Joint symbol detection and channel estimation based on variational Bayesian in MIMO relay system

Rui Wang*, Guosheng Rui, Yang Zhang, Peng Xue

Electronic Information Engineering Department, Naval Aeronautical and Astronautical University, YanTai 264001, China

Abstract

Based on the variational Bayesian inference and tensor decomposition, a joint symbol detection and channel estimation of MIMO relay system under Rayleigh fading channel is proposed. In this algorithm, the symbol matrix and the channel matrix are modeled as unknown matrices by introducing the implicit parameter variables, and recursive formula of the unknown matrix is deduced by the principle of maximizing the model evidence. Compared with the traditional estimation model based on tensor model, this algorithm introduces the a priori information of the channel into symbol estimation and channel estimation process to improve the estimation performance. The simulation results show that the proposed algorithm has high estimation accuracy compared with the existing joint estimation algorithm based on ALS (Alternating Least Square) and PLS (Partial Least Squares).

Keywords: MIMO Relay, Symbol Estimation, Tensor Decomposition, Variational Inference.

1. INTRODUCTION

MIMO relay system has aroused widespread concern for its excellent characteristics in recent years. On the one hand, MIMO technology can improve the effectiveness and reliability of communication systems; on the other hand, relay technology can improve the system network coverage (Cao et al., 2009; Dohler and Li, 2010).

In general, in the MIMO relay system, CSI (Channel State Information) is important for transmitter optimization and symbol detection (Dohler and Li, 2010). The traditional method obtains the channel matrix by supervising the training sequence (Lioliou et al., 2008; Strobach, 1997). At this time, symbol and channel estimation is supervised, and the use of the training sequence reduces spectrum utilization efficiency. Semi-blind estimation algorithm requires only the known coding matrix and relay gain matrix, which can be known in the MIMO Relay system. The algorithm in this paper is a joint semi-blind estimation of channel and symbol based on the above conditions.

Some researchers had found that tensor analysis is extraordinary efficient tool for channel estimation and symbol detection (Costa, 2014; Sidiropoulos, 2002). The supervised algorithm was proposed to estimate one way AF (Amplify and Forward) channel coefficient in (Ximenes et al., 2014; Ximenes et al., 2016). The ALS (Alternating Least Squares) based synchronous estimation of one way partial channel was proposed in (Rong et al., 2012), which needs the strict identifiability conditions. L.R. Ximenes proposed a semi-blind receiver algorithm which is based on NP (Nested PARAFAC) model. The algorithm used KRST (Khatri-Rao Space Time) coding to improve diversity and multiplex gain at source, recode and retransmit the signal at relay, building received signal a nested PARAFAC tensor model at destination. The algorithm used ALS based algorithm to the channel estimation and symbol detection. It should be mentioned that complexity of the algorithm is large. The closed form algorithm named DKRF (Double Khatri-Rao Factorization) was proposed in (Ximenes et al., 2016). Compared with ALS based algorithm, the algorithm avoids the iterations operations, making the complexity lower. However, unique condition of the algorithm is more restrict. Some no-linear algorithm was proposed in the tensor decomposition, for instance, NLS (No-linear Least Squares), PLS (Partial Least Squares) and so on. Compared with ALS based algorithm, these algorithm might not convergent globally. It can be concluded that the existing algorithms of semi-blind channel estimation are deterministic mainly. However, when the prior of channel information is known, the information can be induced in the process of symbol and channel estimation to improve the estimation performance.
In this paper, a joint symbol and channel estimation in MIMO relay system is proposed based on variational inference. Based on tensor decomposition model, implicit parameter variables are established and estimation probability model is established. On the sense of maximizing the model evidence, the variational inference theory is used to derive the recursive formula of channel matrix and symbol matrix. The proposed algorithm can introduce the channel a priori information into the process of joint symbol and channel estimation. The simulation results show that the algorithm has better estimation performance compared with the existing algorithm based on ALS and PLS.

Section 2 introduces MIMO relay communication model. Section 3 describes the probability model of joint estimation based on the variational inference structure, and detail deduced process of unknown factor matrix, latent variables. This section also contains the lower bound of algorithm, initialization and algorithm flow. Section 4 contains simulation of the algorithm.

2. SYSTEM MODEL

The source-relay channel $H^{SR} \in \mathbb{C}^{M_R \times M_S}$, the relay-destination channel $H^{RD} \in \mathbb{C}^{M_D \times M_R}$ are assumed to be flat Rayleigh channel, and constant during the transmission. The transmission is divided into two hops. During the first one, each source antenna transmits the signal after KRST coding; during the second one, the source antenna keeps silent, and the relay will re-code the received signal and retransmit it to the destination.

Assume that the source antenna and relay antenna are independent. The nodes can be regarded as a single terminal with multiple antennas, or multiple terminals with single antenna. The system of fig.1 can be suited to many applications. For instance, the scene $M_S \gg M_R$ is corresponding to cellular mobile communication, which is high centralized relay network. Large number of mobile users share several relays to transmit the signal to base stations. The scene of $J$ is corresponding to Ad hoc network, in which point to point communication needs many relay nodes.

![Figure 1. One-way and two-hop MIMO Relay system](image)

Based on the Ximenes’s works (Ximenes et al., 2014), the source model and the relay channel model are described as follows. At one phase, symbol matrix $S$ is coded by KRST coding matrix $C$. The signal after coding is transmitted through SR channel to the relay, the received signal at the relay is $W \in \mathbb{C}^{M_R \times P N_S}$. At two phase, the signal is coded by matrix $G$. The signal after coding is transmitted through RD channel to the destination, the received signal $X^{RD} \in \mathbb{C}^{M_D \times J P N_S}$ at the destination is

$$X^{RD} = H^{RD}(W \circ C \circ G)^T = H^{RD}((S \odot C)(H^{SR})^T \circ G)^T$$

(1)

In which $J$ is the length of code matrix $G$. $T$ denotes the transpose of matrix, $\circ$ is Khatri-Rao operator. In order to assure the uniqueness of tensor decomposition, it should be mentioned that $C$ and $G$ is full column matrix, and $S$ don’t contains zeros column, and channel is assumed to be flat Rayleigh channel, whose entries are subject to the complex gauss distribution.

Formula 1 can be regarded as mode-1 expansion of four order tensor $\chi^{(SRD)} \in \mathbb{C}^{M_D \times J \times N_S \times P}$, and other expansion can be seen in (Ximenes, 2014).
3. SEMI-BLIND CHANNEL ESTIMATION BASED ON VARIATIONAL INFERENCE ALGORITHM

3.1 Basic model

Assume \( y \) a four order tensor with \( M_D \times J \times N_S \times P \) dimension, which is observed tensor with Gaussian white noise. That is \( y = x + \varepsilon \), in which \( x \) denotes actual received data, \( \varepsilon \) denotes additive Gaussian white noise.

For a four order tensor \( x \), its factor matrices of Nested PARAFAC are

\[
X = [H^{(RD)}] H^{(SR)}, G, S, C
\]

In which encoding matrix \( C \) and \( G \) are known, the dimensions of matrices \( H^{(SR)}, H^{(RD)} \) and \( S \) are \( M_S, M_D \times M_R \) and \( N_S \times M_S \) respectively.

Generally, likelihood function of the observed tensor \( y \) is

\[
p(Y | [H^{(RD)}, H^{(SR)}, S], \tau) = \prod_{i=1}^{L} \prod_{n=1}^{N} \mathcal{N}(y_{i,n} | \{h_{i,n}^{(SR)}, h_{i,n}^{(RD)}; \xi_{i,n}, \xi_{i,n} \}, \tau^{-1})
\]

(3)

In which \( \tau \) is noise precision, \( \mathcal{N}(\cdot) \) denotes Gaussian distribution. Likelihood function in formula (3) indicates that \( Y_{i,n} \) are consist of column vectors of several factor matrices, which is influenced by noise precision. In order to utilize prior information of factor matrices efficiently, introduces

\[
p([H^{(RD)}, H^{(SR)}], S | \lambda) = \prod_{i=1}^{L} \mathcal{N}([h_{i,n}^{(SR)}, h_{i,n}^{(RD)}; \xi_{i,n}, \xi_{i,n} ], 0, \lambda^{-1}) \quad \forall n \in [1, N]
\]

\[
p(\lambda) = \prod_{r=1}^{R} \text{Ga}(\lambda_r | c_r, d_r)
\]

(4)

In which \( \lambda \) is factor precision of factor matrices. With value of \( \lambda \) increases, \( p(\lambda) \) approaches to zero, the validity of factor matrices is lower gradually. When \( p(\lambda) \) is the maximum, the validity of factor matrices is the greatest. In addition, \( R \) is k-rank of the tensor. Assume that hyper parameter \( \lambda \) is subject to Gamma distribution \( \text{Ga}(a, b) = b^a \lambda^{a-1} e^{-b \lambda} / \Gamma(a) \), in which \( \Gamma(a) \) is Gamma function. \( \Lambda = \text{diag}(\lambda) \) denotes inverse covariance matrix, which indicates the interaction between any two of factor matrices. The channel is assumed to be flat Rayleigh, whose entries are subject to Gaussian distribution, whose validity is controlled by the factor precision. It should be mentioned that prior information of factor matrices are specified by the considered environment.

In addition, the prior information of noise precision is also induced by

\[
p(\tau) = \text{Ga}(\tau | \alpha, \beta)
\]

(5)

Lastly, probability graph model of Nested PARAFAC tensor decomposition is illustrated in Fig.2.

For simplicity, factor matrices and hyper parameters are denoted as \( \Theta = \{ H^{(SR)}, H^{(RD)}, S, \lambda, \tau \} \). So the joint distribution probability \( p(Y, \Theta) \) is

\[
p(Y | [H^{(RD)}, H^{(SR)}, S], \tau) p([H^{(RD)}, H^{(SR)}, S] | \lambda) p(\lambda) p(\tau)
\]

(6)

The log-joint posterior distribution of \( \Theta \) can be computed by the maximum likelihood and maximum posterior, which is point estimation. However, the joint probability function is so complicated that can be practical hardly.
we adopt the variational inference methods to compute the joint posterior distribution of Θ, i.e. \( p(\Theta \mid Y) = \frac{p(Y, \Theta)}{\int p(Y, \Theta) d\Theta} \).

\[ \text{Figure 2. Probability graph model of Nested PARAFAC tensor decomposition} \]

### 3.2 Variational inference for channel estimation

As we know, when likelihood function and prior distribution are conjugate distribution; posterior distribution and prior distribution are the identical distribution. So we adopt variational inference algorithm to find the probability distribution, which has the form of factor decomposition, to approach to posterior distribution.

We try to find \( q(\Theta) \) distribution to approach actual posterior distribution \( p(\Theta \mid Y) \), making minimal KL divergence, that is

\[
\text{KL}(q(\Theta) \| p(\Theta \mid Y)) = \ln p(Y) - \mathcal{L}(q)
\]

\[
\mathcal{L}(q) = \int q(\Theta) \ln \left( \frac{p(Y, \Theta)}{q(\Theta)} \right) d\Theta
\]

In which \( \ln p(Y) \) is model evidence, which is constant. Its low bound is \( \mathcal{L}(q) \). The minimal KL divergence means the maximal value of \( \mathcal{L}(q) \). According to the mean field theory (Lathauwer, 2006), the posterior distribution can be factorized into

\[
q(\Theta) = \prod_{\alpha=1}^{N} q(A^{(\alpha)}) q(\lambda) q(\tau)
\]

Minimize the formula (7), can obtain

\[
\ln q(\Theta) = E_{\Theta \mid \Theta} [\ln p(Y, \Theta)]
\]

In which \( E_{\Theta \mid \Theta} [\cdot] \) denotes the expectation of \( q(\Theta_j) \), \( \forall j \neq k q(\Theta_k) \) is optimized by iteration until convergence. We can deduce the estimated parameters by adopting the algorithm above.

### 3.2.1 Posterior distribution of factor matrices

To assure the coherence of factor matrix deduction, we assume \( A = \{A^{(n)}\}_{n=1}^{3} = \{H^{(SR)}, H^{(SR)}, S\} \). According to fig.2, inference of factor matrix \( A^{(n)} \) can be deduced from formula (3), which is combined with prior information in formula (6). The process is listed in appendix A. The estimated posterior parameter of factor matrix \( A^{(n)} \) can be updated by
\[ \tilde{A}^{(n)} = E_q[\tau](Y_{in})E_q[A^{(n)}]V^{(n)} \]

\[ V^{(n)} = (E_q[\tau]E_q[A^{(n)T}A^{(n)}] + E_q[A])^{-1} \]  

(9)

In which \( Y_{in} \) denotes mode-\( n \) matrix of tensor \( Y \), \( V^{(n)} \) is a auxiliary matrix. \( E_q[\cdot] \) denotes posterior expectation. Actually, the complexity of \( E_q[A^{(n)T}A^{(n)}] \) is high, so lemma 1 is induced as follows.

**Lemma 1**: Given a group of matrices \( \{A^{(n)}\}_{n=1}^{N} \), if row vectors \( \{d_{in}^{(n)}\}_{i=1}^{d} \) are independent, and \( \text{cov}[d_{in}^{(n)}] = V^{(n)} \) \( \forall \ n \in [1, I_n] \), we have

\[ E[\sum_{n} A^{(n)T}A^{(n)}] = \sigma(E[A^{(n)T}A^{(n)}]) \]

In which \( E[A^{(n)T}A^{(n)}] = \{E[A^{(n)T}]E[A^{(n)}] + I_n V^{(n)} \} \). Detail proof is listed in appendix B, we have

\[ E[A^{(n)T}A^{(n)}] = \sum_{i=n} \{\tilde{A}^{(n)T}A^{(n)} + I_n V^{(n)} \} \]  

(10)

\( V^{(n)} \) is updated by the prior \( E_q[A] \) and other factor matrices, and the weights of two parts are tuned by \( E_q[\tau] \) automatically. So \( E_q[\tau] \) is in relation with fitness of model. Superior fitness should acquire more information from current model than prior information, and greater value of observed data means that more similarity between model and factor matrices. Then, the estimated \( \tilde{A}^{(n)} \) is rotated by \( V^{(n)} \), and scaled by \( E_q[\tau] \).

### 3.2.2 Posterior distribution of hyper parameter \( \lambda \)

Factor matrices and prior information can be used in the inference of hyper parameter \( \lambda \). Assume posterior of \( \lambda_r, \forall r \in [1, R] \) is subject to Gamma distribution, that is, \( q(\lambda) = \prod_{r=1}^{R} Ga(\lambda_r|c_r, d_r) \). In which \( R \) is k-rank of tensor, \( c_r, d_r \) denote learned posterior parameter for observed data, which are updated by formula below

\[ c_r = c_0 + \frac{1}{2n} \sum_{i=1}^{n} I_n, \quad d_r = d_0 + \frac{1}{2n} \sum_{i=1}^{n} E_q[A^{(n)T}A^{(n)}] \]  

(11)

Posterior expectation items in formula (11) can be computed by posterior parameter in formula (9), we have

\[ E_q[A^{(n)T}A^{(n)}] = \tilde{A}^{(n)T}A^{(n)} + \sum_r (V_r^{(n)})_r \]

So the simplicity of \( d_M = [d_M, \cdots, d_M]^T \) is

\[ d_M = \sum_{r=1}^{R} (\text{diag}(\tilde{A}^{(n)T}A^{(n)})) + I_n V^{(n)} \]  

(12)

### 3.2.3 Posterior distribution of hyper parameter \( \tau \)

The updated formula of noise precious \( \tau \) can be deduced by the prior information and observed data whose posterior is Gamma distribution. that is \( q(\tau) = Ga(\tau|a_M, b_M) \). The posterior parameter is updated by

\[ a_M = a_0 + \frac{M}{2} \]

\[ b_M = b_0 + \frac{1}{2} E_q[V - [H^{(RD)}H^{(SR)}, S, G, C] \]  

(13)
It should be mentioned that posterior expectation of model residual can’t be computed directly, so the lemma 2 is induced as follows.

Lemma 2: Given a group of independent and random matrices\(\{A^{(n)}\}_{n=1}^{N}\), assume that\(\forall n, \forall \eta_n\), row vectors\(\{a_{n,i}\}\) are random, so we have (Rao, 1997; Kolda and Bader, 2009).

\[
E\left[ \left\| A^{(1)} \cdots A^{(N)} \right\| \right] = \sum_{\eta_n} E\left[ a_{n,i} a_{n,i}^{*} \right] \cdots E\left[ a_{n,N} a_{n,N}^{*} \right]
\]

Using lemma 2, posterior expectation of model residual in formula (13) is

\[
E_{q}\left[ \left\| V - \sum_{i=1}^{M} h_{i} h_{i}^{*} \right\| \right] = 2\text{vec}^{T}(U)\text{vec}\left( \sum_{i=1}^{M} h_{i} h_{i}^{*} \right) + \sum_{\eta_n} \left\{ E\left[ h_{i} h_{i}^{*} \right], \cdots , E\left[ h_{i} h_{i}^{*} \right] \right\}
\]

Thus, posterior expectation of is updated by \(E_{q}(\tau) = \frac{a_{M}^{*}}{b_{M}^{*}}\), in which \(a_{M}^{*}\) is in relation with the number of observed data, \(a_{M}^{*}\) is with F-norm of model residual.

3.2.4 Lower bound of model evidence

We can compute the variational lower bound in formula (7). With the estimation progress, the value increases, we can monitor the value to judge the convergence or not. Log- marginal likelihood lower bound is (Lathauwer, 2006).

\[
\mathcal{L}(q) = E_{\psi(\lambda)}[\ln p(Y|\Theta)] + H(q(\Theta))
\]

The first part denotes the posterior expectation of joint probability density, the second part denotes the entropy of \(q\) distribution. According to formula (5)-(8), formula (15) can be obtained

\[
\mathcal{L}(q) = E_{\psi(\lambda)}[\ln p(Y|A^{(\lambda)}, \tau^{-1})] + E_{\psi(\lambda)}\left[ \sum_{n=1}^{N} \ln p(A^{(n)}|\lambda) \right]
\]

\[
+ E_{\psi(\lambda)}[\ln p(\lambda)] + E_{\psi(\lambda)}[\ln p(\tau)] - E_{\psi(\lambda)}\left[ \sum_{n=1}^{N} \ln q(A^{(n)}) \right] - E_{\psi(\lambda)}[\ln q(\lambda)] - E_{\psi(\lambda)}[\ln q(\tau)]
\]

Substituting formula (9), (11) and (13) in the formula (16), the precious solution can be obtained. The detail expansion is no more expressed here.

3.2.5 Initiation of model inference

In order to avoid the tracking stagnation problem, it is of importance to choose initiative values. \(a_{0}^{(n)}, b_{0}^{(n)}, c_{0}, d_{0}\) are setting to \(10^{-4}\), thus making the no informational prior. then \(E[\Lambda] = 1, E[\tau] = 1\). For factor matrices \(E[A^{(n)}]\), two different initiative schemes can be adopted. firstly, each row vector \(\{a_{n,i}\}\) is chosen from \(N(0,1)\); secondly, set \(A^{(n)} = U(0)\Sigma^{(n)} \), in which \(U(0)\) denotes left singular vector, \(\Sigma^{(n)}\) denotes diagonal matrix of singular value. They can be obtained from SVD decomposition of mode-n matrix of observed data. Setting \(V^{(n)}\) to be \(E[\Lambda^{-1}]\), rank \(R\) is the number of transmit antennas.

The entire flow of inference algorithm is listed in table 1.
4. SIMULATION

The complexity, convergence speed, and estimate performance are evaluated in the section. Assume channel matrix \( H^{(SD)} \), \( H^{(RD)} \) and \( H^{(SR)} \) are independent identical distribution, whose entries are subject to Gaussian distribution, whose means are zero, variances are \( 1/M_s,1/M_R,10^{-\alpha/10}/M_s \) respectively, in which \( \alpha \) is a fixed parameter, which denotes the attenuation energy ratio between direct and relay link. Noise is assumed to be additive Gaussian white noise, whose mean is zero, and variance is one. The mean SNR between direct link and relay link is proportional to the symbol energy.

Symbol matrix \( S = \sqrt{E_S}S_0 \) is generated from 8PSK alphabet table, in which \( S_0 \) is symbol matrix with unit energy, \( E_S \) is mean energy of each symbol. First row of \( S \) and \( H^{(RD)} \) is known, which eliminates the scale ambiguous. The coding matrix is truncated DFT matrix, which guarantees that matrix \( C \) is full column rank. Gain matrix \( G \) is a random generated vandermonde matrix, which avoids the permute ambiguous of \( H^{(RD)} \) and \( H^{(SR)} \).

4.1 Complexity

Compare the complexity in each iteration with other two algorithms. Complexity of factor matrices in formula (13) is \( O(R^2M_D|PN\sum_n I_n + M^2_s|\sum_n I_n) \), and complexity of \( \lambda \) is \( O(M^2_s M_D|PN) \). So the total complexity of the algorithm is \( O(M^3_s + M^2_s M_D|PN)(M_D + P + N) \). It can be seen that the complexity is linearly with dimension of observed data, and is polynomial relation with number of transmit antennas. The complexities of NP-PLS and NP-ALS are listed in table 2. It can be concluded that the complexity of proposed algorithm in this paper is slight lower than NP-ALS apparently according to table 2.

4.2 Convergence speed

The subsection will validate convergence of the algorithm, and compare it with NP-ALS and NP-PLS. The initiations of factor matrices \{ \( H^{(SR)} \), \( H^{(RD)} \), \( S \) \} adopt two schemes: (1) chosen from Gaussian distribution randomly. (2) assign them values from SVD decomposition of observed data. The evaluation index \( RE(\text{Reconstruction Error}) \) is brought to make it link with iteration number. RE is defined as

\[
\varepsilon_k = \left\| S^{(SNR)}_k - (S^*_k \otimes G^T)^T (C \otimes \hat{H}_k^{(RD)})_k^{(SR)} \right\|.
\]  

(17)
The simulation results show in fig. 3. It can be concluded that: (1) The convergence speed of NP-PLS is faster than NP-ALS. The reason is that no-linear search method is adopted to avoid the search locally. (2) The convergence speed of algorithm in this paper is faster than NP-ALS, and approaches NP-PLS. It should be mentioned that, although the convergence speed of NP-PLS is faster, the mean RE is greater than other two algorithms. In the view of another aspect, the estimation performance of NP-PLS is worse than the algorithm in this paper.

![Figure 3. Convergence speed comparison with NP-ALS and NP-PLS](image)

4.3 Estimation performance

Compare estimation performance of the algorithm with NP-ALS and NP-PLS. The three algorithms are used to complete the symbol and channel estimation under the CSI known environment, and evaluated the estimated results finally. The evaluation indexes are SER(Symbol Error Rate) and NMSE(Normalized Mean Square Error). SER is mean of multiple Monte Carlo simulation, and NMSE is defined as

$$\text{NMSE} = \frac{1}{K} \sum_{k=1}^{K} \left( \frac{\|\mathbf{H}_k - \hat{\mathbf{H}}_k\|}{\|\mathbf{H}_k\|} \right)$$  \hspace{1cm} (18)

In which $K$ is the run number of Monte Carlo, $K=10000$. Data streams in each run are $NK = 10000N$. $H_k$ is simulated channel of the $k$th run, $\hat{H}_k$ is estimated channel by the algorithms. The simulation parameters are listed in table 3.

<table>
<thead>
<tr>
<th>$M_d$ = $M_R$ = $M_s$ = 4, $N$ = $P$ = $J$ = 4, convergence error (ALS, PLS): $10^{-3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_0, b_0, c_0, d_0 ; 10^{-6}$</td>
</tr>
</tbody>
</table>

We choose two environments of different noise variance ($\sigma^2 = \sigma_{R}^2 = \sigma_{D}^2$). SER and channel NMSE can be shown in Fig.4 and Fig.5. It can be concluded that: (1) In the aspect of NMSE, estimation performance of the algorithm in this paper is superior to the NP-ALS and NP-PLS at the equal symbol energy. The performance of NP-PLS is worst for the local convergence caused by the no-linear search. When noise variance is 10, the performance of NP-PLS worsen drastically. The algorithm in this paper is optimal among the three algorithms. (2) In the aspect of factor matrices, the estimated symbol matrix is the same as the channel matrix, so NP-VB is superior to other two algorithms. Another explanation is that the symbol matrix is assumed to be Gaussian distribution, whose prior information is also used in the process of estimation. It can be seen that NP-VB is also
superior to other two algorithms. The algorithm in this paper is more robust especially under the higher noise variance.

![Figure 4. Channel NMSE comparison with NP-ALS and NP-PLS](image1)

![Figure 5. SER comparison with NP-ALS and NP-PLS](image2)

It should be mentioned that the channel is assumed to be flat Rayleigh. In practical application, the probability model can be revised by the character of channel considered. If no prior information exists or making the estimation objectively, no-information prior might be chosen, such as Jeffrey prior. In addition, the algorithm is based on the Nested-PARAFAC specially, but the deduction is generally. So the algorithms can be extended to channel estimation based on other tensor decomposition, such as Nested Tucker model (Favier et al., 2016).

5. CONCLUSION

A novel channel estimation based on the variational inference is proposed in the paper. In this algorithm, some latent hyper-parameters such as operational accuracy, noise precious was induced into algorithm, and buildup channel estimation probability model based on nested PARAFAC tensor model. The iteration formula of factor matrix, operational accuracy and noise precious were deduced by the idea of variational inference, making the q distribution, which has the factor decomposition form, approaching the unknown parameter post-distribution.
algorithm can utilize the prior information of channel to improve channel estimation performance. The parameter can be tuned automatically, the complexity is linear with the dimension of data. Compared with ALS based algorithm and closed-form algorithm. With generality of deduction, the algorithm can be extended to channel estimation based on other tensor model.

Appendix A : The derivation of formula (9)

\[
\ln(q(A^{(t)})) = \mathbb{E}_{q(\theta^{(t)}, A^{(t)})} [\ln p(Y|A^{(t)}, \lambda, \gamma, \tau)] + \text{const}
\]

\[
= \mathbb{E} \left[ \sum_{t=1}^{T} \left( \frac{1}{2} \ln \tau - \frac{1}{2} \{y_{A, \gamma, \lambda} - \langle A^{(1)}, A^{(2)}, \ldots, A^{(N)} \rangle \}^2 \right) \right] + \text{const}
\]

\[
+ \sum_{t=1}^{T} \left( \frac{1}{2} \ln |A| - \frac{1}{2} (A^{(t)T} \Lambda A^{(t)}) \right) + \text{const}
\]

\[
= \mathbb{E} \left[ \frac{1}{2} \tau \sum_{t=1}^{T} (A^{(1)}, A^{(2)}, \ldots, A^{(N)})^2 - \frac{1}{2} (A^{(t)T} \Lambda A^{(t)}) \right]
\]

\[
+ \tau \sum_{t=1}^{T} (y_{A, \gamma, \lambda} \{ A^{(1)}, A^{(2)}, \ldots, A^{(N)} \}) + \text{const}
\]

\[
= \mathbb{E} \left[ \frac{1}{2} \tau \sum_{t=1}^{T} A^{(t)T} (A^{(1)})^T A^{(t)} - \frac{1}{2} (A^{(t)T} \Lambda A^{(t)}) \right]
\]

\[
+ \tau \sum_{t=1}^{T} (y_{A, \gamma, \lambda} A^{(t)T} (A^{(1)})) + \text{const}
\]

\[
= \frac{1}{2} \left( A^{(t)T} \mathbb{E}[\tau] \mathbb{E} \left[ \left( \sum_{t=1}^{T} A^{(1)} \right)^T \right] \mathbb{E}[A] A^{(t)} + \text{const} \right)
\]

\[
+ A^{(t)T} \mathbb{E}[\tau] \mathbb{E} \left[ \left( \sum_{t=1}^{T} A^{(1)} \right)^T \right] \text{vec}(y_{A, \gamma, \lambda}) + \text{const}
\]

Appendix B: proof of Lemma 1

\[
\mathbb{E} \left[ \left( \sum_{t} A^{(t)} \right)^T \left( \sum_{t} A^{(t)} \right) \right]
\]

\[
= \mathbb{E} \left[ \sum_{t=1}^{T} (A^{(t)})^T (A^{(t)}) \right]
\]

\[
= \mathbb{E} \left[ \sum_{t=1}^{T} A^{(t)T} A^{(t)} \right]
\]

\[
= \sum_{t=1}^{T} \mathbb{E} \left[ A^{(t)T} A^{(t)} \right]
\]

ACKNOWLEDGMENTS

This work was supported by the Natural Science Foundation of China under Grant No.41606117 and National High Technology Research and Development Program of China under grant No.2015AA7015087.
REFERENCES


Favier G., Fernandes C., Almeida A. (2016). Nested Tucker tensor decomposition with application to MIMO relay systems using tensor space time coding(TSTC), Signal Processing, 12(8), 318-331.


