Information Geometry Approach for Ultra-Massive MIMO Signal Detection

Yan Chen[*†], Jiyuan Yang[*†], Xiqi Gao[*†], Dirk Slock[‡], and Xiang-Gen Xia[§]

[*]National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China
[†]Purple Mountain Laboratories, Nanjing 211111, China
[‡]Department of Communication Systems, EURECOM, 06410 Biot, France
[§]Department of Electrical and Computer Engineering, University of Delaware, Newark, DE 19716 USA
Email: {yan_chen, jyyang, xqgao}@seu.edu.cn, Dirk.Slock@eurecom.fr, xxia@ee.udel.edu

Abstract—In this paper, we propose an information geometry approach (IGA) for signal detection in ultra-massive multiple-input multiple-output (MIMO) systems. We formulate the signal detection as obtaining the marginals of the a posteriori probability distribution of the transmitted symbol vector. Then, a maximization of the a posteriori marginals (MPM) for signal detection can be performed. With the information geometry theory, we calculate the approximations of the a posteriori marginals. It is formulated as an iterative m-projection process. We then apply the central-limit-theorem (CLT) to simplify the calculation of the m-projection since the direct calculation of the m-projection is of exponential-complexity. With the CLT, we obtain an approximate solution of the m-projection, which is asymptotically accurate. Simulation results demonstrate that the proposed IGA is a promising and efficient method to implement the signal detector in ultra-massive MIMO systems.

Index Terms—Ultra-massive MIMO, signal detection, Bayesian inference, information geometry.

I. Introduction

Ultra-massive multiple-input multiple-output (MIMO) is able to achieve higher spectral efficiency and energy efficiency than massive MIMO, where the base station (BS) employs an ultra-large array with hundreds or thousands of antennas and serves tens or even hundreds of users simultaneously [1]–[3]. For the realization of the substantial benefits of ultra-massive MIMO, signal detection is of great importance. The optimal detector based on the maximum the a posteriori (MAP) criterion or the maximum-likelihood (ML) criterion performs an exhaustive search and examines all possible symbols, which is shown as non-deterministic polynomial-time hard (NP-hard). On the other hand, the linear detectors, e.g., the linear minimum-mean-squared error (LMMSE) detector, are widely adopted due to the polynomial-time complexity. Nonetheless, the estimation of the transmitted symbols of the LMMSE detector is biased [4], and the performance of the LMMSE detector degrades severely in massive MIMO systems with high-order constellations [5].

Information geometry, which is introduced by Rao [6], and then formally developed by Amari [7] and Cencov [8], has found a wide range of applications. For Bayesian inference, the space defined by the parameters of the a posteriori probability distribution is regarded as a differentiable manifold with a Riemannian structure and the definitions and tools of differential geometry are well applied by Amari et al. [9]. Amari et al. also show the intrinsic geometric insight of some classical Bayesian inference methods, e.g, the belief propagation (BP) [10]. Meanwhile, some optimization methods, such as the concave-convex procedure (CCCP) [11], are also applied to calculate the marginals of the a posteriori distribution. On Bayesian inference in communications applications, [12] analyzes the turbo and low-density parity-check (LDPC) codes from the perspective of information geometry, and an improvement of turbo and LDPC codes is proposed from the geometrical view. The information geometry is extended to complex signal processing and an information geometry approach (IGA) is proposed for massive MIMO channel estimation in [13] and [14].

In this paper, we propose an information geometry approach for ultra-massive MIMO signal detection. We formulate the signal detection as obtaining the marginals of the a posteriori probability distribution of the transmitted symbols. With the information geometry theory, we calculate the approximations of the a posteriori marginals. Furthermore, since the calculation of the m-projection in signal detection is of exponential complexity, we apply the central-limit-theorem (CLT) to simplify its calculation. With the CLT, we are able to find an approximate solution of the m-projection, which is asymptotically accurate. At last, a soft-decision is performed based on the approximation of the a posteriori marginals.

The rest of the paper proceeds as follows. The system configuration and problem statement are presented in Section II. The information geometry approach for ultra-massive MIMO signal detection is proposed in Section III. Simulation results are provided in Section IV. The conclusion is drawn in Section V.

II. System Model and Problem Statement

A. System Configuration

Consider an ultra-massive MIMO system with one base station (BS) serving K single-antenna users within a cell, where the BS comprises a uniform planar array (UPA) of $N_r = N_{r,v} \times N_{r,h}$ antennas, and $N_{r,v}$ and $N_{r,h}$ are the numbers of the antennas at each vertical column and horizontal row, respectively. Denote the transmitted symbol vector of all

users as $\tilde{\mathbf{s}} \triangleq [\tilde{s}_1, \tilde{s}_2, \ldots, \tilde{s}_K]^T \in \tilde{\mathbb{S}}^K$, where $\tilde{s}_k$ is the transmitted symbol of user k. $\tilde{\mathbb{S}}$ is the modulation constellation and let us assume $\tilde{\mathbb{S}} = \left\{ \tilde{s}^{(0)}, \tilde{s}^{(1)}, \ldots, \tilde{s}^{(\tilde{L}-1)} \right\}$, where $\left\{ \tilde{s}^{(\ell)} \right\}_{\ell=0}^{\tilde{L}-1}$ are the constellation points, and $\tilde{L}$ is the modulation order (or constellation size). In this paper, we focus on the $\tilde{L}$-QAM modulation and assume that each user chooses a symbol from $\tilde{\mathbb{S}}$ uniformly at random, and all users use the same alphabet. We also assume that the average power of $\tilde{s}_k$ is normalized to unit, i.e., $\mathbb{E}\left\{ |\tilde{s}_k|^2 \right\} = 1$, $k \in \mathcal{Z}_K^+$, where $\mathcal{Z}_K^+ = \{1, 2, ..., K\}$. The symbol vector $\tilde{\mathbf{s}}$ is then transmitted over a flat-fading complex channel, and the received signal $\tilde{\mathbf{y}} \in \mathbb{C}^{N_r}$ at the BS can be modeled as

$$\tilde{\mathbf{y}} = \tilde{\mathbf{G}}\tilde{\mathbf{s}} + \tilde{\mathbf{z}}, \tag{1}$$

where $\tilde{\mathbf{G}} \in \mathbb{C}^{N_r \times K}$ is the channel matrix, $\tilde{\mathbf{z}}$ is an additive white circular-symmetric complex Gaussian noise vector, $\tilde{\mathbf{z}} \sim \mathcal{CN}\left(\mathbf{0}, \tilde{\sigma}_z^2 \mathbf{I}\right)$ and $\tilde{\sigma}_z^2$ is the noise variance.

B. Problem Statement

We first reformulate the complex-valued received signal model (1) into a real-valued one. Define real vectors $\mathbf{y} \triangleq [\mathcal{R}^T\{\tilde{\mathbf{y}}\}, \mathcal{I}^T\{\tilde{\mathbf{y}}\}]^T \in \mathbb{R}^{2N_r}$, $\mathbf{s} \triangleq [\mathcal{R}^T\{\tilde{\mathbf{s}}\}, \mathcal{I}^T\{\tilde{\mathbf{s}}\}]^T \in \mathbb{R}^{2K}$, $\mathbf{z} \triangleq [\mathcal{R}^T\{\tilde{\mathbf{z}}\}, \mathcal{I}^T\{\tilde{\mathbf{z}}\}]^T \in \mathbb{R}^{2N_r}$ and a real matrix

$$\mathbf{G} \triangleq \begin{bmatrix} \mathcal{R}\{\tilde{\mathbf{G}}\}, & -\mathcal{I}\{\tilde{\mathbf{G}}\} \\ \mathcal{I}\{\tilde{\mathbf{G}}\}, & \mathcal{R}\{\tilde{\mathbf{G}}\} \end{bmatrix} \in \mathbb{R}^{2N_r \times 2K}. \tag{2}$$

Then, the real-valued received signal model is obtained as

$$\mathbf{y} = \mathbf{G}\mathbf{s} + \mathbf{z}, \tag{3}$$

where $\mathbf{s} = [s_1, s_2, \ldots, s_{2K}]^T \in \mathbb{S}^{2K}$, $s_k \in \mathbb{S}, k \in \mathcal{Z}_{2K}^+$, $\mathbb{S} = \left\{ s^{(0)}, s^{(1)}, \ldots, s^{(L-1)} \right\}$ is the alphabet for the real and imaginary components of a symmetric $\tilde{L}$-QAM modulation, $L = \sqrt{\tilde{L}}$, $\mathbf{z} \sim \mathcal{N}\left(\mathbf{0}, \sigma_z^2 \mathbf{I}\right)$ is the noise vector, and $\sigma_z^2 = \tilde{\sigma}_z^2/2$. Assuming that the transmitted signals of all users, as well as the real and imaginary parts of the transmitted signals, are independent of each other and follow a uniform distribution. Given the received signal model (3), the *a posteriori* distribution of $\mathbf{s}$ can be expressed as

$$p\left(\mathbf{s}|\mathbf{y}\right) \propto \prod_{k=1}^{2K} f_k\left(s_k\right) \prod_{n=1}^{2N_r} f\left(y_n|\mathbf{s}\right)$$
$$\propto \prod_{k=1}^{2K} f_k\left(s_k\right) \prod_{n=1}^{2N_r} \exp\left\{ -\frac{\left(y_n - \mathbf{e}_n^T\mathbf{G}\mathbf{s}\right)^2}{2\sigma_z^2} \right\}, \tag{4}$$

where $f_k\left(s_k\right)\big|_{s_k=s^{(\ell)}} = \frac{1}{L}, k \in \mathcal{Z}_{2K}^+, \ell \in \mathcal{Z}_{L-1}$, is the a priori probability of s_k, $\mathcal{Z}_{L-1} = \{0, 1, ..., L-1\}$, y_n is the n-th element of $\mathbf{y}$, $f\left(y_n|\mathbf{s}\right)$ is the probability density function (PDF) of $y_n, n \in \mathcal{Z}_{2N_r}^+$, given $\mathbf{s}$, and $\mathbf{e}_n \in \mathbb{C}^{2N_r}$ is the n-th column of the $2N_r$ dimensional identity matrix. In this work, we propose an information geometry approach for signal detection which aims to obtain the approximations of the marginals, i.e., $p_k\left(s_k|\mathbf{y}\right), k \in \mathcal{Z}_{2K}^+$, of the *a posteriori*

distribution $p\left(\mathbf{s}|\mathbf{y}\right)$, which can be used for the maximization of the *a posteriori* marginals (MPM) detector, i.e., for $k \in \mathcal{Z}_{2K}^+$,

$$s_{k,\mathrm{MPM}} = \arg\max_{s_k \in \mathbb{S}} p_k\left(s_k|\mathbf{y}\right). \tag{5}$$

III. INFORMATION GEOMETRY APPROACH FOR SIGNAL DETECTION

This section applies information geometry into the signal detection. Define a sufficient statistic of s_k as $\mathbf{t}_k \triangleq [t_{k,1}, t_{k,2}, \ldots, t_{k,L-1}]^T \in \mathbb{R}^{L-1}$, where $t_{k,\ell} \triangleq \delta\left(s_k - s^{(\ell)}\right), k \in \mathcal{Z}_{2K}^+, \ell \in \mathcal{Z}_{L-1}^+$, and $\delta(x)$ is equal to 1 when $x = 0$, otherwise 0. Define $\mathbf{d}_k \triangleq [d_{k,1}, d_{k,2}, \ldots, d_{k,L-1}]^T \in \mathbb{R}^{L-1}, k \in \mathcal{Z}_{2K}^+$, and

$$d_{k,\ell} = \ln \frac{f_k\left(s_k\right)\big|_{s_k=s^{(\ell)}}}{f_k\left(s_k\right)\big|_{s_k=s^{(0)}}}, \ell \in \mathcal{Z}_{L-1}^+. \tag{6}$$

The a priori probability $f_k\left(s_k\right), k \in \mathcal{Z}_{2K}^+$, can be expressed as

$$f_k\left(s_k\right) = \exp\left\{ \mathbf{d}_k^T \mathbf{t}_k - \psi\left(\mathbf{d}_k\right) \right\}, \tag{7}$$

where $\psi\left(\mathbf{d}_k\right) = -\ln\left(f_k\left(s_k\right)\big|_{s_k=s^{(0)}} \right)$ is the free energy. Define $c_n\left(\mathbf{s}, y_n\right) \triangleq -\frac{1}{2\sigma_z^2}\left(y_n - \mathbf{e}_n^T\mathbf{G}\mathbf{s}\right)^2$. Then the *a posteriori* distribution $p\left(\mathbf{s}|\mathbf{y}\right)$ in (4) can be expressed as

$$p\left(\mathbf{s}|\mathbf{y}\right) = \exp\left\{ \sum_{k=1}^{2K} \mathbf{d}_k^T \mathbf{t}_k + \sum_{n=1}^{2N_r} c_n\left(\mathbf{s}, y_n\right) - \psi_q \right\}$$
$$= \exp\left\{ \mathbf{d}^T \mathbf{t} + \sum_{n=1}^{2N_r} c_n\left(\mathbf{s}, y_n\right) - \psi_q \right\}, \tag{8}$$

where $\mathbf{d} = \left[\mathbf{d}_1^T, \mathbf{d}_2^T, \ldots, \mathbf{d}_{2K}^T \right]^T \in \mathbb{R}^{2K(L-1)}$, $\mathbf{t} = \left[\mathbf{t}_1^T, \mathbf{t}_2^T, \ldots, \mathbf{t}_{2K}^T \right]^T \in \mathbb{R}^{2K(L-1)}$, and ψ_q is the normalization factor. In (8), $\mathbf{d}^T\mathbf{t}$ only contains the separated random variables $\{s_k\}_{k=1}^{2K}$ (i.e., no cross-terms of $\{s_k\}_{k=1}^{2K}$), and all the interactions (cross-terms) between the random variables $\{s_k\}_{k=1}^{2K}$ are included in $c_n\left(\mathbf{s}, y_n\right), n \in \mathcal{Z}_{2N_r}^+$. IGA aims to approximate $\sum_{n=1}^{2N_r} c_n\left(\mathbf{s}, y_n\right)$ by $\boldsymbol{\theta}_0^T \mathbf{t}$, where $\boldsymbol{\theta}_0 \in \mathbb{R}^{2K(L-1)}$. To obtain $\boldsymbol{\theta}_0$, we construct two types of manifolds, the objective manifold (OBM) and the auxiliary manifold (AM). The OBM $\mathcal{M}_0$ is defined as

$$\mathcal{M}_0 = \left\{ p_0\left(\mathbf{s}; \boldsymbol{\theta}_0\right) \big| \boldsymbol{\theta}_0 \in \mathbb{R}^{2K(L-1)} \right\}, \tag{9a}$$

$$p_0\left(\mathbf{s}; \boldsymbol{\theta}_0\right) = \prod_{k=1}^{2K} p_{0,k}\left(s_k; \boldsymbol{\theta}_{0,k}\right)$$
$$= \exp\left\{ \mathbf{d}^T\mathbf{t} + \boldsymbol{\theta}_0^T\mathbf{t} - \psi_0\left(\boldsymbol{\theta}_0\right) \right\}, \tag{9b}$$

$$p_{0,k}\left(s_k; \boldsymbol{\theta}_{0,k}\right) = \exp\left\{ \mathbf{d}_k^T\mathbf{t}_k + \boldsymbol{\theta}_{0,k}^T\mathbf{t}_k - \psi_0\left(\boldsymbol{\theta}_{0,k}\right) \right\}, \tag{9c}$$

where $\boldsymbol{\theta}_0 = \left[\boldsymbol{\theta}_{0,1}^T, \boldsymbol{\theta}_{0,2}^T, \ldots, \boldsymbol{\theta}_{0,2K}^T \right]^T \in \mathbb{R}^{2K(L-1)}$ is the e-affine coordinate system (EACS) of $p_0\left(\mathbf{s}; \boldsymbol{\theta}_0\right)$, $\boldsymbol{\theta}_{0,k} = \left[\theta_{0,k,1}, \theta_{0,k,2}, \ldots, \theta_{0,k,L-1} \right]^T \in \mathbb{R}^{(L-1)}$ is the EACS of $p_{0,k}\left(s_k; \boldsymbol{\theta}_{0,k}\right)$, $p_{0,k}\left(s_k; \boldsymbol{\theta}_{0,k}\right)$ is the marginal distribution of

s_k, and $\psi_0\left(\boldsymbol{\theta}_0\right)$ and $\psi_0\left(\boldsymbol{\theta}_{0,k}\right)$ are normalization factors. Specifically, $\psi_0\left(\boldsymbol{\theta}_{0,k}\right)$ is given by

$$
\begin{aligned}
\psi_0\left(\boldsymbol{\theta}_{0,k}\right) &= \ln\left(\sum_{s_k \in \mathbb{S}} \exp\left\{\mathbf{d}_k^T \mathbf{t}_k + \boldsymbol{\theta}_{0,k}^T \mathbf{t}_k\right\}\right) \\
&= \ln\left(1 + \sum_{\ell=1}^{L-1} \exp\left\{d_{k,\ell} + \theta_{0,k,\ell}\right\}\right).
\end{aligned}
\tag{10}
$$

Given $p_0\left(\mathbf{s};\boldsymbol{\theta}_0\right)$ and its marginals $p_{0,k}\left(s_k;\boldsymbol{\theta}_{0,k}\right)$, the probability of the signal s_k can be expressed in a more explicit way as

$$
p_{0,k}\left(s_k;\boldsymbol{\theta}_{0,k}\right)\Big|_{s_k=s^{(0)}} \overset{(a)}{=} \frac{1}{1 + \sum_{\ell=1}^{L-1} \exp\left\{d_{k,\ell} + \theta_{0,k,\ell}\right\}},
\tag{11a}
$$

$$
p_{0,k}\left(s_k;\boldsymbol{\theta}_{0,k}\right)\Big|_{s_k=s^{(\ell)}} \overset{(b)}{=} \frac{\exp\left\{d_{k,\ell} + \theta_{0,k,\ell}\right\}}{1 + \sum_{\ell=1}^{L-1} \exp\left\{d_{k,\ell} + \theta_{0,k,\ell}\right\}},
\tag{11b}
$$

where $\ell \in \mathcal{Z}_{L-1}^+$ in (11b), and (a) and (b) come from (9c) and (10). Then $2N_r$ AMs are defined as, for $n \in \mathcal{Z}_{2N_r}^+$,

$$
\mathcal{M}_n = \left\{p_n\left(\mathbf{s};\boldsymbol{\theta}_n\right)\Big|\boldsymbol{\theta}_n \in \mathbb{R}^{2K(L-1)}\right\},
\tag{12a}
$$

$$
p_n\left(\mathbf{s};\boldsymbol{\theta}_n\right) = \exp\left\{\mathbf{d}^T\mathbf{t} + \boldsymbol{\theta}_n^T\mathbf{t} + c_n\left(\mathbf{s},y_n\right) - \psi_n\left(\boldsymbol{\theta}_n\right)\right\},
\tag{12b}
$$

where $\boldsymbol{\theta}_n = \left[\boldsymbol{\theta}_{n,1}^T, \boldsymbol{\theta}_{n,2}^T, \ldots, \boldsymbol{\theta}_{n,2K}^T\right]^T \in \mathbb{R}^{2K(L-1)}$ is the EACS of $p_n\left(\mathbf{s};\boldsymbol{\theta}_n\right)$, $\boldsymbol{\theta}_{n,k} = \left[\theta_{n,k,1}, \theta_{n,k,2}, \ldots, \theta_{n,k,L-1}\right]^T \in \mathbb{R}^{(L-1)}$, and $\psi_n(\boldsymbol{\theta}_n)$ is the normalization factor.

Before proceeding, we further define a manifold called the original manifold (OM), and then show that the OBM and the AMs are its submanifolds. Define the OM as the set of probability distributions of the $2K$ dimensional discrete random vector $\mathbf{s}$ as

$$
\mathcal{S} = \left\{p\left(\mathbf{s}\right)\Big|p\left(\mathbf{s}\right) > 0, \mathbf{s} \in \mathbb{S}^{2K}, \sum_{\mathbf{s}\in\mathbb{S}^{2K}} p\left(\mathbf{s}\right) = 1\right\}.
\tag{13}
$$

$\mathcal{S}$ is then a $L^{2K} - 1$ dimensional manifold and forms an exponential family. It can be obtained that the OBM and the AMs are the submanifolds of the OM, i.e., $\mathcal{M}_0 \subseteq \mathcal{S}$, $\mathcal{M}_n \subseteq \mathcal{S}, n \in \mathcal{Z}_{2N_r}^+$, since the distributions in the OBM and the AMs are all particular probability distributions of $\mathbf{s}$ when the EACSs of them are given.

From (12b), it can be found that only one interaction term $c_n\left(\mathbf{s},y_n\right)$ is remained in $p_n\left(\mathbf{s};\boldsymbol{\theta}_n\right)$, and all the others, i.e., $\sum_{n'\neq n} c_{n'}\left(\mathbf{s},y_{n'}\right)$, are replaced as $\boldsymbol{\theta}_n^T\mathbf{t}$. Assume that $\boldsymbol{\theta}_n$ of $p_n\left(\mathbf{s};\boldsymbol{\theta}_n\right), n \in \mathcal{Z}_{2N_r}^+$, is given, we calculate the approximation of $c_n\left(\mathbf{s},y_n\right)$ from the m-projection of $p_n\left(\mathbf{s};\boldsymbol{\theta}_n\right)$ onto the OBM $\mathcal{M}_0$. Denote the approximation of $c_n\left(\mathbf{s},y_n\right)$ as $\boldsymbol{\xi}_n^T\mathbf{t}$, where $\boldsymbol{\xi}_n = \left[\boldsymbol{\xi}_{n,1}^T, \boldsymbol{\xi}_{n,2}^T, \ldots, \boldsymbol{\xi}_{n,2K}^T\right]^T \in \mathbb{R}^{2K(L-1)}$ and $\boldsymbol{\xi}_{n,k} = \left[\xi_{n,k,1}, \xi_{n,k,2}, \ldots, \xi_{n,k,L-1}\right]^T \in \mathbb{R}^{(L-1)}, k \in \mathcal{Z}_{2K}^+, n \in \mathcal{Z}_{2N_r}^+$. Then the m-projection of $p_n\left(\mathbf{s};\boldsymbol{\theta}_n\right)$ onto $\mathcal{M}_0$ is denoted as $p_0\left(\mathbf{s};\boldsymbol{\theta}_n + \boldsymbol{\xi}_n\right)$. Specifically, the m-projection is calculated by minimizing the following Kullback-Leibler (K-L) divergence,

$$
\boldsymbol{\xi}_n = \arg\min_{\boldsymbol{\xi}_n} D_{\mathrm{KL}}\left\{p_n\left(\mathbf{s};\boldsymbol{\theta}_n\right) : p_0\left(\mathbf{s};\boldsymbol{\theta}_n + \boldsymbol{\xi}_n\right)\right\},
\tag{14}
$$

where the K-L divergence is given by

$$
\begin{aligned}
&D_{\mathrm{KL}}\left\{p_n\left(\mathbf{s};\boldsymbol{\theta}_n\right) : p_0\left(\mathbf{s};\boldsymbol{\theta}_n + \boldsymbol{\xi}_n\right)\right\} \\
&= \mathbb{E}_{p_n(\mathbf{s};\boldsymbol{\theta}_n)}\left\{\ln\frac{p_n\left(\mathbf{s};\boldsymbol{\theta}_n\right)}{p_0\left(\mathbf{s};\boldsymbol{\theta}_n + \boldsymbol{\xi}_n\right)}\right\}.
\end{aligned}
\tag{15}
$$

We now present the properties of the m-projection of any $p\left(\mathbf{s}\right) \in \mathcal{S}$, onto the OBM $\mathcal{M}_0$, which inspires us to approximate the m-projection of $p_n\left(\mathbf{s};\boldsymbol{\theta}_n\right)$ onto the OBM $\mathcal{M}_0$. We then have the following theorem.

Theorem 1: Given $p\left(\mathbf{s}\right) \in \mathcal{S}$, and $\mathcal{M}_0 \subseteq \mathcal{S}$, the m-projection of $p\left(\mathbf{s}\right)$ onto $\mathcal{M}_0$ is unique. Moreover, $p_0\left(\mathbf{s};\boldsymbol{\theta}_0^\star\right)$ is the m-projection of $p\left(\mathbf{s}\right)$ onto $\mathcal{M}_0$ if and only if the following relationship holds,

$$
\boldsymbol{\eta} = \boldsymbol{\eta}_0\left(\boldsymbol{\theta}_0^\star\right),
\tag{16}
$$

where $\boldsymbol{\eta}, \boldsymbol{\eta}_0\left(\boldsymbol{\theta}_0^\star\right) \in \mathbb{R}^{2K(L-1)}$ are the expectations of $\mathbf{t}$ w.r.t. $p\left(\mathbf{s}\right)$ and $p_0\left(\mathbf{s};\boldsymbol{\theta}_0^\star\right)$, respectively.

Define $\mathbf{s}_{\backslash k}$ as the $(2K-1)$-dimensional vector obtained by removing the k-th element from $\mathbf{s}$, $k \in \mathcal{Z}_{2K}^+$. Then, we can obtain the marginal probability distribution of s_k given the joint probability distribution $p\left(\mathbf{s}\right)$ is

$$
p_k\left(s_k\right) \triangleq \sum_{\mathbf{s}_{\backslash k} \in \mathbb{S}^{2K-1}} p\left(\mathbf{s}\right), k \in \mathcal{Z}_{2K}^+.
\tag{17}
$$

From the definition of $p_0\left(\mathbf{s};\boldsymbol{\theta}_0^\star\right)$ in (9b), we denote the marginals of $p_0\left(\mathbf{s};\boldsymbol{\theta}_0^\star\right)$ in Theorem 1 as $p_{0,k}\left(s_k;\boldsymbol{\theta}_{0,k}^\star\right), k \in \mathcal{Z}_{2K}^+$, where $\boldsymbol{\theta}_{0,k}^\star = \left[\theta_{0,k,1}^\star, \theta_{0,k,2}^\star, \ldots, \theta_{0,k,L-1}^\star\right]^T$ and $\boldsymbol{\theta}_0^\star = \left[\left(\boldsymbol{\theta}_{0,1}^\star\right)^T, \left(\boldsymbol{\theta}_{0,2}^\star\right)^T, \ldots, \left(\boldsymbol{\theta}_{0,2K}^\star\right)^T\right]^T$. From Theorem 1, we have the following corollary.

Corollary 1: Given $p\left(\mathbf{s}\right) \in \mathcal{S}$, and $\mathcal{M}_0 \subseteq \mathcal{S}$, $p_0\left(\mathbf{s};\boldsymbol{\theta}_0^\star\right)$ is the m-projection of $p\left(\mathbf{s}\right)$ onto $\mathcal{M}_0$ if and only if the marginals of $p\left(\mathbf{s}\right)$ and the marginals of $p_0\left(\mathbf{s};\boldsymbol{\theta}_0^\star\right)$ are equal, i.e.,

$$
p_k\left(s_k\right) = p_{0,k}\left(s_k;\boldsymbol{\theta}_{0,k}^\star\right), s_k \in \mathbb{S}, k \in \mathcal{Z}_{2K}^+.
\tag{18}
$$

Meanwhile, the EACS of the m-projection is given by $\boldsymbol{\theta}_0^\star = \left[\left(\boldsymbol{\theta}_{0,1}^\star\right)^T, \left(\boldsymbol{\theta}_{0,2}^\star\right)^T, \ldots, \left(\boldsymbol{\theta}_{0,2K}^\star\right)^T\right]^T$, where $\boldsymbol{\theta}_{0,k}^\star = \left[\theta_{0,k,1}^\star, \theta_{0,k,2}^\star, \ldots, \theta_{0,k,L-1}^\star\right]^T$, and

$$
\theta_{0,k,\ell}^\star = \ln\frac{p_k\left(s_k\right)\big|_{s_k=s^{(\ell)}}}{p_k\left(s_k\right)\big|_{s_k=s^{(0)}}} - d_{k,\ell}, \ell \in \mathcal{Z}_{L-1}^+.
\tag{19}
$$

The proofs of Theorem 1 and Corollary 1 can be found in [15]. From Corollary 1, for any $n \in \mathcal{Z}_{2N_r}^+$, the m-projection $p_0\left(\mathbf{s};\boldsymbol{\theta}_n + \boldsymbol{\xi}_n\right)$ is determined by the marginal probability distribution $p_{n,k}\left(s_k;\boldsymbol{\theta}_n\right), k \in \mathcal{Z}_{2K}^+$, where

$$
p_{n,k}\left(s_k;\boldsymbol{\theta}_n\right) = \sum_{\mathbf{s}_{\backslash k} \in \mathbb{S}^{2K-1}} p_n\left(\mathbf{s};\boldsymbol{\theta}_n\right).
\tag{20}
$$

And we have

$$
\xi_{n,k,\ell} = \ln\frac{p_{n,k}\left(s_k;\boldsymbol{\theta}_n\right)\big|_{s_k=s^{(\ell)}}}{p_{n,k}\left(s_k;\boldsymbol{\theta}_n\right)\big|_{s_k=s^{(0)}}} - d_{k,\ell} - \theta_{n,k,\ell},
\tag{21}
$$

where $n \in \mathcal{Z}_{2N_r}^+, k \in \mathcal{Z}_{2K}^+$ and $\ell \in \mathcal{Z}_{L-1}^+$. From the definition of $p_n\,(\mathbf{s};\boldsymbol{\theta}_n)$ in (12b), its marginals can be expressed as

$$p_{n,k}\,(s_k;\boldsymbol{\theta}_n) = \sum_{\mathbf{s}_{\backslash k} \in \mathbb{S}^{2K-1}} \exp\left\{ (\mathbf{d}+\boldsymbol{\theta}_n)^T\,\mathbf{t} + c_n\,(\mathbf{s},y_n) - \psi_n \right\}$$
$$\overset{(a)}{\propto} \exp\left\{ (\mathbf{d}_k + \boldsymbol{\theta}_{n,k})^T\,\mathbf{t}_k \right\} q\,(y_n, s_k), \qquad (22)$$

where $n \in \mathcal{Z}_{2N_r}^+$, $k \in \mathcal{Z}_{2K}^+$, $s_k \in \mathbb{S}$, (a) is obtained by removing the constants that do not vary with the value of s_k, $q\,(y_n, s_k)$ is a function of y_n and s_k, and

$$q\,(y_n, s_k) \qquad (23)$$
$$= \sum_{\mathbf{s}_{\backslash k} \in \mathbb{S}^{2K-1}} \exp\left\{ \sum_{k'=1, k' \neq k}^{2K} (\mathbf{d}_{k'} + \boldsymbol{\theta}_{n,k'})^T\,\mathbf{t}_{k'} + c_n\,(\mathbf{s}, y_n) \right\}.$$

Note that the proportions in the second line of (22) and the third line of (24) next will not affect the calculation of $p_{n,k}(s_k;\boldsymbol{\theta}_n)$ since the constants corresponding to these proportions do not vary with the value of s_k, and thus we can finally normalize $p_{n,k}(s_k;\boldsymbol{\theta}_n)$. In the last line of (22), the calculation of $\exp\left\{ (\mathbf{d}_k + \boldsymbol{\theta}_{n,k})^T\,\mathbf{t}_k \right\}$ is simple, if we can obtain the approximate value of $q\,(s_k, y_n), s_k \in \mathbb{S}$, we can then obtain the approximate value of $p_{n,k}\,(s_k;\boldsymbol{\theta}_n), s_k \in \mathbb{S}$. Hence, our goal now is converted to obtain the approximate value of $q\,(y_n, s_k), s_k \in \mathbb{S}$. From (23), we can obtain

$$q\,(y_n, s_k)$$
$$= \sum_{\mathbf{s}_{\backslash k} \in \mathbb{S}^{2K-1}} \left(\prod_{k'=1, k' \neq k}^{2K} \exp\left\{ (\mathbf{d}_{k'} + \boldsymbol{\theta}_{n,k'})^T\,\mathbf{t}_{k'} \right\} \right.$$
$$\left. \times \exp\left\{ -\frac{1}{2\sigma_z^2} \left(y_n - \mathbf{e}_n^T\mathbf{G}\mathbf{s} \right)^2 \right\} \right) \qquad (24)$$
$$\overset{(a)}{\propto} \sum_{\mathbf{s}_{\backslash k} \in \mathbb{S}^{2K-1}} \left(\prod_{k'=1, k' \neq k}^{2K} p_{0,k'}\,(s_{k'};\boldsymbol{\theta}_{n,k'})\, f_G\,(y_n; \mathbf{e}_n^T\mathbf{G}\mathbf{s}, \sigma_z^2) \right),$$

where $\mathbf{G}$ is defined in (2), (a) is obtained by adding the constants independent with s_k and y_n, $p_{0,k'}\,(s_{k'};\boldsymbol{\theta}_{n,k'})$ is define by (9c) and $f_G\,(x;\mu,\sigma^2)$ denotes the PDF of a real Gaussian distribution $\mathcal{N}(\mu,\sigma^2)$. Inspired by the last line of (24), we consider $2N_r \times 2K$ hybrid random variables $Y_{n,k}, n \in \mathcal{Z}_{2N_r}^+, k \in \mathcal{Z}_{2K}^+$, where the (n,k)-th of them is defined by: for given s_k,

$$Y_{n,k} = \mathbf{e}_n^T\mathbf{G}\mathbf{s} + w = g_{n,k}s_k + \sum_{k'=1, k' \neq k}^{2K} g_{n,k'}s_{k'} + w$$
$$= \sum_{k'=1, k' \neq k}^{2K} g_{n,k'}s_{k'} + w'_{n,k}, \qquad (25)$$

where $g_{n,k}$ is the (n,k)-th element of $\mathbf{G}$, s_k is considered as a deterministic (also known/given) constant, $\{s_{k'}\}_{k' \neq k}$ are considered as the independent discrete random variables, the probability distribution of $s_{k'}, k' \neq k$, is given by $p_{0,k'}\,(s_{k'};\boldsymbol{\theta}_{n,k'})$, the joint probability distribution of $\{s_{k'}\}_{k' \neq k}$ is then given by $p\,(\mathbf{s}_{\backslash k}) = \prod_{k' \neq k} p_{0,k'}\,(s_{k'};\boldsymbol{\theta}_{n',k'})$, $w \sim \mathcal{N}\,(0,\sigma_z^2)$ is a real

Gaussian random variable independent with $\{s_{k'}\}_{k' \neq k}$, and $w'_{n,k} = w + g_{n,k}s_k \sim \mathcal{N}\,(g_{n,k}s_k, \sigma_z^2)$ is also independent with $\{s_{k'}\}_{k' \neq k}$. In this case, it is not difficult to obtain that the PDF of $Y_{n,k}$ is given by [16, Sec. 6.1.2]

$$f\,(Y_{n,k})$$
$$= \sum_{\mathbf{s}_{\backslash k} \in \mathbb{S}^{2K-1}} \left(p\,(\mathbf{s}_{\backslash k})\, f_G\left(Y_{n,k} - \sum_{k' \neq k} g_{n,k'}s_{k'}; g_{n,k}s_k, \sigma_z^2 \right) \right)$$
$$= \sum_{\mathbf{s}_{\backslash k} \in \mathbb{S}^{2K-1}} \left(p\,(\mathbf{s}_{\backslash k})\, f_G\,(Y_{n,k}; \mathbf{e}_n^T\mathbf{G}\mathbf{s}, \sigma_z^2) \right), \qquad (26)$$

which will be equal to the last line of (24) after we set the value of $Y_{n,k}$ as $Y_{n,k} = y_n$.

Before proceeding, we first calculate the expected value and variance of $Y_{n,k}, n \in \mathcal{Z}_{2N_r}^+, k \in \mathcal{Z}_{2K}^+$, in (25). Given a probability distribution $p_{0,k'}\,(s_{k'};\boldsymbol{\theta}_{n,k'})$ of $s_{k'}, k' \in \mathcal{Z}_{2K}^+ \setminus \{k\}$, by using (11) the expected value and the variance of $s_{k'}$ are given by

$$\mu_{n,k'} = \sum_{s_{k'} \in \mathbb{S}} s_{k'} p_{0,k'}\,(s_{k'};\boldsymbol{\theta}_{n,k'})$$
$$= \frac{s^{(0)} + \sum_{\ell=1}^{L-1} s^{(\ell)} \exp\{d_{k',\ell} + \theta_{n,k',\ell}\}}{1 + \sum_{\ell=1}^{L-1} \exp\{d_{k',\ell} + \theta_{n,k',\ell}\}}, \qquad (27a)$$

$$v_{n,k'} = \sum_{s_{k'} \in \mathbb{S}} s_{k'}^2 p_{0,k'}\,(s_{k'};\boldsymbol{\theta}_{n,k'}) - \mu_{n,k'}^2 \qquad (27b)$$
$$= \frac{\left(s^{(0)}\right)^2 + \sum_{\ell=1}^{L-1} \left(s^{(\ell)}\right)^2 \exp\{d_{k',\ell} + \theta_{n,k',\ell}\}}{1 + \sum_{\ell=1}^{L-1} \exp\{d_{k',\ell} + \theta_{n,k',\ell}\}} - \mu_{n,k'}^2$$

Meanwhile, since $\{s_{k'}\}_{k' \neq k}$ and $w'_{n,k}$ are independent in (25), the expected value and variance of $Y_{n,k}, n \in \mathcal{Z}_{2N_r}^+, k \in \mathcal{Z}_{2K}^+$, can be readily expressed as

$$\mathbb{E}\,\{Y_{n,k}\} = \sum_{k'=1, k' \neq k}^{2K} g_{n,k'}\mu_{n,k'} + g_{n,k}s_k, \qquad (28a)$$

$$\mathbb{V}\,\{Y_{n,k}\} = \sum_{k'=1, k' \neq k}^{2K} g_{n,k'}^2 v_{n,k'} + \sigma_z^2. \qquad (28b)$$

From Lyapunov CLT [17], we then have the following theorem, and its detailed proof can be found in [15].

Theorem 2: If the following condition

$$\lim_{K \to \infty} \frac{1}{2K} \sum_{k'=1, k' \neq k}^{2K} g_{n,k'}^2 v_{n,k'} = \zeta > 0 \qquad (29)$$

holds for a positive constant ζ, then $Y_{n,k}$ converges in distribution to a real Gaussian random variable $\tilde{Y}_{n,k}$, as $2K$ goes to infinity, and

$$Y_{n,k} \overset{d}{\to} \tilde{Y}_{n,k} \sim \mathcal{N}\,(\mathbb{E}\,\{Y_{n,k}\}, \mathbb{V}\,\{Y_{n,k}\}). \qquad (30)$$

Intuitively, the condition (29) means that as K tends to infinity, the variance of the random variable $\tilde{s}_{n,k'} \triangleq g_{n,k'}s_{k'}, k' \in \mathcal{Z}_{2K}^+ \setminus \{k\}$, in (25) does not tend to zero, or $\tilde{s}_{n,k'}$ does not tend to be a deterministic value. This guarantees that the CLT holds.

When K is large, from Theorem 2, $q\left(y_n, s_k\right)$ is approximately proportional to $f_{\mathrm{G}}\left(\tilde{Y}_{n,k}; \mathbb{E}\left\{Y_{n,k}\right\}, \mathbb{V}\left\{Y_{n,k}\right\}\right)\big|_{\tilde{Y}_{n,k}=y_n}$, and thus we can obtain

$$
\begin{aligned}
&p_{n,k}\left(s_k; \boldsymbol{\theta}_n\right) \\
&\overset{(a)}{\propto} \exp\left\{\left(\mathbf{d}_k + \boldsymbol{\theta}_{n,k}\right)^T \mathbf{t}_k - \frac{\left(y_n - \mathbb{E}\left\{Y_{n,k}\right\}\right)^2}{2\mathbb{V}\left\{Y_{n,k}\right\}}\right\} \\
&= \exp\left\{\left(\mathbf{d}_k + \boldsymbol{\theta}_{n,k}\right)^T \mathbf{t}_k - \frac{\left(g_{n,k}s_k - \tilde{\mu}_{n,k}\right)^2}{2\mathbb{V}\left\{Y_{n,k}\right\}}\right\},
\end{aligned}
\tag{31}
$$

where $s_k \in \mathbb{S}$, $k \in \mathcal{Z}_{2K}^+$, $n \in \mathcal{Z}_{2N_r}^+$, (a) is obtained by removing the constants independent with s_k and y_n, and $\tilde{\mu}_{n,k}, n \in \mathcal{Z}_{2N_r}^+, k \in \mathcal{Z}_{2K}^+$, is defined as

$$
\tilde{\mu}_{n,k} \triangleq y_n - \sum_{k'=1,k'\neq k}^{2K} g_{n,k'}\mu_{n,k'}.
\tag{32}
$$

As a summary, when K is large we approximately have

$$
p_{n,k}\left(s_k; \boldsymbol{\theta}_n\right)\big|_{s_k=s^{(0)}} = C_{n,k}\exp\left\{-\frac{\left(g_{n,k}s^{(0)} - \tilde{\mu}_{n,k}\right)^2}{2\mathbb{V}\left\{Y_{n,k}\right\}}\right\},
\tag{33a}
$$

$$
\begin{aligned}
&p_{n,k}\left(s_k; \boldsymbol{\theta}_n\right)\big|_{s_k=s^{(\ell)}} \\
&= C_{n,k}\exp\left\{d_{k,\ell} + \theta_{n,k,\ell} - \frac{\left(g_{n,k}s^{(\ell)} - \tilde{\mu}_{n,k}\right)^2}{2\mathbb{V}\left\{Y_{n,k}\right\}}\right\},
\end{aligned}
\tag{33b}
$$

where $C_{n,k}$ is the normalization factor, and $\ell \in \mathcal{Z}_{L-1}^+$ in (33b). Combining (21), we can immediately obtain that

$$
\xi_{n,k,\ell} = \frac{g_{n,k}\left(s^{(0)} - s^{(\ell)}\right)\left[g_{n,k}\left(s^{(0)} + s^{(\ell)}\right) - 2\tilde{\mu}_{n,k}\right]}{2\mathbb{V}\left\{Y_{n,k}\right\}},
\tag{34}
$$

where $n \in \mathcal{Z}_{2N_r}^+$, $k \in \mathcal{Z}_{2K}^+$ and $\ell \in \mathcal{Z}_{L-1}^+$.

After the approximate $\boldsymbol{\xi}_n$ is obtained, we update $\boldsymbol{\theta}_n$ as $\boldsymbol{\theta}_n = \sum_{n'\neq n}\boldsymbol{\xi}_{n'}$ since $\boldsymbol{\theta}_n^T\mathbf{t}$ replaces $\sum_{n'\neq n}c_{n'}\left(\mathbf{s}, y_{n'}\right)$ and each $c_{n'}\left(\mathbf{s}, y_{n'}\right)$ is approximated as $\boldsymbol{\xi}_{n'}^T\mathbf{t}$. Then $\boldsymbol{\theta}_0$ is updated as $\boldsymbol{\theta}_0 = \sum_n \boldsymbol{\xi}_n$ since the ultimate goal is to approximate $\sum_n c_n\left(\mathbf{s}, y_n\right)$ as $\boldsymbol{\theta}_0^T\mathbf{t}$. In practice, to improve the convergence of IGA, the EACSs are usually updated in a damped way. Given $\boldsymbol{\theta}_0(t)$, $\boldsymbol{\theta}_n(t)$ and $\boldsymbol{\xi}_n(t)$ in the t-th iteration, $\boldsymbol{\theta}_0(t+1)$ and $\boldsymbol{\theta}_n(t+1)$ in the $(t+1)$-th iteration are then updated as

$$
\boldsymbol{\theta}_n\left(t+1\right) = \alpha \sum_{n'=1,n'\neq n}^{2N_r} \boldsymbol{\xi}_{n'}\left(t\right) + \left(1-\alpha\right)\boldsymbol{\theta}_n\left(t\right), n \in \mathcal{Z}_{2N_r}^+,
\tag{35a}
$$

$$
\boldsymbol{\theta}_0\left(t+1\right) = \alpha \sum_{n=1}^{2N_r} \boldsymbol{\xi}_n\left(t\right) + \left(1-\alpha\right)\boldsymbol{\theta}_0\left(t\right),
\tag{35b}
$$

where $0 < \alpha \leq 1$ is the damping. Repeat m-projection and the updating process until convergence. We summarize the IGA for signal detection in Algorithm 1. The computational complexity (the number of real-valued multiplications) of the IGA is $\mathcal{O}\left(16N_rK\left(L+1\right)\right)$ per iteration, where N_r is the number of antennas at the BS, K is the number of users, $L = \sqrt{\tilde{L}}$, and $\tilde{L}$ is the modulation order.

Algorithm 1: IGA for Signal Detection

Input: The a priori probability $f_k\left(s_k\right), k \in \mathcal{Z}_{2K}^+$, the received signal $\mathbf{y}$, the noise power σ_z^2 and the maximal iteration number $t_{\max}$.

1 **Initialization:** set $t = 0$, set damping α, where $0 < \alpha \leq 1$, initialize $\boldsymbol{\theta}_n(0), n \in \mathcal{Z}_{2N_r}$;

2 **repeat**

3   1. Calculate
     $\boldsymbol{\xi}_n(t) = \left[\boldsymbol{\xi}_{n,1}^T(t), \boldsymbol{\xi}_{n,2}^T(t), \ldots, \boldsymbol{\xi}_{n,2K}^T(t)\right]^T, n \in \mathcal{Z}_{2N_r}^+$,
     where $\boldsymbol{\xi}_{n,k}(t) = \left[\xi_{n,k,1}(t), \xi_{n,k,2}(t), \ldots, \xi_{n,k,L-1}(t)\right]^T$,
     $\xi_{n,k,\ell}(t)$ is given by (34), and the intermediate variables are given by (27), (28b) and (32);

4   2. Update $\boldsymbol{\theta}_0(t+1)$ and $\boldsymbol{\theta}_n(t+1), \forall n \in \mathcal{Z}_{2N_r}^+$ as (35);

5   3. $t = t + 1$;

6 **until** Convergence or $t > t_{\max}$;

Output: The approximate probability of the *a posteriori* marginals $p_k\left(s_k|\mathbf{y}\right)$ is given by $p_{0,k}\left(s_k; \boldsymbol{\theta}_{0,k}(t)\right)$ in (11), $\forall k \in \mathcal{Z}_{2K}^+$. Then, the MPM detection is given by (5).

IV. Simulation Results

In this section, we provide simulation results to illustrate the performance of the proposed IGA for signal detection. The uncoded bit error rate (BER) is adopted as the performance metric. We average our results for 1000 realizations of the channel matrix $\mathbf{G}$, which is generated by the widely adopted QuaDRiGa [18]. The simulation scenario is set to "3GPP_38.901_UMa_NLOS", and the main parameters for the simulations are summarized in Table I. The channel matrix

TABLE I
PARAMETER SETTINGS OF THE SIMULATION

Parameter	Value
Number of BS antennas $N_{r,v} \times N_{r,h}$	16×64
UT number K	240
Center frequency f_c	4.8GHz
Modulation Mode	QAM
Modulation Order $\tilde{L}$	4, 16 and 64

is normalized as $\mathbb{E}\left\{\|\mathbf{G}\|_F^2\right\} = N_rK$. The average power of the transmitted symbol of each user is normalized to 1, and the SNR is set as $\mathrm{SNR} = \frac{K}{\sigma_z^2}$. Based on the received signal model (3), we compare the proposed IGA with the LMMSE detector, the expectation propagation (EP) detector proposed in [5] and the approximate message passing (AMP) detector proposed in [19].

The computational complexity of the LMMSE detector is $\mathcal{O}\left(8\left(2N_rK^2 + K^3\right)\right)$ [5]. The computational complexity of the EP detector and AMP are $\mathcal{O}\left(8\left(N_rK^2 + K^3\right)\right)$ and $\mathcal{O}\left(8\left(N_rK\right)\right)$ per iteration, respectively [5], [19]. The complexity of EP detector is the highest among all algorithms. When the number of iterations is low (e.g., tens), the complexity of IGA is lower than that of LMMSE detection. The computational complexity of AMP is the lowest.

We first consider 4-QAM modulation. Fig. 1 shows the BER performance of the IGA compared with LMMSE, EP and AMP. The iteration numbers of IGA, EP and AMP are set as 10, 10 and 30, and 10 and 30, respectively. From Fig. 1, we can find that all the iterative algorithms outperform the
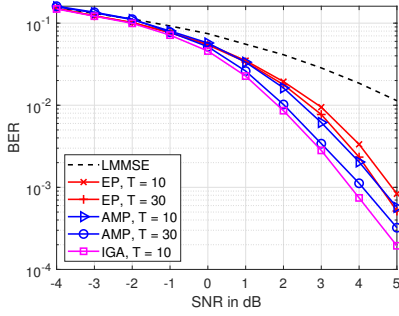
Fig. 1. BER performance of IGA compared with AMP, EP and LMMSE under 4-QAM.
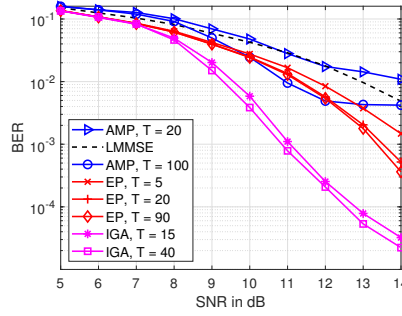
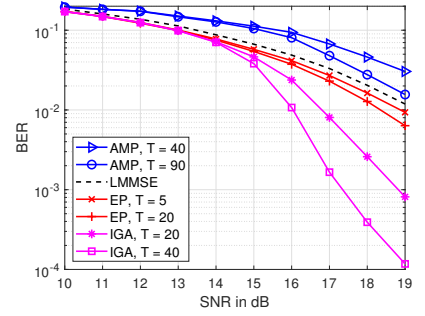Fig. 2. BER performance of IGA compared with AMP, EP and LMMSE under 16-QAM.

Fig. 3. BER performance of IGA compared with AMP, EP and LMMSE under 64-QAM.

LMMSE detector within limited iteration numbers. For BER = 10^{-3}, the SNR gains of the IGA with 10 iterations compared to the AMP with 10 and 30 iterations are around 0.7dB and 0.3dB, respectively. Meanwhile, IGA with 10 iterations can improve the EP performance with 10 and 30 iterations in 1dB and 0.7dB for BER = 10^{-3}, respectively.

Fig. 2 and Fig. 3 show the BER performance for 16-QAM and 64-QAM, respectively. From Fig. 2, we can find that the BER performance of LMMSE outperforms that of the AMP with 20 iterations. Meanwhile, we can find that the gap between IGA and the other algorithms is increasing. For BER =10^{-3}, the SNR gains of the IGA with 15 iterations compared to the EP with 20 and 90 are about 1.2dB and 0.9dB, respectively. The SNR gain for the the IGA with 40 iterations increases by about 0.2dB each over the two gains above. For 64-QAM, from Fig. 3, we can find that the BER performance of the LMMSE detector exceeds that of the AMP after convergence. The gap between IGA and the other algorithms is still increasing. For BER = 10^{-2}, IGA with 20 iterations has improved the EP performance with 5 and 20 iterations in 2.1dB and 1.6dB, respectively. The SNR gain for the the IGA with 40 iterations increases by about 0.7dB each over the two gains above.

V. CONCLUSION

We have proposed an information geometry approach for ultra-MIMO signal detection in this paper. The signal detection is formulated as an MPM detection problem based on the approximation of the *a posteriori* marginals of the transmitted symbols of all users. We convert the calculation of the approximation of the *a posteriori* marginals into an iterative m-projection process. Then, the Lyapunov CLT is applied to have an approximate solution of the m-projection between a probability distribution of the AM and the OBM. Simulation results verify that the IGA can obtain the best BER performance within a limited number of iterations compared with the existing approaches, which demonstrates the superiority of the proposed IGA for ultra-MIMO signal detection.

REFERENCES

[1] C.-X. Wang *et al.*, "On the road to 6G: Visions, requirements, key technologies and testbeds," *IEEE Commun. Surveys Tuts.*, pp. 1–1, 2023.

[2] E. D. Carvalho, A. Ali, A. Amiri, M. Angjelichinoski, and R. W. Heath, "Non-stationarities in extra-large-scale massive MIMO," *IEEE Wireless Commun.*, vol. 27, no. 4, pp. 74–80, Aug. 2020.

[3] M. Cui and L. Dai, "Channel estimation for extremely large-scale MIMO: Far-field or near-field?" *IEEE Trans. Commun.*, vol. 70, no. 4, pp. 2663–2677, April 2022.

[4] S. Yang and L. Hanzo, "Fifty years of MIMO detection: The road to large-scale MIMOs," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 1941–1988, Fourthquarter 2015.

[5] J. Céspedes, P. M. Olmos, M. Sánchez-Fernández, and F. Perez-Cruz, "Expectation propagation detection for high-order high-dimensional MIMO systems," *IEEE Trans. Commun.*, vol. 62, no. 8, pp. 2840–2849, Aug. 2014.

[6] C. R. Rao, "Information and the accuracy attainable in the estimation of statistical parameters," in *Breakthroughs in Statistics*. Springer, 1992, pp. 235–247.

[7] S. Amari and H. Nagaoka, *Methods of Information Geometry*. American Mathematical Soc., 2000, vol. 191.

[8] N. N. Cencov, *Statistical Decision Rules and Optimal Inference*. American Mathematical Soc., 2000, no. 53.

[9] S. Amari, *Information Geometry and Its Applications*. Tokyo, Japan: Springer, 2016.

[10] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann, 1988.

[11] A. L. Yuille and A. Rangarajan, "The concave-convex procedure," *Neural Computation*, vol. 15, no. 4, pp. 915–936, Apr. 2003.

[12] S. Ikeda, T. Tanaka, and S. Amari, "Information geometry of turbo and low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 1097–1114, June 2004.

[13] J. Y. Yang, A.-A. Lu, Y. Chen, X. Q. Gao, X.-G. Xia, and D. T. M. Slock, "Channel estimation for massive MIMO: An information geometry approach," *IEEE Trans. Signal Process.*, vol. 70, pp. 4820–4834, Oct. 2022.

[14] J. Y. Yang, Y. Chen, A.-A. Lu, W. Zhong, X. Q. Gao, X. You, X.-G. Xia, and D. T. M. Slock, "A simplified information geometry approach for massive MIMO-OFDM channel estimation," *submitted to IEEE Trans. Signal Process.*, 2023.

[15] J. Y. Yang, Y. Chen, X. Q. Gao, D. T. M. Slock, and X.-G. Xia, "Signal detection for ultra-massive MIMO: An information geometry approach," *to be submitted*.

[16] H. Pishro-Nik, *Introduction to Probability, Statistics, and Random Processes*. Kappa Research: Galway, Ireland, 2014.

[17] P. Billingsley, *Probability and Measure*. John Wiley & Sons, New York, 2008.

[18] S. Jaeckel, L. Raschkowski, K. Börner, and L. Thiele, "Quadriga: A 3-d multi-cell channel model with time evolution for enabling virtual field trials," *IEEE Trans. Antennas Propag.*, vol. 62, no. 6, pp. 3242–3256, 2014.

[19] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing: I. motivation and construction," in *2010 IEEE Information Theory Workshop on Information Theory (ITW 2010, Cairo)*, Jan 2010, pp. 1–5.