# Channel Estimation for Massive MIMO-OFDM: Simplified Information Geometry Approach

Jiyuan Yang*†, Yan Chen*†, An-An Lu*†, Wen Zhong*,
Xiqi Gao*†, Xiaohu You*†, 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
Email: {jyyang, 213160372, aalu, wzhong, xqgao,xhyu}@seu.edu.cn, xxia@ee.udel.edu, Dirk.Slock@eurecom.fr

***Abstract*—In this paper, we investigate the channel estimation for massive multiple-input multiple-output orthogonal frequency division multiplexing (MIMO-OFDM) systems. We revisit the information geometry approach (IGA) for massive MIMO-OFDM channel estimation. By using the constant magnitude property of the entries of the measurement matrix and the asymptotic analysis, we find that the second-order natural parameters (SONPs) of the distributions on all the auxiliary manifolds (AMs) are equivalent to each other at each iteration of IGA, and the first-order natural parameters (FONPs) of the distributions on all the AMs are asymptotically equivalent to each other at the fixed point. Motivated by these results, we simplify the iterative process of IGA and propose a simplified IGA for massive MIMO-OFDM channel estimation. It is proved that at the fixed point, the *a posteriori* mean obtained by the simplified IGA is asymptotically optimal. The simplified IGA allows efficient implementation with fast Fourier transformation (FFT). Simulations confirm that the simplified IGA can achieve near the optimal performance with low complexity in a limited number of iterations.**

## I. Introduction

Massive MIMO combined with OFDM can provide tremendous gains in both capacity and energy efficiency for communication systems [1]–[3]. To fully reap the various benefits of massive MIMO-OFDM, the accurate acquisition of the channel state information (CSI) is essential. Pilot-aided channel estimation is the common channel estimation approach for practical systems. Given the received pilot signal, the task of channel estimation is to obtain the *a posteriori* information of the channel parameters. With the Gaussian prior, the *a posteriori* distribution of the channel parameters is also Gaussian, of which the *a posteriori* information is determined by the mean vector and the covariance matrix. Nonetheless, the calculation of the optimal estimators, e.g., MMSE estimator, is usually unaffordable for the massive MIMO-OFDM systems due to the large dimension matrix inverse operation.

A differentiable manifold with a Riemannian structure can be regarded as the space defined by the parameters of the *a posteriori* probability density function. Hence, differential geometry definitions and tools can be useful in the calculation of the *a posteriori* distribution. This is exactly one of information geometry's topics [4]–[6]. Thus, it is appropriate to apply information geometry into the channel estimation. Recently, we have introduced the information geometry approach (IGA) to the massive MIMO-OFDM channel estimation [6]. We first provide the space-frequency (SF) beam based channel model for massive MIMO-OFDM system. The channel estimation is then formulated as obtaining the *a posteriori* information of the beam domain channel. By introducing the information geometry theory, we solve this problem through calculating the approximations for the marginals of the *a posteriori* distribution. More precisely, we turn the calculation of the approximations of the marginals into an iterative projection process by treating the set of Gaussian distributions with different constraints as different types of manifolds.

In this paper, we first revisit the proposed IGA. By using the constant magnitude entries of the measurement matrix, we reveal that at each iteration of IGA the second-order natural parameters (SONPs) of the distributions on all the AMs are equivalent to each other, and at the fixed point of IGA the first-order natural parameters (FONPs) of the distributions on all the AMs are asymptotically equivalent to each other. These two results motivate us to set the natural parameters (NPs) of the distributions on all the AMs as a common NP. Based on this, we simplify the iteration of IGA and propose a simplified IGA. It is also shown that at the fixed point, the *a posteriori* mean obtained by the simplified IGA is asymptotically optimal. At last, with the fast Fourier transform (FFT), we provide a low complexity implementation of the simplified IGA.

The rest of this paper is organized as follows. The system configuration and channel model are presented in Section II. We revisit IGA and reveal two new results in Section III. The simplified IGA for massive MIMO-OFDM channel estimation is proposed in Section IV. Simulation results are provided in Section V. The conclusion is drawn in Section VI.

## II. System Model and Problem Statement

### A. System Configuration and Channel Model

Consider a typical massive MIMO-OFDM system working in time division duplexing (TDD) mode with one base station

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.

(BS) serving $K$ single-antenna users within a cell, where the BS comprises a uniform planar array (UPA) of $N_r = N_{r,v} \times N_{r,h}$ antennas, and $N_{r,v}$ and $N_{r,h}$ are the numbers of the antennas at each vertical column and horizontal row, respectively. We focus on the uplink channel estimation since the CSI can be obtained from uplink traning, and then used for UL signal detection and downlink precoding due to channel reciprocity in TDD mode. Standard OFDM modulation with $N_c$ subcarriers is applied, where the cyclic prefix (CP) is $N_g$. $N_p$ training subcarriers are employed, and the set of them are denoted as $\mathcal{N}_p = \{N_1, N_1+1, \cdots, N_2\}$. Assume that the channel is quasi-static. Then, during each OFDM symbol, the SF domain received signal $\mathbf{Y} \in \mathbb{C}^{N_r \times N_p}$ for training at the BS can be expressed as [6]–[8]

$$\mathbf{Y} = \sum\nolimits_{k=1}^{K} \mathbf{G}_k \mathbf{P}_k + \mathbf{Z}, \tag{1}$$

where $\mathbf{G}_k \in \mathbb{C}^{N_r \times N_p}$ is the SF domain channel of user $k$, $\mathbf{P}_k = \mathrm{Diag}\{\mathbf{p}_k\} \in \mathbb{C}^{N_p \times N_p}$ is the pilot signal of user $k$, $\mathbf{p}_k$ is the pilot sequence of user $k$, $\mathrm{Diag}\{\mathbf{x}\}$ denotes the diagonal matrix with $\mathbf{x}$ along its main diagonal, and $\mathbf{Z}$ is the noise matrix with elements identically and independently distributed as $\mathcal{CN}\left(0, \sigma_z^2\right)$. With the SF beam based statistical channel model, $\mathbf{G}_k$ can be expressed as [6]

$$\mathbf{G}_k = \mathbf{V}\mathbf{H}_k\mathbf{F}^T, k \in \mathcal{Z}_K^+, \tag{2}$$

where $\mathbf{V} \triangleq \mathbf{V}_v \otimes \mathbf{V}_h \in \mathbb{C}^{N_r \times F_v F_h N_r}$, $\otimes$ denotes the Kronecker product, $\mathbf{V}_v \in \mathbb{C}^{N_{r,v} \times F_v N_{r,v}}$ and $\mathbf{V}_h \in \mathbb{C}^{N_{r,h} \times F_h N_{r,h}}$ are both partial discrete Fourier transformation (DFT) matrices. Specifically, $\mathbf{V}_v = \tilde{\mathbf{I}}_{N_{r,v} \times F_v N_{r,v}} \tilde{\mathbf{V}}_v$ and $\mathbf{V}_h = \tilde{\mathbf{I}}_{N_{r,h} \times F_h N_{r,h}} \tilde{\mathbf{V}}_h$, where $\tilde{\mathbf{V}}_v$ and $\tilde{\mathbf{V}}_h$ are DFT matrices of $F_v N_{r,v}$ and $F_h N_{r,h}$ points, respectively, and $\tilde{\mathbf{I}}_{N \times FN}$ is a matrix containing the first $N$ rows of the $FN$ dimensional identity matrix. $\mathbf{F} \triangleq \tilde{\mathbf{I}}_{N_p \times F_\tau N_p} \tilde{\mathbf{F}} \tilde{\mathbf{I}}_{F_\tau N_p \times F_\tau N_f} \in \mathbb{C}^{N_p \times N_\tau N_f}$, where $\tilde{\mathbf{F}}$ is a DFT matrix of $F_\tau N_p$ points, and $\tilde{\mathbf{I}}_{F_\tau N_p \times F_\tau N_f}$ is a matrix containing the first $F_\tau N_f$ columns of the $F_\tau N_p$ dimensional identity matrix. $N_f = \lceil N_p N_g / N_c \rceil$, where $\lceil x \rceil$ to denote the largest integer not larger than $x$. $F_v$, $F_h$ and $F_\tau$ are fine factors (FFs). $\mathbf{H}_k \in \mathbb{C}^{F_v F_h N_r \times F_\tau N_f}$ is the SF beam domain channel matrix of user $k$, and the elements in $\mathbf{H}_k$ follow the independent complex Gaussian distributions with zero mean and possibly different variances. We denote the power matrix of beam domain channel as $\mathbf{\Omega}_k = \mathbb{E}\{\mathbf{H}_k \odot \mathbf{H}_k^*\}, k \in \mathcal{Z}_K^+$, where $(\cdot)^*$ denotes the conjugate operator and $\odot$ is the Hadamard product. Due to the channel sparsity, most of the elements in $\mathbf{\Omega}_k$ are (close to) zero. Thus, the BS has sufficient resources to acquire $\mathbf{\Omega}_k$ [6], [7]. We assume that $\{\mathbf{\Omega}_k\}_{k=1}^K$ is known at the BS. By setting the FFs to be greater than 1, the SF beam based channel model allows to sample the angles of arrival as well as the delay more intensively and provides more accurate modeling of the channel in massive MIMO-OFDM systems.

### *B. Problem Statement*

In practice, the pilot sequences with constant magnitude property are preferred for massive MIMO-OFDM systems. In this paper, we use the adjustable phase shift pilots (APSPs) [7] as the training signal. Note that any other pilot sequences with constant magnitude can be adopted. We set the transmit power of the training signal for each user to 1. Then, the APSP for the user $k$ is set to be $\mathbf{P}_k = \mathrm{Diag}\{\mathbf{r}(n_k)\}\mathbf{P}$, where $\mathbf{r}(n_k) \in \mathbb{C}^{N_p \times 1}$,

$$\mathbf{r}(n_k) = \left[\exp\left\{-\bar{\jmath} 2\pi \frac{n_k N_1}{F_\tau N_p}\right\} \cdots \exp\left\{-\bar{\jmath} 2\pi \frac{n_k N_2}{F_\tau N_p}\right\}\right]^T, \tag{3}$$

$n_k \in \{0, 1, \cdots, F_\tau N_p - 1\}$, and $\mathbf{P} = \mathrm{Diag}\{\mathbf{p}\}$ is the basic pilot satisfying $\mathbf{P}\mathbf{P}^H = \mathbf{I}$. Given $\{\mathbf{\Omega}_k\}_{k=1}^K$, we can use [7, Algorithm 1] to determine the value of $n_k$ and thus $\mathbf{P}_k$ for each user. Define a partial DFT matrix of $F\tau N_p$ points as $\mathbf{F}_d \triangleq [\mathbf{r}(0)\ \mathbf{r}(1)\ \cdots\ \mathbf{r}(F_\tau N_p - 1)] \in \mathbb{C}^{N_p \times F_\tau N_p}$ and a permutation matrix as

$$\mathbf{\Pi}_{n_k} \triangleq \begin{bmatrix} \mathbf{O} & \mathbf{I}_{F_\tau N_p - n_k} \\ \mathbf{I}_{n_k} & \mathbf{O} \end{bmatrix} \in \mathbb{C}^{F_\tau N_p \times F_\tau N_p}. \tag{4}$$

Substituting $\mathbf{P}_k$ and (2) into (1), we can obtain

$$\mathbf{Y} = \mathbf{V}\mathbf{H}\mathbf{F}_d^T\mathbf{P} + \mathbf{Z}, \tag{5}$$

where $\mathbf{H} = \sum_{k=1}^K \mathbf{H}_k^{\mathrm{e}} \mathbf{\Pi}_{n_k}$, $\mathbf{H}_k^{\mathrm{e}} = [\mathbf{H}_k\ \mathbf{O}] \in \mathbb{C}^{F_a N_r \times F_\tau N_p}$, and $F_a \triangleq F_v F_h$. Define $\mathbf{\Omega} \triangleq \sum_{k=1}^K \mathbf{\Omega}_k^{\mathrm{e}} \mathbf{\Pi}_{n_k}$ with $\mathbf{\Omega}_k^{\mathrm{e}} \triangleq [\mathbf{\Omega}_k\ \mathbf{O}] \in \mathbb{C}^{F_a N_r \times F_\tau N_p}$. It can be checked that $\mathbf{\Omega}$ is the power matrix of $\mathbf{H}$. Then, we can obtain

$$\mathbf{y} = \mathrm{vec}\left\{\mathbf{Y}\mathbf{P}^H\right\} = \tilde{\mathbf{A}}\tilde{\mathbf{h}} + \mathbf{z}, \tag{6}$$

where $\tilde{\mathbf{A}} = \mathbf{F}_d \otimes \mathbf{V} \in \mathbb{C}^{N \times M_a}$, $\tilde{\mathbf{h}} = \mathrm{vec}\{\mathbf{H}\} \in \mathbb{C}^{M_a \times 1}$, $M_a = F_a F_\tau N$, $\mathbf{z} = \mathrm{vec}\{\mathbf{Z}\mathbf{P}^H\} \in \mathbb{C}^{N \times 1}$, and $\mathbf{z} \sim \mathcal{CN}\left(\mathbf{0}, \sigma_z^2 \mathbf{I}\right)$ since $\mathbf{P}^H$ is unitary. Define the number of non-zero elements in $\boldsymbol{\omega} \triangleq \mathrm{vec}\{\mathbf{\Omega}\}$ as $M$ and the indexes of these non-zero elements as $\mathcal{P} \triangleq \{p_1, p_2, \cdots, p_M\}$. Define an extraction matrix as $\mathbf{E} \triangleq [\mathbf{e}_{p_1}, \mathbf{e}_{p_2}, \cdots, \mathbf{e}_{p_M}] \in \mathbb{C}^{M_a \times M}$, where $\mathbf{e}_i$ is the $i$-th column of the $M_a$ dimensional identity matrix. Then, $\mathbf{y}$ can be re-expressed as

$$\mathbf{y} = \mathbf{A}\mathbf{h} + \mathbf{z}, \tag{7}$$

where $\mathbf{A} = \tilde{\mathbf{A}}\mathbf{E} \in \mathbb{C}^{N \times M}$, $\mathbf{h} = \mathbf{E}^T\tilde{\mathbf{h}} \in \mathbb{C}^{M \times 1}$, $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}, \mathbf{D})$ and $\mathbf{D} \triangleq \mathrm{Diag}\left\{\mathbf{E}^T\boldsymbol{\omega}\right\}$ is positive definite. In this case, the elements in the measurement matrix $\mathbf{A}$ have unit magnitude. Given the observation $\mathbf{y}$, the *a posteriori* distribution of $\mathbf{h}$ is Gaussian, of which the mean and covariance are given by [9]

$$\tilde{\boldsymbol{\mu}} = \mathbf{D}\left(\mathbf{A}^H\mathbf{A}\mathbf{D} + \sigma_z^2\mathbf{I}\right)^{-1}\mathbf{A}^H\mathbf{y}, \tag{8a}$$

$$\tilde{\mathbf{\Sigma}} = \left(\mathbf{D}^{-1} + \mathbf{A}^H\mathbf{A}/\sigma_z^2\right)^{-1}. \tag{8b}$$

$\tilde{\boldsymbol{\mu}}$ is also the MMSE estimate of $\mathbf{h}$. The computational complexity of (8) is $\mathcal{O}\left(M^3 + M^2 N\right)$, of which the application is unaffordable when $N$ and $M$ are large. In this paper, we propose a simplified IGA to compute the marginals of the *a posteriori* distribution.

## III. Revisiting IGA

With the received signal model (7), the PDF of the *a posteriori* distribution can be expressed as

$$p(\mathbf{h}|\mathbf{y}) = \exp\left\{\mathbf{d} \circ \mathbf{t} + \sum\nolimits_{n=1}^{N} c_n(\mathbf{h}) - \psi_q\right\}, \tag{9}$$

where $\mathbf{d} \triangleq \mathbf{f}\left(\mathbf{0}, \mathrm{diag}\left\{-\mathbf{D}^{-1}\right\}\right)$, $\mathrm{diag}\{\mathbf{X}\}$ denotes a vector consisting of the diagonal elements of $\mathbf{X}$, $\mathbf{t} \triangleq \mathbf{f}(\mathbf{h}, (\mathbf{h} \odot \mathbf{h}^*))$, $\mathbf{f}(\mathbf{a}, \mathbf{b}) \triangleq \left[\mathbf{a}^T,\ \mathbf{b}^T\right]^T$, $\mathbf{a} \circ \mathbf{b} \triangleq \left(\mathbf{b}^H\mathbf{a} + \mathbf{a}^H\mathbf{b}\right)/2$, $c_n(\mathbf{h}) =$

$\left(-\mathbf{h}^H\boldsymbol{\gamma}_n\boldsymbol{\gamma}_n^H\mathbf{h}+y_n\mathbf{h}^H\boldsymbol{\gamma}_n+y_n^*\boldsymbol{\gamma}_n^H\mathbf{h}\right)/\sigma_z^2$, $y_n$ is the $n$-th element of $\mathbf{y}$, $\boldsymbol{\gamma}_n=\left[\mathbf{A}^H\right]_{:,n}\in\mathbb{C}^{M\times 1}$, and $\psi_q$ are the normalization factor. IGA [6] constructs two types of manifolds, the objective manifold (OBM) and the auxiliary manifold (AM). The OBM is defined as

$$\mathcal{M}_0=\{p_0(\mathbf{h};\boldsymbol{\vartheta}_0)=\exp\{(\mathbf{d}+\boldsymbol{\vartheta}_0)\circ\mathbf{t}-\psi_0\}\},\tag{10}$$

where $\boldsymbol{\vartheta}_0=\mathbf{f}(\boldsymbol{\theta}_0,\boldsymbol{\nu}_0)$ with $\boldsymbol{\theta}_0\in\mathbb{C}^{M\times 1}$ and $\boldsymbol{\nu}_0\in\mathbb{R}^{M\times 1}$, and $\psi_0$ is the free energy (normalization factor) . We refer to $\boldsymbol{\vartheta}_0$, $\boldsymbol{\theta}_0$ and $\boldsymbol{\nu}_0$ as the NP, the FONP and the SONP of $p_0$. $N$ AMs are defined, where the $n$-th AM is defined as

$$\mathcal{M}_n=\{p_n(\mathbf{h};\boldsymbol{\vartheta}_n)\},n\in\mathcal{Z}_N^+,\tag{11a}$$

$$p_n(\mathbf{h};\boldsymbol{\vartheta}_n)=\exp\{(\mathbf{d}+\boldsymbol{\vartheta}_n)\circ\mathbf{t}+c_n(\mathbf{h})-\psi_n\},\tag{11b}$$

where $\boldsymbol{\vartheta}_n=\mathbf{f}(\boldsymbol{\theta}_n,\boldsymbol{\nu}_n)$, $\boldsymbol{\theta}_n\in\mathbb{C}^{M\times 1}$ and $\boldsymbol{\nu}_n\in\mathbb{R}^{M\times 1}$ are referred to as the NP, the FONP and the SONP of $p_n$, $\mathcal{Z}_N^+=\{1,2,\ldots,N\}$, and $\psi_n$ is the free energy. Based on the OBM and the AMs, IGA approximates $\sum_{n=1}^N c_n(\mathbf{h})$ as $\boldsymbol{\vartheta}_0\circ\mathbf{t}$ in a iterative manner. More precisely, we first initialize the NPs $\boldsymbol{\vartheta}_n,n\in\mathcal{Z}_N$, where $\mathcal{Z}_N=\{0,1,\ldots,N\}$. Then, at the $t$-th iteration, given $\boldsymbol{\vartheta}_n^t$ of $p_n(\mathbf{h};\boldsymbol{\vartheta}_n^t),n\in\mathcal{Z}_N^+$, IGA calculates an approximation item of $c_n(\mathbf{h})$ through $m$-projecting $p_n$ onto the OBM, where $m$-projecting $p_n$ onto the OBM is equivalent to finding the point on the OBM minimizing the K-L divergence from $p_n$ to the OBM, i.e.,

$$\boldsymbol{\vartheta}_{0n}^t=\arg\min_{\boldsymbol{\vartheta}_0}D_{KL}\left\{p_n\left(\mathbf{h};\boldsymbol{\vartheta}_n^t\right):p_0(\mathbf{h};\boldsymbol{\vartheta}_0)\right\},\tag{12a}$$

$$D_{KL}\left\{p_n\left(\mathbf{h};\boldsymbol{\vartheta}_n^t\right):p_0(\mathbf{h};\boldsymbol{\vartheta}_0)\right\}=\mathbb{E}_{p_n}\left\{\ln\frac{p_n(\mathbf{h};\boldsymbol{\vartheta}_n^t)}{p_0(\mathbf{h};\boldsymbol{\vartheta}_0)}\right\}.\tag{12b}$$

$\boldsymbol{\vartheta}_{0n}^t=\mathbf{f}(\boldsymbol{\theta}_{0n}^t,\boldsymbol{\nu}_{0n}^t)$ is given by [6, Equation (47)]. Then, the approximation item is calculated as

$$\boldsymbol{\xi}_n=\boldsymbol{\vartheta}_{0n}-\boldsymbol{\vartheta}_n,n\in\mathcal{Z}_N^+.\tag{13}$$

We update the NPs as the following

$$\boldsymbol{\vartheta}_n^{t+1}=d\sum\nolimits_{n'\neq n}\boldsymbol{\xi}_{n'}^t+(1-d)\,\boldsymbol{\vartheta}_n^t,\tag{14a}$$

$$\boldsymbol{\vartheta}_0^{t+1}=d\sum\nolimits_{n=1}^{N}\boldsymbol{\xi}_n^t+(1-d)\,\boldsymbol{\vartheta}_0^t,\tag{14b}$$

where $0<d\leq 1$ is the damping introduced to increase the convergence of the IGA. Then, repeat the $m$-projections, calculate the approximation items and the updates until convergence or $t$ reaches the preset number. The mean and variance of the approximate marginal, $p(h_i|\mathbf{y}),i\in\mathcal{Z}_M^+$, are given by the $i$-th component of $\boldsymbol{\mu}_0$ and $\operatorname{diag}\{\boldsymbol{\Sigma}_0\}$, respectively, where $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ are given by (22).

We now present two new results on the IGA when $\mathbf{A}$ in (7) has the constant magnitude entries. Unless specified, we assume that this condition holds in the rest of this paper. We omit the proofs for the Theorems in this paper due to space limitation.

**Theorem 1.** *If* $\mathbf{A}$ *in* (7) *has constant magnitude entry property, then at each iteration of IGA, the SONPs of* $p_n,n\in\mathcal{Z}_N^+$, *are independent of* $n$, *i.e.,* $\boldsymbol{\nu}_n^t=\boldsymbol{\nu}_{n'}^t,n,n'\in\mathcal{Z}_N^+$, *when the initializations of the SONPs of* $p_1,p_2,\cdots,p_N$ *are the same.*

Define the arithmetic mean of the SONPs of $\{p_n\}_{n=1}^N$ as $\boldsymbol{\nu}\triangleq\frac{1}{N}\sum_{n=1}^N\boldsymbol{\nu}_n$. From Theorem 1, $\boldsymbol{\nu}_n,n\in\mathcal{Z}_N^+$, in IGA can be replaced by $\boldsymbol{\nu}$ in each iteration, and the two iteration modes are equivalent when $\mathbf{A}$ has constant magnitude entry property. Motivated by this observation, we find that a similar property is satisfied between the FONPs of $p_1,p_2,\ldots,p_N$, in IGA.

Denote the fixed points of the NPs in the IGA as $\boldsymbol{\vartheta}_n^\star=\mathbf{f}(\boldsymbol{\theta}_n^\star,\boldsymbol{\nu}_n^\star),n\in\mathcal{Z}_N$. For an $M\times M$ positive definite diagonal matrix $\mathbf{D}$, define $\|\boldsymbol{\theta}\|_{\mathbf{D}}\triangleq\sqrt{\boldsymbol{\theta}^H\mathbf{D}\boldsymbol{\theta}}$, where $\boldsymbol{\theta}\in\mathbb{C}^{M\times 1}$. Since $\mathbf{D}$ is positive definite diagonal, we have $\|\boldsymbol{\theta}\|_{\mathbf{D}}=\|\mathbf{D}^{\frac{1}{2}}\boldsymbol{\theta}\|$, where $\|\cdot\|$ is the $\ell_2$ norm. And $\|\cdot\|_{\mathbf{D}}$ is a weighted norm on $\mathbb{C}^{M\times 1}$. Then, we have the following result.

**Theorem 2.** *In IGA, the fixed points of all the FONPs of* $p_1,p_2,\cdots,p_N$ *are asymptotically equal to* $\frac{N-1}{N}$ *times the fixed point of the FONP of* $p_0$, *i.e.,*

$$\lim_{N\to\infty}\lim_{M\to\infty}\frac{1}{NM}\sum_{n=1}^{N}\mathbb{E}\left\{\|\boldsymbol{\theta}_n^\star-\frac{N-1}{N}\boldsymbol{\theta}_0^\star\|_{\mathbf{D}}^2\right\}=0.\tag{15}$$

Theorem 2 illustrates that as $N$ and $M$ tend to infinity, the average error between each element in the fixed point of the FONP of $p_n,n\in\mathcal{Z}_N^+$, and each element in the fixed point of the FONP of $p_0$ is asymptotically equal to zero, which indicates that the fixed point of the FONP of $p_n,n\in\mathcal{Z}_N^+$, tends to be equal to each other. In massive MIMO-OFDM channel estimation, $N$ is usually quite large. When the number of users is high, $M$ can be large enough to be comparable to $N$. Define the arithmetic mean of the NPs of $p_1,p_2,\cdots,p_N$, as $\boldsymbol{\vartheta}\triangleq\frac{1}{N}\sum_{n=1}^N\boldsymbol{\vartheta}_n$. Inspired by Theorem 1 and 2, we will use $\boldsymbol{\vartheta}$ instead of $\boldsymbol{\vartheta}_n,n\in\mathcal{Z}_N^+$, to simplify the iteration of IGA.

## IV. Simplified IGA

Instead of (7), we use the following received signal model to develop the simplified IGA in this section,

$$\mathbf{y}=\mathbf{A}\mathbf{h}+\tilde{\mathbf{z}},\tag{16}$$

where $\tilde{\mathbf{z}}\sim\mathcal{CN}\left(\mathbf{0},\tilde{\sigma}_z^2\mathbf{I}\right)$, $\tilde{\sigma}_z^2$ is a positive constant, and the other variables are the same as those in (7). Compared to the real received signal model (7), (16) is a virtual received signal model, where we fictitiously treat the noise vector as $\tilde{\mathbf{z}}$ rather than the true one, i.e., $\mathbf{z}$. By introducing (16), the input noise variance of the simplified IGA is changed from $\sigma_z^2$ to $\tilde{\sigma}_z^2$. We shall see that by determining the value of $\tilde{\sigma}_z^2$ based on the true noise variance $\sigma_z^2$, we can enable the simplified IGA to obtain an estimate of $\mathbf{h}$ which is asymptotically equal to the *a posteriori* mean $\tilde{\boldsymbol{\mu}}$ in (8a).

### A. Simplified IGA

Compared with IGA, the simplified IGA has the same input except that the noise power is replaced by $\tilde{\sigma}_z^2$. At the initialization, we set $t=0$ and choose the damping $d$, where $0<d\leq 1$. We initialize the NP for $p_0$ as $\boldsymbol{\vartheta}_0^t$ and initialize the NP for $\{p_n\}_{n=1}^N$ as $\boldsymbol{\vartheta}^t$ while ensuring that the SONPs in $\boldsymbol{\vartheta}_0^t$ and $\boldsymbol{\vartheta}^t$ satisfy $\boldsymbol{\nu}_0^t,\boldsymbol{\nu}^t<0$. We refer to $\boldsymbol{\vartheta}$ as the common NP of $\{p_n\}_{n=1}^N$ (abbreviated as the common NP). Given the common NP $\boldsymbol{\vartheta}^t=\mathbf{f}(\boldsymbol{\theta}^t,\boldsymbol{\nu}^t)$ at the $t$-th iteration, the simplified IGA $m$-projects $p_n(\mathbf{h};\boldsymbol{\vartheta}^t)$ onto the OBM and obtains the $m$-projection, denoted as $p_0(\mathbf{h};\boldsymbol{\vartheta}_{0n}^t)$,where $n\in\mathcal{Z}_N^+$. The approximation item $\boldsymbol{\xi}_n^t$ is then re-expressed as

$$\boldsymbol{\xi}_n^t=\boldsymbol{\vartheta}_{0n}^t-\boldsymbol{\vartheta}^t,n\in\mathcal{Z}_N^+,\tag{17}$$

since we replace $\boldsymbol{\vartheta}_n^t, n \in \mathcal{Z}_N^+$, with $\boldsymbol{\vartheta}^t$. Then, from (14a), $\left\{\boldsymbol{\vartheta}_n^{t+1}\right\}_{n=1}^N$ can be obtained. Although $\left\{\boldsymbol{\vartheta}_n^{t+1}\right\}_{n=1}^N$, have the same SONPs at each iteration of IGA, the FONPs of $\left\{\boldsymbol{\vartheta}_n^{t+1}\right\}_{n=1}^N$ are only asymptotically equal to each other at the fixed point, and thus, $\left\{\boldsymbol{\vartheta}_n^{t+1}\right\}_{n=1}^N$ are not necessarily equal to each other at each iteration. To update the common NP $\boldsymbol{\vartheta}$ in the simplified IGA, we calculate $\boldsymbol{\vartheta}^{t+1}$ as the arithmetic mean of $\left\{\boldsymbol{\vartheta}_n^{t+1}\right\}_{n=1}^N$,

$$\begin{aligned}\boldsymbol{\vartheta}^{t+1} &= \frac{1}{N}\sum_{n=1}^{N}\boldsymbol{\vartheta}_n^{t+1}\\ &\overset{\text{(a)}}{=} \frac{d}{N}\sum_{n=1}^{N}\sum_{n'=1}^{N}\left(\boldsymbol{\xi}_{n'}^t-\boldsymbol{\xi}_n^t\right)+\frac{1-d}{N}\sum_{n=1}^{N}\boldsymbol{\vartheta}_n^t\\ &\overset{\text{(b)}}{=} \frac{d\left(N-1\right)}{N}\sum_{n=1}^{N}\boldsymbol{\xi}_n^t+\left(1-d\right)\boldsymbol{\vartheta}^t\\ &\overset{\text{(c)}}{=} \frac{d\left(N-1\right)}{N}\sum_{n=1}^{N}\boldsymbol{\vartheta}_{0n}^t+\left(1-dN\right)\boldsymbol{\vartheta}^t,\end{aligned} \tag{18}$$

where (a) comes from (14a), (b) comes from that if $\boldsymbol{\vartheta}$ is updated as above, then at each iteration of the simplified IGA, $\boldsymbol{\vartheta}^t=\frac{1}{N}\sum_{n=1}^{N}\boldsymbol{\vartheta}_n^t$ can be obtained, and (c) comes from (17). From (14b), the update of $\boldsymbol{\vartheta}_0$ can be modified as

$$\begin{aligned}\boldsymbol{\vartheta}_0^{t+1} &= d\sum_{n=1}^{N}\left(\boldsymbol{\vartheta}_{0n}^t-\boldsymbol{\vartheta}^t\right)+\left(1-d\right)\boldsymbol{\vartheta}_0^t\\ &= d\sum_{n=1}^{N}\boldsymbol{\vartheta}_{0n}^t-dN\boldsymbol{\vartheta}^t+\left(1-d\right)\boldsymbol{\vartheta}_0^t.\end{aligned} \tag{19}$$

We now discuss the simplification of the update way of $\boldsymbol{\vartheta}_0$ in (19), which is derived directly from the non-damping version of (18) and (19). Setting $d=1$ in (18) and (19), and after some calculation, we can obtain $\left(N-1\right)\boldsymbol{\vartheta}_0^{t+1}=N\boldsymbol{\vartheta}^{t+1}$. Then, when $0<d<1$, if we constrain $\left(N-1\right)\boldsymbol{\vartheta}_0^t=N\boldsymbol{\vartheta}^t$, $t=0$, at the initialization, at each iteration of (18) and (19), we still have $\left(N-1\right)\boldsymbol{\vartheta}_0^t=N\boldsymbol{\vartheta}^t,\forall t$. In summary, when the initialization satisfies $\left(N-1\right)\boldsymbol{\vartheta}_0=N\boldsymbol{\vartheta}$, the update of the NPs in the simplified IGA can be summarized as

$$\boldsymbol{\vartheta}^{t+1}=\frac{d\left(N-1\right)}{N}\sum_{n=1}^{N}\boldsymbol{\vartheta}_{0n}^t+\left(1-dN\right)\boldsymbol{\vartheta}^t, \tag{20a}$$

$$\boldsymbol{\vartheta}_0^{t+1}=\frac{N}{N-1}\boldsymbol{\vartheta}^t. \tag{20b}$$

We give the detailed expression of $\boldsymbol{\vartheta}^{t+1}=\mathbf{f}\left(\boldsymbol{\theta}^{t+1},\boldsymbol{\nu}^{t+1}\right)$ in (21). In (21), the common NP $\boldsymbol{\vartheta}^{t+1}$ is directly calculated without the step for calculating the approximation item $\boldsymbol{\xi}_n^t$. From (20b), we can see that the NP of $p_0$ in each iteration relies on the common NP only (also vice versa). Therefore, its updating in the iteration process is not necessary. We only need to calculate the NP of $p_0$ with the resulting common NP from the iteration process. We summarize the simplified IGA in Algorithm 1.

We then provide the analysis of the fixed point of the simplified IGA. When converged, denote the fixed point of the common NP as $\boldsymbol{\vartheta}^\star=\mathbf{f}\left(\boldsymbol{\theta}^\star,\boldsymbol{\nu}^\star\right)$. Denote the NP of $p_0$ at the fixed point of the simplified IGA as $\boldsymbol{\vartheta}_0^\star=\mathbf{f}\left(\boldsymbol{\theta}_0^\star,\boldsymbol{\nu}_0^\star\right)\triangleq N/\left(N-1\right)\boldsymbol{\vartheta}^\star$. Denote the mean of $p_0\left(\mathbf{h};\boldsymbol{\vartheta}_0^\star\right)$ as $\boldsymbol{\mu}_0^\star=\boldsymbol{\mu}_0\left(\boldsymbol{\vartheta}_0^\star\right)$, where $\boldsymbol{\mu}_0\left(\cdot\right)$ is given by (22). Then, we have the following theorem.

**Theorem 3.** *If the initialization of the SONP of the common NP in the simplified IGA satisfies $\boldsymbol{\nu}^t<0$, $t=0$, then, the fixed points of the SONPs of the common NP and the NP of $p_0$ satisfy $\boldsymbol{\nu}^\star,\boldsymbol{\nu}_0^\star<0$, and $\boldsymbol{\mu}_0^\star$ satisfies*

$$\boldsymbol{\mu}_0^\star=\mathbf{D}\left[\mathbf{A}^H\mathbf{A}\left(\mathbf{D}-\frac{1}{N}\boldsymbol{\Lambda}^\star\right)+\beta^\star\mathbf{I}\right]^{-1}\mathbf{A}^H\mathbf{y}, \tag{23}$$

*where*

$$\boldsymbol{\Lambda}^\star\triangleq\left(\mathbf{D}^{-1}-\mathrm{Diag}\left\{\boldsymbol{\nu}^\star\right\}\right)^{-1}, \tag{24a}$$

$$\beta^\star\triangleq\tilde{\sigma}_z^2+\mathrm{tr}\left\{\boldsymbol{\Lambda}^\star\right\}>0. \tag{24b}$$

Theorem 3 provides the expression of $\boldsymbol{\mu}_0^\star$ in the simplified IGA. We then show that $\boldsymbol{\mu}_0^\star$ above is asymptotically optimal when $M<N$ and $N$ tends to infinity, where $M$ and $N$ are the numbers of the variables to be estimated and the observations, respectively. We first define an injection as $f:\mathbb{R}^+\to\mathbb{R}$,

$$f\left(x\right)=x-\mathrm{tr}\left\{\left(\mathbf{D}^{-1}+\frac{N-1}{x}\mathbf{I}\right)^{-1}\right\},x>0. \tag{25}$$

**Theorem 4.** *When the initialization of the SONP of the common NP in the simplified IGA satisfies $\boldsymbol{\nu}^t<0$, $t=0$, and $M<N$, the asymptotic values of $\boldsymbol{\Lambda}^\star$ and $f\left(\beta^\star\right)$ as $N$ tends to infinity satisfy*

$$\lim_{N\to\infty}\left[\boldsymbol{\Lambda}^\star\right]_{i,i}=0,i\in\mathcal{Z}_M^+, \tag{26a}$$

$$\lim_{N\to\infty}f\left(\beta^\star\right)=\tilde{\sigma}_z^2. \tag{26b}$$

*Then, if $\tilde{\sigma}_z^2=f\left(\sigma_z^2\right)$, we can obtain $\lim_{N\to\infty}\beta^\star=\sigma_z^2$ and*

$$\lim_{N\to\infty}\boldsymbol{\mu}_0^\star=\tilde{\boldsymbol{\mu}}, \tag{27}$$

*where $\tilde{\boldsymbol{\mu}}$ is the a posteriori mean in* (8a)*. Moreover, when $M$ is fixed, we have $\lim_{N\to\infty}f\left(\sigma_z^2\right)=\sigma_z^2$. In this case,* (27) *holds if either $\tilde{\sigma}_z^2=\sigma_z^2$ or $\tilde{\sigma}_z^2=f\left(\sigma_z^2\right)$ is satisfied.*

Theorem 4 provides the asymptotic values of $\boldsymbol{\Lambda}^\star$ and $f\left(\beta^\star\right)$ when $N$ tends to infinity and $M<N$. It also illustrates that $\boldsymbol{\mu}_0^\star$ is asymptotically optimal as $N$ tends to infinity and $M<N$ when $\tilde{\sigma}_z^2$ is set to be $f\left(\sigma_z^2\right)$. Meanwhile, it can be checked that $f\left(\sigma_z^2\right)<\sigma_z^2$.

**Algorithm 1:** Simplified IGA

**Input:** The covariance $\mathbf{D}$ of the priori distribution $p\left(\mathbf{h}\right)$, the received signal $\mathbf{y}$, the noise power $\tilde{\sigma}_z^2$ and the maximal iteration number $t_{\max}$.

**Initialization:** set $t=0$, set damping $d$, where $0<d\leq1$, initialize the common NP as $\boldsymbol{\vartheta}^t=\mathbf{f}\left(\boldsymbol{\theta}^t,\boldsymbol{\nu}^t\right)$ and ensure $\boldsymbol{\nu}^t<0$;

**repeat**

1. Update $\boldsymbol{\vartheta}=\mathbf{f}\left(\boldsymbol{\theta},\boldsymbol{\nu}\right)$ as (21);
2. $t=t+1$;

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

**Output:** Calculate the NP of $p_0\left(\mathbf{h};\boldsymbol{\vartheta}_0\right)$ as $\boldsymbol{\vartheta}_0=\frac{N}{N-1}\boldsymbol{\vartheta}^t$. The mean and variance of the approximate marginal, $p\left(h_i|\mathbf{y}\right)$, $i\in\mathcal{Z}_M^+$, are given by the $i$-th component of $\boldsymbol{\mu}_0$ and $\mathrm{diag}\left\{\boldsymbol{\Sigma}_0\right\}$, respectively, where $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ are calculated by (22).

$$\boldsymbol{\theta}^{t+1}=\frac{d\left(N-1\right)}{N}\left(\mathbf{I}-\frac{1}{\beta^{t}}\boldsymbol{\Lambda}^{t}\right)^{-1}\left[\frac{1}{\beta^{t}}\mathbf{A}^{H}\left(2\mathbf{y}-\mathbf{A}\boldsymbol{\Lambda}^{t}\boldsymbol{\theta}^{t}\right)+N\boldsymbol{\theta}^{t}\right]+\left(1-dN\right)\boldsymbol{\theta}^{t} \tag{21a}$$

$$\boldsymbol{\nu}^{t+1}=d\left(N-1\right)\mathrm{diag}\left\{\mathbf{D}^{-1}-\left[\boldsymbol{\Lambda}^{t}-\frac{1}{\beta^{t}}\left(\boldsymbol{\Lambda}^{t}\right)^{2}\right]^{-1}\right\}+\left(1-dN\right)\boldsymbol{\nu}^{t} \tag{21b}$$

$$\boldsymbol{\Lambda}^{t}=\left(\mathbf{D}^{-1}-\mathrm{Diag}\left\{\boldsymbol{\nu}^{t}\right\}\right)^{-1},\ \beta\left(t\right)=\tilde{\sigma}_{z}^{2}+\mathrm{tr}\left\{\boldsymbol{\Lambda}^{t}\right\} \tag{21c}$$

$$\boldsymbol{\mu}_{0}\left(\boldsymbol{\vartheta}_{0}\right)=\frac{1}{2}\boldsymbol{\Sigma}_{0}\left(\boldsymbol{\vartheta}_{0}\right)\boldsymbol{\theta}_{0},\ \boldsymbol{\Sigma}_{0}\left(\boldsymbol{\vartheta}_{0}\right)=\left(\mathbf{D}^{-1}-\mathrm{Diag}\left\{\boldsymbol{\nu}_{0}\right\}\right)^{-1} \tag{22}$$

### *B. Efficient Implementation*

The computational complexity of each iteration of the simplified IGA mainly comes from the two matrix-vector multiplications by $\mathbf{A}$ and $\mathbf{A}^H$ in (21a). In this subsection, we focus on (21a) and present an efficient implementation. At each iteration, (21a) can be rewritten as (we omit the counter $t$ on the right-side of the equation for convenience)

$$\boldsymbol{\theta}^{t+1}=\frac{2\mathbf{J}\mathbf{A}^{H}\mathbf{y}-\mathbf{J}\mathbf{A}^{H}\mathbf{A}\boldsymbol{\Lambda}\boldsymbol{\theta}^{t}}{\beta}+\left[N\mathbf{J}+\left(1-dN\right)\mathbf{I}\right]\boldsymbol{\theta}^{t}, \tag{28a}$$

$$\mathbf{J}=\frac{d\left(N-1\right)}{N}\left(\mathbf{I}-\frac{1}{\beta}\boldsymbol{\Lambda}\right)^{-1}. \tag{28b}$$

Since both $\mathbf{J}$ and $\boldsymbol{\Lambda}$ are diagonal, the complexity in (28a) mainly comes from $\mathbf{A}^H\mathbf{y}$, $\mathbf{A}^H\mathbf{s}$ and $\mathbf{A}\mathbf{u}$, where $\mathbf{s}=\mathbf{A}\boldsymbol{\Lambda}\boldsymbol{\theta}^t\in\mathbb{C}^{N\times 1}$ and $\mathbf{u}=\boldsymbol{\Lambda}\boldsymbol{\theta}^t\in\mathbb{C}^{M\times 1}$. For $\mathbf{A}\mathbf{u}$, we have $\mathbf{A}\mathbf{u}=\tilde{\mathbf{A}}\tilde{\mathbf{u}}=\mathrm{vec}\left\{\mathbf{V}\tilde{\mathbf{U}}\mathbf{F}_d^T\right\}$, where $\tilde{\mathbf{u}}=\mathbf{E}\mathbf{u}\in\mathbb{C}^{F_aF_\tau N\times 1}$, $\tilde{\mathbf{U}}\in\mathbb{C}^{F_aN_r\times F_\tau N_p}$ and $\mathrm{vec}\left\{\tilde{\mathbf{U}}\right\}=\tilde{\mathbf{u}}$. Then, $\mathbf{V}\tilde{\mathbf{U}}\mathbf{F}_d^T$ can be calculated by FFT since $\mathbf{V}$ is the Kronecker product of two partial DFT matrices and $\mathbf{F}_d$ is a partial DFT matrix. The complexity of the efficient implementation of $\mathbf{A}\mathbf{u}$ is $\mathcal{O}\left(C\right)$, where

$$\begin{aligned}C=&N\Big[F_aF_\tau\log_2\left(F_vN_{r,v}\right)+F_hF_\tau\log_2\left(F_hN_{r,h}\right)\\&+F_\tau\log_2\left(F_\tau N_p\right)\Big].\end{aligned} \tag{29}$$

For the calculation of $\mathbf{A}^H\mathbf{s}$, we have that $\mathbf{A}^H\mathbf{s}=\mathbf{E}^T\tilde{\mathbf{A}}^H\mathbf{s}=\mathbf{E}^T\mathrm{vec}\left\{\mathbf{V}^H\mathbf{S}\mathbf{F}_d^*\right\}$, where $\mathbf{S}\in\mathbb{C}^{N_r\times N_p}$ and $\mathrm{vec}\left\{\mathbf{S}\right\}=\mathbf{s}$. We first compute $\mathbf{S}'\triangleq\mathbf{S}\mathbf{F}_d^*\in\mathbb{C}^{N_r\times F_\tau N_p}$ and then $\mathbf{V}^H\mathbf{S}'$. Both of the above two calculations can be implemented through inverse FFT (IFFT). Then, $\mathbf{E}^T\tilde{\mathbf{A}}^H\mathbf{s}$ is equivalent to extracting the elements from $\tilde{\mathbf{A}}^H\mathbf{s}$ with the indexes determined by $\mathcal{P}$. The complexity of the efficient implementation of $\tilde{\mathbf{A}}^H\mathbf{s}$ is $\mathcal{O}\left(C\right)$, too. As for the calculation of $\mathbf{A}^H\mathbf{y}$, since it is the same at each iteration, we only need to calculate it once. The calculation of $\mathbf{A}^H\mathbf{y}$ and the corresponding complexities are the same as that of $\mathbf{A}^H\mathbf{s}$ in one iteration.

We compare the simplified IGA (noted as S-IGA in the simulation) with the following algorithms. **AMP**: Approximate message passing algorithm proposed in [10]. **IGA**: The original information geometry approach proposed in [6]. **MMSE**: The MMSE estimation of the beam domain channels based on (8a). The computational complexities of different algorithms are summarized in Table I.

TABLE I
COMPUTATIONAL COMPLEXITIES OF ALGORITHMS

| Algorithm | Complexity |
|---|---|
| MMSE | $\mathcal{O}\left(M^3+M^2N\right)$ |
| AMP/IGA (per iteration) | $\mathcal{O}\left(NM\right)$ |
| simplified IGA (per iteration) | $\mathcal{O}\left(C\right)$ |

## V. Simulation Results

In this section, we provide simulation results to illustrate the complexity and performance of the proposed simplified IGA for massive MIMO-OFDM channel estimation. The widely adopted QuaDRiGa [11] is used to generate the SF domain channel for each user. The simulation scenario is set to "3GPP_38.901_UMa_NLOS", and main parameters for the simulations are summarized in Table II. Note that channels under both LOS and NLOS conditions can be modeled with SF beam based channel model. We locate the BS at $(0,0,25)$ and randomly generate the users in a $120^\circ$ sector with radius $r=200$m around $(0,0,1.5)$. The SF domain channels are normalized as $\mathbb{E}\left\{\|\mathbf{G}_k\|_F^2\right\}=N_rN_p$. The SNR is set as $\mathrm{SNR}=\frac{1}{\sigma_z^2}$. We set the fine factors to $F_v=F_h=F_\tau=2$ in all simulations, which can achieve significant performance gain compared with the case with $F_v=F_h=F_\tau=1$ as shown in [6]. The algorithm proposed in [12] is adopted to obtain the channel power matrices $\boldsymbol{\Omega}_k, k\in\mathcal{Z}_K^+$. And the number of total non-zero elements in $\{\boldsymbol{\Omega}_k\}_{k=1}^{48}$ is calculated as $M=36542$, which is smaller than that of the observations $N=N_{r,v}\times N_{r,h}\times N_p=46080$. We use the normalized mean-squared error (NMSE) as the performance metric for the channel estimation,

TABLE II
PARAMETER SETTINGS OF THE QUADRIGA

| Parameter | Value |
|---|---|
| Number of BS antenna $N_{r,v}\times N_{r,h}$ | $8\times 16$ |
| UT number $K$ | 48 |
| Center frequency $f_c$ | 4.8GHz |
| Number of training subcarriers $N_p$ | 360 |
| Subcarrier spacing $\Delta_f$ | 15kHz |
| Number of subcarriers $N_c$ | 2048 |
| CP length $N_{\mathrm{g}}$ | 144 |

$$\mathrm{NMSE}=\frac{1}{KN_{sam}}\sum_{k=1}^{K}\sum_{n=1}^{N_{sam}}\frac{\|\mathbf{G}_k^{(n)}-\hat{\mathbf{G}}_k^{(n)}\|_F^2}{\|\mathbf{G}_k^{(n)}\|_F^2}, \tag{30}$$

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.

The computational complexity of different algorithms are plotted in Fig. 1 under different numbers of users. The number of iterations for AMP, IGA and simplified IGA are set to 200. We can find that the complexity of MMSE is the highest since

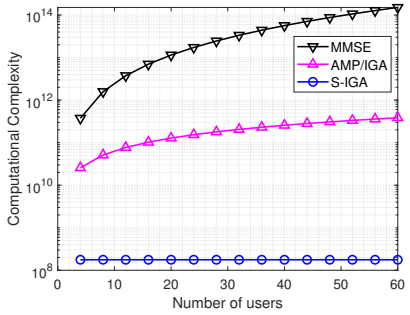


Fig. 1. The complexities of different algorithms versus the number of users.

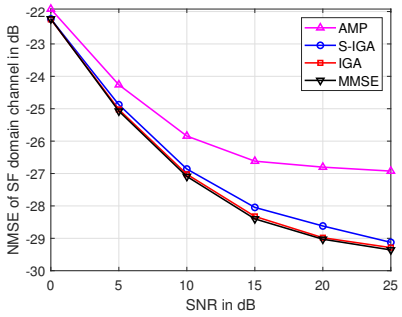


Fig. 2. NMSE performance of simplified IGA compared with AMP, IGA and MMSE.

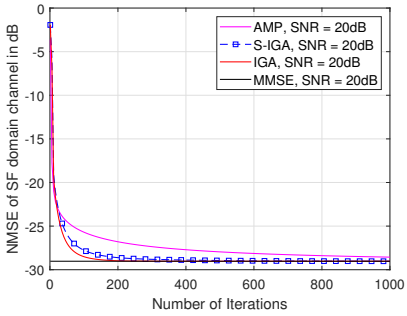


Fig. 3. Convergence performance of simplified IGA compared with AMP and IGA at SNR = 20 dB.

a matrix-inversion is involved. On the other hand, owing to the utilization of the structure of $\mathbf{A}$ and FFT, the complexity of the simplified IGA (S-IGA) is the lowest among all the algorithms.

Fig. 2 shows the NMSE performance of simplified IGA channel estimation compared with AMP, IGA and MMSE. The iteration number of simplified IGA, AMP and IGA is set as 200. We can find that IGA can obtain almost the same NMSE performance as the MMSE estimation at all SNRs. The performance of simplified IGA can approach that of the MMSE estimation with a small gap. The SNR gain of the simplified IGA compared to AMP is about 3dB when the NMSE performance is $-26$dB.

Fig. 3 illustrate the convergence performance of simplified IGA compared with AMP and IGA, where the SNR is set as 20dB. We can find that the simplified IGA converges in about 400 iterations, IGA requires about 200 iterations to converge, while AMP takes more than 1000 iterations to converge. It can also be found that simplified IGA and IGA show similar convergence behavior. This can be attributed to the similarity of the processes of those two approaches. On the other hand, the computational complexity of simplified IGA is much lower than that of IGA. Compared with AMP, simplified IGA converges with a faster rate. The simplified IGA along with the original IGA are developed based on the structure of the *a posteriori* distribution $p(\mathbf{h}|\mathbf{y})$ within the framework of information geometry theory. As a result, we are able to resolve the statistical inference problem from an intrinsic and general standpoint. This might be a significant factor in the improved convergence behavior of the simplified IGA for massive MIMO-OFDM channel estimation.

## VI. Conclusion

In this paper, we have investigated the information geometry approach for channel estimation in massive MIMO-OFDM systems. The original IGA is first revisited. By using the constant magnitude property of the measurement matrix entries and the asymptotic analysis, we reveal that the FONPS of $p_1, p_2, \cdots, p_N$ on the AMs are asymptotically equal at the fixed point of IGA, and the SONPs of $p_1, p_2, \cdots, p_N$ on the AMs are equal to each other at each iteration of IGA. Based on these results, we simplify the iteration of IGA by using the common NP to replace the original NPs of $p_1, p_2, \cdots, p_N$ on the AMs and propose a simplified IGA. In the simplified IGA, the common NP is the only parameter involved for the iteration. Then, we show that at the fixed point, the *a posteriori* mean obtained by the simplified IGA is asymptotically optimal. An FFT-based fast implementation of the simplified IGA is also provided. Simulation results verify that the proposed simplified IGA can obtain near optimal channel estimation performance with much less computational complexity compared with the existing algorithms.

## References


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

[2] E. 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.

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

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

[5] 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.

[6] 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.

[7] 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.

[8] X. Liu, W. Wang, X. Song, X. Q. Gao, and G. Fettweis, "Sparse channel estimation via hierarchical hybrid message passing for massive mimo-ofdm systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 11, pp. 7118–7134, Nov 2021.

[9] S. M. Kay, *Fundamentals of Statistical Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[10] 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.

[11] 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.

[12] A.-A. Lu, Y. Chen, and X. Q. Gao, "2D beam domain statistical CSI estimation for massive MIMO uplink," *arXiv preprint arXiv:2207.04695*, 2022.