

# Signal Detection for Ultra-Massive MIMO: An Information Geometry Approach

Jiyuan Yang, *Student Member, IEEE*, Yan Chen, Xiqi Gao, *Fellow, IEEE*, Dirk Slock, *Fellow, IEEE*, and Xiang-Gen Xia, *Fellow, IEEE*

**Abstract**—In this paper, we propose an information geometry approach (IGA) for signal detection (SD) 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 between submanifolds with different constraints. 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-SD emerges as 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

As one of the critical technologies for 5G, massive multiple-input multiple-output (MIMO) can provide significant gains in both spectral efficiency and energy efficiency for communication systems [1], [2]. In future 6G communications, an ultra-massive MIMO system will employ an ultra-large array with hundreds or thousands of antennas, serving tens or even hundreds of users simultaneously, which is able to achieve higher spectral efficiency and energy efficiency, and wider and more flexible network coverage than ever [3]–[6]. For the realization of the substantial benefits of ultra-massive MIMO, signal detection is of great importance. Based on a received signal, the task of the detector is to determine the transmitted symbol. The optimal detector based on the

maximum *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). Consequently, the computational complexity of the MAP or ML detector rapidly becomes unaffordable as the number of decision symbols increases. 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 estimates of the transmitted symbols of the LMMSE detector are biased [7], and the performance of the LMMSE detector degrades severely in massive MIMO systems with high-order constellations [8].

In the past few decades, many works have been devoted to the massive MIMO signal detection [8]–[13], of which Bayesian inference approaches, e.g., belief propagation (BP), expectation propagation (EP), etc., are of significant interest due to the relatively low computational complexity and higher performance than linear detection. These methods aim to calculate an approximation of the *a posteriori* probability distribution (or its marginals) of the transmitted symbols. A hard-decision based on the *a posteriori* mean or a soft-decision based on the *a posteriori* marginals is then performed. In [8], EP is first introduced into the massive MIMO signal detection with high-order modulation. [11] proposes a beam domain detector based on the layered BP for massive MIMO systems. A variant EP detector is proposed in [12] based on decentralized processing. [13] proposes a MIMO detector for high-order QAM modulation based on the Gaussian tree approximation.

Information geometry, which is introduced by Rao [14], and then formally developed by Amari [15] and Cencov [16], 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. [17], [18]. Amari et al. also show the intrinsic geometric insight of some classical Bayesian inference methods, e.g., the belief propagation (BP) [19]. Meanwhile, some optimization methods, such as the concave-convex procedure (CCCP) [20], are also applied to calculate the marginals of the *a posteriori* distribution. On Bayesian inference in communications, [21] 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

This work was supported by the National Key R&D Program of China under Grant 2018YFB1801103, the Jiangsu Province Basic Research Project under Grant BK20192002, the Fundamental Research Funds for the Central Universities under Grant 2242022k60007, the Key R&D Plan of Jiangsu Province under Grant BE2022067, and the Huawei Cooperation Project. This work was presented at the 15th International Conference on Wireless Communications and Signal Processing (WCSP 2023), Hangzhou, China, November, 2023. (*Corresponding author: Xiqi Gao.*)

Jiyuan Yang, Yan Chen and Xiqi Gao are with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China and are also with Purple Mountain Laboratories, Nanjing 211111, China (e-mail: {jyyang, 213160372, xqgao}@seu.edu.cn).

Xiang-Gen Xia is with the Department of Electrical and Computer Engineering, University of Delaware, Newark, DE 19716 USA (e-mail: xxia@ee.udel.edu).

Dirk Slock is with the Department of Communication Systems, EURECOM, 06410 Biot, France (e-mail: Dirk.Slock@eurecom.fr).

processing and an information geometry approach is proposed for massive MIMO channel estimation in [22], [23].

In addition to the unique insight that the geometric perspective offers, information geometry also provides us with a unified framework where various sets of probability distributions are considered to be endowed with the structure of differential geometry. Hence, we are able to construct a Fisher information matrix (FIM) based distance between two parametrized distributions. Amari [15] also shows that this distance is invariant to non-singular transformation of the parameters. As a result, information geometry is closely related to estimation theory. Due to these advantages, information geometry has recently been applied to many other problems such as the complex network construction [24], the target detection [25], and the clustering [26].

In this paper, we propose an information geometry approach for signal detection (IGA-SD) for ultra-massive MIMO systems. We formulate the signal detection as obtaining the marginals of the *a posteriori* probability distribution of the transmitted symbols. Then, a component-wise decision can be performed based on the *a posteriori* marginals. With the information geometry theory, we calculate the approximations of the *a posteriori* marginals. More precisely, by treating the sets of the probability distributions of discrete random vectors with different constraints as several different (sub)manifolds, the calculation of the marginals is converted into an iterative *m*-projection process. 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. Preliminaries of information geometry is introduced in Section III. The Information geometry approach for ultra-massive MIMO signal detection is proposed in Section IV. Simulation results are provided in Section V. The conclusion is drawn in Section VI.

Notations: The following notations are adopted in this paper. Upper (lower) case boldface letters denote matrices (column vectors).  $\mathcal{R}(\cdot)$  and  $\mathcal{I}(\cdot)$  denote the real and imaginary parts of a complex number (matrix), respectively. The superscripts  $(\cdot)^*$ ,  $(\cdot)^T$  and  $(\cdot)^H$  denote the conjugate, transpose and conjugate-transpose operator, respectively.  $\text{Diag}\{\mathbf{x}\}$  denotes the diagonal matrix with  $\mathbf{x}$  along its main diagonal and  $\text{diag}\{\mathbf{X}\}$  denotes a vector consisting of the diagonal elements of  $\mathbf{X}$ .  $\text{Bdiag}\{\mathbf{X}_1, \mathbf{X}_2, \dots\}$  denotes a block diagonal matrix with matrices  $\mathbf{X}_i$  located along the main diagonal. We use  $a_{i,j}$  to denote the  $(i, j)$ -th element of the matrix  $\mathbf{A}$ , where the element indices start with 1.  $\odot$  and  $\otimes$  denote the Hadamard product and Kronecker product, respectively. Define  $\mathcal{Z}_N \triangleq \{0, 1, \dots, N\}$  and  $\mathcal{Z}_N^+ \triangleq \{1, 2, \dots, N\}$ .  $\setminus$  denotes the set subtraction operation. To avoid confusion,  $p(\cdot)$  and  $f(\cdot)$  denote the probability distribution of discrete random variables and the probability density function (PDF) of continuous random variables, respectively.  $f_{\text{CG}}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$  denotes the PDF

of a complex Gaussian distribution  $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  for vector  $\mathbf{x}$  of complex random variables.  $f_{\text{G}}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$  denotes the PDF of a real Gaussian distribution  $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  for vector  $\mathbf{x}$  of complex random variables.

## II. SYSTEM MODEL AND PROBLEM STATEMENT

### A. System Configuration

Consider an uplink ultra-massive MIMO system where one base station (BS) equipped with an ultra-massive antenna array serves  $K$  single-antenna users, and the BS has  $N_r$  antennas. Denote the transmitted symbol vector of all users as  $\tilde{\mathbf{s}} \triangleq [\tilde{s}_1, \tilde{s}_2, \dots, \tilde{s}_K]^T \in \tilde{\mathbb{S}}^K$  where  $\tilde{s}_k \in \tilde{\mathbb{S}}, k \in \mathcal{Z}_K^+$ , is the transmitted symbol of the user  $k$ .  $\tilde{\mathbb{S}}$  is the signal constellation and let us assume  $\tilde{\mathbb{S}} = \{\tilde{s}^{(0)}, \tilde{s}^{(1)}, \dots, \tilde{s}^{(\tilde{L}-1)}\}$ ,

where  $\{\tilde{s}^{(\ell)}\}_{\ell=0}^{\tilde{L}-1}$  are the constellation points, and  $\tilde{L}$  is the modulation order (or constellation size). In this paper, we focus on the uncoded systems and the symmetric  $\tilde{L}$ -QAM modulation. We assume that each user chooses symbols from  $\tilde{\mathbb{S}}$  uniformly at random, and all users use the same alphabet, although the proposed IGA-SD can be readily extended to arbitrary modulations with different distributions as long as the transmitted symbols of the users are statistically independent and the real and imaginary parts of each user's symbol are statistically independent as well. We also assume that the average power of  $\tilde{s}_k$  is normalized to unit, i.e.,  $\mathbb{E}\{|\tilde{s}_k|^2\} = 1, k \in \mathcal{Z}_K^+$ . 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

$$\tilde{\mathbf{y}} = \tilde{\mathbf{G}}\tilde{\mathbf{s}} + \tilde{\mathbf{z}}, \quad (1)$$

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,  $\tilde{\mathbf{z}} \sim \mathcal{CN}(\mathbf{0}, \tilde{\sigma}_z^2 \mathbf{I})$  and  $\tilde{\sigma}_z^2$  is the noise variance. In this work, we assume that the BS has perfect channel state information. As a note, the reason why, in the above notations, the tildes are added on the tops of the math symbols is that we will later formulate and analyze their real counterparts without the tildes for notational simplicity.

### B. Problem Statement

Assuming that the transmitted symbol vector  $\tilde{\mathbf{s}}$  and the noise vector  $\tilde{\mathbf{z}}$  are independent with each other and the symbols transmitted by different users are independent with each other as well. Then, with the received signal model (1), the *a posteriori* probability distribution of the transmitted symbol vector  $\tilde{\mathbf{s}}$  can be expressed as

$$\begin{aligned} p(\tilde{\mathbf{s}}|\tilde{\mathbf{y}}) &\propto p_{\text{pr}}^c(\tilde{\mathbf{s}}) f(\tilde{\mathbf{y}}|\tilde{\mathbf{s}}) \\ &= \prod_{k=1}^K p_{\text{pr},k}^c(\tilde{s}_k) f_{\text{CG}}(\tilde{\mathbf{y}}; \tilde{\mathbf{G}}\tilde{\mathbf{s}}, \sigma_z^2 \mathbf{I}), \end{aligned} \quad (2)$$

where  $p_{\text{pr}}^c(\tilde{\mathbf{s}})$  is the a priori probability distribution of the complex transmitted symbol vector  $\tilde{\mathbf{s}}$ ,  $f(\tilde{\mathbf{y}}|\tilde{\mathbf{s}})$  is the PDF of the received signal  $\tilde{\mathbf{y}}$  given  $\tilde{\mathbf{s}}$ ,  $p_{\text{pr},k}^c(\tilde{s}_k)$  is the a priori probability of the complex symbol transmitted by user  $k$ ,  $p_{\text{pr},k}^c(\tilde{s}_k)|_{\tilde{s}_k=\tilde{s}^{(\ell)}} = 1/\tilde{L}, k \in \mathcal{Z}_K^+, \ell \in \mathcal{Z}_{\tilde{L}-1}$ . Given the a

*a posteriori* probability distribution  $p(\tilde{\mathbf{s}}|\tilde{\mathbf{y}})$ , the MAP detector (or the ML detector in this case) is given by

$$\tilde{\mathbf{s}}_{\text{MAP}} = \arg \max_{\tilde{\mathbf{s}} \in \tilde{\mathcal{S}}^K} p(\tilde{\mathbf{s}}|\tilde{\mathbf{y}}), \quad (3)$$

which minimizes the error probability that  $\tilde{\mathbf{s}}_{\text{MAP}}$  does not coincide with the true one. The calculation of the MAP detector is unaffordable for practical ultra-massive MIMO systems since the number of candidates of  $\tilde{\mathbf{s}}$  increases exponentially w.r.t.  $K$  and (3) is NP-hard.

Before proceeding, we reformulate the complex-valued received signal model (1) into a real-valued one, which is necessary for developing IGA-SD in this paper. Define real vectors

$$\mathbf{y} \triangleq \begin{bmatrix} \mathcal{R}\{\tilde{\mathbf{y}}\} \\ \mathcal{I}\{\tilde{\mathbf{y}}\} \end{bmatrix}, \mathbf{z} \triangleq \begin{bmatrix} \mathcal{R}\{\tilde{\mathbf{z}}\} \\ \mathcal{I}\{\tilde{\mathbf{z}}\} \end{bmatrix} \in \mathbb{R}^{2N_r} \quad (4a)$$

$$\mathbf{s} \triangleq \begin{bmatrix} \mathcal{R}\{\tilde{\mathbf{s}}\} \\ \mathcal{I}\{\tilde{\mathbf{s}}\} \end{bmatrix} \in \mathbb{R}^{2K}, \quad (4b)$$

and a real matrix

$$\mathbf{G} \triangleq \begin{bmatrix} \mathcal{R}\{\tilde{\mathbf{G}}\}, & -\mathcal{I}\{\tilde{\mathbf{G}}\} \\ \mathcal{I}\{\tilde{\mathbf{G}}\}, & \mathcal{R}\{\tilde{\mathbf{G}}\} \end{bmatrix} \in \mathbb{R}^{2N_r \times 2K}. \quad (5)$$

Then, we can obtain the real-valued received signal model as

$$\mathbf{y} = \mathbf{G}\mathbf{s} + \mathbf{z}, \quad (6)$$

where  $\mathbf{s} = [s_1, s_2, \dots, s_{2K}]^T \in \mathbb{S}^{2K}$ ,  $s_k \in \mathbb{S}$ ,  $k \in \mathcal{Z}_{2K}^+$ ,  $\mathbb{S} = \{s^{(0)}, s^{(1)}, \dots, s^{(L-1)}\}$  is the alphabet for the real and imaginary components of a symmetric  $\tilde{L}$ -QAM modulation where the index starts from 0 is for the representation convenience of the likelihood ratio detection later,  $L = \sqrt{\tilde{L}}$ ,  $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \sigma_z^2 \mathbf{I})$  is the noise vector, and  $\sigma_z^2 = \tilde{\sigma}_z^2/2$ . Given the received signal model (6), the *a posteriori* distribution of  $\mathbf{s}$  can be expressed as

$$\begin{aligned} p(\mathbf{s}|\mathbf{y}) &\propto \prod_{k=1}^{2K} p_{\text{pr},k}(s_k) \prod_{n=1}^{2N_r} f(y_n|\mathbf{s}) \\ &\propto \prod_{k=1}^{2K} p_{\text{pr},k}(s_k) \prod_{n=1}^{2N_r} \exp \left\{ -\frac{(y_n - \mathbf{e}_n^T \mathbf{G}\mathbf{s})^2}{2\sigma_z^2} \right\}, \end{aligned} \quad (7)$$

where  $p_{\text{pr},k}(s_k)|_{s_k=s^{(\ell)}} = \frac{1}{L}$ ,  $k \in \mathcal{Z}_{2K}^+$ ,  $\ell \in \mathcal{Z}_{L-1}$ , is the a priori probability of  $s_k$ ,  $y_n$  is the  $n$ -th element of  $\mathbf{y}$ ,  $f(y_n|\mathbf{s})$  is the 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.,  $p_k(s_k|\mathbf{y})$ ,  $k \in \mathcal{Z}_{2K}^+$ , of the *a posteriori* distribution  $p(\mathbf{s}|\mathbf{y})$ , which can be used for the maximization of the *a posteriori* marginals (MPM) detector, i.e., for  $k \in \mathcal{Z}_{2K}^+$ ,

$$s_{k,\text{MPM}} = \arg \max_{s_k \in \mathbb{S}} p_k(s_k|\mathbf{y}). \quad (8)$$

### III. PRELIMINARIES OF INFORMATION GEOMETRY

In this section, we briefly introduce the information geometry approach (IGA), where more details can be found in [17], [18], [21], [22]. We begin with the exponential family. Consider a discrete random vector  $\mathbf{x} \in \mathbb{X}$  with finite dimension, where each scalar random variable in  $\mathbf{x}$  takes finite values and  $\mathbb{X}$  is a finite set. The probability distribution of  $\mathbf{x}$  is said to belong to the exponential family if it can be expressed as

$$p(\mathbf{x}; \boldsymbol{\theta}) = \exp \{ \boldsymbol{\theta}^T \mathbf{t} - \psi(\boldsymbol{\theta}) \}, \quad (9)$$

where  $\mathbf{t}$  is a sufficient statistic of random vector  $\mathbf{x}$ ,  $\boldsymbol{\theta}$  is the natural parameter (NP) of  $p(\mathbf{x}; \boldsymbol{\theta})$ , and  $\psi(\boldsymbol{\theta})$  is the free energy, which makes  $p(\mathbf{x}; \boldsymbol{\theta})$  a probability distribution, i.e.,  $\sum_{\mathbf{x}} p(\mathbf{x}; \boldsymbol{\theta}) = 1$ . We then introduce the  $e$ -flat manifold that is needed for  $m$ -projections later [17], [18], [21]. Consider a manifold  $\mathcal{U}$ , which is defined as a set of probability distributions of  $\mathbf{x}$ , e.g.,  $\mathcal{U} = \{p(\mathbf{x})\}$ , where each element in  $\mathcal{U}$ , i.e.,  $p(\mathbf{x})$ , is a particular probability distribution of  $\mathbf{x}$ .  $\mathcal{U}$  is said to be  $e$ -flat if for all  $0 \leq d \leq 1$ ,  $p_i(\mathbf{x}), p_j(\mathbf{x}) \in \mathcal{U}$ , the following  $q(\mathbf{x}; d)$  belongs to  $\mathcal{U}$ ,

$$q(\mathbf{x}; d) = (1-d) \ln p_i(\mathbf{x}) + d \ln p_j(\mathbf{x}) + c_{\text{na}}(d), \quad (10)$$

where  $c_{\text{na}}(d)$  is a normalization constant makes  $q(\mathbf{x}; d)$  a probability distribution. From the definition, any exponential family is  $e$ -flat. Suppose  $\mathcal{V} = \{q(\mathbf{x})\} \subseteq \mathcal{U}$  is an  $e$ -flat submanifold. Given  $p(\mathbf{x}) \in \mathcal{U}$ , the point (probability distribution) in  $\mathcal{V}$  that minimizes the Kullback-Leibler (K-L) divergence from  $p(\mathbf{x})$  to  $\mathcal{V}$ , i.e.,

$$q^*(\mathbf{x}) = \arg \min_{q(\mathbf{x}) \in \mathcal{V}} D_{\text{KL}} \{p(\mathbf{x}) : q(\mathbf{x})\}, \quad (11)$$

is called the  $m$ -projection of  $p(\mathbf{x})$  onto  $\mathcal{V}$ , where the K-L divergence is defined by

$$D_{\text{KL}} \{p(\mathbf{x}) : q(\mathbf{x})\} = \sum_{\mathbf{x} \in \mathbb{X}} p(\mathbf{x}) \ln \left( \frac{p(\mathbf{x})}{q(\mathbf{x})} \right). \quad (12)$$

We now give the preliminaries of IGA in Bayesian inference. Let  $\mathbf{x}_h \in \mathbb{R}^{N_h}$  and  $\mathbf{y}_o \in \mathbb{R}^{N_o}$  be hidden and observed random vectors, respectively. Denote the *a posteriori* distribution as  $p(\mathbf{x}_h|\mathbf{y}_o)$ . Our goal is to calculate the approximations of the *a posteriori* marginals, i.e.,  $p(x_{h,i}|\mathbf{y}_o)$ , where  $x_{h,i}$  is the  $i$ -th component of  $\mathbf{x}_h$  and  $i \in \mathcal{Z}_{N_h}^+$ . In this paper, we focus on the following case: all the components of  $\mathbf{x}_h$  are independent and all the components of  $\mathbf{y}_o$  given  $\mathbf{x}_h$  are independent as well. The *a posteriori* distribution can be then expressed as

$$p(\mathbf{x}_h|\mathbf{y}_o) \propto p(\mathbf{x}_h) p(\mathbf{y}_o|\mathbf{x}_h) = \prod_{i=1}^{N_h} p_i(x_{h,i}) \prod_{n=1}^{N_o} p_n(y_{o,n}|\mathbf{x}_h), \quad (13)$$

where  $p_i(x_{h,i})$  and  $p_n(y_{o,n}|\mathbf{x}_h)$  are the marginals of  $p(\mathbf{x}_h)$  and  $p(\mathbf{y}_o|\mathbf{x}_h)$ , respectively, and  $y_{o,n}$  is the  $n$ -th component of  $\mathbf{y}_o$ . Suppose that the a priori marginals  $\{p_i(x_{h,i})\}_{i=1}^{N_h}$  belong to an exponential family, and each of them can be expressed as

$$p_i(x_{h,i}) = p_i(x_{h,i}; \mathbf{d}_{h,i}) = \exp \{ \mathbf{d}_{h,i}^T \mathbf{t}_{h,i} - \psi(\mathbf{d}_{h,i}) \}, \quad (14)$$

where  $\mathbf{d}_{h,i} \in \mathbb{R}^{N_i}$  is the NP of  $p_i(x_{h,i}; \mathbf{d}_{h,i})$ ,  $\mathbf{t}_{h,i} \in \mathbb{R}^{N_i}$  is a sufficient statistic of the single random variable  $x_{h,i}$ , e.g.,  $x_{h,i}$  and  $x_{h,i}^2$ , and  $\psi(\mathbf{d}_{h,i})$  is the free energy. As we shall see later in this section, the probability distributions of discrete random vectors belong to the exponential family. Meanwhile, suppose that the marginals of the conditional probability distribution can be expressed as

$$p_n(y_{o,n}|\mathbf{x}_h) = \exp\{c_n(\mathbf{x}_h, y_{o,n}) - \psi_n\}, n \in \mathcal{Z}_{N_o}^+, \quad (15)$$

where  $c_n(\mathbf{x}_h, y_{o,n})$  is a polynomial of  $\mathbf{x}_h$  which is parameterized by the variables including  $y_{o,n}$ , and  $\psi_n$  is the normalization factor.  $c_n(\mathbf{x}_h, y_{o,n})$  above often contains the interactions between the random variables of  $\mathbf{x}_h$ , e.g., the cross-terms  $x_{h,i}x_{h,j}$ ,  $i \neq j$ . A more detailed  $c_n(\mathbf{x}_h, y_{o,n})$  will occur in the next section. In this case, the *a posteriori* probability distribution can be expressed as

$$p(\mathbf{x}_h|\mathbf{y}_o) = \exp\left\{\mathbf{d}_h^T \mathbf{t}_h + \sum_{n=1}^{N_o} c_n(\mathbf{x}_h, y_{o,n}) - \psi_q\right\}, \quad (16)$$

where  $\mathbf{d}_h = [\mathbf{d}_{h,1}^T, \mathbf{d}_{h,2}^T, \dots, \mathbf{d}_{h,N_h}^T]^T \in \mathbb{R}^{N_a}$ ,  $\mathbf{t}_h = [\mathbf{t}_{h,1}^T, \mathbf{t}_{h,2}^T, \dots, \mathbf{t}_{h,N_h}^T]^T \in \mathbb{R}^{N_a}$ ,  $N_a = \sum_{i=1}^{N_h} N_i$ , and  $\psi_q$  is the normalization factor. In (16),  $\mathbf{t}_h$  only contains the separated random variables (i.e., no cross-terms of them), and all the interactions (cross-terms) between the random variables are included in  $c_n(\mathbf{x}_h, y_{o,n})$ ,  $n \in \mathcal{Z}_{N_o}^+$ . IGA aims to approximate  $\sum_{n=1}^{N_o} c_n(\mathbf{x}_h, y_{o,n})$  as  $\boldsymbol{\theta}_0^T \mathbf{t}_h$ , where  $\boldsymbol{\theta}_0 \in \mathbb{R}^{N_a}$ , i.e., IGA aims to approximate the summation of all the cross-terms into a summation of non-cross-terms of the random variables, when  $N_a$  is large. In this case, we have

$$p(\mathbf{x}_h|\mathbf{y}_o) \approx p_0(\mathbf{x}_h; \boldsymbol{\theta}_0) = \exp\left\{(\mathbf{d}_h + \boldsymbol{\theta}_0)^T \mathbf{t}_h - \psi_0(\boldsymbol{\theta}_0)\right\}, \quad (17)$$

where  $\psi_0(\boldsymbol{\theta}_0)$  is the normalization factor. The marginals of  $p_0(\mathbf{x}_h; \boldsymbol{\theta}_0)$ , i.e.,  $p_0(x_{h,i}; \boldsymbol{\theta}_0)$ ,  $i \in \mathcal{Z}_{N_h}^+$ , can be calculated easily since  $p_0(\mathbf{x}_h; \boldsymbol{\theta}_0)$  contains no interactions between the random variables  $\{x_{h,i}\}_{i=1}^{N_h}$ . To obtain  $\boldsymbol{\theta}_0$ , we construct two types of manifolds and compute the approximation for each  $c_n(\mathbf{x}_h, y_{o,n})$  in an iterative manner, which is denoted as  $\boldsymbol{\xi}_n^T \mathbf{t}_h$ . At last,  $\boldsymbol{\theta}_0$  is calculated as  $\boldsymbol{\theta}_0 = \sum_{n=1}^{N_o} \boldsymbol{\xi}_n$ . The two types of manifolds are the objective manifold (OBM) and the auxiliary manifold (AM). The OBM  $\mathcal{M}_0$  is defined as the set of probability distributions of random vector  $\mathbf{x}_h$ , of which all the components are independent with each other, i.e.,

$$\mathcal{M}_0 = \{p_0(\mathbf{x}_h; \boldsymbol{\theta}_0) | \boldsymbol{\theta}_0 \in \mathbb{R}^{N_a}\}, \quad (18a)$$

$$\begin{aligned} p_0(\mathbf{x}_h; \boldsymbol{\theta}_0) &= \prod_{i=1}^{N_h} p_{0,i}(x_{h,i}; \boldsymbol{\theta}_{0,i}) \\ &= \exp\left\{(\mathbf{d}_h + \boldsymbol{\theta}_0)^T \mathbf{t}_h - \psi_0(\boldsymbol{\theta}_0)\right\}, \end{aligned} \quad (18b)$$

$$p_{0,i}(x_{h,i}; \boldsymbol{\theta}_{0,i}) = \exp\left\{(\mathbf{d}_{h,i} + \boldsymbol{\theta}_{0,i})^T \mathbf{t}_{h,i} - \psi_0(\boldsymbol{\theta}_{0,i})\right\}, \quad (18c)$$

where  $\boldsymbol{\theta}_0 = [\boldsymbol{\theta}_{0,1}^T, \boldsymbol{\theta}_{0,2}^T, \dots, \boldsymbol{\theta}_{0,N_h}^T]^T \in \mathbb{R}^{N_a}$ ,  $\boldsymbol{\theta}_{0,i} \in \mathbb{R}^{N_i}$ ,  $p_{0,i}(x_{h,i}; \boldsymbol{\theta}_{0,i})$  is the marginal distribution of  $p_0(\mathbf{x}_h; \boldsymbol{\theta}_0)$ ,  $\psi_0(\boldsymbol{\theta}_0) = \sum_{i=1}^{N_h} \psi_0(\boldsymbol{\theta}_{0,i})$  is the free energy (normalization factor) of  $p_0(\mathbf{x}_h; \boldsymbol{\theta}_0)$ , and  $\psi_0(\boldsymbol{\theta}_{0,i})$  is the free energy

of  $p_{0,i}(x_{h,i}; \boldsymbol{\theta}_{0,i})$ .  $\boldsymbol{\theta}_0$  above is referred as to the  $e$ -affine coordinate system or the natural parameter of  $p_0(\mathbf{x}_h; \boldsymbol{\theta}_0)$ . And  $\boldsymbol{\theta}_{0,i}$  is referred as to the  $e$ -affine coordinate system or the natural parameter of  $p_{0,i}(x_{h,i}; \boldsymbol{\theta}_{0,i})$ . To avoid confusion with the natural parameter of the exponential family, we refer to  $\boldsymbol{\theta}_0$  as the  $e$ -affine coordinate system (abbreviated as EACS) of  $p_0(\mathbf{x}_h; \boldsymbol{\theta}_0)$  in this paper (similar with  $\boldsymbol{\theta}_{0,i}$  and  $p_{0,i}(x_{h,i}; \boldsymbol{\theta}_{0,i})$ ). Then,  $N_o$  AMs are defined, where the  $n$ -th of them is expressed as

$$\mathcal{M}_n = \{p_n(\mathbf{x}_h; \boldsymbol{\theta}_n) | \boldsymbol{\theta}_n \in \mathbb{R}^{N_a}\}, \quad (19a)$$

$$p_n(\mathbf{x}_h; \boldsymbol{\theta}_n) = \exp\left\{(\mathbf{d}_h + \boldsymbol{\theta}_n)^T \mathbf{t}_h + c_n(\mathbf{x}_h, y_{o,n}) - \psi_n(\boldsymbol{\theta}_n)\right\}, \quad (19b)$$

where  $\boldsymbol{\theta}_n$  is referred as to the EACS of  $p_n(\mathbf{x}_h; \boldsymbol{\theta}_n)$  and  $\psi_n(\boldsymbol{\theta}_n)$  is the free energy. It can be readily checked that the OBM and the AMs are all  $e$ -flat. Only one interaction term  $c_n(\mathbf{x}_h, y_{o,n})$  is remained in  $p_n(\mathbf{x}_h; \boldsymbol{\theta}_n)$ , and all the others, i.e.,  $\sum_{n' \neq n} c_{n'}(\mathbf{x}_h, y_{o,n'})$  are replaced as  $\boldsymbol{\theta}_n^T \mathbf{t}_h$ . Assume that the EACS  $\boldsymbol{\theta}_n$  of  $p_n(\mathbf{x}_h; \boldsymbol{\theta}_n)$ ,  $n \in \mathcal{Z}_{N_o}^+$ , is given, we calculate the approximation of  $c_n(\mathbf{x}_h, y_{o,n})$  from the  $m$ -projection of  $p_n(\mathbf{x}_h; \boldsymbol{\theta}_n)$  onto the OBM  $\mathcal{M}_0$ . Denote the  $m$ -projection of  $p_n(\mathbf{x}_h; \boldsymbol{\theta}_n)$  onto  $\mathcal{M}_0$  as  $p_0(\mathbf{x}_h; \boldsymbol{\theta}_{0n})$ , where  $\boldsymbol{\theta}_{0n} \in \mathbb{R}^{N_a}$ , and

$$\boldsymbol{\theta}_{0n} = \arg \min_{\boldsymbol{\theta}_0 \in \mathbb{R}^{N_a}} D_{\text{KL}}\{p_n(\mathbf{x}_h; \boldsymbol{\theta}_n) : p_0(\mathbf{x}_h; \boldsymbol{\theta}_0)\}. \quad (20)$$

We shall see a more specific example about the calculation of the  $m$ -projection in the next section. After  $\boldsymbol{\theta}_{0n}$  is obtained, we express the  $m$ -projection  $p_0(\mathbf{x}_h; \boldsymbol{\theta}_{0n})$  as

$$\begin{aligned} p_0(\mathbf{x}_h; \boldsymbol{\theta}_{0n}) &= \exp\left\{(\mathbf{d}_h + \boldsymbol{\theta}_{0n})^T \mathbf{t}_h - \psi_0(\boldsymbol{\theta}_{0n})\right\} \\ &= \exp\left\{(\mathbf{d}_h + \boldsymbol{\theta}_n + \boldsymbol{\xi}_n)^T \mathbf{t}_h - \psi_0(\boldsymbol{\theta}_{0n})\right\}, \end{aligned} \quad (21)$$

where the EACS  $\boldsymbol{\theta}_{0n}$  of  $p_0(\mathbf{x}_h; \boldsymbol{\theta}_{0n})$  is regarded as the sum of the EACS  $\boldsymbol{\theta}_n$  of  $p_n(\mathbf{x}_h; \boldsymbol{\theta}_n)$  and an extra item  $\boldsymbol{\xi}_n$ . If we compare the last equation of (21) and  $p_n(\mathbf{x}_h; \boldsymbol{\theta}_n)$  in (19b), it can be found that in the  $m$ -projection  $p_0(\mathbf{x}_h; \boldsymbol{\theta}_{0n})$ , the interaction item  $c_n(\mathbf{x}_h, y_{o,n})$  is replaced by  $\boldsymbol{\xi}_n^T \mathbf{t}_h$ . Hence,  $\boldsymbol{\xi}_n^T \mathbf{t}_h$  is regarded as the approximation of  $c_n(\mathbf{x}_h, y_{o,n})$ , and we calculate the approximation item  $\boldsymbol{\xi}_n$  as

$$\boldsymbol{\xi}_n = \boldsymbol{\theta}_{0n} - \boldsymbol{\theta}_n, n \in \mathcal{Z}_{N_o}^+. \quad (22)$$

Then,  $p_0(\mathbf{x}_h; \boldsymbol{\theta}_0)$  with  $\boldsymbol{\theta}_0 = \sum_{n=1}^{N_o} \boldsymbol{\xi}_n$  is considered as the approximation of the *a posteriori* distribution  $p(\mathbf{x}_h|\mathbf{y}_o)$ . Meanwhile, note that the whole process is proceeded in an iterative manner since the EACSs  $\{\boldsymbol{\theta}_n\}_{n=1}^{N_o}$  are not known at first. To be specific, we first initialize the EACSs as  $\{\boldsymbol{\theta}_n(0)\}_{n=0}^{N_o}$ . Given the EACS  $\boldsymbol{\theta}_0(t)$  of  $p_0(\mathbf{x}_h; \boldsymbol{\theta}_0(t))$  and the EACS  $\boldsymbol{\theta}_n(t)$  of  $p_n(\mathbf{x}_h; \boldsymbol{\theta}_n(t))$ ,  $n \in \mathcal{Z}_{N_o}^+$ , at the  $t$ -th time, we calculate  $\boldsymbol{\theta}_{0n}(t)$  and  $\boldsymbol{\xi}_n(t)$ ,  $n \in \mathcal{Z}_{N_o}^+$ , as (20) and (22), respectively. We then update the EACS of  $p_n(\mathbf{x}_h; \boldsymbol{\theta}_n(t))$ ,  $n \in \mathcal{Z}_{N_o}^+$ , as

$$\boldsymbol{\theta}_n(t+1) = \sum_{n'=1, n' \neq n}^{N_o} \boldsymbol{\xi}_{n'}(t), \quad (23)$$

since  $\boldsymbol{\theta}_n^T(t+1) \mathbf{t}_h$  replaces  $\sum_{n' \neq n} c_{n'}(\mathbf{x}_h, y_{o,n})$  in  $p_n(\mathbf{x}_h; \boldsymbol{\theta}_n(t+1))$  and each interaction term  $c_n(\mathbf{x}_h, y_{o,n})$  is approximated as  $\boldsymbol{\xi}_n^T(t) \mathbf{t}_h$  at the  $t$ -th time. The EACS of

$p_0(\mathbf{x}_n; \boldsymbol{\theta}_0(t))$  is updated as  $\boldsymbol{\theta}_0(t+1) = \sum_{n=1}^{N_o} \boldsymbol{\xi}_n(t)$  as mentioned above. Then, repeat the  $m$ -projection, calculate the approximation terms  $\{\boldsymbol{\xi}_n\}_{n=1}^{N_o}$  and the updates until convergence. We now discuss about the damped updating. In practice, to improve the convergence of the IGA, the EACSSs  $\{\boldsymbol{\theta}_n\}_{n=0}^{N_o}$  are usually updated in a damped way, i.e.,

$$\boldsymbol{\theta}_n(t+1) = \alpha \sum_{n'=1, n' \neq n}^{N_o} \boldsymbol{\xi}_{n'}(t) + (1-\alpha)\boldsymbol{\theta}_n(t), n \in \mathcal{Z}_{N_o}^+, \quad (24a)$$

$$\boldsymbol{\theta}_0(t+1) = \alpha \sum_{n=1}^{N_o} \boldsymbol{\xi}_n(t) + (1-\alpha)\boldsymbol{\theta}_0(t), \quad (24b)$$

where  $0 < \alpha \leq 1$  is the damping.

At the end of this section, we formulate a probability distribution of discrete random vectors as one in the exponential family. Consider an  $N$  dimensional discrete random vector  $\mathbf{x} \in \mathbb{X}$ , where each component of  $\mathbf{x}$  takes only finite values,  $\mathbb{X} = \{\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N_x-1)}\}$ , and  $N_x \geq 2$  is the number of all possible vectors of  $\mathbf{x}$ . Denote the probability distribution of  $\mathbf{x}$  as  $p(\mathbf{x})$  and the probability of  $\mathbf{x}$  taking the value  $\mathbf{x}^{(i)}$  as  $p(\mathbf{x})|_{\mathbf{x}=\mathbf{x}^{(i)}} = p_i > 0, i \in \mathcal{Z}_{N_x-1}$ . Denote the set of probability distributions of  $\mathbf{x}$  as

$$\mathcal{X} = \left\{ p(\mathbf{x}) \mid p(\mathbf{x}) > 0, \mathbf{x} \in \mathbb{X}, \sum_{\mathbf{x} \in \mathbb{X}} p(\mathbf{x}) = 1 \right\}. \quad (25)$$

For the discrete probability distributions, let

$$t_{x,i} = \delta(\mathbf{x} - \mathbf{x}^{(i)}) = \begin{cases} 1, & \text{when } \mathbf{x} = \mathbf{x}^{(i)}, \\ 0, & \text{otherwise,} \end{cases} \quad (26)$$

where  $i \in \mathcal{Z}_{N_x-1}$ . Then, the probability distribution of  $\mathbf{x}$  can be rewritten as

$$p(\mathbf{x}) = \sum_{\mathbf{x} \in \mathbb{X}} p(\mathbf{x})|_{\mathbf{x}=\mathbf{x}^{(i)}} \delta(\mathbf{x} - \mathbf{x}^{(i)}) = \sum_{i=0}^{N_x-1} p_i t_{x,i}, \quad (27)$$

where  $\{p_i\}_{i=0}^{N_x-1}$  are positive values and constrained by  $\sum_{i=0}^{N_x-1} p_i = 1$ . Hence,  $\mathcal{X}$  has  $N_x - 1$  degrees of freedom and is a  $N_x - 1$  dimensional manifold [21]. Since the dimension of  $\mathcal{X}$  is  $N_x - 1$ , we define an  $N_x - 1$  dimensional parameter vector as  $\boldsymbol{\theta}_x = [\theta_{x,1}, \theta_{x,2}, \dots, \theta_{x,N_x-1}]^T$ , where each component is given by

$$\theta_{x,i} = \ln\left(\frac{p_i}{p_0}\right), i \in \mathcal{Z}_{N_x-1}^+. \quad (28)$$

Then,

$$p(\mathbf{x}) = \exp\{\boldsymbol{\theta}_x^T \mathbf{t}_x - \psi(\boldsymbol{\theta}_x)\}, \quad (29)$$

where  $\mathbf{t}_x = [t_{x,1}, t_{x,2}, \dots, t_{x,N_x-1}]^T$  is a random vector of  $N_x - 1$  dimension, and

$$\psi(\boldsymbol{\theta}_x) = -\ln p_0. \quad (30)$$

The above expresses  $\mathcal{X}$  is expressed in terms of an exponential family, and  $\boldsymbol{\theta}_x$  is the NP of  $p(\mathbf{x})$ .

---

### Algorithm 1: IGA-SD

---

**Input:** The a priori probability  $p_{\text{pr},k}(s_k), k \in \mathcal{Z}_{2K}^+$ , the received signal  $\mathbf{y}$ , the channel matrix  $\mathbf{G}$ , the alphabet  $\mathbb{S} = \{s^{(0)}, s^{(1)}, \dots, s^{(L-1)}\}$  for the components of  $\mathbf{s}$ , the noise power  $\sigma_z^2$  and the maximal iteration number  $t_{\text{max}}$ .

**Initialization:** set  $t = 0$ , set damping  $\alpha$ , where  $0 < \alpha \leq 1$ , initialize the EACSSs  $\boldsymbol{\theta}_n, n \in \mathcal{Z}_{2N_r}$ , which are defined in (36) and (42), zeros are sufficient for their initializations in general, calculate the NP  $d_{k,\ell}, k \in \mathcal{Z}_{2K}^+, \ell \in \mathcal{Z}_{L-1}^+$ , as (31);

**repeat**

1. Calculate  $\boldsymbol{\xi}_n(t), n \in \mathcal{Z}_{2N_r}^+$ , as (70) and (71);
2. Update the EACSSs as (72);
3.  $t = t + 1$ ;

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

**Output:** The probability of the approximate marginal,  $p_k(s_k|\mathbf{y})$ , is given by the probability of  $p_{0,k}(s_k; \boldsymbol{\theta}_{0,k}), k \in \mathcal{Z}_{2K}^+$ , which is given by (39). Then, the MPM detection is given by (8).

---

## IV. INFORMATION GEOMETRY APPROACH FOR SIGNAL DETECTION

As discussed in Sec. III,  $p_{\text{pr},k}(s_k), k \in \mathcal{Z}_{2K}^+$ , belong to the exponential family. Define a sufficient statistic as  $\mathbf{t}_k \triangleq [t_{k,1}, t_{k,2}, \dots, t_{k,L-1}]^T \in \mathbb{R}^{(L-1)}$ , where  $t_{k,\ell} \triangleq \delta(s_k - s^{(\ell)})$ ,  $k \in \mathcal{Z}_{2K}^+, \ell \in \mathcal{Z}_{L-1}^+$ . Define the NP as  $\mathbf{d}_k \triangleq [d_{k,1}, d_{k,2}, \dots, d_{k,L-1}]^T \in \mathbb{R}^{(L-1)}, k \in \mathcal{Z}_{2K}^+$ , and

$$d_{k,\ell} = \ln \frac{p_{\text{pr},k}(s_k)|_{s_k=s^{(\ell)}}}{p_{\text{pr},k}(s_k)|_{s_k=s^{(0)}}}, \ell \in \mathcal{Z}_{L-1}^+. \quad (31)$$

Then,  $p_{\text{pr},k}(s_k), k \in \mathcal{Z}_{2K}^+$ , can be expressed as

$$p_{\text{pr},k}(s_k) = \exp\{\mathbf{d}_k^T \mathbf{t}_k - \psi(\mathbf{d}_k)\}, \quad (32)$$

where  $\psi(\mathbf{d}_k) = -\ln(p_{\text{pr},k}(s_k)|_{s_k=s^{(0)}})$  is the free energy. Combining with (32), the *a posteriori* distribution  $p(\mathbf{s}|\mathbf{y})$  can be expressed as

$$\begin{aligned} p(\mathbf{s}|\mathbf{y}) &= \exp\left\{ \sum_{k=1}^{2K} \mathbf{d}_k^T \mathbf{t}_k + \sum_{n=1}^{2N_r} c_n(\mathbf{s}, y_n) - \psi_q \right\} \\ &= \exp\left\{ \mathbf{d}^T \mathbf{t} + \sum_{n=1}^{2N_r} c_n(\mathbf{s}, y_n) - \psi_q \right\}, \end{aligned} \quad (33)$$

where  $\mathbf{d} = [\mathbf{d}_1^T, \mathbf{d}_2^T, \dots, \mathbf{d}_{2K}^T]^T \in \mathbb{R}^{2K(L-1)}$ ,  $\mathbf{t} = [\mathbf{t}_1^T, \mathbf{t}_2^T, \dots, \mathbf{t}_{2K}^T]^T \in \mathbb{R}^{2K(L-1)}$ ,  $\psi_q$  is the normalization factor, and

$$c_n(\mathbf{s}, y_n) = -\frac{1}{2\sigma_z^2} (y_n - \mathbf{e}_n^T \mathbf{G} \mathbf{s})^2, \quad (34a)$$

$$\psi_q = \ln \left( \sum_{\mathbf{s} \in \mathbb{S}^{2K}} \exp \left\{ \mathbf{d}^T \mathbf{t} + \sum_{n=1}^{2N_r} c_n(\mathbf{s}, y_n) \right\} \right). \quad (34b)$$

According to (33), we can immediately define the OBM and the AMs as in the previous section. The OBM is defined as

$$\mathcal{M}_0 = \left\{ p_0(\mathbf{s}; \boldsymbol{\theta}_0) \mid \boldsymbol{\theta}_0 \in \mathbb{R}^{2K(L-1)} \right\}, \quad (35a)$$

$$\begin{aligned} p_0(\mathbf{s}; \boldsymbol{\theta}_0) &= \prod_{k=1}^{2K} p_{0,k}(s_k; \boldsymbol{\theta}_{0,k}) \\ &= \exp \left\{ \mathbf{d}^T \mathbf{t} + \boldsymbol{\theta}_0^T \mathbf{t} - \psi_0(\boldsymbol{\theta}_0) \right\}, \end{aligned} \quad (35b)$$

$$\begin{aligned} p_{0,k}(s_k; \boldsymbol{\theta}_{0,k}) &= \exp \left\{ \mathbf{d}_k^T \mathbf{t}_k + \boldsymbol{\theta}_{0,k}^T \mathbf{t}_k - \psi_0(\boldsymbol{\theta}_{0,k}) \right\} \\ &= \exp \left\{ \sum_{\ell=1}^{L-1} (d_{k,\ell} + \theta_{0,k,\ell}) \delta(s_k - s^{(\ell)}) \right\} \\ &\quad \times \exp \left\{ -\psi_0(\boldsymbol{\theta}_{0,k}) \right\}, \end{aligned} \quad (35c)$$

where

$$\boldsymbol{\theta}_0 = [\boldsymbol{\theta}_{0,1}^T, \boldsymbol{\theta}_{0,2}^T, \dots, \boldsymbol{\theta}_{0,2K}^T]^T \in \mathbb{R}^{2K(L-1)} \quad (36)$$

is the EACS of  $p_0(\mathbf{s}; \boldsymbol{\theta}_0)$ ,

$$\boldsymbol{\theta}_{0,k} = [\theta_{0,k,1}, \theta_{0,k,2}, \dots, \theta_{0,k,L-1}]^T \in \mathbb{R}^{(L-1)} \quad (37)$$

is the EACS of  $p_{0,k}(s_k; \boldsymbol{\theta}_{0,k})$ ,  $p_{0,k}(s_k; \boldsymbol{\theta}_{0,k})$  is the marginal distribution of  $s_k$ , the free energies  $\psi_0(\boldsymbol{\theta}_0)$  and  $\psi_0(\boldsymbol{\theta}_{0,k})$  are given by

$$\begin{aligned} \psi_0(\boldsymbol{\theta}_0) &= \sum_{k=1}^K \psi_0(\boldsymbol{\theta}_{0,k}) \\ &= \ln \left( \sum_{\mathbf{s} \in \mathbb{S}^{2K}} \exp \left\{ \mathbf{d}^T \mathbf{t} + \boldsymbol{\theta}_0^T \mathbf{t} \right\} \right), \end{aligned} \quad (38a)$$

$$\begin{aligned} \psi_0(\boldsymbol{\theta}_{0,k}) &= \ln \left( \sum_{s_k \in \mathbb{S}} \exp \left\{ \mathbf{d}_k^T \mathbf{t}_k + \boldsymbol{\theta}_{0,k}^T \mathbf{t}_k \right\} \right) \\ &= \ln \left( 1 + \sum_{\ell=1}^{L-1} \exp \left\{ d_{k,\ell} + \theta_{0,k,\ell} \right\} \right). \end{aligned} \quad (38b)$$

Given  $p_0(\mathbf{s}; \boldsymbol{\theta}_0)$  and its marginals  $p_{0,k}(s_k; \boldsymbol{\theta}_{0,k})$ , the probability of signal  $s_k$ ,  $k \in \mathcal{Z}_{2K}^+$ , can be expressed in a more explicit way as

$$p_{0,k}(s_k; \boldsymbol{\theta}_{0,k}) \Big|_{s_k=s^{(0)}} \stackrel{(a)}{=} \frac{1}{1 + \sum_{\ell=1}^{L-1} \exp \{d_{k,\ell} + \theta_{0,k,\ell}\}}, \quad (39a)$$

$$p_{0,k}(s_k; \boldsymbol{\theta}_{0,k}) \Big|_{s_k=s^{(\ell)}} \stackrel{(b)}{=} \frac{\exp \{d_{k,\ell} + \theta_{0,k,\ell}\}}{1 + \sum_{\ell=1}^{L-1} \exp \{d_{k,\ell} + \theta_{0,k,\ell}\}}, \quad (39b)$$

where  $\ell \in \mathcal{Z}_{L-1}^+$  in (39b), and (a) and (b) come from (35c) and (38b). The probability of  $p_0(\mathbf{s}; \boldsymbol{\theta}_0)$  can be then expressed more explicitly by using (35b). Also, from (39), we can conversely use the marginal probability of  $s_k$  to express the EACS  $\boldsymbol{\theta}_{0,k}$  of  $p_{0,k}(s_k; \boldsymbol{\theta}_{0,k})$ ,  $k \in \mathcal{Z}_{2K}^+$ , i.e.,

$$\theta_{0,k,\ell} = \ln \frac{p_{0,k}(s_k; \boldsymbol{\theta}_{0,k}) \Big|_{s_k=s^{(\ell)}}}{p_{0,k}(s_k; \boldsymbol{\theta}_{0,k}) \Big|_{s_k=s^{(0)}}} - d_{k,\ell}, \ell \in \mathcal{Z}_{L-1}^+. \quad (40)$$

Then, the EACS  $\boldsymbol{\theta}_0$  of  $p_0(\mathbf{s}; \boldsymbol{\theta}_0)$  can be also obtained. This relationship will be used later in this section.  $2N_r$  AMs are defined, where the  $n$ -th of them is given by

$$\mathcal{M}_n = \left\{ p_n(\mathbf{s}; \boldsymbol{\theta}_n) \mid \boldsymbol{\theta}_n \in \mathbb{R}^{2K(L-1)} \right\}, \quad (41a)$$

$$p_n(\mathbf{s}; \boldsymbol{\theta}_n) = \exp \left\{ \mathbf{d}^T \mathbf{t} + \boldsymbol{\theta}_n^T \mathbf{t} + c_n(\mathbf{s}, y_n) - \psi_n(\boldsymbol{\theta}_n) \right\}, \quad (41b)$$

where

$$\boldsymbol{\theta}_n = [\boldsymbol{\theta}_{n,1}^T, \boldsymbol{\theta}_{n,2}^T, \dots, \boldsymbol{\theta}_{n,2K}^T]^T \in \mathbb{R}^{2K(L-1)} \quad (42)$$

is the EACS of  $p_n(\mathbf{s}; \boldsymbol{\theta}_n)$ ,

$$\boldsymbol{\theta}_{n,k} = [\theta_{n,k,1}, \theta_{n,k,2}, \dots, \theta_{n,k,L-1}]^T \in \mathbb{R}^{(L-1)}, \quad (43)$$

and the free energy  $\psi_n$  is given by

$$\psi_n(\boldsymbol{\theta}_n) = \ln \left( \sum_{\mathbf{s} \in \mathbb{S}^{2K}} \exp \left\{ \mathbf{d}^T \mathbf{t} + \boldsymbol{\theta}_n^T \mathbf{t} + c_n(\mathbf{s}, y_n) \right\} \right). \quad (44)$$

From the definitions, it is not difficult to check that the OBM and the AMs are all  $e$ -flat.

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

$$\mathcal{S} = \left\{ p(\mathbf{s}) \mid p(\mathbf{s}) > 0, \mathbf{s} \in \mathbb{S}^{2K}, \sum_{\mathbf{s} \in \mathbb{S}^{2K}} p(\mathbf{s}) = 1 \right\}. \quad (45)$$

$\mathcal{S}$  is then a  $L^{2K} - 1$  dimensional manifold and forms an exponential family. Then, it can be readily checked that the *a posteriori* distribution  $p(\mathbf{s}|\mathbf{y})$  belongs to  $\mathcal{S}$  since  $p(\mathbf{s}|\mathbf{y})$  is a particular probability distribution of  $\mathbf{s}$ . Similarly, 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.

We now present the properties of the  $m$ -projection of any  $p(\mathbf{s}) \in \mathcal{S}$ , onto the OBM  $\mathcal{M}_0$ , which inspires us to approximate the  $m$ -projection of  $p_n(\mathbf{s}; \boldsymbol{\theta}_n)$  onto the OBM  $\mathcal{M}_0$ . According to the Section III, given  $p(\mathbf{s}) \in \mathcal{S}$  and the OBM  $\mathcal{M}_0$ , which is an  $e$ -flat submanifold of  $\mathcal{S}$ , the  $m$ -projection of  $p(\mathbf{s})$  onto  $\mathcal{M}_0$  is obtained by the following minimization problem,

$$\boldsymbol{\theta}_0^* = \arg \min_{\boldsymbol{\theta}_0} D_{\text{KL}} \{p(\mathbf{s}) : p_0(\mathbf{s}; \boldsymbol{\theta}_0)\}, \quad (46)$$

where the K-L divergence is given by

$$\begin{aligned} D_{\text{KL}} \{p(\mathbf{s}) : p_0(\mathbf{s}; \boldsymbol{\theta}_0)\} &= \mathbb{E}_{p(\mathbf{s})} \left\{ \ln \frac{p(\mathbf{s})}{p_0(\mathbf{s}; \boldsymbol{\theta}_0)} \right\} \\ &= C_p - \sum_{\mathbf{s} \in \mathbb{S}^{2K}} p(\mathbf{s}) \ln(p_0(\mathbf{s}; \boldsymbol{\theta}_0)), \end{aligned} \quad (47)$$

where  $C_p = \sum_{\mathbf{s} \in \mathbb{S}^{2K}} p(\mathbf{s}) \ln p(\mathbf{s})$  is a constant independent of  $\boldsymbol{\theta}_0$ . We then have the following theorem.

**Theorem 1.** *Given  $p(\mathbf{s}) \in \mathcal{S}$ , and the  $e$ -flat  $\mathcal{M}_0 \subseteq \mathcal{S}$ , the  $m$ -projection of  $p(\mathbf{s})$  onto  $\mathcal{M}_0$  is unique. Moreover,  $p_0(\mathbf{s}; \boldsymbol{\theta}_0^*)$  is*

the  $m$ -projection of  $p(\mathbf{s})$  onto  $\mathcal{M}_0$  if and only if the following relationship holds,

$$\boldsymbol{\eta} = \boldsymbol{\eta}_0(\boldsymbol{\theta}_0^*), \quad (48)$$

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

$$\boldsymbol{\eta} = \mathbb{E}_{p(\mathbf{s})} \{\mathbf{t}\} = \sum_{\mathbf{s} \in \mathbb{S}^{2K}} \mathbf{t} p(\mathbf{s}), \quad (49a)$$

$$\boldsymbol{\eta}_0(\boldsymbol{\theta}_0^*) = \mathbb{E}_{p_0(\mathbf{s}; \boldsymbol{\theta}_0^*)} \{\mathbf{t}\} = \sum_{\mathbf{s} \in \mathbb{S}^{2K}} \mathbf{t} p_0(\mathbf{s}; \boldsymbol{\theta}_0^*). \quad (49b)$$

*Proof.* See Appendix A.  $\square$

Define  $2K$  discrete random vectors of  $2K - 1$  dimensions, where the  $k$ -th of them, denoted as  $\mathbf{s}_{\setminus k}$ , is obtained by removing the  $k$ -th element, i.e.,  $s_k$ , of  $\mathbf{s}$ ,  $k \in \mathcal{Z}_{2K}^+$ . Then, we can obtain  $\mathbf{s}_{\setminus k} \in \mathbb{S}^{2K-1}$ ,  $k \in \mathcal{Z}_{2K}^+$ , and the marginal probability distribution of  $s_k$  given the joint probability distribution  $p(\mathbf{s})$  is

$$\begin{aligned} p_k(s_k) &\triangleq \sum_{\mathbf{s}_{\setminus k} \in \mathbb{S}^{2K-1}} p(\mathbf{s}) \\ &= \sum_{s_1 \in \mathbb{S}} \cdots \sum_{s_{k-1} \in \mathbb{S}} \sum_{s_{k+1} \in \mathbb{S}} \cdots \sum_{s_{2K} \in \mathbb{S}} p(\mathbf{s}), k \in \mathcal{Z}_{2K}^+. \end{aligned} \quad (50)$$

From the definition of  $p_0(\mathbf{s}; \boldsymbol{\theta}_0)$  in (35b), we denote the marginals of  $p_0(\mathbf{s}; \boldsymbol{\theta}_0^*)$  in Theorem 1 as  $p_{0,k}(s_k; \boldsymbol{\theta}_{0,k}^*)$ ,  $k \in \mathcal{Z}_{2K}^+$ , where  $\boldsymbol{\theta}_{0,k}^* = [\theta_{0,k,1}^*, \theta_{0,k,2}^*, \dots, \theta_{0,k,L-1}^*]^T \in \mathbb{R}^{(L-1)}$  and  $\boldsymbol{\theta}_0^* = [(\boldsymbol{\theta}_{0,1}^*)^T, (\boldsymbol{\theta}_{0,2}^*)^T, \dots, (\boldsymbol{\theta}_{0,2K}^*)^T]^T$ . Combining Theorem 1, we have the following corollary.

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

$$p_k(s_k) = p_{0,k}(s_k; \boldsymbol{\theta}_{0,k}^*), s_k \in \mathbb{S}, k \in \mathcal{Z}_{2K}^+. \quad (51)$$

Meanwhile, the EACS of the  $m$ -projection is given by  $\boldsymbol{\theta}_0^* = [(\boldsymbol{\theta}_{0,1}^*)^T, (\boldsymbol{\theta}_{0,2}^*)^T, \dots, (\boldsymbol{\theta}_{0,2K}^*)^T]^T$ , where  $\boldsymbol{\theta}_{0,k}^* = [\theta_{0,k,1}^*, \theta_{0,k,2}^*, \dots, \theta_{0,k,L-1}^*]^T$ ,  $k \in \mathcal{Z}_{2K}^+$ , and

$$\theta_{0,k,\ell}^* = \ln \frac{p_k(s_k) \big|_{s_k=s(\ell)}}{p_k(s_k) \big|_{s_k=s(0)}} - d_{k,\ell}, \ell \in \mathcal{Z}_{L-1}^+. \quad (52)$$

*Proof.* See Appendix B.  $\square$

Given  $p_n(\mathbf{s}; \boldsymbol{\theta}_n) \in \mathcal{M}_n$ ,  $n \in \mathcal{Z}_{2N_r}^+$ , from Theorem 1 we can obtain that its  $m$ -projection onto  $\mathcal{M}_0$  is unique since  $p_n(\mathbf{s}; \boldsymbol{\theta}_n) \in \mathcal{S}$ . Denote the  $m$ -projection of  $p_n(\mathbf{s}; \boldsymbol{\theta}_n)$  onto  $\mathcal{M}_0$  as  $p_0(\mathbf{s}; \boldsymbol{\theta}_{0n})$ ,  $n \in \mathcal{Z}_{2N_r}^+$ , where  $\boldsymbol{\theta}_{0n} = [\boldsymbol{\theta}_{0n,1}^T, \boldsymbol{\theta}_{0n,2}^T, \dots, \boldsymbol{\theta}_{0n,2K}^T]^T \in \mathbb{R}^{2K(L-1)}$  and  $\boldsymbol{\theta}_{0n,k} = [\theta_{0n,k,1}, \theta_{0n,k,2}, \dots, \theta_{0n,k,L-1}]^T \in \mathbb{R}^{(L-1)}$ ,  $k \in \mathcal{Z}_{2K}^+$ . From Corollary 1, for any  $n \in \mathcal{Z}_{2N_r}^+$ , the  $m$ -projection

$p_0(\mathbf{s}; \boldsymbol{\theta}_{0n})$  is determined by the marginal probability distribution  $p_{n,k}(s_k; \boldsymbol{\theta}_n)$ ,  $k \in \mathcal{Z}_{2K}^+$ , where

$$p_{n,k}(s_k; \boldsymbol{\theta}_n) \triangleq \sum_{\mathbf{s}_{\setminus k} \in \mathbb{S}^{2K-1}} p_n(\mathbf{s}; \boldsymbol{\theta}_n). \quad (53)$$

And we have

$$\theta_{0n,k,\ell} = \ln \frac{p_{n,k}(s_k; \boldsymbol{\theta}_n) \big|_{s_k=s(\ell)}}{p_{n,k}(s_k; \boldsymbol{\theta}_n) \big|_{s_k=s(0)}} - d_{k,\ell}, \quad (54)$$

where  $n \in \mathcal{Z}_{2N_r}^+$ ,  $k \in \mathcal{Z}_{2K}^+$  and  $\ell \in \mathcal{Z}_{L-1}^+$ . Nevertheless, it is relatively difficult to obtain the closed-form solution of the marginal probability distribution  $p_{n,k}(s_k; \boldsymbol{\theta}_n)$  since the calculation is of exponential-complexity. In this work, we solve this problem by calculating the approximations of the marginals  $p_{n,k}(s_k; \boldsymbol{\theta}_n)$ ,  $k \in \mathcal{Z}_{2K}^+$ ,  $n \in \mathcal{Z}_{2N_r}^+$ , using the central-limit-theorem (CLT).

From the definition of  $p_n(\mathbf{s}; \boldsymbol{\theta}_n)$  in (41b), its marginals can be expressed as

$$\begin{aligned} p_{n,k}(s_k; \boldsymbol{\theta}_n) &= \sum_{\mathbf{s}_{\setminus k} \in \mathbb{S}^{2K-1}} \exp \left\{ (\mathbf{d} + \boldsymbol{\theta}_n)^T \mathbf{t} + c_n(\mathbf{s}, y_n) - \psi_n \right\} \\ &\stackrel{(a)}{\propto} \exp \left\{ (\mathbf{d}_k + \boldsymbol{\theta}_{n,k})^T \mathbf{t}_k \right\} q(y_n, s_k), \end{aligned} \quad (55)$$

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

$$\begin{aligned} q(y_n, s_k) & \\ &= \sum_{\mathbf{s}_{\setminus k} \in \mathbb{S}^{2K-1}} \exp \left\{ \sum_{k'=1, k' \neq k}^{2K} (\mathbf{d}_{k'} + \boldsymbol{\theta}_{n,k'})^T \mathbf{t}_{k'} + c_n(\mathbf{s}, y_n) \right\}. \end{aligned} \quad (56)$$

Note that the proportions in the second line of (55) and the third line of (57) 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  $p_{n,k}(s_k; \boldsymbol{\theta}_n)$ . We will not repeat this property when a similar situation arises in the rest of this paper. In the last line of (55), 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 then can 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 (56), we can obtain

$$\begin{aligned} q(y_n, s_k) & \\ &= \sum_{\mathbf{s}_{\setminus k} \in \mathbb{S}^{2K-1}} \left( \prod_{k'=1, k' \neq k}^{2K} \exp \left\{ (\mathbf{d}_{k'} + \boldsymbol{\theta}_{n,k'})^T \mathbf{t}_{k'} \right\} \right) \\ &\quad \times \exp \left\{ -\frac{1}{2\sigma_z^2} (y_n - \mathbf{e}_n^T \mathbf{G} \mathbf{s})^2 \right\} \\ &\stackrel{(a)}{\propto} \sum_{\mathbf{s}_{\setminus k} \in \mathbb{S}^{2K-1}} \left( \prod_{k'=1, k' \neq k}^{2K} p_{0,k'}(s_{k'}; \boldsymbol{\theta}_{n,k'}) f_G(y_n; \mathbf{e}_n^T \mathbf{G} \mathbf{s}, \sigma_z^2) \right), \end{aligned} \quad (57)$$

where  $\mathbf{G}$  is defined in (5), (a) is obtained by adding the constant independent with  $s_k$  and  $y_n$ ,  $p_{0,k'}(s_{k'}; \boldsymbol{\theta}_{n,k'})$  is defined by (35c), and  $f_G(x; \mu, \sigma^2)$  denotes the PDF of a real

Gaussian distribution  $\mathcal{N}(\mu, \sigma^2)$  for a real random variable  $x$ . Inspired by the last line of (57), 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 a given  $s_k$ ,

$$\begin{aligned} Y_{n,k} &= \mathbf{e}_n^T \mathbf{G} \mathbf{s} + w = g_{n,k} s_k + \sum_{k'=1, k' \neq k}^{2K} g_{n,k'} s_{k'} + w \\ &= \sum_{k'=1, k' \neq k}^{2K} g_{n,k'} s_{k'} + w'_{n,k}, \end{aligned} \quad (58)$$

where  $s_k$  is considered as a determinate (also known/given) constant,  $g_{n,k}$  is the  $(n, k)$ -th component of  $\mathbf{G}$ ,  $g_{n,k}$  is also considered as a determinate and known 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  $p(\mathbf{s}_{\setminus k}) = \prod_{k' \neq k} p_{0,k'}(s_{k'}; \boldsymbol{\theta}_{n,k'})$ ,  $w \sim \mathcal{N}(0, \sigma_z^2)$  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  $\{s_{k'}\}_{k' \neq k}$ . Briefly, for  $Y_{n,k}$  the subscript  $n$  determines which row of  $\mathbf{G}$  is multiplied by  $\mathbf{s}$ , and the subscript  $k$  determines which component of  $\mathbf{s}$  is considered deterministic. In this case, it is not difficult to obtain that the PDF of  $Y_{n,k}$  is given by [27, Sec. 6.1.2]

$$\begin{aligned} & f(Y_{n,k}) \\ &= \sum_{\mathbf{s}_{\setminus k} \in \mathbb{S}^{2K-1}} \left( p(\mathbf{s}_{\setminus k}) f_G \left( Y_{n,k} - \sum_{k' \neq k} g_{n,k'} s_{k'}, g_{n,k} s_k, \sigma_z^2 \right) \right) \\ &= \sum_{\mathbf{s}_{\setminus k} \in \mathbb{S}^{2K-1}} \left( p(\mathbf{s}_{\setminus k}) f_G(Y_{n,k}; \mathbf{e}_n^T \mathbf{G} \mathbf{s}, \sigma_z^2) \right), \end{aligned} \quad (59)$$

which will be equal to the last line of (57) after we set the value of  $Y_{n,k}$  as  $Y_{n,k} = y_n$ . Since although the terms in the summation in (58) are independent each other, they do not have the same distribution. Thus, the conventional CLT may not apply directly. We next apply Lyapunov CLT to impose a condition on the values of  $g_{n,k}$  in matrix  $\mathbf{G}$  and the variances of the random variables in (58) so that  $Y_{n,k}$  converges in distribution to a real Gaussian random variable. To do so, let us first see Lyapunov CLT.

**Lemma 1** (Lyapunov central-limit-theorem [28]). *Suppose  $\{X_n\}_{n=1}^N$  are independent real random variables, each with finite expected value  $\mu_n$  and variance  $\sigma_n^2$ . Denote the random variable  $S = \sum_{n=1}^N X_n$  and its expected value and variance as  $\tilde{\mu} = \sum_{n=1}^N \mu_n$  and  $\tilde{\sigma}^2 = \sum_{n=1}^N \sigma_n^2$ , respectively. Suppose for some positive  $\delta$ , Lyapunov's condition*

$$\lim_{N \rightarrow \infty} \frac{1}{\tilde{\sigma}^{2+\delta}} \sum_{n=1}^N \mathbb{E} \left\{ |X_n - \mu_n|^{2+\delta} \right\} = 0 \quad (60)$$

*holds. Then,  $S$  converges in distribution to a real Gaussian random variable  $\tilde{S}$ , as  $N$  tends to infinity, and*

$$S \xrightarrow{d} \tilde{S} \sim \mathcal{N}(\tilde{\mu}, \tilde{\sigma}^2). \quad (61)$$

Given the probability distribution  $p_{0,k'}(s_{k'}; \boldsymbol{\theta}_{n,k'})$  of  $s_{k'}, k' \in \mathcal{Z}_{2K}^+ \setminus \{k\}$ , in (58), by using (39) the expected value and the variance of  $s_{k'}$  are given by

$$\begin{aligned} \mu_{n,k'} &= \sum_{s_{k'} \in \mathbb{S}} s_{k'} p_{0,k'}(s_{k'}; \boldsymbol{\theta}_{n,k'}) \\ &= \frac{s^{(0)} + \sum_{\ell=1}^{L-1} s^{(\ell)} \exp\{d_{k',\ell} + \theta_{n,k',\ell}\}}{1 + \sum_{\ell=1}^{L-1} \exp\{d_{k',\ell} + \theta_{n,k',\ell}\}}, \end{aligned} \quad (62a)$$

$$\begin{aligned} v_{n,k'} &= \sum_{s_{k'} \in \mathbb{S}} s_{k'}^2 p_{0,k'}(s_{k'}; \boldsymbol{\theta}_{n,k'}) - \mu_{n,k'}^2 \\ &= \frac{(s^{(0)})^2 + \sum_{\ell=1}^{L-1} (s^{(\ell)})^2 \exp\{d_{k',\ell} + \theta_{n,k',\ell}\}}{1 + \sum_{\ell=1}^{L-1} \exp\{d_{k',\ell} + \theta_{n,k',\ell}\}} - \mu_{n,k'}^2. \end{aligned} \quad (62b)$$

Meanwhile, since  $\{s_{k'}\}_{k' \neq k}$  and  $w'_{n,k}$  are independent in (58), 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

$$\mathbb{E}\{Y_{n,k}\} = \sum_{k'=1, k' \neq k}^{2K} g_{n,k'} \mu_{n,k'} + g_{n,k} s_k, \quad (63a)$$

$$\mathbb{V}\{Y_{n,k}\} = \sum_{k'=1, k' \neq k}^{2K} g_{n,k'}^2 v_{n,k'} + \sigma_z^2, \quad (63b)$$

We then have the following theorem.

**Theorem 2.** *If the following condition*

$$\lim_{K \rightarrow \infty} \frac{1}{2K} \sum_{k'=1, k' \neq k}^{2K} g_{n,k'}^2 v_{n,k'} = \zeta > 0 \quad (64)$$

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

$$Y_{n,k} \xrightarrow{d} \tilde{Y}_{n,k} \sim \mathcal{N}(\mathbb{E}\{Y_{n,k}\}, \mathbb{V}\{Y_{n,k}\}). \quad (65)$$

*Proof.* See Appendix C.  $\square$

Intuitively, the condition (64) means that as  $K$  (or, equivalently,  $2K - 1$ ) 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 (58) does not tend to zero, where  $g_{n,k'}$  is the  $(n, k')$ -th component of  $\mathbf{G}$  defined in (5), or  $\tilde{s}_{n,k'}$  does not tend to be a deterministic value. This guarantees that the CLT holds. When  $2K$  is large, from Theorem 2,  $q(y_n, s_k)$  is approximately proportional to  $f_G(\tilde{Y}_{n,k}; \mathbb{E}\{Y_{n,k}\}, \mathbb{V}\{Y_{n,k}\})|_{\tilde{Y}_{n,k}=y_n}$ , and thus we can obtain

$$\begin{aligned} & p_{n,k}(s_k; \boldsymbol{\theta}_n) \\ & \stackrel{(a)}{\propto} \exp \left\{ (\mathbf{d}_k + \boldsymbol{\theta}_{n,k})^T \mathbf{t}_k - \frac{(y_n - \mathbb{E}\{Y_{n,k}\})^2}{2\mathbb{V}\{Y_{n,k}\}} \right\} \\ & = \exp \left\{ (\mathbf{d}_k + \boldsymbol{\theta}_{n,k})^T \mathbf{t}_k - \frac{(g_{n,k} s_k - \tilde{\mu}_{n,k})^2}{2\mathbb{V}\{Y_{n,k}\}} \right\}, \end{aligned} \quad (66)$$

where  $s_k \in \mathbb{S}, k \in \mathcal{Z}_{2K}^+, n \in \mathcal{Z}_{2N_r}^+$ , (a) is obtained by removing the constant 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

$$\tilde{\mu}_{n,k} \triangleq y_n - \sum_{k'=1, k' \neq k}^{2K} g_{n,k'} \mu_{n,k'}, \quad (67)$$



As a summary, when  $2K$  is large we approximately have

$$p_{n,k}(s_k; \boldsymbol{\theta}_n) \Big|_{s_k=s^{(0)}} = C_{n,k} \exp \left\{ -\frac{(g_{n,k}s^{(0)} - \tilde{\mu}_{n,k})^2}{2\mathbb{V}\{Y_{n,k}\}} \right\}, \quad (68a)$$

$$p_{n,k}(s_k; \boldsymbol{\theta}_n) \Big|_{s_k=s^{(\ell)}} = C_{n,k} \exp \left\{ d_{k,\ell} + \theta_{n,k,\ell} - \frac{(g_{n,k}s^{(\ell)} - \tilde{\mu}_{n,k})^2}{2\mathbb{V}\{Y_{n,k}\}} \right\}, \quad (68b)$$

where  $C_{n,k}$  is the normalization factor, and  $\ell \in \mathcal{Z}_{L-1}^+$  in (68b). Combining (54), we can immediately obtain that

$$\theta_{0n,k,\ell} = \frac{g_{n,k}(s^{(0)} - s^{(\ell)}) [g_{n,k}(s^{(0)} + s^{(\ell)}) - 2\tilde{\mu}_{n,k}]}{2\mathbb{V}\{Y_{n,k}\}} + \theta_{n,k,\ell}, \quad (69)$$

where  $\ell \in \mathcal{Z}_{L-1}^+$ ,  $k \in \mathcal{Z}_{2K}^+$ , and  $n \in \mathcal{Z}_{2N_r}^+$ . Hence, we obtain an approximate solution of the  $m$ -projection  $p_0(\mathbf{s}; \boldsymbol{\theta}_{0n})$ ,  $n \in \mathcal{Z}_{2N_r}^+$ . From Theorem 2, it is not difficult to check that when the condition (64) holds, (69) is asymptotically accurate as  $K$  goes infinity. From (22), the approximation term  $\boldsymbol{\xi}_n = \boldsymbol{\theta}_{0n} - \boldsymbol{\theta}_n$  can be then expressed as

$$\boldsymbol{\xi}_n = [\boldsymbol{\xi}_{n,1}^T, \boldsymbol{\xi}_{n,2}^T, \dots, \boldsymbol{\xi}_{n,2K}^T]^T \quad (70a)$$

$$\boldsymbol{\xi}_{n,k} = [\xi_{n,k,1}, \xi_{n,k,2}, \dots, \xi_{n,k,L-1}]^T, \quad (70b)$$

$$\xi_{n,k,\ell} = \frac{g_{n,k}(s^{(0)} - s^{(\ell)}) [g_{n,k}(s^{(0)} + s^{(\ell)}) - 2\tilde{\mu}_{n,k}]}{2\mathbb{V}\{Y_{n,k}\}}, \quad (70c)$$

where  $n \in \mathcal{Z}_{2N_r}^+$ ,  $k \in \mathcal{Z}_{2K}^+$ , and  $\ell \in \mathcal{Z}_{L-1}^+$ . We give the detailed expression of  $\xi_{n,k,\ell}$  in (71), where  $g_{n,k}$  is the  $(n, k)$ -th component of the real-valued channel matrix  $\mathbf{G}$  in (6),  $\{s^{(\ell)}\}_{\ell=0}^{L-1}$  defined below (6) are the constellation points for the components of  $\mathbf{s}$ ,  $y_n$  is the  $n$ -th component of the received signal  $\mathbf{y}$  in (6),  $d_{k,\ell}$  is the NP defined by (31), and  $\sigma_z^2$  is the noise variance of  $\mathbf{z}$  in (6).

After the approximate  $\boldsymbol{\xi}_n$  is obtained, we update the EACSSs of  $p_n(\mathbf{s}; \boldsymbol{\theta}_n)$ ,  $n \in \mathcal{Z}_{2N_r}$ , as

$$\boldsymbol{\theta}_n(t+1) = \alpha \sum_{n'=1, n' \neq n}^{2N_r} \boldsymbol{\xi}_{n'}(t) + (1-\alpha) \boldsymbol{\theta}_n(t), n \in \mathcal{Z}_{2N_r}^+, \quad (72a)$$

$$\boldsymbol{\theta}_0(t+1) = \alpha \sum_{n=1}^{2N_r} \boldsymbol{\xi}_n(t) + (1-\alpha) \boldsymbol{\theta}_0(t), \quad (72b)$$

where  $0 < \alpha \leq 1$  is the damping, and repeat the  $m$ -projections, calculating  $\boldsymbol{\xi}_n$  and updating until convergence. We summarize the IGA-SD in Algorithm 1. The computational complexity (the number of real-valued multiplications) of the IGA-SD is  $\mathcal{O}(16N_r K(L+1))$  (the number of real-valued multiplications) 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.

## V. SIMULATION RESULTS

In this section, we provide simulation results to illustrate the performance of the proposed IGA-SD. The uncoded bit error rate (BER) is adopted as the performance metric. In our simulations, 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. We average our results for 1000 realizations of the channel matrix  $\mathbf{G}$ , which is generated by the widely adopted QuaDRiGa [29]. We set the simulation scenario to "3GPP\_38.901\_UMa\_NLOS", and the main parameters for the simulations are summarized in Table I. The BS is located at

TABLE I  
PARAMETER SETTINGS OF THE SIMULATION

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

$(0, 0, 25)$ . The users are randomly generated in a  $120^\circ$  sector with radius  $r = 200\text{m}$  around  $(0, 0, 1.5)$ . The channel matrix is normalized as  $\mathbb{E}\{\|\mathbf{G}\|_F^2\} = N_r K$ . The average power of the transmitted symbol of each user is normalized to 1, and the SNR is set as  $\text{SNR} = \frac{K}{\sigma_z^2}$ . Based on the received signal model (6), we compare the proposed IGA-SD with the following detectors.

**LMMSE:** The linear minimum-mean-squared error (LMMSE) detector with hard-decision. The LMMSE detector is given by

$$\mathbf{s}_{\text{LMMSE}} = (\mathbf{G}^T \mathbf{G} + \sigma_z^2 \mathbf{I})^{-1} \mathbf{G}^T \mathbf{y}. \quad (73)$$

Then, a component-wise hard-decision is performed as

$$s_{k,\text{MMSE}} = \arg \min_{s_k \in \mathbb{S}} |s_k - [\mathbf{s}_{\text{LMMSE}}]_k|^2, k \in \mathcal{Z}_{2K}^+. \quad (74)$$

**EP:** The expectation propagation detector proposed in [8], where the hard-decision is also performed.

**AMP:** Approximate message passing algorithm proposed in [30]. AMP can obtain the approximations of the marginals of the *a posteriori* distribution  $p(\mathbf{s}|\mathbf{y})$ . Thus, AMP is used as an MPM detector ((8)).

The computational complexity of the LMMSE detector is  $\mathcal{O}(8(2N_r K^2 + K^3))$  [8]. The computational complexity of the EP detector and AMP are  $\mathcal{O}(8(N_r K^2 + K^3))$  and  $\mathcal{O}(8(N_r K))$  per iteration, respectively [8], [30]. 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-SD 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-SD compared with LMMSE, EP and AMP. The iteration numbers of IGA-SD, EP and AMP are set as 10, 10 and 30, and 10 and 30, respectively. Meanwhile, the convergence performance of the iterative algorithms at  $\text{SNR} = 5\text{dB}$  is shown in Fig. 2. From Fig. 1, we can find that all the iterative algorithms outperform the LMMSE detector within

$$\xi_{n,k,\ell} = \frac{g_{n,k}(s^{(0)} - s^{(\ell)}) \left\{ g_{n,k}(s^{(0)} + s^{(\ell)}) - 2 \left[ y_n - \sum_{k'=1, k' \neq k}^{2K} g_{n,k'} \left( \frac{s^{(0)} + \sum_{\ell=1}^{L-1} s^{(\ell)} \exp\{d_{k',\ell} + \theta_{n,k',\ell}\}}{1 + \sum_{\ell=1}^{L-1} \exp\{d_{k',\ell} + \theta_{n,k',\ell}\}} \right) \right] \right\}}{2\mathbb{V}\{Y_{n,k}\}} \quad (71a)$$

$$\mathbb{V}\{Y_{n,k}\} = \sum_{k'=1, k' \neq k}^{2K} g_{n,k'}^2 \left[ \frac{(s^{(0)})^2 + \sum_{\ell=1}^{L-1} (s^{(\ell)})^2 \exp\{d_{k',\ell} + \theta_{n,k',\ell}\}}{1 + \sum_{\ell=1}^{L-1} \exp\{d_{k',\ell} + \theta_{n,k',\ell}\}} - \left( \frac{s^{(0)} + \sum_{\ell=1}^{L-1} s^{(\ell)} \exp\{d_{k',\ell} + \theta_{n,k',\ell}\}}{1 + \sum_{\ell=1}^{L-1} \exp\{d_{k',\ell} + \theta_{n,k',\ell}\}} \right)^2 \right] + \sigma_z^2 \quad (71b)$$

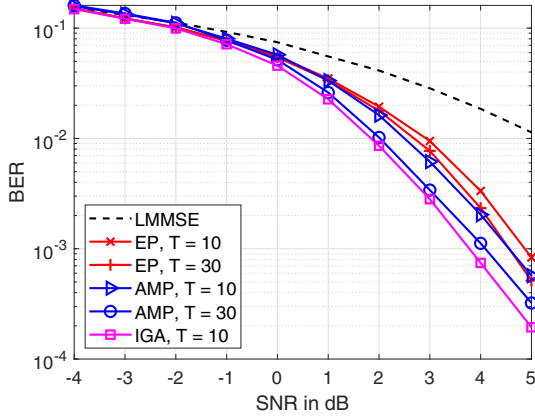


Fig. 1. BER performance of IGA compared with AMP, EP and LMMSE under 4-QAM.

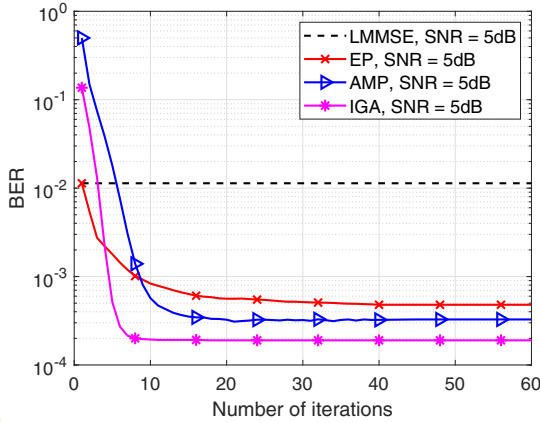


Fig. 2. Convergence performance of IGA compared with EP and AMP at SNR = 5 dB under 4-QAM.

limited iteration numbers. For BER =  $10^{-3}$ , the SNR gains of the IGA-SD with 10 iterations compared to the AMP with 10 and 30 iterations are around 0.7dB and 0.3dB, respectively. Meanwhile, IGA-SD with 10 iterations can improve the EP performance with 10 and 30 iterations in 1dB and 0.7dB for BER =  $10^{-3}$ , respectively. From Fig. 2, it can be found that in the case with 4-QAM and SNR = 5dB, the IGA-SD requires around 10 iterations to converge and achieves the lowest BER performance. AMP and EP require about 25 and 45 iterations to converge, respectively. The decrease in BER

is minor after 30 iterations for EP. Moreover, we can find that the BER performance of EP with one iteration is equal to that of LMMSE detector. This can be attributed to that the EP detector with one iteration is equivalent to the LMMSE detector [8].

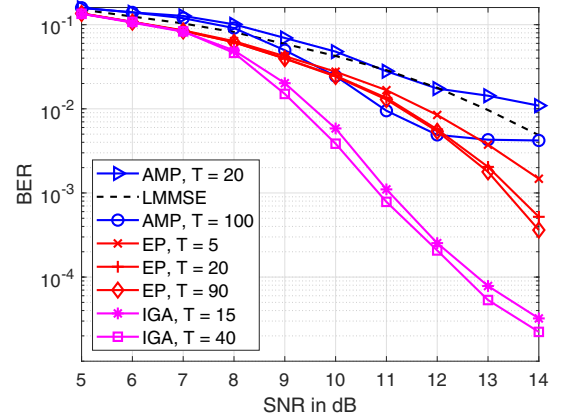


Fig. 3. BER performance of IGA compared with AMP, EP and LMMSE under 16-QAM.

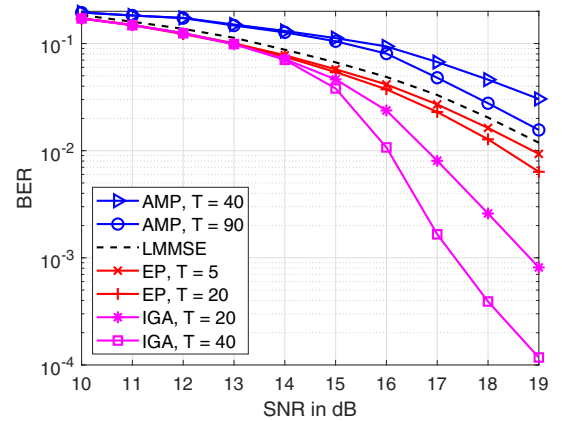


Fig. 4. BER performance of IGA compared with AMP, EP and LMMSE under 64-QAM.

Fig. 3 and 4 show the BER performance for 16-QAM and 64-QAM, respectively. From Fig. 3, we can find that the BER performance of LMMSE outperforms that of the AMP with 20 iterations. Meanwhile, we can find that the

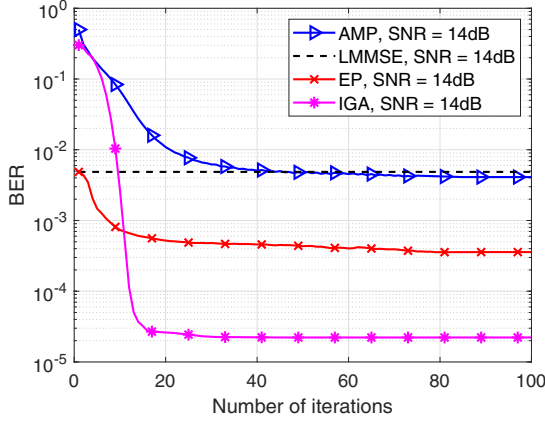


Fig. 5. Convergence performance of IGA compared with EP and AMP at SNR = 14 dB under 16-QAM.

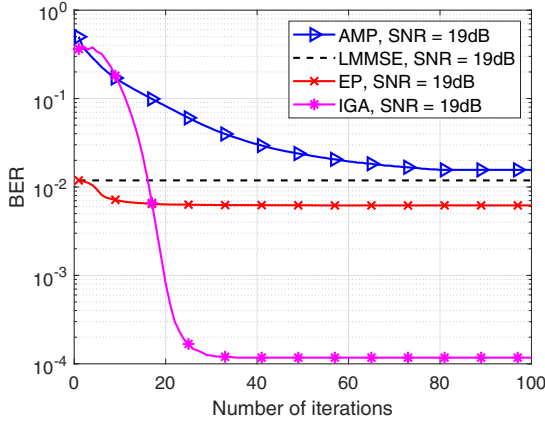


Fig. 6. Convergence performance of IGA compared with EP and AMP at SNR = 19 dB under 64-QAM.

gap between IGA-SD and the other algorithms is increasing. For  $\text{BER} = 10^{-3}$ , the SNR gains of the IGA-SD 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-SD with 40 iterations increases by about 0.2dB each over the two gains above. For 64-QAM, from Fig. 4, we can find that the BER performance of the LMMSE detector exceeds that of the AMP after convergence. The gap between IGA-SD and the other algorithms is still increasing. For  $\text{BER} = 10^{-2}$ , IGA-SD 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-SD with 40 iterations increases by about 0.7dB each over the two gains above.

We then show the convergence performances for the cases with 16-QAM and 64-QAM in Fig. 5 and 6, respectively. From Fig. 5, it can be found that in the case with 16-QAM and SNR = 14dB, the IGA-SD requires around 30 iterations to converge and achieves the best BER performance. Both AMP and EP require about 90 iterations to converge. From Fig. 6, we can find that in the case with 64-QAM and SNR = 19dB, the IGA-SD requires around 30 iterations to converge and achieves the best BER performance. AMP and EP require about 80 and

20 iterations to converge, respectively. We can also find that compared to 16-QAM, AMP and EP require fewer iterations to converge at 64-QAM modulation. This could be attributed to the fact that compared to 16-QAM, the converged BER performances of both AMP and EP have severely degraded in the case with 64-QAM.

## VI. 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. To obtain the approximation of the *a posteriori* marginals, the information geometry theory is introduced. Specifically, 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-SD can obtain the lowest BER performance within a limited number of iterations compared with the existing approaches, which demonstrates the superiority of the proposed IGA-SD for ultra-massive MIMO systems.

## APPENDIX A PROOF OF THEOREM I

We prove this theorem by showing that problem (46) is strictly convex w.r.t.  $\theta_0$ . More precisely, we show that the Hessian of the objective function defined by the K-L divergence (47) is a positive definite matrix. Before proceeding, we first introduce the expectation parameter (EP) and the Fisher information matrices (FIM) of  $p_0(\mathbf{s}; \theta_0)$  and  $p_n(\mathbf{s}; \theta_n)$ ,  $n \in \mathcal{Z}_{2N_r}^+$  in (35b) and (41b), respectively.  $\{p_n(\mathbf{s}; \theta_n)\}_{n=0}^{2N_r}$  can be expressed as

$$p_n(\mathbf{s}; \theta_n) = \exp \{ \theta_n^T \mathbf{t} + m_n(\mathbf{s}) - \psi_n(\theta_n) \}, \quad (75)$$

where  $n \in \mathcal{Z}_{2N_r}$  and  $m_n(\mathbf{s})$  is a function independent of  $\theta_n$ . Specifically, we have  $m_0(\mathbf{s}) \triangleq \mathbf{d}^T \mathbf{t}$  for  $p_0(\mathbf{s}; \theta_0)$ , and  $m_n(\mathbf{s}) = \mathbf{d}^T \mathbf{t} + c_n(\mathbf{s}, y_n)$  for  $p_n(\mathbf{s}; \theta_n)$ ,  $n \in \mathcal{Z}_{2N_r}^+$ . Since the free energy  $\psi_n(\theta_n)$  is constrained by the normalization condition, we have

$$1 = \sum_{\mathbf{s} \in \mathbb{S}^{2K}} \exp \{ \theta_n^T \mathbf{t} + m_n(\mathbf{s}) - \psi_n(\theta_n) \}, n \in \mathcal{Z}_{2N_r}, \quad (76)$$

from which we can obtain

$$\psi_n(\theta_n) = \ln \left( \sum_{\mathbf{s} \in \mathbb{S}^{2K}} \exp \{ \theta_n^T \mathbf{t} + m_n(\mathbf{s}) \} \right), n \in \mathcal{Z}_{2N_r}. \quad (77)$$

Then, from (77), the partial derivative of  $\psi_n(\theta_n)$  is

$$\begin{aligned} & \frac{\partial \psi_n(\theta_n)}{\partial \theta_n} \\ &= \frac{1}{\sum_{\mathbf{s} \in \mathbb{S}^{2K}} \exp \{ \theta_n^T \mathbf{t} + m_n(\mathbf{s}) \}} \sum_{\mathbf{s} \in \mathbb{S}^{2K}} \exp \{ \theta_n^T \mathbf{t} + m_n(\mathbf{s}) \} \mathbf{t} \\ &\stackrel{(a)}{=} \exp \{ -\psi_n(\theta_n) \} \sum_{\mathbf{s} \in \mathbb{S}^{2K}} \exp \{ \theta_n^T \mathbf{t} + m_n(\mathbf{s}) \} \mathbf{t} \end{aligned}$$

$$= \sum_{\mathbf{s} \in \mathbb{S}^{2K}} p_n(\mathbf{s}; \boldsymbol{\theta}_n) \mathbf{t} = \mathbb{E}_{p_n(\mathbf{s}; \boldsymbol{\theta}_n)} \{\mathbf{t}\} = \boldsymbol{\eta}_n(\boldsymbol{\theta}_n), \quad (78)$$

where (a) comes from (76).  $\boldsymbol{\eta}_n(\boldsymbol{\theta}_n) \in \mathbb{R}^{2K(L-1)}$  above is referred as to the EP of  $p_n(\boldsymbol{\theta}_n)$ ,  $n \in \mathcal{Z}_{2N_r}$ . Then, the Hessian of  $\psi_n(\boldsymbol{\theta}_n)$  is

$$\begin{aligned} \mathcal{I}_n(\boldsymbol{\theta}_n) &\triangleq \frac{\partial^2 \psi_n(\boldsymbol{\theta}_n)}{\partial \boldsymbol{\theta}_n \partial \boldsymbol{\theta}_n^T} = \frac{\partial \boldsymbol{\eta}_n(\boldsymbol{\theta}_n)}{\partial \boldsymbol{\theta}_n^T} = \sum_{\mathbf{s} \in \mathbb{S}^{2K}} \mathbf{t} \frac{\partial p_n(\mathbf{s}; \boldsymbol{\theta}_n)}{\partial \boldsymbol{\theta}_n^T} \\ &= \sum_{\mathbf{s} \in \mathbb{S}^{2K}} \mathbf{t} p_n(\mathbf{s}; \boldsymbol{\theta}_n) \left( \mathbf{t}^T - \frac{\partial \psi_n(\boldsymbol{\theta}_n)}{\partial \boldsymbol{\theta}_n^T} \right) \\ &= \sum_{\mathbf{s} \in \mathbb{S}^{2K}} \mathbf{t} p_n(\mathbf{s}; \boldsymbol{\theta}_n) (\mathbf{t} - \boldsymbol{\eta}_n(\boldsymbol{\theta}_n))^T \quad (79) \\ &= \sum_{\mathbf{s} \in \mathbb{S}^{2K}} p_n(\mathbf{s}; \boldsymbol{\theta}_n) (\mathbf{t} - \boldsymbol{\eta}_n(\boldsymbol{\theta}_n)) (\mathbf{t} - \boldsymbol{\eta}_n(\boldsymbol{\theta}_n))^T \\ &= \mathbb{E}_{p_n(\mathbf{s}; \boldsymbol{\theta}_n)} \left\{ (\mathbf{t} - \boldsymbol{\eta}_n(\boldsymbol{\theta}_n)) (\mathbf{t} - \boldsymbol{\eta}_n(\boldsymbol{\theta}_n))^T \right\}, \end{aligned}$$

where  $n \in \mathcal{Z}_{2N_r}$ .  $\mathcal{I}_n(\boldsymbol{\theta}_n) \in \mathbb{R}^{2K(L-1) \times 2K(L-1)}$  above is referred as to the FIM of  $p_n(\mathbf{s}; \boldsymbol{\theta}_n)$ ,  $n \in \mathcal{Z}_{2N_r}$ . From the definition, we can readily show that  $\mathcal{I}_n(\boldsymbol{\theta}_n)$ ,  $n \in \mathcal{Z}_{2N_r}$ , is positive semi-definite. Particularly, the FIM of  $p_0(\mathbf{s}; \boldsymbol{\theta}_0)$  in (35b) is positive definite. The reason is as follows. Since  $\{s_k\}_{k=1}^{2K}$  are independent with each other given the joint probability distribution  $p_0(\mathbf{s}; \boldsymbol{\theta}_0)$ , we can readily obtain  $p_0(\mathbf{s}; \boldsymbol{\theta}_0) = \prod_{k=1}^{2K} p_{0,k}(s_k; \boldsymbol{\theta}_{0,k})$ , where  $p_{0,k}(s_k; \boldsymbol{\theta}_{0,k})$  defined by (35c) is the probability distribution of  $s_k$ ,  $k \in \mathcal{Z}_{2K}^+$ . Then,  $\boldsymbol{\eta}_0(\boldsymbol{\theta}_0) \in \mathbb{R}^{2K(L-1)}$  can be expressed as

$$\boldsymbol{\eta}_0(\boldsymbol{\theta}_0) = [\boldsymbol{\eta}_{0,1}^T(\boldsymbol{\theta}_{0,1}), \boldsymbol{\eta}_{0,2}^T(\boldsymbol{\theta}_{0,2}), \dots, \boldsymbol{\eta}_{0,2K}^T(\boldsymbol{\theta}_{0,2K})]^T, \quad (80)$$

where  $\boldsymbol{\eta}_{0,k}(\boldsymbol{\theta}_{0,k}) \in \mathbb{R}^{(L-1)}$  is given by,

$$\boldsymbol{\eta}_{0,k}(\boldsymbol{\theta}_{0,k}) = \mathbb{E}_{p_{0,k}(\mathbf{s}; \boldsymbol{\theta}_{0,k})} \{\mathbf{t}_k\} = \mathbb{E}_{p_{0,k}(s_k; \boldsymbol{\theta}_{0,k})} \{s_k\}. \quad (81)$$

From the last equation in (81), we refer to  $\boldsymbol{\eta}_{0,k}(\boldsymbol{\theta}_{0,k})$  as the EP of  $p_{0,k}(s_k; \boldsymbol{\theta}_{0,k})$ . Denote the  $\ell$ -th component in  $\boldsymbol{\eta}_{0,k}(\boldsymbol{\theta}_{0,k})$  as  $\eta_{0,k,\ell}(\boldsymbol{\theta}_{0,k})$ ,  $\ell \in \mathcal{Z}_{L-1}^+$ ,  $k \in \mathcal{Z}_{2K}^+$ . From  $\mathbf{t}_k = [t_{k,1}, t_{k,2}, \dots, t_{k,L-1}]^T$ ,  $k \in \mathcal{Z}_{2K}^+$ , and  $t_{k,\ell} = \delta(s_k - s^{(\ell)})$ ,  $\ell \in \mathcal{Z}_{L-1}^+$ , we can obtain

$$\begin{aligned} \eta_{0,k,\ell}(\boldsymbol{\theta}_{0,k}) &= \mathbb{E}_{p_{0,k}(s_k; \boldsymbol{\theta}_{0,k})} \left\{ \delta(s_k - s^{(\ell)}) \right\} \\ &= p_{0,k}(s_k; \boldsymbol{\theta}_{0,k}) \Big|_{s_k=s^{(\ell)}} > 0. \quad (82) \end{aligned}$$

We then define  $(L-1) \times (L-1)$  dimensional covariance matrices  $\mathbf{C}(\mathbf{t}_k, \mathbf{t}_{k'})$  as

$$\begin{aligned} \mathbf{C}(\mathbf{t}_k, \mathbf{t}_{k'}) &\triangleq \\ &\mathbb{E}_{p_0(\mathbf{s}; \boldsymbol{\theta}_0)} \left\{ (\mathbf{t}_k - \boldsymbol{\eta}_{0,k}(\boldsymbol{\theta}_{0,k})) (\mathbf{t}_{k'} - \boldsymbol{\eta}_{0,k'}(\boldsymbol{\theta}_{0,k'}))^T \right\}, \quad (83) \end{aligned}$$

where  $k, k' \in \mathcal{Z}_{2K}^+$ . Since  $\{s_k\}_{k=1}^{2K}$  are independent with each other given  $p_0(\mathbf{s}; \boldsymbol{\theta}_0)$ , we can obtain

$$\begin{aligned} \mathbf{C}(\mathbf{t}_k, \mathbf{t}_{k'}) &\quad (84) \\ &= \begin{cases} \mathbf{R}(\mathbf{t}_k) - \boldsymbol{\eta}_{0,k}(\boldsymbol{\theta}_{0,k}) \boldsymbol{\eta}_{0,k}^T(\boldsymbol{\theta}_{0,k}), & \text{when } k = k', \\ \mathbf{0}, & \text{otherwise,} \end{cases} \end{aligned}$$

where  $\mathbf{0} \in \mathbb{R}^{(L-1) \times (L-1)}$  is the zero matrix,  $\mathbf{R}(\mathbf{t}_k) \in \mathbb{R}^{(L-1) \times (L-1)}$  is given by

$$\mathbf{R}(\mathbf{t}_k) = \mathbb{E}_{p_{0,k}(\mathbf{s}; \boldsymbol{\theta}_{0,k})} \{\mathbf{t}_k \mathbf{t}_k^T\} = \mathbb{E}_{p_{0,k}(s_k; \boldsymbol{\theta}_{0,k})} \{s_k s_k\}. \quad (85)$$

From the definition of  $\mathbf{t}_k$ , the  $(i, j)$ -th element in  $\mathbf{R}(\mathbf{t}_k)$  can be expressed as

$$\begin{aligned} [\mathbf{R}(\mathbf{t}_k)]_{i,j} &= \mathbb{E}_{p_{0,k}(s_k; \boldsymbol{\theta}_{0,k})} \left\{ \delta(s_k - s^{(i)}) \delta(s_k - s^{(j)}) \right\} \\ &= \begin{cases} p_{0,k}(s_k; \boldsymbol{\theta}_{0,k}) \Big|_{s_k=s^{(i)}} = \eta_{0,k,i}(\boldsymbol{\theta}_{0,k}), & \text{when } i = j, \\ 0, & \text{otherwise,} \end{cases} \quad (86) \end{aligned}$$

where  $i, j \in \mathcal{Z}_{L-1}^+$ . Hence, we can obtain  $\mathbf{R}(\mathbf{t}_k) = \text{Diag}\{\boldsymbol{\eta}_{0,k}(\boldsymbol{\theta}_{0,k})\}$ ,  $k \in \mathcal{Z}_{2K}^+$ . From (82), we can readily check that  $\mathbf{R}(\mathbf{t}_k)$  is positive definite. Also, we refer to  $\mathbf{C}(\mathbf{t}_k, \mathbf{t}_k)$  as the FIM of  $p_{0,k}(s_k; \boldsymbol{\theta}_{0,k})$ ,  $k \in \mathcal{Z}_{2K}^+$ , and we denote  $\mathcal{I}_{0,k}(\boldsymbol{\theta}_{0,k}) \triangleq \mathbf{C}(\mathbf{t}_k, \mathbf{t}_k)$ . The FIM of  $p_0(\mathbf{s}; \boldsymbol{\theta}_0)$  can be then expressed as

$$\begin{aligned} \mathcal{I}_0(\boldsymbol{\theta}_0) &= \\ &\begin{bmatrix} \mathcal{I}_{0,1}(\boldsymbol{\theta}_{0,1}) & \mathbf{C}(\mathbf{t}_1, \mathbf{t}_2) & \cdots & \mathbf{C}(\mathbf{t}_1, \mathbf{t}_{2K}) \\ \mathbf{C}(\mathbf{t}_2, \mathbf{t}_1) & \mathcal{I}_{0,2}(\boldsymbol{\theta}_{0,2}) & \cdots & \mathbf{C}(\mathbf{t}_2, \mathbf{t}_{2K}) \\ \vdots & \cdots & \ddots & \vdots \\ \mathbf{C}(\mathbf{t}_{2K}, \mathbf{t}_1) & \cdots & \cdots & \mathcal{I}_{0,2K}(\boldsymbol{\theta}_{0,2K}) \end{bmatrix}. \quad (87) \end{aligned}$$

From (84), we can obtain that the FIM  $\mathcal{I}_0(\boldsymbol{\theta}_0)$  of  $p_0(\mathbf{s}; \boldsymbol{\theta}_0)$  is a block diagonal matrix with the FIMs of  $\{p_{0,k}(s_k; \boldsymbol{\theta}_{0,k})\}_{k=1}^{2K}$  located along its main diagonal, i.e.,

$$\begin{aligned} \mathcal{I}_0(\boldsymbol{\theta}_0) &\quad (88) \\ &= \text{Bdiag}\{\mathcal{I}_{0,1}(\boldsymbol{\theta}_{0,1}), \mathcal{I}_{0,2}(\boldsymbol{\theta}_{0,2}), \dots, \mathcal{I}_{0,2K}(\boldsymbol{\theta}_{0,2K})\}. \end{aligned}$$

We then show that each FIM  $\mathcal{I}_{0,k}(\boldsymbol{\theta}_{0,k})$ ,  $k \in \mathcal{Z}_{2K}^+$ , is positive definite. From the definition, we have

$$\begin{aligned} \mathcal{I}_{0,k}(\boldsymbol{\theta}_{0,k}) &= \mathbf{R}(\mathbf{t}_k) - \boldsymbol{\eta}_{0,k}(\boldsymbol{\theta}_{0,k}) \boldsymbol{\eta}_{0,k}^T(\boldsymbol{\theta}_{0,k}) \\ &= \text{Diag}\{\boldsymbol{\eta}_{0,k}(\boldsymbol{\theta}_{0,k})\} - \boldsymbol{\eta}_{0,k}(\boldsymbol{\theta}_{0,k}) \boldsymbol{\eta}_{0,k}^T(\boldsymbol{\theta}_{0,k}). \quad (89) \end{aligned}$$

Given a non-zero vector  $\mathbf{a} = [a_1, a_2, \dots, a_{L-1}] \in \mathbb{R}^{(L-1)}$ , we have (abbreviate  $\eta_{0,k,\ell}(\boldsymbol{\theta}_{0,k})$  to  $\eta_{0,k,\ell}$  starting from the second equation)

$$\begin{aligned} &\mathbf{a}^T \mathcal{I}_{0,k}(\boldsymbol{\theta}_{0,k}) \mathbf{a} \\ &= \sum_{\ell=1}^{L-1} \eta_{0,k,\ell}(\boldsymbol{\theta}_{0,k}) a_\ell^2 - \left( \sum_{\ell=1}^{L-1} \eta_{0,k,\ell}(\boldsymbol{\theta}_{0,k}) a_\ell \right)^2 \\ &= \sum_{\ell=1}^{L-1} \eta_{0,k,\ell} a_\ell^2 - \left( \sum_{\ell=1}^{L-1} \sqrt{\eta_{0,k,\ell}} \sqrt{\eta_{0,k,\ell} a_\ell^2} \right)^2 \\ &\geq \sum_{\ell=1}^{L-1} \eta_{0,k,\ell} a_\ell^2 - \left( \sum_{\ell=1}^{L-1} \eta_{0,k,\ell} \right) \left( \sum_{\ell=1}^{L-1} \eta_{0,k,\ell} a_\ell^2 \right) \\ &= \left( 1 - \sum_{\ell=1}^{L-1} \eta_{0,k,\ell} \right) \sum_{\ell=1}^{L-1} \eta_{0,k,\ell} a_\ell^2 \\ &\stackrel{(a)}{=} p_{0,k}(s_k; \boldsymbol{\theta}_{0,k}) \Big|_{s_k=s^{(0)}} \times \mathbf{a}^T \mathbf{R}(\mathbf{t}_k) \mathbf{a} \stackrel{(b)}{>} 0, \quad (90) \end{aligned}$$

where (a) comes from that from (82) we have  $\eta_{0,k,\ell}(\boldsymbol{\theta}_{0,k}) = p_{0,k}(s_k; \boldsymbol{\theta}_{0,k}) \Big|_{s_k=s^{(\ell)}}$ ,  $\ell \in \mathcal{Z}_{L-1}^+$ , and  $\sum_{\ell=0}^{L-1} p_{0,k}(s_k; \boldsymbol{\theta}_{0,k}) \Big|_{s_k=s^{(\ell)}} = 1$ , and (b) comes from that  $p_{0,k}(s_k; \boldsymbol{\theta}_{0,k}) \Big|_{s_k=s^{(0)}} > 0$ , and  $\mathbf{R}(\mathbf{t}_k)$  is positive definite. Hence,  $\mathcal{I}_{0,k}(\boldsymbol{\theta}_{0,k})$ ,  $k \in \mathcal{Z}_{2K}^+$ , is positive definite. From (88), we can readily obtain that  $\mathcal{I}_0(\boldsymbol{\theta}_0)$  is also positive definite. Also, we can obtain that  $\psi_0(\boldsymbol{\theta}_0)$  is a strictly convex function of  $\boldsymbol{\theta}_0$ .

We now show that the problem in (46) has a minimum and the solution of it is unique, which satisfies (48). From (47), the partial derivative of  $D_{\text{KL}}\{p(\mathbf{s}) : p_0(\mathbf{s}; \boldsymbol{\theta}_0)\}$  is

$$\begin{aligned} \frac{\partial D_{\text{KL}}}{\partial \boldsymbol{\theta}_0} &= - \sum_{\mathbf{s} \in \mathbb{S}^K} p(\mathbf{s}) \frac{\partial \ln p_0(\mathbf{s}; \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0} \\ &= - \sum_{\mathbf{s} \in \mathbb{S}^K} p(\mathbf{s}) \frac{\partial (\boldsymbol{\theta}_0^T \mathbf{t} - \psi_0(\boldsymbol{\theta}_0))}{\partial \boldsymbol{\theta}_0} \\ &= - \sum_{\mathbf{s} \in \mathbb{S}^K} p(\mathbf{s}) (\mathbf{t} - \boldsymbol{\eta}_0(\boldsymbol{\theta}_0)) = -\boldsymbol{\eta} + \boldsymbol{\eta}_0(\boldsymbol{\theta}_0). \end{aligned} \quad (91)$$

The Hessian of  $D_{\text{KL}}\{p(\mathbf{s}) : p_0(\mathbf{s}; \boldsymbol{\theta}_0)\}$  is

$$\frac{\partial^2 D_{\text{KL}}}{\partial \boldsymbol{\theta}_0 \partial \boldsymbol{\theta}_0^T} = \frac{\partial (-\boldsymbol{\eta} + \boldsymbol{\eta}_0(\boldsymbol{\theta}_0))}{\partial \boldsymbol{\theta}_0^T} = \mathcal{I}_0(\boldsymbol{\theta}_0). \quad (92)$$

Since the FIM  $\mathcal{I}_0(\boldsymbol{\theta}_0)$  is positive definite,  $D_{\text{KL}}\{p(\mathbf{s}) : p_0(\mathbf{s}; \boldsymbol{\theta}_0)\}$  is a strictly convex function of  $\boldsymbol{\theta}_0$  and (46) has a unique solution  $\boldsymbol{\theta}_0^*$ , which satisfies the the first order sufficient condition, i.e.,

$$\boldsymbol{\eta}_0(\boldsymbol{\theta}_0^*) = \boldsymbol{\eta}. \quad (93)$$

This completes the proof.

## APPENDIX B

### PROOF OF COROLLARY I

Denote  $\boldsymbol{\eta}$  as  $\boldsymbol{\eta} \triangleq [\boldsymbol{\eta}_1^T, \boldsymbol{\eta}_2^T, \dots, \boldsymbol{\eta}_{2K}^T]^T \in \mathbb{R}^{2K(L-1)}$ , where  $\boldsymbol{\eta}_k \triangleq \mathbb{E}_{p(\mathbf{s})}\{\mathbf{t}_k\} \in \mathbb{R}^{(L-1)}$ . Denote the  $\ell$ -th element in  $\boldsymbol{\eta}_k$  as  $\eta_{k,\ell}$ , where  $k \in \mathcal{Z}_{2K}^+$  and  $\ell \in \mathcal{Z}_{L-1}^+$ . Then, from  $\mathbf{t}_k = [t_{k,1}, t_{k,2}, \dots, t_{k,L-1}]^T$  and  $t_{k,\ell} = \delta(s_k - s^{(\ell)})$ , we can obtain

$$\begin{aligned} \eta_{k,\ell} &= \mathbb{E}_{p(\mathbf{s})} \left\{ \delta(s_k - s^{(\ell)}) \right\} = \sum_{\mathbf{s} \in \mathbb{S}^{2K}} p(\mathbf{s}) \delta(s_k - s^{(\ell)}) \\ &= \sum_{\mathbf{s} \setminus k \in \mathbb{S}^{2K-1}} p(\mathbf{s}) \Big|_{s_k=s^{(\ell)}} \stackrel{(a)}{=} p_k(s_k) \Big|_{s_k=s^{(\ell)}}, \end{aligned}$$

where  $k \in \mathcal{Z}_{2K}^+$ ,  $\ell \in \mathcal{Z}_{L-1}^+$ , (a) comes from (50), and  $p_k(s_k)$  is the marginal distribution of  $s_k$  given  $p(\mathbf{s})$ . Similar with the process in Appendix A,  $\boldsymbol{\eta}_0(\boldsymbol{\theta}_0^*)$  can be expressed as  $\boldsymbol{\eta}_0(\boldsymbol{\theta}_0^*) \triangleq [\boldsymbol{\eta}_{0,1}^T(\boldsymbol{\theta}_{0,1}^*), \boldsymbol{\eta}_{0,2}^T(\boldsymbol{\theta}_{0,2}^*), \dots, \boldsymbol{\eta}_{0,2K}^T(\boldsymbol{\theta}_{0,2K}^*)]^T \in \mathbb{R}^{2K(L-1)}$ , where  $\boldsymbol{\eta}_{0,k}(\boldsymbol{\theta}_0^*) \triangleq \mathbb{E}_{p_{0,k}(s_k; \boldsymbol{\theta}_{0,k}^*)}\{\mathbf{t}_k\} \in \mathbb{R}^{(L-1)}$ . Denote the  $\ell$ -th element in  $\boldsymbol{\eta}_{0,k}(\boldsymbol{\theta}_0^*)$  as  $\eta_{0,k,\ell}(\boldsymbol{\theta}_0^*)$ . We can obtain

$$\eta_{0,k,\ell}(\boldsymbol{\theta}_0^*) = p_{0,k}(s_k; \boldsymbol{\theta}_{0,k}^*) \Big|_{s_k=s^{(\ell)}}, \quad (94)$$

through a process the same as that of (82), where  $p_{0,k}(s_k; \boldsymbol{\theta}_{0,k}^*)$  is the (marginal) probability distribution of

$s_k$  given the joint probability distribution  $p_0(\mathbf{s}; \boldsymbol{\theta}_0^*)$ . Thus,  $\boldsymbol{\eta} = \boldsymbol{\eta}_0(\boldsymbol{\theta}_0^*)$  is equivalent to

$$p_k(s_k) = p_{0,k}(s_k; \boldsymbol{\theta}_{0,k}^*), \quad s_k \in \mathbb{S}, k \in \mathcal{Z}_{2K}^+. \quad (95)$$

From Theorem 1,  $\boldsymbol{\eta} = \boldsymbol{\eta}_0(\boldsymbol{\theta}_0^*)$  is a necessary and sufficient condition for  $p_0(\mathbf{s}; \boldsymbol{\theta}_0^*)$  being the  $m$ -projection of  $p(\mathbf{s})$  onto  $\mathcal{M}_0$ . Thus, (95) is also a necessary and sufficient condition for  $p_0(\mathbf{s}; \boldsymbol{\theta}_0^*)$  being the  $m$ -projection of  $p(\mathbf{s})$  onto  $\mathcal{M}_0$ . Then, combining (40) and (51), we can immediately obtain (52) This completes the proof.

## APPENDIX C

### PROOF OF THEOREM 2

From the last equation of (58) we have

$$Y_{n,k} = \sum_{k'=1, k' \neq k}^{2K} g_{n,k'} s_{k'} + w'_{n,k} \quad (96)$$

Given  $n$  and  $k$ , let  $\{X_{k'}\}_{k'=1}^{2K}$  be a sequence of random variables, of which each element is defined as

$$\begin{cases} X_{k'} = w'_{n,k}, & k' = k, \\ X_{k'} = g_{n,k'} s_{k'}, & \text{otherwise.} \end{cases} \quad (97)$$

Then,  $Y_{n,k}$  is the sum of the sequence  $\{X_{k'}\}_{k'=1}^{2K}$ . From the probability distribution of  $s_{k'}$  and  $w'_{n,k}$  in (58) (also in (96)) we have

$$\mathbb{E}\{X_{k'}\} = \begin{cases} g_{n,k} s_k, & k' = k, \\ g_{n,k'} \mu_{n,k'}, & \text{otherwise,} \end{cases} \quad (98a)$$

$$\mathbb{V}\{X_{k'}\} = \begin{cases} \sigma_z^2, & k' = k, \\ g_{n,k'}^2 v_{n,k'}, & \text{otherwise.} \end{cases} \quad (98b)$$

Next, we show that when (64) holds, the sequence  $\{X_{k'}\}_{k'=1}^{2K}$  satisfies the Lyapunov's condition (60), where  $\delta = 1$ . When  $k' = k$ ,  $X_{k'}$  is a real Gaussian random variable, and its third central absolute moment is given by [31]

$$\mathbb{E}\{|X_{k'} - \mathbb{E}\{X_{k'}\}|^3\} = 2\sigma_z^3 \sqrt{\frac{2}{\pi}} = 2\sigma_z \sqrt{\frac{2}{\pi}} \mathbb{V}\{X_{k'}\}, \quad (99)$$

which is bounded. When  $k' \neq k$ , we have  $X_{k'} = g_{n,k'} s_{k'}$ . Since  $s_{k'} \in \mathbb{S}$  and  $\mu_{n,k'} = \mathbb{E}\{s_{k'}\}$  are bounded, we can obtain that  $X_{k'}$ ,  $\mathbb{E}\{X_{k'}\}$  and  $(X_{k'} - \mathbb{E}\{X_{k'}\})$  are also bounded when  $g_{n,k'}$  is bounded. Suppose that  $|X_{k'} - \mathbb{E}\{X_{k'}\}| \leq \epsilon$ , we can readily obtain that

$$\mathbb{E}\{|X_{k'} - \mathbb{E}\{X_{k'}\}|^3\} \leq \epsilon \mathbb{V}\{X_{k'}\}. \quad (100)$$

Let  $\varepsilon = \max(\epsilon, 2\sigma_z \sqrt{2/\pi})$ , then for  $k' \in \mathcal{Z}_{2K}^+$  we can obtain

$$\mathbb{E}\{|X_{k'} - \mathbb{E}\{X_{k'}\}|^3\} \leq \varepsilon \mathbb{V}\{X_{k'}\}. \quad (101)$$

Let  $\delta = 1$ , the Lyapunov's condition for  $Y_{n,k} = \sum_{k'=1}^{2K} X_{k'}$  can be expressed as

$$0 \leq \frac{1}{(\mathbb{V}\{Y_{n,k}\})^{\frac{3}{2}}} \sum_{k'=1}^{2K} \mathbb{E}\{|X_{k'} - \mathbb{E}\{X_{k'}\}|^3\}$$

$$\leq \frac{\varepsilon}{(\mathbb{V}\{Y_{n,k}\})^{\frac{3}{2}}} \sum_{k'=1}^{2K} \mathbb{V}\{X_{k'}\} \stackrel{(a)}{=} \frac{\varepsilon}{\sqrt{\mathbb{V}\{Y_{n,k}\}}}, \quad (102)$$

where (a) comes from (63b) and (98b). Meanwhile, from (63b) and (64) we can obtain

$$\lim_{K \rightarrow \infty} \mathbb{V}\{Y_{n,k}\} = \lim_{K \rightarrow \infty} 2\zeta K + \sigma_z^2 \rightarrow \infty. \quad (103)$$

Thus,

$$\lim_{2K \rightarrow \infty} \frac{1}{(\mathbb{V}\{Y_{n,k}\})^{\frac{3}{2}}} \sum_{k'=1}^{2K} \mathbb{E}\{|X_{k'} - \mathbb{E}\{X_{k'}\}|^3\} = 0. \quad (104)$$

Hence,  $\{X_{k'}\}_{k'=1}^{2K}$  satisfies the Lyapunov's condition. This completes the proof.

## REFERENCES

- [1] E. Björnson, L. Van der Perre, S. Buzzi, and E. G. Larsson, "Massive MIMO in sub-6 GHz and mmwave: Physical, practical, and use-case differences," *IEEE Wireless Commun.*, vol. 26, no. 2, pp. 100–108, Apr. 2019.
- [2] H. Jin *et al.*, "Massive MIMO evolution towards 3GPP release 18," *IEEE J. Sel. Areas Commun.*, pp. 1–1, 2023.
- [3] C.-X. Wang *et al.*, "On the road to 6g: Visions, requirements, key technologies and testbeds," *IEEE Commun. Surveys Tuts.*, pp. 1–1, 2023.
- [4] E. Björnson, L. Sanguinetti, H. Wymeersch, J. Hoydis, and T. L. Marzetta, "Massive MIMO is a reality—what is next?: Five promising research directions for antenna arrays," *Digit. Signal Process.*, vol. 94, pp. 3–20, Nov. 2019.
- [5] E. D. Carvalho, A. Ali, A. Amiri, M. Angelichinoski, and R. W. Heath, "Non-stationarities in extra-large-scale massive MIMO," *IEEE Wireless Commun.*, vol. 27, no. 4, pp. 74–80, Aug. 2020.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] Y. Wei, M.-M. Zhao, M. Hong, M.-J. Zhao, and M. Lei, "Learned conjugate gradient descent network for massive MIMO detection," *IEEE Trans. Signal Process.*, vol. 68, pp. 6336–6349, 2020.
- [10] Q. Zhou and X. Ma, "Element-based lattice reduction algorithms for large mimo detection," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 274–286, Feb. 2013.
- [11] T. Takahashi, A. Tölli, S. Ibi, and S. Sampei, "Low-complexity large MIMO detection via layered belief propagation in beam domain," *IEEE Trans. Wireless Commun.*, vol. 21, no. 1, pp. 234–249, Jan. 2022.
- [12] Z. Zhang, H. Li, Y. Dong, X. Wang, and X. Dai, "Decentralized signal detection via expectation propagation algorithm for uplink massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 11 233–11 240, Oct. 2020.
- [13] J. Goldberger and A. Leshem, "MIMO detection for high-order QAM based on a Gaussian tree approximation," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 4973–4982, Aug. 2011.
- [14] C. R. Rao, "Information and the accuracy attainable in the estimation of statistical parameters," in *Breakthroughs in Statistics*. Springer, 1992, pp. 235–247.
- [15] S. Amari and H. Nagaoka, *Methods of Information Geometry*. American Mathematical Soc., 2000, vol. 191.
- [16] N. N. Cencov, *Statistical Decision Rules and Optimal Inference*. American Mathematical Soc., 2000, no. 53.
- [17] S. Amari, *Information Geometry and Its Applications*. Tokyo, Japan: Springer, 2016.
- [18] S. Ikeda, T. Tanaka, and S. Amari, "Stochastic reasoning, free energy, and information geometry," *Neural Computation*, vol. 16, no. 9, pp. 1779–1810, Sep. 2004.
- [19] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann, 1988.
- [20] A. L. Yuille and A. Rangarajan, "The concave-convex procedure," *Neural Computation*, vol. 15, no. 4, pp. 915–936, Apr. 2003.
- [21] 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.
- [22] 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.
- [23] 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.
- [24] J. Sun, Y. Yang, N. N. Xiong, L. Dai, X. Peng, and J. Luo, "Complex network construction of multivariate time series using information geometry," *IEEE Trans. Syst., Man, Cybern.*, vol. 49, no. 1, pp. 107–122, Jan. 2019.
- [25] X. Hua, Y. Ono, L. Peng, Y. Cheng, and H. Wang, "Target detection within nonhomogeneous clutter via total Bregman divergence-based matrix information geometry detectors," *IEEE Trans. Signal Process.*, vol. 69, pp. 4326–4340, Jul. 2021.
- [26] S. Zhang, Y. Wang, Y. Zhang, P. Wan, and J. Zhuang, "A novel clustering algorithm based on information geometry for cooperative spectrum sensing," *IEEE Syst. Jour.*, vol. 15, no. 2, pp. 3121–3130, Jun. 2021.
- [27] H. Pishro-Nik, *Introduction to Probability, Statistics, and Random Processes*. Kappa Research: Galway, Ireland, 2014.
- [28] P. Billingsley, *Probability and Measure*. John Wiley & Sons, New York, 2008.
- [29] 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.
- [30] 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.
- [31] A. Winkelbauer, "Moments and absolute moments of the normal distribution," *CoRR*, vol. abs/1209.4340, 2012. [Online]. Available: <http://arxiv.org/abs/1209.4340>