Information Geometry Approach for Massive MIMO Channel Estimation

Jiyuan Yang*†, An-An Lu*†, Yan Chen*†, Xiqi Gao*†, Xiang-Gen Xia‡, and Dirk Slock§
*National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China
†Purple Mountain Laboratories, Nanjing 211111, China
‡Department of Electrical and Computer Engineering, University of Delaware, Newark, DE 19716 USA
§Department of Communication Systems, EURECOM, 06410 Biot, France
{jyyang, aalu, 213160372, xqgao}@seu.edu.cn, xxia@ee.udel.edu, Dirk.Slock@eurecom.fr

***Abstract*—In this paper, we investigate the channel estimation for massive multi-input multi-output orthogonal frequency division multiplexing (MIMO-OFDM) systems. Using the sampled steering vectors in the space and frequency domains, we first establish a space-frequency (SF) beam based statistical channel model. Based on the channel model, the channel estimation is formulated as obtaining the $a\ posteriori$ information of the beam domain channel. We solve this problem by calculating an approximation of the $a\ posteriori$ distribution's marginals within the information geometry framework. Specifically, by viewing the set of Gaussian distributions and the set of the marginals as a manifold and its submanifold, respectively, we turn the calculation of the marginals into an iterative projection process between submanifolds with different constraints. We derive the information geometry approach (IGA) for channel estimation by calculating the solutions of projections. We prove that the mean of the approximate marginals at the fixed point of IGA is equal to that of the $a\ posteriori$ distribution. Simulations demonstrate that the proposed IGA can accurately estimate the beam domain channel within limited iterations.**

***Index Terms*—Massive MIMO, channel estimation, information geometry**

I. Introduction

Massive multiple-input multiple-output (MIMO) [1]–[3] is known as one of the key techniques of the fifth generation (5G) cellular systems. In a massive MIMO system, the base station (BS) equipped with a large number of antennas can serve tens of users on the same time and frequency resource simultaneously, which provides tremendous capacity gains potentially and increases the energy efficiency significantly. Orthogonal frequency division multiplexing (OFDM) [4] is a multicarrier modulation technique, which can reduce the severe effects of frequency selective fading for wideband wireless communications. Massive MIMO-OFDM plays an essential role in 5G systems and receives increasing attention for the future sixth generation (6G) systems.

In massive MIMO-OFDM systems, channel estimation plays a vital role since the system performance is highly dependent on the quality of the estimated channel. In realistic systems, pilot-aided channel estimation, i.e., the transmitter periodically sends the pilot signals, and the receiver obtains channel state information (CSI) based on the received pilot signals, is the common channel estimation approach [5]. Given the received pilot signals, channel estimation is to obtain the $a\ posteriori$ information of the channel parameters. When the prior distribution of channel parameters is Gaussian, the $a\ posteriori$ distribution of them is also Gaussian, of which the $a\ posteriori$ information is given by the mean and covariance matrix. Nevertheless, calculating the $a\ posteriori$ mean and covariance is challenging due to the large dimension of the channel in massive MIMO-OFDM systems. The calculation of the conventional estimators, such as MMSE estimator, is not affordable since a large dimension matrix inverse is usually required.

The space defined by the parameters of the $a\ posteriori$ PDF can be regarded as a differentiable manifold with a Riemannian structure. Hence, the definitions and tools of differential geometry can be well applied. This is exactly one of the subjects of information geometry [6]–[8]. Thus, it is appropriate to apply information geometry into the channel estimation. The main idea of information geometry is to investigate the intrinsic geometrical structures of the specific sets of PDFs by regarding the parametric space of them as differentiable manifolds. It has been recently applied in multisensor estimation fusion [9], false alarm rate detection [10] and generalized Bayesian prediction [11].

In this paper, we propose an information geometry approach for channel estimation for massive MIMO-OFDM systems. We first derive a space-frequency (SF) beam based statistical channel model by using the sampled steering vectors in the space and frequency domains. The channel estimation is then formulated as obtaining the $a\ posteriori$ information of the beam domain channel. We solve this problem by calculating an approximation of the $a\ posteriori$ distribution's marginals within the information geometry framework. By viewing the set of Gaussian distributions and the set of the marginals as a manifold and its submanifold, respectively, we turn the calculation of the marginals into an iterative projection process between submanifolds with different constraints. By calculating the solution of m-projections, we derive the information geometry approach (IGA) for the channel estimation. At last, it is shown that the mean of the approximate $a\ posteriori$

This work was supported by the National Key R&D Program of China under Grants 2018YFB1801103, the Jiangsu Province Basic Research Project under Grant BK20192002, the Fundamental Research Funds for the Central Universities, and the National Natural Science Foundation of China under Grant 61801113.

marginals at the fixed point of IGA is equal to that of the *a posteriori* distribution.

II. Channel Model and Problem Formulation

In this section, we derive the SF beam based statistical channel model for massive MIMO-OFDM systems by using sampled steering vectors in the space and frequency domains. Then, the problem of the channel estimation is formulated.

A. System Model

Consider a typical massive MIMO-OFDM system working in time division duplexing (TDD) mode, where the base station (BS) equipped with UPA of $N_r = N_{r,v} \times N_{r,h}$ antennas communicates with K single-antenna users within a cell, and $N_{r,v}$ and $N_{r,h}$ are the numbers of the antennas at each vertical column and horizontal row, respectively. In TDD mode, due to the channel reciprocity, channel state information (CSI) can be obtained from uplink (UL) training and then used for UL signal detection and DL precoding. Hence, we focus on the uplink channel estimation. The standard OFDM modulation with N_c subcarriers is applied. The system sampling interval and the length of the cyclic prefix (CP) are denoted as T_s and N_{g}, respectively. The set of the training subcarriers is denoted as $\mathcal{N}^t = \{N_1, N_1+1, \ldots, N_2\}$. Then, the number of training subcarriers is $N_p \triangleq |\mathcal{N}^t|$, and $N_p \le N_c$, where $|\mathcal{N}^t|$ denotes the cardinality of set $\mathcal{N}^t$. We assume the channel is quasi-static. Thus, we only consider the UL training within one OFDM symbol. Then, during each symbol, the received signal on the n-th subcarrier at the BS, denoted as $\mathbf{y}_n \in \mathbb{C}^{N_r \times 1}$, can be expressed as [12]

$$\mathbf{y}_n = \sum_{k=1}^{K} \mathbf{g}_{n,k} x_k[n] + \mathbf{z}_n, n \in \mathcal{N}^t, \tag{1}$$

where $\mathbf{g}_{n,k} \in \mathbb{C}^{N_r \times 1}$ is the space-domain channel on the n-th subcarrier between the user k and BS, $x_k[n]$ is the training sequence transmitted by user k and $\mathbf{z}_n \in \mathbb{C}^{N_r \times 1}$ is the circularly symmetric Gaussian noise with $\mathbb{E}\left\{\mathbf{z}_n \mathbf{z}_n^H\right\} = \sigma_z^2 \mathbf{I}$. Denote the channel of user k over all training subcarriers as

$$\mathbf{G}_k = [\mathbf{g}_{N_1,k} \ \cdots \ \mathbf{g}_{N_2,k}] \in \mathbb{C}^{N_r \times N_p}, \tag{2}$$

which is referred as the SF domain channel coefficient matrix of user k. Then, let $\mathbf{Y} = [\mathbf{y}_{N_1} \ \cdots \ \mathbf{y}_{N_2}] \in \mathbb{C}^{N_r \times N_p}$, $\mathbf{X}_k = \mathrm{diag}(\mathbf{x}_k) \in \mathbb{C}^{N_p \times N_p}$ with $\mathbf{x}_k = [x_k[N_1], \ldots, x_k[N_2]]^T$ and $\mathbf{Z} = [\mathbf{z}_{N_1} \ \cdots \ \mathbf{z}_{N_2}] \in \mathbb{C}^{N_r \times N_p}$, and we can obtain

$$\mathbf{Y} = \sum_{k=1}^{K} \mathbf{G}_k \mathbf{X}_k + \mathbf{Z}. \tag{3}$$

B. SF Beam Based Statistical Channel Model

Assumed that the antennas of UPA on the BS are separated by one-half wavelength. Then, the space steering vector $\mathbf{v}(u,v) \in \mathbb{C}^{N_r \times 1}$ can be expressed as $\mathbf{v}(u,v) \triangleq \mathbf{v}_v(u) \otimes \mathbf{v}_h(v)$ [13], where

$$\mathbf{v}_v(u) \triangleq [p(1) \ \ p(2) \ \cdots p(N_{r,v})]^T \in \mathbb{C}^{N_{r,v} \times 1}, \tag{4a}$$

$$\mathbf{v}_h(v) = [q(1) \ \ q(2) \ \cdots \ q(N_{r,h})]^T \in \mathbb{C}^{N_{r,h} \times 1}, \tag{4b}$$

$p(n) = \exp\{-\jmath\pi(n-1)u\}$ and $q(n) = \exp\{-\jmath\pi(n-1)v\}$. u and v above are the directional cosines, i.e., $u = \sin\theta$, $v = \cos\theta\sin\phi$, where $\theta, \phi \in [-\pi/2, \pi/2]$ are the vertical and the horizontal angles of arrival at BS, respectively. Then, the space-domain channel $\mathbf{g}_{n,k}$ in (1) can be expressed as [12], [14]

$$\mathbf{g}_{n,k} = \sum_{p=1}^{P_k} \alpha_{p,k} \exp\{-\jmath 2\pi \tau_{p,k}(f_c + n\Delta_f)\} \mathbf{v}(u_{p,k}, v_{p,k}), \tag{5}$$

where P_k is the number of paths between user k and BS, $\alpha_{p,k}$ is the complex-valued channel gain of the p-th path of user k, $\tau_{p,k}$ is the delay of path p between user k and the first antenna of UPA, f_c is the carrier frequency, $\Delta_f = \frac{1}{N_c T_s}$ is the subcarrier interval, and $u_{p,k}$ and $v_{p,k}$ are the directional cosines of path p between user k and BS. We define the frequency steering vector as

$$\mathbf{u}(\tau) \triangleq [r(N_1) \ \cdots \ r(N_2)]^T \in \mathbb{C}^{N_p \times 1}, \tag{6}$$

where $r(n) = \exp\{-\jmath 2\pi \Delta_f n \tau\}$. Then, the SF domain channel coefficient matrix $\mathbf{G}_k$ (2) can be expressed as

$$\mathbf{G}_k = \sum_{p=1}^{P_k} h_{p,k} \mathbf{v}(u_{p,k}, v_{p,k}) \mathbf{u}^T(\tau_{p,k}), \tag{7}$$

where $h_{p,k} = \alpha_{p,k}\exp(-\jmath 2\pi f_c \tau_{p,k})$. We derive the SF beam based statistical channel model of the massive MIMO systems by discretizing the directional cosines u, v and the delay τ. Define $h_k(u,v,\tau) \triangleq \sum_{p=1}^{P_k} h_{p,k}\delta(u-u_{p,k})\delta(v-v_{p,k})\delta(\tau-\tau_{p,k})$, where the parameters $u_{p,k}$, $v_{p,k}$ and $\tau_{p,k}$ are the same as before. We then define the sampled directional cosines and delays as

$$u_i = \frac{2(i-1)-N_v}{N_v}, i \in \mathcal{Z}_{N_v}^+ \tag{8a}$$

$$v_j = \frac{2(j-1)-N_h}{N_h}, j \in \mathcal{Z}_{N_h}^+ \tag{8b}$$

$$\tau_\ell = \frac{(\ell-1)N_f}{N_\tau N_p \Delta_f}, \ell \in \mathcal{Z}_{N_\tau}^+, \tag{8c}$$

respectively, where $N_f = \lceil N_p N_{\mathrm{g}}/N_{\mathrm{c}} \rceil$ and $\mathcal{Z}_N^+ = \{1, 2, \ldots, N\}$. $N_v \triangleq F_v N_{r,v}$, $N_h \triangleq F_h N_{r,h}$ and $N_\tau \triangleq F_\tau N_f$, where F_v, F_h ,and F_τ are the fine factors (FFs). When N_v,N_h and N_τ are sufficiently large, the SF domain channel matrix (7) can be well approximated as

$$\mathbf{G}_k \approx \sum_{i,j,\ell} \tilde{h}_k(u_i, v_j, \tau_\ell) \mathbf{v}(u_i, v_j) \mathbf{u}^T(\tau_\ell), \tag{9}$$

where

$$\tilde{h}_k(u_i, v_j, \tau_\ell) = \iiint\limits_{\substack{u \in \mathcal{B}_{u,i}, v \in \\ \mathcal{B}_{v,j}, \tau \in \mathcal{B}_{\tau,\ell}}} h_k(u,v,\tau)\, \mathrm{d}u \mathrm{d}v \mathrm{d}\tau, \tag{10}$$

$\mathcal{B}_{u,i} = [u_i, u_{i+1}), i \in \mathcal{Z}_{N_v}^+$, $\mathcal{B}_{v,j} = [v_j, v_{j+1}), j \in \mathcal{Z}_{N_h}^+$ and $\mathcal{B}_{\tau,\ell} = [\tau_\ell, \tau_{\ell+1}), \ell \in \mathcal{Z}_{N_\tau}^+$. Let $\mathbf{V} \in \mathbb{C}^{N_r \times N_v N_h} = \mathbf{V}_v \otimes \mathbf{V}_h$ and $\mathbf{F} = [\mathbf{u}(\tau_1) \ \ldots \ \mathbf{u}(\tau_{N_\tau})] \in \mathbb{C}^{N_p \times N_\tau}$, where

$\mathbf{V}_v = [\mathbf{v}_v(u_1) \cdots \mathbf{v}_v(u_{N_v})] \in \mathbb{C}^{N_{r,v} \times N_v}$ and $\mathbf{V}_h = [\mathbf{v}_h(v_1) \cdots \mathbf{v}_h(v_{N_h})] \in \mathbb{C}^{N_{r,h} \times N_h}$. Then, the SF domain channel matrix $\mathbf{G}_k$ can be expressed as,

$$\mathbf{G}_k = \mathbf{V}\mathbf{H}_k\mathbf{F}^T, \tag{11}$$

where $\mathbf{H}_k \in \mathbb{C}^{N_v N_h \times N_\tau}$ and $[\mathbf{H}_k]_{m,n} = \tilde{h}_k(u_i, v_j, \tau_n)$ with $i = \lfloor \frac{m-1}{N_h} + 1 \rfloor$ and $j = m - (i-1)N_h$. With the assumption of wide-sense stationary uncorrelated scattering Rayleigh fading channel, the elements in $\mathbf{H}_k$ follow the independent complex Gaussian distributions with zero mean and possibly different variances, and we define the beam domain channel power matrix as:

$$\mathbf{\Omega}_k = \mathbb{E}\left\{\mathbf{H}_k \odot \overline{\mathbf{H}}_k\right\}. \tag{12}$$

Due to the channel sparsity, most of the elements in $\mathbf{\Omega}_k$ are close to zero [12]. Therefore, there are sufficient resources to acquire $\mathbf{\Omega}_k, \forall k$. For instance, the authors of [13] propose a method that can obtain the estimate of $\mathbf{\Omega}_k$. Thus, we assume that $\mathbf{\Omega}_k$ of all users are known at the BS in the rest of the paper.

C. Problem Formulation

During the UL training, the task of channel estimation is to obtain the *a posteriori* information of the SF domain channel matrix $\mathbf{G}_k, \forall k$. From (11), the *a posteriori* information of $\mathbf{G}_k$ can be calculated from that of the beam domain channel matrix $\mathbf{H}_k$. Thus, we focus on the estimation of $\mathbf{H}_k, \forall k$. By substituting (11) into the UL received signal model (3), we have

$$\mathbf{Y} = \mathbf{V}\mathbf{H}_a\mathbf{M} + \mathbf{Z}, \tag{13}$$

where $\mathbf{H}_a = [\mathbf{H}_1 \ \mathbf{H}_2 \ \cdots \ \mathbf{H}_K] \in \mathbb{C}^{N_v N_h \times K N_\tau}$ and $\mathbf{M} = [\mathbf{X}_1\mathbf{F} \ \mathbf{X}_2\mathbf{F} \ \cdots \ \mathbf{X}_K\mathbf{F}]^T \in \mathbb{C}^{K N_\tau \times N_p}$. After the vectorization of (13) and removing the elements in $\mathrm{vec}(\mathbf{H}_a)$ with zero variance, denoted as $\mathbf{h}$, and the corresponding columns in $\mathbf{M}^T \otimes \mathbf{V}$, we have,

$$\mathbf{y} = \mathbf{A}\mathbf{h} + \mathbf{z}, \tag{14}$$

where $\mathbf{A} \in \mathbb{C}^{N \times M}$ is a deterministic matrix extracted from $\mathbf{M}^T \otimes \mathbf{V}$, $N = N_r N_p$, M is the number of elements in $\mathbf{H}_a$ with non-zero variance, $\mathbf{y}$ and $\mathbf{z}$ are the vectorizations of $\mathbf{Y}$ and $\mathbf{Z}$, respectively, $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}, \mathbf{D})$ with positive definite and diagonal $\mathbf{D}$, and $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}, \sigma_z^2\mathbf{I})$. $\mathbf{h}$ and $\mathbf{z}$ are assumed to be independent with each other. Then, the *a posteriori* distribution is also Gaussian, i.e., $p(\mathbf{h}|\mathbf{y}) = p_G\left(\mathbf{h}; \tilde{\boldsymbol{\mu}}, \tilde{\mathbf{\Sigma}}\right)$, where $p_G(\mathbf{h}; \boldsymbol{\mu}, \mathbf{\Sigma})$ denotes the PDF of a complex Gaussian distribution $\mathcal{CN}(\boldsymbol{\mu}, \mathbf{\Sigma})$. The *a posteriori* mean $\tilde{\boldsymbol{\mu}}$ and covariance $\tilde{\mathbf{\Sigma}}$ are given by [15]

$$\tilde{\boldsymbol{\mu}} = \mathbf{D}\left(\mathbf{A}^H\mathbf{A}\mathbf{D} + \sigma_z^2\mathbf{I}\right)^{-1}\mathbf{A}^H\mathbf{y}, \tag{15a}$$

$$\tilde{\mathbf{\Sigma}} = \left(\mathbf{D}^{-1} + \frac{1}{\sigma_z^2}\mathbf{A}^H\mathbf{A}\right)^{-1}. \tag{15b}$$

The computational complexity of the *a posteriori* information in (15) is $\mathcal{O}\left(M^3 + M^2N\right)$. Note that when the number of users is relatively large, e.g., tens, M can be comparable to N even though channel sparsity exists. In this case, it may be unaffordable to apply (15) in practice when both N and M are large.

III. Information Geometry Approach for Channel Estimation

This section applies information geometry into the beam domain channel estimation. We propose the information geometry approach (IGA) for channel estimation, and show that at the fixed point of IGA, the approximate *a posteriori* mean obtained by IGA is equal to the *a posteriori* mean (15a).

Recalling the received signal model (14), with prior distributions of $\mathbf{h}$ and $\mathbf{z}$ being independent Gaussian distributions, i.e., $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}, \mathbf{D})$ with positive definite and diagonal $\mathbf{D}$, and $\mathbf{z} \sim \mathcal{CN}\left(\mathbf{0}, \sigma_z^2\mathbf{I}\right)$, the *a posteriori* distribution is also Gaussian with its PDF

$$\begin{aligned} p(\mathbf{h}|\mathbf{y}) &= C\prod_{i=1}^{M} p_i(h_i)\prod_{n=1}^{N} p_n(y_n|\mathbf{h}) \\ &= C'\exp\left\{-\mathbf{h}^H\mathbf{D}^{-1}\mathbf{h}\right\}\prod_{n=1}^{N}\exp\left\{-\frac{\left|y_n - \boldsymbol{\gamma}_n^H\mathbf{h}\right|^2}{\sigma_z^2}\right\} \\ &= \exp\left\{\mathbf{d}_h \circ \mathbf{t}_h + \sum_{n=1}^{N} c_n(\mathbf{h}) - \psi_q\right\}, \end{aligned} \tag{16}$$

where y_n is the n-th element of $\mathbf{y}$, $\mathbf{d}_h = \mathbf{f}\left(\mathbf{0}, -\mathbf{D}^{-1}\right)$, $\mathbf{t}_h = \mathbf{f}\left(\mathbf{h}, \mathbf{I} \odot \left(\mathbf{h}\mathbf{h}^H\right)\right)$, C, C' and ψ_q are the normalization factors and $c_n(\mathbf{h})$ is given by

$$c_n(\mathbf{h}) = -\mathbf{h}^H\frac{\boldsymbol{\gamma}_n\boldsymbol{\gamma}_n^H}{\sigma_z^2}\mathbf{h} + \mathbf{h}^H\frac{\boldsymbol{\gamma}_n y_n}{\sigma_z^2} + \frac{y_n^H\boldsymbol{\gamma}_n^H}{\sigma_z^2}\mathbf{h}, \tag{17a}$$

$$\boldsymbol{\gamma}_n = \left[\mathbf{A}^H\right]_{:,n} = \left[\overline{a}_{n1} \ \cdots \ \overline{a}_{nM}\right]^T \in \mathbb{C}^{M \times 1}. \tag{17b}$$

$\mathbf{f}(\mathbf{a}, \mathbf{A})$ above is defined as $\mathbf{f}(\mathbf{a}, \mathbf{A}) = \left[\mathbf{a}^T, \mathrm{vec}^T(\mathbf{A})\right]^T \in \mathbb{C}^{\left(P+Q^2\right)\times 1}$ with $\mathbf{a} \in \mathbb{C}^{P\times 1}$ and $\mathbf{A} \in \mathbb{C}^{Q \times Q}$. It can be shown that $\mathbf{f}(\mathbf{a}_1, \mathbf{A}_1) \circ \mathbf{f}(\mathbf{a}_2, \mathbf{A}_2) = \mathbf{a}_1 \circ \mathbf{a}_2 + \mathbf{A}_1 \circ \mathbf{A}_2$. Note that $\mathbf{t}_h$ above only contains the statistics of single random variables, i.e., h_i as well as $|h_i|^2$, and no interactions terms of $h_i, i = 1, 2, \ldots, M$. All the interactions, i.e., $w_{i,j}h_i\overline{h}_j$ are in the $\sum_{n=1}^{M} c_n(\mathbf{h})$. If we can somehow approximate $\sum c_n$ as $\boldsymbol{\vartheta}_0 \circ \mathbf{t}_h$, where $\boldsymbol{\vartheta}_0 = \mathbf{f}(\boldsymbol{\theta}_0, \mathbf{\Theta}_0)$ with $\boldsymbol{\theta}_0 \in \mathbb{C}^{M\times 1}$ and $\mathbf{\Theta}_0 \in \mathbb{D}^M$, and $\mathbb{D}^M$ is the set of $M \times M$ real diagonal matrices, then we can obtain

$$p(\mathbf{h}|\mathbf{y}) \approx p_0\left((\mathbf{d}_h + \boldsymbol{\vartheta}_0) \circ \mathbf{t}_h - \psi_0\right), \tag{18}$$

where ψ_0 is the normalization factor. The marginals of p_0 can be obtained easily since it contains no interactions. Based on the observation above, we define two sets of Gaussian distributions, which are called the objective manifold M_0 (OBM) and the auxiliary manifolds M_n (AMs), respectively,

$$M_0 = \{p_0(\mathbf{h}; \boldsymbol{\vartheta}_0)\}, \tag{19a}$$

$$p_0(\mathbf{h}; \boldsymbol{\vartheta}_0) = \exp\left\{\mathbf{d}_h \circ \mathbf{t}_h + \boldsymbol{\vartheta}_0 \circ \mathbf{t}_h - \psi_0(\boldsymbol{\vartheta}_0)\right\}, \tag{19b}$$

$$M_n = \{p_n(\mathbf{h};\boldsymbol{\vartheta}_n)\}, n \in \mathcal{Z}_N^+, \tag{20a}$$

$$p_n(\mathbf{h};\boldsymbol{\vartheta}_n) = \exp\{(\mathbf{d}_h + \boldsymbol{\vartheta}_n)\circ \mathbf{t}_h + c_n(\mathbf{h}) - \psi_n(\boldsymbol{\vartheta}_n)\}, \tag{20b}$$

where $\boldsymbol{\vartheta}_n = \mathbf{f}(\boldsymbol{\theta}_n, \boldsymbol{\Theta}_n)$ is the parameter of p_n, $\boldsymbol{\theta}_n \in \mathbb{C}^{M\times 1}$, $\boldsymbol{\Theta}_n \in \mathbb{D}^M$ is a diagonal parameter matrix of p_n, $\psi_n(\boldsymbol{\vartheta}_n)$ is the free energy, the subscript $n \in \mathcal{Z}_N$, $\mathcal{Z}_N \triangleq \{0, 1, \ldots, N\}$, and $\mathbf{d}_h$, $\mathbf{t}_h$ as well as $c_n(\mathbf{h})$ are the same as before. Observing $p_n(\mathbf{h};\boldsymbol{\vartheta}_n)$ in (20), we can find that only one interaction item $c_n(\mathbf{h})$ is maintained, and all others, i.e., $\sum_{n'\neq n} c_{n'}(\mathbf{h})$ are replaced with $\boldsymbol{\vartheta}_n \circ \mathbf{t}_h$. Assume that the parameter $\boldsymbol{\vartheta}_n$ of $p_n, n \in \mathcal{Z}_N^+$, is given, then, the approximation of $c_n(\mathbf{h})$ can be calculated through the m-projection of p_n on M_0. To be specific, the m-projection of p_n on M_0 is calculated by minimizing the following Kullback-Leibler (K-L) divergence,

$$\boldsymbol{\vartheta}_{0n} = \arg\min_{\boldsymbol{\vartheta}_0} D_{KL}\{p_n(\mathbf{x};\boldsymbol{\vartheta}_n) : p_0(\mathbf{x};\boldsymbol{\vartheta}_0)\}, \tag{21}$$

where

$$D_{KL}\{p_n(\mathbf{x};\boldsymbol{\vartheta}_n) : p_0(\mathbf{x};\boldsymbol{\vartheta}_0)\} = \mathbb{E}_{p_n}\left\{\ln\frac{p_n(\mathbf{x};\boldsymbol{\vartheta}_n)}{p_0(\mathbf{x};\boldsymbol{\vartheta}_0)}\right\}. \tag{22}$$

After some calculations, $\boldsymbol{\vartheta}_{0n} = \mathbf{f}(\boldsymbol{\theta}_{0n}, \boldsymbol{\Theta}_{0n})$ is given by (23). To calculate the approximation of $c_n(\mathbf{h})$, we express the m-projection $p_0(\mathbf{h};\boldsymbol{\vartheta}_{0n})$ as

$$\begin{aligned} p_0(\mathbf{h};\boldsymbol{\vartheta}_{0n}) &= \exp\{(\mathbf{d}_h + \boldsymbol{\vartheta}_{0n})\circ \mathbf{t}_h - \psi_0(\boldsymbol{\vartheta}_{0n})\} \\ &= \exp\{(\mathbf{d}_h + \boldsymbol{\vartheta}_n + \boldsymbol{\xi}_n)\circ \mathbf{t}_h - \psi_0\}, \end{aligned} \tag{24}$$

where $\boldsymbol{\vartheta}_{0n}$ is regarded as the sum of the parameter of p_n and the parameter of the approximation of $c_n(\mathbf{h})$ since the difference between $p_n(\mathbf{h};\boldsymbol{\vartheta}_n)$ and $p_0(\mathbf{h};\boldsymbol{\vartheta}_{0n})$ is that $c_n(\mathbf{h})$ in $p_n(\mathbf{h};\boldsymbol{\vartheta}_n)$ is replaced by $\boldsymbol{\xi}_n \circ \mathbf{t}_h$ in $p_0(\mathbf{h};\boldsymbol{\vartheta}_{0n})$. Thus, we approximate $c_n(\mathbf{h})$ as $\boldsymbol{\xi}_n \circ \mathbf{t}_h$, and the $\boldsymbol{\xi}_n$ is given by

$$\boldsymbol{\xi}_n = \boldsymbol{\vartheta}_{0n} - \boldsymbol{\vartheta}_n, n \in \mathcal{Z}_N^+. \tag{25}$$

Then, $\boldsymbol{\vartheta}_0$ is calculated as $\boldsymbol{\vartheta}_0 = \sum_{n=1}^N \boldsymbol{\xi}_n$ since our ultimate goal is to approximate $\sum_{n=1}^N c_n$ as $\boldsymbol{\vartheta}_0 \circ \mathbf{t}_h$ and each c_n is approximated as $\boldsymbol{\xi}_n \circ \mathbf{t}_h$. It should be also noted that the complete algorithm is an iterative procedure. Specifically, we first initialize $\boldsymbol{\vartheta}_n$ with any value. Then, calculate the m-projection $\boldsymbol{\vartheta}_{0n}$ and the approximation item $\boldsymbol{\xi}_n$. The parameter of p_n in M_n, $n \in \mathcal{Z}_N^+$, is then updated as $\boldsymbol{\vartheta}_n = \sum_{n'\neq n} \boldsymbol{\xi}_{n'}$ since $\boldsymbol{\vartheta}_n \circ \mathbf{t}_h$ replaces $\sum_{n'\neq n} c_{n'}(\mathbf{x}_h)$ in p_n and each interaction item $c_n(\mathbf{x}_h)$ is approximated as $\boldsymbol{\xi}_n \circ \mathbf{t}_h$. Then, the parameter of $p_0(\mathbf{h};\boldsymbol{\vartheta}_0)$ in M_0 is updated as $\boldsymbol{\vartheta}_0 = \sum_{n=1}^N \boldsymbol{\xi}_n$. Repeat the m-projections, calculate the approximation items and the updates until convergence. In practice, to improve the convergence of IGA, we can also use the following damped updating:

$$\boldsymbol{\vartheta}_n^{t+1} = \alpha \sum_{n'\neq n} \boldsymbol{\xi}_{n'}^t + (1-\alpha)\boldsymbol{\vartheta}_n^t, n \in \mathcal{Z}_N^+, \tag{26a}$$

$$\boldsymbol{\vartheta}_0^{t+1} = \alpha \sum_{n=1}^N \boldsymbol{\xi}_n^t + (1-\alpha)\boldsymbol{\vartheta}_0^t, \tag{26b}$$

where $0 < \alpha \leq 1$ is the damping. We summarize the IGA in Algorithm 1. The computational complexity of the IGA is $\mathcal{O}(TNM)$, where T is the number of the iterations, $N = N_r N_p$ is the product of the number of antennas at the BS and the number of transmitting subcarriers, and M is the number of variables to be estimated for the beam domain channel. Then, at the fixed point of IGA, we have the following theorem.

Theorem 1. *When converged, the mean of $p_0(\mathbf{h};\boldsymbol{\vartheta}_0^*)$ obtained by IGA is equal to that of the a posteriori distribution $p(\mathbf{h}|\mathbf{y})$.*

The proof is omitted due to space limitations. Note that the above relationship holds for arbitrary matrix $\mathbf{A}$. For the most popular AMP algorithms, the same conclusion can be obtained when $\mathbf{A}$ is a zero-mean i.i.d. sub-Gaussian matrix satisfying the large-system limit, and the fixed point (equilibrium) of AMP is unique [16].

Algorithm 1: IGA for Channel Estimation

Input: The covariance $\mathbf{D}$ of the priori distribution $p(\mathbf{h})$, the received signal $\mathbf{y}$, the noise power σ_z^2 and the maximal iteration number $t_{\max}$.
Initialization: set $t = 0$, set α, initialize the parameters of $p_n(\mathbf{h};\boldsymbol{\vartheta}_n(t)), n = 0, 1, \ldots, N$;
repeat
  Calculate the m-projection as (23);
  Update the parameters of AMs and OBM as (26a) and (26b), respectively;
  $t = t + 1$;
until Convergence or $t > t_{\max}$;
Output: The mean and variance of the approximate marginal, $p(h_i|\mathbf{y}), i = 1, 2, \ldots, M$, are given by the i-th component of $\boldsymbol{\mu}_0$ and $\mathrm{diag}(\boldsymbol{\Sigma}_0)$, respectively, where $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ of $p_0(\mathbf{h};\boldsymbol{\vartheta}_0(t))$ are given by (27)

IV. Simulation Results

This section provides some simulation results to illustrate the performance of the proposed information geometry approach for massive MIMO-OFDM channel estimation. We adopt the widely used QuaDRiGa channel model [17]. The main parameters for the simulations are summarized in Table I. The simulation scenario is set to "3GPP_38.901_UMa_NLOS". The layout of the massive MIMO-OFDM system is plotted in Fig. 1. The BS is located at

TABLE I
Parameter Settings of the QuaDRiGa

Parameter	Value
Number of BS antenna $N_{r,v} \times N_{r,h}$	8×16
UT number K	48
Center frequency f_c	4.8GHz
Number of training subcarriers N_p	360
Subcarrier spacing Δ_f	15kHz
Number of subcarriers N_c	2048
CP length N_{g}	144
Fine factors F_h, F_v, F_τ	1, 2 or 4

$$\boldsymbol{\theta}_{0n} = \left(\mathbf{I} - \frac{\left(\mathbf{D}^{-1} - \boldsymbol{\Theta}_n\right)^{-1} \mathbf{I} \odot \left(\boldsymbol{\gamma}_n \boldsymbol{\gamma}_n^H\right)}{\sigma_z^2 + \boldsymbol{\gamma}_n^H \left(\mathbf{D}^{-1} - \boldsymbol{\Theta}_n\right)^{-1} \boldsymbol{\gamma}_n}\right)^{-1} \left(\frac{2y_n - \boldsymbol{\gamma}_n^H \left(\mathbf{D}^{-1} - \boldsymbol{\Theta}_n\right)^{-1} \boldsymbol{\theta}_n}{\sigma_z^2 + \boldsymbol{\gamma}_n^H \left(\mathbf{D}^{-1} - \boldsymbol{\Theta}_n\right)^{-1} \boldsymbol{\gamma}_n} \boldsymbol{\gamma}_n + \boldsymbol{\theta}_n\right) \tag{23a}$$

$$\boldsymbol{\Theta}_{0n} = \mathbf{D}^{-1} - \left[\left(\mathbf{D}^{-1} - \boldsymbol{\Theta}_n\right)^{-1} - \frac{\left(\left(\mathbf{D}^{-1} - \boldsymbol{\Theta}_n\right)^{-1}\right)^2 \mathbf{I} \odot \left(\boldsymbol{\gamma}_n \boldsymbol{\gamma}_n^H\right)}{\sigma_z^2 + \boldsymbol{\gamma}_n^H \left(\mathbf{D}^{-1} - \boldsymbol{\Theta}_n\right)^{-1} \boldsymbol{\gamma}_n}\right]^{-1}, n \in \mathcal{Z}_N^+ \tag{23b}$$

$$\boldsymbol{\mu}_0 = \frac{1}{2} \left(\mathbf{D}^{-1} - \boldsymbol{\Theta}_0\right)^{-1} \boldsymbol{\theta}_0, \ \boldsymbol{\Sigma}_0 = \left(\mathbf{D}^{-1} - \boldsymbol{\Theta}_0\right)^{-1} \tag{27}$$

$(0, 0, 25)$. The users are randomly generated in a $120°$ sector with radius $r = 200$m around $(0, 0, 1.5)$. We normalize the channel as $\mathbb{E}\left\{\|\mathbf{G}_k\|_F^2\right\} = N_r N_p$. We adopt the adjustable phase shift pilots [12] as the training signal. The transmit power of the training signal for each user is set to 1. It should be noted that any other training signal can be applied. The SNR is set as $\text{SNR} = 1/\sigma_z^2$. Furthermore, we use the algorithm proposed in [13] to obtain the channel power matrices $\boldsymbol{\Omega}_k, \forall k$. The normalized mean-squared error (NMSE) is used as the performance metric for the channel estimation,

$$\text{NMSE} = \frac{1}{K N_{sam}} \sum_{k=1}^{K} \sum_{n=1}^{N_{sam}} \frac{\|\mathbf{G}_k^{(n)} - \hat{\mathbf{G}}_k^{(n)}\|_F^2}{\|\mathbf{G}_k^{(n)}\|_F^2}, \tag{28}$$

where N_{sam} is the number of the channel samples, $\mathbf{G}_k^{(n)}$ is the n-th channel sample of user k, $\hat{\mathbf{G}}_k^{(n)}$ is the estimate of the $\mathbf{G}_k^{(n)}$ and $\|\cdot\|_F$ is the F-norm. We set $N_{sam} = 200$ in our simulations. We compare the proposed IGA with the following algorithms.
GAMP: Generalized approximate message passing algorithm proposed in [18].
VEP: A low-complexity variant of the EP algorithm proposed in [19].
MMSE: The MMSE estimation of the beam domain channels based on (15a). Fig. 2 shows the NMSE performance of IGA

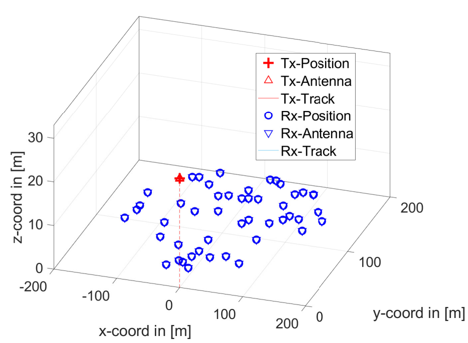

Fig. 1. The layout of the massive MIMO-OFDM system.

channel estimation compared with GAMP, VEP and MMSE with $F_v = F_h = F_\tau = 2$. The maximal iteration number of IGA, GAMP and VEP is set as 100. It can be found that the

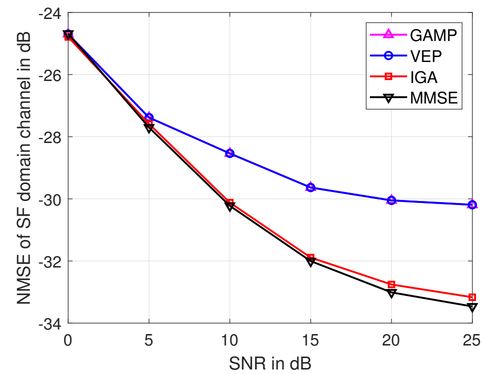

Fig. 2. NMSE performance of IGA channel estimation compared with GAMP, VEP and MMSE.

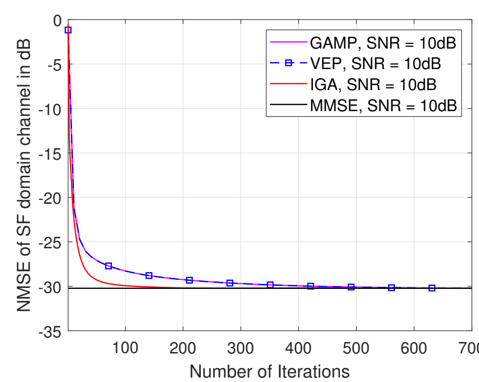

Fig. 3. Convergence performance of IGA channel estimation compared with GAMP and VEP at SNR = 10 dB.

IGA can obtain almost the same NMSE performance as the MMSE estimation at all SNRs. The SNR gain of the IGA compared to GAMP and VEP is about 5dB when the NMSE performance is -29dB.

Fig. 3 and Fig. 4 shows the convergence performance of IGA channel estimation compared with GAMP and VEP, where the SNR is set as 10dB and 20dB respectively, and $F_v = F_h = F_\tau = 2$. We can find that in the case with $\text{SNR} = 10$dB, IGA requires about 150 iterations to converge and achieves the optimal solution as that by the MMSE estimation, while the GAMP and VEP converge in more than 450 iterations.

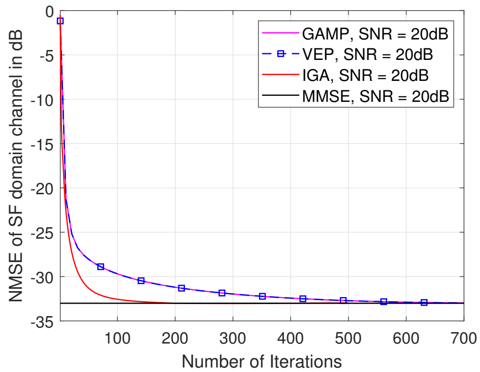

Fig. 4. Convergence performance of IGA channel estimation compared with GAMP and VEP at SNR = 20 dB.

In the case with SNR $= 20$dB, IGA converge in about 200 iterations, while GAMP and VEP take more than 550 iterations to converge. We can also find that VEP and GAMP show almost identical convergence behavior in all the simulations. This might be caused by the similarity in the processes of VEP and GAMP. The key difference between VEP and GAMP lies in the computation of one set of intermediate variables, which is denoted as $\{\tilde{\tau}_{m,n}\}$ in [19]. Other than that, they are nearly identical to each other [19]. Apart from the simulations in this work, VEP and GAMP also show almost identical convergence behavior in the simulations of [19], see, e.g., Fig. 3 (b) therein. Compared with GAMP and VEP, IGA has a faster convergence rate. The proposed IGA is derived based on the structure of the $a\ posteriori$ distribution $p\left(\mathbf{h}|\mathbf{y}\right)$ within the information geometry framework. The geometrical perspective provides an intuitive understanding of the statistical model, and thus allows to solve the statistical inference problem from an intrinsic and general standpoint. This might be a key reason for the improved convergence behavior of IGA compared with GAMP and VEP in massive MIMO channel estimation.

V. Conclusion

We have proposed an information geometry approach for channel estimation in massive MIMO-OFDM systems. We first derive the SF beam based statistical channel model for massive MIMO-OFDM systems by using sampled steering vectors in space and frequency domain. The accuracy of the beam based channel model is guaranteed by sufficiently large number of sampled steering vectors. With the established channel model, the channel estimation is formulated as calculating the $a\ posteriori$ information of the beam domain channel. We calculate approximate marginals of the $a\ posteriori$ distribution within the information geometry framework. Specifically, the calculation of the marginals is formulated as an iterative m-projection process. We derive the IGA for channel estimation by finding the solution of the m-projection. We show that the mean of the obtained marginals at the IGA's equilibrium equals the $a\ posteriori$ mean. Simulation results verify that the proposed IGA can obtain high channel estimation accuracy with much less number of iterations compared with the existing approaches. This demonstrates the superiority of our proposed channel estimation approach for massive MIMO-OFDM systems.

References

[1] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

[2] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

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

[4] L. Cimini, "Analysis and simulation of a digital mobile channel using orthogonal frequency division multiplexing," *IEEE Trans. Commun.*, vol. 33, no. 7, pp. 665–675, Jul. 1985.

[5] L. Tong, B. Sadler, and M. Dong, "Pilot-assisted wireless transmissions: general model, design criteria, and signal processing," *IEEE Signal Processing Magazine*, vol. 21, no. 6, pp. 12–25, Nov. 2004.

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

[7] 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.

[8] S. Ikeda, "Information geometrical framework to analyze the belief propagation algorithm," in *Workshop on Mathematics of Statistical Inference, December 2004, Tohoku University, Sendai, Japan*.

[9] M. Tang, Y. Rong, J. Zhou, and X. R. Li, "Information geometric approach to multisensor estimation fusion," *IEEE Trans. Signal Process.*, vol. 67, no. 2, pp. 279–292, Jan. 2019.

[10] L. Ye, Q. Yang, Q. Chen, and W. Deng, "Multidimensional joint domain localized matrix constant false alarm rate detector based on information geometry method with applications to high frequency surface wave radar," *IEEE Access*, vol. 7, pp. 28 080–28 088, 2019.

[11] F. Zhang, Y. Shi, H. K. T. Ng, and R. Wang, "Information geometry of generalized bayesian prediction using α -divergences as loss functions," *IEEE Trans. Inf. Theory*, vol. 64, no. 3, pp. 1812–1824, Mar, 2018.

[12] L. You, X. Q. Gao, A. L. Swindlehurst, and W. Zhong, "Channel acquisition for massive MIMO-OFDM with adjustable phase shift pilots," *IEEE Trans. Signal Process.*, vol. 64, no. 6, pp. 1461–1476, Mar. 2016.

[13] A.-A. Lu, X. Q. Gao, and C. Xiao, "Robust precoder design for 3D massive MIMO downlink with a posteriori channel model," *arXiv preprint arXiv:2004.04331*, 2020.

[14] O. Simeone, Y. Bar-Ness, and U. Spagnolini, "Pilot-based channel estimation for OFDM systems by tracking the delay-subspace," *IEEE Trans. Wireless Commun.*, vol. 3, no. 1, pp. 315–325, Jan. 2004.

[15] S. M. Kay, *Fundamentals of statistical signal processing*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[16] S. Rangan, P. Schniter, A. K. Fletcher, and S. Sarkar, "On the convergence of approximate message passing with arbitrary matrices," *IEEE Trans. Inf. Theory*, vol. 65, no. 9, pp. 5339–5351, 2019.

[17] 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.

[18] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *IEEE ISIT, St. Petersburg, Russia, July 31 - August 5, 2011*, pp. 2168–2172.

[19] D. Zhang, X. Song, W. Wang, G. Fettweis, and X. Q. Gao, "Unifying message passing algorithms under the framework of constrained bethe free energy minimization," *IEEE Trans. Wireless Commun.*, vol. 20, no. 7, pp. 4144–4158, Jul. 2021.