term parameter correlation which may exist between the downlink and uplink channels to steer the beam [9][13], but cannot take advantage of the fading diversity which the multiple antennas may be able to provide. Conversely, when space time codes are used and there is only one receive antenna [1][2][10], one can obtain fading diversity from the multiple transmit antennas, but no gain for fully correlated fading and no beam steering reduction of radiated interference. Hence, in FDD systems full realization of the possible benefits of adaptive transmit antennas requires receive unit to transmit unit feedback in order to provide channel state information for the selection of the transmit weights [3][5][7].

This paper will present a system for the adaptation of transmit antenna weights using gradient based feedback from the receiving unit. The usable power delivered to the mobile is considered to be an inverse cost, to be maximized, and an efficient gradient method is defined using weight vector perturbations and feedback to extract a stochastic estimate of the gradient of this inverse cost, which is then used to adjust the transmit antenna weights. While stochastic gradient algorithms, especially LMS, have been applied to receive antenna systems before [4][6][11][12], it appears that this is the first proposal for their application to the adaptation of transmit antenna weights.

Utilizing the algorithm, the weights converge to those which deliver the maximum usable power subject to a total transmit power constraint, which in the rank 1 case gives the "matched filter" weights [8], equal to the conjugate of the channel gains. These weights give gains from both fading diversity and beam steering, such as is not provided by blind techniques and space time codes respectively. The gradient sign feedback mechanism is simple and robust. One substantial benefit of the system is that the receiving unit does not need to know the details of the transmit antenna system. That is, the receiver need not know the number of transmit antennas, a separate code or time slot for measurement each transmit antenna channel, or the precise nature of the transmit system's weight update algorithm. The receiver only needs to know the appropriate measurement time intervals and feedback format.

2. PRELIMINARIES

2.1. System Model
The system is considered to use Nyquist pulse shaping so that a discrete time representation of the waveforms is adequate. The environment contains N transmission antennas at the base station and a maximum of L delay paths, spaced integer sample times apart. While a DS-CFDMA system might provide several such paths, a narrow band system would give L=1. The signal received by the mobile is

\[ r(n) = \sum_{l=0}^{L-1} \sqrt{\beta_l} \cdot w^H \cdot \mathbf{c}_l \cdot s(n-l) + n(n) \]

where \( s(n) \) is the modulation sequence, \( \mathbf{c}_l \) is the received noise and interference, \( \beta_l \) is the transmit power, \( w \) is the normalized transmit weight vector and \( \mathbf{c}_l \) is the channel response vector of the \( l \)-th path.

Defining the mobile's gain matrix as

\[ \mathbf{R} = \sum_{l=0}^{L-1} \mathbf{c}_l \mathbf{c}_l^H \]

the total usable signal power at the receiver is given by

\[ p(\mathbf{R}) = \beta \cdot \mathbf{w}^H \cdot \mathbf{R} \cdot \mathbf{w} \]

A simple transmit algorithm will ignore the cross interference and potential for null-forming and simply select the weights to maximize this power. This result is simply the eigenvector corresponding to the largest eigenvalue, indexed 0, so that for arbitrary \( \phi \)

\[ \frac{\mathbf{w}}{||\mathbf{w}||} = e^{i\phi} \mathbf{q}_0 \]

In the case where \( L=1 \), we get the "matched filter" weights (5).

\[ \frac{\mathbf{w}}{||\mathbf{w}||} = e^{i\phi} \mathbf{c}_0 \]
2.2. Inverse Cost Function, Gradient and Eigenanalysis

In the application of a gradient algorithm it is necessary to define the cost function which is to be minimized. In this application, the received signal power is the inverse cost function, and is to be maximized subject to a total transmission power constraint. In order to simplify the analysis, the weight vector is considered to be tracked in a non-normalized sense and normalization is applied just prior to transmission. Thus, the inverse cost function is given by

\[ J = \frac{p^{\text{error}}}{p^{\text{traffic}}} = w^H R w \]

(6)

The gradient of \( J \) with respect to \( w \) is given by

\[ \nabla w (J) = \begin{bmatrix} R w \frac{w^H R w}{w^H w} \end{bmatrix} - \begin{bmatrix} w \frac{w^H R w}{w^H w} \end{bmatrix} \]

(7)

It is convenient to consider the eigenmodes of the system. The mobile's gain matrix is decomposed as

\[ R = Q \Lambda Q^H = \sum_{k=0}^{\text{rank}(R)-1} q_k q_k^H \]

(8)

The weight vector is portrayed in terms of the eigenmodes as

\[ u = Q^H w \]

(9)

It will be convenient to define the vector \( v \) comprised of the squared magnitudes of \( u \), so that

\[ v_k = |u_k|^2 \]

(10)

3. BINARY FEEDBACK GRADIENT ADAPTATION

3.1. System Description

The system will be considered to be a CDMA system. The base station transmits the data with a weight vector \( w \), while the pilot is transmitted using a pair of time multiplexed weight vectors, \( w_{\text{even}} \) and \( w_{\text{odd}} \), which are perturbed from the tracked weight vector \( w_{\text{base}} \). All of these weight vectors are constant during a measurement interval. The transmitted waveform for the specific mobile is given by

\[ t(n) = \begin{cases} \begin{bmatrix} \frac{w_{\text{even}}}{w_{\text{base}}} & s_p(n) \end{bmatrix} & \text{if } \left\lceil \frac{n}{M} \right\rceil = \text{even} \\ \begin{bmatrix} \frac{w_{\text{even}}}{w_{\text{even}}} & s_p(n) \end{bmatrix} & \text{if } \left\lceil \frac{n}{M} \right\rceil = \text{odd} \end{cases} \]

(11)

Here, \( n \) is the Nyquist sampling time index, \( M \) is the duration of each even/odd perturbation slot, \( s_p(n) \) is the information bearing modulated symbols, \( s_p(n) \) is the pilot sequence modulation, \( w_{\text{even}} \) is the mean traffic channel transmission power, and \( w_{\text{pilot}} \) is the mean pilot channel transmission power. The sequences \( s_p(n) \) and \( s_p(n) \) would typically be orthogonal, by code multiplexing or time multiplexing.

The transmitter block diagram for a two antenna example is shown in Figure 1. The weight vector applied to the data channel is the mean of even and odd pilot weight vectors to facilitate coherent demodulation using the pilot channel.

\[ w = \frac{w_{\text{even}} + w_{\text{odd}}}{2} \]

(12)

Figure 1: Example of 2 antenna transmission system

An efficient receiver for this mechanism generates channel estimates for even and odd channel estimates, \( \hat{c}_{\text{even}} \) and \( \hat{c}_{\text{odd}} \), the mean of which is used for coherent demodulation. The difference between the powers of these even and odd channel estimates, \( T \), is used to generate a decision feedback sign bit, with '1' selecting the even channel or '0' selecting the odd channel. Note that in the case of resolvable multipath (\( \text{rank}(R) > 1 \)) the mobile will be tracking several versions of the received pilot, making channel estimates for each path and combining these channel estimate powers prior to doing the decision comparison, as is implied by (2) and (6). For the purpose of this discussion, it is assumed that the base station receives the feedback bit instantaneously after the completion of the measurement period. Given this feedback, the base station generates new transmit weights as follows.

When feedback received at beginning of new test interval

\[ \text{if feedback} \iff 1, \text{indicating even channel was better} \]

\[ w_{\text{even}} \leftarrow w_{\text{even}} \]

\[ \text{else} \]

\[ w_{\text{odd}} \leftarrow w_{\text{odd}} \]

\[ \text{end} \]

\[ p \leftarrow \text{normalized test perturbation function} \]

\[ w_{\text{even}} \leftarrow w_{\text{even}} + [w_{\text{even}} - w_{\text{base}}] \beta \frac{\pi}{2} \cdot p \]

\[ w_{\text{odd}} \leftarrow w_{\text{odd}} - [w_{\text{odd}} - w_{\text{base}}] \beta \frac{\pi}{2} \cdot p \]

\[ w \leftarrow \frac{w_{\text{even}} + w_{\text{odd}}}{2} \]

The test perturbation \( p \) is generated as a zero mean complex random Gaussian vector with an autocorrelation matrix \( 2I \). The parameter \( \beta \) is the adaptation rate, with larger \( \beta \) giving faster but noisier tracking. The normalization of the even/odd weight vectors in (11) is applied so that the power transmitted in even and odd time slots is equal, and hence the mobile selects the better weight vector rather than a larger transmission power.

3.2. Convergence Performance (Learning Curve)

Indexing the measurement and feedback intervals \( i \), the decision statistic \( T(i) \), the difference between the delivered powers in the even and odd time slots, is
\[ T(i) = \rho_{\text{pol}}^{(L)} \left( \frac{w_{\text{mean}}^H(i)Rw_{\text{mean}}(i)}{||w_{\text{mean}}(i)||^2} - \frac{w_{\text{old}}^H(i)Rw_{\text{old}}(i)}{||w_{\text{old}}(i)||^2} \right) \]  

(13)

If we assume that $\beta$ is small, then the test statistic $T(i)$ can be considered to be generated as a weighted differential step, given by the gradient weighted by the perturbation vector.

\[ T(i) \approx P_{\text{pol}}(i) \beta \sqrt{\frac{\pi}{2}} ||w(i)|| \left( |p^H g(w(i)) + a^H_n w(i) p| \right) \]  

(14)

The decision and the subsequent weight update are given by

\[ w(i+1) = w(i) + \text{sign}(T) \cdot \beta \sqrt{\frac{\pi}{2}} ||w(i)|| \cdot p(i) \]  

(15)

The expectation of the update (15) is the scaled gradient vector, and hence the update can be portrayed as the sum of this scaled gradient and a zero mean error vector $e(i)$.

\[ w(i+1) = w(i) + \beta ||w(i)|| \frac{g(w(i))}{||g(w(i))||} + \beta \sqrt{\frac{\pi}{2}} ||w(i)|| e(i) \]  

(16)

The expectation of the eigenmode energies incorporates the variance of the error vector $e$; with the ones vector $1 \equiv [1 \ldots 1]$ this eigenmodal energy update is

\[ v(i+1) = \left( I + \frac{4\beta}{||w(i)||^2 ||g(w(i))||} (\lambda - \bar{J}1) + \beta^2 \pi 11^T \right) v(i) \]  

(17)

The iteratively computed eigenmode energies $v$ of (17) applied appropriately to (6) give the learning curve of the inverse cost function.

A lower bound which does not require iterative computation is found by noting that

\[ \beta_{\text{min}} = \frac{\beta}{\lambda_0 - \lambda_{N-1}} \leq \beta_{\text{eff}}(i) = \frac{\beta}{||w(i)|| ||g(w(i))||} \]  

(18)

Hence, from (17) we can define

\[ \tilde{v}(i) = (I + 4\beta_{\text{min}} \cdot \Lambda + \beta^2 \pi 11^T) v(0) \]  

and a lower bound on the performance metric is

\[ \tilde{J}(i) = \left( \frac{1}{\tilde{v}(i)} \right) \]  

(20)

3.3. Discussion

Equations (17) and (19) show that the weight vector’s projection into the eigenspace corresponding to the largest eigenvalue will always grow at a faster rate than the other eigenmodes. The noise term assures that the desired eigenmode cannot stay at zero weighting even if it starts as such. Since this eigenvector provides the desired transmit weight vector, this shows that the algorithm converges, though final convergence is limited by weight noise and large $\beta$ will introduce large weight noise.

It has been assumed that the delay in implementing the feedback decision is negligible, though in practice some latency will be introduced. If this latency is much smaller than the measurement interval then it will not have any significant effect on the next feedback decision. Similarly, for small $\beta$ the
latex will not have a significant effect on the next decision since the weight change is small. For large $\beta$ and large latency more consideration would have to be made of this effect.

3.4. Numerical examples and simulation

Simulations of convergence were performed assuming no feedback bit errors and noiseless measurements by the mobile.

As an example a system with $N=3$ antennas is considered, with $\beta$ of 0.05. A two path case will be considered, with path gains given by the matrix $C$, which is normalized by the Frobenius norm.

\[
C = \begin{bmatrix}
1 & 0.8 \\
0.7 & 1 \\
1 & -0.8 \\
\end{bmatrix}
\]

This gives rise to the eigenvalue matrix

\[
\Lambda = \text{diag} \left[ 0.70, 0.30, 0 \right]
\]

The results are displayed as the eigenvector energies, $\nu$, normalized by the norm of $w$. For the first case, the initial weight vector is given by equal starting eigenvector energies, per (23).

\[
\nu = \frac{1}{3} \begin{bmatrix} 1 & 1 \end{bmatrix}
\]

In the second case, the initial weight vector has a very small component in the primary eigenmode, and is given by (24).

\[
\nu = \frac{1}{1.160004} \begin{bmatrix} 0.00004 & 0.16 \end{bmatrix}
\]

Figure 4 and Figure 5 display the learning curves obtained by simulation, iterative application of (17) which includes noise, iterative application of (16) with a noiseless update ($\beta=8$), and the lower bound (20). We see that for both examples the iteration of (17) gives very good results for both the resultant inverse cost metric $J$ and the values of the weights vector's eigenvector energies $\nu$, shown normalized by $|w|^2$ in Figure 2 and Figure 3. The noiseless update of equation (16) gives reasonable results in the first example, until weight noise starts to dominate when the weights approach the optimum, but in the second example the noiseless calculation provides a pessimistic adaptation rate, as it experiences a kind of stalling phenomenon. The lower bound of (20) appears to give a reasonably tight prediction of the performance.

In principle, regions with a small gradient could give rise to stalling in a gradient algorithm. However, this gradient sign algorithm obtains twofold stall mitigation, from the gradient normalization in the update and from the eigenmode excitation due to the noisy update. These effects are clear in Figure 3 for the second example. The simulations and calculations from (17) match, showing a large immediate step of $\nu_0$ from about 10e-5 to about 10e-2, which is the excitation of the desired eigenmode by the noisy update. On the other hand, the noiseless update calculation for this example remains somewhat stalled, as is visible by the slow update of $\nu_0$ and the slight plateau visible around sample 20 of Figure 5.

4. CONCLUSION

A new gradient sign algorithm for transmit antenna array adaptation has been defined and the convergence performance of the algorithm has been analyzed. The algorithm makes use of gradient sign feedback from the receiver to generate a coarse gradient estimate used by the transmitter to adjust the transmit weights. The mechanism employed by the receiver to generate the feedback is simple and can be employed with no knowledge of the specifics of the transmitter antenna algorithm, i.e. the receiver need not know how many antennas are employed or exactly how the update is performed. It was shown that the algorithm converges. Eigenmodal excitation from the noisy weight update guarantees immunity to stalling conditions. The convergence performance was found to match the analysis through simulation verification.

5. REFERENCES


