Article

A Robust Semi-Blind Receiver for Joint Symbol and Channel Parameter Estimation in Multiple-Antenna Systems

Jianhe Du [1],* ![ORCID]**, Meng Han [1], Yan Hua [1], Yuanzhi Chen [1] and Heyun Lin [2]**

[1]   School of Information and Communication Engineering, Communication University of China,
     Beijing 100024, China; hm593@cuc.edu.cn (M.H.); huayan@cuc.edu.cn (Y.H.); chenyuanzhi@cuc.edu.cn (Y.C.)
[2]   Power Dispatching Control Center, Guangxi Power Grid, Nanning 530023, China; forlan@bupt.edu.cn
*   Correspondence: dujianhe1@gmail.com

check for
updates

Abstract: For multiple-antenna systems, the technologies of joint symbol and channel parameter estimation have been developed in recent works. However, existing technologies have a number of problems, such as performance degradation and the large cost of prior information. In this paper, a tensor space-time coding scheme in multiple-antenna systems was considered. This scheme allowed spreading, multiplexing, and allocating information symbols associated with multiple transmitted data streams. We showed that the received signal was formulated as a third-order tensor satisfying a Tucker-2 model, and then a robust semi-blind receiver was developed based on the optimized Levenberg–Marquardt (LM) algorithm. Under the assumption that the instantaneous channel state information (CSI) is unknown at the receiving end, the proposed semi-blind receiver jointly estimates the information symbol and channel parameters efficiently. The proposed receiver had a better estimation performance compared with existing semi-blind receivers, and still performed well when the channel became strongly correlated. Moreover, the proposed semi-blind receiver could be extended to the multi-user massive multiple-input multiple-output (MIMO) system for joint symbol and channel estimation. Computer simulation results were shown to demonstrate the effectiveness of the proposed receiver.

Keywords: multiple-antenna systems; third-order tensor; Tucker-2 model; semi-blind receiver; optimized LM algorithm

1. Introduction

Multiple-antenna techniques are well known to provide spatial diversity and multiplexing gains [1–3]. Over the last few decades, the benefits of multiple-antenna communications have been verified in both theory and practice. On the other hand, tensor-based signaling approaches that utilize several signal dimensions such as time, space, and code, are seen as good technologies for improving the information transmission rate and enhancing communication reliability [4–6]. Against this background, the problem of joint symbol and channel estimation is resolved by using tensor-based signaling approaches, and a number of semi-blind or blind receivers have been proposed for multiple-input multiple-output (MIMO) systems.

A parallel factor (PARAFAC) [7] based receiver is proposed in [8] by using the Khatri–Rao space-time (KRST) coding scheme, which can achieve a flexible tradeoff between error performance and transmission efficiency. In [9], the authors extend the KRST coding scheme by using the linear constellation precoding, and then developing several semi-blind receivers. These semi-blind receivers allow a joint symbol and channel estimation without requiring pilot sequences for the

instantaneous channel state information (CSI) acquisition. In [10], the authors develop a new tensor-based receiver in MIMO relay systems for channel estimation by using PARAFAC analysis. A low complexity PARAFAC-based channel estimation scheme for non-regenerative MIMO relay systems is developed in [11]. In [12], a novel semi-blind receiver is derived using a multiple KRST coding scheme for joint symbol and channel estimation. More recently, a nested PARAFAC-based receiver for cooperative MIMO communications is proposed in [13], and three-step and double two-step alternating least squares (ALS) algorithms are proposed to fit the nested PARAFAC model for estimating system parameters. For millimeter wave (mmWave) massive MIMO systems, a PARAFAC decomposition-based algorithm is developed in [14] to jointly estimate channel parameters of multiple users. In [15], the algorithm in [14] is extended to mmWave MIMO orthogonal frequency division multiplexing (MIMO-OFDM) systems for channel estimation, and Cramér–Rao bound (CRB) results for channel parameters are also derived. Considering the channel estimation issue in the presence of pilot contamination for multi-cell massive MIMO systems, a new PARAFAC-based approach is proposed in [16] to jointly estimate directions of arrival, fading coefficients, and delays. Although these works [8–16] consider different design approaches, their common feature is using the PARAFAC model, which needs to know the first column or row of one loading matrix to eliminate scaling ambiguity. Furthermore, the ALS algorithm used in these receivers exhibits a convergence problem when ill-conditioned factor matrices exist [17].

In contrast to the ALS algorithm, the Levenberg–Marquardt (LM) algorithm updates all the parameters to be estimated at the same time. The LM algorithm is successfully used to fit some tensor models, adapt to collinearity problems, and provide quadratic convergence [18–20]. A LM algorithm is first proposed for fitting PARAFAC model in [18]. In [19], the authors present a LM algorithm to the decomposition of the Block Component Model (BCM) in the uplink of a wideband direct-sequence code-division multiple access (DS-CDMA) systems. Recently, a LM algorithm was developed in [20] to jointly estimate information symbol and channel matrices for a generalized PARATUCK2 model. As an iterative algorithm, the LM algorithm is also sensitive to initialization. Thus, the optimization of the initial value is important to improve the performance of the LM algorithm.

In [21], a tensor-based space-coding scheme using PARATUCK2 model is developed. For the PARATUCK2 model, the number of channel uses can be different from one transmitted data stream to another. In [22], a generalized PARATUCK2 model is proposed by exploiting a tensor space-time (TST) coding. Recently, a Kronecker product least squares (KPLS) receiver is proposed in [23] to estimate the symbol and channel matrices. More recently, it is shown in [24] that a KPLS receiver can be extended to all the tensor-based systems. Although the KPLS receiver is a non-iterative and low-complexity solution, it needs the related core tensor unfolding to be right-invertible, which is a relatively harsh condition in signal design.

Inspired by [21] and [22], we considered a simple tensor space-time coding scheme for multiple-antenna systems, along with an efficient receiver. The allocation factor and the space-time code factor in the TST coding scheme in [22] are independent, while the allocation factor in our coding scheme is also a three-dimensional space-time code factor. Thanks to the special structure of the proposed coding scheme, the received signal can be constructed as a Tucker-2 model [25,26], which has uniqueness property under some suitable conditions. Then, a robust semi-blind receiver based on optimized LM algorithm is presented for joint channel and symbol estimation. Uniqueness and identifiability issues for the constructed Tucker-2 model are also discussed in this paper. Compared with existing receivers, the proposed receiver has a better estimation performance. Moreover, the proposed semi-blind receiver can be extended to the multi-user massive MIMO system. For the low-rank channel,the proposed receiver still has good performance for joint symbol and channel estimation even in the shorter length of code and information symbol, and larger number of data streams.

The organization of this paper is as follows. Section 2 presents a brief overview of the Tucker model. In Section 3, the system model is presented and the associated tensor signal model is formulated.

Section 4 briefly reviews the receiver with the ALS algorithm and describes the proposed semi-blind receiver based on the optimized LM algorithm. Section 5 extends the proposed semi-blind receiver to multi-user massive MIMO systems for joint symbol and channel estimation. In Section 6, some simulation results are shown to demonstrate the performance of our semi-blind receiver. Conclusions are drawn in Section 7.

Notation: Scalars, vectors, matrices, and tensors are denoted by lower-case letters $(a, b, \cdots)$, boldface lower-case letters $(\mathbf{a}, \mathbf{b}, \cdots)$, boldface capitals $(\mathbf{A}, \mathbf{B}, \cdots)$, and underlined boldface capitals $(\underline{\mathbf{A}}, \underline{\mathbf{B}}, \cdots)$, respectively. $\mathbf{A}^T$, $\mathbf{A}^H$, $\mathbf{A}^{-1}$, and $\mathbf{A}^\dagger$ represent transpose, conjugate transpose, inverse, and Moore–Penrose pseudo-inverse of the matrix $\mathbf{A}$, respectively. $\|\mathbf{A}\|_F$ denotes the Frobenius norm of $\mathbf{A}$. $\mathbf{I}_M$ denotes the $M \times M$ identity matrix. The operator $vec\,(\cdot)$ stacks the columns of its matrix argument to a vector, while $unvec\,(\cdot)$ represents the inverse vectorization operation. The Kronecker matrix product is denoted by $\otimes$. The term $D_i\,(\mathbf{A})$ corresponds the diagonal matrix out of the i-th row of $\mathbf{A}$.

2. Tucker Model

This section first presents a brief overview of the Tucker model, and then focuses on the Tucker-2 model used in this work. For an Nth-order tensor $\underline{\mathbf{T}} \in \mathbb{C}^{I_1 \times \cdots \times I_N}$, a Tucker-$N$ model or Tucker model is defined in the following scalar form as [26]:

$$
\begin{aligned}
t_{i_1,\dots,i_N} &= \sum_{r_1=1}^{R_1} \cdots \sum_{r_N=1}^{R_N} g_{r_1,\dots,r_N} \left(a_{i_1,r_1}^{(1)} \times \cdots \times a_{i_N,r_N}^{(N)} \right) \\
&= \sum_{r_1=1}^{R_1} \cdots \sum_{r_N=1}^{R_N} g_{r_1,\dots,r_N} \prod_{n=1}^{N} a_{i_n,r_n}^{(n)}
\end{aligned}
\tag{1}
$$

where $i_n = 1, \dots, I_n$ for $n = 1, \dots, N$, $a_{i_n,r_n}^{(n)}$ and $g_{r_1,\dots,r_N}$ stand for typical elements of the matrix factor $\mathbf{A}^{(n)} \in \mathbb{C}^{I_n \times R_n}$ and the core tensor $\underline{\mathbf{G}} \in \mathbb{C}^{R_1 \times \cdots \times R_N}$, respectively. Using the mode-n product representation, the model (1) can be written as:

$$
\begin{aligned}
\underline{\mathbf{T}} &= \underline{\mathbf{G}} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \times_3 \cdots \times_N \mathbf{A}^{(N)} \\
&= \underline{\mathbf{G}} \times_{n=1}^{N} \mathbf{A}^{(n)}
\end{aligned}
\tag{2}
$$

where $\underline{\mathbf{G}} \times_n \mathbf{A}^{(n)}$ denotes the mode-n product of $\underline{\mathbf{G}}$ and $\mathbf{A}^{(n)}$ along the Nth mode, gives a tensor $\underline{\mathbf{W}}$ of dimensions $R_1 \times \cdots \times R_{n-1} \times I_n \times R_{n+1} \times \cdots \times R_N$ such that:

$$
w_{r_1,\dots,r_{n-1},i_n,r_{n+1},\dots,r_N} = \sum_{r_n=1}^{R_n} a_{i_n,r_n}^{(n)} g_{r_1,\dots,r_{n-1},r_n,r_{n+1},\dots,r_N}
\tag{3}
$$

where $w_{r_1,\dots,r_{n-1},i_n,r_{n+1},\dots,r_N}$ is a typical element of the tensor $\underline{\mathbf{W}}$. It has been known that the Tucker model is not essentially unique [26], which restricts its application. Their matrix factors can be only determined up to nonsingular transformations characterized by nonsingular matrices. However, some low-order Tucker models with special structures are unique up to permutation and/or scaling ambiguities.

Assuming $N = 3$ and $\mathbf{A}^{(3)} = \mathbf{I}_{I_3}$ for the third-order tensor $\underline{\mathbf{T}} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$, we have:

$$
\begin{aligned}
t_{i_1,i_2,i_3} &= \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} \sum_{r_3=1}^{R_3} g_{r_1,r_2,r_3} \prod_{n=1}^{3} a_{i_n,r_n}^{(n)} \\
&= \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} g_{r_1,r_2,i_3} a_{i_1,r_1}^{(1)} a_{i_2,r_2}^{(2)}.
\end{aligned}
\tag{4}
$$

This model is called Tucker-2 model or Tucker-(2, 3) model, and is widely applied in data analysis and parameter estimation [4]. $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$ are the two loading matrices, and $\underline{\mathbf{G}}$ is the core tensor. In the same way, such a model can be written in terms of mode-n product as:

$$\begin{aligned}\underline{\mathbf{T}} &= \underline{\mathbf{G}} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \times_3 \mathbf{I}_{I_3} \\ &= \underline{\mathbf{G}} \times_{n=1}^{2} \mathbf{A}^{(n)}.\end{aligned} \tag{5}$$

3. System Model

Consider a multiple-antenna system with M_S transmit antennas and M_D receive antennas as shown in Figure 1. h_{m_D,m_S} represents the channel coefficient between the m_S-th transmit antenna and the m_D-th receive antenna ($m_S = 1, \ldots, M_S, m_D = 1, \ldots, M_D$). $s_{n,r}$ represents the n-th symbol of the r-th data stream ($n = 1, \ldots, N, r = 1, \ldots, R$), with each data stream being formed of N information symbols. Each symbol $s_{n,r}$ is coded by a three-dimensional space-time code $b_{m_S,r,p}$ ($p = 1, \ldots, P$), whose dimensions are the numbers of transmit antennas, data streams, and chips, respectively. We then define the antenna-to-slot allocation factor q_{p,m_S}, which is 0 or 1. Both the transmitter and the receiver know these factors $b_{m_S,r,p}$ and q_{p,m_S}.
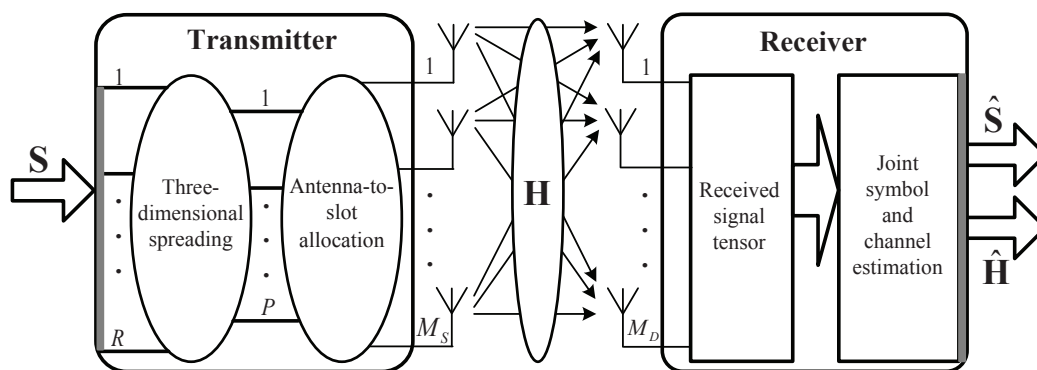

Figure 1. Block-diagram of the system model.

The signal transmitted from m_S-th transmit antenna, during the n-th symbol period of the p-th chip, is given by:

$$x_{m_S,n,p} = \sum_{r=1}^{R} q_{p,m_S} b_{m_S,r,p} s_{n,r} \tag{6}$$

where $s_{n,r}$ and q_{p,m_S} are (n, r)-th and (p, m_S)-th elements of signal matrix $\mathbf{S} \in \mathbb{C}^{N \times R}$ and the antenna-to-slot allocation matrix $\mathbf{Q} \in \mathbb{C}^{P \times M_S}$, respectively. $x_{m_S,n,p}$ and $b_{m_S,r,p}$ are typical elements of the transmitted signal tensor $\underline{\mathbf{X}} \in \mathbb{C}^{M_S \times N \times P}$ and the coding tensor $\underline{\mathbf{B}} \in \mathbb{C}^{M_S \times R \times P}$, respectively. The elements in $\underline{\mathbf{B}}$ are chosen as $e^{\sqrt{-1}\varsigma/2\pi}$, where ς is taken from random uniformly distributed pseudorandom numbers. In our tensor coding scheme, the number of transmitted data streams is not restricted to be equal to that of transmit antennas, and the data streams can be allocated to an arbitrary set of transmitted antennas. Without considering the allocation of stream-to-slot, the coding scheme in [21] can be regarded as a special case of our tensor coding scheme with a fixed two-dimensional space-time code.

Assuming Rayleigh flat fading channels, then the discrete-time baseband signal at the m_D-th receive antenna can be written as:

$$
\begin{aligned}
y_{m_D,n,p} &= \sum_{m_S=1}^{M_S} h_{m_D,m_S} x_{m_S,n,p} + v_{m_D,n,p} \\
&= \sum_{m_S=1}^{M_S} \sum_{r=1}^{R} h_{m_D,m_S} q_{p,m_S} b_{m_S,r,p} s_{n,r} + v_{m_D,n,p}
\end{aligned}
\tag{7}
$$

where h_{m_D,m_S} is the (m_D, m_S)-th element of channel matrix $\mathbf{H} \in \mathbb{C}^{M_D \times M_S}$, $y_{m_D,n,p}$ and $v_{m_D,n,p}$ are typical elements of the received signal tensor $\underline{\mathbf{Y}} \in \mathbb{C}^{M_D \times N \times P}$ and the noise tensor $\underline{\mathbf{V}} \in \mathbb{C}^{M_D \times N \times P}$, respectively.

3.1. Constructed Tucker-2 Model

Let us define $c_{m_S,r,p} = q_{p,m_S} b_{m_S,r,p}$, where $c_{m_S,r,p}$ is the typical element of the compound tensor $\underline{\mathbf{C}} \in \mathbb{C}^{M_S \times R \times P}$. So Equation (7) can be written as:

$$
y_{m_D,n,p} = \sum_{m_S=1}^{M_S} \sum_{r=1}^{R} h_{m_D,m_S} c_{m_S,r,p} s_{n,r} + v_{m_D,n,p}.
\tag{8}
$$

By comparing Equation (4) with Equation (8), the received signal tensor $\underline{\mathbf{Y}} \in \mathbb{C}^{M_D \times N \times P}$ of noiseless signals satisfies a Tucker-2 model , with the following correspondences:

$$
\left(\underline{\mathbf{G}}, \mathbf{A}^{(1)}, \mathbf{A}^{(2)} \right) \Leftrightarrow (\underline{\mathbf{C}}, \mathbf{H}, \mathbf{S})
\tag{9}
$$

$$
(I_1, I_2, R_1, R_2, I_3) \Leftrightarrow (M_D, N, M_S, R, P).
\tag{10}
$$

Using the mode-n product representation, the model (8) can be written as:

$$
\underline{\mathbf{Y}} = \underline{\mathbf{C}} \times_1 \mathbf{H} \times_2 \mathbf{S} + \underline{\mathbf{V}}
\tag{11}
$$

where $\mathbf{S}$ and $\mathbf{H}$ represent the two loading matrices, and $\underline{\mathbf{C}}$ is the core tensor.

Let us define $\mathbf{Y}_{..p} \in \mathbb{C}^{M_D \times N}$, $\mathbf{B}_{..p} \in \mathbb{C}^{M_S \times R}$, $\mathbf{C}_{..p} \in \mathbb{C}^{M_S \times R}$ and $\mathbf{V}_{..p} \in \mathbb{C}^{M_D \times N}$ as the p-th matrix slice of $\underline{\mathbf{Y}} \in \mathbb{C}^{M_D \times N \times P}$, $\underline{\mathbf{B}} \in \mathbb{C}^{M_S \times R \times P}$, $\underline{\mathbf{C}} \in \mathbb{C}^{M_S \times R \times P}$, and $\underline{\mathbf{V}} \in \mathbb{C}^{M_D \times N \times P}$, respectively. We have $\mathbf{C}_{..p} = D_p(\mathbf{Q}) \mathbf{B}_{..p}$. By defining $\mathbf{Y}_1 = \left[\mathbf{Y}_{..1}^T, \ldots, \mathbf{Y}_{..P}^T \right]^T \in \mathbb{C}^{PM_D \times N}$, $\mathbf{Y}_2 = \left[\mathbf{Y}_{..1}, \ldots, \mathbf{Y}_{..P} \right]^T \in \mathbb{C}^{PN \times M_D}$, $\mathbf{Y}_3 = \left[vec(\mathbf{Y}_{..1}), \ldots, vec(\mathbf{Y}_{..P}) \right] \in \mathbb{C}^{M_D N \times P}$, and $\mathbf{Y}_4 = \left[vec(\mathbf{Y}_{..1}^T), \ldots, vec(\mathbf{Y}_{..P}^T) \right] \in \mathbb{C}^{N M_D \times P}$, we can obtain four compact forms of the Tucker-2 model (11):

$$
\mathbf{Y}_1 = (\mathbf{I}_P \otimes \mathbf{H}) \mathbf{F}_1 \mathbf{S}^T + \mathbf{V}_1
\tag{12}
$$

$$
\mathbf{Y}_2 = (\mathbf{I}_P \otimes \mathbf{S}) \mathbf{F}_2 \mathbf{H}^T + \mathbf{V}_2
\tag{13}
$$

$$
\mathbf{Y}_3 = (\mathbf{S} \otimes \mathbf{H}) \mathbf{F}_3 + \mathbf{V}_3
\tag{14}
$$

$$
\mathbf{Y}_4 = (\mathbf{H} \otimes \mathbf{S}) \mathbf{F}_4 + \mathbf{V}_4
\tag{15}
$$

with,

$$\mathbf{F}_1 = \left[\mathbf{B}^T_{..1}D_1\left(\mathbf{Q}\right),\ldots,\mathbf{B}^T_{..P}D_P\left(\mathbf{Q}\right)\right]^T$$
$$\mathbf{F}_2 = \left[D_1\left(\mathbf{Q}\right)\mathbf{B}_{..1},\ldots,D_P\left(\mathbf{Q}\right)\mathbf{B}_{..P}\right]^T$$
$$\mathbf{F}_3 = \left[vec\left(D_1\left(\mathbf{Q}\right)\mathbf{B}_{..1}\right),\ldots,vec\left(D_P\left(\mathbf{Q}\right)\mathbf{B}_{..P}\right)\right]$$
$$\mathbf{F}_4 = \left[vec\left(\mathbf{B}^T_{..1}D_1\left(\mathbf{Q}\right)\right),\ldots,vec\left(\mathbf{B}^T_{..P}D_P\left(\mathbf{Q}\right)\right)\right]$$

(16)

and,

$$\mathbf{V}_1 = \left[\mathbf{V}^T_{..1},\ldots,\mathbf{V}^T_{..P}\right]^T$$
$$\mathbf{V}_2 = \left[\mathbf{V}_{..1},\ldots,\mathbf{V}_{..P}\right]^T$$
$$\mathbf{V}_3 = \left[vec\left(\mathbf{V}_{..1}\right),\ldots,vec\left(\mathbf{V}_{..P}\right)\right]$$
$$\mathbf{V}_4 = \left[vec\left(\mathbf{V}^T_{..1}\right),\ldots,vec\left(\mathbf{V}^T_{..P}\right)\right].$$

(17)

In this paper, two following assumptions are satisfied.

(a) The antenna-to-slot allocation matrix $\mathbf{Q}$ does not have an all-zero column. This means that at least one transmit antenna is used during each time slot;

(b) Both the transmitter and receiver know the allocation matrix $\mathbf{Q}$ and the coding tensor $\underline{\mathbf{B}}$.

3.2. Uniqueness Issue

Due to the loading matrices factors being unique up to nonsingular matrices, the generalized Tucker-2 model is not essentially unique. This consequence can be verified by using the property of the mode-n product:

$$\underline{\mathbf{C}}\times_1\mathbf{H}\times_2\mathbf{S}=\underline{\mathbf{C}}\times_1\left(\mathbf{H}\mathbf{\Theta_H}\mathbf{\Theta_H}^{-1}\right)\times_2\left(\mathbf{S}\mathbf{\Theta_S}\mathbf{\Theta_S}^{-1}\right)$$
$$=\underline{\mathbf{C}}\times_1\mathbf{\Theta_H}^{-1}\times_2\mathbf{\Theta_S}^{-1}\times_1\left(\mathbf{H}\mathbf{\Theta_H}\right)\times_2\left(\mathbf{S}\mathbf{\Theta_S}\right)$$

(18)

where the noise tensor $\underline{\mathbf{V}}$ has been omitted for convenience of notation, $\mathbf{\Theta_S} \in \mathbb{C}^{R\times R}$ and $\mathbf{\Theta_H} \in \mathbb{C}^{M_S\times M_S}$ are nonsingular matrices.

It is shown that applying the uniqueness theorem of the Tucker model in [25], if the core tensor $\underline{\mathbf{C}}$ is known, then $\mathbf{S}$ and $\mathbf{H}$ are unique to a scaling ambiguity, i.e.,

$$(\mathbf{S},\ \mathbf{H}) = \left(\bar{\mathbf{S}}\mathbf{\Theta_S}^{-1},\ \bar{\mathbf{H}}\mathbf{\Theta_H}^{-1}\right)$$

(19)

where $\bar{\mathbf{S}}$ and $\bar{\mathbf{H}}$ are alternative solutions for $\mathbf{S}$ and $\mathbf{H}$, respectively, $\mathbf{\Theta_S} = \beta\mathbf{I}_R$ and $\mathbf{\Theta_H}= \left(1/\beta\right)\mathbf{I}_{M_S}$. Consequently, the priori knowledge of only one symbol is enough to resolve this scaling ambiguity factor β. Compared to the PARAFAC model used in existing receivers, the constructed Tucker-2 model only needs a priori knowledge of one symbol to eliminate the scaling ambiguity. Therefore, our scheme has higher spectral efficiency.

3.3. Identifiability Conditions

The identifiability for the constructed Tucker-2 model is an assignable problem for recovering the parameters to be estimated. In this paper, it is directly linked to the estimation of the signal matrix $\mathbf{S}$ and channel matrix $\mathbf{H}$ from the received signal tensor $\underline{\mathbf{Y}}$. Conditions of parameter identifiability is given in the following theorem.

Theorem 1. *(Sufficient Conditions): Assuming that* **H** *has independent and identically distributed (i.i.d.) entries, and* **S** *has a full column rank.* P_1 *denotes the number of nonzero elements in* **Q**. *Then sufficient conditions for identifiability of signal matrix* **S** *and channel matrix* **H** *are:*

$$P_1 \geqslant R \quad and \quad \min(M_D,\ R) \geqslant M_S. \tag{20}$$

Proof of Theorem 1. From Equation (12) and Equation (13), necessary and sufficient conditions for identifiability of **S** and **H** requires that $(\mathbf{I}_P \otimes \mathbf{H})\,\mathbf{F}_1$ and $(\mathbf{I}_P \otimes \mathbf{S})\,\mathbf{F}_2$ have a full column rank, i.e.

$$Rank\left((\mathbf{I}_P \otimes \mathbf{H})\,\mathbf{F}_1\right) = R \tag{21}$$

$$Rank\left((\mathbf{I}_P \otimes \mathbf{S})\,\mathbf{F}_2\right) = M_S. \tag{22}$$

Under the assumptions in Theorem 1 that **H** has i.i.d. entries, $M_D \geqslant M_S$ can ensure **H** has a full column rank. Since $\mathbf{I}_P$ and **H** have a full column-rank, then $\mathbf{I}_P \otimes \mathbf{H}$ has a full column rank, i.e., $Rank(\mathbf{I}_P \otimes \mathbf{H}) = PM_S$. Therefore, Equation (21) is satisfied if $\mathbf{F}_1$ has a full column rank. We rewrite $\mathbf{F}_1$ from Equation (15) as:

$$\mathbf{F}_1 = diag\left(vec\left(\mathbf{Q}^T\right)\right)\mathbf{B}_1 \tag{23}$$

where $\mathbf{B}_1 = \left[\mathbf{B}_{..1}^T, \ldots, \mathbf{B}_{..P}^T\right]^T$. Due to $\mathbf{B}_1$ being a specially constructed matrix with different generators, any two rows (or two columns) of $\mathbf{B}_1$ are linearly independent. If the number of non-zero rows of $\mathbf{F}_1$ is greater than or equal to the number of columns of $\mathbf{F}_1$ (i.e. $P_1 \geqslant R$), then $\mathbf{F}_1$ is full column rank. Thus, $M_D \geqslant M_S$ and $P_1 \geqslant R$ can ensure that condition (21) is satisfied.

Since **S** has a full column rank, we can deduce that $Rank(\mathbf{I}_P \otimes \mathbf{S}) = PR$. Thus, condition (22) is satisfied if $\mathbf{F}_2$ has a full column rank. We rewrite $\mathbf{F}_2$ from Equation (16) as:

$$\mathbf{F}_2 = \mathbf{B}_2^T \mathbf{Q}_P \tag{24}$$

where $\mathbf{B}_2 = \text{blk}\left[\mathbf{B}_{..1}, \ldots, \mathbf{B}_{..P}\right]$ and $\mathbf{Q}_P = \left[D_1(\mathbf{Q}), \ldots, D_P(\mathbf{Q})\right]^T$. Recall that **Q** does not have an all-zero column, which means that $\mathbf{Q}_P$ has a full column rank. We have that $\mathbf{F}_2$ is full column rank if $\mathbf{B}_2$ is full row rank. Since $\mathbf{B}_2$ has the block diagonal structure and $\mathbf{B}_{..p}$ has different generators, $R \geqslant M_S$ ensures that $\mathbf{B}_2$ has full row rank. Therefore, $R \geqslant M_S$ can ensure that condition (22) is satisfied. This ends the proof. □

Remark 1. *The conditions in Theorem 1 is sufficient but not necessary for parameter identifiability. Sufficient condition (21) and condition (22) also concern the ALS algorithm. In fact, identifiability of signal and channel parameters is possible in our simulation results when $M_S > R$. Necessary conditions for parameter identifiability is based on the dimensions of $(\mathbf{I}_P \otimes \mathbf{H})\,\mathbf{F}_1$ and $(\mathbf{I}_P \otimes \mathbf{S})\,\mathbf{F}_2$. If the channel matrix* **H** *does not have a full column or row rank, i.e., $L < \min(M_S, M_D)$, where L is the rank of the channel matrix* **H**. *Thus, identifiability conditions of Theorem 1 are no longer applicable because of the low-rank property of* **H**. *However, we can also deduce identifiability conditions based on Equations (21) and (22), i.e., necessary and sufficient conditions for identifiability of* **S** *and* **H** *requires that $(\mathbf{I}_P \otimes \mathbf{H})\,\mathbf{F}_1$ and $(\mathbf{I}_P \otimes \mathbf{S})\,\mathbf{F}_2$ have full column rank. For this case, we will do further analysis in Section 5.*

4. Semi-Blind Receiver

The ALS algorithm is a classical solution for fitting tensor models. However, it is well known that the ALS algorithm exhibits a convergence problem when collinearity is present in one or more modes [27,28]. The LM algorithm is successfully used to fit the PARAFAC and PARATUCK2 models, adapted to collinearity problems, and provide quadratic convergence [19,20]. As an iterative algorithm, the LM algorithm is also sensitive to initialization. Thus, the optimization of the initial value is important to improving the performance of the LM algorithm.

In this section, a novel semi-blind receiver based on the optimized LM algorithm is developed for joint symbol and channel estimation. The basic principle of the optimized LM algorithm is to first resort to a LSK approximation problem [29,30], based on the singular value decomposition (SVD) of rank-1 matrix to initialize the symbol and channel matrices, and then update these two matrices at the same time in each iteration. Finally, the modified singular value projection (SVP) based algorithm [31,32] is used to further improve the performance of channel estimation.

The proposed optimal initialization method is based on the Kronecker least squares algorithm, which exploits SVD-based rank-one approximations to get an initial estimation of **S** and **H** from their Kronecker matrix product.

By post-multiplying Equation (14) with $\mathbf{F}_3^\dagger$, we get $\mathbf{Z} = \left(\hat{\mathbf{S}}^{(0)} \otimes \hat{\mathbf{H}}^{(0)} \right) = \mathbf{Y}_3 \mathbf{F}_3^\dagger \in \mathbb{C}^{NM_D \times RM_S}$, where $\hat{\mathbf{S}}^{(0)}$ and $\hat{\mathbf{H}}^{(0)}$ are initial estimates of **S** and **H**. According to the Theorem 2.1 in [29], we have:

$$\left\| \mathbf{Z} - \hat{\mathbf{S}}^{(0)} \otimes \hat{\mathbf{H}}^{(0)} \right\|_F^2 = \left\| \mathbf{\Xi} - vec\left(\hat{\mathbf{H}}^{(0)} \right) \left(vec\left(\hat{\mathbf{S}}^{(0)} \right) \right)^T \right\|_F^2 \tag{25}$$

where $\mathbf{\Xi} = unvec(\mathbf{\Delta}) \in \mathbb{C}^{M_D M_S \times NR}$ is a rank-one matrix, and $\mathbf{\Delta} \in \mathbb{C}^{NM_D RM_S \times 1}$ is, given that:

$$\mathbf{\Delta} = \begin{bmatrix} vec(\mathbf{Z}(1:M_D, \ 1:M_S)) \\ \vdots \\ vec(\mathbf{Z}((N-1)M_D + 1:NM_D, \ 1:M_S)) \\ \vdots \\ vec(\mathbf{Z}(1:M_D, \ (R-1)M_S + 1:RM_S)) \\ \vdots \\ vec(\mathbf{Z}((N-1)M_D + 1:NM_D, \ (R-1)M_S + 1:RM_S)) \end{bmatrix} \tag{26}$$

In this case, the Kronecker product matrix **Z** has been rearranged into a rank-one matrix $\mathbf{\Xi}$. Applying SVD to the rank-one matrix $\mathbf{\Xi}$, the vectors $vec\left(\hat{\mathbf{S}}^{(0)} \right)$ and $vec\left(\hat{\mathbf{H}}^{(0)} \right)$ can be estimated by using a rank-one approximation method, i.e., by computing its largest singular value and the corresponding left and right singular vectors. $\hat{\mathbf{S}}^{(0)}$ and $\hat{\mathbf{H}}^{(0)}$ are determined up to a scaling factor, which can be removed by setting $s_{1,1} = 1$ as in [27,30]. The detailed process is shown below.

By applying SVD to the rank-one matrix $\mathbf{\Xi}$, we have:

$$\mathbf{\Xi} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H \tag{27}$$

where $\mathbf{\Sigma} \in \mathbb{C}^{M_D M_S \times NR}$ is a diagonal matrix containing singular values of $\mathbf{\Xi}$, $\mathbf{U} \in \mathbb{C}^{M_D M_S \times M_D M_S}$ and $\mathbf{V} \in \mathbb{C}^{NR \times NR}$ are unitary matrices. Using the rank-one approximation of $\mathbf{\Xi}$, we have:

$$\mathbf{\Xi} \approx \mathbf{U}_{.1}\sigma_1 \mathbf{V}_{.1}^H \tag{28}$$

where $\sigma_{.1}$ is the largest singular value, and $\mathbf{U}_{.1}$ and $\mathbf{V}_{.1}$ are the corresponding left and right singular vectors. Thus, vectors $vec\left(\hat{\mathbf{S}}^{(0)} \right)$ and $vec\left(\hat{\mathbf{H}}^{(0)} \right)$ can be estimated as:

$$vec\left(\hat{\mathbf{S}}^{(0)} \right) = \alpha \mathbf{V}_{.1}^*, \quad vec\left(\hat{\mathbf{H}}^{(0)} \right) = \frac{\sigma_1}{\alpha} \mathbf{U}_{.1} \tag{29}$$

where α is the scalar factor. $vec\left(\hat{\mathbf{S}}^{(0)} \right)$ and $vec\left(\hat{\mathbf{H}}^{(0)} \right)$ are determined up to this scalar factor. In practical communication systems, this scalar factor α can be removed by setting $s_{1,1} = 1$. Thus, the value α in this paper is equal to $\frac{1}{v_{1,1}^*}$ when $s_{1,1} = 1$. Note that we can also choose Equation (14) to implement the above optimal initialization procedure.

Define a parameter vector stacking all the unknowns as:

$$\mathbf{u} = \left[\mathbf{u}_\mathbf{S}^T, \ \mathbf{u}_\mathbf{H}^T\right]^T \in \mathbb{C}^{Q \times 1} \tag{30}$$

where $\mathbf{u_S} = vec\left(\mathbf{S}^T\right) \in \mathbb{C}^{NR \times 1}$, $\mathbf{u_H} = vec\left(\mathbf{H}^T\right) \in \mathbb{C}^{M_D M_S \times 1}$, and $Q = NR + M_D M_S$. The cost function to be minimized is given by:

$$\begin{aligned} \phi\left(\mathbf{u}\right) &= \frac{1}{2} \sum_{m_D=1}^{M_D} \sum_{n=1}^{N} \sum_{p=1}^{P} \left|\tilde{y}_{m_D,n,p}\left(\mathbf{u}\right) - y_{m_D,n,p}\right|^2 \\ &= \frac{1}{2} \sum_{l=1}^{L} \left|z_l\left(\mathbf{u}\right)\right|^2 = \frac{1}{2}\mathbf{z}^H\left(\mathbf{u}\right)\mathbf{z}\left(\mathbf{u}\right) \end{aligned} \tag{31}$$

where $\tilde{y}_{m_D,n,p}\left(\mathbf{u}\right)$ is the typical element of the tensor $\underline{\tilde{\mathbf{Y}}}\left(\mathbf{u}\right) \in \mathbb{C}^{M_D \times N \times P}$, which denotes the output tensor in absence of noise. $\mathbf{z}\left(\mathbf{u}\right) = vec\left(\underline{\tilde{\mathbf{Y}}}\left(\mathbf{u}\right)\right) - vec\left(\underline{\mathbf{Y}}\right) = \tilde{\mathbf{y}}\left(\mathbf{u}\right) - \mathbf{y} \in \mathbb{C}^{NPM_D \times 1}$ denotes the vector of residuals and $L = NPM_D$.

Let the $\mathbf{J} \in \mathbb{C}^{L \times Q}$ be the Jacobian matrix of $\mathbf{z}\left(\mathbf{u}\right)$ with respect to $\mathbf{u}$, and $\mathbf{g}$ be the gradient of $\phi\left(\mathbf{u}\right)$ with respect to $\mathbf{u}$. $\mathbf{J}$ and $\mathbf{g}$ are respectively defined by:

$$J_{l,q} = \partial z_l\left(\mathbf{u}\right)/\partial \mathbf{u}_q = \partial \tilde{y}_l\left(\mathbf{u}\right)/\partial \mathbf{u}_q \tag{32}$$

$$\mathbf{g} = \partial \phi\left(\mathbf{u}\right)/\partial \mathbf{u} = \mathbf{J}^H\left(\mathbf{u}\right)\mathbf{z}\left(\mathbf{u}\right). \tag{33}$$

The optimized LM algorithm consists in optimizing $\mathbf{u}^{(0)}$, and estimating $\mathbf{u}^{(i+1)}$ at the $(i+1)$-th iteration from $\mathbf{u}^{(i)}$ at the i-th iteration via $\mathbf{u}^{(i+1)} = \mathbf{u}^{(i)} + \Delta\mathbf{u}^{(i)}$. The step $\Delta\mathbf{u}^{(i)} \in \mathbb{C}^{NPM_D \times 1}$ is updated by solving the following modified normal equations:

$$\left(\mathbf{J}^{(i)^H}\mathbf{J}^{(i)} + \lambda^{(i)}\mathbf{I}_Q\right)\Delta\mathbf{u}^{(i)} = -\mathbf{g}^{(i)} \tag{34}$$

where $\lambda^{(i)}$ is the damping parameter to ensure that $\Delta\mathbf{u}^{(i)}$ is a descent direction. The whole procedure of the optimized LM algorithm used in our semi-blind receiver is listed in Algorithm 1.

Due to the partitioned structure of $\mathbf{u}$, the Jacobian matrix $\mathbf{J}$ can be written as $\mathbf{J} = [\mathbf{J}_{\mathbf{u_S}}, \ \mathbf{J}_{\mathbf{u_H}}]$, where $\mathbf{J}_{\mathbf{u_S}} \in \mathbb{C}^{NPM_D \times NR}$ and $\mathbf{J}_{\mathbf{u_H}} \in \mathbb{C}^{NPM_D \times M_D M_S}$ are respectively given by:

$$\mathbf{J}_{\mathbf{u_S}} = \mathbf{I}_N \otimes \left(\left(\mathbf{I}_P \otimes \mathbf{H}\right)\mathbf{F}_1\right) \tag{35}$$

$$\mathbf{J}_{\mathbf{u_H}} = \mathbf{\Pi}\left(\mathbf{I}_{M_D} \otimes \left(\left(\mathbf{I}_P \otimes \mathbf{S}\right)\mathbf{F}_2\right)\right). \tag{36}$$

The permutation matrix $\mathbf{\Pi} \in \mathbb{C}^{NPM_D \times M_D PN}$ is given by:

$$\mathbf{\Pi} = \sum_{m_D=1}^{M_D} \sum_{n=1}^{N} \mathbf{e}_n^{(N)}\left(\mathbf{e}_{m_D}^{(M_D)}\right)^T \otimes \mathbf{I}_P \otimes \mathbf{e}_{m_D}^{(M_D)}\left(\mathbf{e}_n^{(N)}\right)^T \tag{37}$$

where $\mathbf{e}_n^{(N)}$ and $\mathbf{e}_{m_D}^{(M_D)}$ are the n-th and m_D-th column vectors of the identity matrices $\mathbf{I}_N$ and $\mathbf{I}_{M_D}$, respectively.

Algorithm 1 The optimized LM algorithm

First stage:

• Compute the LS estimate of $\mathbf{Z}$: $\mathbf{Z} = \mathbf{Y}_3 \left(\mathbf{F}_3\right)^{\dagger}$;

• Rearrange $\mathbf{Z}$ to a rank-one matrix Ξ;

• Apply the SVD on Ξ: $SVD\left(\Xi\right) = \mathbf{U}\Sigma\mathbf{V}^H$;

• Calculate initialization matrices $\mathbf{S}^{(0)}$ and $\mathbf{H}^{(0)}$:

$\mathbf{S}^{(0)} = unvec\left(\mathbf{V}_{.1}^{*}/v_{1,1}^{*}\right)$, $\mathbf{H}^{(0)} = unvec\left(\sigma_1 \mathbf{U}_{.1} v_{1,1}^{*}\right)$.

Second stage :

Initialization: Initialize $\mathbf{u}^{(0)} = \left[\mathbf{u}_{\hat{\mathbf{S}}^{(0)}}^T, \mathbf{u}_{\hat{\mathbf{H}}^{(0)}}^T\right]^T$, $\lambda^{(0)} = \max\left(\text{diag}\left(\mathbf{J}^{(0)H}\mathbf{J}^{(0)}\right)\right)$ and $\tau = 2$; set $\varepsilon = 10^{-5}$ and $i = 1$;

while $\left|\phi\left(\mathbf{u}^{(i)}\right) - \phi\left(\mathbf{u}^{(i-1)}\right)\right|/\left|\phi\left(\mathbf{u}^{(i)}\right)\right| \geqslant \varepsilon$ **do**

Step 1. Compute $\mathbf{J}^{(i)H}\mathbf{J}^{(i)}$ and $\mathbf{g}^{(i)}$ respectively;

Step 2. Compute $\Delta\mathbf{u}^{(i)}$:

$\quad \Delta\mathbf{u}^{(i)} = -\left(\mathbf{J}^{(i)H}\mathbf{J}^{(i)} + \lambda^{(i)}\mathbf{I}_Q\right)^{-1}\mathbf{g}^{(i)}$;

Step 3. Update $\mathbf{u}^{(i+1)}$: $\mathbf{u}^{(i+1)} = \mathbf{u}^{(i)} + \Delta\mathbf{u}^{(i)}$;

Step 4. Calculate the gain rate α: $\alpha = \frac{\phi\left(\mathbf{u}^{(i+1)}\right) - \phi\left(\mathbf{u}^{(i)}\right)}{\delta^{(i)}}$, where $\delta^{(i)} = \left(\mathbf{J}^{(i)}\Delta\mathbf{u}^{(i)}\right)^H \mathbf{z}\left(\mathbf{u}^{(i)}\right) + \frac{1}{2\left\|\mathbf{J}^{(i)}\Delta\mathbf{u}^{(i)}\right\|_F^2}$;

Step 5. Update λ : If $\alpha \geqslant 0$, $\mathbf{u}^{(i+1)}$ is ture, and set $\lambda^{(i+1)} = \lambda^{(i)}\max\left(1 - (2\alpha - 1)^3, 1/3\right)$ and $\tau = 2$. Otherwise, $\mathbf{u}^{(i+1)}$ is invalid, and set $\lambda^{(i+1)} = \tau\lambda^{(i)}$ and $\tau \leftarrow 2\tau$;

Step 6. $i \leftarrow i + 1$;

end

Acquire $\mathbf{S}^{(\infty)}$ and $\mathbf{H}^{(\infty)}$: $\mathbf{S}^{(\infty)} = \left(unvec\left(\mathbf{u}_{\mathbf{S}}^{(\infty)}\right)\right)^T$, $\mathbf{H}^{(\infty)} = \left(unvec\left(\mathbf{u}_{\mathbf{H}}^{(\infty)}\right)\right)^T$.

Compute $\mathbf{H}_{new}^{(\infty)}$: If $L < \min\left(M_S, M_D\right)$, $\mathbf{H}_{new}^{(\infty)} = SVP\left(\mathbf{H}^{(\infty)}\right)$. Otherwise, $\mathbf{H}_{new}^{(\infty)} = \mathbf{H}^{(\infty)}$.

Remove the scaling ambiguity: $\hat{\mathbf{S}}^{(final)} = \frac{\hat{\mathbf{S}}^{(\infty)}}{\hat{s}_{1,1}^{(\infty)}}$, $\hat{\mathbf{H}}^{(final)} = \hat{s}_{1,1}^{(\infty)}\mathbf{H}_{new}^{(\infty)}$.

We can then build the blocks of $\mathbf{J}^H\mathbf{J}$ as follows:

$$\mathbf{J}^H\mathbf{J} = \begin{bmatrix} \mathbf{J}_{\mathbf{u_S}}^H\mathbf{J}_{\mathbf{u_S}} & \mathbf{J}_{\mathbf{u_S}}^H\mathbf{J}_{\mathbf{u_H}} \\ \left(\mathbf{J}_{\mathbf{u_S}}^H\mathbf{J}_{\mathbf{u_H}}\right)^H & \mathbf{J}_{\mathbf{u_H}}^H\mathbf{J}_{\mathbf{u_H}} \end{bmatrix} \tag{38}$$

The terms $\mathbf{J}_{\mathbf{u_S}}^H\mathbf{J}_{\mathbf{u_S}}$, $\mathbf{J}_{\mathbf{u_H}}^H\mathbf{J}_{\mathbf{u_H}}$ and $\mathbf{J}_{\mathbf{u_S}}^H\mathbf{J}_{\mathbf{u_H}}$ can be respectively written as:

$$\mathbf{J}_{\mathbf{u_S}}^H\mathbf{J}_{\mathbf{u_S}} = \mathbf{I}_N \otimes \left(\mathbf{F}_1^H\left(\mathbf{I}_P \otimes \mathbf{H}^H\mathbf{H}\right)\mathbf{F}_1\right) \tag{39}$$

$$\mathbf{J}_{\mathbf{u_H}}^H\mathbf{J}_{\mathbf{u_H}} = \mathbf{I}_{M_D} \otimes \left(\mathbf{F}_2^H\left(\mathbf{I}_P \otimes \mathbf{S}^H\mathbf{S}\right)\mathbf{F}_2\right) \tag{40}$$

$$\mathbf{J}_{\mathbf{u_S}}^H\mathbf{J}_{\mathbf{u_H}} = \left(\mathbf{I}_N \otimes \left(\mathbf{I}_P \otimes \mathbf{H}\right)\mathbf{F}_1\right)^H\Pi\left(\mathbf{I}_{M_D} \otimes \left(\mathbf{I}_P \otimes \mathbf{S}\right)\mathbf{F}_2\right). \tag{41}$$

Similarly, the partitioned structure of $\mathbf{u}$ allows us to write $\mathbf{g}$ as the concatenation of the following two gradients:

$$\mathbf{g} = \begin{bmatrix} \partial\phi\left(\mathbf{u}\right)/\partial\mathbf{u_S} \\ \partial\phi\left(\mathbf{u}\right)/\partial\mathbf{u_H} \end{bmatrix} = \begin{bmatrix} \mathbf{g}_{\mathbf{u_S}} \\ \mathbf{g}_{\mathbf{u_H}} \end{bmatrix} \tag{42}$$

where $\mathbf{g_{u_S}} \in \mathbb{C}^{NR \times 1}$ and $\mathbf{g_{u_H}} \in \mathbb{C}^{M_D M_S \times 1}$ are respectively given by:

$$\mathbf{g_{u_S}} = \mathbf{J}_{\mathbf{u_S}}^H \mathbf{J}_{\mathbf{u_S}} \mathbf{u_S} - \mathbf{J}_{\mathbf{u_S}}^H \mathbf{y} \tag{43}$$

$$\mathbf{g_{u_H}} = \mathbf{J}_{\mathbf{u_H}}^H \mathbf{J}_{\mathbf{u_H}} \mathbf{u_H} - \mathbf{J}_{\mathbf{u_H}}^H \mathbf{y}. \tag{44}$$

In Algorithm 1, the estimated matrix $\mathbf{H}^{(\infty)}$ is projected onto a low rank estimated matrix $\mathbf{H}_{new}^{(\infty)}$ by the SVP based algorithm when $L < \min(M_S, M_D)$. Here $\mathbf{H}_{new}^{(\infty)}$ is calculated as $\mathbf{H}_{new}^{(\infty)} = SVP\left(\mathbf{H}^{(\infty)}\right) = \sum_{l=1}^{L} \beta_l \mathbf{U}_{\cdot l}^{(C)} \left(\mathbf{V}_{\cdot l}^{(C)}\right)^H$, where β_l denotes the l-th largest singular value of $\mathbf{H}^{(\infty)}$, $\mathbf{U}_{\cdot l}^{(C)}$ and $\mathbf{V}_{\cdot l}^{(C)}$ are the corresponding left and right singular vectors. The overall complexity of the optimized LM algorithm mainly depends on the per-iteration complexity and the numbers of iterations. The per-iteration complexity of this algorithm can be estimated as $\mathcal{O}\left((NR + M_D M_S)^3\right)$. Since the antenna-to-slot allocation matrix and the coding tensor are fixed and known at the receiver, the convergence of the optimized LM algorithm is usually achieved in only a few iterations. The average number of iterations for the optimized LM algorithm will be further analyzed in Section 6.

5. Extension to Multi-User Massive Mimo Systems

In the following section, we show that the developed algorithm can be applied to multi-user massive MIMO systems with hybrid precoding architecture for joint symbol and channel estimation. We consider a fully-connected hybrid precoding architecture, which is the typical model of massive MIMO systems. The base station communicates with M users simultaneously, and each mobile station is equipped with M_D antennas. The base station is equipped with M_S antennas and M_{RF} independent radio frequency chains to transmit R streams for M_D receive antennas in each mobile station. In the considered downlink system, each symbol $s_{n,r}$ is coded by a three-dimensional baseband code $b_{m_{RF},r,p}^{(M)}$ followed by a radio frequency code $e_{m_S,m_{RF}}$ in the base station. At the m-th ($m = 1, \ldots, M$) mobile station, the discrete-time baseband signal at the m_D-th receive antenna is written as:

$$y_{m_D,n,p}^{(m)} = \sum_{m_S=1}^{M_S} \sum_{m_{RF}=1}^{M_{RF}} \sum_{r=1}^{R} h_{m_D,m_S}^{(m)} q_{p,m_S} e_{m_S,m_{RF}} b_{m_{RF},r,p} s_{n,r} + v_{m_D,n,p}^{(m)} \tag{45}$$

where $e_{m_S,m_{RF}}$ and $h_{m_D,m_S}^{(m)}$ are (m_S, m_{RF})-th and (m_D, m_S)-th elements of the radio frequency precoder matrix $\mathbf{E} \in \mathbb{C}^{M_S \times M_{RF}}$ and the massive MIMO channel matrix $\mathbf{H}^{(m)} \in \mathbb{C}^{M_D \times M_S}$, respectively. $y_{m_D,n,p}^{(m)}$ is the typical element of the received signal tensor $\underline{\mathbf{Y}}^{(m)} \in \mathbb{C}^{M_D \times N \times P}$. Then Equation (45) can be rewritten as:

$$y_{m_D,n,p}^{(m)} = \sum_{m_S=1}^{M_S} \sum_{r=1}^{R} h_{m_D,m_S}^{(m)} c_{m_S,r,p} s_{n,r} + v_{m_D,n,p}^{(m)} \tag{46}$$

where,

$$c_{m_S,r,p} = \sum_{m_{RF}=1}^{M_{RF}} q_{p,m_S} e_{m_S,m_{RF}} b_{m_{RF},r,p} \tag{47}$$

Following [33,34], we also adopt a geometric channel model with L_m scatterers between the base station and the m-th mobile station, $L_m = 1, \ldots, L_M$. Under this model, the channel matrix $\mathbf{H}^{(m)}$ is expressed as:

$$\mathbf{H}^{(m)} = \sum_{l=1}^{L_m} \alpha_l^{(m)} \mathbf{\Lambda}_{MS}\left(\theta_l^{(m)}\right) \mathbf{\Lambda}_{BS}\left(\phi_l^{(m)}\right) \mathbf{a}_{MS}\left(\theta_l^{(m)}\right) \mathbf{a}_{BS}^H\left(\phi_l^{(m)}\right) \tag{48}$$

where $\alpha_l^{(m)}$ denotes the complex gain of *l*-th path, $\theta_l^{(m)}$, and $\phi_l^{(m)}$ are *l*-th 's azimuth angles of arrival and departure (AoAs/AoDs) of the mobile station and base station, respectively. $\boldsymbol{\Lambda}_{MS}\left(\theta_l^{(m)}\right)$ and $\boldsymbol{\Lambda}_{BS}\left(\phi_l^{(m)}\right)$ are receive and transmit antenna array at a specific AoA and AoD, respectively. Finally, $\mathbf{a}_{BS}\left(\phi_l^{(m)}\right)$ and $\mathbf{a}_{MS}\left(\theta_l^{(m)}\right)$ are the steering vectors at the base station and mobile station, respectively. If uniform linear arrays are considered, the steering vectors $\mathbf{a}_{BS}\left(\phi_l^{(m)}\right)$ and $\mathbf{a}_{MS}\left(\theta_l^{(m)}\right)$ are respectively given by:

$$\mathbf{a}_{BS}\left(\phi_l^{(m)}\right) = \frac{1}{M_S}[1,\ e^{j\frac{2\pi}{\lambda}d\sin\phi_l^{(m)}},\cdots,\ e^{j\frac{2\pi}{\lambda}d(M_S-1)\sin\phi_l^{(m)}}]^T \tag{49}$$

$$\mathbf{a}_{MS}\left(\theta_l^{(m)}\right) = \frac{1}{M_D}[1,\ e^{j\frac{2\pi}{\lambda}d\sin\theta_l^{(m)}},\cdots,\ e^{j\frac{2\pi}{\lambda}d(M_D-1)\sin\theta_l^{(m)}}]^T \tag{50}$$

where λ denotes the signal wavelength, and d is the distance between two neighboring antenna elements.

Similar to the analysis of Section 3.1, the received signal tensor $\underline{\mathbf{Y}}^{(m)}$ of noiseless signal also satisfies the Tucker-2 model, and the proposed algorithm in Section 4 remains suitable for joint symbol and channel estimation at each mobile station. However, two points are important to note here. First, identifiability conditions of Theorem 1 are no longer applicable because of the low-rank property of $\mathbf{H}^{(m)}$. However, we can deduce new identifiability conditions based on Equations (21) and (22), i.e., necessary and sufficient conditions for identifiability of $\mathbf{S}$ and $\mathbf{H}^{(m)}$ require that $\left(\mathbf{I}_P\otimes\mathbf{H}^{(m)}\right)\mathbf{F}_1$ and $(\mathbf{I}_P\otimes\mathbf{S})\mathbf{F}_2$ have full column rank. For convenience of analysis, we assume that the antenna-to-slot allocation matrix is all-ones matrix. Then, we have the following theorem.

Theorem 2. *Assuming that the path gains of the low-rank channel $\mathbf{H}^{(m)}$ are Rayleigh distributed, and N and R are large enough. Then sufficient conditions for identifiability of $\mathbf{H}^{(m)}$ and $\mathbf{S}$ are:*

$$P \geqslant \max\left(\frac{R}{L_m},\ \frac{M_S}{N},\ \frac{M_S}{R}\right) \tag{51}$$

Proof of Theorem 2. The channel model $\mathbf{H}^{(m)}$ is expressed as Equation (48). The rank of $\mathbf{H}^{(m)}$ is L_m, and the path gains of the $\mathbf{H}^{(m)}$ are Rayleigh distributed. $\mathbf{F}_1$ is a full rank matrix, which contains different generators. Consequently, $\min\left(PM_D,\ PL_m,\ PM_S\right)\geqslant R$ ensures that $(\mathbf{I}_P\otimes\mathbf{H})\mathbf{F}_1$ have full column rank. Since $\mathbf{H}^{(m)}$ is a low-rank, i.e., $L_m < \min\left(M_S,\ M_D\right)$, $P\geqslant\frac{R}{L_m}$ can ensure $(\mathbf{I}_P\otimes\mathbf{H})\mathbf{F}_1$ has the full column rank. Since N and R are large enough, and $\mathbf{S}$ has the random nature, the rank of $\mathbf{S}$ is equal to N or R. Moreover, $\mathbf{F}_2$ is also a full rank matrix because of its special structure. We deduce that $(\mathbf{I}_P\otimes\mathbf{S})\mathbf{F}_2$ is full column rank if $\min\left(PN,\ PR\right)\geqslant M_S$, i.e., $P\geqslant\max\left(\frac{M_S}{N},\ \frac{M_S}{R}\right)$. Therefore, condition (51) can ensure identifiability of $\mathbf{H}^{(m)}$ and $\mathbf{S}$. This ends the proof of Theorem 2. □

Second, the low-rank property of the mmWave massive MIMO channel should be exploited. Due to very limited scattering of the mmWave channel and larger quantities of transmitting and receiving antennas, L_m is usually less than M_S and M_D. Different from the conventional MIMO channel matrix that usually has full column or row rank, the rank of the mmWave massive MIMO channel matrix is much smaller than its dimension. This is called 'low-rank property' of the mmWave massive MIMO channel matrix. Therefore, the final part of the proposed **Algorithm 1** takes advantage of this low-rank constraint $rank\left(\mathbf{H}^{(m)}\right)\leqslant L_m$ to further improve the estimation accuracy of the channel.

6. Simulation Results and Discussion

We studied the performance of the proposed semi-blind receiver through numerical simulations. The channel matrix $\mathbf{H}$ has independent and identically distributed (i.i.d.) complex Gaussian entries with zero-mean and unit variance. The default values of the system parameters are set to

$M_S = M_D = 4$, and the antenna-to-slot allocation matrix is all-ones matrix. Throughout the simulation, the coding tensor $\underline{\mathbf{C}}$ is known at the receiver. Quadrature phase-shift keying (QPSK) constellations are used to modulate the transmitted symbols. All results are averaged over 10,000 independent Monte Carlo simulations. As in [8,9], the signal-to-noise ratio (SNR) at the receiver is defined as:

$$SNR = 10\log_{10}\left(\left\|\underline{\tilde{\mathbf{Y}}}\right\|_F^2 \Big/ \left\|\underline{\mathbf{V}}\right\|_F^2\right) \quad dB \tag{52}$$

where $\underline{\tilde{\mathbf{Y}}}$ denotes the noise-free signal tensor (the tensor-of-interest) containing both symbol and channel parameters. For each channel realization, the normalized mean square error (NMSE) for different receivers is computed as $\left\|\mathbf{H} - \hat{\mathbf{H}}\right\|_F^2 \Big/ \left\|\mathbf{H}\right\|_F^2$, where $\hat{\mathbf{H}}$ is the estimation of $\mathbf{H}$ at convergence.

In the first example, we evaluate the convergence performance of the optimized LM algorithm, which is used in our semi-blind receiver. We assume the system design parameters $N = P = 5$ and $R = 3$. In Figure 2, the average value of the cost function is plotted versus the number of iterations, for three SNR values. We observe from Figure 2 that for each SNR value, the cost function decreases as the number of iterations increases until the algorithm converges. We can also see that for the same number of iterations, the cost function decreases as SNR increases. The proposed algorithm needs few iterations to converge. For instance, the optimized LM algorithm achieves convergence in about 10 iterations at the SNR of 20 dB.
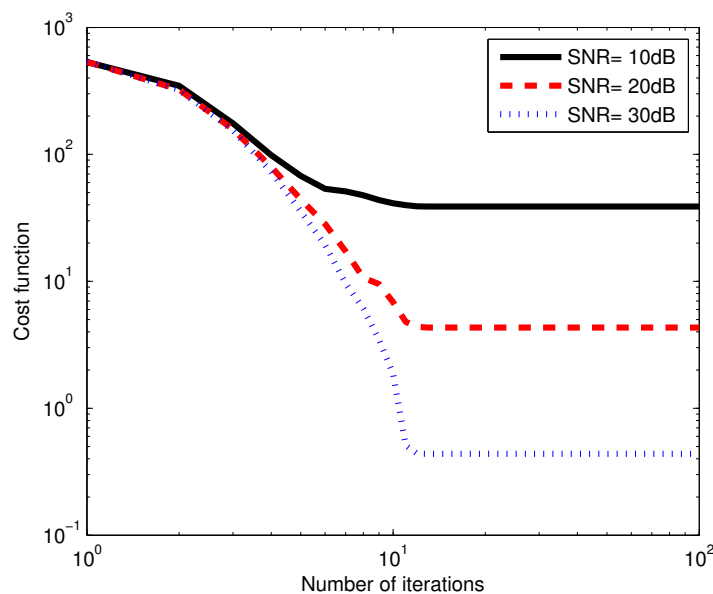
**Figure 2.** Cost function versus the number of iterations.

In the second example, we evaluated the estimation performance of the proposed semi-blind receiver in terms of bit error rate (BER) and the NMSE of channel estimation. In particular, we compared the PARAFAC-based receiver with KRST (P-KRST) coding scheme in [8] and the training-based receiver with the space-time (TB-ST) coding scheme. For the TB-ST scheme, the symbol matrix is composed of two parts as in [9], i.e., the training symbol matrix and the unknown data symbol matrix. N_{tr} denotes the length of the channel training sequence in the TB-ST receiver.

The transmission rates for the proposed coding scheme and the KRST coding scheme are $\frac{RN}{PN} = \frac{R}{P}$ and $\frac{M_S N}{PN} = \frac{M_S}{P}$ (data symbols per symbol period), respectively. However, the KRST coding scheme needs to know the first column of signal matrix $\mathbf{S}$ to eliminate the scaling ambiguity, while the proposed coding scheme only needs to know $s_{1,1}$ to eliminate the scaling ambiguity. Thus, the efficient transmission rates for the proposed coding scheme and the KRST coding scheme are $\frac{RN-1}{PN}$ and $\frac{M_S(N-1)}{PN}$, respectively. To ensure a fair comparison, the proposed coding scheme and the KRST coding

scheme should keep the same efficient transmission rate, i.e., $N = \frac{M_S - 1}{M_S - R}$. Thus, the system design parameters in this example are set equal to $M_S = 4$, $P = 7$, and $R = N = 3$. For TB-ST coding scheme, we divide $P = P_{tr} + P_d$, where blocks $P_{tr} = 2$ and $P_{tr} = 5$ are used for channel training and data transmitting. Therefore, the length of the channel training sequence in the TB-ST receiver is $N_{tr} = P_{tr}N = 6$.

The BER performance of different receivers versus SNR is shown in Figure 3. It can be seen that the proposed semi-blind receiver outperforms the P-KRST and TB-ST receiver. The NMSE performance of the different receivers is demonstrated in Figure 4. It can be seen from Figure 4 that the P-KRST receiver has the best performance of channel estimation, and the proposed semi-blind receiver yields a smaller NMSE compared with the TB-ST receiver. From [8], the per-iteration complexity in the PARAFAC based receiver is $\mathcal{O}\left(M_S M_D P N\right)$. The complexity of the TB-ST scheme can be estimated as $\mathcal{O}\left(N_r M_S\left(M_D + N_r\right) + RPM_D\left(N + R\right)\right)$. The per-iteration complexity of the proposed O-LM algorithm is given at the end of Section 4. The TB-ST scheme has the least computational complexity due to the use of the channel training sequence. Due to the adoption of the simple KRST coding scheme, the PARAFAC based receiver has lower complexity than that of the proposed receiver. However, the TB-ST receiver requires a long channel training sequence, the PARAFAC-based receiver needs to know the first column or row of the signal matrix to eliminate the scaling ambiguity, but the proposed receiver only needs to know one symbol of the signal matrix.

In the third example, we evaluated and compared the performance of the traditional ALS (T-ALS) and optimized LM (O-LM) algorithms. We assume the system design parameters $N = P = L$ and $R = 5$. Correlated MIMO channel is considered in this example, and the channel matrix **H** is modeled as in [35], where ρ denotes the normalized correlation coefficient with magnitude $|\rho| \leq 1$. We consider $\rho = 0$ (non-correlation) and $\rho = 0.8$ (strong correlation), respectively. For each Monte Carlo run, the T-ALS algorithm is initialized with ten different random matrices as in [20,36]. The estimation performance is evaluated after selecting the best initialization, which is the one that results in the minimum value of $\delta^{(j)}$.
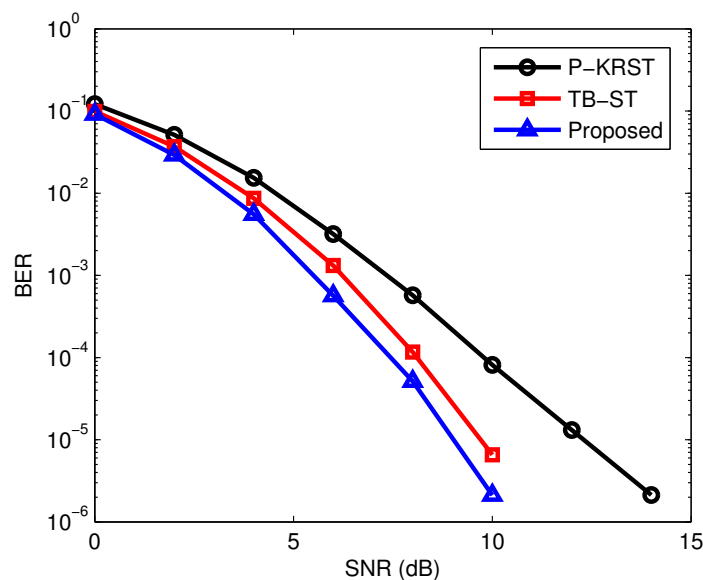
Figure 3. Bit error rate (BER) performance of different receivers versus signal-to-noise ratio (SNR).

We observe from Figure 5 that the T-ALS and O-LM algorithms give a similar BER and NMSE performance, which means that these two algorithms converge to the same point. For the right subfigure of Figure 5, the NMSE of the T-ALS and O-LM algorithms is also shown in Table 1 for the sake of comparison. We can also observe from Figure 5 that for these two algorithms, BER and NMSE performance degrade when the channel becomes strongly correlated. The overall complexities of

the O-LM algorithm and the ALS algorithm depend on the per-iteration complexity and the number of iterations. The per-iteration complexity of the O-LM algorithm is higher than that of the T-ALS algorithm. However, because of the robustness of the O-LM algorithm, the O-LM algorithm needs fewer iterations compared with the T-ALS algorithm. Therefore, the proposed algorithm has lower complexity compared with the existing T-ALS algorithm. The mean processing times required in the T-ALS and O-LM algorithms are shown in Figure 6. We observe that the mean processing time required in the O-LM algorithm is shorter than that of the T-ALS algorithm, especially when the channel becomes strongly correlated. From Figure 6 we can also observe that the advantage of the O-LM algorithm is obvious as *L* decreases from 8 to 7 compared with the T-ALS algorithm.
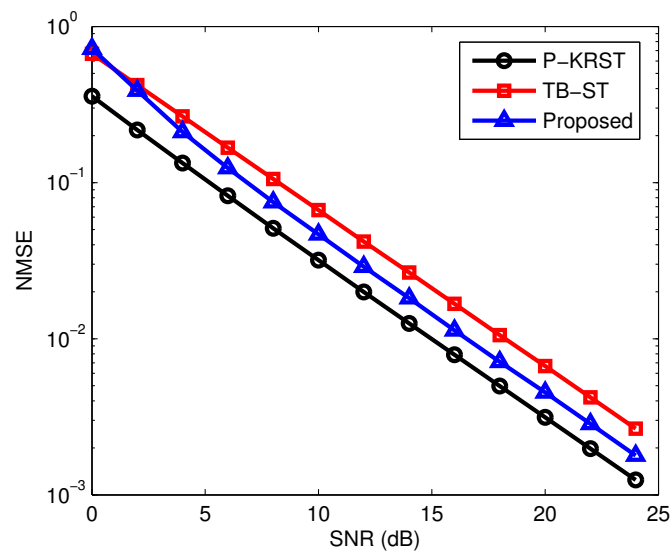
Figure 4. Normalized mean square error (NMSE) performance of different receivers versus SNR.
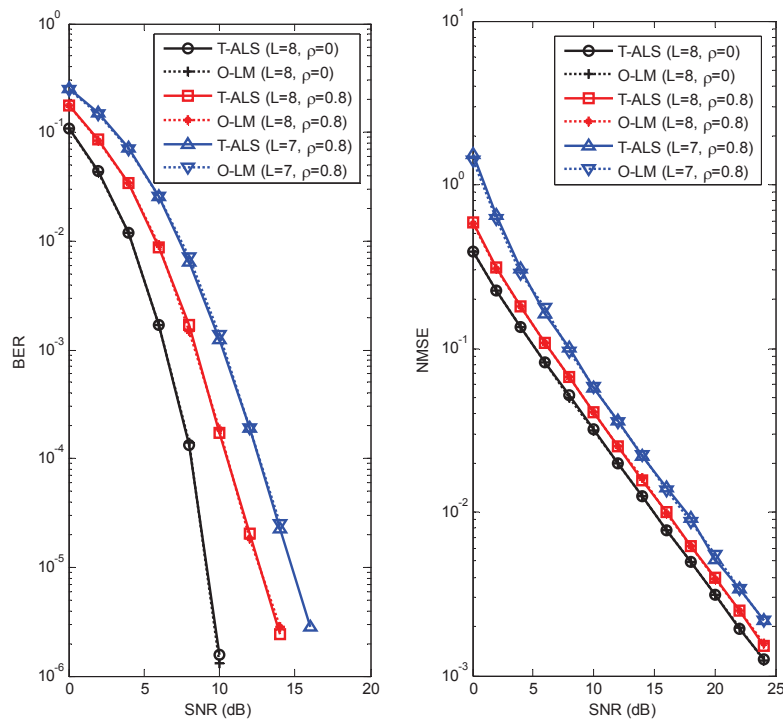
Figure 5. BER and NMSE performance of traditional alternating least squares (T-ALS) and O-LM algorithms for different *L* and *ρ*.
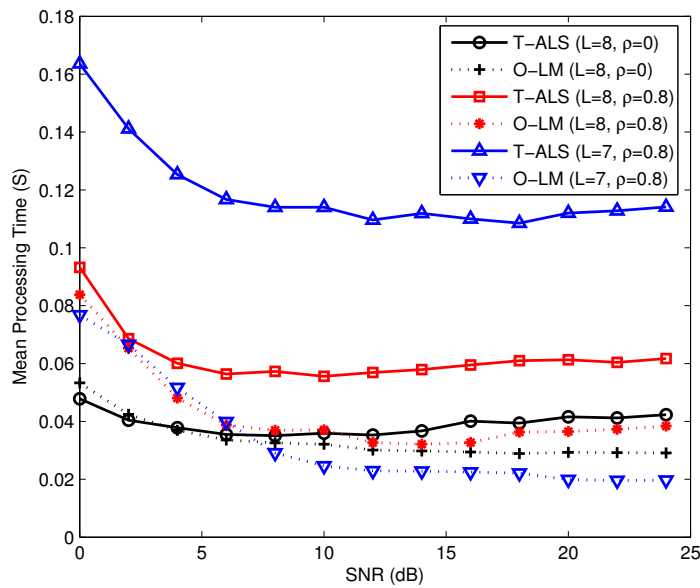
Figure 6. The mean processing times required in T-ALS and O-LM algorithms versus SNR.

Table 1. NMSE of the T-ALS and O-LM algorithms.

SNR (dB)	0	2	4	6	8	10	12	14	16	18	20	22	24
TALS, $L = 8$, $\rho = 0$	0.3917	0.2265	0.1348	0.0822	0.0514	0.0318	0.0199	0.0124	0.0077	0.0049	0.0030	0.0019	0.0012
$O - LM$, $L = 8$, $\rho = 0$	0.3900	0.2253	0.1355	0.0814	0.0503	0.0317	0.0198	0.0124	0.0077	0.0049	0.0030	0.0019	0.0012
TALS, $L = 8$, $\rho = 0.8$	0.5920	0.3113	0.1803	0.1079	0.0666	0.0407	0.0251	0.0156	0.0100	0.0061	0.0039	0.0024	0.0015
$O - LM$, $L = 8$, $\rho = 0.8$	0.5777	0.3090	0.1814	0.1078	0.0664	0.0404	0.0251	0.0161	0.0098	0.0062	0.0039	0.0024	0.0015
TALS, $L = 7$, $\rho = 0.8$	1.5305	0.6519	0.3055	0.1623	0.1004	0.0570	0.0360	0.0221	0.0139	0.0091	0.0051	0.0034	0.0021
$O - LM$, $L = 7$, $\rho = 0.8$	1.4047	0.6228	0.2849	0.1770	0.0952	0.0580	0.0354	0.0221	0.0136	0.0087	0.0055	0.0034	0.0021

In the fourth example, the influence of design parameters (P, R) for the proposed receiver is studied. In the left subfigure of Figure 7, it can be seen that the BER decreases when P increases, which expounds the performance gain brought by the time diversity. It can also be seen from this subfigure that the BER increases as the number of data streams R increases. The impact of design parameters (P, R) on the NMSE performance is shown in the right subfigure of Figure 7. As expected, we can observe that the NMSE decreases linearly as a function of P, and increases as R increases. Hence, appropriate values for the design parameters P and R can be selected according to requirements of the system performance and transmission rate.

In the fifth example, we assume $M_S = 3$, $R = 4$, and $N = P = 8$ for our semi-blind receiver. The influence of the receive antenna was analyzed. We also compared the performance of our chosen coding tensor (OCCT) **B** with the random coding tensor (RCT) whose entries are circularly-symmetric Gaussian random variables. In Figure 8, it can be seen that both the BER and NMSE decrease when M_D increases, which expounds the performance gain brought by the receive diversity. We also observed from Figure 8 that OCCT has a better performance than RCT. Although OCCT is suboptimal, this choice has good symbol and channel identifiability properties, which is advantageous from a receiver design viewpoint.
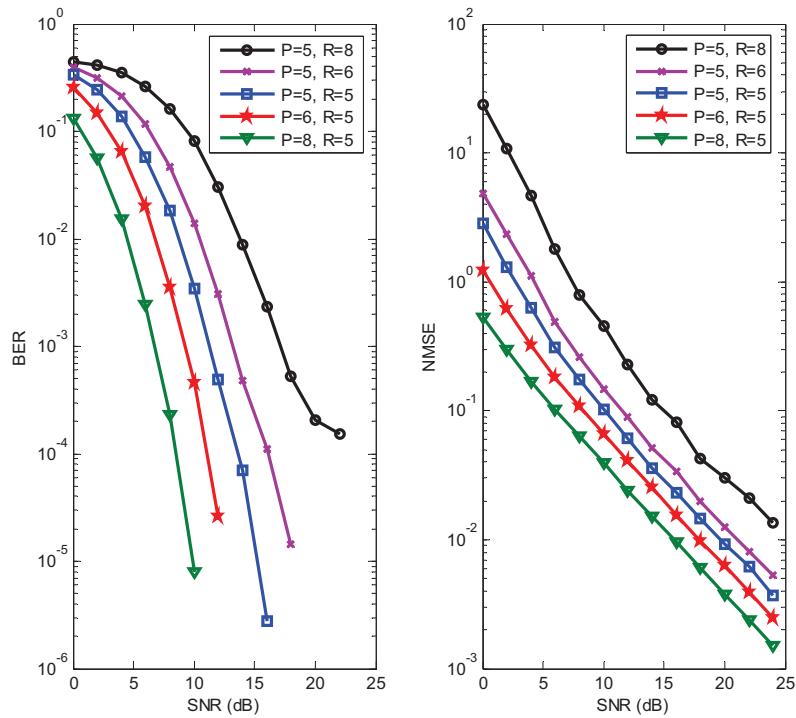
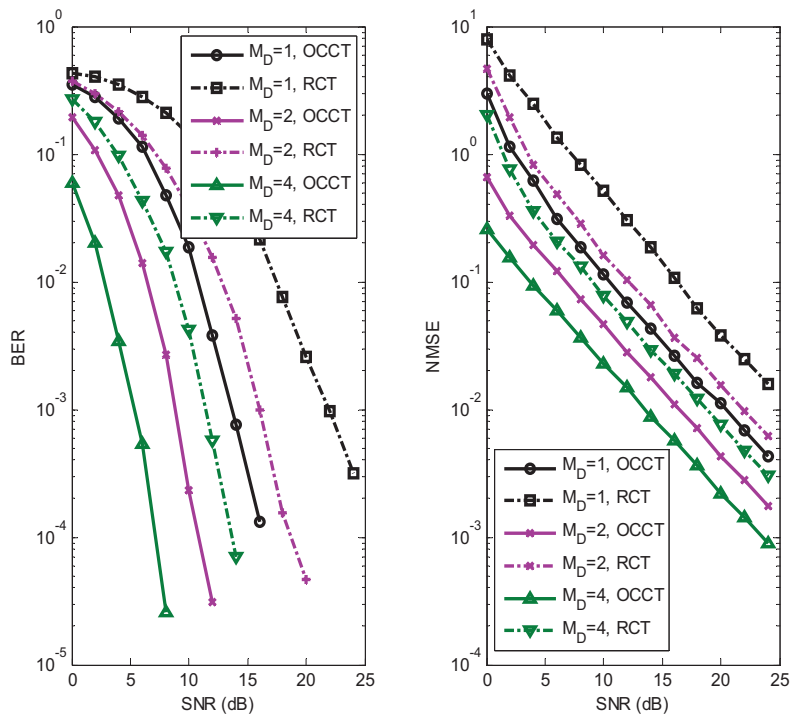**Figure 7.** BER and NMSE performance of the proposed receiver for different *P* and *R*.

Figure 8. Influence of the receive antenna and the coding tensor.

In the sixth example, we studied the estimation performance of two different transmission schemes for our semi-blind receiver. The default values of the system parameters were set to $M_D = 5$ and $N = 6$. In scheme 1, we assume $M_S = 2$, $R = 5$, and $P = 6$. Three different antenna-to-slot allocation matrices are given as follows:

$$\mathbf{Q}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \mathbf{Q}_2 = \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{bmatrix}, \quad \mathbf{Q}_3 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}$$

In scheme 2, we assume $M_S = 3$, $R = 3$, and $P = 7$. Three different antenna-to-slot allocation matrixes are given as follows:

$$\mathbf{Q}_4 = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \quad \mathbf{Q}_5 = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

The BER and NMSE performance of the proposed receiver for different schemes is shown in Figure 9. For scheme 1, the proposed receiver with $\mathbf{Q}_2$ has a better BER and NMSE performance than that of the proposed receiver with $\mathbf{Q}_1$. The reason is that the allocation matrix $\mathbf{Q}_2$ provides a higher transmit spatial diversity gain than the allocation matrix $\mathbf{Q}_1$. For the same reason, the allocation matrix $\mathbf{Q}_3$ outperforms $\mathbf{Q}_2$, and the allocation matrix $\mathbf{Q}_5$ outperforms $\mathbf{Q}_4$. We also observe in Figure 9 that scheme 2 has a better BER and NMSE performance than scheme 1. The reason is that scheme 2 can provide a higher coding diversity than scheme 1. It is worth noting that scheme 1 has higher spectral efficiency compared with scheme 2. The transmission rates for scheme 1 and scheme 2 are about 5/6 and 3/7 (data symbols per symbol period), respectively. In summary, a desired tradeoff between estimation performance and transmission rate can be obtained by designing a suitable scheme.
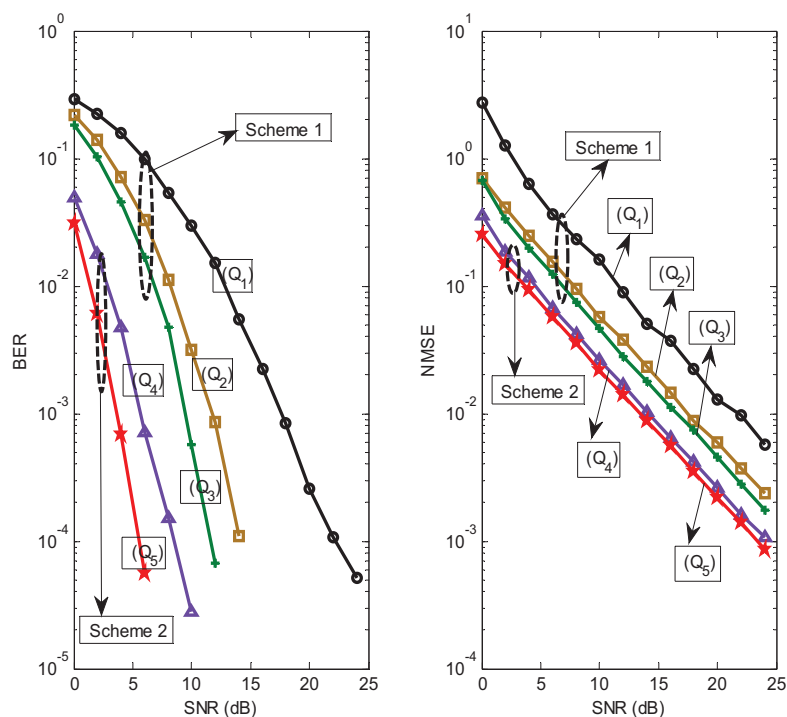

Figure 9. Performances of the proposed receiver for different schemes.

In final example, the multi-user massive MIMO system with a fully-connected hybrid precoding architecture was considered, where $M_S = 48$, $M \times M_D = 6 \times 6$, and $L_m = 2$ for all $m = 1, \ldots, M$. The carrier frequency of this system is set as 28 GHz [37], and $d = \lambda/2$. We assume that AoAs/AoDs are uniformly distributed in $[0, 2\pi]$. For the considering multi-user massive MIMO system, we also evaluate the estimation performance of the proposed receiver in terms of BER and NMSE of channel estimation. It can be seen from Figures 10 and 11 that the BER and NMSE of the proposed semi-blind receiver decrease as P and N increase, and increase as R increases. The increase of P will reduce the transmission rate, but the increase or decrease of N has no effect on the transmission rate. That means that we can improve the estimation performance of the proposed semi-blind receiver by increasing N if the channel is constant over a long time interval before changing to another realization. We also observed from Figures 10 and 11 that the proposed semi-blind receiver still has a good performance for joint symbol and channel estimation even in a shorter length of code and information symbol, and a larger number of data streams, i.e., $P = 24$, $N = 6$, and $R = 12$.
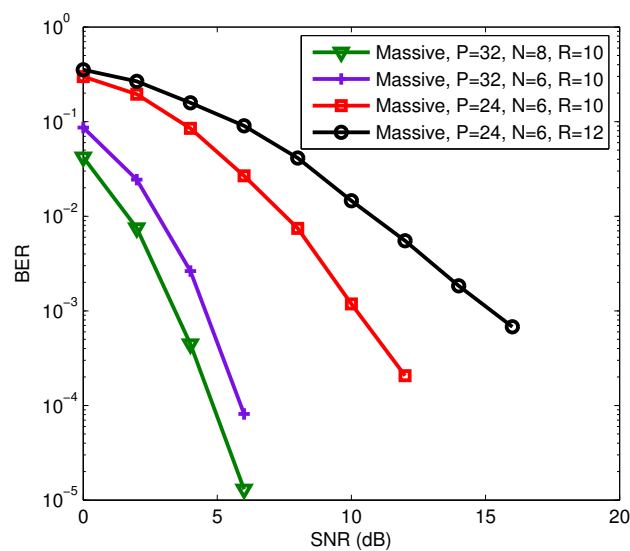

Figure 10. BER performance of the proposed receiver for the multi-user massive multiple-input multiple-output (MIMO) system.
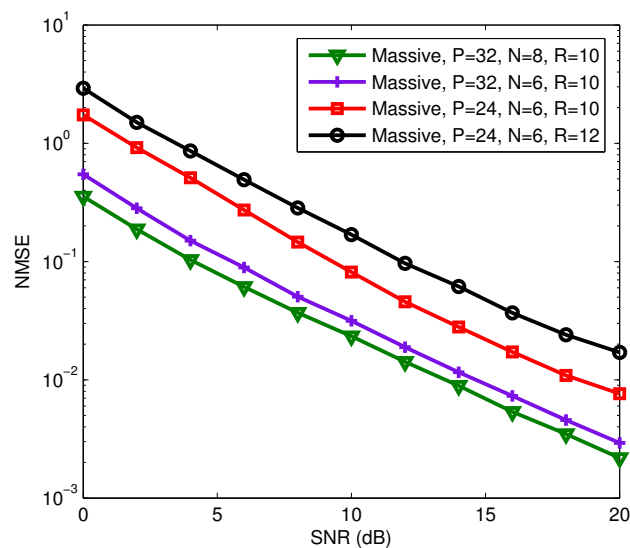

Figure 11. NMSE performance of the proposed receiver for the multi-user massive MIMO system.

7. Conclusions

We have developed a robust semi-blind receiver combined with the Tucker-2 model in multiple-antenna systems. The proposed receiver could jointly estimate the information symbol and channel parameters. Compared with existing semi-blind receivers, the proposed one gave better estimation performance, and had a higher spectral efficiency Moreover, the proposed semi-blind receiver was also applicable to multi-user massive MIMO systems. Perspectives of this work include an extension to relay-assisted massive MIMO systems by applying the antenna allocation matrix at the relays. Since both the source-relay and the relay-destination channel matrices have low-rank property, new identifiability conditions and efficient fitting algorithms will be deduced and developed, respectively. Another perspective considers extending the proposed robust semi-blind receiver into mmwave MIMO systems for joint channel parameter estimation, which includes AOAs, fading coefficients and time delays [38,39].

References

1. Jin, S.; McKay, M.R.; Wong, K.K.; Li, X. Low-SNR capacity of multiple-antenna systems with statistical channel-state information. *IEEE Trans. Veh. Technol.* **2010**, *59*, 2874–2884. [CrossRef]
2. Shafi, M.; Molisch, A.F.; Smith, P.J.; Haustein, T.; Zhu, P.; De Silva, P.; Tufvesson, F.; Benjebbour, A.; Wunder, G. 5G: A tutorial overview of standards, trials, challenges, deployment, and practice. *IEEE J. Sel. Areas Commun.* **2017**, *35*, 1201–1221. [CrossRef]
3. Collins, A.; Polyanskiy, Y. Coherent multiple-antenna block-fading channels at finite blocklength. *IEEE Trans. Inf. Theory* **2019**, *65*, 380–405. [CrossRef]
4. Lahat, D.; Adali, T.; Jutten, C. Multimodal data fusion: An overview of methods, challenges, and prospects. *Proc. IEEE* **2015**, *103*, 1449–1477. [CrossRef]
5. Sidiropoulos, N.D.; De Lathauwer, L.; Fu, X.; Huang, K.; Papalexakis, E.E.; Faloutsos, C. Tensor decomposition for signal processing and machine learning. *IEEE Trans. Signal Process.* **2017**, *65*, 3551–3582. [CrossRef]
6. Freitas, W.d.C.; Favier, G.; de Almeida, A.L. Tensor-Based Joint Channel and Symbol Estimation for Two-Way MIMO Relaying Systems. *IEEE Signal Process. Lett.* **2019**, *26*, 227–231. [CrossRef]
7. Harshman, R.A. Foundations of the PARAFAC Procedure: Models and Conditions for an "Explanatory" Multimodal Factor Analysis. *UCLA Working Pap. Phon.* **1970**, *16*, 1–84.
8. Sidiropoulos, N.D.; Budampati, R.S. Khatri–Rao space-time codes. *IEEE Trans. Signal Process.* **2002**, *50*, 2396–2407. [CrossRef]
9. Liu, K.; Da Costa, J.P.C.; So, H.C.; De Almeida, A.L. Semi-blind receivers for joint symbol and channel estimation in space-time-frequency MIMO-OFDM systems. *IEEE Trans. Signal Process.* **2013**, *61*, 5444–5457. [CrossRef]
10. Rong, Y.; Khandaker, M.R.; Xiang, Y. Channel estimation of dual-hop MIMO relay system via parallel factor analysis. *IEEE Trans. Wirel. Commun.* **2012**, *11*, 2224–2233. [CrossRef]
11. Du, J.; Yuan, C.; Zhang, J. Low complexity PARAFAC-based channel estimation for non-regenerative MIMO relay systems. *IET Commun.* **2014**, *8*, 2193–2199. [CrossRef]

12. Du, J.; Yuan, C.; Zhang, J. Semi-blind parallel factor based receiver for joint symbol and channel estimation in amplify-and-forward multiple-input multiple-output relay systems. *IET Commun.* **2015**, *9*, 737–744. [CrossRef]

13. Ximenes, L.R.; Favier, G.; de Almeida, A.L. Semi-blind receivers for non-regenerative cooperative MIMO communications based on nested PARAFAC modeling. *IEEE Trans. Signal Process.* **2015**, *63*, 4985–4998. [CrossRef]

14. Zhou, Z.; Fang, J.; Yang, L.; Li, H.; Chen, Z.; Li, S. Channel estimation for millimeter-wave multiuser MIMO systems via PARAFAC decomposition. *IEEE Trans. Wirel. Commun.* **2016**, *15*, 7501–7516. [CrossRef]

15. Zhou, Z.; Fang, J.; Yang, L.; Li, H.; Chen, Z.; Blum, R.S. Low-rank tensor decomposition-aided channel estimation for millimeter wave MIMO-OFDM systems. *IEEE J. Sel. Areas Commun.* **2017**, *35*, 1524–1538. [CrossRef]

16. Wei, X.; Peng, W.; Chen, D.; Ng, D.W.K.; Jiang, T. Joint Channel Parameter Estimation in Multi-cell Massive MIMO System. *IEEE Trans. Commun.* **2019**. [CrossRef]

17. Comon, P.; Luciani, X.; De Almeida, A.L. Tensor decompositions, alternating least squares and other tales. *J. Chemom.* **2009**, *23*, 393–405. [CrossRef]

18. Tomasi, G.; Bro, R. A comparison of algorithms for fitting the PARAFAC model. *Comput. Stat. Data Anal.* **2006**, *50*, 1700–1734. [CrossRef]

19. Nion, D.; De Lathauwer, L. A block component model-based blind DS-CDMA receiver. *IEEE Trans. Signal Process.* **2008**, *56*, 5567–5579. [CrossRef]

20. De Almeida, A.L.; Favier, G.; Ximenes, L.R. Space-time-frequency (STF) MIMO communication systems with blind receiver based on a generalized PARATUCK2 model. *IEEE Trans. Signal Process.* **2013**, *61*, 1895–1909. [CrossRef]

21. De Almeida, A.L.; Favier, G.; Mota, J.C. Space–time spreading–multiplexing for MIMO wireless communication systems using the PARATUCK-2 tensor model. *Signal Process.* **2009**, *89*, 2103–2116. [CrossRef]

22. Favier, G.; Da Costa, M.N.; De Almeida, A.L.; Romano, J.M.T. Tensor space–time (TST) coding for MIMO wireless communication systems. *Signal Process.* **2012**, *92*, 1079–1092. [CrossRef]

23. Favier, G.; de Almeida, A.L. Tensor space-time-frequency coding with semi-blind receivers for MIMO wireless communication systems. *IEEE Trans. Signal Process.* **2014**, *62*, 5987–6002. [CrossRef]

24. Da Costa, M.N.; Favier, G.; Romano, J.M.T. Tensor modelling of MIMO communication systems with performance analysis and Kronecker receivers. *Signal Process.* **2018**, *145*, 304–316. [CrossRef]

25. Tucker, L.R. Some mathematical notes on three-mode factor analysis. *Psychometrika* **1966**, *31*, 279–311. [CrossRef] [PubMed]

26. Favier, G.; de Almeida, A.L. Overview of constrained PARAFAC models. *EURASIP J. Adv. Signal Process.* **2014**, *2014*, 142. [CrossRef]

27. Du, J.; Yuan, C.; Hu, Z.; Lin, H. A novel tensor-based receiver for joint symbol and channel estimation in two-hop cooperative MIMO relay systems. *IEEE Commun. Lett.* **2015**, *19*, 1961–1964. [CrossRef]

28. Chen, Y.; Han, D.; Qi, L. New ALS methods with extrapolating search directions and optimal step size for complex-valued tensor decompositions. *IEEE Trans. Signal Process.* **2011**, *59*, 5888–5898. [CrossRef]

29. Van Loan, C.F.; Pitsianis, N. Approximation with Kronecker products. In *Linear Algebra for Large Scale and Real-Time Applications*; Springer: Dordrecht, The Netherlands, 1993; pp. 293–314.

30. Du, J.; Tian, P.; Lin, H. Tucker-2 Model Based Scheme for Joint Signal Detection and Channel Estimation in MIMO Systems. *J. Beijing Univ. Posts Telecommun.* **2016**, *39*, 6–10.

31. Jain, P.; Meka, R.; Dhillon, I.S. Guaranteed rank minimization via singular value projection. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 6–9 December 2010; pp. 937–945.

32. Shen, W.; Dai, L.; Shim, B.; Mumtaz, S.; Wang, Z. Joint CSIT acquisition based on low-rank matrix completion for FDD massive MIMO systems. *IEEE Commun. Lett.* **2015**, *19*, 2178–2181. [CrossRef]

33. Alkhateeb, A.; El Ayach, O.; Leus, G.; Heath, R.W. Hybrid precoding for millimeter wave cellular systems with partial channel knowledge. In Proceedings of the 2013 Information Theory and Applications Workshop (ITA), San Diego, CA, USA, 10–15 February 2013; pp. 1–5.

34. Dai, L.; Gao, X.; Quan, J.; Han, S.; Chih-Lin, I. Near-optimal hybrid analog and digital precoding for downlink mmWave massive MIMO systems. In Proceedings of the 2015 IEEE International Conference on Communications (ICC), London, UK, 8–12 June 2015; pp. 1334–1339.

35. Shiu, D.S.; Foschini, G.J.; Gans, M.J.; Kahn, J.M. Fading correlation and its effect on the capacity of multielement antenna systems. *IEEE Trans. Commun.* **2000**, *48*, 502–513. [CrossRef]

36. Nion, D.; De Lathauwer, L. An enhanced line search scheme for complex-valued tensor decompositions. Application in DS-CDMA. *Signal Process.* **2008**, *88*, 749–755. [CrossRef]

37. Alkhateeb, A.; El Ayach, O.; Leus, G.; Heath, R.W. Channel estimation and hybrid precoding for millimeter wave cellular systems. *IEEE J. Sel. Top. Signal Process.* **2014**, *8*, 831–846. [CrossRef]

38. Hu, C.; Dai, L.; Mir, T.; Gao, Z.; Fang, J. Super-resolution channel estimation for mmWave massive MIMO with hybrid precoding. *IEEE Trans. Veh. Technol.* **2018**, *67*, 8954–8958. [CrossRef]

39. Srivastava, S.; Mishra, A.; Rajoriya, A.; Jagannatham, A.K.; Ascheid, G. Quasi-Static and Time-Selective Channel Estimation for Block-Sparse Millimeter Wave Hybrid MIMO Systems: Sparse Bayesian Learning (SBL) Based Approaches. *IEEE Trans. Signal Process.* **2019**, *67*, 1251–1266. [CrossRef]