Blind Adaptive Channel Estimation For Decoding of Orthogonal Space-Time Block Codes

Yanmin Sheng
Department of Automation,
University of Science and Technology of China,
Hefei, China
Email: ymsheng@mail.ustc.edu.cn

HongSheng Xi*
Department of Automation,
University of Science and Technology of China,
Hefei, China
Email: xihs@ustc.edu.cn

Feng Yang
and Zilei Wang
Department of Automation,
University of Science and Technology of China,
Hefei, China

Abstract—in this paper, we provide a blind adaptive channel estimation method for decoding of orthogonal space-time block codes (OSTBC), which is built on the minimizing of an unconstrained cost function. The global minima of the cost function are analyzed. Both LMS and RLS algorithm realizations are given. Our method can adaptively track the time varying channel and decode transferred symbols without pilot symbols.

I. INTRODUCTION

Recently Multiple-input-Multiple-output Space-Time System has been a promising approach to exploit spatial diversity in Wireless Communication System. Space-Time Block Code based on Orthogonal design is an attractive code because it can provide full diversity gain and make receiver decode simple [1][2]. When the receiver knows the Channel State Information (CSI), the optimum ML receiver can be viewed as simple linear receiver followed by the symbol-by-symbol decoder. The linear receiver also makes the output signal-to-noise ratio (SNR) maximized. Pilot Symbols (training sequence) can help receiver estimate the CSI. However, when the channel changes, another training sequence must be sent to the receiver again. Thus, exploiting training signals reduces the bandwidth efficiency. Hence, blind techniques which do not use any training sequence are of great interest.

The blind methods can be divided into two different kinds: directly decoding without estimating the CSI and estimating the CSI first, then decoding. In [3] Zero-Forcing Blind Method is applied for called Generalized STBC without estimating the CSI. But the method needs several pilot symbols and amplitudes the noise. In [4] a closed-form blind method is provided and has a good global convergence, but it needs a lot of receive signals to estimate the covariance of the receive sequence and cannot track the changes of the channel. In [5], a blind method based on subspace is provided. However, the method estimates and decomposes the covariance of the signal sequence, which is complex and sensitive to the estimation of the covariance.

In this paper, we provide a blind adaptive on-line method to channel estimation for decoding of OSTBC. The method applies in most OSTBC except for a few exceptions. We give both LMS and RLS algorithm realizations. Our method can adaptively track the dynamic channel with low complexity and decode transferred symbols symbol by symbol without pilot symbols.

Notations: $Re[s]$ and $Im[s]$ denote the real and imaginary parts of $s$ respectively. For any complex Matrix $P_{ij}$, the underline operator for $P$ is defined as a $2IJ \times 1$ column vector by stacking the real and imaginary parts of all the elements of $P$ in a single column, so that:

$$
P = [Re(P_{11}), Re(P_{21}), \ldots, Re(P_{IJ}), IM(P_{11}), IM(P_{21}), \ldots, IM(P_{IJ})]^T \tag{1}
$$

And $e_k$ is a $k \times 1$ vector having one in the $k$th position and zeros elsewhere. $\|\|_F$ denotes the Frobenius norm and $\|\|$ denotes the Euclidean norm. $I_{k}$ is the $k \times k$ identity matrix. $\otimes$ means Kronecker product. $(.)^T$ stands for the transpose, and $(.)^H$ stands for the Hermitian transpose.

II. SYSTEM MODEL

The relationship between the input and the output of a Space-Time MIMO system with $N_t$ transmit and $N_r$ receive antennas in flat block-fading channel can be expressed as [1]:

$$
y(t) = x(t)h_s + \epsilon(t) \tag{2}
$$

Where $y(t) = [y_1(t), \ldots, y_{N_r}(t)]$, $x(t) = [x_1(t), \ldots, x_{N_t}(t)]$ and $n(t) = [n_1(t), \ldots, n_{N_r}(t)]$ are the complex row vectors of the received signals, transmitted signals and noise at time $t$, respectively. $H$ is the $N_r \times N_t$ complex channel matrix. Assume that the noise is both spatially and temporally white with the variance of $\sigma^2$ per complex dimension.

Collecting all the signals of one data block results in

$$
Y = HX + N \tag{3}
$$

Where $Y = [y^T(1), \ldots, y^T(T)]^T$, $X = [x^T(1), \ldots, x^T(T)]^T$ and $N = [n^T(1), \ldots, n^T(T)]^T$ are the matrices of the received signals, transmitted signals, and noise, respectively, $T$ is the data block length. The case of the slow fading channel is considered, i.e., the channel coherence time is assumed to be substantially larger than the data block length $T$. The $T \times N_t$ matrix $X(s)$ is called an orthogonal STBC if [1]

All elements of $X(s)$ are linear functions of the $K$ complex
input signal variables $s_1, s_2, \ldots, s_K$ and their complex conjugates; 
. For any $s$, it satisfies $X^H(s)X(s) = |s|^2 I_{N_t}$. 
It has been verfied [6] that the matrix $X(s)$ can be written as 
$$X(s) = \sum_{k=1}^{K} \{C_k Re[s_k] + D_k Im[s_k]\}$$
(4) Where the matrix $C_k$ and $D_k$ are defined as $C_k = X(e_k)$, $D_k = X(\jmath e_k)$. In fact, any orthogonal STBC can be defined by its matrices $C_k$ and $D_k(k = 1, \ldots, K)$. Using (4), model (3) can be rewritten as [6]
$$Y = A(H)s + N$$
The $2NmT \times 2K$ Matrix $A(H)$ is given by [6]
$$A(H) = [C_1I, \ldots, C_KI, D_1I, \ldots, D_KI]$$
(6) For any OSTBC and any channel matrix $H$. The columns of $A(H)$ are orthogonal to each other and have equal norms independent of the values of the entries of $H$ [6].
$$A^T(H)A(H) = ||H||^2 I_{2K}$$
(7) The ML decoder can also be viewed as the MF receiver whose output SNR is maximized [2]. In the orthogonal STBC case, the ML decoder can be interpreted as the linear MF receiver which computes the following estimate of $s$ [6]:
$$s = \frac{1}{||H||^2}A^T(H)Y$$
(8) To use the ML receiver (8), receiver needs to know the channel matrix. This matrix can be obtained by sending training sequence which are known to the receiver.

III. BLIND ADAPTIVE ALGORITHM

In this section, we will give a blind adaptive method for channel estimation for decoding for OSTBC.

A. Preliminaries of System Model

Here we give some preliminaries for adaptive algorithm. First of all, one can rewrite $A(H)$ as following
$$A(h) = [a_1h, \ldots, a_Kh, a_{K+1}h, \ldots, a_{2K}h]$$
(9) Where $h = H$ is a $2NmT \times 1$ column vector and $a_i(i = 1, \ldots, 2K)$ are $2NmT \times 2$ matrices. These can be easily verified because $C_1I, \ldots, C_KI, D_1I, \ldots, D_KI$ are all linear in $h$. Let $\Gamma = [a_1^T, \ldots, a_K^T, a_{K+1}^T, \ldots, a_{2K}^T]^T$, $R = E[YY^T]$ Thus $\text{vec}(A(h)) = \Theta h$. [4] has given an conjecture that $\text{range}(A(h)) = \text{range}(A(h))$ if and only if $h = \alpha h$ and proven true via simulations for most OSTBC. Based on this conjecture, the conclusion $\Gamma(\Gamma^T_R \Gamma)^{-1} \Gamma$ has only one principal normalized eigenvector $\pm \vec{h}$ has been made.

Here, we give a lemma for OSTBC first, which is the special character of OSTBC:

**Lemma 1:**
$$a_i^T a_j + a_j^T a_i = 0 \quad \forall i \neq j$$
(10)

**Proof:** From (7) $A^T(h)A(h) = ||h||^2 I_{2K}$ Thus
$$\begin{pmatrix}
(a_1h)^T \\
(a_2h)^T \\
(a_{K+1}h)^T \\
(a_{2K}h)^T
\end{pmatrix} =
\begin{pmatrix}
|a_1h|^2 & a_1h^T a_2h & \ldots & a_1h^T a_{2K}h \\
|a_2h|^2 & a_2h^T a_2h & \ldots & a_2h^T a_{2K}h \\
|a_{K+1}h|^2 & a_{K+1}h^T a_{K+1}h & \ldots & a_{K+1}h^T a_{2K}h \\
|a_{2K}h|^2 & a_{2K}h^T a_{2K}h & \ldots & a_{2K}h^T a_{2K}h
\end{pmatrix} = ||h||^2 I_{2K}
$$
because $h$ is an arbitrary vector, So for any $2NmT \times 1$ vector $t = (a_1^T a_2 a_1) = \delta_{ij}||t||^2$ where $\delta_{ij}$ is a Kronecker delta. Thus for any $2NmT \times 1$ vector $t^T = (a_1^T a_2 a_1) = (a_1^T a_2 a_1) = (a_1^T a_2 a_1) = 0$. Since $a_1^T a_2 a_1$ is a symmetric real matrix, So $a_1^T a_2 a_1 = 0$.

For System Model (5) And the following scalar function
$$J(h) = E[|Y - A(h)X_1X_1^T H^T|^2] = tr(R) - 2tr[A^T(h)RA(h)]$$
(11)
$$+ 2tr[A^T(h)RA(h)A^T(h)A(h)]$$
We have two following theorems

**Theorem 1:** $h$ is a stationary point of $J(h)$ if and only if $\vec{h}$ is a normalized eigenvector of $\Gamma(\Gamma_R \Gamma)^{-1}$. 
**Proof:** Let $\nabla J$ be the gradient operator with respect to $\vec{h}$
$$\frac{1}{2} \nabla J = \vec{h}^T (\Gamma_R I - 2\Gamma_R \Gamma) (\vec{h}^T - 2^I_R \Gamma_R \Gamma \vec{h})$$
(12)
if $\Gamma_R \Gamma \Gamma_R \vec{h} = \lambda \vec{h}$ and $||\vec{h}|| = 1$, it is straightforward to show $\nabla J = 0$. Conversely, $\nabla J = 0$ means that
$$\vec{h}^T (\Gamma_R I - 2\Gamma_R \Gamma) (\vec{h}^T - 2^I_R \Gamma_R \Gamma \vec{h}) = 0$$
(13)
Since $\Gamma_R \Gamma \Gamma_R \vec{h}$ is positive defined, we conclude that $||\vec{h}|| = 1$. Taking this result into $\nabla J = 0$ leads to
$$\Gamma_R \Gamma \Gamma_R \vec{h} = \vec{h}^T (\Gamma_R I - 2 \Gamma_R \Gamma) (\vec{h}^T - 2^I_R \Gamma_R \Gamma \vec{h})$$
(14)
So $\vec{h}$ is an normalized eigenvector of $\Gamma_R \Gamma \Gamma_R \vec{h}$ with correspond eigenvalue $\vec{h}^T (\Gamma_R I - 2 \Gamma_R \Gamma) (\vec{h}^T - 2^I_R \Gamma_R \Gamma \vec{h})$.

**Theorem 2:** All stationary points of $J(h)$ are saddle points except when $h$ is the principal normalized eigenvector of $\Gamma_R \Gamma \Gamma_R \vec{h}$. In the case, $J(h)$ attains the global minimum. If the conjecture [4] holds true, the global minimum is $\pm \vec{h}$. And at the global minimum $J(h) = (N_mT - K)\vec{h}^2$. 
**Proof:** Let $\vec{h}$ be the $2NmT \times 2NmT$ Hessian Matrix of $J(h)$ with respect to $\vec{h}$
$$\frac{1}{2} H = \frac{1}{2} \nabla \nabla J = h^T \Psi h + 2 \vec{h}^T \Psi + 2 \vec{h}^T \Psi - \Psi$$
(15)
Where $\Psi = \Gamma_R \Gamma$. Let $\lambda_1$ be the eigenvalue correspond with eigenvector $\vec{h}$. And $\Psi = Qdiag(\lambda_1, \ldots, \lambda_{2NmT})Q^T$ be the ED, $Q = [q_1, \ldots, q_{2NmT}]$ and $q_1$ is the eigenvector the same as $\vec{h}$. 
An evaluation of $H$ at the stationary point results in
$$\frac{1}{2} H = \lambda_1 I + 4 \lambda_1 q_1 q_1^T - \Psi$$
(16)
We have following equations:

$$ I = QQ^T $$

$$ 4\lambda q_i q_i^T = Q \text{diag}(4\lambda_1, 0, \ldots, 0) Q^T $$

Thus

$$ \frac{1}{2} = \text{diag}(4\lambda_1, \lambda_1 - \lambda_2, \ldots, \lambda_1 - \lambda_{2N-1}) Q^T $$

If and only if \( \lambda_1 \) is the maximum eigenvalue, \( \mathbf{H} \) is positive define. Otherwise, \( \mathbf{H} \) is neither negative nor positive, which means \( J(\mathbf{h}) \) has the global minimum when \( \mathbf{h} \) is the principal normalized eigenvector. If the conjecture [4] holds true, the global minimum is \( \pm \frac{\mathbf{h}}{||\mathbf{h}||} \). And it is easy to verify \( J(\pm \frac{\mathbf{h}}{||\mathbf{h}||}) = (N_r T - K) \sigma^2 \)

We give the following remarks:

- \( J(\mathbf{h}) \) has a global minimum point at which \( \mathbf{h} = \pm \frac{\mathbf{h}}{||\mathbf{h}||} \) and no other local minima, a global convergence is guaranteed if seeks for \( \mathbf{h} \) via iterative methods.
- We do not need \( \mathbf{h} \) to be normalized, from the Theorems above, minimizing \( J(\mathbf{h}) \) automatically leads to a normalized vector \( \mathbf{h} \). So we need not to make \( \mathbf{h} \) normalized at each iteration step.
- An amplitude/phase ambiguity exists, which inherent to all blind estimation approaches using second order statistics and cannot be resolved without further side information. Here we can use

$$ \text{tr} \{ \mathbf{R} \} = 2K \text{tr} \{ \Sigma_s \} ||\mathbf{h}||^2 + N_r T \sigma^2 $$

And

$$ J(\mathbf{h})_{	ext{min}} = (N_r T - K) \sigma^2 $$

It is known that the amplitude of \( \mathbf{h} \) when \( \Sigma_s = E[\mathbf{s} \mathbf{s}^T] \) is . But for constant modulus constellation signals , only phase ambiguity should be resolved.

B. Adapted Cost Function

Replacing the exception of \( \mathbf{J}(\mathbf{h}) \) with the exponentially weighted sum, we have

$$ J(\mathbf{h})(t) = \sum_{i=1}^{t} \beta^{t-i} [\mathbf{y}(i) - \mathbf{h}(i) \mathbf{A}^T(\mathbf{h}(i)) \mathbf{Y}(i)]^2 $$

$$ = \text{tr} \{ \mathbf{R}(t) \} - 2 \text{tr} \{ \mathbf{A}(\mathbf{h}(i)) \mathbf{R}(t) \mathbf{A}(\mathbf{h}(i)) \} $$

$$ + 2 \text{tr} \{ \mathbf{A}^T(\mathbf{h}(i)) \mathbf{R}(t) \mathbf{A}(\mathbf{h}(i)) \mathbf{A}^T(\mathbf{h}(i)) \mathbf{A}(\mathbf{h}(i)) \} $$

(22)

\( J(\mathbf{h})(t) \) is identical to \( J(\mathbf{h}) \) except that the sample correlation matrix now is

$$ \mathbf{R}(t) = \sum_{i=1}^{t} \beta^{t-i} [\mathbf{y}(i) \mathbf{y}^T(i)] $$

(23)

Therefore both theorems hold true for \( J(\mathbf{h})(t) \). We can not use normal RLS rule to derive RLS-based algorithm because the cost function is forth-order about \( \mathbf{h} \). But we can use \( \mathbf{A}^T(\mathbf{h}(i - 1)) \mathbf{Y}(i) \) to approximate \( \mathbf{A}^T(\mathbf{h}(i)) \mathbf{Y}(i) \) in sum of \( J(\mathbf{h})(t) \) for \( 1 < i \leq t \). Because both of them are of little difference for stationary environment, particularly when \( i \) is close to \( t \). The difference may be large when \( i \gg t \) but the exponential affect decreasing with the forgetting factor in the sum value makes little influence on \( J(\mathbf{h}(t)) \). Therefore we have the following cost function as a good estimate for \( J(\mathbf{h}(i)) \):

$$ J(\mathbf{h})(i) = \sum_{j=1}^{t} \beta^{t-i} \| \mathbf{y}(i) - \mathbf{A}(\mathbf{h}(i)) \mathbf{A}(\mathbf{h}(i)) \|^2 $$

(24)

C. LMS Adaptive Algorithm

The gradient method about \( \mathbf{h}(i) \) to \( J(\mathbf{h}(i)) \) leads to

$$ \nabla J(\mathbf{h})(i) = - \sum_{j=1}^{t} \beta^{t-i} [\mathbf{y}(i) - \mathbf{A}(\mathbf{h}(i)) \mathbf{A}(\mathbf{h}(i)) \mathbf{y}(i)] $$

$$ = \sum_{j=1}^{t} \beta^{t-i} [\mathbf{z}(i) \mathbf{z}^T(i)] \mathbf{y}(i) $$

$$ + \sum_{j=1}^{t} \beta^{t-i} [\mathbf{z}(i) \mathbf{z}^T(i)] \mathbf{y}(i) $$

$$ = \sum_{j=1}^{t} \beta^{t-i} [\mathbf{z}(i) \mathbf{z}^T(i)] \mathbf{y}(i) $$

$$ - \sum_{j=1}^{t} \beta^{t-i} [\mathbf{z}(i) \mathbf{z}^T(i)] \mathbf{y}(i) $$

$$ = \sum_{j=1}^{t} \beta^{t-i} [\mathbf{z}(i) \mathbf{z}^T(i)] \mathbf{y}(i) $$

(25)

Where

$$ \mathbf{B}(i) = (\mathbf{z}(i) \otimes \mathbf{I}_{2N_T}) \mathbf{G} $$

(27)

To track the fast varying channel, Here we omit the sum of \( \beta^{t-i} [\mathbf{z}(i) \mathbf{b}(i)] \mathbf{b}(i) \mathbf{y}(i) \mathbf{y}(i) \mathbf{G}^T \). The instance value \( \mathbf{b}(i) \mathbf{b}(i) \mathbf{y}(i) \mathbf{y}(i) \mathbf{G} \mathbf{z}(i) \mathbf{y}(i) \mathbf{y}(i) \mathbf{G}^T \) is used instead of the sum value in (26):

$$ \nabla J(\mathbf{h})(i) = \mathbf{B}(i) \mathbf{h}(i) $$

(28)

Thus the adaptive update of \( \mathbf{h}(i) \) is given by:

$$ \mathbf{h}(i) = \mathbf{h}(i) - \mu \nabla J(\mathbf{h})(i) $$

$$ = \mathbf{h}(i) + \mu (\mathbf{z}(i) \otimes \mathbf{y}(i)) - \mathbf{B}(i) \mathbf{b}(i) \mathbf{h}(i) $$

(29)
D. RLS Adaptive Algorithm

Now let the cost function \( J(\tilde{h}(i)) \) be zero, that is \( \nabla J = 0 \) we have

\[
\sum_{i=1}^{t} \beta^{-i} [(z(i) \otimes I_{2NRT})^T [(z(i) \otimes I_{2NRT})] \tilde{h}(i) = \sum_{i=1}^{t} \beta^{-i} [(z^T(i) \otimes Y^T(i)) \Gamma]^T \]
\]

\[
(30)
\]

(30) can be rewritten if we use (27)

\[
\sum_{i=1}^{t} \beta^{-i} B^T(i) B(i) \tilde{h}(i) = \sum_{i=1}^{t} \beta^{-i} [T(z(i) \otimes Y(i))] \]

\[
(31)
\]

We can not directly use normal RLS algorithm here because \( B(i) \) is not a column vector. But we can verify the following theorem using lemma 1.

**Theorem 3:**

\[
B^T(i) B(i) = \|z(i)\|^2 I
\]

**Proof:** From (9)(25) and (27) \( B(i) \) can be expressed as:

\[
B(i) = \sum_{j=1}^{2K} (Y^T(j)a_j \tilde{h}(t-1)a_j)
\]

Thus

\[
B^T(i) B(i) = \left[ \sum_{j=1}^{2K} (Y^T(j)a_j \tilde{h}(t-1)a_j)^T \right] \left[ \sum_{k=1}^{2K} (Y^T(k)a_k \tilde{h}(t-1)a_k) \right]
\]

\[
\sum_{j=1}^{2K} (Y^T(j)a_j \tilde{h}(t-1))^T (Y^T(j)a_j \tilde{h}(t-1))
\]

\[
\sum_{j=1}^{2K} \|Y^T(j)a_j \tilde{h}(t-1)\|^2 a_j^T a_j
\]

\[
(34)
\]

From lemma 1 (34) can be written:

\[
B^T(i) B(i) = \sum_{j=1}^{2K} \|Y^T(j)a_j \tilde{h}(t-1)\|^2 I = \|z(i)\|^2 I
\]

\[
(35)
\]

The LMS and RLS adaptive algorithm are given in TABLE I and TABLE II respectively.

The initial value of \( d(0) \) and \( h(0) \) should be chosen suitably. And the forgetting factor also should be chosen suitably to track the change of the channel. The step size \( \mu \) is an important value to choose in LMS adaptive algorithm, and RLS algorithm has an adaptive step size \( d(n) \) compared to LMS, which shows a better convergence via simulations later.

E. Practical Issues

Both adaptive algorithms have a much low complexity with \( O(8N_{r}N_{t}K^{2}T^{2} \min(N_{r}, N_{t})) \) compared to other blind batch algorithms. The main calculation burden is on \( z(n) \) and \( g(n) \). Note that \( z(n) \) is the estimation of \( s(n) \) and can be used for decoding of the transferred symbols at each iteration step.

![Fig. 1. Track ability in static environment. Angle versus iterator number](image1)

IV. Simulations

In the simulation, the full-rate orthogonal STBC with \( N_{r} = 3, N_{t} = 4, K = T = 4 \) is chosen. The code can be found in [1], And BPSK signals are used. It is well known that the forgetting factor grows bigger, the track ability in dynamic environment becomes better while steady state performance goes worse. Here we choose a moderate value 0.900. The step size of LMS algorithm is chosen to be 0.0003, which is also a moderate value. If the value grows bigger, it can track faster but the steady state performance goes worse. The initial value of \( d(0) \) and \( h(0) \) can be arbitrarily chosen and will not influence the dynamic performance. Here, \( d(0) = 1 \) and \( h(0) = [1, 0, \ldots, 0]^T \) just for simplicity.

To evaluate the mismatch between the estimated \( \tilde{h} \) and the
varying channel both the adaptive algorithms can track the change of the channel. RLS algorithm shows better convergence rate and performance than LMS algorithm because RLS algorithm has a better track ability.

Fig 3 shows the Symbol error rates(SER) versus Signal-to-noise ratio(SNR) in static environment, SNR is defined $\frac{\sigma_H^2}{\sigma^2}$ as [3], and $\sigma_H^2$ is the variance of $H$. To compare the blind techniques, the Performance when the CSI is known to the receiver is also listed. Blind Adaptive techniques are used for decoding after receiving 20 and 50 block codes. And the total block codes is 1000 each one simulation.

Fig 4 shows the SER versus SNR in dynamic environment. The channel changes suddenly and the blind techniques can adaptively change the estimated $H$. Similarly the receiver begin to decode after receiving 20 and 50 block codes. And the total block codes is 1000 each one simulation. From the figure it is true that the performance of RLS algorithm is nearly the Performance when the CSI is known to the receiver.