Adaptive MIMO Antenna Selection

Inaki Berenguer[1]
Lab. for Commun. Engineering
University of Cambridge
Cambridge, England

Xiaodong Wang
Dept. of Electrical Engineering
Columbia University
New York, NY

Vikram Krishnamurthy
Dept. of Electrical Engineering
University of British Columbia
Vancouver, Canada

Abstract

The performance of multiple-input multiple-output (MIMO) systems can be improved by employing a larger number of antennas than actually used and selecting the optimal subset based on the channel state information. Existing antenna selection algorithms assume perfect channel knowledge and optimize criteria such as Shannon capacity or various bounds on error rate. This paper examines MIMO antenna selection algorithms when the set of possible solutions is large and only a noisy estimate of the channel is available. We propose discrete stochastic approximation algorithms to adaptively select a better antenna subset using criteria such as maximum channel capacity, minimum error rate, etc. We also consider scenarios of time-varying channels for which the antenna selection algorithms can track the time-varying optimal antenna configuration. We present numerical examples to show the convergence of these algorithms and the excellent tracking capabilities.

I. Introduction

Multiple-input multiple-output (MIMO) systems can offer significant capacity gains over traditional single-input single-output (SISO) systems [7]. However, multiple antennas require multiple RF chains which consist of amplifiers, analog to digital converters, mixers, etc., that are typically very expensive. An approach for reducing the cost while maintaining the high potential data rate of a MIMO system is to employ a reduced number of RF chains at the receiver (or transmitter) and attempt to optimally allocate each chain to one of a larger number of receive (transmit) antennas which are usually cheaper elements. In this way, only the best set of antennas is used, while the remaining antennas are not employed, thus reducing the number of required RF chains.

Recently, several algorithms have been developed for selecting the optimal antenna subset given a channel realization. In [6] it is proposed to select the subset of transmit or receive antennas based on the Shannon capacity criterion. Antenna selection algorithms that minimize the bit error rate (BER) of linear receivers in spatial multiplexing systems are presented in [3]. In [2], antenna selection algorithms are proposed to minimize the symbol error rate when orthogonal space-time block coding is used in MIMO systems.

The main problem in the algorithms appeared in the literature is that they assume perfect channel knowledge to find the optimal antenna configuration. This paper presents discrete stochastic approximation algorithms for selecting the optimal antenna subset based on advanced discrete stochastic optimization techniques found in the recent operations research literature [1]. These techniques optimize an objective function (e.g., maximum capacity or minimum error rate) over a set of feasible parameters (e.g., antenna subsets to be used) when the objective function cannot be evaluated analytically but can only be estimated. The methods are in the same spirit as traditional adaptive filtering algorithms such as the least mean-squares (LMS) algorithm in which at each iteration, the algorithm makes a move towards a better solution. But in this case, the parameters to be optimized take discrete values (i.e., antenna indices to be used). The most important property of the proposed algorithms is their self-learning capability - most of the computational effort is spent at the global or a local optimizer of the objective function.

When the MIMO channel is time-varying, the optimal antenna subset is no longer fixed. To cope with this situation we make use of a discrete stochastic approximation algorithm with a fixed step size which acts as a forgetting factor which enables it to track the time-varying optimal antenna subset.

The remainder of this paper is organized as follows. In Section 2, the MIMO system model with antenna selection is presented. We also formulate the antenna selection problem as a discrete stochastic optimization problem. In Section 3, a general discrete stochastic optimization algorithm is presented and its convergence properties are summarized. In Section 4, several antenna selection criteria are presented, including maximum capacity, minimum bound on error rate and *minimum error rate*. In Section 5, antenna selection in time-varying channels is addressed. Section 6 contains the conclusions.

II. Problem Formulation

A. MIMO System with Antenna Selection

Consider a MIMO system as shown in Figure 1 with n_T transmit and n_R receive RF chains and suppose that there are $N_R \geq n_R$ receive antennas. Without loss of generality, in this paper we assume that antenna selection is implemented only at the receiver. The channel is represented by an $(N_R \times n_T)$ matrix $\boldsymbol{H}$ whose element h_{ij} represents the complex gain of the channel between the jth transmit antenna and the ith receive antenna. We assume a flat fading channel remaining constant over several bursts. The subset of $n_R \leq N_R$ receive antennas to be employed is determined by the selection algorithm operating at the receiver which selects the optimal subset

[1]I. Berenguer is supported by a Fulbright Fellowship.

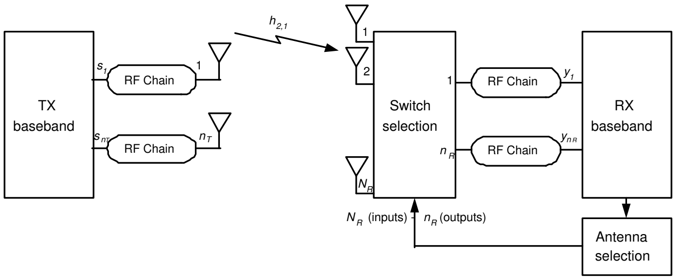

Fig. 1. Schematic representation of a MIMO system with antenna selection.

ω of all possible $\binom{N_R}{n_R}$ subsets of n_R receive antennas. Denote $\boldsymbol{H}[\omega]$ as the $(n_R \times n_T)$ channel submatrix corresponding to the receive antenna subset ω, i.e., rows of $\boldsymbol{H}$ corresponding to the selected antennas. The corresponding received signal is then

$$\boldsymbol{y} = \sqrt{\frac{\rho}{n_T}}\boldsymbol{H}[\omega]\boldsymbol{s} + \boldsymbol{n} \tag{1}$$

where $\boldsymbol{s} = [s_1, s_2, ..., s_{n_T}]^T$ is the $(n_T \times 1)$ transmitted signal vector, $\boldsymbol{y} = [y_1, y_2, ..., y_{n_R}]^T$ is the $(n_R \times 1)$ received signal vector, $\boldsymbol{n}$ is the $(n_R \times 1)$ received noise vector, and ρ is the total signal-to-noise ratio independent of the number of transmit antennas. The entries of $\boldsymbol{n}$ are i.i.d. circularly symmetric complex Gaussian variables with unit variance, i.e., $n_i \sim \mathcal{N}_c(0, 1)$.

For the problems that we are looking at in this paper, the receiver is required to know the channel. One way to perform channel estimation at the receiver is to use a training preamble [5]. Suppose $T \geq n_T$ MIMO training symbols $\boldsymbol{s}(1)$, $\boldsymbol{s}(2)$, ..., $\boldsymbol{s}(T)$ are used to probe the channel. The received signals corresponding to these training symbols are

$$\boldsymbol{y}(i) = \sqrt{\frac{\rho}{n_T}}\boldsymbol{H}[\omega]\boldsymbol{s}(i) + \boldsymbol{n}(i), \quad i = 1, 2, ..., T. \tag{2}$$

Denote $\boldsymbol{Y} = [\boldsymbol{y}(1), \boldsymbol{y}(2), ..., \boldsymbol{y}(T)]$, $\boldsymbol{S} = [\boldsymbol{s}(1), \boldsymbol{s}(2), ..., \boldsymbol{s}(T)]$ and $\boldsymbol{N} = [\boldsymbol{n}(1), \boldsymbol{n}(2), ..., \boldsymbol{n}(T)]$. Then (2) can be written as

$$\boldsymbol{Y} = \sqrt{\frac{\rho}{n_T}}\boldsymbol{H}[\omega]\boldsymbol{S} + \boldsymbol{N} \tag{3}$$

and the maximum likelihood estimate of the channel matrix $\boldsymbol{H}[\omega]$ is given by

$$\hat{\boldsymbol{H}}[\omega] = \sqrt{\frac{n_T}{\rho}}\boldsymbol{Y}\boldsymbol{S}^H(\boldsymbol{S}\boldsymbol{S}^H)^{-1}. \tag{4}$$

According to [5], the optimal training symbol sequence $\boldsymbol{S}$ that minimizes the channel estimation error should be orthogonal, i.e., $\boldsymbol{S}\boldsymbol{S}^H = T \cdot \boldsymbol{I}_{n_T}$. In an uncorrelated MIMO channel, the channel estimates $\hat{\boldsymbol{H}}_{i,j}[\omega]$ computed using (4) with orthogonal training symbols are statistically independent Gaussian variables with [5]

$$\hat{\boldsymbol{H}}_{i,j}[\omega] \sim \mathcal{N}_c\Big(\boldsymbol{H}_{i,j}[\omega], \frac{\rho}{Tn_T}\Big). \tag{5}$$

B. Discrete Optimization Problem

We now formulate the antenna selection problem as a discrete stochastic optimization problem. Denote each of the antenna subsets as $\omega = \{Ant(1), Ant(2), ..., Ant(n_R)\}$ (e.g., selecting the first, second and sixth antennas is equivalent to $\omega = \{1, 2, 6\}$). Denote the set of all $P = \binom{N_R}{n_R}$ possible antenna subsets as $\Omega = \{\omega_1, \omega_2, ..., \omega_P\}$. Then, the receiver selects one of the antenna subsets in Ω to optimize a certain objective function $\Phi(\boldsymbol{H}[\omega])$ according to some specific criterion, e.g., maximum capacity between the transmitter and the receiver or minimum error rate. Thus, the discrete optimization problem becomes

$$\omega^* = \arg\max_{\omega\in\Omega} \Phi(\boldsymbol{H}[\omega]), \tag{6}$$

where we use ω^* to denote the global maximizer of the objective function. In practice, however, the exact value of the channel $\boldsymbol{H}[\omega]$ is not available. Instead, we typically have a noisy estimate $\hat{\boldsymbol{H}}[\omega]$ of the channel.

Suppose that at time n the receiver obtains an estimate of the channel, $\hat{\boldsymbol{H}}[n, \omega]$, and computes a noisy estimate of the objective function $\Phi(\boldsymbol{H}[\omega])$ denoted as $\phi[n, \omega]$. Given a sequence of i.i.d. random variables $\{\phi[n, \omega],\ n = 1, 2, ...\}$, if each $\phi[n, \omega]$ is an unbiased estimate of the objective function $\Phi(\boldsymbol{H}[\omega])$, then (6) can be reformulated as the following discrete stochastic optimization problem

$$\omega^* = \arg\max_{\omega\in\Omega} \Phi(\boldsymbol{H}[\omega]) = \arg\max_{\omega\in\Omega} E\left\{\phi[n, \omega]\right\}. \tag{7}$$

Note that existing works on antenna selection assume perfect channel knowledge and therefore treat deterministic combinatorial optimization problems. On the other hand, we assume that only noisy estimates of the channel are available and hence the corresponding antenna selection problem becomes a discrete stochastic optimization problem. In what follows we first discuss a general discrete stochastic approximation method to solve the discrete stochastic optimization problem in (7) and then we treat different forms of the objective function under different criteria, e.g., maximum capacity or minimum error rate.

III. Discrete Stochastic Approximation Algorithm

There are several methods that can be used to solve the discrete stochastic optimization problem in (7). A way to approach the discrete stochastic optimization problem is to use iterative algorithms that resemble a stochastic approximation algorithm in the sense that they generate a sequence of estimates of the solution where each new estimate is obtained from the previous one by taking a small step in a good direction toward the global optimizer.

We present an stochastic approximation algorithm based on [1]. We use the $P = \binom{N_R}{n_R}$ unit vectors as labels for the P possible antenna subsets, i.e., $\xi = \{\boldsymbol{e}_1, \boldsymbol{e}_2, ..., \boldsymbol{e}_P\}$, where $\boldsymbol{e}_i$ denotes the $(P \times 1)$ vector with a one in the ith position and zeros elsewhere. At each iteration, the algorithm updates the $(P \times 1)$ probability vector $\boldsymbol{\pi}[n] = \Big[\pi[n, 1], ..., \pi[n, P]\Big]^T$ representing the state occupation probabilities with elements

$\pi[n,i] \in [0,1]$ and $\sum_i \pi[n,i] = 1$. Let $\omega^{(n)}$ be the antenna subset chosen at the n-th iteration. For notational simplicity, it is convenient to map the sequence of antenna subsets $\{\omega^{(n)}\}$ to the sequence $\{\boldsymbol{D}[n]\} \in \xi$ of unit vectors where $\boldsymbol{D}[n] = \boldsymbol{e}_i$ if $\omega^{(n)} = \omega_i$, $i = 1, ..., P$.

Algorithm 1 Discrete stochastic approximation algorithm for antenna selection

```
☐ Initialization
    n ⇐ 0
    select initial antenna subset ω^(0) ∈ Ω
    and set π[0, ω^(0)] = 1
    set π[0, ω] = 0 for all ω ≠ ω^(0)
for n = 0, 1, ... do
☐ Sampling and evaluation
    given ω^(n) at time n, obtain φ[n, ω^(n)]
    choose another ω̃^(n) ∈ Ω\ω^(n) uniformly
    obtain an independent observation
    φ[n, ω̃^(n)]
☐ Acceptance
    if φ[n, ω̃^(n)] > φ[n, ω^(n)] then
      set ω^(n+1) = ω̃^(n)
    else
      ω^(n+1) = ω^(n)
    end if
☐ Adaptive filter for updating state
occupation probabilities
    π[n+1] = π[n] + μ[n+1](D[n+1] − π[n])
    with the step size μ[n] = 1/n
☐ Computing the maximum
    if π[n+1, ω^(n+1)] > π[n+1, ω̂^(n)] then
      ω̂^(n+1) = ω^(n+1)
    else
      set ω̂^(n+1) = ω̂^(n)
    end if
end for
```

We assume that in a realistic communications scenario, each iteration of the above algorithm occurs within a data frame where the training symbols are used to obtain the channel estimates $\hat{\boldsymbol{H}}[n, \omega^{(n)}]$ and hence the noisy estimate of the cost $\phi[n, \omega^{(n)}]$. At the end of each iteration, antenna subset $\hat{\omega}^{(n)}$ will be selected for the rest of the data frame.

In the `Sampling and Evaluation` step in Algorithm 1, the candidate antenna subset $\tilde{\omega}^{(n)}$ is chosen uniformly from $\Omega\backslash\omega^{(n)}$. There are several variations for selecting a candidate antenna subset $\tilde{\omega}^{(n)}$. One possibility is to select a new antenna subset by replacing only one antenna in $\omega^{(n)}$. Define the distance between two antenna subsets as the number of different antennas between them. Hence, we can select $\tilde{\omega}^{(n)} \in \Omega\backslash\omega^{(n)}$ such that $d(\tilde{\omega}^{(n)}, \omega^{(n)}) = 1$. More generally we can select a new subset with distance D from the current antenna subset, where $1 \leq D \leq \min(n_R, N_R - n_R)$.

To obtain the independent observations in the `Sampling and Evaluation` step in Algorithm 1 we proceed as follows. At time n, we collect training symbols to estimate the channel and compute $\phi[n, \omega]$. Now, collect other training symbols from another antenna subset and compute $\phi[n, \tilde{\omega}]$. Therefore, $\phi[n, \omega]$ and $\phi[n, \tilde{\omega}]$ are independent observations. Although the independence of the samples is a condition for the analysis of the convergence of the algorithm, it is observed through simulations that the algorithm also converges under correlated observations of the objective function, i.e., we can use the same channel estimate to compute several observations of the objective function under different antenna configurations.

The sequence $\{\omega^{(n)}\}$ generated by Algorithm 1 is a Markov chain on the state space Ω which is not expected to converge and may visit each element in Ω infinitely often. Instead, under certain conditions, the sequence $\{\hat{\omega}^{(n)}\}$ converges almost surely to the global maximizer ω^*. Therefore, $\hat{\omega}^{(n)}$ denotes the estimate at time n of the optimal antenna subset ω^*.

In the `Adaptive filter for updating state occupation probabilities` step in Algorithm 1, $\boldsymbol{\pi}[n] = \left[\pi[n,1], \pi[n,2], ..., \pi[n,P]\right]$ denotes the empirical state occupation probability at time n of the Markov chain $\{\omega^{(n)}\}$. If we denote $W^{(n)}[\omega]$ for each $\omega \in \Omega$ as a counter of the number of times the Markov chain has visited antenna subset $\omega^{(n)}$ by time n, we can observe that $\boldsymbol{\pi}[n] = \frac{1}{n}\left[\; W^{(n)}[\omega_1], \quad ..., \quad W^{(n)}[\omega_P] \;\right]^T$. Therefore, the algorithm chooses the antenna subset which has been visited most often by the Markov chain $\{\omega^{(n)}\}$ so far.

As shown in [1], a sufficient condition to converge to the global maximizer of the objective function $\Phi(\boldsymbol{H}[\omega])$ is that the estimator of the objective function used in Algorithm 1 at two different antenna subsets are unbiased, i.e.,

$$\phi[n, \omega] = \Phi(\boldsymbol{H}[w]) + v[\omega, n], \tag{8}$$

where $\{v[\omega, n], \omega \in \Omega\}$ are i.i.d. and each $v[\omega, n]$, $\omega \in \Omega$, has a symmetric continuous probability density function with zero mean.

IV. Antenna Selection Under Different Criteria

In this section we use Algorithm 1 to optimize different objective functions $\Phi(\boldsymbol{H}[\omega])$.

A. Maximum MIMO Capacity

Assuming that the channel matrix $\boldsymbol{H}[\omega]$ is known at the receiver, but not at the transmitter, the instantaneous capacity of the MIMO channel is given by [7]

$$C[\omega] = \log\det\left(\boldsymbol{I}_{n_T} + \frac{\rho}{n_T}\boldsymbol{H}^H[\omega]\boldsymbol{H}[\omega]\right) \text{bit/s/Hz}. \tag{9}$$

One criterion for selecting the antennas is to maximize the above instantaneous capacity [6], i.e., choosing the objective function $\Phi(\boldsymbol{H}[\omega]) = C[\omega]$. We now present an implementation of Algorithm 1 and prove that it converges to the global maximum of the capacity (9).

In the `Sampling and Evaluation` step of Algorithm 1 choose

$$\phi[n, \omega] = \det\left(\boldsymbol{I}_{n_T} + \frac{\rho}{n_T}\hat{\boldsymbol{H}}_1^H[n, \omega]\hat{\boldsymbol{H}}_2[n, \omega]\right) \tag{10}$$

where the channel estimates $\hat{\boldsymbol{H}}_1[n,\omega]$ and $\hat{\boldsymbol{H}}_2[n,\omega]$ are obtained from independent training blocks. We consider the case in which $\hat{\boldsymbol{H}}_1[n,\omega]$ and $\hat{\boldsymbol{H}}_2[n,\omega]$ satisfy (5).

Theorem 1*: With $\phi[n,\omega]$ computed according to (10), the sequence $\{\hat{\omega}^{(n)}\}$ generated by Algorithm 1 converges to the antenna subset ω^* corresponding to the global maximizer of the MIMO capacity in (9).*

Proof: Since the logarithm is a monotonically increasing function, the antenna subset ω^* maximizing $\log\det(\cdot)$ is identical to that maximizing $\det(\cdot)$. To prove global convergence, we only need to show that $\phi[n,\omega]$ of (10) satisfies (8). For convenience define

$$\begin{aligned} \boldsymbol{M}[\omega] &= \boldsymbol{I}_{n_T} + \frac{\rho}{n_T}\boldsymbol{H}^H[\omega]\boldsymbol{H}[\omega] \\ \hat{\boldsymbol{M}}[n,\omega] &= \boldsymbol{I}_{n_T} + \frac{\rho}{n_T}\hat{\boldsymbol{H}}_1^H[n,\omega]\hat{\boldsymbol{H}}_2[n,\omega] \end{aligned} \quad (11)$$

and denote the elements of $\hat{\boldsymbol{M}}[n,\omega]$ as $\hat{m}_{i,j}$.

Consider (11). Since $\hat{\boldsymbol{H}}_1[n,\omega]$ and $\hat{\boldsymbol{H}}_2[n,\omega]$ are statistically independent samples, clearly $\hat{\boldsymbol{M}}[n,\omega]$ is an unbiased estimator of $\boldsymbol{M}[\omega]$. Moreover, due to (5), clearly

$$\hat{\boldsymbol{M}}[n,\omega] = \boldsymbol{M}[\omega] + \boldsymbol{T}[n,\omega] \quad (12)$$

where $\boldsymbol{T}[n,\omega]$ is a random variable with zero mean symmetric density function.

Now consider $\det(\hat{\boldsymbol{M}}[n,\omega])$. From [4, p.8]

$$\det(\hat{\boldsymbol{M}}[n,\omega]) = \sum_{\sigma}\text{sign}(\sigma)\prod_{i=1}^{n_T}\hat{m}_{i,\sigma(i)} \quad (13)$$

where the sum runs over all $n_T!$ permutations σ of the n_T items $\{1,...,n_T\}$ and $\text{sign}(\sigma)$ is $+1$ or -1. Omitting the sign, each term in the summation is of the form $\hat{m}_{1,\sigma(1)}\hat{m}_{2,\sigma(2)}...\hat{m}_{n_T,\sigma(n_T)}$. Thus, each term in the summation is multilinear and involves the product of elements of $\hat{\boldsymbol{M}}[n,\omega]$ from different rows and columns.

Next, due to the independence assumption in (5), it follows that for the matrix $\hat{\boldsymbol{M}}[n,\omega]$, the elements $\hat{m}_{i,j}$ and $\hat{m}_{p,q}$ are independent for $i \neq p$ and $j \neq q$, i.e., elements of $\hat{\boldsymbol{M}}[n,\omega]$ from distinct rows and columns are statistically independent. Hence $\hat{m}_{1,\sigma(1)}, \hat{m}_{1,\sigma(1)}, ..., \hat{m}_{n_T,\sigma(n_T)}$ are statistically independent. This independence together with the multilinear property and (12) implies that $\det(\hat{\boldsymbol{M}}[n,\omega])$ is an unbiased sample of $\det(\boldsymbol{M}[\omega])$ and satisfies

$$\det(\hat{\boldsymbol{M}}[n,\omega]) = \det(\boldsymbol{M}[\omega]) + v[n,\omega] \quad (14)$$

where $v[n,\omega]$ is a zero mean random variable with symmetric density function. We have demonstrated that (8) holds and therefore, Algorithm 1 converges to ω^*. □

To reduce the training symbols needed to estimate the channel in Algorithm 1, in numerical studies we used a single sample of the channel $\hat{\boldsymbol{H}}[n,\omega]$ and chose

$$\phi[n,\omega] = \log\det\left(\boldsymbol{I}_{n_T} + \frac{\rho}{n_T}\hat{\boldsymbol{H}}[n,\omega]^H\hat{\boldsymbol{H}}[n,\omega]\right). \quad (15)$$

Although this sample is biased, numerical results presented below show that Algorithm 1 still converges to the global optimum and uses only one channel estimate per iteration.

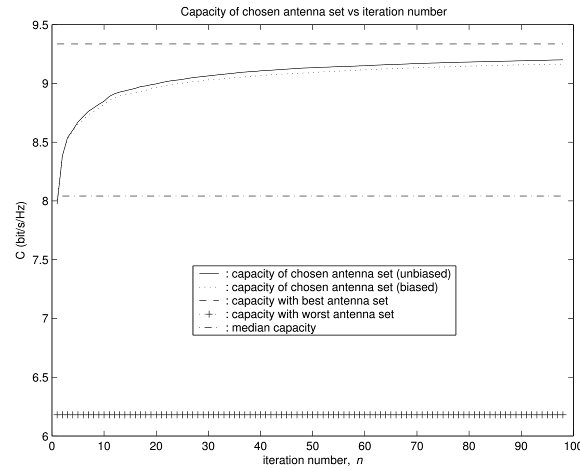

Fig. 2. The average of the capacities of chosen antenna subsets (over 3000 runs) versus iteration number n.

Simulation Results: We consider the performance of Algorithm 1 which selects the antenna subset maximizing the channel capacity using (15) as an estimate of the objective function. We consider $n_T = 2$, $N_R = 8$ and $n_R = 4$ antennas. We use the ML channel estimate in (4) with $T = 4$ orthogonal training symbols. We set $\rho = 10$dB. The $(N_R \times n_T)$ channel $\boldsymbol{H}$ is randomly generated and fixed during the whole simulation. In Figure 2 we consider 100 iterations per execution and we average the capacities of the antenna subset selected at all iterations over 3000 channel realizations. In the same figure we show the capacities of the best antenna subset and the worst antenna subset, as well as the median capacity among the $\binom{8}{4} = 70$ antenna configurations, found by exhaustive search. For comparison, we also show the performance of the original unbiased implementation in (10). It is seen that the unbiased implementation has slightly better convergence behavior than the biased one although it needs two channel estimates per iteration. From both figures, it is seen that the algorithm adaptively moves to the best antenna subset. We observe that although the algorithm takes some time to converge to the optimal antenna subset, it moves very fast to an antenna subset inducing high channel capacity.

B. Minimum Bounds on Error Rate

Consider the system in Figure 1 where the transmitted data $\boldsymbol{s}$ is multiplexed into the n_T transmit antennas. The input-output relationship is expressed in (1) where in this case, the transmitted symbols s_i belong to a finite constellation $\mathcal{A}$ of size $|\mathcal{A}|$. The receive antennas see the superposition of all transmitted signals. The task of the receiver is to recover the transmitted data $\boldsymbol{s}$. The ML detection rule is given by

$$\hat{\boldsymbol{s}} = \arg\min_{\boldsymbol{s}\in\mathcal{A}^{n_T}}\left\|\boldsymbol{y} - \sqrt{\frac{\rho}{n_T}}\boldsymbol{H}[\omega]\boldsymbol{s}\right\|^2. \quad (16)$$

At high signal-to-noise ratio, we can upper bound the probability of error of the ML detector using the union bound [3] which is a function of the squared minimum distance $d^2_{\min,r}$

of the received constellation given by [3]

$$d^2_{\min,r}[\omega] = \min_{\substack{\boldsymbol{s}_i,\boldsymbol{s}_j \in \mathcal{A}^{n_T} \\ \boldsymbol{s}_i \neq \boldsymbol{s}_j}} \|\boldsymbol{H}[\omega]\,(\boldsymbol{s}_i - \boldsymbol{s}_j)\|^2. \tag{17}$$

Therefore, minimizing the union bound on error probability is equivalent to maximizing $d^2_{\min,r}$.

***Theorem 2**: With*

$$\phi[n,\omega] = \min_{\substack{\boldsymbol{s}_i,\boldsymbol{s}_j \in \mathcal{A}^{n_T} \\ \boldsymbol{s}_i \neq \boldsymbol{s}_j}} \Big[\hat{\boldsymbol{H}}_1[n,\omega]\,(\boldsymbol{s}_i - \boldsymbol{s}_j)\Big]^H \Big[\hat{\boldsymbol{H}}_2[n,\omega]\,(\boldsymbol{s}_i - \boldsymbol{s}_j)\Big] \tag{18}$$

the sequence $\{\hat{\omega}^{(n)}\}$ generated by Algorithm 1 converges to the global maximizer ω^* of (17).

Note that the computation of $d^2_{\min,r}[\omega]$ is performed over $|\mathcal{A}|^{n_T}(|\mathcal{A}|^{n_T} - 1)$ possibilities for each antenna subset which can be prohibitive for large $|\mathcal{A}|$ or n_T. Let $\lambda_{\min}[\omega]$ be the smallest singular value of $\boldsymbol{H}[\omega]$ and let the minimum squared distance of the transmit constellation be $d^2_{\min,t} = \min_{\boldsymbol{s}_i,\boldsymbol{s}_j \in \mathcal{A}^{n_T}} \|(\boldsymbol{s}_i - \boldsymbol{s}_j)\|^2$. Then, it is shown in [3] that $d^2_{\min,r}[\omega] \geq \lambda^2_{\min}[\omega] d^2_{\min,t}$. Therefore, a selection criterion can be simplified to select the antenna subset $\omega \in \Omega$ whose associated channel matrix $\boldsymbol{H}[\omega]$ has the largest minimum singular value. In Algorithm 1, based on an estimate of the channel at time n we let $\phi[n,\omega] = \hat{\lambda}_{\min}[n,\omega]$.

For the simpler linear receivers, different objective functions have been developed in [3] which bound the error rate and their estimates can be used in Algorithm 1.

Another case of interest is when the orthogonal space-time block codes are employed. The post processing signal-to-noise ratio of the data stream is given by [2]

$$\text{SNR}[\omega] = \frac{\rho}{n_T}\text{trace}\left(\boldsymbol{H}^H[\omega]\boldsymbol{H}[\omega]\right). \tag{19}$$

***Theorem 3**: The sequence $\{\hat{\omega}^{(n)}\}$ generated by Algorithm 1 with $\phi[n,\omega] = trace\left[\hat{\boldsymbol{H}}_1^H[n,\omega]\hat{\boldsymbol{H}}_2^H[n,\omega]\right]$ converges to the global maximizer of (19).*

C. Minimum Error Rate

The antenna subset chosen by the different criteria based on bounds do not necessarily choose the antenna subset minimizing the bit error rate (BER). In this section, we propose an antenna selection algorithm that directly minimizes the symbol or bit error rate of the system under any type of receivers.

In the proposed method, a noisy estimate of the *simulated* error rate is used as the cost function in the stochastic approximation algorithm instead of a noisy estimate of a bound. The method proceeds as follows. Assume for example that the ML decoding algorithm in (16) is used. At time n, estimate the channel $\hat{\boldsymbol{H}}[n,\omega]$ with antenna subset ω. At the receiver, generate m *fake* random symbol vectors $\boldsymbol{S}_f = [\boldsymbol{s}_f(1),...,\boldsymbol{s}_f(m)]$ with $s_{f,k}(i) \in \mathcal{A}$ and perform a simulation of the form

$$\boldsymbol{Y}_f = \sqrt{\frac{\rho}{n_T}}\hat{\boldsymbol{H}}[n,\omega]\boldsymbol{S}_f + \boldsymbol{N} \tag{20}$$

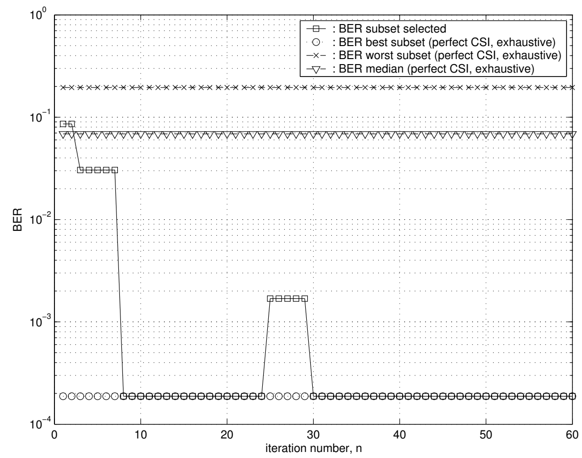

Fig. 3. Single run of the algorithm: BER of the of the chosen antenna subset versus iteration number n employing an ML receiver.

where the $(n_R \times m)$ matrix $\boldsymbol{N}$ contains i.i.d. $\mathcal{N}_c(0,1)$ samples. Perform the ML detection on (20) to obtain

$$\hat{\boldsymbol{S}}_f = \arg\min_{\boldsymbol{S} \in \mathcal{A}^{n_T \times m}} \left\| \boldsymbol{Y}_f - \sqrt{\frac{\rho}{n_T}}\hat{\boldsymbol{H}}[n,\omega]\boldsymbol{S} \right\|^2 \tag{21}$$

and estimate the bit error rate $\widehat{\text{BER}}[n,\omega]$ by comparing $\hat{\boldsymbol{S}}_f$ and $\boldsymbol{S}_f$. In this way, at time n, an estimate of the real $\text{BER}[\omega]$ has been obtained. Therefore, in Algorithm 1 we use $\phi[n,\omega] = -\widehat{\text{BER}}[n,\omega]$ as an observation of the cost function. The number of fake symbol vectors required to obtain a good estimate of the BER depends on the signal-to-noise ratio ρ of the channel. For low signal-to-noise ratio, only short fake sequences are needed.

It has been observed that the antenna subset minimizing the error rate at high signal-to-noise ratio usually corresponds to the same antenna subset minimizing the error rate at low signal-to-noise ratio. Therefore, to reduce the complexity of this method, we may perform the fake simulation with a lower signal-to-noise ratio and therefore, less fake symbols are needed to obtain a good estimate of the error rate.

Note that the fake symbols $\boldsymbol{s}_f$ are *not* actually sent through the channel. They are merely generated at the receiver to estimate the BER. Moreover, it is important to point that this method uses an *estimate* of the BER and a closed-form BER expression is not needed, which makes it appealing for other receivers for which even a tight bound is difficult to find. Obviously, the same method can be used in antenna selection for MIMO systems employing various space-time coding schemes. Moreover, it is straightforward to modify the algorithm to minimize the symbol error rate or frame error rate as well.

Simulation Results: To show the performance of this method we consider an ML receiver. We use QPSK symbols and we consider $N_R = 6$, $n_R = 2$ and $n_T = 2$. The $(N_R \times n_T)$ channel $\boldsymbol{H}$ is randomly generated and fixed during the whole simulation. We set $\rho = 9$dB and we use $T = 6$ orthogonal training symbols to estimate the channel. Before starting the algorithm, long simulations are performed assuming perfect channel knowledge over all antenna configurations to find the BER associated with each antenna subset. We run $n = 60$ iterations of the algorithm with $m = 500$ fake symbols per iteration. Figure 3 shows the BER of the antenna selected by the algorithm comparing it with the median, the best and the worst BER. It is seen that the algorithm converges to the optimal antenna subset. Moreover, it is observed that antenna selection at the receiver can improve the BER by more than two orders of magnitude with respect to the median BER.

V. Antenna Selection in Time-varying Channels

Now we consider nonstationary environments for which the optimum antenna subset takes on a time-varying form, $\omega^*[n] \in \Omega$, since the MIMO channel is time-varying. Consequently, the MIMO antenna selection algorithms should be able to track the best antenna subset if the variation of the channel is *slow* for tracking to be feasible.

In the static channel environment discussed in the previous section, in order for the method to converge, it was necessary for the method to become progressively more and more conservative as the number of iterations grew. Consequently, a decreasing step size, $\mu[n] = 1/n$, was used, in order to avoid moving away from a promising point unless there was a strong evidence that the move will result in an improvement. In the time-varying case, we require a step size that permits moving away from a state as the optimal antenna subset changes. Therefore, to track the optimal antenna subset, we replace the `Adaptive filter for updating state occupation probabilities` step in Algorithm 1 by

$$\boldsymbol{\pi}[n+1] = \boldsymbol{\pi}[n] + \mu(\boldsymbol{D}[n+1] - \boldsymbol{\pi}[n]) \qquad (22)$$

where $0 < \mu \leq 1$. A fixed step size μ in (22) introduces an exponential forgetting factor of the past occupation probabilities and allows to track slowly time-varying optimal antenna subset $\omega^*[n]$.

It has been observed that time-varying channels modify the optimal antenna subset over the time although most of the antennas in the optimal antenna subset remain the same. Hence, in time-varying channels, we can modify the `Sampling and Evaluation` step in Algorithm 1 to select a candidate solution $\tilde{\omega}^{(n)}$ uniformly from $\Theta \backslash \omega^{(n)}$ where Θ is defined as the set of antenna subsets $\tilde{\omega}^{(n)} \in \Omega$ such that $d(\tilde{\omega}^{(n)}, \omega^{(n)}) = D$, where we choose $D < \min(n_R, N_R - n_R)$.

Simulation Results: We demonstrate the tracking performance of this version of the algorithm under the maximum capacity criterion in time-varying channels. We assume that each channel gain $h_{i,j}$ between a transmit and receive antenna remains constant over τ frame intervals (we assume that each frame interval corresponds to one iteration of the algorithm) and follows a first order AR dynamics over τ written as

$$h_{i,j}(t) = \alpha h_{i,j}(t-1) + \beta v_{i,j}(t) \qquad (23)$$

where α and β are the fixed parameters of the model related through $\beta = (1-\alpha^2)^{1/2}$ and $v_{i,j} \sim \mathcal{N}_c(0,1)$. In the simulations we set $\alpha = 0.9$, $\tau = 500$ and the constant step size $\mu = 0.002$. We consider $N_R = 12$, $n_R = 6$ and $n_T = 2$. We set $\rho = 10$dB and we use the ML channel estimate with $T = 6$ orthogonal training symbols. It has also been observed that in most cases $d(\omega^*[n], \omega^*[n-\tau]) \leq 2$ and therefore we set $D = 2$. The tracking performance of the algorithm is shown in Figure 4. It is seen that the algorithm closely tracks the best antenna subset.

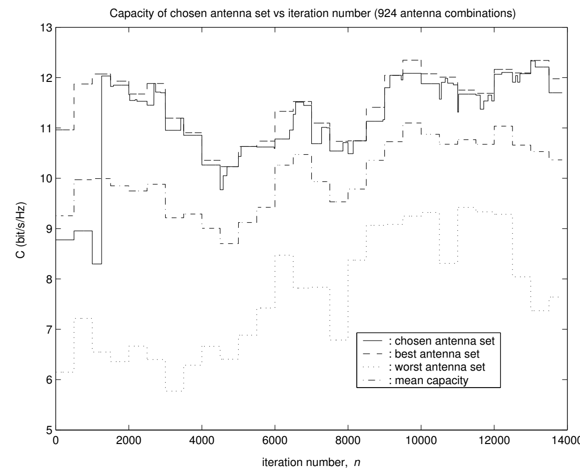

Fig. 4. Single run of the algorithm: the capacities of the chosen antenna subsets versus iteration number n.

VI. Conclusions

We have developed MIMO antenna selection algorithms based on various performance criteria in situations where only noisy estimates of the channels are available. The proposed techniques are based on discrete stochastic approximation algorithms, which generate a sequence of antenna subsets where each new subset is obtained from the previous one by taking a small step in a good direction towards the global optimizer. One salient feature of the proposed approach is that no closed-form expression for the objective function is needed and only an estimate of it is sufficient. We have also developed antenna selection algorithms for time-varying scenarios where the optimal antenna subset is slowly varying. Simulation results have demonstrated the excellent performance of these new MIMO antenna selection algorithms.

References

[1] S. Andradottir, "A global search method for discrete stochastic optimization," *SIAM J. Optimization*, 6(2):513–530, May 1996.

[2] D.A. Gore and A.J. Paulraj, "MIMO antenna subset selection with space-time coding," *IEEE Trans. Sig. Proc.*, 50(10):2580–2588, Oct. 2002.

[3] R. Heath and A. Paulraj, "Antenna selection for spatial multiplexing systems based on minimum error rate," In *Proc. IEEE Int. Conf. Commun. (ICC)*, Helsinki, Finland, Jun. 2001.

[4] R.A. Horn and C.R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, U.K., 1985.

[5] T.L. Marzetta, "BLAST training: estimation channel characteristics for high-capacity space-time wireless," In *Proc. 37th Annual Allerton Conf. Commun., Comput. & Control*, Monticello, IL, Sep. 1999.

[6] R. Nabar, D. Gore, and A. Paulraj, "Optimal selection and use of transmit antennas in wireless systems," In *Proc. Int. Conf. Telecommun. (ICT)*, Acapulco, Mexico, May 2000.

[7] I.E. Telatar, "Capacity of multi-antenna Gaussian channels," *Euro. Trans. Telecommun.*, 10(6):585–595, Nov. 1999.