Convergence Condition of Simplified Information Geometry Approach for Massive MIMO-OFDM Channel Estimation

Mingrui Fan*†, Jiyuan Yang*†, Yan Chen*†, An-An Lu*†, 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
{220220758, jyyang, 213160372, aalu, xqgao}@seu.edu.cn, xxia@ee.udel.edu, Dirk.Slock@eurecom.fr

***Abstract*—In this paper, we prove the convergence of the simplified information geometry approach (SIGA), which was proposed for massive MIMO-OFDM channel estimation. For a general Bayesian inference problem, we first show that the iteration of the common second-order natural parameter (SONP) is separated from that of the common first-order natural parameter (FONP). Hence, the convergence of the common SONP can be checked independently. We show that with the initialization satisfying a specific but large range, the common SONP is convergent regardless of the value of the damping factor. For the common FONP, we establish a sufficient condition of its convergence and prove that the convergence of the common FONP relies on the spectral radius of a particular matrix related to the damping factor. We give the range of the damping factor that guarantees the convergence in the worst case. Further, we determine the range of the damping factor for massive MIMO-OFDM channel estimation by using the specific properties of the measurement matrices. Simulation results are provided to confirm the theoretical results.**

***Index Terms*—Convergence, Bayesian inference, information geometry, damping factor.**

I. Introduction

Numerous problems in signal processing eventually come to the issue of computing marginal probability density functions (PDF) from a high dimensional joint PDF. In general, the computation of direct marginalization could be unaffordable since complicated integration calculations are involved. In the past decades, many works have been devoted to providing an efficient way to compute the (approximate) marginal PDFs under various cases. [1] proposes the Gaussian belief propagation (BP) and shows that Gaussian BP is able to compute true marginal mean. In [2], the powerful approximate message passing (AMP) algorithm is proposed. It has been shown that AMP with Bayes-optimal denoiser can be treated as an exact approximation of loopy BP in the large system limit. When the underlying factor graph is a tree, the expectation propagation (EP) in [3] is convergent and can exactly achieve the Bayes-optimal performance.

Recently, we have introduced the information geometry approach (IGA) to the massive multiple-input multiple-output (MIMO) channel estimation [4]. We also improve the stability of IGA by introducing the damping factor and show that IGA can obtain accurate a posteriori mean at its fixed point. On the basis of IGA, two new results of IGA are revealed when the constant magnitude pilots are adopted. Based on these new results, we propose a simplified IGA (SIGA) in [5], [6]. Although proposed for the massive MIMO-OFDM channel estimation, SIGA itself could serve as a generic Bayesian inference method which is suitable for Gaussian priors and constant magnitude measurement matrix. It has been shown that at the fixed point, the a posteriori mean obtained by SIGA is asymptotically optimal. Furthermore, SIGA can be implemented efficiently when the measurement matrix has special structure. For example, in massive MIMO channel estimation, the measurement matrix can be constructed by partial DFT matrices, SIGA can be then implemented by fast Fourier transform (FFT), which significantly reduces its computational complexity.

On the other hand, one standard limitation of Bayesian inference approaches is that they work well only when the iteration converges. However, for a variety of problems, a unified way of proving the convergence of Bayesian inference approaches has not yet been found. Thus, the convergence needs to be proved for individual approaches as well as individual problems. So far, there are only a few methods whose convergence has been relatively well revealed. One sufficient convergence condition named the walk-summability is proposed in [1] for Gaussian BP. Then, several extended conditions are proposed by the authors in [7], [8]. In [9], the convergence of orthogonal/vector AMP is proved based on the idea of "convergence in principle". In [10], the convergence of generalized AMP is proved in the special case with Gaussian priors. Similar to most other Bayesian inference approaches,

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.

SIGA suffers from divergence. However, by adding damping factor, its convergence can be significantly improved. This is an interesting observation, since in many iterative Bayesian inference approaches, such as, e.g., AMP, damped updating likewise plays an important role in convergence.

In this paper, we give a theoretical analysis of the convergence of SIGA. The role of the damping factor in the iteration will also be clarified. We first prove the convergence of the common second-order natural parameter (SONP) in SIGA for a general Bayesian inference problem with the measurement matrix of constant magnitude property. It is then proved that given the initialization satisfying a specific and large range, the common SONP is guaranteed to converge regardless of the value of damping factor. Then, we show that the iteration of the common SONP is separated from that of the common first-order natural parameter (FONP) and establish a sufficient condition of the convergence of the common FONP in SIGA where it depends on the spectral radius of the iterating system matrix. On this basis, we give the range of the damping factor that guarantees the convergence of the common FONP in the worst case. We then apply the above general convergence results to the case of massive MIMO-OFDM channel estimation and determine a range of the damping factor that guarantees the convergence of the common FONP in SIGA from the properties of the measurement matrices.

II. Convergence of SIGA

In this section, we first introduce SIGA proposed in [5], more details can be found therein. Then, through re-expressing its iterations, we prove its convergence.

A. Review of SIGA

Considering the following Bayesian inference problem:

$$\mathbf{y} = \mathbf{A}\mathbf{h} + \mathbf{z}, \tag{1}$$

where $\mathbf{y} \in \mathbb{C}^N$ is the observed vector, $\mathbf{A} \in \mathbb{C}^{N\times M}$ is a deterministic measurement matrix, and entries of $\mathbf{A}$ have constant magnitude, $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}, \mathbf{D})$ is the M-dimensional complex Gaussian random vector to be estimated, its covariance matrix $\mathbf{D}$ is deterministic, known, positive definite and diagonal, $\mathbf{z} \sim \mathcal{CN}\left(\mathbf{0}, \sigma_z^2\mathbf{I}\right)$ is the N-dimensional noise vector where σ_z^2 is known, and $\mathbf{h}$ and $\mathbf{z}$ are independent with each other. Without loss of generality, assuming that the components of $\mathbf{A}$ have unit magnitude. From (1), the *posteriori* PDF $p(\mathbf{h}|\mathbf{y})$ is Gaussian with a given $\mathbf{y}$ and thus its *a posteriori* mean $\tilde{\boldsymbol{\mu}}$ and covariance matrix $\tilde{\boldsymbol{\Sigma}}$ are given by [11]

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

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

where the computational complexity of (2) is $\mathcal{O}\left(M^3 + M^2N\right)$ and it is unaffordable when M and N are large.

The aim of SIGA is calculating the approximations of the marginals of the *a posteriori* distribution, i.e., the approximations of $p_i(h_i|\mathbf{y}), i \in \mathcal{Z}_M^+$, with a lower computational complexity than (2). Then, the *a posteriori* mean and variance can be obtained. We begin with some essential definitions in SIGA. Given $\mathbf{a}, \mathbf{b} \in \mathbb{C}^M$, define a vector function as $\mathbf{f}(\mathbf{a}, \mathbf{b}) \triangleq \left[\mathbf{a}^T, \mathbf{b}^T\right]^T \in \mathbb{C}^{2M}$, the operator $\circ$ as $\mathbf{a} \circ \mathbf{b} \triangleq \frac{1}{2}\left(\mathbf{b}^H\mathbf{a} + \mathbf{a}^H\mathbf{b}\right)$ and $\mathbf{a} < \mathbf{b}$ means that each component in vector $\mathbf{a}$ is smaller than the component in the corresponding position in vector $\mathbf{b}$. Let $\mathbf{d} = \mathbf{f}\left(\mathbf{0}, \mathrm{diag}\left\{-\mathbf{D}^{-1}\right\}\right) \in \mathbb{C}^{2M}$ and $\mathbf{t} = \mathbf{f}(\mathbf{h}, \mathbf{h} \odot \mathbf{h}^*) \in \mathbb{C}^{2M}$ where $\mathrm{diag}\left\{-\mathbf{D}^{-1}\right\}$ denotes a vector consisting of the diagonal elements of $-\mathbf{D}^{-1}$. Then, $p(\mathbf{h}|\mathbf{y})$ can be expressed as [4], [5]

$$p(\mathbf{h}|\mathbf{y}) = \exp\left\{\mathbf{d} \circ \mathbf{t} + \sum_{n=1}^{N} c_n(\mathbf{h}) - \psi_q\right\}, \tag{3a}$$

$$c_n(\mathbf{h}) = \frac{1}{\sigma_z^2}\left(-\mathbf{h}^H\boldsymbol{\gamma}_n\boldsymbol{\gamma}_n^H\mathbf{h} + y_n\mathbf{h}^H\boldsymbol{\gamma}_n + y_n^*\boldsymbol{\gamma}_n^H\mathbf{h}\right), \tag{3b}$$

where $\boldsymbol{\gamma}_n$ is the n-th column of $\mathbf{A}^H$, y_n is the n-th element of $\mathbf{y}$, and ψ_q is the normalization factor. In (3a), $\mathbf{t}$ only contains the statistics of single random variables, i.e., h_i and $|h_i|^2, i \in \mathcal{Z}_M^+$, and all the interactions (cross terms), $h_ih_j^*, i \neq j$, are contained in $c_n(\mathbf{h}), n \in \mathcal{Z}_N^+$. SIGA is to approximate each $c_n(\mathbf{h})$ as $\boldsymbol{\xi}_n \circ \mathbf{t}$ in an iterative manner, where $\boldsymbol{\xi}_n \in \mathbb{C}^M$ is referred as to the approximation item. In this way, we have

$$p(\mathbf{h}|\mathbf{y}) \approx p_0(\mathbf{h}; \boldsymbol{\vartheta}_0) = \exp\left\{(\mathbf{d} + \boldsymbol{\vartheta}_0) \circ \mathbf{t} - \psi_0\right\}, \tag{4}$$

where $\boldsymbol{\vartheta}_0 = \sum_{n=1}^{N}\boldsymbol{\xi}_n \in \mathbb{C}^{2M}$ and ψ_0 is the normalization factor. The marginals of $p_0(\mathbf{h}; \boldsymbol{\vartheta}_0)$ can be calculated directly since it contains no cross terms between random variables. To obtain $\boldsymbol{\xi}_n, n \in \mathcal{Z}_N^+$, and $\boldsymbol{\vartheta}_0$, SIGA constructs the following two types of manifolds: the objective manifold (OBM) and the auxiliary manifold (AM). The OBM containing any potential $p_0(\mathbf{h}; \boldsymbol{\vartheta}_0)$ is defined as

$$\mathcal{M}_0 = \left\{p_0(\mathbf{h}; \boldsymbol{\vartheta}_0) = \exp\left\{(\mathbf{d} + \boldsymbol{\vartheta}_0) \circ \mathbf{t} - \psi_0(\boldsymbol{\vartheta}_0)\right\}\right\}, \tag{5}$$

where $\boldsymbol{\vartheta}_0 = \mathbf{f}(\boldsymbol{\theta}_0, \boldsymbol{\nu}_0)$, $\boldsymbol{\theta}_0 \in \mathbb{C}^M$ and $\boldsymbol{\nu}_0 \in \mathbb{R}^M$ are referred to the first-order natural parameter (FONP) and the second-order natural parameter (SONP) of p_0, respectively, and $\psi_0(\boldsymbol{\vartheta}_0)$ is the normalization factor. To obtain all approximation items $\boldsymbol{\xi}_n \in \mathbb{C}^{2M}$, N auxiliary manifolds (AMs) are defined, where the n-th AM is given by

$$\mathcal{M}_n = \left\{p_n(\mathbf{h}; \boldsymbol{\vartheta})\right\}, n \in \mathcal{Z}_N^+, \tag{6a}$$

$$p_n(\mathbf{h}; \boldsymbol{\vartheta}) = \exp\left\{(\mathbf{d} + \boldsymbol{\vartheta}) \circ \mathbf{t} + c_n(\mathbf{h}) - \psi_n(\boldsymbol{\vartheta})\right\}, \tag{6b}$$

where $\boldsymbol{\vartheta} = \mathbf{f}(\boldsymbol{\theta}, \boldsymbol{\nu})$, $\boldsymbol{\theta} \in \mathbb{C}^M$ and $\boldsymbol{\nu} \in \mathbb{R}^M$ are referred to the common FONP and the common SONP of p_n, respectively, and $\psi_n(\boldsymbol{\vartheta})$ is the normalization factor.

We now introduce the iteration of the SIGA. Mathematically, the iteration of SIGA can be summarized as the iterative calculation of $\boldsymbol{\vartheta}$ until convergence. When $\boldsymbol{\vartheta}$ is converged, we calculate $\boldsymbol{\vartheta}_0$ as $\boldsymbol{\vartheta}_0 = \frac{N}{N-1}\boldsymbol{\vartheta}$. Then, $p_0(\mathbf{h}; \boldsymbol{\vartheta}_0)$ in (5) is referred as to the approximation of the product of the marginals of the *a posreriori* distributions, i.e., $\prod_{i=1}^{M} p_i(h_i|\mathbf{y})$. And the *a posteriori* mean and variance are given by $\boldsymbol{\mu}_0(\boldsymbol{\vartheta}_0)$ and the diagonal of $\boldsymbol{\Sigma}_0(\boldsymbol{\vartheta}_0)$, respectively, where

$$\boldsymbol{\mu}_0(\boldsymbol{\vartheta}_0) = \frac{1}{2}\boldsymbol{\Sigma}_0(\boldsymbol{\vartheta}_0)\boldsymbol{\theta}_0, \tag{7a}$$

$$\mathbf{\Sigma}_0\left(\boldsymbol{\vartheta}_0\right)=\left(\mathbf{D}^{-1}-\mathrm{Diag}\left\{\boldsymbol{\nu}_0\right\}\right)^{-1}, \tag{7b}$$

where $\mathrm{Diag}\{\boldsymbol{\nu}_0\}$ denotes the diagonal matrix with $\boldsymbol{\nu}_0$ along its main diagonal. Specifically, given $\boldsymbol{\vartheta}(t)=\mathbf{f}(\boldsymbol{\theta}(t),\boldsymbol{\nu}(t))$ at the t-th iteration, $\boldsymbol{\vartheta}(t+1)=\mathbf{f}(\boldsymbol{\theta}(t+1),\boldsymbol{\nu}(t+1))$ is then calculated as (9) [5], where $0<d\leq 1$ is the damping factor, and $\mathbf{\Lambda}(\boldsymbol{\nu}(t))$ and $\beta(\boldsymbol{\nu}(t))$ are given by

$$\mathbf{\Lambda}(\boldsymbol{\nu}(t))=\left(\mathbf{D}^{-1}-\mathrm{Diag}\left\{\boldsymbol{\nu}(t)\right\}\right)^{-1}, \tag{8a}$$

$$\beta(\boldsymbol{\nu}(t))=\sigma_z^2+\mathrm{tr}\left\{\mathbf{\Lambda}(\boldsymbol{\nu}(t))\right\}. \tag{8b}$$

Briefly, $\boldsymbol{\vartheta}(t+1)$ is obtained by the m-projections of $p_n(\mathbf{h};\boldsymbol{\vartheta}(t)), n\in\mathcal{Z}_N^+$, onto the OBM. The detailed calculation can be found in the Sec. IV of [5]. We summarize the process of SIGA in Algorithm 1.

Algorithm 1: SIGA

Input: The covariance $\mathbf{D}$ of the a 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 damping d, where $0<d\leq 1$, initialize the common NP as $\boldsymbol{\vartheta}(0)=\mathbf{f}(\boldsymbol{\theta}(0),\boldsymbol{\nu}(0))$ and ensure $\tilde{\mathbf{g}}_{\min}\leq\boldsymbol{\nu}(0)\leq\mathbf{0}$;

repeat

1. Update $\boldsymbol{\vartheta}=\mathbf{f}(\boldsymbol{\theta},\boldsymbol{\nu})$ as (9), where $\mathbf{\Lambda}(\boldsymbol{\nu}(t))$ and $\beta(\boldsymbol{\nu}(t))$ are given by (8a) and (8b);
2. $t=t+1$;

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

Output: Calculate the NP of $p_0(\mathbf{h};\boldsymbol{\vartheta}_0)$ as $\boldsymbol{\vartheta}_0=\frac{N}{N-1}\boldsymbol{\vartheta}(t)$.The mean and variance of the approximate marginal, $p_i(h_i|\mathbf{y}),\ i\in\mathcal{Z}_M^+$, are given by the i-th component of $\boldsymbol{\mu}_0$ and $\mathrm{diag}\{\mathbf{\Sigma}_0\}$, respectively, where $\boldsymbol{\mu}_0$ and $\mathbf{\Sigma}_0$ are calculated by (7).

B. Convergence

From (9), we can see that $\boldsymbol{\nu}(t+1)$ only depends on $\boldsymbol{\nu}(t)$ and does not depend on $\boldsymbol{\theta}(t)$, while $\boldsymbol{\theta}(t+1)$ depends on both $\boldsymbol{\theta}(t)$ and $\boldsymbol{\nu}(t)$. This shows that the iterating system of $\boldsymbol{\nu}$ is separated from that of $\boldsymbol{\theta}$, and hence, the convergence of $\boldsymbol{\nu}(t)$ can be checked individually. Thus, (9b) can be rewritten as

$$\boldsymbol{\nu}(t+1)=\tilde{\mathbf{g}}(\boldsymbol{\nu}(t))\triangleq d\mathbf{g}(\boldsymbol{\nu}(t))+(1-d)\boldsymbol{\nu}(t), \tag{10a}$$

$$\mathbf{g}(\boldsymbol{\nu}(t))=-(N-1)\,\mathrm{diag}\left\{\left(\beta(\boldsymbol{\nu}(t))\mathbf{I}-\mathbf{\Lambda}(\boldsymbol{\nu}(t))\right)^{-1}\right\}, \tag{10b}$$

where $\tilde{\mathbf{g}},\mathbf{g}:\mathbb{R}^M\to\mathbb{R}^M$ are vector functions about $\boldsymbol{\nu}(t)$. Then, we first present the following lemma about $\tilde{\mathbf{g}}$.

Lemma 1. *Given $\boldsymbol{\nu}\leq\mathbf{0}$, $\tilde{\mathbf{g}}(\boldsymbol{\nu})$ satisfies the following two properties.*
1. Monotonicity: If $\boldsymbol{\nu}<\boldsymbol{\nu}'\leq\mathbf{0}$, then $\tilde{\mathbf{g}}(\boldsymbol{\nu})<\tilde{\mathbf{g}}(\boldsymbol{\nu}')\leq\mathbf{0}$.
2. Scalability: Given a positive constant $0<\alpha<1$, we have $\tilde{\mathbf{g}}(\alpha\boldsymbol{\nu})<\alpha\tilde{\mathbf{g}}(\boldsymbol{\nu})$.
Moreover, if $\tilde{\mathbf{g}}_{min}\leq\boldsymbol{\nu}\leq\mathbf{0}$ with $\tilde{\mathbf{g}}_{min}\triangleq-\frac{N-1}{\sigma_z^2}\mathbf{1}\in\mathbb{R}^M$, we have $\tilde{\mathbf{g}}_{min}<\tilde{\mathbf{g}}(\boldsymbol{\nu})<\mathbf{0}$.

The detailed proofs of all the results in this paper are referred to [12]. Based on Lemma 1, the convergence of SONP $\boldsymbol{\nu}(t)$ can be proved in following theorem.

Theorem 1. *Given initialization $\boldsymbol{\nu}(0)$ with $\tilde{\mathbf{g}}_{min}\leq\boldsymbol{\nu}(0)\leq\mathbf{0}$, the sequence $\boldsymbol{\nu}(t+1)=\tilde{\mathbf{g}}(\boldsymbol{\nu}(t))$ converges to a finite fixed point $\boldsymbol{\nu}^\star$, where $\tilde{\mathbf{g}}_{min}<\boldsymbol{\nu}^\star<\mathbf{0}$.*

From Theorem 1, we can find that $\boldsymbol{\nu}(t)$ converges to a finite fixed point as long as the initialization satisfies $\tilde{\mathbf{g}}_{\min}\leq\boldsymbol{\nu}(0)\leq\mathbf{0}$, and this range can be quite large. Furthermore, the convergence of $\boldsymbol{\nu}(t)$ is not related to the choice of damping factor. Since the iteration of $\boldsymbol{\nu}(t)$ is separated from $\boldsymbol{\theta}(t)$, we can analyze the convergence of $\boldsymbol{\theta}(t)$ assuming that the iteration of $\boldsymbol{\nu}(t)$ converges to the fixed point $\boldsymbol{\nu}^\star$. Then, the intermideate variable in (8) can be redefined as

$$\mathbf{\Lambda}^\star=\left(\mathbf{D}^{-1}-\mathrm{Diag}\left\{\boldsymbol{\nu}^\star\right\}\right)^{-1}, \tag{11a}$$

$$\beta^\star=\sigma_z^2+\mathrm{tr}\left\{\mathbf{\Lambda}^\star\right\}. \tag{11b}$$

The above diagonal matrix $\mathbf{\Lambda}^\star$ is positive definite and $\beta^\star$ is positive from Theorem 1. Given $\boldsymbol{\theta}(t)$ at the t-th iteration, then $\boldsymbol{\theta}(t+1)$ in (9a) can be rewritten as

$$\boldsymbol{\theta}(t+1)=\tilde{\mathbf{B}}(\boldsymbol{\nu}(t))\,\boldsymbol{\theta}(t)+\mathbf{b}(\boldsymbol{\nu}(t)), \tag{12a}$$

$$\tilde{\mathbf{B}}(\boldsymbol{\nu}(t))\triangleq d\mathbf{B}(\boldsymbol{\nu}(t))+(1-d)\,\mathbf{I}, \tag{12b}$$

$$\begin{aligned}\mathbf{B}(\boldsymbol{\nu}(t))=&\frac{N-1}{\beta(\boldsymbol{\nu}(t))}\left(\mathbf{I}-\frac{1}{\beta(\boldsymbol{\nu}(t))}\mathbf{\Lambda}(\boldsymbol{\nu}(t))\right)^{-1}\\&\times\left(\mathbf{I}-\frac{1}{N}\mathbf{A}^H\mathbf{A}\right)\mathbf{\Lambda}(\boldsymbol{\nu}(t)),\end{aligned} \tag{12c}$$

$$\mathbf{b}(\boldsymbol{\nu}(t))=\frac{2d(N-1)}{N\beta(\boldsymbol{\nu}(t))}\left(\mathbf{I}-\frac{\mathbf{\Lambda}(\boldsymbol{\nu}(t))}{\beta(\boldsymbol{\nu}(t))}\right)^{-1}\mathbf{A}^H\mathbf{y}, \tag{12d}$$

where $\tilde{\mathbf{B}}$ and $\mathbf{B}$ are two matrix functions with $\boldsymbol{\nu}(t)$ being the only variable, i.e., $\tilde{\mathbf{B}},\mathbf{B}:\mathbb{R}^M\to\mathbb{C}^{M\times M}$, and $\mathbf{b}$ is a vector function with $\boldsymbol{\nu}(t)$ being the only variable, i.e., $\mathbf{b}:\mathbb{R}^M\to\mathbb{C}^M$, and $\tilde{\mathbf{g}},\mathbf{g}:\mathbb{R}^M\to\mathbb{R}^M$. Let $t\to\infty$ and define

$$\tilde{\mathbf{B}}^\star=\tilde{\mathbf{B}}(\boldsymbol{\nu}^\star)=d\mathbf{B}^\star+(1-d)\,\mathbf{I}, \tag{13}$$

where $\mathbf{B}^\star=\mathbf{B}(\boldsymbol{\nu}^\star)$ and $\mathbf{b}^\star=\mathbf{b}(\boldsymbol{\nu}^\star)$. From the definition, $\tilde{\mathbf{B}}^\star$ is determined by the fixed point of the common SONP $\boldsymbol{\nu}^\star$ and matrix $\mathbf{A}$, which does not vary with iterations. To avoid any ambiguity, the iterating system matrix refers to $\tilde{\mathbf{B}}^\star$ in the rest of the paper, since the convergence condition for the iterating system of $\boldsymbol{\theta}(t)$ depends only on the spectral radius of $\tilde{\mathbf{B}}^\star$.

Lemma 2. *Given a finite initialization $\boldsymbol{\theta}(0)\in\mathbb{C}^{M\times 1}$ and $\boldsymbol{\nu}(0)$ with $\tilde{\mathbf{g}}_{min}\leq\boldsymbol{\nu}(0)\leq\mathbf{0}$. Then, $\boldsymbol{\theta}(t)$ in (12) converges to its fixed point if the spectral radius of $\tilde{\mathbf{B}}^\star$ is less than 1, i.e., $\rho\left(\tilde{\mathbf{B}}^\star\right)<1$, with $\rho\left(\tilde{\mathbf{B}}^\star\right)=\max\left\{|\lambda|:\lambda\text{ is an eigenvalue of }\tilde{\mathbf{B}}^\star\right\}$.*

Next, we investigate the convergence of $\boldsymbol{\theta}$ by analyzing the eigenvalues of $\tilde{\mathbf{B}}^\star$. At the fixed point where $\boldsymbol{\nu}^\star=\boldsymbol{\nu}(t+1)=\boldsymbol{\nu}(t)$ in (9b), we can reformulate $\mathbf{B}^\star$ as

$$\mathbf{B}^\star=\left(\mathbf{I}-\frac{1}{N}\mathbf{D}^{-1}\mathbf{\Lambda}^\star\right)\left(N\mathbf{I}-\mathbf{A}^H\mathbf{A}\right)\left(\frac{1}{\beta^\star}\mathbf{\Lambda}^\star\right). \tag{14}$$

$$\boldsymbol{\theta}(t+1)=\frac{d(N-1)}{N}\left(\mathbf{I}-\frac{1}{\beta(\boldsymbol{\nu}(t))}\boldsymbol{\Lambda}(\boldsymbol{\nu}(t))\right)^{-1}\left[\frac{1}{\beta(\boldsymbol{\nu}(t))}\mathbf{A}^H(2\mathbf{y}-\mathbf{A}\boldsymbol{\Lambda}(\boldsymbol{\nu}(t))\boldsymbol{\theta}(t))+N\boldsymbol{\theta}(t)\right]+(1-dN)\boldsymbol{\theta}(t), \tag{9a}$$

$$\boldsymbol{\nu}(t+1)=d(N-1)\mathrm{diag}\left\{\mathbf{D}^{-1}-\left(\boldsymbol{\Lambda}(\boldsymbol{\nu}(t))-\frac{1}{\beta(\boldsymbol{\nu}(t))}\boldsymbol{\Lambda}^2(\boldsymbol{\nu}(t))\right)^{-1}\right\}+(1-dN)\boldsymbol{\nu}(t). \tag{9b}$$

where $\mathbf{B}^\star$ is the product of three Hermitian matrices. By [13, Exercise below Theorem 5.6.9], we have

$$\rho(\mathbf{B}^\star)\le\rho\left(\mathbf{I}-\frac{\mathbf{D}^{-1}\boldsymbol{\Lambda}^\star}{N}\right)\rho\left(N\mathbf{I}-\mathbf{A}^H\mathbf{A}\right)\rho\left(\frac{\boldsymbol{\Lambda}^\star}{\beta^\star}\right). \tag{15}$$

Then, we first present some properties about the matrices inside the right hand side term of (15).

Lemma 3. *The matrices inside the term of the right hand side of* (15) *satisfy following properties:*
1. The spectral radius of $\mathbf{I}-\frac{1}{N}\mathbf{D}^{-1}\boldsymbol{\Lambda}^\star$ *is less than* 1.
2. The spectral radius of $\boldsymbol{\Lambda}^*$ *satisfies* $\rho(\boldsymbol{\Lambda}^\star)<\frac{\beta^\star}{N}$.
3. The spectral radius of $N\mathbf{I}-\mathbf{A}^H\mathbf{A}$ *satisfies*

$$\rho\left(N\mathbf{I}-\mathbf{A}^H\mathbf{A}\right)\le NM-N. \tag{16}$$

Based on Lemma 3, we have the following Lemma about the eigenvalues of $\mathbf{B}^\star$.

Lemma 4. *Denote the eigenvalues of* $\mathbf{B}^\star$ *as* $\lambda_{B,i}, i\in\mathcal{Z}_M^+$. *Then,* $\{\lambda_{B,i}\}_{i=1}^M$ *are all real and* $-\frac{\rho(N\mathbf{I}-\mathbf{A}^H\mathbf{A})}{N}<\lambda_{B,i}<1$.

Denote the eigenvalues of $\tilde{\mathbf{B}}^\star$ as $\tilde{\lambda}_i, i\in\mathcal{Z}_M^+$ and we have $\tilde{\lambda}_i=d\lambda_{B,i}+1-d, i\in\mathcal{Z}_M^+$ from (12b). Then, the eigenvalues of $\tilde{\mathbf{B}}^\star$ are all real and satisfy

$$1-d\left(1+\frac{\rho\left(N\mathbf{I}-\mathbf{A}^H\mathbf{A}\right)}{N}\right)<\tilde{\lambda}_i<1. \tag{17}$$

Combining Lemma 2 and (17), we have the following Theorem

Theorem 2. *Given a finite initialization* $\boldsymbol{\theta}(0)\in\mathbb{C}^{M\times1}$ *and* $\boldsymbol{\nu}(0)$ *with* $\tilde{\mathbf{g}}_{min}\le\boldsymbol{\nu}(0)\le\mathbf{0}$. *Then,* $\boldsymbol{\theta}(t)$ *in* (12) *converges to its fixed point if the damping factor satisfies*

$$d<\frac{2}{1+\frac{\rho(N\mathbf{I}-\mathbf{A}^H\mathbf{A})}{N}}. \tag{18}$$

Combining (16) *and* (18)*, it can be shown that in the worst case where* $\mathrm{rank}(\mathbf{A})=1$ *and* $\rho\left(N\mathbf{I}-\mathbf{A}^H\mathbf{A}\right)=NM-N$*.,* $d<\frac{2}{M}$ *can ensure the convergence of* $\boldsymbol{\theta}(t)$.

From Theorem 2, we can find that SIGA will always converge with a sufficiently small damping factor and the range of d is mainly determined by $\rho\left(N\mathbf{I}-\mathbf{A}^H\mathbf{A}\right)$. The spectral radius $\rho\left(N\mathbf{I}-\mathbf{A}^H\mathbf{A}\right)$ depends on the measurement matrix $\mathbf{A}$ and even in the worst case where $\mathrm{rank}(\mathbf{A})=1$, SIGA converges if $d<\frac{2}{M}$. The range of $\rho\left(N\mathbf{I}-\mathbf{A}^H\mathbf{A}\right)$ and the corresponding range of damping factor in massive MIMO-OFDM channel estimation will be discussed in the next section.

III. Application to Massive MIMO-OFDM Channel Estimation

In this section, we will discuss the range of $\rho\left(N\mathbf{I}-\mathbf{A}^H\mathbf{A}\right)$ in massive MIMO-OFDM channel estimation, where the range of damping factor d can be expanded. Consider the following uplink massive MIMO-OFDM channel estimation problem: A base station equipped with $N_r=N_{rv}\times N_{rh}$ uniform planar array (UPA) serves K single antenna users, where N_{rv} and N_{rh} are the numbers of the antennas at each vertical column and horizontal row, respectively. The number of subcarriers, efficient subcarriers and cyclic prefix (CP) length of OFDM modulation are N_c ,N_p and N_g, respectively. Let $\mathbf{Y}\in\mathbb{C}^{N_r\times N_p}$ and $\mathbf{Z}\in\mathbb{C}^{N_r\times N_p}$ be the space-frequency domain received signal and noise, respectively, then the received signal model is given by [5]

$$\mathbf{Y}=\sum_{k=1}^K\mathbf{G}_k\mathbf{X}_k+\mathbf{Z}=\sum_{k=1}^K\mathbf{V}\mathbf{H}_k\mathbf{F}^T\mathbf{X}_k+\mathbf{Z}, \tag{19}$$

where $\mathbf{G}_k\in\mathbb{C}^{N_r\times N_p}$ is the space-frequency domain channel matrix of user k, the diagonal matrix $\mathbf{X}_k$ is the training signal of user k satisfying $\mathbf{X}_k^H\mathbf{X}_k=\mathbf{I}$ and $\mathbf{Z}$ is the noise matrix whose components are independent and identically distributed (i.i.d.) complex Gaussian random variables with zero mean and variance σ_z^2. Define F_v, F_h and F_τ as fine (oversampling) factors, $\tilde{\mathbf{I}}_{N\times FN}$ as a matrix composed of the first N rows of the FN dimensional identity matrix and parameter $N_f\triangleq\lceil N_pN_g/N_c\rceil$. Then, $\mathbf{V}_v$, $\mathbf{V}_h$ and $\mathbf{F}$ are partial discrete Fourier transformation (DFT) matrices given by

$$\begin{aligned}\mathbf{V}_v=&\tilde{\mathbf{I}}_{N_{rv}\times F_vN_{rv}}\tilde{\mathbf{V}}_v,\quad \mathbf{V}_h=\tilde{\mathbf{I}}_{N_{rh}\times F_hN_{rh}}\tilde{\mathbf{V}}_h,\\ &\mathbf{F}=\tilde{\mathbf{I}}_{N_p\times F_\tau N_p}\tilde{\mathbf{F}}\tilde{\mathbf{I}}_{F_\tau N_f\times F_\tau N_p}^T\end{aligned} \tag{20}$$

where $\tilde{\mathbf{V}}_v$, $\tilde{\mathbf{V}}_h$ and $\tilde{\mathbf{F}}$ are $F_vN_{r,v}$, $F_hN_{r,h}$ and $F_\tau N_p$ dimensional DFT matrices, respectively, $\mathbf{V}\in\mathbb{C}^{N_r\times F_vF_hN_r}$ is the Kronecker product of $\mathbf{V}_v$ and $\mathbf{V}_h$ and $\mathbf{H}_k\in\mathbb{C}^{F_vF_hN_r\times F_\tau N_f}$ is the beam domain channel matrix of user k whose components follow the independent complex Gaussian distributions with zero mean. Then, (19) can be rewritten as

$$\mathbf{Y}=\mathbf{V}\mathbf{H}\mathbf{M}+\mathbf{Z}, \tag{21}$$

where $\mathbf{H}=[\mathbf{H}_1,\mathbf{H}_2,\cdots,\mathbf{H}_K]\in\mathbb{C}^{F_vF_hN_r\times KF_\tau N_f}$ and $\mathbf{M}=[\mathbf{X}_1\mathbf{F},\mathbf{X}_2\mathbf{F},\cdots,\mathbf{X}_K\mathbf{F}]^T\in\mathbb{C}^{KF_\tau N_f\times N_p}$. After vectorization, we have

$$\mathbf{y}=\tilde{\mathbf{A}}\tilde{\mathbf{h}}+\mathbf{z}, \tag{22}$$

where $\mathbf{y},\mathbf{z}\in\mathbb{C}^{N\times1}$ and $\tilde{\mathbf{h}}\in\mathbb{C}^{\tilde{M}\times1}$ are the vectorizations of $\mathbf{Y}$, $\mathbf{Z}$ and $\mathbf{H}$, respectively, $\tilde{\mathbf{A}}\in\mathbb{C}^{N\times\tilde{M}}$ is the Kronecker product of $\mathbf{M}^T$ and $\mathbf{V}$, $N = N_r N_p$, $\tilde{M} = K F_a F_\tau N_r N_f$, and $F_a = F_v F_h$. Since most components in $\tilde{\mathbf{h}}$ are zeros, we can reduce the dimension of variables by extracting non-zero components. Then, (22) can be reformulated as

$$\mathbf{y} = \mathbf{A}\mathbf{h} + \mathbf{z}, \tag{23}$$

where $\mathbf{h} \in \mathbb{C}^{M\times 1}$ is extracted from $\tilde{\mathbf{h}}$ when there are $M << \tilde{M}$ nono-zeros elements and $\mathbf{A} \in \mathbb{C}^{N\times M}$ is extracted from $\tilde{\mathbf{A}}$ by column. Thus, $\mathbf{h} \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{D}\right)$ with diagonal and positive definite $\mathbf{D}$ and $\mathbf{z} \sim \mathcal{CN}\left(\mathbf{0}, \sigma_z^2 \mathbf{I}\right)$. Combining Lemma 3 and Theorem 2, we have

$$\rho\left(N\mathbf{I} - \mathbf{A}^H\mathbf{A}\right) \leq \left(K F_v F_h F_\tau - 1\right) N. \tag{24}$$

and then the damping factor needs to satisfy

$$d < \frac{2}{K F_v F_h F_\tau} \tag{25}$$

to ensure the convergence of SIGA with initialization $\tilde{\mathbf{g}}_{\text{min}} \leq \boldsymbol{\nu}\left(0\right) \leq \mathbf{0}$. For the case with $K = 48$, $M = 29277$, and $\left(F_v, F_h, F_\tau\right) = \left(2, 2, 2\right)$, when general pilot sequences with constant magnitude property are adopted, $d < 0.0052$ is sufficient to ensure the convergence of SIGA. Note that this range is much larger than the worst case $d < \frac{2}{M} = 6.8\times 10^{-5}$. We finally consider the special case, where the adjustable phase shift pilots (APSPs) [14], are used. In this case, $\mathbf{A}$ is extracted from $\mathbf{A}_p$ where $\tilde{\mathbf{A}}_p = \mathbf{F}_d \otimes \mathbf{V} \in \mathbb{C}^{N\times F_v F_h F_\tau N}$. Combining Lemma 3 and Theorem 2, we have

$$\rho\left(N\mathbf{I} - \mathbf{A}^H\mathbf{A}\right) \leq \left(F_v F_h F_\tau - 1\right) N, \tag{26}$$

and then the damping factor needs to satisfy

$$d < \frac{2}{F_v F_h F_\tau}, \tag{27}$$

to ensure the convergence of SIGA with initialization $\tilde{\mathbf{g}}_{\text{min}} \leq \boldsymbol{\nu}\left(0\right) \leq \mathbf{0}$. For the case with $F_v = F_h = F_\tau = 2$, $d < 0.25$ is sufficient for SIGA to converge.

IV. Simulation Results

In this section, we present numerical simulations to illustrate the theoretical results in this paper. For the massive MIMO-OFDM channel estimation, two types of pilots are used in simulation: one is the general constant magnitude pilot and the other is the APSPs. In the simulation, we focus on the convergence of parameters $\boldsymbol{\nu}$ and $\boldsymbol{\theta}$ in massive MIMO channel estimation. The widely used QuaDRiGa channel model is adopted and the main parameters for the simulations are summarized in Table I.

In the case of general constant magnitude pilot, the pilots of different users in (19) are generated as $\mathbf{X}_k = \text{diag}\left\{\mathbf{x}_k\right\}, k \in \mathcal{Z}_K^+$, where the components of $\mathbf{x}_k$ are drawn from i.i.d. $\mathcal{CN}\left(0, 1\right)$ and then normalized to the unit magnitude. Thus, $\mathbf{A}$ can be calculated from (21) to (23), and $\mathbf{A} \in \mathbb{C}^{N\times M}$ where $N = 46080$ and $M = 29277$ under the system parameters in Table I. The variance of the noise $\mathbf{z}$ is chosen to $\sigma_z^2 = 0.01$, and this value could achieve an signal to noise ratio (SNR) of 20 dB, where the SNR is defined as $\text{SNR} \triangleq \frac{1}{\sigma_z^2}$ in

TABLE I
System Parameter

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_v, F_h, F_τ	$2, 2, 2$
Mobile velocity of users	$3 - 10$kmph

this experiment [5]. The convergence performance of $\boldsymbol{\nu}$ is presented in Fig. 1. To verify Theorem 1, the initialization of $\boldsymbol{\nu}$ is set to be $\boldsymbol{\nu}\left(0\right) = \mathbf{0}$, $\boldsymbol{\nu}\left(0\right) = \tilde{\mathbf{g}}_{\text{min}} = -\frac{N-1}{\sigma_z^2}\mathbf{1}$, and $\boldsymbol{\nu}\left(0\right) = -\mathbf{1}$, repectively. The damping factor is the same as the previous experiment. From Fig. 1, it can be found that $\|\boldsymbol{\nu}\|_2$ converges to the same fixed point in all settings, where the fixed point satisfies $0 < \|\boldsymbol{\nu}^\star\|_2 < \|\tilde{\mathbf{g}}_{\text{min}}\|_2$. And the fixed point $\boldsymbol{\nu}^\star$ is not related to the choice of damping factor. This observation is in exact agreement with Theorem 1. Under the condition that initialization of $\boldsymbol{\nu}$ is set to be

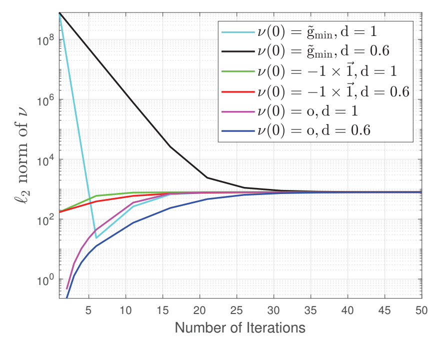

Fig. 1. Convergence of $\|\boldsymbol{\nu}\left(t\right)\|$ for different initializations and damping factors for general constant magnitude pilot.

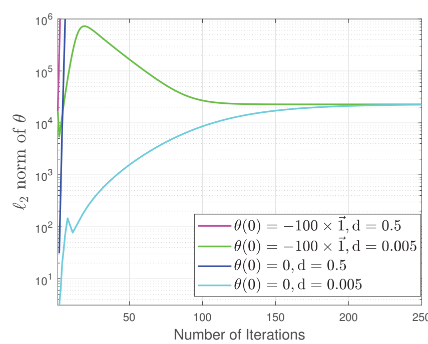

Fig. 2. Convergence and divergence of $\|\boldsymbol{\theta}\left(t\right)\|$ for different initializations and damping factors for general constant magnitude pilot.

$\boldsymbol{\nu}\left(0\right) = \mathbf{0}$, the initialization of $\boldsymbol{\theta}$ is set to be $\boldsymbol{\theta}\left(0\right) = \mathbf{0}$ and $\boldsymbol{\theta}\left(0\right) = -100 \times \mathbf{1}$, respectively and the damping factor is set to be $d = 0.5$ and $d = 0.005$. From Fig. 2, it can be found that $\|\boldsymbol{\theta}\|_2$ diverges when $d = 0.5$ and converges to the same fixed point in case of $d = 0.05$ regardless of the initialization of $\boldsymbol{\theta}$. From Theorem 2, the worst case is that damping factor has to less than $\frac{2}{M} = 6.8 \times 10^{-5}$ but $d = 0.05$ is enough to ensure the convergence of $\boldsymbol{\theta}$. This is consistent with (25).

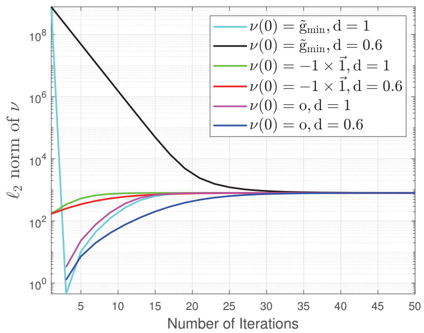

Fig. 3. Convergence of $\|\boldsymbol{\nu}(t)\|$ for different initializations and damping factors for APSPs.

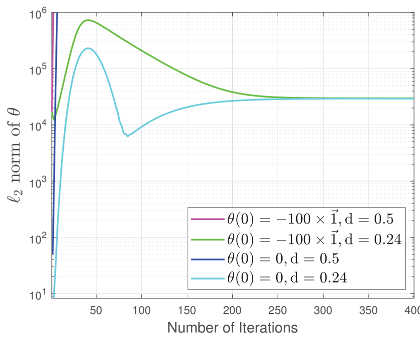

Fig. 4. Convergence and divergence of $\|\boldsymbol{\theta}(t)\|$ for different initializations and damping factors for APSPs.

When the APSPs are used, the pilot of the user k is set to be $\mathbf{X}_k = \mathrm{Diag}\{\mathbf{r}(n_k)\}\mathbf{P}$, where

$$\begin{aligned}&\mathbf{r}(n_k)\\ &= \left[\exp\left\{-\jmath 2\pi \frac{n_k N_1}{F_\tau N_p}\right\}, \cdots, \exp\left\{-\jmath 2\pi \frac{n_k N_2}{F_\tau N_p}\right\}\right]^T \in \mathbb{C}^{N_p},\end{aligned}$$

$n_k \in \{0, 1, \cdots, F_\tau N_p - 1\}$ is the phase shift scheduled for the user k, and $\mathbf{P} = \mathrm{Diag}\{\mathbf{p}\}$ is the basic pilot satisfying $\mathbf{P}\mathbf{P}^H = \mathbf{I}$. We can use [14, Algorithm 1] to determine the value of n_k and thus $\mathbf{X}_k, k \in \mathcal{Z}_K^+$. In this experiment, the dimension of $\mathbf{A} \in \mathbb{C}^{N \times M}$ is also $N = 46080$ and $M = 29277$. The noise variance σ_z^2 is set to the same value in the previous one. The convergence performance of $\boldsymbol{\nu}$ is plotted in Fig. 3 and similar conclusions can be drawn, where the initialization of $\boldsymbol{\nu}$ and the damping factor are the same as in the previous experiment. The convergence performance of $\boldsymbol{\theta}$ is plotted in Fig. 4, where the damping factor is set to be $d = 0.5$ and $d = 0.24$, and the initialization of $\boldsymbol{\nu}$ and $\boldsymbol{\theta}$ is set the same as in the previous experiment. From Fig. 4, $\|\boldsymbol{\theta}\|_2$ diverges when $d = 0.5$ and converges to the same fixed point in case of $d = 0.24$, which is consistent with (27).

V. Conclusion

In this paper, we investigated the convergence of SIGA for massive MIMO-OFDM channel estimation. We analyzed the convergence of SIGA for a general Bayesian inference problem with the measurement matrix of constant magnitude property. We then applied the general theories to the case of massive MIMO-OFDM channel estimation. Through revisiting its iteration, we found that the iterating system of the common SONP is independent of that of the common FONP. Hence, we can check the convergence of the common SONP separately. It was then proved that given the initialization satisfying a particular and large range, the common SONP is convergent no matter the size of the damping factor is. For the convergence of the common FONP, we established a sufficient condition. More specifically, we found that the convergence of the common FONP depends on the spectral radius of the iterating system matrix $\tilde{\mathbf{B}}^\star$. On this basis, we obtained a condition for the damping factor that guarantees the convergence of $\boldsymbol{\theta}(t)$ in the worst case. Further, we determined the range of the damping factor for massive MIMO-OFDM channel estimation by using the specific properties of the measurement matrices. Simulation results confirmed the theoretical results.

References

[1] D. M. Malioutov, J. K. Johnson, and A. S. Willsky, "Walk-sums and belief propagation in gaussian graphical models," *The Journal of Machine Learning Research*, vol. 7, pp. 2031–2064, 2006.

[2] 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.

[3] T. P. Minka, "Expectation propagation for approximate bayesian inference," *arXiv preprint arXiv:1301.2294*, 2013.

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

[5] J. Y. Yang, Y. Chen, A.-A. Lu, W. Zhong, X. Q. Gao, X. You, X.-G. Xia, and D. Slock, "Simplified information geometry approach for massive MIMO-OFDM channel estimation – Part I: Algorithm and fixed point analysis," *arXiv preprint arXiv:2401.02035*, 2024.

[6] ——, "Channel estimation for massive MIMO-OFDM: Simplified information geometry approach," in *2023 IEEE 98th Vehicular Technology Conference (VTC2023-Fall)*, 2023, pp. 1–6.

[7] Q. Su and Y.-C. Wu, "Convergence analysis of the variance in Gaussian belief propagation," *IEEE Trans. Signal Process.*, vol. 62, no. 19, pp. 5119–5131, Oct. 2014.

[8] ——, "On convergence conditions of Gaussian belief propagation," *IEEE Trans. Signal Process.*, vol. 63, no. 5, pp. 1144–1155, Mar. 2015.

[9] K. Takeuchi, "On the convergence of orthogonal/vector amp: Long-memory message-passing strategy," *IEEE Trans. Inf. Theory*, vol. 68, no. 12, pp. 8121–8138, Dec. 2022.

[10] 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, Sep. 2019.

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

[12] J. Y. Yang, Y. Chen, M. R. Fan, X. Q. Gao, X.-G. Xia, and D. Slock, "Simplified information geometry approach for massive MIMO-OFDM channel estimation - Part II: Convergence analysis," *arXiv preprint arXiv:2401.02037*, 2024.

[13] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York, NY, USA: Cambridge Univ. press, 2012.

[14] 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.