MULTIACCESS INTERFERENCE REDUCTION IN OSTBC-MIMO SYSTEMS BY ADAPTIVE PROJECTED SUBGRADIENT METHOD

Renato L. G. Cavalcante and Isao Yamada

Department of Communications and Integrated Systems S3-60
Tokyo Institute of Technology, Tokyo 152-8552, Japan
Email: {renato,isao}@comm.ss.titech.ac.jp

ABSTRACT

We introduce adaptive linear filters based on the adaptive projected subgradient method that are suitable for online implementation of multiple access interference (MAI) suppression in OSTBC-MIMO systems. The proposed adaptive filters track the optimal solution of a new cost function that is robust against channel state information (CSI) mismatch. The adaptive update algorithm is based on projections onto closed convex sets that contain the optimal solution with high probability. The main features of the adaptive filters are that no matrix inversion of a sample covariance matrix is required and that a low-complexity recursive implementation is possible. Convergence analysis and simulation results show the effectiveness of the proposed schemes.

Index Terms— Adaptive filters, Interference suppression, Multiuser channels, MIMO systems.

1. INTRODUCTION

Orthogonal space-time block codes (OSTBCs) are efficient codes to implement low-complexity receivers enjoying full diversity [1]. When only one user is present in an additive white Gaussian noise (AWGN) channel, maximum likelihood (ML) detection can be realized with a simple linear filter that can be interpreted as a space-time matched filter (STMF) [2]. However, if multiple users access the channel at the same time, a STMF, which is matched to one specific user, does not take into account the structure of this multiple access interference (MAI), which cannot be modeled as white noise and severely degrades the system performance. Unfortunately, ML detection has high complexity and requires knowledge of the channels of all interfering users [2, Ch. 11]. Therefore, suboptimal receivers that exploit the structure of MAI have been proposed as an alternative to the conventional STMF and ML receivers [1]–[3] .

Blind linear receivers that reduce MAI and optionally completely eliminate self-interference without excessively amplifying noise have been introduced in [1]. These receivers are based on the inversion of a sample covariance matrix of a block of received symbols, hence they can be hard to be implemented for online adaptation with low computational complexity. (The size of the matrix to be inverted increases if either the number of times that the channel is used to transmit one OSTBC or the number of receiver antennas increases.) Moreover, in practical situations, an *ad-hoc* choice of a diagonal loading factor or a certain worst case optimization [3], which further increases the computational complexity, is necessary to provide robustness against imperfect channel state information (CSI).

We introduce a new cost function of which the solution is a linear filter that is robust against CSI mismatch and reduces to the one proposed in [1] when CSI is perfect. Instead of using standard adaptive algorithms that try to minimize directly our proposed robust cost function, we use the adaptive projected subgradient method to derive a recursion that minimizes a series of cost functions. The minimum of this series is achieved in the intersection of closed convex sets that contain the optimal solution of our robust cost function with high probability. The resulting algorithms are suitable for online adaptation, provide good steady-state performance and convergence speed, and give rise to an efficient recursive implementation that does not require any matrix inversion and converges in the mean sense to the desired filter. Simulation results are provided to illustrate the effectiveness of our algorithms. Due to the space limitation, the proofs of this paper are omitted.

2. SYSTEM MODEL

Consider a MIMO system with P users using the same OSTBC (these assumptions can be relaxed to users that use different linear codes) and N transmitter antennas. The code rate is $R = K/T$, where K is the number of symbols transmitted at T time instants by each user. The receiver has M antennas, and user one ($p = 1$) is the desired user, where p is the user number. The received signal $\boldsymbol{Y}$, at time $i \in \mathbb{N}^*$, can be expressed as in [1]

$$\boldsymbol{Y}[i] = \sum_{p=1}^{P} \boldsymbol{X}_p(\boldsymbol{s}_p[i])\boldsymbol{H}_p + \boldsymbol{V}[i] \in \mathbb{C}^{T\times M}, \tag{1}$$

where $\boldsymbol{H}_p \in \mathbb{C}^{N\times M}$ is the channel matrix between the pth transmitter and the receiver, $\boldsymbol{s}_p[i] \in \mathbb{C}^K$ is the vector of K symbols transmitted by the pth user, and $\boldsymbol{V}[i] \in \mathbb{C}^{T\times M}$ is the noise matrix. The xth row and yth column of $\boldsymbol{H}_p$ is the complex channel gain between the xth antenna of the pth user and the yth antenna of the receiver. The matrix $\boldsymbol{X}_p(\boldsymbol{s}_p[i])$ is the OSTBC of the pth user associated with the vector $\boldsymbol{s}_p[i]$ and is formed according to

$$\boldsymbol{X}_p(\boldsymbol{s}_p[i]) = \sum_{k=1}^{K} [\boldsymbol{C}_k \mathrm{Re}\,(s_{p,k}[i]) + \boldsymbol{D}_k \mathrm{Im}\,(s_{p,k}[i])] \in \mathbb{C}^{T\times N},$$

where $\{\boldsymbol{C}_k, \boldsymbol{D}_k\}$ are code matrices of dimension $T \times N$ [2, Ch. 7] and $\boldsymbol{s}_p[i] =: [s_{p,1}[i]\ s_{p,2}[i] \ldots s_{p,K}[i]]^T$.

Defining the "underline" operator for any given matrix $\boldsymbol{M}$, $\underline{\boldsymbol{M}} = \left[\mathrm{vec}[\mathrm{Re}\,(\boldsymbol{M})]^T\ \mathrm{vec}[\mathrm{Im}\,(\boldsymbol{M})]^T\right]^T$, where vec is the column-stacking operator, (1) can be rewritten as in [2]

$$\underline{\boldsymbol{Y}}[i] = \sum_{p=1}^{P} \boldsymbol{A}_p \underline{\boldsymbol{s}}_p[i] + \underline{\boldsymbol{V}}[i] \in \mathbb{R}^{2MT}, \tag{2}$$

where

$$\begin{aligned}\boldsymbol{A}_p &= [\underline{\boldsymbol{C}_1\boldsymbol{H}_p} \cdots \underline{\boldsymbol{C}_K\boldsymbol{H}_p}\ \underline{\boldsymbol{D}_1\boldsymbol{H}_p} \cdots \underline{\boldsymbol{D}_K\boldsymbol{H}_p}] \\ &= [\boldsymbol{a}_{p,1}\ \boldsymbol{a}_{p,2} \cdots \boldsymbol{a}_{p,2K}] \in \mathbb{R}^{2MT\times 2K},\end{aligned}$$

and $\underline{\boldsymbol{s}}_p[i] =: [b_{p,1}[i] \quad \ldots \quad b_{p,2K}[i]]^T \in \mathbb{R}^{2K}$. (Note that $b_{p,a}[i] = \mathrm{Re}\,(s_{p,a}[i])$ for $1 \le a \le K$ and $b_{p,a}[i] = \mathrm{Im}\,(s_{p,a-K}[i])$ for $K+1 \le a \le 2K$.) The matrix $\boldsymbol{A}_p$ satisfies $\boldsymbol{A}_p^T\boldsymbol{A}_p = \|\boldsymbol{H}_p\|_F^2 \boldsymbol{I}_{2K}$ [2], where $\boldsymbol{I}_{2K}$ is the identity matrix of dimension $2K \times 2K$. The following assumptions are used.

Assumption 1 *The data symbols $\boldsymbol{s}_{p,k}[i]$, $1 \le p \le P$, $1 \le k \le K$, are chosen from a quadrature phase shift keying (QPSK) constellation and belong to the set $\{-1-1j, -1+1j, 1-1j, 1+1j\}$, where $j = \sqrt{-1}$. Additionally, these symbols are modeled as independent, zero-mean random variables with equal probability.*

Assumption 2 *The data symbols from different users and noise are mutually uncorrelated. Additionally, the elements of the noise matrix $\boldsymbol{V}[i]$ are independent, circularly symmetric, complex zero-mean Gaussian random variables with variance σ_V^2.*

Assumption 3 *The condition $MT > PK$ is satisfied.*

3. MULTIACCESS MIMO LINEAR RECEIVER

In communication systems, in order to reduce BER, low-complexity receivers usually apply a linear filter $\boldsymbol{W}$, which is the solution of a fairly simple optimization problem, to the received signal $\underline{\boldsymbol{Y}}[i]$. Under Assumption 1, the transmitted symbols are usually estimated by $\underline{\widetilde{\boldsymbol{s}}}_p[i] = \mathrm{sgn}(\boldsymbol{W}^T\underline{\boldsymbol{Y}}[i])$, where sgn is the element-wise signum function and $\underline{\widetilde{\boldsymbol{s}}}_p[i]$ is the estimate of $\underline{\boldsymbol{s}}_p[i]$. In [1], the following filter has been proposed to reduce MAI: $\boldsymbol{W}^\star_{\mathrm{opt}} = [\boldsymbol{w}^\star_{\mathrm{opt},1} \ \ldots \ \boldsymbol{w}^\star_{\mathrm{opt},2K}]$, where

$$\boldsymbol{w}^\star_{\mathrm{opt},k} \in \arg\min_{\boldsymbol{w}_k \in C_k^\star} \boldsymbol{w}_k^T \boldsymbol{R} \boldsymbol{w}_k, \ C_k^\star := \{\boldsymbol{w} \in \mathbb{R}^{2MT} | \boldsymbol{A}_1^T \boldsymbol{w} = \boldsymbol{e}_k\}, \tag{3}$$

$\boldsymbol{e}_k$ is the vector of zeros, except for the kth entry, which is one, and $\boldsymbol{R} = E\{\underline{\boldsymbol{Y}}[i] \ \underline{\boldsymbol{Y}}[i]^T\}$. The idea behind the above filter is to minimize the filter output energy under the constraint $C_k^\star$ ($k = 1, \ldots, 2K$), which eliminates self-interference and keeps the desired user's power. Unless stated explicitly, we assume that $\boldsymbol{R}$ has full rank. In such a case, $\boldsymbol{w}^\star_{\mathrm{opt},k}$ is unique and is given by

$$\boldsymbol{w}^\star_{\mathrm{opt},k} = \boldsymbol{R}^{-1}\boldsymbol{A}_1(\boldsymbol{A}_1^T\boldsymbol{R}^{-1}\boldsymbol{A}_1)^{-1}\boldsymbol{e}_k. \tag{4}$$

Unfortunately, the performance of receivers based on (4) with $\boldsymbol{A}_1$ and $\boldsymbol{R}$ replaced by their estimates can be very sensitive to estimation errors [1].

The transmitted symbols belong to a small finite set in communication systems, so it is reasonable to use estimates of these symbols to improve the performance against MAI and imperfect CSI. Assuming for the moment that the estimates are perfect (this assumption will be dropped later), to accommodate this additional information, we propose a filter that minimizes the mean-square error between the filter output and the desired output scaled by a constant under a practical constraint C_k. More precisely, the kth column of the proposed filter $\boldsymbol{W}_{\mathrm{opt}}(\boldsymbol{\alpha}) := [\boldsymbol{w}_{\mathrm{opt},1}(\alpha_1) \ \ldots \ \boldsymbol{w}_{\mathrm{opt},2K}(\alpha_{2K})]$ is given by

$$\boldsymbol{w}_{\mathrm{opt},k}(\alpha_k) \in \arg\min_{\boldsymbol{w}_k \in C_k} E\{|\boldsymbol{w}_k^T\underline{\boldsymbol{Y}}[i] - \alpha_k b_{1,k}[i]|^2\} \tag{5}$$
$$C_k := \{\boldsymbol{w} \in \mathbb{R}^{2MT} | \widetilde{\boldsymbol{A}}_1^T \boldsymbol{w}_k = \boldsymbol{e}_k\},$$

where $\widetilde{\boldsymbol{A}}_1$ is a possibly erroneous estimate of $\boldsymbol{A}_1$, $\boldsymbol{\alpha} = [\alpha_1 \ \ldots \ \alpha_{2K}]^T$, and $\alpha_k \geq 0$ is a properly chosen constant. The relation between the estimate of the channel $\widetilde{\boldsymbol{H}}_1$ and $\widetilde{\boldsymbol{A}}_1$ is given by $\widetilde{\boldsymbol{A}}_1 = [\underline{\boldsymbol{C}_1\widetilde{\boldsymbol{H}}_1} \ \cdots \ \underline{\boldsymbol{C}_K\widetilde{\boldsymbol{H}}_1} \ \underline{\boldsymbol{D}_1\widetilde{\boldsymbol{H}}_1} \ \cdots \ \underline{\boldsymbol{D}_K\widetilde{\boldsymbol{H}}_1}]$. A reasonable choice for α_k is addressed in Sect. 4. A closed form expression for the filter in (5) is given by the following proposition.

Proposition 1 *If $\boldsymbol{R}$ has full rank, the filter in (5) is uniquely given by*

$$\boldsymbol{w}_{\mathrm{opt},k}(\alpha_k) = \alpha_k\boldsymbol{R}^{-1}\boldsymbol{a}_{1,k} - \alpha_k\boldsymbol{Q}\widetilde{\boldsymbol{A}}_1^T\boldsymbol{R}^{-1}\boldsymbol{a}_{1,k} + \boldsymbol{Q}\boldsymbol{e}_k, \tag{6}$$

where $\boldsymbol{Q} = \boldsymbol{R}^{-1}\widetilde{\boldsymbol{A}}_1(\widetilde{\boldsymbol{A}}_1^T\boldsymbol{R}^{-1}\widetilde{\boldsymbol{A}}_1)^{-1}$. In particular, if $\boldsymbol{A}_1 = \widetilde{\boldsymbol{A}}_1$, $\boldsymbol{w}_{\mathrm{opt},k}(\alpha_k)$ is reduced to the filter in (3), i.e., $\boldsymbol{w}_{\mathrm{opt},k}(\alpha) = \boldsymbol{w}^\star_{\mathrm{opt},k}$ $\forall\alpha_k \geq 0$ (the solution does not depend on α_k). [1]

[1]Whenever we assume perfect CSI, we define $\boldsymbol{w}_{\mathrm{opt},k} := \boldsymbol{w}_{\mathrm{opt},k}(\alpha_k)$ for notational simplicity.

4. PROPOSED ALGORITHM

Instead of directly minimizing a cost function related to (5), our adaptive filters suppress the following function of $\boldsymbol{w}_k$ under the constraint $\boldsymbol{w}_k \in C_k$ at time i,

$$\Theta[i](\boldsymbol{w}_k) := \begin{cases} \sum_{j=0}^{q[i]-1} \dfrac{\omega[i,j]}{L[i]} \|\boldsymbol{w}_k[i] - P_{\mathcal{K}_k[i-j]}(\boldsymbol{w}_k[i])\| \cdot \\ \quad \cdot\|\boldsymbol{w}_k - P_{\mathcal{K}_k[i-j]}(\boldsymbol{w}_k)\|, \quad \text{if} \quad L[i] \neq 0 \\ 0, \quad \text{otherwise}, \end{cases} \tag{7}$$

where $\mathcal{K}_k[i-j]$ ($j = 0, \ldots, q[i]-1$) are closed convex sets that contain the optimal filter $\boldsymbol{w}_{\mathrm{opt},k}(\alpha_k)$ with high probability, $\omega[i,j]$ is the weighting factor of the set $\mathcal{K}_k[i-j]$ ($\sum_{j=0}^{q[i]-1}\omega[i,j] = 1$ and $\omega[i,j] > 0$), $L[i] = \sum_{j=0}^{q[i]-1}\omega[i,j]\|\boldsymbol{w}_k[i] - P_{\mathcal{K}_k[i-j]}(\boldsymbol{w}_k[i])\|$, and $P_{\mathcal{K}_k[i]}$ is the projection onto $\mathcal{K}_k[i]$. It is clear that $\Theta[i]$ achieves 0 when the filter $\boldsymbol{w}_k$ belongs to the intersection $\bigcap_{j=0}^{q[i]-1}\mathcal{K}_k[i-j]$. Therefore, if $\boldsymbol{w}_{\mathrm{opt},k}(\alpha_k) \in \bigcap_{j=0}^{q[i]-1}\mathcal{K}_k[i]$, and this intersection is fairly small, a filter that suppresses $\Theta[i]$ is expected to be a good approximation of $\boldsymbol{w}_{\mathrm{opt},k}(\alpha_k)$.

As a low-complexity algorithm, we use the adaptive projected subgradient method, which minimizes asymptotically over C_k the sequence of non-negative functions $\Theta[i]$ ($i = 1, 2, \ldots$) in (7). Substitution of (7) into [4, Eq. (11)] with $C := C_k$ yields the following algorithm.

Algorithm 1

$$\boldsymbol{w}_k[i+1] = P_{C_k}\left(\boldsymbol{w}_k[i] + \lambda[i]\left(\sum_{j=0}^{q[i]-1}\omega[i,j]P_{\mathcal{K}_k[i-j]}(\boldsymbol{w}_k[i]) - \boldsymbol{w}_k[i]\right)\right),$$

where $\lambda[i] \in \Big(0, 2\mathcal{M}[i](\boldsymbol{w}_k[i])\Big)$ is the step size, $\sum_{j=0}^{q[i]-1}\omega[i,j] = 1$, $\omega[i,j] > 0$, $P_{C_k}(\boldsymbol{w}) = \boldsymbol{P}\boldsymbol{w} + \widetilde{\boldsymbol{A}}_1(\widetilde{\boldsymbol{A}}_1^T\widetilde{\boldsymbol{A}}_1)^{-1}\boldsymbol{e}_k$, $P_{\mathcal{K}_k[i]}$ is the projection onto $\mathcal{K}_k[i]$, $\boldsymbol{P} = \boldsymbol{I} - \widetilde{\boldsymbol{A}}_1\widetilde{\boldsymbol{A}}_1^T/\|\widetilde{\boldsymbol{H}}_1\|_F^2$, and

$$\mathcal{M}[i](\boldsymbol{w}) = \begin{cases} \dfrac{\sum_{j=0}^{q[i]-1}\omega[i,j] \ \|P_{\mathcal{K}_k[i-j]}(\boldsymbol{w}) - \boldsymbol{w}\|^2}{\|\sum_{j=0}^{q[i]-1}\omega[i,j]P_{\mathcal{K}_k[i-j]}(\boldsymbol{w}) - \boldsymbol{w}\|^2}, \\ \qquad \text{if} \quad \boldsymbol{w} \notin \bigcap_{j=0}^{q[i]-1}\mathcal{K}_k[i-j] \\ 1, \quad \text{otherwise}. \end{cases}$$

A natural choice for the set $\mathcal{K}_k[i]$ can be easily obtained by dropping the expectation operator in (5),

$$\mathcal{K}_k[i] := \{\boldsymbol{w}_k \in \mathbb{R}^{2MT} | \ |\boldsymbol{w}_k^T\underline{\boldsymbol{Y}}[i] - \alpha_k b_{1,k}[i]|^2 \leq \rho^2\}, \quad k = 1, \ldots, 2K, \tag{8}$$

where $\rho > 0$ increases the probability that $\boldsymbol{w}_{\mathrm{opt},k}(\alpha_k) \in \mathcal{K}_k[i]$ when noise is present. The set $\mathcal{K}_k[i]$ is a closed hyperslab, and thus the projection onto this set is very simple [5]. If $\alpha_k \neq 0$ is used, we have to replace $b_{1,k}[i]$ by its tentative estimate $\widetilde{b}_{1,k}[i] := \mathrm{sgn}(\boldsymbol{w}_k[i]^T\underline{\boldsymbol{Y}}[i])$ because $b_{1,k}[i]$ is not usually available. Indeed, we find through numerical simulations in Sect. 5 that the replacement is practically effective and does not degrade severely the steady-state performance as compared to the case when training data are available. If the estimated channel matrix $\widetilde{\boldsymbol{H}}_1$ and the true channel matrix $\boldsymbol{H}_1$ are sufficiently close to each other, the set C_k provides robustness against symbol estimation errors. The influence of the parameters α_k and ρ on the set-membership probability of $\boldsymbol{w}_{\mathrm{opt},k}(\alpha_k) \in \mathcal{K}_k[i]$ is given by Proposition 2.

Proposition 2 *On the set-membership probability of $\boldsymbol{w}_k$ ($\in C_k$) in $\mathcal{K}_k[i]$, the following holds*

(a) $P\{|\boldsymbol{w}_k^T \underline{\boldsymbol{Y}}[i] - \alpha_k b_{1,k}[i]|^2 \le \rho^2\} \ge Q_{[i]}(\boldsymbol{w}_k, \alpha_k)$, *where* $Q_{[i]}(\boldsymbol{w}_k, \alpha_k) := 1 - E\{|\boldsymbol{w}_k^T \underline{\boldsymbol{Y}}[i] - \alpha_k b_{1,k}[i]|^2\}/\rho^2$, $(\forall \boldsymbol{w}_k \in C_k, \forall \alpha_k \ge 0)$.

(b) Among all filters $\boldsymbol{w}_k$ *in* C_k, $\boldsymbol{w}_{\mathrm{opt},k}(\alpha_k)$ *maximizes the lower bound* $Q_{[i]}(\boldsymbol{w}_k, \alpha_k)$ *of the set-membership probability, i.e.,* $Q_{[i]}(\boldsymbol{w}_{\mathrm{opt},k}(\alpha_k), \alpha_k) \ge Q_{[i]}(\boldsymbol{w}_k, \alpha_k)$.

(c) If CSI is perfect, although the optimal solution does not depend on α_k *[i.e,* $\boldsymbol{w}_{\mathrm{opt},k}(\alpha_k) = \boldsymbol{w}_{\mathrm{opt},k}$, $\forall \alpha_k \ge 0$ *(see Sect. 3)], the lower bound of the set-membership probability is maximized with* $\alpha_k = 1$ *and* $\boldsymbol{w}_k = \boldsymbol{w}_{\mathrm{opt},k}$, *i.e.,* $Q_{[i]}(\boldsymbol{w}_{\mathrm{opt},k}, 1) \ge Q_{[i]}(\boldsymbol{w}_k, \alpha_k)$ $(\forall \boldsymbol{w}_k \in C_k, \forall \alpha_k \ge 0)$.

4.1. Recursive implementation

If a large number of sets are desired, an efficient recursive algorithm can be devised from Algorithm 1 by proper selection of the parameters. The recursion in Algorithm 1 can be simplified as follows if the weights satisfy a geometric series with larger weights given to more recent data.

Algorithm 2 *For* $\rho = 0$, $\lambda[i] = \mu \in (0, 2)$, $q[i] = i$, $\omega[i,j] = \gamma^j / \sum_{m=0}^{i-1} \gamma^m$, $j = 0, \ldots, q[i]-1$, *where* $0 \le \gamma < 1$ *is the forgetting factor, Algorithm 1 reduces to*

$$\boldsymbol{w}_k[i+1] = \left(\boldsymbol{I} - \mu \boldsymbol{P} \frac{\boldsymbol{F}[i]}{S[i]}\right) \boldsymbol{w}_k[i] + \mu \alpha_k \boldsymbol{P} \frac{\boldsymbol{g}_k[i]}{S[i]},$$

where $S[i] = 1 + \gamma S[i-1]$, $\boldsymbol{F}[i] = \gamma \boldsymbol{F}[i-1] + \dfrac{\underline{\boldsymbol{Y}}[i]\underline{\boldsymbol{Y}}[i]^T}{\|\underline{\boldsymbol{Y}}[i]\|^2}$, $\boldsymbol{g}_k[i] = \gamma \boldsymbol{g}_k[i-1] + \dfrac{\underline{\boldsymbol{Y}}[i] b_{1,k}[i]^T}{\|\underline{\boldsymbol{Y}}[i]\|^2}$, $S[1] = 1$, $\boldsymbol{F}[1] = \underline{\boldsymbol{Y}}[1]\underline{\boldsymbol{Y}}[1]^T / \|\underline{\boldsymbol{Y}}[1]\|^2$, *and* $\boldsymbol{g}_k[1] = \underline{\boldsymbol{Y}}[1] b_{1,k}[1] / \|\underline{\boldsymbol{Y}}[1]\|^2$.

Unlike RLS-based algorithms, Algorithm 2 does not require a matrix inversion and can achieve good steady-state performance even with small forgetting factor values.

4.2. Convergence of the algorithm

We assume that the filter provides reliable estimates of $b_{1,k}[i]$ if $\alpha_k \ne 0$ although this assumption does not always hold in our simulations. We also use the following common assumptions in the analysis of adaptive filters.

Assumption 4 $\sum_{j=0}^{q[i]-1} \omega[i,j] \underline{\boldsymbol{Y}}[i-j] \underline{\boldsymbol{Y}}[i-j]^T / \|\underline{\boldsymbol{Y}}[i-j]\|^2$ *and* $\boldsymbol{w}_k[i]$ *are mutually independent* $\forall i$.

This assumption only holds exactly for $q[i] = 1$. However, it is a good approximation for large $q[i]$ because $\sum_{j=0}^{q[i]-1} \omega[i,j] \underline{\boldsymbol{Y}}[i-j] \underline{\boldsymbol{Y}}[i-j]^T / \|\underline{\boldsymbol{Y}}[i-j]\|^2$ can be approximated by its mean value, which is a constant, and thus independent of $\boldsymbol{w}_k[i]$, $\forall i$ (see, for example, [6, Sect. 6.9.2]).

Assumption 5

$E\left\{\underline{\boldsymbol{Y}}[i]\underline{\boldsymbol{Y}}[i]^T / \|\underline{\boldsymbol{Y}}[i]\|^2\right\} \approx E\{\underline{\boldsymbol{Y}}[i]\underline{\boldsymbol{Y}}[i]^T\} / E\{\|\underline{\boldsymbol{Y}}[i]\|^2\}$ *and* $E\left\{b_{1,k}[i]\underline{\boldsymbol{Y}}[i] / \|\underline{\boldsymbol{Y}}[i]\|^2\right\} \approx E\{b_{1,k}[i]\underline{\boldsymbol{Y}}[i]\} / E\{\|\underline{\boldsymbol{Y}}[i]\|^2\}$.

Closed form expressions for $E\{\underline{\boldsymbol{Y}}[i]\underline{\boldsymbol{Y}}[i]^T / \|\underline{\boldsymbol{Y}}[i]\|^2\}$ and $E\{b_{1,k}\underline{\boldsymbol{Y}}[i] / \|\underline{\boldsymbol{Y}}[i]\|^2\}$ are difficult to obtain, but similar approximations have been widely used in current literature [6, p. 301].

Proposition 3 *On the convergence of the proposed method, the following holds*

(a) In the noiseless case with perfect CSI, if $\alpha_k = 1$ *and* $\bigcap_{p=2}^{P} \mathrm{null}(\boldsymbol{A}_p^T) \bigcap C_k \ne \emptyset$, *Algorithms 1 and 2 satisfy* $\lim_{i\to\infty} \Theta[i](\boldsymbol{w}_k[i]) = 0$, *regardless of the possible choices of other parameters*[2].

(b) (Unbiasedness) When noise is present, under Assumptions 4 and 5, Algorithm 2 satisfies $\lim_{i\to\infty} E\{\boldsymbol{W}[i]\} = \boldsymbol{W}_{\mathrm{opt}}(\boldsymbol{\alpha})$, *where* $\boldsymbol{\alpha} = [\alpha_1 \ \ldots \ \alpha_{2K}]^T$, *and* $\alpha_k = \alpha \ge 0$, $k = 1, \ldots, 2K$.

5. SIMULATION RESULTS

We assume a system with four users ($P = 4$) and a receiver with $M = 8$ antennas. All users use the 1/2 rate OSTBC [1] with $K = 4$, $T = 8$, $N = 3$, and employ QPSK modulation (see Assumption 1). Ensemble average curves are obtained by averaging the mismatch $\|\boldsymbol{W}[i-1]^T \underline{\boldsymbol{Y}}[i] - \underline{\boldsymbol{s}}_1[i]\|^2$ over 500 realizations. Since the receivers in [1] are batch receivers [3], for fairness, final symbol detection for our receivers is made using $\boldsymbol{W}[i = L]$, where L is the number of blocks considered at each realization. Symbol error rate (SER) curves are obtained by considering $L = 500$ blocks of transmitted symbols (in 100 realizations). The signal-to-noise ratio (SNR) for each element of $b_{1,k}[i]$ is given by $\mathrm{SNR} = E\{\mathrm{tr}\left(\boldsymbol{X}\boldsymbol{X}^{\boldsymbol{H}}\right)\} E\{\|\boldsymbol{H}_1\|_F^2\} / (N\sigma_V^2 K)$ [2]. Initially, the elements of the channel matrices are circularly symmetric, complex zero-mean Gaussian random variables with unit variance, and then the channels are normalized to yield $\|\boldsymbol{H}_p\|_F^2 = 1, \forall p$.

We consider three different versions of Algorithm 2 (Proposed-a, Proposed-b, and Proposed-c), the parameters of which are chosen as follows: $\mu = 1$ and $\alpha_k = 1$ $(\forall k)$. They differ in the choice of γ, which is set to 0, 0.9, and 0.99, for Proposed-a, Proposed-b, and Proposed-c, respectively. The true transmitted symbols are replaced by their rough estimates. We compare our algorithms with the optimal filter $\boldsymbol{W}_{\mathrm{opt}}^{\star}$, STMF ($\boldsymbol{W}[i] = \boldsymbol{A}_1^T / \|\boldsymbol{H}_1\|_F^2$), an "adaptive" version of the diagonally loaded minimum variance (DLMV) receiver of [1] [$\boldsymbol{W}[i] = (\widetilde{\boldsymbol{R}}[i] + \beta \boldsymbol{I})^{-1} \boldsymbol{A}_1 (\boldsymbol{A}_1^T (\widetilde{\boldsymbol{R}}[i] + \beta \boldsymbol{I})^{-1} \boldsymbol{A}_1)^{-1}$, where $\widetilde{\boldsymbol{R}}[i] := 1/i \sum_{j=1}^{i} \underline{\boldsymbol{Y}}[j] \underline{\boldsymbol{Y}}[j]^T$ and $\beta = 5\sigma_V^2$], and the direct matrix inversion (DMI) detector, which is essentially the DLMV receiver with a small, fixed diagonal loading factor ($\beta = 10^{-6}$). These "adaptive" versions are used to show the tracking characteristics of the receivers. The DMI algorithm is actually tracking our proposed optimal linear filter with $\alpha_k = 0$ $(\forall k)$[4].

Figures 1 and 2 show the performance of the algorithms when CSI is perfect at the receiver. As γ increases, the steady-state performance improves due to the increased memory of the algorithm, and smaller mismatch results in lower SER.

In Fig. 3 we consider imperfect CSI, i.e., $\widetilde{\boldsymbol{H}}_1 = \boldsymbol{H}_1 + \boldsymbol{E}$, where the elements of $\boldsymbol{E}$ are complex, zero-mean Gaussian random variables with unit variance. Then the matrix $\boldsymbol{E}$ is normalized so that the mismatch $\|\boldsymbol{E}\|_F^2 / \|\boldsymbol{H}_1\|_F^2$ is set to 0.1 in every realization. The proposed adaptive algorithms outperform the DLMV and DMI algorithms in this scenario. The poor performance of the DLMV is due to the finite sample effect of the sample covariance matrix. The DLMV algorithm provides robustness by adjusting the diagonal loading factor, whereas our algorithms provide robustness by using rough estimates of the desired user's symbols.

In order to show good convergence properties of the DLMV algorithm, in Fig. 4 we increase the interference level ($\|\boldsymbol{H}_p\|_F = 10\|\boldsymbol{H}_1\|_F$, $p \ne 1$). Other parameters are the same as in Fig. 1. We see that the convergence properties of the DLMV algorithm greatly improve in highly interfered channels. Proposed-c is slower than Proposed-b and Proposed-a because the symbol estimates at the beginning of the simulation are not reliable and the memory of the

[2]This property holds without any approximation and even when $\boldsymbol{R}$ does not have full rank.

[3]The receivers of [1] first receive all symbol blocks, estimate the matrix $\boldsymbol{R}$, calculate the approximate optimal filter with the available information (estimates of $\boldsymbol{R}$ and $\boldsymbol{A}_1$), and finally decode all symbols with the same filter.

[4]To see this, we only need to substitute $\alpha_k = 0$ and $\boldsymbol{R} = \widetilde{\boldsymbol{R}}$ into (6).

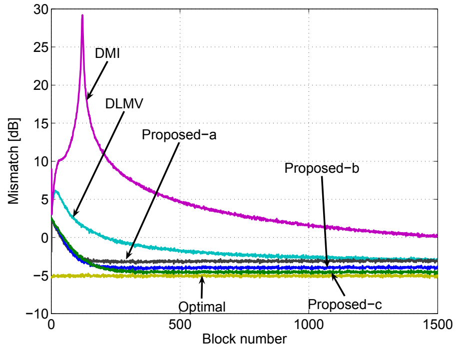

Fig. 1. Mismatch as a function of the number of received blocks. Perfect CSI and SNR 15 dB.

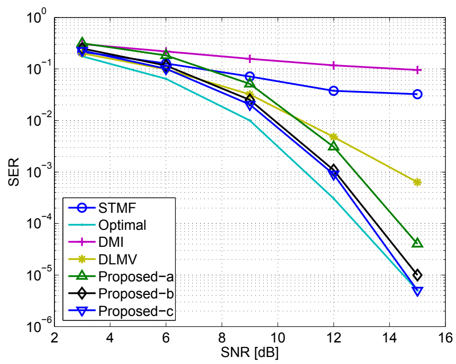

Fig. 2. SER as a function of SNR. Perfect CSI.

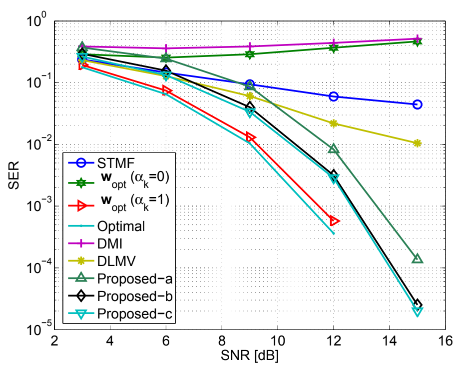

Fig. 3. SER as a function of SNR. Imperfect CSI.

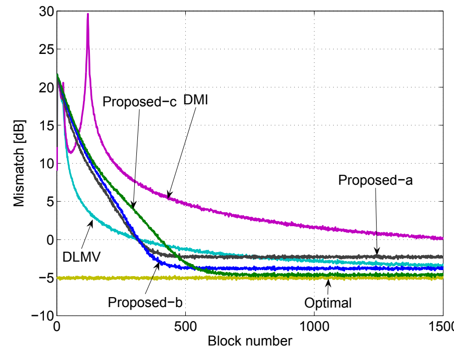

Fig. 4. Mismatch as a function of the number of received blocks. Perfect CSI, high interference level, and SNR 15 dB.

algorithm is large. If we replace the rough estimates by the true transmitted symbols, the convergence speed of Proposed-c is recovered in relation to Proposed-a and Proposed-b.

6. CONCLUSIONS

We have proposed algorithms based on the adaptive projected subgradient method to MIMO systems. Our algorithms converge to a set that includes the desired filter(s) when the covariance matrix of the input signal does not have full rank. Additionally, our recursive filters can provide good steady-state performance whether the memory of the algorithm is small or large. These characteristics contrast with RLS-based algorithms, which usually require a large memory to provide good steady-state performance and additional techniques to stabilize the filter when the covariance matrix does not have full rank. We have also proved that the recursive implementation of the algorithm converges in the mean sense to the optimal filter. Finally, the convergence speed of all proposed schemes can be further improved with a technique called p-times acceleration [4, Remark 1(c)], and our results can be straightforwardly extended to CDMA systems and adaptive beamforming.

Acknowledgment: The authors would like to thank Prof. Kohichi Sakaniwa of Tokyo Institute of Technology and Mr. António Trindade from Chipidea Microelectrónica for helpful suggestions and comments. This work was partially supported by JSPS Grants-in-Aid (C-17500139).

7. REFERENCES

[1] S. Shahbazpanahi, M. Beheshti, A. B. Gershman, M. Gharavi-Alkhansari, and K. M. Wong, “Minimum variance linear receivers for multiaccess MIMO wireless systems with space-time block coding,” *IEEE Trans. Signal Processing*, vol. 52, no. 12, pp. 3306–3313, Dec. 2004.

[2] E. G. Larsson and P. Stoica, *Space-Time Block Coding for Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[3] Y. Rong, S. Shahbazpanahi, and A. B. Gershman, “Robust linear receivers for space-time block coded multiaccess MIMO systems with imperfect channel state information,” *IEEE Trans. Signal Processing*, vol. 53, no. 8, pp. 3081–3090, Aug. 2005.

[4] I. Yamada and N. Ogura, “Adaptive subgradient method for asymptotic minimization of sequence of nonnegative convex functions,” *Numerical Functional Analysis and Optimization*, vol. 25, no. 7/8, pp. 593–617, 2004.

[5] M. Yukawa, R. L. G. Cavalcante, and I. Yamada, “Efficient blind MAI suppression in DS/CDMA systems by embedded constraint parallel projection techniques,” *IEICE Trans. Fundamentals*, vol. E88-A, no. 8, pp. 2062–2071, Aug. 2005.

[6] A. H. Sayed, *Fundamentals of Adaptive Filtering*. Hoboken, NJ: Wiley, 2003.