Semi-blind joint channel estimation and data detection on sphere manifold for MIMO with high-order QAM signaling

Semi-Blind Joint Channel Estimation and Data Detection on Sphere Manifold for MIMO with high-order QAM Signaling

Xia Hong, Junbin Gao and Sheng Chen

***Abstract*—A low-complexity semi-blind scheme is proposed for joint channel estimation and data detection on sphere manifold for multiple-input multiple-output (MIMO) systems with high-order quadrature amplitude modulation signaling. Specifically, the optimal channel estimator is expressed in the least squares form in terms of the received signals and unknown transmitted data, and by splitting the channel and transmitted data into their real parts and imaginary parts, the data detection becomes a problem defined on a scaled sphere manifold in the real domain. Our semi-blind algorithm consists of three stages: (i) a few training symbols are employed to provide a rough initial MIMO channel estimate which in turn yields the initial zero-forcing (ZF) estimate of data samples; (ii) the Riemannian conjugate gradient algorithm is used to estimate the data samples in real domain, and the detected data samples are used to estimate the final MIMO channel matrix; and (iii) the final ZF data detection is carried out based on the final MIMO channel estimate. In particular, we present the first order Riemannian geometry of the sphere manifold which is utilized in the Riemannian conjugate gradient algorithm for solving (ii). Simulation results are employed to demonstrate the effectiveness of the proposed approach.**

***Index Terms*—Multiple-input multiple-output, high-order quadrature amplitude modulation, channel estimation, data detection, sphere manifold, Riemannian conjugate gradient algorithm**

I. Introduction

Multiple-input multiple-output (MIMO) technology is capable of dramatically enhancing system's reliability and capacity [1]–[4]. In order to fully utilize the MIMO capacity, accurate channel estimation is necessary. MIMO channel estimation methods can be classified into three categories: training-based methods, blind methods, and semi-blind methods. For pure training-based schemes, a large training overhead is necessary in order to obtain a reliable MIMO channel estimate, which reduces the system's effective throughput considerably. Blind methods do not require any training symbols and are capable of maintaining high system throughput at the expense of high computational complexity. In addition, pure blind methods suffer from an intractable ambiguity problem in MIMO estimation and detection. Semi-blind schemes are attractive for practical implementation, since they are capable of resolving the ambiguity problem otherwise unsolvable by blind methods as well as require less computational complexity than blind methods and fewer training symbols than training-based methods.

Joint blind channel estimation and data detection has been proposed based on the iterative least squares (LS) with projection [5]–[7]. In this approach, the channel and data are estimated iteratively but the convergence of the iterative process depends on the initialization of the channel model. A pure blind approach is often computationally prohibitive, particularly for high-dimensional MIMO. Moreover, unlike the usual phase ambiguity in blindly detected data symbols, which may be resolved by differential encoding, in blind joint MIMO channel estimation and data detection, permutation ambiguity corresponding to reordering the detected transmitted data and estimated channel matrix columns cannot be easily resolved. One way of solving this permutation ambiguity is to employ a few pilot training symbols, leading to semi-blind schemes. In the context of MIMO, semi-blind schemes have been developed [8]–[15], which use a few training symbols to provide the initial MIMO channel estimation and then exchange the information between the data detector and the decision-directed (DD) channel estimator iteratively. Typically, 5 to 10 iterations are required for the iterative MIMO data detector and DD channel estimator to converge.

Given the MIMO channel matrix, the optimal data detector is the maximum likelihood (ML) detector. Therefore, given the initial rough training-based channel estimate, the optimal semi-blind scheme, in terms of achievable performance, is based on the iterative procedure between the ML data detector and DD channel estimator [12]–[14]. However, such an optimal procedure is impractical to implement, owing to the prohibitively high complexity of the ML detector, particularly for high-dimensional MIMO with spectral efficient high-order signaling schemes. For example, for 6 by 6 MIMO with 64-quadrature amplitude modulation (QAM), the ML detector needs to search the candidate data set of the size $64^6 \approx 7 \times 10^{10}$, which is difficult if not impossible to realize. Therefore, low-complexity suboptimal detection schemes must be adopted in practice [15]. A low-complexity semi-blind scheme suitable for MIMO systems with high-order QAM signaling is the semi-blind iterative zero-forcing (ZF) data detection and DD LS channel estimation, which achieves good performance. In this paper, we advocate a low-complexity semi-blind alternative for high-dimensional MIMO with high-order QAM signaling based on sphere manifold.

Recent years have witnessed great development in Rie-

X. Hong is with Department of Computer Science, School of Mathematical and Physical Sciences, University of Reading, Reading, RG6 6AY, UK (E-mail: x.hong@reading.ac.uk).

J. Gao is with Discipline of Business Analytics, University of Sydney Business School, University of Sydney, Camperdown NSW 2006, Australia (E-mail: junbin.gao@sydney.edu.au).

S. Chen is with School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K. (E-mail: sqc@ecs.soton.ac.uk), and also with King Abdulaziz University, Jeddah 21589, Saudi Arabia.

TABLE I
NOTATIONS FOR SPHERE MANIFOLD

$\{\mathbb{S}^{M'\cdot N-1}, g\}$	Sphere manifold for parameter matrix $\boldsymbol{X}$ and the inner product of the manifold
$T_{\boldsymbol{X}}\mathbb{S}^{M'\cdot N-1}$	Tangent space of the sphere manifold
$\boldsymbol{U_X}, \boldsymbol{V_X}$	Tangent matrices at $\boldsymbol{X}$
$\text{Proj}_{\boldsymbol{X}}(\boldsymbol{Z})$	Orthogonal projector from matrix $\boldsymbol{Z}$ in ambient space onto the tangent space at $\boldsymbol{X}$
$\text{grad}F(\boldsymbol{X})$	Riemannian gradient of $F(\boldsymbol{X})$ on the manifold $\mathbb{S}^{M'\cdot N-1}$
$\text{Grad}F(\boldsymbol{X})$	Classical gradient of $F(\boldsymbol{X})$ as seen in Euclidean space
$\text{Exp}_{\boldsymbol{X}}$	Retraction mapping
$\mathcal{T}_{\boldsymbol{X}_{k+1}\leftarrow\boldsymbol{X}_k}(\boldsymbol{U}_{\boldsymbol{X}_k})$	Matrix transport

mannian optimization algorithms on many types of matrix manifolds, such as Stiefel manifold, Grassmann manifold and the manifold of positive definite matrices, see for example [16]. Since Riemannian optimization is directly based on the curved manifolds, one can eliminate those constraints such as orthogonality to obtain an unconstrained optimization problem that, by construction, will only use feasible points. This allows one to incorporate Riemannian geometry in the resulting optimization problems, thus producing far more accurate numerical results. The recent successful applications of Riemannian optimization in machine learning, computer vision and data mining, citing a few, include fixed low rank optimization [17], Riemannian dictionary learning [18]–[21], and computer vision tasks [22].

Against this background, in this paper we propose a low-complexity semi-blind scheme for joint channel estimation and data detection for MIMO systems with spectral efficient high-order QAM signaling. By representing the LS channel estimator in terms of the received signals and unknown transmitted data, the variables of the optimization cost function for blind data detection become the transmitted symbols alone which must satisfy the constraint of the known modulation scheme whose symbol constellation is symmetric in real and imaginary parts. Furthermore, we formulate our blind data detection problem into an equivalent one by splitting the channel and data into their real parts and imaginary parts, so that under the reasonable assumption of a sufficient long data block, the constraint on the data samples becomes a scaled sphere manifold in the real domain. This allows us to apply Riemannian optimization technique to solve the resulting data detection problem. More specifically, our proposed algorithm consists of three parts: i) a few training symbols are employed to yield an initial LS MIMO channel estimate in order to provide a rough ZF detected data samples for initializing the data detection; ii) the Riemannian conjugate gradient algorithm is used for solving the corresponding blind data detection problem to estimate the date samples, and the detected data symbols are then used to provide the final DD LS MIMO channel estimate; and iii) the data samples are re-estimated using the ZF detector based on the final MIMO channel estimate. In particular, in terms of stage ii), the first order Riemannian geometry of the sphere manifold is presented and then utilized in the Riemannian trust-region algorithm on the sphere manifold.

The remaining paper is organized as follows. Section II introduces the concept of sphere manifold and necessary notations for Riemannian optimization. Section III details the proposed semi-blind scheme for joint MIMO channel estimation and data detection on sphere manifold, including novel signal detection using Riemannian conjugate gradient algorithm. Simulation results are provided in Section IV to demonstrate the effectiveness of the proposed approach. Our conclusions are drawn in Section V.

Throughout our discussions, a complex-valued (CV) number $x \in \mathbb{C}$ is represented by $x = x^{(\text{R})} + \text{j}\cdot x^{(\text{I})}$, where $\text{j} = \sqrt{-1}$, while $x^{(\text{R})}$ and $x^{(\text{I})}$ are the real and imaginary parts of x, respectively. The transpose and conjugate transpose operators are denoted by $(\ )^{\text{T}}$ and $(\ )^{\text{H}}$, respectively, and $(\ )^*$ denotes the conjugate operation, while $(\ )^{-1}$ stands for the inverse operation and the expectation operator is given by $\text{E}\{\ \}$. Matrix trace operator is denoted by trace$(\ )$, and $\boldsymbol{I}$ denotes the identity matrix with an appropriate dimension, while $\boldsymbol{B}[a:b;:]$ is the matrix consists of the ath to bth rows of $\boldsymbol{B}$.

II. SPHERE MANIFOLD

This section briefly introduces the concept of sphere manifold and the necessary ingredients used in the retraction based framework of Riemannian optimization. As a reference, the main notations on Riemannian geometry on sphere manifold in this section is summarized in Table I. We refer the readers to [16] for the general concepts of manifolds.

The sphere manifold is the set of unit Frobenius norm matrices of size $M' \times N$, denoted as

$$\mathbb{S}^{M'\cdot N-1} = \big\{\boldsymbol{X} \in \mathbb{R}^{M'\cdot N} : \|\boldsymbol{X}\|_F = 1\big\}. \tag{1}$$

It is endowed with a Riemannian manifold structure by considering it as a Riemannian submanifold of the embedding Euclidean space $\mathbb{R}^{M'\cdot N}$ endowed with the usual inner product

$$g(\boldsymbol{U_X}, \boldsymbol{V_X}) = \text{trace}\big(\boldsymbol{U}_{\boldsymbol{X}}^{\text{T}}\boldsymbol{V_X}\big), \tag{2}$$

where $\boldsymbol{U_X}, \boldsymbol{V_X} \in T_{\boldsymbol{X}}\mathbb{S}^{M'\cdot N-1} \subset \mathbb{R}^{M'\cdot N}$ are tangent 'vectors' or matrices[1] to $\mathbb{S}^{M'\cdot N-1}$ at $\boldsymbol{X}$. The inner product on $\mathbb{S}^{M'\cdot N-1}$ determines the geometry such as distance, angle, curvature on $\mathbb{S}^{M'\cdot N-1}$. Note that the tangent space $T_{\boldsymbol{X}}\mathbb{S}^{M'\cdot N-1}$ at element $\boldsymbol{X}$ can be described by

$$T_{\boldsymbol{X}}\mathbb{S}^{M'\cdot N-1} = \big\{\boldsymbol{U_X} : \text{trace}\big(\boldsymbol{U}_{\boldsymbol{X}}^{\text{T}}\boldsymbol{X}\big) = 0\big\}. \tag{3}$$

Riemannian gradient: Let the Riemannian gradient of a scalar function $F(\boldsymbol{X})$ on $\mathbb{S}^{M'\cdot N-1}$ be denoted by $\text{grad}F(\boldsymbol{X})$,

[1]Here we change the notion tangent 'vector' in Riemannian geometry to tangent 'matrix', as it is in the same matrix form as seen in classical Euclidean geometry. Similarly, the notion 'vector' in Riemannian geometry is changed to 'matrix'.

and its classical gradient as seen in the Euclidean space as Grad$F(\boldsymbol{X})$. Then we have

$$\text{grad}F(\boldsymbol{X}) = \text{Proj}_{\boldsymbol{X}}\big(\text{Grad}F(\boldsymbol{X})\big), \tag{4}$$

where $\text{Proj}_{\boldsymbol{X}}(\boldsymbol{Z})$ is the orthogonal projection onto the tangent space, which can be computed as

$$\text{Proj}_{\boldsymbol{X}}(\boldsymbol{Z}) = \boldsymbol{Z} - \text{trace}\big(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Z}\big)\boldsymbol{X} \tag{5}$$

in which $\boldsymbol{Z}$ represents a matrix in the ambient space.

Retraction mapping: An important concept in the recent retraction-based framework of Riemannian optimization is the retraction mapping, see Section 4.1 of [16]. The exponential map $\text{Exp}_{\boldsymbol{X}}$, defined by

$$\begin{aligned}\text{Exp}_{\boldsymbol{X}}\big(\alpha\boldsymbol{U}_{\boldsymbol{X}}\big) =& \cos\big(\|\alpha\boldsymbol{U}_{\boldsymbol{X}}\|_F\big)\boldsymbol{X} \\ &+ \frac{\sin(\|\alpha\boldsymbol{U}_{\boldsymbol{X}}\|_F)}{\|\boldsymbol{U}_{\boldsymbol{X}}\|_F}\boldsymbol{U}_{\boldsymbol{X}},\end{aligned} \tag{6}$$

is the canonical choice for the retraction mapping, where the scalar α is a chosen step size. The retraction mapping is used to locate the next iterate on the manifold along a specified tangent matrix, such as a search direction in the line search of Newton's algorithm or the suboptimal tangent direction in the trust-region algorithm, see Chapter 7 of [16]. For example, the line search algorithm is simply given by

$$\boldsymbol{X}_{k+1} = \text{Exp}_{\boldsymbol{X}_k}\big(\alpha_k\boldsymbol{U}_{\boldsymbol{X}_K}\big) \tag{7}$$

in which the search direction $\boldsymbol{U}_{\boldsymbol{X}_K} \in T_{\boldsymbol{X}_k}\mathbb{S}^{M'\cdot N-1}$ and α_k is a chosen step size at iteration step k.

Matrix Transport: In Riemannian optimization algorithms, the second derivatives can be approximated by comparing the first-order information (tangent matrices) at distinct points on the manifold. The notion of matrix transport $\mathcal{T}_{\boldsymbol{X}_{k+1}\leftarrow\boldsymbol{X}_k}\big(\boldsymbol{U}_{\boldsymbol{X}_k}\big)$ on a manifold, roughly speaking, specifies how to transport a tangent matrix $\boldsymbol{U}_{\boldsymbol{X}_k}$ from a point $\boldsymbol{X}_k$ to another point $\boldsymbol{X}_{k+1}$ on the manifold. The matrix transport for the sphere manifold is calculated as

$$\mathcal{T}_{\boldsymbol{X}_{k+1}\leftarrow\boldsymbol{X}_k}\big(\boldsymbol{U}_{\boldsymbol{X}_k}\big) = \text{Proj}_{\boldsymbol{X}_{k+1}}\big(\boldsymbol{U}_{\boldsymbol{X}_k}\big). \tag{8}$$

III. The Proposed Semi-blind Scheme for MIMO on Sphere Manifold

In this section, we begin by introducing the MIMO system model and the existing low-complexity semi-blind iterative ZF data detection and DD LS channel estimation scheme for high-dimensional MIMO with high-order QAM signaling [8]–[15]. Then our alternative low-complexity semi-blind scheme for MIMO on sphere manifold is detailed .

A. System Model and Existing Low-Complexity Semi-blind Scheme

Consider a MIMO system with n_T transmit antennas and n_R receive antennas. It is assumed that the channel coherence bandwidth is larger than the transmitted signal bandwidth so that the channel can be considered as narrowband or flat fading. The sampled received signal vector $\boldsymbol{y}(k) = \big[y_1(k)\ y_2(k)\cdots y_{n_R}(k)\big]^{\mathrm{T}}$ at symbol index k is given by the well-known MIMO model

$$\boldsymbol{y}(k) = \boldsymbol{H}\boldsymbol{s}(k) + \boldsymbol{\varepsilon}(k), \tag{9}$$

where $\boldsymbol{s}(k) = \big[s_1(k)\ s_2(k)\cdots s_{n_T}(k)\big]^{\mathrm{T}}$ and $s_m(k)$ is the kth transmitted symbol from transmit antenna m with $\mathrm{E}\big\{|s_m(k)|^2\big\} = \mathrm{E}\big\{s_m(k)s_m^*(k)\big\} = \sigma_s^2$ for $1 \le m \le n_T$, and the additive white Gaussian noise (AWGN) vector $\boldsymbol{\varepsilon}(k) = \big[\varepsilon_1(k)\ \varepsilon_2(k)\cdots\varepsilon_{n_R}(k)\big]^{\mathrm{T}}$ with $\mathrm{E}\big\{|\varepsilon_l(k)|^2\big\} = 2\sigma_\varepsilon^2$ for $1 \le l \le n_R$, while $\boldsymbol{H} \in \mathbb{C}^{n_R\cdot n_T}$ denotes the MIMO channel matrix whose lth row and mth column element $h_{l,m}$ is the channel coefficient connecting transmit antenna m to receive antenna l. Furthermore, the channel is assumed to be quasi-static. Specifically, over the transmission period of a data frame, which contains N_f symbols, all the entries of $\boldsymbol{H}$ remain unchanged. From frame to frame, $h_{l,m}$ obeys the CV zero-mean Gaussian distribution $\mathcal{CN}(0,1)$ with a variance of $1/2$ per dimension. The modulation scheme adopted is L-ary QAM (L-QAM), which is defined by the symbol constellation set

$$\mathcal{S} = \big\{s^{(i)}, 1 \le i \le L\big\}. \tag{10}$$

Thus, the transmitted data $s_m(k) \in \mathcal{S}$, $\forall m, k$.

By collecting the received data, the transmitted data and the channel noise over a frame of $1 \le k \le N_f$ respectively as

$$\boldsymbol{Y} = \big[\boldsymbol{y}(1)\ \boldsymbol{y}(2)\cdots\boldsymbol{y}(N_f)\big] \in \mathbb{C}^{n_R\cdot N_f}, \tag{11}$$

$$\boldsymbol{S} = \big[\boldsymbol{s}(1)\ \boldsymbol{s}(2)\cdots\boldsymbol{s}(N_f)\big] \in \mathbb{C}^{n_T\cdot N_f}, \tag{12}$$

$$\boldsymbol{E} = \big[\boldsymbol{\varepsilon}(1)\ \boldsymbol{\varepsilon}(2)\cdots\boldsymbol{\varepsilon}(N_f)\big] \in \mathbb{C}^{n_R\cdot N_f}, \tag{13}$$

the overall MIMO system can be represented by

$$\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{S} + \boldsymbol{E}. \tag{14}$$

If the channel $\boldsymbol{H}$ is known, the low-complexity ZF data detection solution is given by

$$\widetilde{\boldsymbol{S}}_{\text{ZF}} = \boldsymbol{W}_{\text{ZF}}\boldsymbol{Y}, \tag{15}$$

with the detector weight matrix given by

$$\boldsymbol{W}_{\text{ZF}} = \big(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\big)^{-1}\boldsymbol{H}^{\mathrm{H}}. \tag{16}$$

Every element of $\widetilde{\boldsymbol{S}}_{\text{ZF}}$ is then quantized to the nearest symbol point in $\mathcal{S}$ to yield the hard-decision detected symbol matrix $\widehat{\boldsymbol{S}}_{\text{ZF}}$. On the other hand, if the data $\boldsymbol{S}$ is known, the LS channel estimate is readily given by

$$\widehat{\boldsymbol{H}}_{\text{LS}} = \boldsymbol{Y}\boldsymbol{S}^{\mathrm{H}}\big(\boldsymbol{S}\boldsymbol{S}^{\mathrm{H}}\big)^{-1}. \tag{17}$$

Blind joint channel estimation and data detection is to find $\boldsymbol{H}$ and $\boldsymbol{S}$ purely based on the received $\boldsymbol{Y}$, and this task suffers from the well-known scaling and permutation ambiguity. For example, (14) can also be represented as

$$\boldsymbol{Y} = \boldsymbol{H}^*\boldsymbol{S}^* + \boldsymbol{E}, \tag{18}$$

where $\boldsymbol{H}^* = \boldsymbol{H}\boldsymbol{T}$ and $\boldsymbol{S}^* = \boldsymbol{T}^{\mathrm{H}}\boldsymbol{S}$, and $\boldsymbol{T}$ is the unitary $n_T\times n_T$ permutation and scaling matrix with only one nonzero element in each row and in each column. In order to resolve this ambiguity problem, a semi-blind approach is desirable in which a few training symbols of $N_t \ll N_f$ are used to

provide an initial estimate of the channel $\boldsymbol{H}$. For convenience, collecting the N_t training data together as

$$\boldsymbol{S}_t = \big[\boldsymbol{s}(1)\ \boldsymbol{s}(2)\cdots\boldsymbol{s}(N_t)\big] \in \mathbb{C}^{n_T\cdot N_t}, \tag{19}$$

$$\boldsymbol{Y}_t = \big[\boldsymbol{y}(1)\ \boldsymbol{y}(2)\cdots\boldsymbol{y}(N_t)\big] \in \mathbb{C}^{n_R\cdot N_t}. \tag{20}$$

The LS channel estimate based on the initial training data $\{\boldsymbol{S}_t, \boldsymbol{Y}_t\}$ is readily given by

$$\widehat{\boldsymbol{H}}_{\rm LS}^{(\rm t)} = \boldsymbol{Y}_t\boldsymbol{S}_t^{\rm H}\big(\boldsymbol{S}_t\boldsymbol{S}_t^{\rm H}\big)^{-1}. \tag{21}$$

In order to maintain the system throughput, N_t should be as small as possible. On the other hand, to guarantee $\boldsymbol{S}_t\boldsymbol{S}_t^{\rm H}$ having a full rank, it is necessary that $N_t \geq n_T$.

Existing Practical Semi-Blind Scheme: In the existing semi-blind scheme of iterative ZF data detection and DD LS channel estimation (LSCE), e.g., [15], a minimum number of training data with $N_t = n_T$ or very close to it, are employed to provide the initial training based LS channel estimate of (21). With this rough LS channel estimate, the ZF data detection is carried out, and the detected data matrix is employed in the DD LSCE. The improved channel estimate is then used for the next ZF data detection. A few iterations, typically 5 to 10, between the ZF data detector and the DD LS channel estimator are sufficient for this semi-blind scheme to converge to the ZF detection solution associated with the perfect channel state information (CSI) $\boldsymbol{H}$, provided that the MIMO system's signal to noise ratio (SNR) is over certain threshold. This low-complexity semi-blind scheme offers a practical means suitable for implementation in MIMO systems with high-order QAM signaling, and it will be used as a benchmark for comparison with our proposed semi-blind scheme based on sphere manifold.

B. Proposed Semi-blind Scheme

Our alternative semi-blind scheme also offers a low-complexity solution suitable for MIMO systems with high-order QAM signaling by exploiting Riemannian optimization on sphere manifold. We begin by noting that the probability density function of the received signal matrix $\boldsymbol{Y}$ conditioned on the MIMO channel matrix $\boldsymbol{H}$ and the transmitted data matrix $\boldsymbol{S}$ is given by

$$p(\boldsymbol{Y}|\boldsymbol{H},\boldsymbol{S}) = \frac{1}{\big(2\pi\sigma_e^2\big)^{n_R\cdot N_f}}\exp\left(-\frac{1}{(2\pi\sigma_e^2)}\|\boldsymbol{Y}-\boldsymbol{H}\boldsymbol{S}\|_F^2\right). \tag{22}$$

The joint ML estimation of $\boldsymbol{S}$ and $\boldsymbol{H}$ is obtained in theory by maximizing $p(\boldsymbol{Y}|\boldsymbol{H},\boldsymbol{S})$ jointly over the continuous space for $\boldsymbol{H}$ and the discrete space for $\boldsymbol{S}$, which is computationally impossible to achieve. Hence, suboptimal solution has to be sought.

Note that maximizing $p(\boldsymbol{Y}|\boldsymbol{H},\boldsymbol{S})$ is equivalent to minimizing the cost function

$$J(\boldsymbol{H},\boldsymbol{S}) = \|\boldsymbol{Y}-\boldsymbol{H}\boldsymbol{S}\|_F^2 = \|\bar{\boldsymbol{Y}}-\bar{\boldsymbol{H}}\bar{\boldsymbol{S}}\|_F^2, \tag{23}$$

where

$$\bar{\boldsymbol{H}} = \begin{bmatrix} \boldsymbol{H}^{(\rm R)} & -\boldsymbol{H}^{(\rm I)} \\ \boldsymbol{H}^{(\rm I)} & \boldsymbol{H}^{(\rm R)} \end{bmatrix} \in \mathbb{R}^{\bar{n}_R\cdot\bar{n}_T} \tag{24}$$

in which $\bar{n}_T = 2n_T$, $\bar{n}_R = 2n_R$ and $\boldsymbol{H} = \boldsymbol{H}^{(\rm R)} + \mathrm{j}\cdot\boldsymbol{H}^{(\rm I)}$, while $\bar{\boldsymbol{Y}} \in \mathbb{R}^{\bar{n}_R\cdot N_f}$ with

$$\bar{\boldsymbol{Y}} = \begin{bmatrix} \boldsymbol{Y}^{(\rm R)} \\ \boldsymbol{Y}^{(\rm I)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{y}^{(\rm R)}(1) & \boldsymbol{y}^{(\rm R)}(2) & \cdots & \boldsymbol{y}^{(\rm R)}(N_f) \\ \boldsymbol{y}^{(\rm I)}(1) & \boldsymbol{y}^{(\rm I)}(2) & \cdots & \boldsymbol{y}^{(\rm I)}(N_f) \end{bmatrix} \tag{25}$$

in which $\boldsymbol{y}(k) = \boldsymbol{y}^{(\rm R)}(k) + \mathrm{j}\cdot\boldsymbol{y}^{(\rm I)}(k)$, and $\bar{\boldsymbol{S}} \in \mathbb{R}^{\bar{n}_T\cdot N_f}$ with

$$\bar{\boldsymbol{S}} = \begin{bmatrix} \boldsymbol{S}^{(\rm R)} \\ \boldsymbol{S}^{(\rm I)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{s}^{(\rm R)}(1) & \boldsymbol{s}^{(\rm R)}(2) & \cdots & \boldsymbol{s}^{(\rm R)}(N_f) \\ \boldsymbol{s}^{(\rm I)}(1) & \boldsymbol{s}^{(\rm I)}(2) & \cdots & \boldsymbol{s}^{(\rm I)}(N_f) \end{bmatrix} \tag{26}$$

in which $\boldsymbol{s}(k) = \boldsymbol{s}^{(\rm R)}(k) + \mathrm{j}\cdot\boldsymbol{s}^{(\rm I)}(k)$. The minimizer of $J(\boldsymbol{H},\boldsymbol{S})$ given $\bar{\boldsymbol{S}}$ is readily expressed by the closed-form LS solution

$$\bar{\boldsymbol{H}} = \bar{\boldsymbol{Y}}\bar{\boldsymbol{S}}^{\rm T}\big(\bar{\boldsymbol{S}}\bar{\boldsymbol{S}}^{\rm T}\big)^{-1}, \tag{27}$$

which is plugged back to (23) to yield

$$\begin{aligned} J(\bar{\boldsymbol{S}}) &= \big\|\bar{\boldsymbol{Y}} - \bar{\boldsymbol{Y}}\bar{\boldsymbol{S}}^{\rm T}\big(\bar{\boldsymbol{S}}\bar{\boldsymbol{S}}^{\rm T}\big)^{-1}\bar{\boldsymbol{S}}\big\|_F^2 \\ &= \mathrm{trace}\Big(\big(\bar{\boldsymbol{Y}} - \bar{\boldsymbol{Y}}\bar{\boldsymbol{S}}^{\rm T}\big(\bar{\boldsymbol{S}}\bar{\boldsymbol{S}}^{\rm T}\big)^{-1}\bar{\boldsymbol{S}}\big)\big(\bar{\boldsymbol{Y}} - \bar{\boldsymbol{Y}}\bar{\boldsymbol{S}}^{\rm T}\big(\bar{\boldsymbol{S}}\bar{\boldsymbol{S}}^{\rm T}\big)^{-1}\bar{\boldsymbol{S}}\big)^{\rm T}\Big) \\ &= \mathrm{trace}\big(\bar{\boldsymbol{Y}}\bar{\boldsymbol{Y}}^{\rm T}\big) - \mathrm{trace}\big(\bar{\boldsymbol{Y}}\bar{\boldsymbol{S}}^{\rm T}\big(\bar{\boldsymbol{S}}\bar{\boldsymbol{S}}^{\rm T}\big)^{-1}\bar{\boldsymbol{S}}\bar{\boldsymbol{Y}}^{\rm T}\big). \end{aligned} \tag{28}$$

Since the first term in the righthand side of (28) does not depend on $\bar{\boldsymbol{S}}$, minimizing $J(\bar{\boldsymbol{S}})$ is equivalent to minimizing the following cost function

$$\begin{aligned} F(\bar{\boldsymbol{S}}) &= -\frac{1}{2}\mathrm{trace}\big(\bar{\boldsymbol{Y}}\bar{\boldsymbol{S}}^{\rm T}\big(\bar{\boldsymbol{S}}\bar{\boldsymbol{S}}^{\rm T}\big)^{-1}\bar{\boldsymbol{S}}\bar{\boldsymbol{Y}}^{\rm T}\big) \\ &= -\frac{1}{2}\mathrm{trace}\big(\bar{\boldsymbol{Y}}^{\rm T}\bar{\boldsymbol{Y}}\bar{\boldsymbol{S}}^{\rm T}\big(\bar{\boldsymbol{S}}\bar{\boldsymbol{S}}^{\rm T}\big)^{-1}\bar{\boldsymbol{S}}\big). \end{aligned} \tag{29}$$

Note that the CV constellation set $\mathcal{S} = \mathcal{S}^{(\rm R)} + \mathrm{j}\cdot\mathcal{S}^{(\rm I)}$ of (10) is symmetric in both real and imaginary dimensions, which means that $\mathcal{S}^{(\rm R)}$ and $\mathcal{S}^{(\rm I)}$ are identical, both containing the same $\sqrt{L}$ real-valued (RV) constellation symbol points. Let us define the RV constellation set $\mathcal{D}$ that contains these $\sqrt{L}$ RV symbol points, namely,

$$\mathcal{D} = \mathcal{S}^{(\rm R)} = \mathcal{S}^{(\rm I)}. \tag{30}$$

Since $s_m^{(\rm R)}(k) \in \mathcal{D}$ and $s_m^{(\rm I)}(k) \in \mathcal{D}$ $\forall m,k$, we have $\bar{\boldsymbol{S}} \in \mathcal{D}^{\bar{n}_T\cdot N_f}$. The above optimization problem can then be expressed as

$$\begin{aligned} &\min_{\bar{\boldsymbol{S}}} \Big\{F\big(\bar{\boldsymbol{S}}\big) = -\tfrac{1}{2}\mathrm{trace}\big(\bar{\boldsymbol{Y}}^{\rm T}\bar{\boldsymbol{Y}}\bar{\boldsymbol{S}}^{\rm T}\big(\bar{\boldsymbol{S}}\bar{\boldsymbol{S}}^{\rm T}\big)^{-1}\bar{\boldsymbol{S}}\big)\Big\}, \\ &\text{s.t. } \bar{\boldsymbol{S}} \in \mathcal{D}^{\bar{n}_T\cdot N_f}, \end{aligned} \tag{31}$$

which is of course computationally prohibitive to solve directly.

We now seek approximation or suboptimal solution. Note that although $\|\bar{\boldsymbol{S}}\|_F$ is frame dependent, it can be approximated by a constant, specifically,

$$\|\bar{\boldsymbol{S}}\|_F \approx \sqrt{\bar{n}_T N_f\frac{\sigma_s^2}{2}} = \sqrt{n_T N_f\sigma_s^2}, \tag{32}$$

where we have $\mathrm{E}\big\{\big(s_m^{(\rm R)}(k)\big)^2\big\} = \mathrm{E}\big\{\big(s_m^{(\rm I)}(k)\big)^2\big\} = \frac{\sigma_s^2}{2}$. The above approximation is very accurate for a sufficiently large N_f, since $\big\{s_m(k) = s_m^{(\rm R)}(k) + \mathrm{j}\cdot s_m^{(\rm I)}(k)\big\}_{k=1}^{N_f}$ for $1 \leq m \leq$

n_T are independent sequences. Therefore, the solution of the following optimization problem

$$\begin{aligned}&\min_{\bar{\boldsymbol{S}}}\left\{F(\bar{\boldsymbol{S}})=-\tfrac{1}{2}\text{trace}\big(\bar{\boldsymbol{Y}}^{\text{T}}\bar{\boldsymbol{Y}}\bar{\boldsymbol{S}}^{\text{T}}\big(\bar{\boldsymbol{S}}\bar{\boldsymbol{S}}^{\text{T}}\big)^{-1}\bar{\boldsymbol{S}}\big)\right\},\\&\text{s.t. }\|\bar{\boldsymbol{S}}\|_F=\sqrt{n_TN_f\sigma_s^2},\end{aligned}\tag{33}$$

provides an approximate solution to (31). More specifically, let the solution of (33) be $\widetilde{\bar{\boldsymbol{S}}}$. By mapping every element of $\widetilde{\bar{\boldsymbol{S}}}$ to the nearest point in $\mathcal{D}$, the resulting hard-decision solution $\widehat{\bar{\boldsymbol{S}}}$ is a suboptimal solution of (31). Noting that the constraint $\|\bar{\boldsymbol{S}}\|_F=\sqrt{n_TN_f\sigma_s^2}$ is equivalent to the scaled version of the sphere manifold (1) by a factor of $\sqrt{n_TN_f\sigma_s^2}$, we propose a new algorithm utilizing a sphere manifold based optimization algorithm, specifically, Riemannian conjugate gradient algorithm, to efficiently solve the optimization (33), which is detailed in Section III-C.

Using Riemannian conjugate gradient algorithm to solve the optimization (33) requires an initial point $\bar{\boldsymbol{S}}_{\text{ini}}$. In order to resolve the permutation and scaling ambiguity as mentioned in Section III-A, we adopt the training based scheme with a minimum number of training samples $N_t\approx n_T$ to obtain the initial LS channel estimate of (21). Based on this initial channel estimate $\widehat{\boldsymbol{H}}_{\text{LS}}^{(\text{t})}$, the ZF data detection can be carried out according to (17), yielding the initial hard-decision solution $\widehat{\boldsymbol{S}}_{\text{ini}}$ which can be turned into $\bar{\boldsymbol{S}}_{\text{ini}}$ according to (26). Given the initial point $\bar{\boldsymbol{S}}_{\text{ini}}$, the proposed semi-blind scheme uses Riemannian conjugate gradient algorithm of Section III-C to obtain a solution of (33), denoted as $\widetilde{\bar{\boldsymbol{S}}}_{\text{smo}}$. By quantizing every element of $\widetilde{\bar{\boldsymbol{S}}}_{\text{smo}}$, the resulting hard-decision solution $\widehat{\bar{\boldsymbol{S}}}_{\text{smo}}$ is an approximate solution of the optimization problem (31). Then the data detection solution is constructed as

$$\widehat{\boldsymbol{S}}_{\text{smo}}=\widehat{\bar{\boldsymbol{S}}}_{\text{smo}}[1:n_T;:]+\text{j}\cdot\widehat{\bar{\boldsymbol{S}}}_{\text{smo}}[n_T+1:\bar{n}_T;:],\tag{34}$$

which is a suboptimal data detection solution. The final channel estimate is then given by the DD LS estimate based on $\widehat{\boldsymbol{S}}_{\text{smo}}$ as

$$\widehat{\boldsymbol{H}}_{\text{fin}}=\boldsymbol{Y}\widehat{\boldsymbol{S}}_{\text{smo}}^{\text{H}}\big(\widehat{\boldsymbol{S}}_{\text{smo}}\widehat{\boldsymbol{S}}_{\text{smo}}^{\text{H}}\big)^{-1}.\tag{35}$$

To take the advantage of improved accuracy of the final channel estimate $\widehat{\boldsymbol{H}}_{\text{fin}}$ over the initial rough channel estimate $\widehat{\boldsymbol{H}}_{\text{LS}}^{(\text{t})}$, we may utilize $\widehat{\boldsymbol{H}}_{\text{fin}}$ to carry out the final ZF detection to yield the final data detection solution $\widehat{\boldsymbol{S}}_{\text{ZF}}^{(\text{f})}$. This semi-blind scheme is summarized in Algorithm 1.

C. Riemannian Conjugate Gradient Algorithm for Data Detection

Recall that the constraint $\|\bar{\boldsymbol{S}}\|_F=\sqrt{n_TN_f\sigma_s^2}$ of the optimization problem (33) is the scaled version of the sphere manifold (1) by a factor of $\sqrt{n_TN_f\sigma_s^2}$. Thus, the solution of (33) can be found via an equivalent Riemannian optimization problem. Specifically, let the solution of this equivalent Riemannian optimization problem be $\boldsymbol{X}_{\text{smo}}$, i.e.,

$$\begin{aligned}\boldsymbol{X}_{\text{smo}}=\arg\min_{\boldsymbol{X}\in\mathbb{S}^{\bar{n}_T\cdot N_f-1}}\Bigg\{&F(\boldsymbol{X})=\\&-\frac{1}{2}\text{trace}\big(\bar{\boldsymbol{Y}}^{\text{T}}\bar{\boldsymbol{Y}}\boldsymbol{X}^{\text{T}}\big(\boldsymbol{X}\boldsymbol{X}^{\text{T}}\big)^{-1}\boldsymbol{X}\big)\Bigg\}.\end{aligned}\tag{36}$$

Algorithm 1 Proposed semi-blind scheme for MIMO

Initialize: Training data $\{\boldsymbol{S}_t,\boldsymbol{Y}_t\}$, observation data $\boldsymbol{Y}$;

Output: $\widehat{\boldsymbol{S}}_{\text{ZF}}^{(\text{f})}$ and $\widehat{\boldsymbol{H}}_{\text{smo}}$ are suboptimal joint data detection and channel estimation solution.

1: **Stage i)-Initialization**. Given $\{\boldsymbol{S}_t,\boldsymbol{Y}_t\}$,
2: Calculate initial training-based LS channel estimate $\widehat{\boldsymbol{H}}_{\text{LS}}^{(\text{t})}$ according to (21);
3: Calculate initial ZF soft data detection solution $\widetilde{\boldsymbol{S}}_{\text{ini}}=\Big(\big(\widehat{\boldsymbol{H}}_{\text{LS}}^{(\text{t})}\big)^{\text{H}}\widehat{\boldsymbol{H}}_{\text{LS}}^{(\text{t})}\Big)^{-1}\big(\widehat{\boldsymbol{H}}_{\text{LS}}^{(\text{t})}\big)^{\text{H}}\boldsymbol{Y}$;
4: Quantize every element of $\widetilde{\boldsymbol{S}}_{\text{ini}}$ to produce hard-decision data detection solution, denoted as $\widehat{\boldsymbol{S}}_{\text{ini}}=\widehat{\boldsymbol{S}}_{\text{ini}}^{(\text{R})}+\text{j}\cdot\widehat{\boldsymbol{S}}_{\text{ini}}^{(\text{I})}$;
5: Obtain
$$\bar{\boldsymbol{S}}_{\text{ini}}=\begin{bmatrix}\widehat{\boldsymbol{S}}_{\text{ini}}^{(\text{R})}\\\widehat{\boldsymbol{S}}_{\text{ini}}^{(\text{I})}\end{bmatrix}.$$
6: **Stage ii)-Riemannian optimization for data detection**. Given $\bar{\boldsymbol{S}}_{\text{ini}}$,
7: Use Algorithm 2 of Section III-C to find the solution $\widetilde{\bar{\boldsymbol{S}}}_{\text{smo}}$ of the optimization problem (33);
8: Quantize every element of $\widetilde{\bar{\boldsymbol{S}}}_{\text{smo}}$ to produce hard-decision solution $\widehat{\bar{\boldsymbol{S}}}_{\text{smo}}$ as an suboptimal solution to the optimization problem (31);
9: Obtain the data detection solution $\widehat{\boldsymbol{S}}_{\text{smo}}$ according to (34);
10: Obtain the final DD-LS channel estimate $\widehat{\boldsymbol{H}}_{\text{fin}}$ according to (35).
11: **Stage iii)-Final data detection**. Given $\widehat{\boldsymbol{H}}_{\text{fin}}$,
12: Obtain the final ZF data detection solution $\widehat{\boldsymbol{S}}_{\text{ZF}}^{(\text{f})}$ according to
$$\big(\widehat{\boldsymbol{H}}_{\text{fin}}^{\text{H}}\widehat{\boldsymbol{H}}_{\text{fin}}\big)^{-1}\widehat{\boldsymbol{H}}_{\text{fin}}^{\text{H}}\boldsymbol{Y}=\widetilde{\boldsymbol{S}}_{\text{ZF}}^{(\text{f})}\xrightarrow{\text{quantize}}\widehat{\boldsymbol{S}}_{\text{ZF}}^{(\text{f})}.$$
13: **Return**. Joint channel estimation and data detection solution $\big\{\widehat{\boldsymbol{H}}_{\text{fin}},\widehat{\boldsymbol{S}}_{\text{ZF}}^{(\text{f})}\big\}$.

We have $\widetilde{\bar{\boldsymbol{S}}}_{\text{smo}}=\sqrt{n_TN_f\sigma_s^2}\boldsymbol{X}_{\text{smo}}$.

Riemannian conjugate gradient algorithm generalizes the classical conjugate gradient algorithm [23] to optimization problems over Riemannian manifolds [24]. For our objective function $F(\boldsymbol{X})$, it is easy to check that Euclidean gradient, Grad$F(\boldsymbol{X})$, can be calculated according to

$$\begin{aligned}\text{Grad}F(\boldsymbol{X})=&-\big(\boldsymbol{X}\boldsymbol{X}^{\text{T}}\big)^{-1}\boldsymbol{X}\bar{\boldsymbol{Y}}^{\text{T}}\bar{\boldsymbol{Y}}\\&+\big(\boldsymbol{X}\boldsymbol{X}^{\text{T}}\big)^{-1}\boldsymbol{X}\bar{\boldsymbol{Y}}^{\text{T}}\bar{\boldsymbol{Y}}\boldsymbol{X}^{\text{T}}\big(\boldsymbol{X}\boldsymbol{X}^{\text{T}}\big)^{-1}\boldsymbol{X}.\end{aligned}\tag{37}$$

Based on Grad$F(\boldsymbol{X})$, Riemannian gradient of the objective function $F(\boldsymbol{X})$ and the matrix transport on the sphere manifold can be calculated according to (4) and (5).

With all the ingredients available, we form the algorithm for solving (33) in Algorithm 2, which can easily be implemented using the Manifold Optimization Toolbox, Manopt, available at: `http://www.manopt.org`, see [24]. We use the default parameter settings in Manopt, so that in Step 4 of Algorithm 2, α_k is based on the line search backtracking procedure as described in p.63 of Section 4 in Chapter 4 of [16]. In Step 5 of Algorithm 2, β_k is based on the default option 'Hestenes-Stiefel's modified rule' in Manopt.

Algorithm 2 Riemannian conjugate gradient algorithm for solving (33)

Initialize: Observation data $\bar{\boldsymbol{Y}}$, initial point $\boldsymbol{X}_0 = \bar{\boldsymbol{S}}_{\text{ini}}/\|\bar{\boldsymbol{S}}_{\text{ini}}\|_F$ which is on the sphere manifold $\mathbb{S}^{\bar{n}_T \cdot N_f - 1}$, and default threshold ϖ, e.g., $\varpi = 10^{-6}$;

Output: $\widetilde{\bar{\boldsymbol{S}}}$ that yields the minimum $F(\widetilde{\bar{\boldsymbol{S}}})$.

1: Set $\boldsymbol{\eta}_0 = -\text{grad}F(\boldsymbol{X}_0)$ and $k = 0$;
2: **while** $\|\text{grad}F(\boldsymbol{X}_k)\|_F < \varpi$ **do**
3: $k = k + 1$;
4: Compute a step size α_k and set

$$\boldsymbol{X}_k = \text{Exp}_{\boldsymbol{X}_{k-1}}(\alpha_k \boldsymbol{\eta}_{k-1}); \tag{38}$$

5: Compute β_k and set

$$\boldsymbol{\eta}_k = -\text{grad}F(\boldsymbol{X}_k) + \beta_k \mathcal{T}_{\boldsymbol{X}_k \leftarrow \boldsymbol{X}_{k-1}}(\boldsymbol{\eta}_{k-1}); \tag{39}$$

6: **end while**
7: Return $\widetilde{\bar{\boldsymbol{S}}} = \sqrt{n_T N_f \sigma_s^2}\boldsymbol{X}_k$.

Specifically, we have

$$\beta_k = \max\left\{0, \frac{g(\text{grad}F(\boldsymbol{X}_k), \boldsymbol{\Gamma}_k)}{g(\mathcal{T}_{\boldsymbol{X}_k \leftarrow \boldsymbol{X}_{k-1}}(\boldsymbol{\eta}_{k-1}), \boldsymbol{\Gamma}_k)}\right\}, \tag{40}$$

with

$$\boldsymbol{\Gamma}_k = \text{grad}F(\boldsymbol{X}_k) - \mathcal{T}_{\boldsymbol{X}_k \leftarrow \boldsymbol{X}_{k-1}}(\text{grad}F(\boldsymbol{X}_{k-1})). \tag{41}$$

Remark 1. *The solution (36) is an approximate or suboptimal solution to the intractable optimization problem (31). It would be desirable to examine how 'close' the solution (36) is to the optimal solution of the problem (31). This is however very challenging if not impossible to do analytically. It is even difficult to compare the two solutions by simulation, as it is intractable to realize the optimal solution for the problem (31). As mentioned previously, for MIMO systems with high-order QAM signaling, it is only practical to implement low-complexity suboptimal solutions, such as the semi-blind scheme of iterative ZF data detection and DD LSCE, described in Section III-A. Our proposed semi-blind scheme based on sphere manifold offers an alternative to this existing low-complexity practical semi-blind scheme. Thus, it is more reasonable to compare our proposed semi-blind scheme based on sphere manifold with this existing semi-blind scheme. This is what we will do in the following simulation study.*

Remark 2. *Algorithm 2 is a local optimization algorithm that aims to find a local minimum near the initial solution $\boldsymbol{X}_0$. Whereas the convergence theory of linear conjugate gradient algorithm is well understood, nonlinear conjugate gradient methods have convergence properties that depend on the choice of α_k and β_k. Theoretically analyzing the convergence properties of this conjugate gradient algorithm is challenging. However, the conjugate gradient algorithm of Manopt is known to converge very fast [24]. In our application, we observe that Algorithm 2 always converges in 2 to 4 iterations. The convergence property of the existing semi-blind scheme of iterative ZF data detection and DD LSCE is well known [15]. As mentioned in Section III-A, typically 5 to 10 iterations between the ZF data detector and the DD LS channel estimator are sufficient for this semi-blind scheme to converge.*

TABLE II
COMPLEXITY ANALYSIS OF EVALUATING $F(\boldsymbol{X})$.

multiplication $\boldsymbol{X}\boldsymbol{X}^{\text{T}}$	$\mathsf{O}(\bar{n}_T N_f^2)$
inversion $(\boldsymbol{X}\boldsymbol{X}^{\text{T}})^{-1}$	$\mathsf{O}(\bar{n}_T^3)$
multiplication $(\boldsymbol{X}\boldsymbol{X}^{\text{T}})^{-1}\boldsymbol{X}$	$\mathsf{O}(\bar{n}_T^2 N_f)$
multiplication $\bar{\boldsymbol{Y}}^{\text{T}}\bar{\boldsymbol{Y}}$	$\mathsf{O}(\bar{n}_R^2 N_f)$
multiplication $\bar{\boldsymbol{Y}}^{\text{T}}\bar{\boldsymbol{Y}}\boldsymbol{X}^{\text{T}}$	$\mathsf{O}(N_f^2 \bar{n}_T)$
multiplication $\bar{\boldsymbol{Y}}^{\text{T}}\bar{\boldsymbol{Y}}\boldsymbol{X}^{\text{T}}(\boldsymbol{X}\boldsymbol{X}^{\text{T}})^{-1}\boldsymbol{X}$	$\mathsf{O}(\bar{n}_T^2 N_f)$

D. Complexity Analysis

The main computational cost of the proposed scheme lies in evaluating (36) and (37). Note that it is important to apply matrix multiplication association rule to realize efficient implementation. For example, the term $(\boldsymbol{X}\boldsymbol{X}^{\text{T}})^{-1}\boldsymbol{X} \in \mathbb{R}^{\bar{n}_T \cdot N_f}$ is evaluated only once and can be used in both $F(\boldsymbol{X})$ and Grad$F(\boldsymbol{X})$. Also it is well-known that the number of operations (real-valued multiplications and/or additions) required to compute matrix multiplication $\boldsymbol{X}\boldsymbol{X}^{\text{T}}$ is on the order of $\bar{n}_T N_f^2$, denoted as $\mathsf{O}(\bar{n}_T N_f^2)$, and the cost of matrix inversion $(\boldsymbol{X}\boldsymbol{X}^{\text{T}})^{-1}$ is on the order of $\mathsf{O}(\bar{n}_T^3)$.

Evaluation of $F(\boldsymbol{X})$: The total computational cost of evaluating $F(\boldsymbol{X})$ is detailed in Table II. From Table II, we conclude that the complexity of evaluating $F(\boldsymbol{X})$ is on the order of $\mathsf{O}(\bar{n}_T N_f^2 + (\bar{n}_T^2 + \bar{n}_R^2)N_f + \bar{n}_T^3)$.

Evaluation of Grad$F(\boldsymbol{X})$: Note that $(\boldsymbol{X}\boldsymbol{X}^{\text{T}})^{-1}\boldsymbol{X}$, $\bar{\boldsymbol{Y}}^{\text{T}}\bar{\boldsymbol{Y}}$ and $\bar{\boldsymbol{Y}}^{\text{T}}\bar{\boldsymbol{Y}}\boldsymbol{X}^{\text{T}}(\boldsymbol{X}\boldsymbol{X}^{\text{T}})^{-1}\boldsymbol{X}$ have already been computed. Hence, evaluating Grad$F(\boldsymbol{X})$ imposes an addition number of operations on the order of $\mathsf{O}(\bar{n}_T N_f^2)$.

Complexity of proposed algorithm: The complexity of **Stage i)-Initialization** is small by comparison. Also this initialization is needed in the existing scheme and, therefore, its complexity can be omitted in the comparison.

According to the above analysis, the number of operations (real-valued multiplications and/or additions) required by each iteration of Algorithm 2 is on the order of $\mathsf{O}(\bar{n}_T N_f^2 + (\bar{n}_T^2 + \bar{n}_R^2)N_f + \bar{n}_T^3)$. Hence the complexity of **Stage ii)-Riemannian optimization for data detection** is $\mathsf{O}(\bar{n}_T N_f^2 + (\bar{n}_T^2 + \bar{n}_R^2)N_f + \bar{n}_T^3)$, scaled by the number of iterations for Algorithm 2.

Stage iii)-Final data detection is optional and is performed at most once. By converting the complex-valued matrix multiplication and inversion into the equivalent real-valued ones, it can be seen that the computational cost of this stage is smaller than $\mathsf{O}(\bar{n}_T N_f^2 + (\bar{n}_T^2 + \bar{n}_R^2)N_f + \bar{n}_T^3)$.

Remark 3. *The computational complexity of the proposed algorithm is on the order of $\mathsf{O}(\bar{n}_T N_f^2 + (\bar{n}_T^2 + \bar{n}_R^2)N_f + \bar{n}_T^3)$, scaled by the number of iterations for Algorithm 2, which is 2 to 4. By contrast, it is well known that the computational complexity of the existing scheme of iterative ZF data detection and DD LSCE is also on the order of $\mathsf{O}(\bar{n}_T N_f^2 + (\bar{n}_T^2 + \bar{n}_R^2)N_f + \bar{n}_T^3)$, scaled by the number of iterations between the ZF data detector and the DD LS channel estimator, which is typically more than 4. It can be seen that the complexity per iteration are roughly the same for the both schemes. Hence, the proposed scheme offers some computational advantage as it*

requires a smaller number of iterations. This will be confirmed by the simulation study.

IV. Simulation Study

Three MIMO systems employing high-order modulation schemes were simulated with the length of data frame set to $N_f = 200$ and the number of initial training samples N_t. The proposed semi-blind scheme based on sphere manifold was compared with the following schemes: (a) 'initial LSCE', which is the ZF detection solution based on the initial training based LS channel estimate $\widehat{\boldsymbol{H}}_{\rm LS}^{(\rm t)}$, (b) 'existing semi-blind scheme', which is the semi-blind scheme of iterative ZF data detection and DD LSCE discussed in Section III-A, and (c) 'perfect channel knowledge', which is the ZF detection solution associated with the perfect CSI $\boldsymbol{H}$. In addition to the bit error rate (BER) performance, the mean absolute deviation (MAD), defined as

$$\text{MAD} = \text{E}\left\{\sum_{l=1}^{n_R}\sum_{m=1}^{n_T}\left|\widehat{h}_{l,m} - h_{l,m}\right|\right\}, \qquad (42)$$

was also used as a performance measure, where $\widehat{h}_{l,m}$ denotes the estimate of the true channel coefficient $h_{l,m}$, and the expectation is approximated using the time average over 100 random realizations of the MIMO channel. The average symbol power was set to $\sigma_s^2 = 1$ and the system's SNR was defined as SNR $= E_{\rm b}/N_{\rm o}$, where $E_{\rm b} = \sigma_s^2/\log_2 L$ is the average energy per bit and $N_{\rm o} = 2\sigma_\varepsilon^2$. The number of iterations for Algorithm 2 was set to 2, as this was observed to be sufficient for the conjugate gradient algorithm to converge.

Example 1: The system employed $n_T = 6$ transmit antennas and $n_R = 6$ receive antennas with 16-QAM signaling, i.e., $L = 16$. The number of initial training data was $N_t = n_T+2 = 8$. Fig. 1 compares the BER achieved by the proposed semi-blind scheme on sphere manifold with those of the initial LSCE based detection, the existing semi-blind scheme with 2 iterations and 10 iterations, as well as the detection based on perfect CSI. The achievable MAD of the proposed semi-blind scheme on sphere manifold is depicted in Fig. 2 in comparison with the MAD performance of the initial LSCE based scheme as well as the existing semi-blind scheme with 2 iterations and 10 iterations, respectively. From the results of Figs. 1 and 2, it can be seen that the performance of the proposed semi-blind scheme on sphere manifold is better than that of the existing semi-blind scheme with 2 iterations, while the performance of the existing semi-blind scheme with 10 iterations is better than the proposed semi-blind scheme. This is significant, as the complexity of the proposed semi-blind scheme on sphere

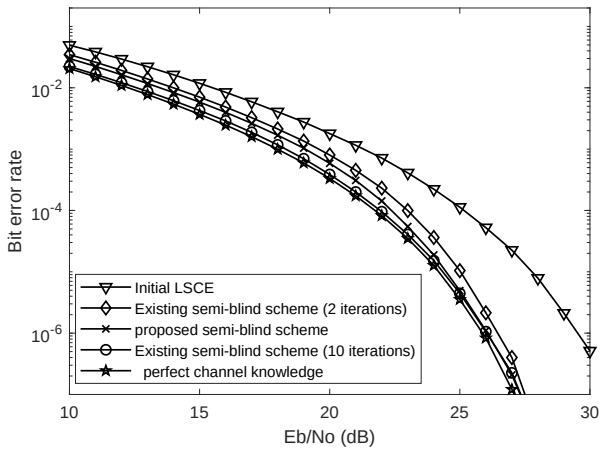

Fig. 1. Bit error rate performance comparison for the MIMO system with $N_T = N_R = 6$ employing 16-QAM modulation scheme.

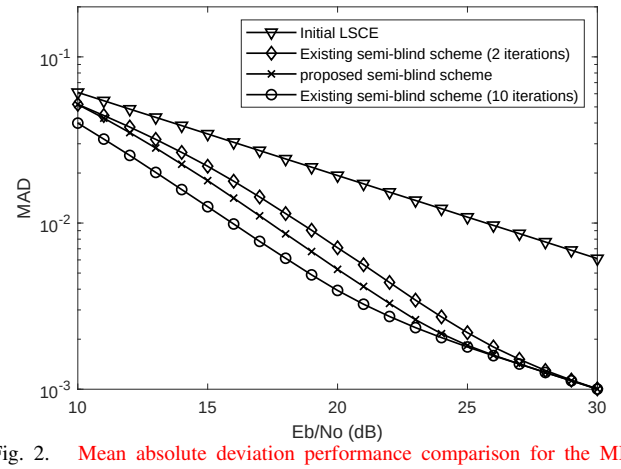

Fig. 2. Mean absolute deviation performance comparison for the MIMO system with $N_T = N_R = 6$ employing 16-QAM modulation scheme.

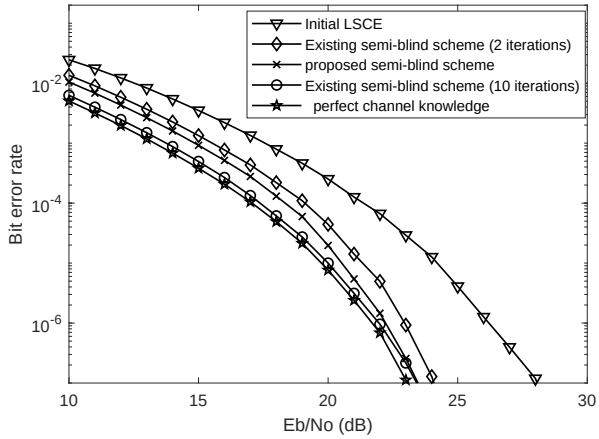

Fig. 3. Bit error rate performance comparison for the MIMO system with $N_T = N_R = 8$ employing 16-QAM modulation scheme.

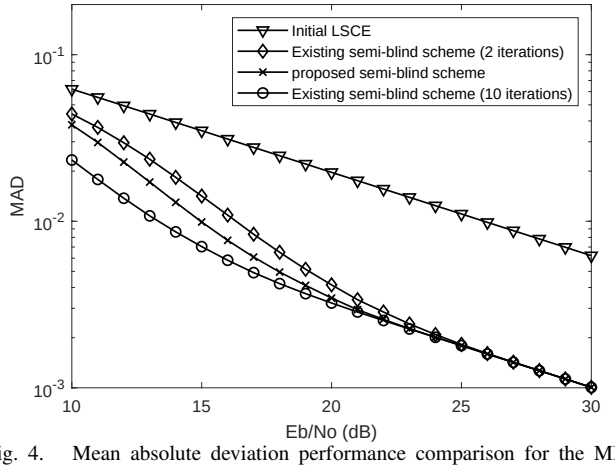

Fig. 4. Mean absolute deviation performance comparison for the MIMO system with $N_T = N_R = 8$ employing 16-QAM modulation scheme.

manifold is roughly comparable to that of the existing semi-blind scheme with 2 iterations but lower than that of the existing semi-blind scheme with 10 iterations.

Example 2: The system employed $n_T = 8$ transmit antennas and $n_R = 8$ receive antennas with 16-QAM signaling, i.e., $L = 16$. The number of initial training data was $N_t = n_T + 2 = 10$. Fig. 3 compares the BER achieved by the proposed semi-blind scheme on sphere manifold with those of the initial LSCE based detection, the existing semi-blind scheme with 2 iterations and 10 iterations, as well as the detection based on perfect CSI. The achievable MAD of the proposed semi-blind scheme on sphere manifold is depicted in Fig. 4 in comparison with the MAD performance of the initial LSCE based scheme as well as the existing semi-blind scheme with 2 iterations and 10 iterations, respectively. Again the results obtained show that the performance of the proposed semi-blind scheme on sphere manifold is better than that of the existing semi-blind scheme with 2 iterations, while the performance of the existing semi-blind scheme with 10 iterations is better than the proposed semi-blind scheme. Furthermore, the achievable BER performance is better than *Example 1*, since there are more antennas in this example.

Example 3: The system employed 64-QAM signaling and had $n_T = 6$ transmit antennas and $n_R = 6$ receive antennas. The number of initial training data was set to $N_t = n_T+2 = 8$.

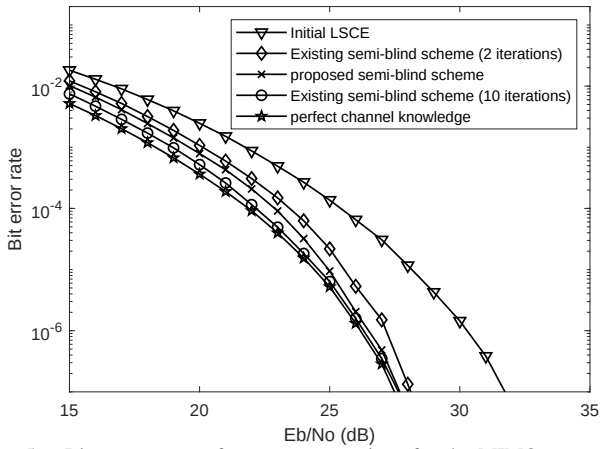

Fig. 5. Bit error rate performance comparison for the MIMO system with $N_T = N_R = 6$ employing 64-QAM modulation scheme.

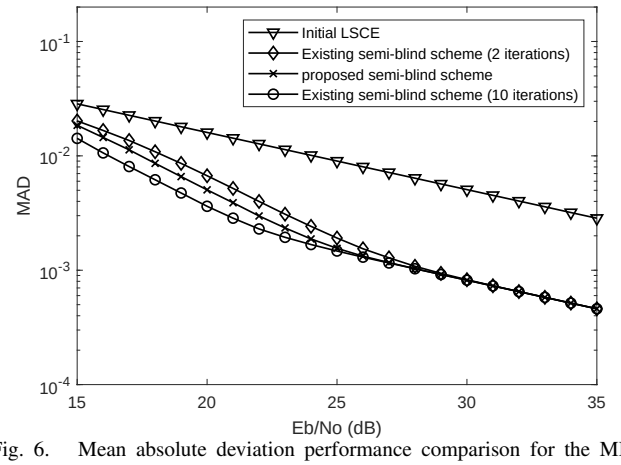

Fig. 6. Mean absolute deviation performance comparison for the MIMO system with $N_T = N_R = 6$ employing 64-QAM modulation scheme.

Fig. 5 compares the BER performance of the proposed semi-blind scheme on sphere manifold with those of the initial LSCE based scheme, the existing semi-blind scheme with 2 iterations and 10 iterations, as well as the scheme based on perfect CSI, while the MAD performance of the proposed semi-blind scheme on sphere manifold, the initial LSCE based scheme as well as the existing semi-blind scheme with 2 iterations and 10 iterations, respectively, are shown in Fig. 6. Based on the results of Figs. 5 and 6, we can draw the same conclusions as those observed in *Example 1* and *Example 2*.

V. Conclusions

In this paper we have introduced a new approach of semi-blind joint channel estimation and data detection on sphere manifold for MIMO systems employing high-order QAM signaling schemes, which consists of three stages. Specifically, in stage (i), a few training symbols are employed to provide a rough initial MIMO channel estimate which yields the initial ZF estimate of data samples; in stage (ii), the Riemannian conjugate gradient algorithm is used to estimate the data samples in real domain, and the detected data samples are used to estimate the final MIMO channel matrix; and in stage (iii), the final ZF data detection is carried out based on the final MIMO channel estimate. Our novel contributions include the derivation of the first order Riemannian geometry of sphere manifold that is necessary for solving stage (ii) of blind data detection. Our proposed semi-blind scheme on sphere manifold offers a viable and practical alternative to existing low-complexity semi-blind joint channel estimation and data detection schemes for high-dimensional MIMO with high-order QAM signaling. Simulation results have been provided to demonstrate the effectiveness of the proposed semi-blind scheme on sphere manifold.

References

[1] A. J. Paulraj, D. A. Gore, R. U. Nabar, and H. Bolcskei, "An overview of MIMO communications – a key to gigabit wireless," *Proc. IEEE*, vol. 92, no. 2, pp. 198–218, Feb. 2004.

[2] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Proces. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[3] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, " Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[4] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[5] S. Talwar, M. Viberg, and A. Paulraj, "Blind separation of synchronous co-channel digital signals using an antenna array–Part I: Algorithms," *IEEE Trans. Signal Proces.*, vol. 44, no. 5, pp. 1184–1197, May 1996.

[6] A. Ranheim, "A decoupled approach to adaptive signal separation using an antenna array," *IEEE Trans. Vehicular Technology*, vol. 48, no. 3, pp. 676–682, May 1999.

[7] A. Dogandzic and A. Nehorai, "Generalized multivariate analysis of variance – a unified framework for signal processing in correlated noise," *IEEE Signal Proces. Mag.*, vol. 20, no. 5, pp. 39–54, Sep. 2003.

[8] A. Medles and D. T. M. Slock, "Semiblind channel estimation for MIMO spatial multiplexing systems," in *Proc. VTC2001-Fall* (Atlantic City, NJ), Oct. 7-11, 2001, vol.2, pp. 1240–1244.

[9] C. Cozzo and B. L. Hughes, "Joint channel estimation and data detection in space-time communications," *IEEE Trans. Commun.*, vol. 51, no. 8, pp. 1266–1270, Aug. 2003.

[10] S. Buzzi, M. Lops, and S. Sardellitti, "Performance of iterative data detection and channel estimation for single-antenna and multiple-antennas wireless communications," *IEEE Trans. Vehicular Technology*, vol. 53, no. 4, pp. 1085–1104, Jul. 2004.

[11] T. Wo, P. A. Hoeher, A. Scherb, and K.-D. Kammeyer, "Performance analysis of maximum-likelihood semiblind estimation of MIMO channels," in *Proc. VTC2006-Spring* (Melbourne, Australia), May 7-10, 2006, pp. 1738–1742.

[12] M. Abuthinien, S. Chen, and L. Hanzo, "Semi-blind joint maximum likelihood channel estimation and data detection for MIMO systems," *IEEE Signal Proces. Let.*, vol. 15, pp. 202–205, 2008.

[13] S. Chen, S. Sugiura, and L. Hanzo, "Semi-blind joint channel estimation and data detection for space-time shift keying systems," *IEEE Signal Proces. Let.*, vol. 17, no. 12, pp. 993–996, Dec. 2010.

[14] I. Dey, G. G. Messier, S. Magierowski, and S. Chen, "Semi-blind iterative joint channel estimation and K-best sphere decoding for MIMO," in *Proc. 2013 IEEE Pacific Rim Conf Commun., Computers and Signal Proces.* (Victoria, Canada), Aug. 27-29, 2013, pp. 37–44.

[15] S. Chen and L. Hanzo, "Fast converging semi-blind space-time equalisation for dispersive QAM MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 3969–3974, Aug. 2009.

[16] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2008.

[17] B. Mishra, G. Meyer, F. Bach, and R. Sepulchre, "Low-rank optimization with trace norm penalty," *SIAM J. Optimization*, vol. 23, no. 4, pp. 2124–2149, 2013.

[18] Y. Xie, J. Ho, and B. Vemuri, "On a nonlinear generalization of sparse coding and dictionary learning," in *Proc. 30th Int Conf. Machine Learning* (Atlanta, GA), June 16-21, 2013, pp. 1480–1488.

[19] M. Harandi, R. Hartley, C. Shen, B. Lovell, and C. Sanderson, "Extrinsic methods for coding and dictionary learning on Grassmann manifolds," *arXiv:1401.8126*, pp. 1–41, 2014.

[20] X. Hong, J. Gao, S. Chen, and T. Zia, "Sparse density estimation on the multinomial manifold," *IEEE Trans. Neural Networks and Learning Systems*, vol. 26, no. 11, pp. 2972–2977, Nov. 2015.

[21] X. Hong and J. Gao, "Sparse density estimation on multinomial manifold combined with local component analysis," in *Proc. IJCNN 2015* (Killarney, Ireland), Jul. 12-17, 2015, pp. 1–7.

[22] Y. M. Lui, "Advances in matrix manifolds for computer vision," *Image and Vision Computing*, vol. 30, nos.6-7, pp. 380–388, 2012.

[23] W. W. Hager and H. Zhang, "A survey of nonlinear conjugate gradient methods," *Pacific J. Optimization*, vol. 2, no. 1, pp. 35–58, 2006.

[24] N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre, "Manopt, a Matlab toolbox for optimization on manifolds," *J. Machine Learning research*, vol. 15, pp. 1455–1459, 2014.