Non-Circular Combined Expectation Propagation and Variational Bayes for Semi-Blind Channel Estimation in Cell-Free Systems

Zilu Zhao, Dirk Slock
Communication Systems Department, EURECOM, France
zilu.zhao@eurecom.fr, dirk.slock@eurecom.fr

***Abstract*—In this work, we investigate uplink communication in Semi-Blind Cell-Free (CF) Massive Multiple-Input Multiple-Output (MaMIMO) systems. A key challenge in CF MaMIMO systems is pilot contamination, which arises when multiple user terminals (UTs) share the same pilot sequence due to an imbalance between the number of UTs and the length of the pilot sequence. Semi-blind approaches have been proposed to address this issue by enabling access points (APs) to jointly estimate both the channel and user data. This joint estimation, however, leads to a bilinear problem. To develop a tractable algorithm that exploits the finite alphabet of the user data, we analyze the constrained Bethe Free Energy (BFE) of the bilinear system and propose a message-passing algorithm based on minimizing the constrained BFE. We decompose complex quantities into in-phase and quadrature components, removing circular Gaussian constraints. Additionally, we find that replacing normal covariance constraints with average covariance constraints significantly reduces computational complexity. Simulation results indicate that this simplification incurs negligible performance loss.**

I. Introduction

In Cell-Free (CF) Massive Multiple-Input Multiple-Output (MaMIMO) systems, user terminals (UTs) are simultaneously served by all access points (APs) in a given region. A significant challenge in CF MaMIMO systems is pilot contamination, which occurs when the number of users exceeds the length of the pilot sequences. Consequently, APs cannot estimate the channel solely based on pilot sequences. To address this issue, semi-blind channel estimation is employed [1]. In Semi-Blind settings, APs jointly estimate the channel and user data based on received signals and limited pilot sequences.

A. Prior Work

Bayesian estimation in semi-blind structures holds significant potential [1], but it also presents challenges due to high-dimensional and intractable integrals. Message-passing algorithms, particularly Expectation Propagation (EP) [2] and Belief Propagation (BP) [3], are widely used in Bayesian estimation. Both EP and BP assume a factored joint probability density function and simplify high-complexity global inference problems into manageable local inference tasks. EP further reduces complexity by approximating the factors of the joint pdf with simpler forms, such as Gaussian distributions.

Variable-Level EP (VL-EP) was introduced for Gaussian input data by combining Expectation-Maximization (EM) with EP [4]. To improve its convergence properties, hybrid EM-EP and loop-free EM-EP algorithms were proposed in [5]. However, these approaches are not designed to handle user symbols from finite alphabets.

1) Expectation Propagation for Gaussian Mixture Models: The bilinear combination of a Gaussian distribution (e.g., channel distribution) and a discrete distribution (e.g., input data distribution) leads to a Gaussian Mixture Model. To address the limitations of VL-EP, a distributed bilinear-EP algorithm was proposed in [6], which adopts a brute-force approach to inference over finite alphabets, avoiding high-dimensional computations by considering only one data symbol at a time. Inspired by [6] and [7], the authors of [8] proposed a simplified decentralized bilinear-EP algorithm.

2) Bethe Free Energy: The Bethe Free Energy (BFE) is another powerful tool for Bayesian inference. It represents the variational energy of a factored joint pdf under a specific trial distribution, whose form is determined by the factorization scheme of the joint pdf. It has been demonstrated [9] that various message-passing algorithms, such as EP and BP, can be derived by optimizing BFE under different constraints on the trial distribution.

Hybrid Vector Message Passing (HVMP) [10] was proposed based on BFE optimization, introducing a mean-field constraint for the bilinear factor. However, this method does not account for finite alphabets and entirely neglects the correlation between the channel and data.

B. Main Contributions

We propose a low-complexity algorithm for semi-blind channel and data estimation, leveraging a framework based on BFE constrained optimization. To effectively handle posterior interference, we introduce an auxiliary variable, which enables a more tractable optimization process. Unlike prior works such as [6] and [8], our method treats the entire data sequence of a single user as a single atomic variable, significantly simplifying the estimation procedure.

To address non-analytical integrals that arise during posterior estimation, we incorporate mean-field assumptions into the belief factors by including delta functions. This approximation streamlines the derivations and reduces computational complexity. Furthermore, we derive the algorithm by separating complex quantities into their in-phase and quadrature components.

Lastly, we introduce a covariance averaging operation to further minimize complexity while maintaining robust performance. This combination of methods ensures a computationally efficient and scalable approach to semi-blind estimation.

II. System Model

We examine the uplink cell-free semi-blind network containing K single-antenna user terminals (UTs) and L access points (APs). Each AP is equipped with M antennas. The received signals of the l-th AP is

$$\begin{bmatrix}\widehat{\mathbf{Y}}_{p,l} & \widehat{\mathbf{Y}}_l\end{bmatrix} = \widehat{\mathbf{H}}_l \begin{bmatrix}\widehat{\mathbf{X}}_p^\intercal & \widehat{\mathbf{X}}^\intercal\end{bmatrix} + \begin{bmatrix}\widehat{\mathbf{V}}_{p,l} & \widehat{\mathbf{V}}_l\end{bmatrix} \in \mathbb{C}^{M\times(P+T)}, \quad (1)$$

where $\widehat{\mathbf{H}}_l \in \mathbb{C}^{M\times K}$ models the channel matrix, $\widehat{\mathbf{X}}_p \in \mathbb{C}^{P\times K}$ models the pilot sequences of all the users, $\widehat{\mathbf{X}} \in \mathbb{C}^{T\times K}$ models the transmitted data sequences.

The complex model can be transformed into in-phase and quadrature representation

$$\begin{bmatrix}\Re[\widehat{\mathbf{Y}}_{p,l}] & \Re[\widehat{\mathbf{Y}}_l]\\ \Im[\widehat{\mathbf{Y}}_{p,l}] & \Im[\widehat{\mathbf{Y}}_l]\end{bmatrix} = \begin{bmatrix}\Re[\widehat{\mathbf{H}}_l] & -\Im[\widehat{\mathbf{H}}_l]\\ \Im[\widehat{\mathbf{H}}_l] & \Re[\widehat{\mathbf{H}}_l]\end{bmatrix} \begin{bmatrix}\Re[\widehat{\mathbf{X}}_p^\intercal] & \Re[\widehat{\mathbf{X}}^\intercal]\\ \Im[\widehat{\mathbf{X}}_p^\intercal] & \Im[\widehat{\mathbf{X}}^\intercal]\end{bmatrix} + \mathbf{V}, \quad (2)$$

where $\mathbf{V} = \begin{bmatrix}\Re[\widehat{\mathbf{V}}_{p,l}] & \Re[\widehat{\mathbf{V}}_l]\\ \Im[\widehat{\mathbf{V}}_{p,l}] & \Im[\widehat{\mathbf{V}}_l]\end{bmatrix}$.

979-8-3315-2096-0/25/$31.00 ©2025 IEEE

For simplicity, we define the following notations

$$\mathbf{Y}_{p,l} = \begin{bmatrix} \Re[\widehat{\mathbf{Y}}_{p,l}] \\ \Im[\widehat{\mathbf{Y}}_{p,l}] \end{bmatrix};\ \mathbf{Y}_l = \begin{bmatrix} \Re[\widehat{\mathbf{Y}}_l] \\ \Im[\widehat{\mathbf{Y}}_l] \end{bmatrix};\ \mathbf{H}_l = \begin{bmatrix} \Re[\widehat{\mathbf{H}}_l] \\ \Im[\widehat{\mathbf{H}}_l] \end{bmatrix};$$

$$\boldsymbol{x}_{k,\Re} = \Re[\widehat{\boldsymbol{x}}_k];\ \boldsymbol{x}_{k,\Im} = \Im[\widehat{\boldsymbol{x}}_k];\ \mathbf{X}_k^\top = \begin{bmatrix} \boldsymbol{x}_{k,\Re}^\top \\ \boldsymbol{x}_{k,\Im}^\top \end{bmatrix};\ \boldsymbol{x}_k = \mathrm{vec}[\mathbf{X}_k^\top]$$

We assume that $\mathbf{H}_l$ is column-wise independent, with the distribution of its k-th column given by $\mathbf{h}_{lk} \sim \mathcal{N}(\mathbf{h}_{lk}|\mathbf{0}, \boldsymbol{\Xi}_{\mathbf{h}_{lk}})$. Let $\mathcal{S} = \{\mathbf{s}_1, \dots, \mathbf{s}_{|\mathcal{S}|}\}$ denote the set of symbol constellations. Each data symbol is assumed to be independent of others, with a distribution $p(\boldsymbol{x}_{kt})$, where $\boldsymbol{x}_{kt}$ represents the t-th column of $\mathbf{X}_k$.
The power of each pilot sequence (complex) is assumed to be $P\sigma_x^2$. The noise is modeled as additive white Gaussian noise (AWGN), with each entry in $\mathbf{V}$ having power $\sigma_v^2/2$. For simplicity, we define $\mathbf{C}_v = \frac{\sigma_v^2}{2}\mathbf{I}$.

A. Orthogonal Pilots

When orthogonal pilots are used, we correlate the received pilot signals $Y_{p,l}$ with the g-th pilot sequence $\widetilde{\boldsymbol{x}}_{p,g}$ (not to confuse with the pilot sequence of the g-th user) to obtain the correlated version of the received pilot signals $\widetilde{\boldsymbol{y}}_{p,lg}$:

$$\widehat{\widetilde{\boldsymbol{y}}}_{p,lg} = \widehat{\mathbf{Y}}_{p,l}\widehat{\widetilde{\boldsymbol{x}}}_{p,g}^* = P\sigma_x^2\widehat{\mathbf{H}}_{lG_g}\mathbf{1}_{|G_g|} + \widehat{\widetilde{\boldsymbol{v}}}_{p,lg}, \tag{3}$$

where we use G_g to denote the UTs groups using the g-th pilot sequence. The columns of $\widehat{\mathbf{H}}_{lG_g}$ are composed of the complex channel coefficients corresponding to the users using the g-th pilot, i.e., $\widehat{\mathbf{h}}_{lk}$ is a column of $\widehat{\mathbf{H}}_{lG_g}$ if $\widehat{\boldsymbol{x}}_{p,k} = \widehat{\widetilde{\boldsymbol{x}}}_{p,g}$. We denote $\widehat{\widetilde{\boldsymbol{v}}}_{p,lg} = \widehat{\mathbf{V}}_{p,l}\widehat{\widetilde{\boldsymbol{x}}}_{p,g}^*$ which is the transformed noise following a distribution $\mathcal{CN}(\mathbf{0}, \sigma_x^2\sigma_v^2 P\mathbf{I}_M)$. The complex expression is transformed into in-phase and quadrature expression. Denote

$$\widetilde{\boldsymbol{y}}_{p,lg} = \begin{bmatrix} \Re[\widehat{\widetilde{\boldsymbol{y}}}_{p,lg}] \\ \Im[\widehat{\widetilde{\boldsymbol{y}}}_{p,lg}] \end{bmatrix};\ \widetilde{\boldsymbol{v}}_g = \begin{bmatrix} \Re[\widehat{\widetilde{\boldsymbol{v}}}_g] \\ \Im[\widehat{\widetilde{\boldsymbol{v}}}_g] \end{bmatrix} \tag{4}$$

Thus, we have

$$\widetilde{\boldsymbol{y}}_{p,lg} = P\sigma_x^2 \sum_{k\in G_g} \mathbf{h}_{lk} + \widetilde{\boldsymbol{v}}_g, \tag{5}$$

where $\widetilde{\boldsymbol{v}}_g \sim \mathcal{N}(\boldsymbol{v}_g|\mathbf{0}_{2M}, \frac{\sigma_v^2\sigma_x^2 P}{2}\mathbf{I}_{2M})$. For simplicity, denote $\mathbf{C}_{v,p} = \frac{\sigma_v^2\sigma_x^2 P}{2}\mathbf{I}_{2M}$ and define $\mathbf{H}_{lG_g}$ to be a matrix whose columns are $\mathbf{h}_{lk}$ such that $k \in G_g$ and its vectorization $\mathbf{h}_{lG_g} = \mathrm{vec}(\mathbf{H}_{lG_g})$.

B. Factored Joint Distribution

Introduce an auxiliary variable

$$\mathbf{Z}_{lk} = \widetilde{\mathbf{H}}_{lk}\mathbf{X}_k^\top = \mathbf{h}_{lk}\boldsymbol{x}_{\Re,k}^\top + \mathbf{S}\mathbf{h}_{lk}\boldsymbol{x}_{\Im,k}^\top \tag{6}$$

where

$$\mathbf{S} = \begin{bmatrix} & -\mathbf{I}_M \\ \mathbf{I}_M & \end{bmatrix};\ \widetilde{\mathbf{H}}_{lk} = \begin{bmatrix} \mathbf{h}_{lk} & \mathbf{S}\mathbf{h}_{lk} \end{bmatrix}. \tag{7}$$

The vectorization $\mathbf{Z}_{lk}$ can be represented as $\mathbf{z}_{lk} = \mathrm{vec}[\mathbf{Z}_{lk}]$. Therefore, the likelihood of $\mathbf{Z}_{lk}$ is captured by Dirac function $p(\mathbf{Z}_{lk}|\mathbf{h}_{lk}, \boldsymbol{x}_k) = \delta(\mathbf{Z}_{lk} - \mathbf{h}_{lk}\boldsymbol{x}_{\Re,k}^\top - \mathbf{S}\mathbf{h}_{lk}\boldsymbol{x}_{\Im,k}^\top)$. The joint probability density function (PDF) can be derived as

$$\begin{aligned} &p(\widehat{\mathbf{Y}}_{p,\{l\}}, \widehat{\mathbf{Y}}_{\{l\}}, \widehat{\mathbf{Z}}_{\{l\}\{k\}}, \widehat{\mathbf{H}}_{\{l\}}, \widehat{\mathbf{X}}) \\ &= \prod_l p(\mathbf{Y}_l|\mathbf{Z}_{l\{k\}}) \prod_l \prod_k p(\mathbf{Z}_{lk}|\mathbf{h}_{lk}, \boldsymbol{x}_k) \\ &\prod_l \prod_g p(\widetilde{\boldsymbol{y}}_{p,lg}, \mathbf{H}_{lG_g}) \prod_k p(\boldsymbol{x}_k). \end{aligned} \tag{8}$$

For simplicity, we define

$$\begin{aligned} &f_{\mathbf{z}_l}(\mathbf{z}_{l\{k\}}) \propto p(\mathbf{Y}_l|\mathbf{Z}_{l\{k\}});\ f_{\mathbf{h}_{lG_g}}(\mathbf{h}_{lG_g}) \propto p(\widetilde{\boldsymbol{y}}_{p,lg}, \mathbf{H}_{lG_g}) \\ &f_{\boldsymbol{x}_k}(\boldsymbol{x}_k) = p(\boldsymbol{x}_k);\ f_{\delta_{lk}}(\mathbf{z}_{lk}, \mathbf{h}_{lk}, \boldsymbol{x}_k) \propto p(\mathbf{Z}_{lk}|\mathbf{h}_{lk}, \boldsymbol{x}_k). \end{aligned} \tag{9}$$

The factorization given by (8) admits a factor graph [3]. We denote $\mathbb{F} = \{f_{\mathbf{z}_l}, f_{\mathbf{h}_{lG_g}}, f_{\boldsymbol{x}_k}, \delta_{lk}\}$ as the set of all factor nodes and $\mathbb{V} = \{\mathbf{z}_{lk}, \mathbf{h}_{lk}, \boldsymbol{x}_k\}$ as the set of all variable nodes.

III. Bethe Free Energy Optimization Framework

Bethe free energy is the approximated variational free energy between the true probability (8) and a constrained Bethe approximation trial function. For a given factored pdf p, its trial pdf b is obtained by:

$$p(\boldsymbol{\theta}) \propto \prod_\alpha f_\alpha(\boldsymbol{\theta}_\alpha) \Rightarrow b(\boldsymbol{\theta}) = \frac{\prod_\alpha b_{f_\alpha}(\boldsymbol{\theta}_\alpha)}{\prod_i b_{\theta_i}(\theta_i)^{|N_i|-1}}, \tag{10}$$

s.t.

$$\forall \alpha, \theta_i \in \boldsymbol{\theta}_\alpha,\ \int b_{f_\alpha}(\boldsymbol{\theta}_\alpha) d\boldsymbol{\theta}_{\bar{i}} = b_{\theta_i}(\theta_i) \tag{11}$$

where $|N_i|$ denotes the number of factors f_α that contain θ_i and $\boldsymbol{\theta}_{\bar{i}}$ denotes all the variables except θ_i.
With (10)-(11), the BFE can be obtained as

$$\mathrm{BFE}{=}D[b(\boldsymbol{\theta})\| \prod_\alpha f_\alpha(\boldsymbol{\theta}_\alpha)]{=}\sum_\alpha D(b_{f_\alpha}\|f_\alpha) + \sum_i (|N_i|-1)H(b_{\theta_i}), \tag{12}$$

where we define $D(b\|q) = \int b(\boldsymbol{\theta}) \ln \frac{b(\boldsymbol{\theta})}{q(\boldsymbol{\theta})} d\boldsymbol{\theta}$, and $H(\cdot)$ as entropy. It is worth noticing that (12) only holds if the factorization (10) is loop-free and strict constraints (11) are applied. Otherwise, (12) is only an approximation.

A. Bethe Approximation with Constraints

Following [9], the BFE of (8) is:

$$\begin{aligned} \mathrm{BFE} = &\sum_l D[b_{f_{\mathbf{z}_l}}(\mathbf{z}_{l\{k\}})\|f_{\mathbf{z}_l}(\mathbf{z}_{l\{k\}})] \\ &+\sum_{l,g} D[b_{f_{\mathbf{h}_{lG_g}}}(\mathbf{h}_{lG_g})\|f_{\mathbf{h}_{lG_g}}(\mathbf{h}_{lG_g})] + \sum_k D[b_{f_{\boldsymbol{x}_k}}(\boldsymbol{x}_k)\|f_{\boldsymbol{x}_k}(\boldsymbol{x}_k)] \\ &+ \sum_{l,k} D[b_{\delta_{lk}}(\mathbf{z}_{lk}, \mathbf{h}_{lk}, \boldsymbol{x}_k)\|f_{\delta_{lk}}(\mathbf{z}_{lk}, \mathbf{h}_{lk}, \boldsymbol{x}_k)] + \sum_{l,k} H[b_{\mathbf{z}_{lk}}(\mathbf{z}_{lk})] \\ &+ \sum_{l,k} H[b_{\mathbf{h}_{lk}}(\mathbf{h}_{lk})] + \sum_k L \cdot H[b_{\boldsymbol{x}_k}(\boldsymbol{x}_k)]. \end{aligned} \tag{13}$$

where all the factor-level beliefs $b_{f_{\mathbf{z}_l}}$, $b_{\delta_{\mathbf{z}_{lk}}}$, $b_{f_{\mathbf{h}_{lG_g}}}$, $b_{f_{\boldsymbol{x}_k}}$, and variable-level beliefs $b_{\mathbf{h}_{lk}}$, $b_{\mathbf{z}_{lk}}$, $b_{\boldsymbol{x}_k}$ are proper distributions normalized to one. Furthermore, to make all these factors consistent, the variable-level beliefs must be the marginal distribution of the factor-level beliefs. For all $l \in [1, L]$, $k \in [1, K]$, the constraints for the $\boldsymbol{x}_k$ are

$$\int b_{\delta_{lk}}(\mathbf{z}_{lk}, \mathbf{h}_{lk}, \boldsymbol{x}_k) d\mathbf{z}_{lk} d\mathbf{h}_{lk} = b_{\boldsymbol{x}_k}(\boldsymbol{x}_k) \tag{14}$$

$$b_{f_{\boldsymbol{x}_k}}(\boldsymbol{x}_k) = b_{\boldsymbol{x}_k}(\boldsymbol{x}_k). \tag{15}$$

However, satisfying the strict constraints of $\mathbf{h}_{lk}$ and $\mathbf{z}_{lk}$ will lead to an intractable problem. Therefore, we relax the strict constraints to first and second-order moment constraints (specifically, mean and covariance constraints). W.l.o.g., we denote those sufficient statistics as $\phi_{\mathbf{h}_{lk}}(\mathbf{h}_{lk})$, $\phi_{\mathbf{z}_{lk}}(\mathbf{z}_{lk})$

$$\mathbb{E}_{b_{f_{\mathbf{z}_l}}}[\phi_{\mathbf{z}_{lk}}(\mathbf{z}_{lk})] = \mathbb{E}_{b_{\mathbf{z}_{lk}}}[\phi_{\mathbf{z}_{lk}}(\mathbf{z}_{lk})] \tag{16}$$

$$\mathbb{E}_{\delta_{lk}}[\phi_{\mathbf{z}_{lk}}(\mathbf{z}_{lk})] = \mathbb{E}_{b_{\mathbf{z}_{lk}}}[\phi_{\mathbf{z}_{lk}}(\mathbf{z}_{lk})] \tag{17}$$

$$\mathbb{E}_{b_{f_{\mathbf{h}_{lG_g}}}}[\phi_{\mathbf{h}_{lk}}(\mathbf{h}_{lk})] = \mathbb{E}_{b_{\mathbf{h}_{lk}}}[\phi_{\mathbf{h}_{lk}}(\mathbf{h}_{lk})] \tag{18}$$

$$\mathbb{E}_{b_{\delta_{lk}}}[\phi_{\mathbf{h}_{lk}}(\mathbf{h}_{lk})] = \mathbb{E}_{b_{\mathbf{h}_{lk}}}[\phi_{\mathbf{h}_{lk}}(\mathbf{h}_{lk})] \tag{19}$$

Moreover, to make the further derivation tractable with finite input $\mathbf{X}$, we only consider the average covariance constraints of elements within every size-$2M$ block $\forall t \in [1, T]$, $[\mathbf{z}_{lk}]_{2M(t-1)+1:2Mt}$.

B. Bethe Free Energy Optimization

The optimization criteria can be concluded by

$$\begin{aligned} &\min_{b} \text{BFE} \\ &\text{s.t.} \;\; (14) \sim (19). \end{aligned} \tag{20}$$

We observe the term $D[b_{\delta_{lk}}(\mathbf{z}_{lk}, \mathbf{h}_{lk}, \boldsymbol{x}_k)\|\delta(\mathbf{Z}_{lk} - \mathbf{h}_{lk}\boldsymbol{x}_k^{\mathsf{T}})]$ in (13). Since we need to minimize the BFE, the posterior factor $b_{\delta_{lk}}$ must contain the factor $\delta(\mathbf{Z}_{lk} - \mathbf{h}_{lk}\boldsymbol{x}_k^{\mathsf{T}})$ to avoid infinity BFE value. In order to have an analytical algorithm, we use the following mean-field approximation for the joint belief $b_{\delta_{lk}}$:

$$b_{\delta_{lk}}(\mathbf{z}_{lk}, \mathbf{h}_{lk}, \boldsymbol{x}_k) = b_{\delta_{\mathbf{h},lk}}(\mathbf{h}_{lk}) b_{\delta_{\boldsymbol{x},lk}}(\boldsymbol{x}_k) p(\mathbf{Z}_{lk}|\mathbf{h}_{lk}, \boldsymbol{x}_k), \tag{21}$$

where the belief $b_{\delta_{\mathbf{h},lk}}$ and $b_{\delta_{\boldsymbol{x},lk}}$ are beliefs normalized to one. By using Lagrangian method, we can obtain the following message-passing style system of equations along with (21):

$$b_{f_{\mathbf{z}_l}}(\mathbf{z}_{l\{k\}}) = p(\mathbf{Y}_l|\mathbf{z}_{l\{k\}}) \prod_k \mu_{\mathbf{z}_{lk};f_{\mathbf{z}_l}}(\mathbf{z}_{lk}) \tag{22}$$

$$b_{f_{\mathbf{h}_{lG_g}}}(\mathbf{h}_{lG_g}) = p(\widetilde{\boldsymbol{y}}_{p,lg}, \mathbf{h}_{lG_g}) \prod_{k \in G_g} \mu_{\mathbf{h}_{lk};f_{\mathbf{h}_{lG_g}}}(\mathbf{h}_{lk}) \tag{23}$$

$$b_{f_{\boldsymbol{x}_k}}(\boldsymbol{x}_k) = p(\boldsymbol{x}_k)\mu_{\boldsymbol{x}_k;f_{\boldsymbol{x}_k}}(\boldsymbol{x}_k) \tag{24}$$

$$b_{\delta_{\mathbf{h},lk}}(\mathbf{h}_{lk}) = \mu_{\mathbf{h}_{lk};\delta_{lk}}(\mathbf{h}_{lk}) e^{\int b_{\delta_{\boldsymbol{x},lk}}(\boldsymbol{x}_k) \ln \mu_{\mathbf{z}_{lk};\delta_{lk}}(\text{vec}(\mathbf{h}_{lk}\boldsymbol{x}_k^{\mathsf{T}}))d\boldsymbol{x}_k} \tag{25}$$

$$b_{\delta_{\boldsymbol{x},lk}}(\boldsymbol{x}_k) = \mu_{\boldsymbol{x}_k;\delta_{lk}}(\boldsymbol{x}_k) e^{\int b_{\delta_{\mathbf{h},lk}}(\mathbf{h}_{lk}) \ln \mu_{\mathbf{z}_{lk};\delta_{lk}}(\text{vec}(\mathbf{h}_{lk}\boldsymbol{x}_k^{\mathsf{T}}))d\mathbf{h}_{lk}} \tag{26}$$

$$b_{\mathbf{z}_{lk}}(\mathbf{z}_{lk}) = \mu_{\mathbf{z}_{lk};f_{\mathbf{z}_l}}(\mathbf{z}_{lk})\mu_{\mathbf{z}_{lk};\delta_{lk}}(\mathbf{z}_{lk}) \tag{27}$$

$$b_{\mathbf{h}_{lk}}(\mathbf{h}_{lk}) = \mu_{\mathbf{h}_{lk};f_{\mathbf{h}_{lG_g}}}(\mathbf{h}_{lk})\mu_{\mathbf{h}_{lk};\delta_{lk}}(\mathbf{h}_{lk}) \tag{28}$$

$$b_{\boldsymbol{x}_k}(\boldsymbol{x}_k) = [\mu_{\boldsymbol{x}_k;f_{\boldsymbol{x}_k}}(\boldsymbol{x}_k) \prod_l \mu_{\boldsymbol{x}_k;\delta_{lk}}(\boldsymbol{x}_k)]^{1/L}, \tag{29}$$

The equations (21)$\sim$ (26) describes the factor level beliefs while (27)$\sim$(29) are variable level beliefs. For all $f \in \mathbb{F}$, $\boldsymbol{\theta} \in \mathbb{V}$, we interpret $\mu_{\boldsymbol{\theta};f}$ as the variable to factor message. Furthermore, we can define the factor to variable messages such that the following relation holds [11]

$$\forall f \in N(\boldsymbol{\theta}), \mu_{\boldsymbol{\theta};f}(\boldsymbol{\theta}) = \prod_{f' \in N(\boldsymbol{\theta})/\{f\}} \mu_{f';\boldsymbol{\theta}}(\boldsymbol{\theta}), \tag{30}$$

where $N(\boldsymbol{\theta})$ denotes the neighborhood around $\boldsymbol{\theta}$ in the corresponding factor graph. Thus, (29) can be rewritten into the message passing form

$$b_{\boldsymbol{x}_k}(\boldsymbol{x}_k) = \mu_{f_{\boldsymbol{x}_k};\boldsymbol{x}_k}(\boldsymbol{x}_k) \prod_l \mu_{\delta_{lk};\boldsymbol{x}_k}(\boldsymbol{x}_k) \tag{31}$$

Since the sufficient statistics we consider here are first and second-order moments, the messages $\mu_{f_{\mathbf{h}_{lG_g}};\mathbf{h}_{lk}}$, $\mu_{\delta_{lk};\mathbf{h}_{lk}}$, $\mu_{f_{\mathbf{z}_l};\mathbf{z}_{lk}}$ and $\mu_{\delta_{lk};\mathbf{z}_{lk}}$ are all (unnormalized) Gaussian distributions.

IV. Algorithm Derivation

A. Message from δ_{lk} to $\boldsymbol{x}_k$

By satisfying the constraint (14) between (26) and (31), we obtain the iterative updating expression for the message from δ_{lk} to $\boldsymbol{x}_k$:

$$\mu_{\delta_{lk};\boldsymbol{x}_k}(\boldsymbol{x}_k) = e^{\int b_{\delta_{\mathbf{h},lk}}(\mathbf{h}_{lk}) \ln \mu_{\mathbf{z}_{lk};\delta_{lk}}[\mathbf{A}(\mathbf{h}_{lk})\boldsymbol{x}_k]d\mathbf{h}}, \tag{32}$$

where $\mu_{\mathbf{z}_{lk};\delta_{lk}}(\mathbf{z}_{lk}) = \mu_{f_{\boldsymbol{y}_l};\mathbf{z}_{lk}}(\mathbf{z}_{lk})$, and we define the linear transformation matrix

$$\mathbf{A}(\mathbf{h}_{lk}) = \mathbf{I}_T \otimes \begin{bmatrix}\mathbf{h}_{lk} & \mathbf{S}\mathbf{h}_{lk}\end{bmatrix}. \tag{33}$$

For simplicity, denote

$$\widetilde{\mathbf{H}}_{lk} = \begin{bmatrix}\mathbf{h}_{lk} & \mathbf{S}\mathbf{h}_{lk}\end{bmatrix}. \tag{34}$$

The message $\mu_{\mathbf{z}_{lk};\delta_{lk}}$ can be explicitly described as

$$\mu_{\mathbf{z}_{lk};\delta_{lk}}(\mathbf{z}_{lk}) = \mathcal{N}(\mathbf{z}_{lk}|\mathbf{m}_{\mathbf{z}_{lk};\delta_{lk}}, \mathbf{C}_{\mathbf{z}_{lk};\delta_{lk}}), \tag{35}$$

where $\mathbf{C}_{\mathbf{z}_{lk};\delta_{lk}}$ is a $2TM \times 2TM$ block diagonal matrix

$$\mathbf{C}_{\mathbf{z}_{lk};\delta_{lk}} = \text{blkdiag}(\mathbf{C}_{\mathbf{z}_{lk1};\delta_{lk}}, \ldots, \mathbf{C}_{\mathbf{z}_{lkt};\delta_{lk}}, \ldots, \mathbf{C}_{\mathbf{z}_{lkT};\delta_{lk}}), \tag{36}$$

with blocks of size $2M \times 2M$ and the operation blkdiag$(\cdot)$ forms a block diagonal matrix the same way as defined in MatLab. The expression in (32) expands as a Gaussian distribution of $\boldsymbol{x}_k$.
We denote the first and second-order moments of the belief of $\mathbf{h}_{lk}$ at factor node δ_{lk} as

$$\mathbb{E}_{b_{\delta_{lk}}}[\mathbf{h}_{lk}] = \mathbf{m}_{b_{\delta_{\mathbf{h}_{lk}}}}; \; \mathbb{E}_{b_{\delta_{lk}}}[\mathbf{h}_{lk}\mathbf{h}_{lk}^{\mathsf{T}}] = \boldsymbol{R}_{b_{\delta_{\mathbf{h}_{lk}}}}, \tag{37}$$

The corresponding covariance matrix of $\boldsymbol{x}_k$ can be computed from (32) as

$$\mathbf{C}_{\delta_{lk};\boldsymbol{x}_k} = \text{blkDiag}(\mathbf{C}_{\delta_{lk},\boldsymbol{x}_{k1}}, \ldots, \mathbf{C}_{\delta_{lk},\boldsymbol{x}_{kt}}, \ldots, \mathbf{C}_{\delta_{lk},\boldsymbol{x}_{kT}}) \tag{38}$$

where

$$\mathbf{C}_{\delta_{lk};\boldsymbol{x}_{kt}} = \begin{bmatrix}\gamma_{\delta_{lk};x_{kt,\Re\Re}} & \gamma_{\delta_{lk};x_{kt,\Re\Im}} \\ \gamma_{\delta_{lk};x_{kt,\Im\Re}} & \gamma_{\delta_{lk};x_{kt,\Im\Im}}\end{bmatrix}^{-1}, \tag{39}$$

$$\begin{aligned} \gamma_{\delta_{lk};x_{kt,\Re\Re}} &= \text{tr}[\mathbf{C}^{-1}_{\mathbf{z}_{lkt};\delta_{lk}}\boldsymbol{R}_{b_{\delta_{\mathbf{h}_{lk}}}}] \\ \gamma_{\delta_{lk};x_{kt,\Re\Im}} &= \text{tr}[\mathbf{C}^{-1}_{\mathbf{z}_{lkt};\delta_{lk}}\mathbf{S}\boldsymbol{R}_{b_{\delta_{\mathbf{h}_{lk}}}}] \\ \gamma_{\delta_{lk};x_{kt,\Im\Re}} &= \text{tr}[\mathbf{C}^{-1}_{\mathbf{z}_{lkt};\delta_{lk}}\boldsymbol{R}_{b_{\delta_{\mathbf{h}_{lk}}}}\mathbf{S}^{\mathsf{T}}] \\ \gamma_{\delta_{lk};x_{kt,\Im\Im}} &= \text{tr}[\mathbf{C}^{-1}_{\mathbf{z}_{lkt};\delta_{lk}}\mathbf{S}\boldsymbol{R}_{b_{\delta_{\mathbf{h}_{lk}}}}\mathbf{S}^{\mathsf{T}}]. \end{aligned} \tag{40}$$

Similarly, the mean of the feedback Gaussian message is obtained by the integral

$$\mathbf{m}_{\delta_{lk};\boldsymbol{x}_k} = \begin{bmatrix}\mathbf{m}^{\mathsf{T}}_{\delta_{lk};\boldsymbol{x}_{k1}}, \ldots, \mathbf{m}^{\mathsf{T}}_{\delta_{lk};\boldsymbol{x}_{kT}}\end{bmatrix}^{\mathsf{T}}, \tag{41}$$

where

$$\mathbf{m}_{\delta_{lk};\boldsymbol{x}_{kt}} = \mathbf{C}_{\delta_{lk};\boldsymbol{x}_{kt}} \boldsymbol{M}^{\mathsf{T}}_{b_{\delta_{\widetilde{\mathbf{H}}_{lk}}}} \mathbf{C}^{-1}_{\mathbf{z}_{lkt};\delta_{lk}} \mathbf{m}_{\mathbf{z}_{lkt};\delta_{lk}}, \tag{42}$$

with $\boldsymbol{M}_{b_{\delta_{\widetilde{\mathbf{H}}_{lk}}}} = \begin{bmatrix}\mathbf{m}_{b_{\delta_{\mathbf{h}_{lk}}}} & \mathbf{S}\mathbf{m}_{b_{\delta_{\mathbf{h}_{lk}}}}\end{bmatrix}$.

B. Message from $\boldsymbol{x}_k$ to δ_{lk} and the Belief $b_{\boldsymbol{x},k}$

Due to (14), (26) and (31), the belief of $\boldsymbol{x}_k$ at δ_{lk} is

$$\begin{aligned} &b_{\delta_{\boldsymbol{x},lk}}(\boldsymbol{x}_k) = b_{\boldsymbol{x}_k}(\boldsymbol{x}_k) = \mu_{\boldsymbol{x}_k;\delta_{lk}}(\boldsymbol{x}_k)\mu_{\delta_{lk};\boldsymbol{x}_k}(\boldsymbol{x}_k) \\ &= p(\boldsymbol{x}_k) \prod_l \mu_{\delta_{lk};\boldsymbol{x}_k}(\boldsymbol{x}_k) \propto p(\boldsymbol{x}_k)\mathcal{N}(\boldsymbol{x}_k|\mathbf{m}_{\boldsymbol{y}|\boldsymbol{x}_k}, \mathbf{C}_{\boldsymbol{y}|\boldsymbol{x}_k}) \end{aligned} \tag{43}$$

where

$$\mathbf{C}_{\boldsymbol{y}|\boldsymbol{x}_k} = \left(\sum_l \mathbf{C}^{-1}_{\delta_{lk};\boldsymbol{x}_k}\right)^{-1}; \; \mathbf{m}_{\boldsymbol{y}|\boldsymbol{x}_k} = \mathbf{C}_{\boldsymbol{y}|\boldsymbol{x}_k}\left(\sum_l \mathbf{C}^{-1}_{\delta_{lk};\boldsymbol{x}_k}\mathbf{m}_{\delta_{lk};\boldsymbol{x}_k}\right) \tag{44}$$

Due to the block diagonal structure of the covariance matrix, we denote the t-th block as

$$\begin{aligned} \mathbf{C}_{\boldsymbol{y}|\boldsymbol{x}_{kt}} &= [\mathbf{C}_{\boldsymbol{y}|\boldsymbol{x}_k}]_{2(t-1)+1:2t,2(t-1)+1:2t} \\ \mathbf{m}_{\boldsymbol{y}|\boldsymbol{x}_{kt}} &= [\mathbf{m}_{\boldsymbol{y}|\boldsymbol{x}_k}]_{2(t-1)+1:2t} \end{aligned} \tag{45}$$

The symbol-wise posterior mean and covariance matrix for $\boldsymbol{x}_{kt}$ according to (43) are

$$\mathbf{m}_{b_{\boldsymbol{x},kt}} = \frac{1}{Z}\sum_i \mathbf{c}_i p(\boldsymbol{x}_{kt}=\mathbf{s}_i)\mathcal{N}(\mathbf{s}_i|\mathbf{m}_{\boldsymbol{y}|\boldsymbol{x}_{kt}}, \mathbf{C}_{\boldsymbol{y}|\boldsymbol{x}_{kt}})$$
$$\mathbf{C}_{b_{\boldsymbol{x},kt}} = \frac{1}{Z}\sum_i \mathbf{c}_i \mathbf{s}_i^\top p(\boldsymbol{x}_{kt}=\mathbf{s}_i)\mathcal{N}(\mathbf{s}_i|\mathbf{m}_{\boldsymbol{y}|\boldsymbol{x}_{kt}}, \mathbf{C}_{\boldsymbol{y}|\boldsymbol{x}_{kt}}) - \mathbf{m}_{b_{\boldsymbol{x},kt}}\mathbf{m}_{b_{\boldsymbol{x},kt}}^\top,$$

where $Z=\sum_i p(\boldsymbol{x}_{kt}=\mathbf{s}_i)\mathcal{N}(\mathbf{s}_i|\mathbf{m}_{\boldsymbol{y}|\boldsymbol{x}_{kt}}, \mathbf{C}_{\boldsymbol{y}|\boldsymbol{x}_{kt}})$.
Combine the symbol-wise means and covariance matrices into sequence-wise mean and covariance matrix

$$\begin{aligned}\mathbf{m}_{b_{\boldsymbol{x},k}} &= \left[\mathbf{m}_{b_{\boldsymbol{x},k1}}^\top, \ldots, \mathbf{m}_{b_{\boldsymbol{x},kT}}^\top\right]^\top \\ \mathbf{C}_{b_{\boldsymbol{x},k}} &= \text{blkdiag}(\mathbf{C}_{b_{\boldsymbol{x},k1}}, \ldots, \mathbf{C}_{b_{\boldsymbol{x},kT}})\end{aligned} \tag{46}$$

Thus, the posterior means and covariance matrices of the real and imaginary parts become

$$\begin{aligned}&\mathbf{m}_{b_{\boldsymbol{x},k,\Re}} = \boldsymbol{F}_1\mathbf{m}_{b_{\boldsymbol{x},k}};\ \mathbf{m}_{b_{\boldsymbol{x},k,\Im}} = \boldsymbol{F}_2\mathbf{m}_{b_{\boldsymbol{x},k}} \\ &\mathbf{C}_{b_{\boldsymbol{x},k,\Re\Re}} = \boldsymbol{F}_1\mathbf{C}_{b_{\boldsymbol{x},k}}\boldsymbol{F}_1^\top;\ \mathbf{C}_{b_{\boldsymbol{x},k,\Re\Im}} = \boldsymbol{F}_1\mathbf{C}_{b_{\boldsymbol{x},k}}\boldsymbol{F}_2^\top \\ &\mathbf{C}_{b_{\boldsymbol{x},k,\Im\Re}} = \boldsymbol{F}_2\mathbf{C}_{b_{\boldsymbol{x},k}}\boldsymbol{F}_1^\top;\ \mathbf{C}_{b_{\boldsymbol{x},k,\Im\Im}} = \boldsymbol{F}_2\mathbf{C}_{b_{\boldsymbol{x},k}}\boldsymbol{F}_2^\top\end{aligned} \tag{47}$$

where $\boldsymbol{F}_1 = \mathbf{I}_T \otimes \begin{bmatrix}1 & 0\end{bmatrix}$, $\boldsymbol{F}_2 = \mathbf{I}_T \otimes \begin{bmatrix}0 & 1\end{bmatrix}$. It is easy to see that $\mathbf{C}_{b_{\boldsymbol{x},k,\Re\Re}}$ is a $T\times T$ diagonal matrix where the T-th diagonal entry corresponds to the top-left elements of the T-th block matrix in the block diagonal matrix $\mathbf{C}_{b_{\boldsymbol{x},k}}$. Similarly, $\mathbf{C}_{b_{\boldsymbol{x},k,\Re\Im}}$, $\mathbf{C}_{b_{\boldsymbol{x},k,\Im\Re}}$ and $\mathbf{C}_{b_{\boldsymbol{x},k,\Im\Im}}$ correspond to the top-right, bottom left and bottom right elements of the T-th block matrix in $\mathbf{C}_{b_{\boldsymbol{x},k}}$ respectively.

C. Message from δ_{lk} to $\mathbf{h}_{lk}$

We consider the feedback message

$$\mu_{\delta_{lk};\mathbf{h}_{lk}}(\mathbf{h}_{lk}) = e^{\int b_{\boldsymbol{x},k}(\boldsymbol{x}_k)\ln \mu_{\mathbf{z}_{lk};\delta_{lk}}[\mathbf{B}(\boldsymbol{x}_k)\mathbf{h}_{lk}]d\boldsymbol{x}_k}, \tag{48}$$

where $\mathbf{B}(\boldsymbol{x}_k) = \boldsymbol{x}_{\Re,k}\otimes\mathbf{I}_{2M} + \boldsymbol{x}_{\Im,k}\otimes\mathbf{S}$. The expression (48) result to a Gaussian distribution of $\mathbf{h}_{lk}$. Due to the expectation (integral) operation, we get

$$\begin{aligned}&\mathbf{C}_{\delta_{lk};\mathbf{h}_{lk}} \\ &= \Big(\sum_t r_{b_{x,kt,\Re\Re}}\mathbf{C}_{\mathbf{z}_{lkt};\delta_{lk}}^{-1} + r_{b_{x,kt,\Im\Im}}\mathbf{S}^\top\mathbf{C}_{\mathbf{z}_{lkt};\delta_{lk}}^{-1}\mathbf{S} \\ &\quad r_{b_{x,kt,\Re\Im}}\mathbf{C}_{\mathbf{z}_{lkt};\delta_{lk}}^{-1}\mathbf{S} + r_{b_{x,kt,\Im\Re}}\mathbf{S}^\top\mathbf{C}_{\mathbf{z}_{lkt};\delta_{lk}}^{-1}\Big)^{-1},\end{aligned} \tag{49}$$

and the mean

$$\begin{aligned}&\mathbf{m}_{\delta_{lk};\mathbf{h}_{lk}} = \mathbf{C}_{\delta_{lk};\mathbf{h}_{lk}} \\ &\cdot\sum_t (m_{b_{x,kt,\Re}}\mathbf{I}_{2M} + m_{b_{x,kt,\Im}}\mathbf{S}^\top)\mathbf{C}_{\mathbf{z}_{lkt};\delta_{lk}}^{-1}\mathbf{m}_{\mathbf{z}_{lkt};\delta_{lk}}.\end{aligned} \tag{50}$$

D. Message from δ_{lk} to $\mathbf{z}_{lk}$

The vectorization of $\mathbf{Z}_{lk}$ can be expressed as

$$\mathbf{z}_{lk} = \boldsymbol{x}_{\Re,k}\otimes\mathbf{h}_{lk} + \boldsymbol{x}_{\Im,k}\otimes(\mathbf{S}\mathbf{h}_{lk}). \tag{51}$$

The posterior (belief) mean according to the relation (21) becomes

$$\mathbf{m}_{b_{\delta_{\mathbf{z},lk}}} = \mathbf{m}_{b_{\boldsymbol{x},k,\Re}}\otimes\mathbf{m}_{b_{\delta_{\mathbf{h},lk}}} + \mathbf{m}_{b_{\boldsymbol{x},k,\Im}}\otimes(\mathbf{S}\mathbf{m}_{b_{\delta_{\mathbf{h},lk}}}) \tag{52}$$

The correlation matrix of $\mathbf{z}_{lk}$ is:

$$\begin{aligned}\boldsymbol{R}_{b_{\delta_{\mathbf{z},lk}}} &= \boldsymbol{R}_{b_{\boldsymbol{x},k,\Re\Re}}\otimes\boldsymbol{R}_{b_{\delta_{\mathbf{h},lk}}} + \boldsymbol{R}_{b_{\boldsymbol{x},k,\Re\Im}}\otimes(\boldsymbol{R}_{b_{\delta_{\mathbf{h},lk}}}\mathbf{S}^\top) \\ &+ \boldsymbol{R}_{b_{\boldsymbol{x},k,\Im\Re}}\otimes(\mathbf{S}\boldsymbol{R}_{b_{\delta_{\mathbf{h},lk}}}) + \boldsymbol{R}_{b_{\boldsymbol{x},k,\Im\Im}}\otimes(\mathbf{S}\boldsymbol{R}_{b_{\delta_{\mathbf{h},lk}}}\mathbf{S}^\top)\end{aligned} \tag{53}$$

To make the algorithm trackable, we project $b_{\delta_{\mathbf{z},lk}}$ to Gaussian with block diagonal covariance matrix (block size $2M$).

$$\mathbf{C}_{b_{\delta_{\mathbf{z},lk}}} = \text{blkdiag}_{2M}[\boldsymbol{R}_{b_{\delta_{\mathbf{z},lk}}} - \mathbf{m}_{b_{\delta_{\mathbf{z},lk}}}\mathbf{m}_{b_{\delta_{\mathbf{z},lk}}}^\top], \tag{54}$$

where the operation $\text{blkdiag}_{2M}[\cdot] : \mathbb{R}^{2MT\times 2MT} \to \mathbb{R}^{2MT\times 2MT}$ keeps only the block matrices of size $2M$ along the diagonal while setting all the other elements to zero. Another treatment is to project to Gaussian with identical block diagonal matrices:

$$\mathbf{C}_{b_{\delta_{\mathbf{z},lk}}} = \mathbf{I}_T\otimes\frac{1}{T}\text{blktrace}_{2M}[\boldsymbol{R}_{b_{\delta_{\mathbf{z},lk}}} - \mathbf{m}_{b_{\delta_{\mathbf{z},lk}}}\mathbf{m}_{b_{\delta_{\mathbf{z},lk}}}^\top], \tag{55}$$

where $\text{blktrace}_{2M}[\cdot] : \mathbb{R}^{2MT\times 2MT} \to \mathbb{R}^{2M\times 2M}$ sums over all the block matrices of size $2M$ along the diagonal. Thus, the feedback message becomes

$$\mu_{\delta_{lk};\mathbf{z}_{lk}}(\mathbf{z}_{lk}) = \frac{\mathcal{N}(\mathbf{z}_{lk}|\mathbf{m}_{b_{\delta_{\mathbf{z},lk}}}, \mathbf{C}_{b_{\delta_{\mathbf{z},lk}}})}{\mathcal{N}(\mathbf{z}_{lk}|\mathbf{m}_{\mathbf{z}_{lk};\delta_{lk}}, \mathbf{C}_{\mathbf{z}_{lk};\delta_{lk}})}. \tag{56}$$

Note that the above division may lead to a message that cannot be normalized to a distribution.
Define the eigenvalue matrix $\boldsymbol{\Lambda}_{\delta_{lk};\mathbf{z}_{lk}}$ and unitary eigenvector $\mathbf{U}_{\delta_{lk};\mathbf{z}_{lk}}$ such that $\mathbf{C}_{\delta_{lk};\mathbf{z}_{lk}} = \mathbf{U}_{\delta_{lk};\mathbf{z}_{lk}}\boldsymbol{\Lambda}_{\delta_{lk};\mathbf{z}_{lk}}\mathbf{U}_{\delta_{lk};\mathbf{z}_{lk}}^\top$. We propose the following correction: for all $\lambda\in\text{diag}[\boldsymbol{\Lambda}_{\delta_{lk};\mathbf{z}_{lk}}]$, we clip λ^{-1} to the range $[10^{-8}, 10^{8}]$. Since we are using the iterative algorithm to find the fixed point of the BFE, resetting the value will not change the final result.

E. Message from $f_{\mathbf{h}_{lG_g}}$ to $\mathbf{h}_{lk}$ and Belief at $f_{\mathbf{h}_{lG_g}}$

The posterior belief at $f_{\mathbf{h}_{lG_g}}$ is

$$\begin{aligned}b_{f_{\mathbf{h}_{lG_g}}}(\mathbf{h}_{lG_g}) &= \mathcal{N}(\widetilde{\boldsymbol{y}}_{p,lg}|P\sigma_x^2\sum_{k'\in G_g}\mathbf{h}_{lk'}, \mathbf{C}_{v,p}) \\ &\prod_{k'\in G_g}\mu_{\mathbf{h}_{lk'};f_{\mathbf{h}_{lG_g}}}(\mathbf{h}_{lk'})p(\mathbf{h}_{lk'}),\end{aligned} \tag{57}$$

where $\mu_{\mathbf{h}_{lk};f_{\mathbf{h}_{lG_g}}}(\mathbf{h}_{lk}) = \mu_{\delta_{lk};\mathbf{h}_{lk}}(\mathbf{h}_{lk})$. The feedback message is

$$\begin{aligned}&\mu_{f_{\mathbf{h}_{lG_g}};\mathbf{h}_{lk}}(\mathbf{h}_{lk}) = \int b_{f_{\mathbf{h}_{lG_g}}}(\mathbf{h}_{lG_g})d\mathbf{h}_{l\bar{k}}/\mu_{f_{\mathbf{h}_{lG_g}};\mathbf{h}_{lk}}(\mathbf{h}_{lk}) \\ &= \int \mathcal{N}(\mathbf{h}_{lk}|\frac{\widetilde{\boldsymbol{y}}_{p,lg}}{P\sigma_x^2} - \sum_{k'\in G_g/\{k\}}\mathbf{h}_{lk'}, \frac{\mathbf{C}_{v,p}}{(P\sigma_x^2)^2}) \\ &\cdot\underbrace{\prod_{k'\in G_g/\{k\}}\mu_{\mathbf{h}_{lk'};f_{\mathbf{h}_{lG_g}}}(\mathbf{h}_{lk'})p(\mathbf{h}_{lk'})}_{\text{Gaussian interference distribution}}\,d\mathbf{h}_{l\bar{k}}\end{aligned} \tag{58}$$

We denote the interference plus noise mean and covariance as

$$\begin{aligned}\mathbf{C}_{\mathbf{h}_{l\bar{k}}} &= \frac{\mathbf{C}_{v,p}}{(P\sigma_x^2)^2} + \sum_{k'\in G_g/\{k\}}\mathbf{C}_{\mathbf{h}_{lk'}|\boldsymbol{y}} \\ \mathbf{m}_{\mathbf{h}_{l\bar{k}}} &= \sum_{k'\in G_g/\{k\}}\mathbf{m}_{\mathbf{h}_{lk'}|\boldsymbol{y}},\end{aligned} \tag{59}$$

where

$$\begin{aligned}\mathbf{C}_{\mathbf{h}_{lk}|\boldsymbol{y}} &= (\boldsymbol{\Xi}_{\mathbf{h}_{lk}}^{-1} + \mathbf{C}_{\mathbf{h}_{lk};f_{\mathbf{h}_{lG_g}}}^{-1})^{-1} \\ \mathbf{m}_{\mathbf{h}_{lk}|\boldsymbol{y}} &= \mathbf{C}_{\mathbf{h}_{lk}|\boldsymbol{y}}(\mathbf{C}_{\mathbf{h}_{lk};f_{\mathbf{h}_{lG_g}}}^{-1}\mathbf{m}_{\mathbf{h}_{lk};f_{\mathbf{h}_{lG_g}}}).\end{aligned} \tag{60}$$

Due to the Gaussian reproduction lemma [], the feedback message can be obtained as

$$\begin{aligned}\mathbf{C}_{f_{\mathbf{h}_{lG_g}};\mathbf{h}_{lk}} &= (\mathbf{C}_{\mathbf{h}_{l\bar{k}}}^{-1} + \boldsymbol{\Xi}_{\mathbf{h}_{lk}}^{-1})^{-1} \\ \mathbf{m}_{f_{\mathbf{h}_{lG_g}};\mathbf{h}_{lk}} &= \mathbf{C}_{f_{\mathbf{h}_{lG_g}};\mathbf{h}_{lk}}\left[\mathbf{C}_{\mathbf{h}_{l\bar{k}}}^{-1}\left(\frac{\widetilde{\boldsymbol{y}}_{p,lg}}{P\sigma_x^2} - \mathbf{m}_{\mathbf{h}_{l\bar{k}}}\right)\right].\end{aligned} \tag{61}$$

Algorithm 1 Proposed Method

Require: $\forall l, g, k,\ \tilde{\boldsymbol{y}}_{p,lg},\ \boldsymbol{y}_l,\ p(\boldsymbol{x}_k),\ p(\mathbf{h}_{lk}),\ p(\boldsymbol{y}_l|\mathbf{z}_{l1},\ldots,\mathbf{z}_{lK})$

Initialize: All the factor-to-variable messages are set to zero mean and unit covariance matrices.
[$\forall l$, At AP l, execute the following loop]
repeat [$\forall l' \in N(l)k, g$]
    $\mathbf{C}_{f_{\mathbf{h}_{lG_g}};\mathbf{h}_{lk}}$ and $\mathbf{m}_{f_{\mathbf{h}_{lG_g}};\mathbf{h}_{lk}}$ by (60) $\to$ (59) $\to$(61)
    $\mathbf{C}_{f_{\boldsymbol{y}_l};\mathbf{z}_{lk}}$ and $\mathbf{m}_{f_{\boldsymbol{y}_l};\mathbf{z}_{lk}}$ by (65)
    $\mathbf{C}_{b_{\delta_{\mathbf{h},lk}}}$ and $\mathbf{m}_{b_{\delta_{\mathbf{h},lk}}}$ by (62)
    $\boldsymbol{R}_{b_{\delta_{\mathbf{h},lk}}} = \mathbf{C}_{b_{\delta_{\mathbf{h},lk}}} + \mathbf{m}_{b_{\delta_{\mathbf{h},lk}}}\mathbf{m}^{\mathsf{T}}_{b_{\delta_{\mathbf{h},lk}}}$
    $\mathbf{C}_{\delta_{lk};\boldsymbol{x}_k}$ and $\mathbf{m}_{\delta_{lk};\boldsymbol{x}_k}$ by (40) $\to$ (39) $\to$ (42) $\to$ (41) $\to$ (38)
    Compute $\mathbf{m}_{b_{\boldsymbol{x},k,\Re}}$, $\mathbf{m}_{b_{\boldsymbol{x},k,\Im}}$, $\mathbf{C}_{b_{\boldsymbol{x},k,\Re\Re}}$, $\mathbf{C}_{b_{\boldsymbol{x},k,\Re\Im}}$, $\mathbf{C}_{b_{\boldsymbol{x},k,\Im\Re}}$, $\mathbf{C}_{b_{\boldsymbol{x},k,\Im\Im}}$ via (45)-(47)
    $\mathbf{C}_{\delta_{lk};\mathbf{h}_{lk}}$ and $\mathbf{m}_{\delta_{lk};\mathbf{h}_{lk}}$ by (49) and (50)
    $\mathbf{C}_{\delta_{lk};\mathbf{z}_{lk}}$ and $\mathbf{m}_{\delta_{lk};\mathbf{z}_{lk}}$ via (52) $\to$ (53) $\to$ {(54) or (55)} $\to$ (56)
until Convergence

According to relation (58), the mean and covariance matrix of the belief are

$$\mathbf{C}_{b_{\delta_{\mathbf{h},lk}}} = \left(\mathbf{C}^{-1}_{f_{\mathbf{h}_{lG_g}};\mathbf{h}_{lk}} + \mathbf{C}^{-1}_{\delta_{lk};\mathbf{h}_{lk}}\right)^{-1}$$
$$\mathbf{m}_{b_{\delta_{\mathbf{h},lk}}} = \mathbf{C}_{b_{\delta_{\mathbf{h},lk}}}\left(\mathbf{C}^{-1}_{f_{\mathbf{h}_{lG_g}};\mathbf{h}_{lk}}\mathbf{m}_{f_{\mathbf{h}_{lG_g}};\mathbf{h}_{lk}} + \mathbf{C}^{-1}_{\delta_{lk};\mathbf{h}_{lk}}\mathbf{m}_{\delta_{lk};\mathbf{h}_{lk}}\right) \tag{62}$$

F. Message from $f_{\boldsymbol{y}_l}$ to $\mathbf{z}_{lk}$

The belief at $f_{\boldsymbol{y}_l}$ is

$$b_{f_{\boldsymbol{y}_l}}(\mathbf{z}_{l\{k\}}) = \mathcal{N}(\boldsymbol{y}_l|\sum_k \mathbf{z}_{lk}, \mathbf{C}_v)\prod_k \mu_{\mathbf{z}_{lk};f_{\boldsymbol{y}_l}}(\mathbf{z}_{lk}), \tag{63}$$

where $\mu_{\mathbf{z}_{lk};f_{\boldsymbol{y}_l}}(\mathbf{z}_{lk}) = \mu_{\delta_{lk};\mathbf{z}_{lk}}(\mathbf{z}_{lk})$. The feedback message is then obtained by

$$\mu_{f_{\boldsymbol{y}_l};\mathbf{z}_{lk}}(\mathbf{z}_{lk}) = \frac{\int b_{f_{\boldsymbol{y}_l}}(\mathbf{z}_{l\{k\}})d\mathbf{z}_{l\overline{k}}}{\mu_{\mathbf{z}_{lk};f_{\boldsymbol{y}_l}}(\mathbf{z}_{lk})} \tag{64}$$

which can be computed to be a Gaussian with mean and covariance matrix:

$$\mathbf{m}_{f_{\boldsymbol{y}_l};\mathbf{z}_{lk}} = \boldsymbol{y}_l - \sum_{k'\neq k}\mathbf{m}_{\mathbf{z}_{lk};f_{\boldsymbol{y}_l}};\ \mathbf{C}_{f_{\boldsymbol{y}_l};\mathbf{z}_{lk}} = \mathbf{C}_v + \sum_{k'\neq k}\mathbf{C}_{\mathbf{z}_{lk};f_{\boldsymbol{y}_l}}. \tag{65}$$

For the algorithm with covariance averaging, the overall complexity at every AP is $O(M^3K + M^2KT)$. While the CPU has a complexity of $O(LKT + KT|\mathcal{S}|)$. For algorithms using (54) instead of (55), the complexity at the APs will be $O(M^3KT)$.

V. Simulation Results

In this section, we verify the algorithm using numerical simulations. We consider a $400m \times 400m$ area with $M = 16$ APs and $K = 8$ UTs. The APs are located at the coordinates $(\frac{400}{3}i, \frac{400}{3}j)$, where $i, j \in \{0, \ldots, 3\}$. The UTs are uniformly randomly distributed over this area. The fading model we use is [6],

$$\sigma^2_{l,k}[\text{dB}] = -30.5 - 36.7\log_{10}(d_{lk}), \tag{66}$$

where d_{lk} is the distance between AP l and UT k. To induce pilot contamination, the pilot sequence length is set to $P = 4$. Furthermore, the pilots are randomly assigned to the users. We utilize 4-QAM and 16-QAM modulation schemes to generate $\mathbf{X}$, with the data length set to $T = 16$. To ensure fairness in the simulations, power control is applied for each UT, maintaining equal total received power across all users. Our proposed method is compared to an alternative BFE-based approach derived under circular complex Gaussian constraints. In Genie-Aided scenarios, we assume the user data to be known for performance benchmarking.

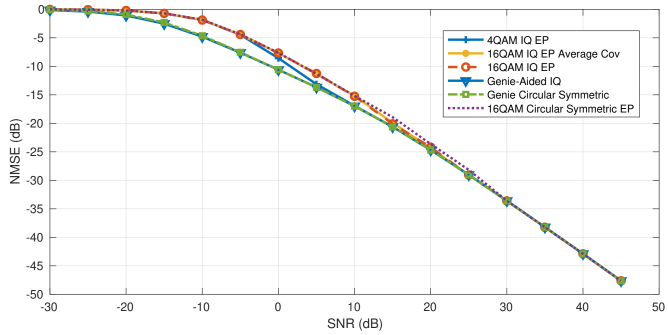

Fig. 1. NMSE vs SNR

Compared to the complex version of EP [12] with a complexity of $O[(M^3 + |\mathcal{S}|)KT]$ at each AP, the complexity of our proposed method with covariance averaging is $O(M^3K + M^2KT)$.

VI. Conclusions

We propose a BFE-constrained optimization algorithm that separates complex quantities into in-phase and quadrature components, enabling it to handle a wider range of messages for approximating the true posterior. Furthermore, the complexity of the algorithm is effectively managed by employing averaged covariance matrices.

Acknowledgements EURECOM's research is partially supported by its industrial members: ORANGE, BMW, SAP, iABG, Norton LifeLock, and by the Franco-German project CellFree6G.

References

[1] R. Gholami, L. Cottatellucci, and D. Slock, "Tackling Pilot Contamination in Cell-Free Massive MIMO by Joint Channel Estimation and Linear Multi-User Detection," in *IEEE International Symposium on Information Theory (ISIT)*, 2021.
[2] T. P. Minka, "A Family of Algorithms for Approximate Bayesian Inference," Ph.D. dissertation, Massachusetts Institute of Technology, 2001.
[3] M. J. Wainwright, M. I. Jordan *et al.*, "Graphical Models, Exponential Families, and Variational Inference," *Foundations and Trends® in Machine Learning*, vol. 1, no. 1–2, pp. 1–305, 2008.
[4] R. Gholami, L. Cottatellucci, and D. Slock, "Message Passing for a Bayesian Semi-Blind Approach to Cell-Free Massive MIMO," in *Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2021.
[5] Z. Zhao and D. Slock, "Bilinear Hybrid Expectation Naximization and Expectation Propagation for Semi-Blind Channel Estimation," in *European Signal Processing Conference (EUSIPCO)*, 2024.
[6] A. Karataev, C. Forsch, and L. Cottatellucci, "Bilinear Expectation Propagation for Distributed Semi-Blind Joint Channel Estimation and Data Detection in Cell-Free Massive MIMO," *IEEE Open Journal of Signal Processing*, 2024.
[7] J. T. Parker, P. Schniter, and V. Cevher, "Bilinear Generalized Approximate Message Passing—Part I: Derivation," *IEEE Transactions on Signal Processing*, 2014.
[8] Z. Zhao and D. Slock, "Decentralized Expectation Propagation for Semi-Blind Channel Estimation in Cell-Free Networks," 2024. [Online]. Available: https://arxiv.org/abs/2410.01303
[9] D. Zhang, X. Song, W. Wang, G. Fettweis, and X. Gao, "Unifying Message Passing Algorithms Under the Framework of Constrained Bethe Free Energy Minimization," *IEEE Transactions on Wireless Communications*, 2021.
[10] H. Jiang, X. Yuan, and Q. Guo, "Hybrid Vector Message Passing for Generalized Bilinear Factorization," *arXiv preprint arXiv:2401.03626*, 2024.
[11] T. Heskes, M. Opper, W. Wiegerinck, O. Winther, and O. Zoeter, "Approximate Inference Techniques with Expectation Constraints," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2005, no. 11, 2005.
[12] Z. Zhao and D. Slock, "Decentralized Message-Passing for Semi-Blind Channel Estimation in Cell-Free Systems Based on Bethe free energy Optimization," in *Asilomar Conference on Signals, Systems, and Computers*, 2024.