Cascaded Channel Estimation for Large Intelligent Metasurface Assisted Massive MIMO

Zhen-Qing He and Xiaojun Yuan, *Senior Member, IEEE*

***Abstract*—In this letter, we consider the problem of channel estimation for large intelligent metasurface (LIM) assisted massive multiple-input multiple-output (MIMO) systems. The main challenge of this problem is that the LIM integrated with a large number of low-cost metamaterial antennas can only passively reflect the incident signals by certain phase shifts, and does not have any signal processing capability. To deal with this, we introduce a general framework for the estimation of the transmitter-LIM and LIM-receiver cascaded channel, and propose a two-stage algorithm that includes a sparse matrix factorization stage and a matrix completion stage. Simulation results illustrate that the proposed method can achieve accurate channel estimation for LIM-assisted massive MIMO systems.**

***Index Terms*—Bilinear factorization, channel estimation, large intelligent metasurface, massive MIMO, matrix completion.**

I. Introduction

MASSIVE multiple-input multiple-output (MIMO), as a promising technology for future wireless systems, has attracted growing research interest in both academia and industry over recent years. Although massive MIMO exhibits huge potentials to support a significantly large amount of mobile data traffic and wireless connections, implementing this system with large-scale antenna arrays in practice remains very challenging due to high hardware cost and increased power consumption. To achieve green and sustainable wireless networks, researchers have started looking into energy efficient techniques to improve the system performance, ranging from the utilization of energy efficient hardware components to the design of green resource allocation and transceiver signal processing algorithms.

To reduce energy consumption and enhance communication quality in wireless networks, the large intelligent metasurface (LIM) [1], also known as the intelligent reflecting surface [2] or the reconfigurable intelligent surface [3], has been recently proposed as an innovative technology that conceptually goes beyond contemporary massive MIMO communications. Metamaterials, as an emerging technology known for its flexibility in manipulating electromagnetic waves, have found applications such as in radar imaging [4], etc. As a potential application of the metamaterial in wireless communications, the LIM with integrated electronics retains almost all the advantages of massive MIMO such as allowing for an unprecedented focusing of energy that enables highly efficient wireless charging and remote sensing. Although traditional reflecting surfaces have a variety of applications in radar and satellite communications, their application in terrestrial wireless communication was not possible earlier. This is because these reflecting surfaces only had fixed phase shifters that could not adapt the induced phases with time-varying channels which generally constitute the wireless propagation environments.

To achieve full potential of the LIM-assisted massive MIMO systems, the channel state information (CSI) between the base station (BS) and the LIM and between the LIM and the receiver is essential in reflect beamforming [2], energy-efficient design [3], as well as in simultaneous passive beamforming and information transfer [5]. The main challenge of the CSI acquisition of the LIM-assisted MIMO systems is that the LIM, unlike the BS or the receiver, only passively reflects the electromagnetic waves, and does not have any signal processing capability. By leveraging the programmable property of the LIM and the rank-deficient structure of the massive MIMO channel, we formulate the BS-LIM and LIM-receiver cascaded channel estimation as a combined sparse matrix factorization and matrix completion problem. To solve the problem, we present a two-stage algorithm which includes the sparse matrix factorization stage and the matrix completion stage. The proposed two-stage algorithm includes the bilinear generalized approximate message passing (BiG-AMP) [6] for sparse matrix factorization and the Riemannian manifold gradient-based algorithm for matrix completion [7]. Numerical results demonstrate that our algorithm is able to attain accurate channel estimation for the LIM-assisted massive MIMO systems.

To the best of our knowledge, this is the first attempt to tackle the cascaded channel estimation problem for the LIM-assisted massive MIMO systems with all passive elements in the LIM.[1] Notice that the authors in [8] proposed a compressive sensing and training based deep learning approach for the LIM-assisted channel estimation. Nevertheless, in [8] a few active antenna elements of the LIM are required to circumvent the challenging cascaded channel estimation problem by utilizing conventional channel estimation techniques. In addition, it should be noted that the BiG-AMP algorithm has been previously exploited for blind signal detection in conventional massive MIMO systems [9], where the transmitter does not need to send pilot signals to the receiver, and the channel and

[1]The LIM-based MIMO technology is still immature in terms of practical implementations and a number of follow-up works are necessary to prove its practicality and utility. At this stage, this letter can be seen as an academic exercise for channel estimation by assuming the practicality of the LIM-based MIMO technology.

Manuscript received August 8, 2019; revised October 10, 2019; accepted October 14, 2019. Date of publication October 22, 2019; date of current version February 7, 2020. This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB1801105, and in part by the National Natural Science Foundation of China under Grant 61801084. The associate editor coordinating the review of this article and approving it for publication was V. Raghavan. *(Corresponding author: Xiaojun Yuan.)*

The authors are with the Center for Intelligent Networking and Communications, National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: zhenqinghe@uestc.edu.cn; xjyuan@uestc.edu.cn).

Digital Object Identifier 10.1109/LWC.2019.2948632

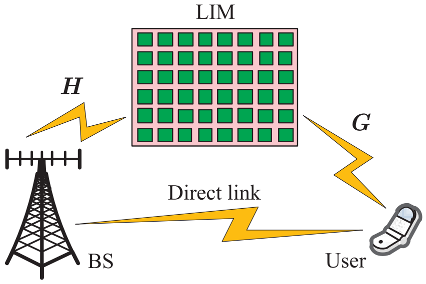

Fig. 1. A LIM-assisted massive MIMO system.

the data signal can be simultaneously estimated by factorizing the received signal matrix. In contrast, this letter aims to factorize the cascaded channels of the LIM-assisted MIMO systems by using pilot signals, which is a totally different problem from the one considered in [9].

Notations: $\mathbb{E}\{\cdot\}$, $\mathbb{V}\mathrm{ar}\{\cdot\}$, $\delta(\cdot)$, $\mathbb{C}$, $(\cdot)^T$, $(\cdot)^H$, $(\cdot)^*$, and $\|\cdot\|_F$ denote the expectation operator, the variance operator, the Dirac delta function, the space of complex number, the transpose, the conjugate transpose, the conjugate, and the Frobenius norm, respectively. The (i, j)-th entry, the i-th row, and the j-th column of a matrix $\mathbf{A}$, are denoted by $a_{i,j}$, $\mathbf{a}_i^T$, and $\boldsymbol{a}_j$, respectively. We use $\odot$ and $\mathcal{CN}(\boldsymbol{a}, \boldsymbol{C})$ to stand for the Hadamard product and the circularly-symmetric complex Gaussian distribution with mean vector $\boldsymbol{a}$ and covariance matrix $\boldsymbol{C}$, respectively.

II. System Model

Consider a LIM-assisted massive MIMO system, as shown in Fig. 1, where the LIM consists of N low-cost passive elements and the BS is equipped with M transmit antennas to serve a number of user terminals, each equipped with L receive antennas. In particular, the LIM is deployed to assist the BS in communicating with the users. Without loss of generality, we consider the communication from the BS to a reference user. The baseband equivalent channels from the BS to the LIM and from the LIM to the user are respectively denoted by $\boldsymbol{H} \in \mathbb{C}^{N\times M}$ and $\boldsymbol{G} \in \mathbb{C}^{L\times N}$. The channel component of the direct link between the BS and the user is neglected due to unfavorable propagation conditions or can be estimated (and cancelled from the model) via conventional massive MIMO channel estimation methods by turning off the LIM. We assume a block-fading[2] channel with coherence time T, i.e., the channel remains unchanged within each transmission block of length T and varies from block to block. Then, the received signal of the reference user can be expressed as

$$\boldsymbol{y}[t] = \boldsymbol{G}(\boldsymbol{s}[t] \odot (\boldsymbol{H}\boldsymbol{x}[t])) + \boldsymbol{w}[t],\ t = 1, \ldots, T \tag{1}$$

where $\boldsymbol{x}[t] \in \mathbb{C}^M$ and $\boldsymbol{w}[t] \in \mathbb{C}^L$ are respectively the transmitted signal and the additive noise drawn from $\mathcal{CN}(\mathbf{0}, \sigma^2\boldsymbol{I})$ at time t with σ^2 being the noise power. The phase shift vector $\boldsymbol{s}[t]$ of the LIM is defined as $\boldsymbol{s}[t] \triangleq [s_{1,t}e^{\jmath\theta_{1,t}}, \ldots, s_{N,t}e^{\jmath\theta_{N,t}}]^T$ where $\jmath \triangleq \sqrt{-1}$, with $\theta_{n,t} \in (0, 2\pi]$ and $s_{n,t} \in \{0, 1\}$ representing the phase shift and the on/off state[3] of the n-th LIM reflect element at time t, respectively. By summarizing all the T samples in a transmission block, the received signal can be recast as

$$\boldsymbol{Y} = \boldsymbol{G}(\boldsymbol{S} \odot (\boldsymbol{H}\boldsymbol{X})) + \boldsymbol{W} \tag{2}$$

where $\boldsymbol{Y} \triangleq [\boldsymbol{y}[1], \ldots, \boldsymbol{y}[T]] \in \mathbb{C}^{L\times T}$, $\boldsymbol{W} \triangleq [\boldsymbol{w}[1], \ldots, \boldsymbol{w}[T]] \in \mathbb{C}^{L\times T}$, $\boldsymbol{S} \triangleq [\boldsymbol{s}[1], \ldots, \boldsymbol{s}[T]] \in \mathbb{C}^{N\times T}$, and $\boldsymbol{X} \triangleq [\boldsymbol{x}[1], \ldots, \boldsymbol{x}[T]] \in \mathbb{C}^{M\times T}$. In general, a smart programmable controller is built in the LIM to adaptively adjust the states and phases of the LIM based on environmental changes, which is usually referred to as reflect beamforming or passive beamforming [2], [5].

The BS-LIM channel matrix $\boldsymbol{H}$ and the LIM-user channel matrix $\boldsymbol{G}$ are assumed to be rank-deficient, i.e., $\mathrm{rank}(\boldsymbol{H}) < \min\{N, M\}$ and $\mathrm{rank}(\boldsymbol{G}) < \min\{L, N\}$. The rank-deficiency property commonly arises in millimeter wave MIMO channels under far-field and limited-scattering assumptions.

III. Problem Statement

The objective of this letter is to estimate the channel matrices $\boldsymbol{H}$ and $\boldsymbol{G}$ based on the observation $\boldsymbol{Y}$ in (2) by assuming that the pilot symbols $\boldsymbol{X}$ and $\boldsymbol{S}$ are known to the receiver. We note that there exists an ambiguity problem in the estimates of $\boldsymbol{G}$ and $\boldsymbol{H}$, since the following equality holds for any full-rank diagonal matrix $\boldsymbol{\Phi} \in \mathbb{C}^{N\times N}$:

$$\boldsymbol{G}(\boldsymbol{S} \odot (\boldsymbol{H}\boldsymbol{X})) = \boldsymbol{G}'\big(\boldsymbol{S} \odot (\boldsymbol{H}'\boldsymbol{X})\big) \tag{3}$$

where $\boldsymbol{G}' \triangleq \boldsymbol{G}\boldsymbol{\Phi}$ and $\boldsymbol{H}' \triangleq \boldsymbol{\Phi}^{-1}\boldsymbol{H}$. Nevertheless, it is sufficient to acquire the knowledge of the alternative $\boldsymbol{G}'$ and $\boldsymbol{H}'$, rather than the ground-truth $\boldsymbol{G}$ and $\boldsymbol{H}$. This is because $\boldsymbol{G}'$ and $\boldsymbol{H}'$ can be viewed as effective channels in the sense that they would not affect the reflect/passive beamforming design (see Remark 1 below).

Remark 1: In reflect/passive beamforming [2], [3], [5], the phase shift vectors $\{\boldsymbol{s}[t]\}$ are generally set as a constant vector in each transmission block,[4] i.e.,

$$\boldsymbol{s}[t] = \boldsymbol{s}, t = 1, \ldots, T.$$

Then, the signal model in (2) reduces to

$$\boldsymbol{Y} = \boldsymbol{G}\mathrm{diag}\{\boldsymbol{s}\}\boldsymbol{H}\boldsymbol{X} + \boldsymbol{W}.$$

Clearly, we have

$$\boldsymbol{G}\mathrm{diag}\{\boldsymbol{s}\}\boldsymbol{H} = \boldsymbol{G}'\mathrm{diag}\{\boldsymbol{s}\}\boldsymbol{H}'.$$

This implies that the design of $\boldsymbol{s}$ for reflect beamforming based on $\boldsymbol{G}$ and $\boldsymbol{H}$ yields the same result as that based on $\boldsymbol{G}'$ and $\boldsymbol{H}'$.

IV. Two-Stage Channel Estimation Approach

In this section, we propose a two-stage channel estimation approach for the LIM-assisted massive MIMO systems. We first present the outline of the proposed scheme and then elaborate the implementation details.

[2]The block-fading channel model naturally arises when any two of the three nodes (i.e., the BS, the LIM, and the user) are in a moving state. In practice, both the BS and the LIM may be fixed at known positions. In this situation, the block fading assumption is still valid when considering moving scatterers or considering wireless communications over certain millimeter wave bands in bad weather conditions (e.g., heavy rain, downpour, and monsoon) [10].

[3]The state "off" of a LIM reflect element means that there is only structure-mode reflection generated as if the element is a conducting object, whereas the state "on" means that there are both structure-mode reflection and antenna-mode reflection [11]. Note that the structure-mode can be absorbed into the direct link in channel modelling.

[4]In reflect/passive beamforming, there is no need to change the value of the phase shift vector $\boldsymbol{s}[t]$ unless the channel changes.

A. Outline of the Proposed Scheme

Based on the discussions in Section III, we need to estimate $\boldsymbol{G}$ and $\boldsymbol{H}$ up to an ambiguity in (3) by appropriate design of the training signals $\{\boldsymbol{s}[t]\}$ and $\{\boldsymbol{x}[t]\}$. To begin with, let us recast the received signal (2) as a bilinear matrix factorization model

$$\boldsymbol{Y} = \boldsymbol{G}\boldsymbol{Z} + \boldsymbol{W} \tag{4}$$

where $\boldsymbol{Z} \triangleq \boldsymbol{S} \odot (\boldsymbol{H}\boldsymbol{X})$. This motivates us to propose a two-stage scheme: first estimate $\boldsymbol{G}$ and $\boldsymbol{Z}$, and then determine $\boldsymbol{H}$ based on the estimated $\boldsymbol{Z}$. It follows from the theories of matrix factorization and matrix completion that the proposed two-stage scheme is workable when $\boldsymbol{Z}$ is a sparse matrix and $\boldsymbol{H}$ is a low-rank matrix. The sparse structure of $\boldsymbol{Z}$ can be guaranteed when the pilot symbol $\boldsymbol{S}$ of the LIM is designed to be a sparse matrix with a large number of zero elements. The low-rank structure of $\boldsymbol{H}$ is assumed in the system model.

We now discuss the design of the pilot signals $\boldsymbol{S}$ and $\boldsymbol{X}$. As mentioned above, the pilot signal $\boldsymbol{S}$ of the LIM should be set to a sparse matrix. In this way, we generate $\{s_{t,n}\}$ independently from a Bernoulli distribution $\mathrm{Bernoulli}(\lambda)$ with (small) λ being the probability of taking the value of 1. The transmitted pilot signal $\boldsymbol{X}$ is designed to be a full-rank matrix, i.e., $\mathrm{rank}(\boldsymbol{X}) = M\,(M < T)$, which ensures the successful recovery of BS-LIM channel matrix $\boldsymbol{H}$ via matrix completion and will become clear in Section IV-C.

We now describe the algorithm design at the receiver side via two separate stages, i.e., the matrix factorization stage and the matrix completion stage. In the matrix factorization stage, we adopt the state-of-the-art BiG-AMP algorithm [6] to recover $\boldsymbol{G}$ and $\boldsymbol{Z}$ from $\boldsymbol{Y}$. Then, based on the estimated $\widehat{\boldsymbol{Z}}$ in the first stage and the pilot signal $\boldsymbol{S}$, the matrix completion stage is to retrieve $\boldsymbol{H}$ by using the available Riemannian manifold gradient-based algorithm [7] (based on the fact that $\boldsymbol{H}$ is rank-deficient). Such a two-stage channel estimation procedure is referred to as the joint bilinear factorization and matrix completion (JBF-MC) algorithm which is presented in Algorithm 1. The details of the JBF-MC algorithm are elaborated in the following subsections.

B. Sparse Matrix Factorization

Based on the observation $\boldsymbol{Y}$, we use the BiG-AMP algorithm [6] to approximately calculate the minimum mean-squared error estimates of $\boldsymbol{G}$ and $\boldsymbol{Z}$, i.e., the means of the marginal posteriors $\{p(g_{l,n}|\boldsymbol{Y})\}$ and $\{p(z_{n,t}|\boldsymbol{Y})\}$. Specifically, the BiG-AMP implements the sum-product loopy belief propagation over a factor graph induced by the posterior distribution

$$p(\boldsymbol{G}, \boldsymbol{Z}|\boldsymbol{Y}) \propto p(\boldsymbol{Y}|\boldsymbol{B}) p(\boldsymbol{G}) p(\boldsymbol{Z}) \tag{5}$$

where $\boldsymbol{B} \triangleq \boldsymbol{G}\boldsymbol{Z}$ and the likelihood $p(\boldsymbol{Y}|\boldsymbol{B})$ is given as

$$\begin{aligned} p(\boldsymbol{Y}|\boldsymbol{B}) &= \prod_{l=1}^{L}\prod_{t=1}^{T} p(y_{l,t}|b_{l,t}) \\ &= \prod_{l=1}^{L}\prod_{t=1}^{T} \exp\Big(-|y_{l,t} - b_{l,t}|^2/\sigma^2\Big). \end{aligned} \tag{6}$$

Algorithm 1 : JBF-MC Algorithm

Input: $\boldsymbol{Y}$, $\boldsymbol{S}$, $\boldsymbol{X}$, prior distributions $p(\boldsymbol{G})$ and $p(\boldsymbol{Z})$

% sparse matrix factorization via BiG-AMP

Initialization: $\forall l, n, t$: generate $\hat{g}_{l,n}(1)$ from $p(g_{l,n})$, $v^g_{l,n}(1) = \nu_g^2$, $\hat{z}_{n,t}(1) = \mathbb{E}\{z_{n,t}\}$, $v^z_{n,t}(1) = \lambda\nu_z^2$, and $\hat{u}_{l,t}(0) = 0$

for $i = 1, \ldots, I_{\max}$ % outer iteration

  for $j = 1, \ldots, J_{\max}$ % inner iteration

  $\forall l, t$: $\bar{v}^p_{l,t}(i) = \sum_{n=1}^{N} |\hat{g}_{l,n}(i)|^2 v^z_{n,t}(i) + v^g_{l,n}(i)|\hat{z}_{n,t}(i)|^2$

  $\forall l, t$: $\bar{p}_{l,t}(i) = \sum_{n=1}^{N} \hat{g}_{l,n}(i)\hat{z}_{n,t}(i)$

  $\forall l, t$: $v^p_{l,t}(i) = \bar{v}^p_{l,t}(i) + \sum_{n=1}^{N} v^g_{l,n}(i) v^z_{n,t}(i)$

  $\forall l, t$: $\hat{p}_{l,t}(i) = \bar{p}_{l,t}(i) - \hat{u}_{l,t}(i-1)\bar{v}^p_{l,t}(i)$

  $\forall l, t$: $v^b_{l,t}(i) = \sigma^2 v^p_{l,t}(i) / \big(v^p_{l,t}(i) + \sigma^2\big)$

  $\forall l, t$: $\hat{b}_{l,t}(i) = v^p_{l,t}(i)(y_{l,t} - \hat{p}_{l,t}(i))/(v^p_{l,t}(i) + \sigma^2) + \hat{p}_{l,t}(i)$

  $\forall l, t$: $v^u_{l,t}(i) = \big(1 - v^b_{l,t}(l)/v^p_{l,t}(i)\big)/v^p_{l,t}(i)$

  $\forall l, t$: $\hat{u}_{l,t}(i) = \big(\hat{b}_{l,t}(i) - \hat{p}_{l,t}(i)\big)/v^p_{l,t}(i)$

  $\forall l, n$: $v^q_{l,n}(i) = \big(\sum_{t=1}^{T} |\hat{z}_{n,t}(i)|^2 v^u_{l,t}(i)\big)^{-1}$

  $\forall l, n$: $\hat{q}_{l,n}(i) = \hat{g}_{l,n}(i)\big(1 - v^q_{l,n}(i)\sum_{t=1}^{T} v^z_{n,t}(i) v^u_{l,t}(i)\big) + v^q_{l,n}(i)\sum_{t=1}^{T} \hat{z}^*_{n,t}(i)\hat{u}_{l,t}(i)$

  $\forall n, t$: $v^r_{n,t}(i) = \big(\sum_{l=1}^{L} |\hat{g}_{l,n}(i)|^2 v^u_{l,t}(i)\big)^{-1}$

  $\forall n, t$: $\hat{r}_{n,t}(i) = \hat{z}_{n,t}(i)\big(1 - v^r_{n,t}(i)\sum_{l=1}^{L} v^g_{l,n}(i) v^u_{l,t}(i)\big) + v^r_{n,t}(i)\sum_{l=1}^{L} \hat{g}^*_{l,n}(i)\hat{u}_{l,t}(i)$

  $\forall l, n$: $\hat{g}_{l,n}(i+1) = \mathbb{E}\big\{g_{l,n}|\hat{q}_{l,n}(i), v^q_{l,n}(i)\big\}$

  $\forall l, n$: $v^g_{l,n}(i+1) = \mathbb{Var}\big\{g_{l,n}|\hat{q}_{l,n}(i), v^q_{l,n}(i)\big\}$

  $\forall n, t$: $\hat{z}_{n,t}(i+1) = \mathbb{E}\big\{z_{n,t}|\hat{r}_{n,t}(i), v^r_{n,t}(i)\big\}$

  $\forall n, t$: $v^z_{n,t}(i+1) = \mathbb{Var}\big\{z_{n,t}|\hat{r}_{n,t}(i), v^r_{n,t}(i)\big\}$

  if *a certain stopping criterion is met*, **stop**

  end for

  $\forall l, n, t, \hat{g}_{l,n}(i) = \hat{g}_{l,n}(i+1), v^g_{l,n}(i) = v^g_{l,n}(i+1)$, $\hat{z}_{n,t}(i) = \hat{z}_{n,t}(1), v^z_{n,t}(i) = v^z_{n,t}(1)$

end for

$\widehat{\boldsymbol{G}} \leftarrow \widehat{\boldsymbol{G}}(i+1)$, $\widehat{\boldsymbol{Z}} \leftarrow \widehat{\boldsymbol{Z}}(i+1)$

% matrix completion via RGrad

Initialization: $\boldsymbol{A}(0) = \boldsymbol{0}$

for $k = 1, \ldots, K_{\max}$

  $\boldsymbol{Q}(k) = \boldsymbol{S}^* \odot \big(\widehat{\boldsymbol{Z}} - \boldsymbol{A}(k)\big)$

  $\alpha(k) = \frac{\|\mathcal{P}_{\mathcal{S}(k)}(\boldsymbol{Q}(k))\|_F^2}{\|\boldsymbol{S}^* \odot (\mathcal{P}_{\mathcal{S}(k)}(\boldsymbol{Q}(k)))\|_F^2}$

  $\widetilde{\boldsymbol{W}}(k) = \boldsymbol{A}(k) + \alpha_k \mathcal{P}_{\mathcal{S}(k)}(\boldsymbol{Q}(k))$

  $\boldsymbol{A}(k+1) = \mathcal{H}_r\big(\widetilde{\boldsymbol{W}}(k)\big)$

  if *a certain stopping criterion is met*, **stop**

end for

$\widehat{\boldsymbol{H}} \leftarrow \widehat{\boldsymbol{A}}\boldsymbol{X}^{\dagger}$ with $\widehat{\boldsymbol{A}} = \boldsymbol{A}(k+1)$

Output: $\widehat{\boldsymbol{G}}$ and $\widehat{\boldsymbol{H}}$

We choose independent Gaussian priors for $\boldsymbol{G}$ and independent Bernoulli-Gaussian priors for $\boldsymbol{Z}$, i.e.,

$$p(\boldsymbol{G}) = \prod_{l=1}^{L}\prod_{n=1}^{N} p(g_{l,n}) = \prod_{l=1}^{L}\prod_{n=1}^{N} \mathcal{CN}(g_{l,n}; 0, \nu_g^2) \tag{7}$$

$$\begin{aligned} p(\boldsymbol{Z}) = \prod_{n=1}^{N}\prod_{t=1}^{T} p(z_{n,t}) = \prod_{n=1}^{N}\prod_{t=1}^{T} \Big[& s_{n,t}\,\mathcal{CN}(z_{n,t}; 0, \nu_z^2) \\ & + (1 - s_{n,t})\delta(z_{n,t})\Big] \end{aligned} \tag{8}$$

where ν_g^2 and ν_z^2 are respectively the average variances of the LIM-user channel matrix $\boldsymbol{G}$ and the non-zero elements of the sparse matrix $\boldsymbol{Z}$; the on/off state $s_{n,t} \in \{0,1\}$ at time t for the n-th reflect element of the LIM is known by the receiver because $\boldsymbol{S}$ is a pilot signal.

To achieve computational efficiency, the BiG-AMP leverages the central limit theorem and the Gaussian approximation for the involved messages. Details of the BiG-AMP algorithm can be found from Lines 1 to 24 of Algorithm 1. At the i-th iteration of the BiG-AMP, the means and variances of $\{g_{l,n}\}$ and $\{z_{n,t}\}$ from Lines 16 to 19 are respectively calculated with respect to the approximate marginal posterior distributions:

$$\hat{p}^{(i)}(g_{l,n}) \propto p(g_{l,n})\, \mathcal{CN}\Big(g_{l,n}; \hat{q}_{l,n}(i), v^{q}_{l,n}(i)\Big) \tag{9}$$

$$\hat{p}^{(i)}(z_{n,t}) \propto p(z_{n,t})\, \mathcal{CN}\Big(z_{n,t}; \hat{r}_{n,t}(i), v^{r}_{n,t}(i)\Big) \tag{10}$$

where $p(g_{l,n})$ and $p(z_{n,t})$ are the prior distributions defined in (7) and (8), respectively.

It is worth noting that the K-SVD (K-means singular value decomposition) [12] and the SPAMS (SPArse Modeling Software) [13] can be exploited to solve the bilinear sparse matrix factorization problem (4). However, these approaches perform much worse than the BiG-AMP algorithm, as will be seen from the numerical results presented in Section V. Additionally, note that the likelihood $p(\boldsymbol{Y}|\boldsymbol{B})$ and the priors $p(\boldsymbol{G})$ and $p(\boldsymbol{Z})$ are characterized by the parameters σ^2, ν_g^2, and ν_z^2, respectively. These parameters can be estimated by utilizing the standard expectation-maximization methodology [6], [14]. In this letter, we do not elaborate on how to estimate these parameters due to space limitation.

C. Matrix Completion

To facilitate the estimate of $\boldsymbol{H}$, we now recover the missing entries of $\widehat{\boldsymbol{Z}}$ by using the rank-deficient property of the original $\boldsymbol{H}$. We employ the Riemannian gradient (RGrad) algorithm [7] to solve the matrix completion problem, which is summarized from Lines 25 to 32 of Algorithm 1. The RGrad algorithm is to solve the following matrix completion problem:

$$\min_{\boldsymbol{A}} \ \frac{1}{2}\big\|\boldsymbol{S}^* \odot (\boldsymbol{A} - \widehat{\boldsymbol{Z}})\big\|_F^2 \ \text{ subject to } \operatorname{rank}(\boldsymbol{A}) = r. \tag{11}$$

In Lines 28 and 29 of the JBF-MC algorithm, $\mathcal{P}_{\mathcal{S}(k)}(\cdot)$ stands for the projection operation to the left singular vector subspace (denoted by $\mathcal{S}(k)$) of the current estimate $\boldsymbol{A}(k)$, corresponding to the first r eigenvalues of $\boldsymbol{A}(k)$. In Line 30 of Algorithm 1, $\mathcal{H}_r(\widetilde{\boldsymbol{W}})$ is the hard-thresholding operator for the best rank-r approximation of the associated SVD, i.e.,

$$\mathcal{H}_r\Big(\widetilde{\boldsymbol{W}}\Big) \triangleq \boldsymbol{U}\boldsymbol{\Sigma}_r\boldsymbol{V}^H,\ \Sigma_r(i,i) = \begin{cases} \Sigma(i,i), & i \le r \\ 0, & i > r \end{cases} \tag{12}$$

where $\widetilde{\boldsymbol{W}} \triangleq \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^H$ is the SVD of $\widetilde{\boldsymbol{W}}$ and $\Sigma(i,i)$ is the (i, i)-th entry of $\boldsymbol{\Sigma}$. Finally, the estimate of the channel matrix $\boldsymbol{H}$ can be computed as

$$\widehat{\boldsymbol{H}} = \widehat{\boldsymbol{A}}\boldsymbol{X}^{\dagger} \tag{13}$$

where $\boldsymbol{X}^{\dagger} = (\boldsymbol{X}\boldsymbol{X}^H)^{-1}\boldsymbol{X}$ is the Moore-Penrose inverse and $\widehat{\boldsymbol{A}}$ is output of the RGrad algorithm. Here, we assume that the pilot length T is no less than the number of transmit antennas M and rank $(\boldsymbol{X}) = M$, so as to ensure the existence of $\boldsymbol{X}^{\dagger}$. It is worth noting that the final estimates $\widehat{\boldsymbol{G}}$ and $\widehat{\boldsymbol{H}}$ have a diagonal ambiguity resulting from the sparse matrix factorization stage. Nevertheless, recall the discussions from Section III that there is no need to eliminate the diagonal ambiguity.

D. Computational Complexity

We now offer a brief discussion on the computational complexity of the proposed JBF-MC algorithm. Note that the total computational complexity of the JBF-MC consists of implementations of the BiG-AMP for matrix factorization and the RGrad for matrix completion and the computation involved in (13). We thus sketch the respective complexity as follows. First, the complexity of the BiG-AMP is dominated by basic matrix multiplications in Lines 4–6 and Lines 12–15 of Algorithm 1, requiring $\mathcal{O}(LNT)$ flops per iteration. Consequently, the complexity of the BiG-AMP is at most $I_{\max}J_{\max}\mathcal{O}(LNT)$ flops, where $I_{\max}$ and $J_{\max}$ are the maximum numbers of the outer and inner iterations of the BiG-AMP, respectively. Second, the complexity of the RGrad algorithm is dominated by the calculations in Lines 27 and 29, requiring $\mathcal{O}(rNT)$ and $\mathcal{O}(r^3)$ flops, respectively. Therefore, the total cost of the RGrad is at most $K_{\max}\mathcal{O}(rNT)$ flops by noticing that $r < \min\{N, T\}$, where $K_{\max}$ denotes the maximum number of iterations of the RGrad. Finally, the computation of (13) in Line 33 requires $\mathcal{O}(M^2T) + \mathcal{O}(M^3) + \mathcal{O}(MNT)$ flops.

V. Numerical Studies

We now carry out numerical experiments to corroborate the effectiveness of the proposed JBF-MC algorithm for the cascaded channel estimation of the LIM-assisted massive MIMO systems. We assume that the antenna elements form a half-wavelength uniform linear array (ULA) configuration at the BS, the LIM, and the receiver side. Following the superposition principle of different paths in the prorogation environment [9], [14], the baseband BS-LIM and LIM-user channel matrices $\boldsymbol{H}$ and $\boldsymbol{G}$ are respectively modeled as

$$\boldsymbol{H} = \sum_{k=1}^{K_h} \alpha_k\, \boldsymbol{a}_L(\vartheta_k)\boldsymbol{a}_T^H(\omega_k),\ \boldsymbol{G} = \sum_{k=1}^{K_g} \beta_k\, \boldsymbol{a}_R(\psi_k)\boldsymbol{a}_L^H(\vartheta_k)$$

where K_h and K_g stand for the number of propagation paths of radio signals in the BS-LIM channel and the LIM-user channel, respectively; $\boldsymbol{a}_L(\vartheta_k) \in \mathbb{C}^N$, $\boldsymbol{a}_T(\omega_k) \in \mathbb{C}^M$, and $\boldsymbol{a}_R(\psi_k) \in \mathbb{C}^L$ are the steering vectors of the ULA at the LIM, the BS, and the receiver side, respectively. In each trial, the angular parameters ψ_r, ϑ_r, and ω_k independently follow from the uniform distribution within (0, 1], and the path gain coefficients $\{\alpha_k\}$ and $\{\beta_k\}$ are independently drawn from $\mathcal{CN}(0,1)$. The number of paths in the channel matrix $\boldsymbol{H}$, i.e., K_h, is set to be enough small such that it has a low-rank structure to facilitate its estimation in the matrix completion stage. In addition, our empirical experiments suggest that choosing K_g to be a positive integer more than half of the receive antenna number L or the passive antenna number N in the LIM can lead to a better performance. The sampling matrix $\boldsymbol{S}$, i.e., the on/off states of the elements in the LIM, are set to be a random 0-1 matrix.

The pilot symbols in $\boldsymbol{X}$ are generated from $\mathcal{CN}(0,1)$ and the signal-to-noise ratio (SNR) is defined as $10\log_{10}(1/\sigma^2)$ dB. The estimation performance is evaluated in terms of the normalized mean-square-error (NMSE). All the simulation results are obtained by averaging 200 independent trials. Notice that the outputs $\widehat{\boldsymbol{G}}$ and $\widehat{\boldsymbol{H}}$ of the JBF-MC algorithm still contain diagonal ambiguities. These ambiguities are eliminated based on the true values of $\boldsymbol{G}$ and $\boldsymbol{H}$ in the calculation of the NMSEs. In the sparse matrix factorization stage, to benchmark

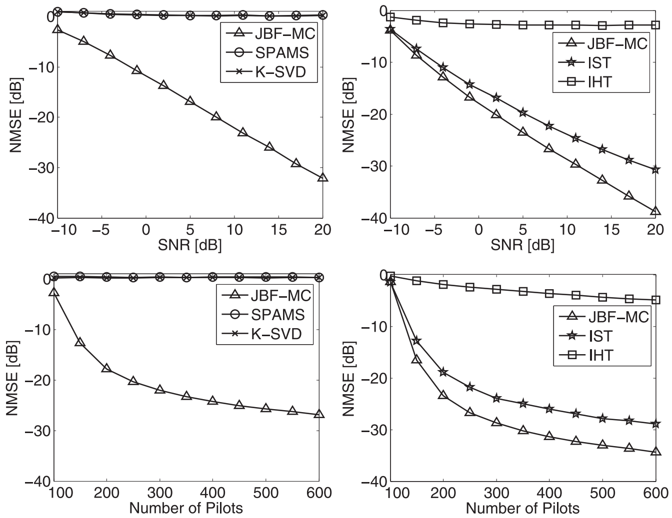

Fig. 2. NMSEs of $\boldsymbol{G}$ (left subplots) and $\boldsymbol{H}$ (right subplots) versus the SNR and the number of pilots with $N = 70$, $M = L = 64$, and $\lambda = 0.2$.

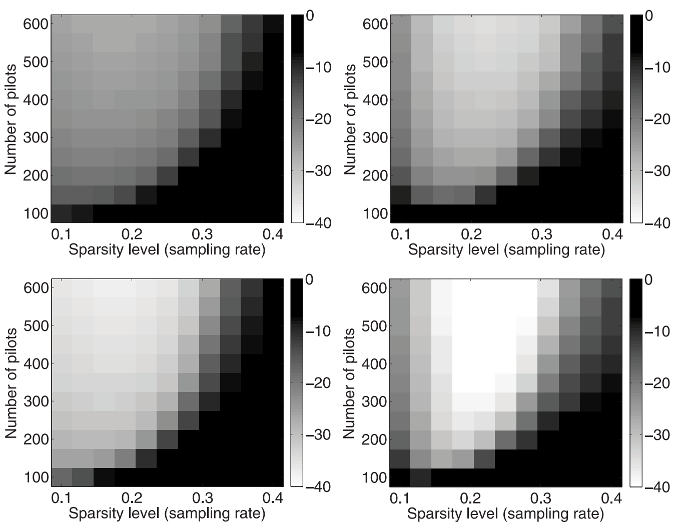

Fig. 3. Phase transitions (in NMSE) of $\boldsymbol{G}$ (left subplots) and $\boldsymbol{H}$ (right subplots) versus the sparsity level and the number of pilots. The upper subplots are for SNR $= 10$ dB and the lower subplots for SNR $= 20$ dB.

our JBF-MC algorithm for the estimation of $\boldsymbol{G}$, we adopt the K-SVD [12] and the SPAMS [13] for comparison. In all simulations, we set $N = 70$, $M = L = 64$, $K_h = r = 4$, and $K_g = 35$. For comparison, in the matrix completion stage we also adopt the IHT (iterative hard thresholding) and the IST (iterative soft thresholding) [15] to estimate the BS-LIM channel $\boldsymbol{H}$.

The NMSEs versus the SNR (with $T = 300$) and the number of pilots (with SNR $= 10$ dB) are depicted in Fig. 2 under the sparsity level (sampling rate) $\lambda = 0.2$. It is observed that the proposed JBF-MC algorithm has a significant performance gain over the baseline methods, especially for the estimation of $\boldsymbol{G}$. We also observe that the NMSE of $\boldsymbol{H}$ is smaller than that of $\boldsymbol{G}$. This is because $\boldsymbol{Z}$ is a sparse matrix with a much smaller number of non-zero variables to be estimated. Fig. 3 shows the phase transitions of the NMSEs of the channel estimation versus the sparsity level and the number of pilots. It is seen from Fig. 3 that there is a performance tradeoff between the sparse matrix factorization and the matrix completion: The matrix completion fails if λ is too small; the sparse matrix factorization fails if λ is too large. This is because when a larger number of samples (i.e., large λ) are available for matrix completion, the performance of the BiG-AMP algorithm becomes worse as a large number of non-zero variables are needed to be estimated in the sparse matrix factorization stage.

VI. Conclusion

In this letter, we considered the problem of cascaded channel estimation of the LIM-assisted MIMO systems. We introduced a general framework for this problem by leveraging a combined bilinear spare matrix factorization and matrix completion, and presented a two-stage algorithm that includes the bilinear generalized message passing algorithm for sparse matrix factorization and the Riemannian manifold gradient-based algorithm for matrix completion. We provided experimental evidences that the proposed approach achieves an accurate channel estimation for the LIM-assisted MIMO systems.

References

[1] C. Liaskos, S. Nie, A. Tsioliaridou, A. Pitsillides, S. Ioannidis, and I. Akyildiz, "A new wireless communication paradigm through software-controlled metasurfaces," *IEEE Commun. Mag.*, vol. 56, no. 9, pp. 162–169, Sep. 2018.

[2] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Trans. Wireless Commun.*, to be published.

[3] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4157–4170, Aug. 2019.

[4] T. Sleasman *et al.*, "Experimental synthetic aperture radar with dynamic metasurfaces," *IEEE Trans. Antennas Propag.*, vol. 65, no. 12, pp. 6864–6877, Dec. 2017.

[5] W. Yan, X. Kuai, and X. Yuan. (May 2019). *Passive Beamforming and Information Transfer via Large Intelligent Metasurface*. [Online]. Available: https://arxiv.org/abs/1905.01491

[6] J. T. Parker, P. Schniter, and V. Cevher, "Bilinear generalized approximate message passing—Part I: Derivation," *IEEE Trans. Signal Process.*, vol. 62, no. 22, pp. 5839–5853, Nov. 2014.

[7] K. Wei, J.-F. Cai, T. F. Chan, and S. Leung, "Guarantees of Riemannian optimization for low rank matrix recovery," *SIAM J. Matrix Anal. Appl.*, vol. 37, no. 3, pp. 1198–1222, Sep. 2016.

[8] A. Taha, M. Alrabeiah, and A. Alkhateeb. (Apr. 2019). *Enabling Large Intelligent Surfaces With Compressive Sensing and Deep Learning*. [Online]. Available: https://arxiv.org/abs/1904.10136

[9] J. Zhang, X. Yuan, and Y.-J. A. Zhang, "Blind signal detection in massive MIMO: Exploiting the channel sparsity," *IEEE Trans. Commun.*, vol. 66, no. 2, pp. 700–712, Feb. 2018.

[10] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, Jun. 2011.

[11] R. B. Green, "The general theory of antenna scattering," Ph.D. dissertation, Elect. Comput. Eng. Dept., Ohio State Univ., Columbus, OH, USA, 1963.

[12] M. Aharon, M. Elad, and A. M. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[13] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *J. Mach. Learn. Res.*, vol. 11, pp. 19–60, Mar. 2010.

[14] C. Huang, L. Liu, C. Yuen, and S. Sun, "Iterative channel estimation using LSE and sparse message passing for mmWave MIMO systems," *IEEE Trans. Signal Process.*, vol. 67, no. 1, pp. 245–259, Jan. 2019.

[15] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, Mar. 2010.