Expectation Propagation based Analysis of 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 the uplink communication in cell-free (CF) massive multiple-input multiple-output (MaMIMO) systems. One of the major problems in CF MaMIMO systems is pilot contamination, i.e., the number of users is larger than the length of the pilot sequences. Researchers have proposed a semi-blind approach to overcome this problem. The semi-blind joint estimation of the channel and the data leads to a bilinear problem. We propose a vector-level Expectation Propagation (EP) method to avoid the high-dimensional intractable integrals of Bayesian posterior distributions and estimate both the multi-user data sequences and channels. We exploit the property of orthogonal pilots to decompose the pilot measurement into multiple smaller, mutually independent equivalent measurements. We also assume the data sequences to be randomly chosen from a codebook (possibly with structure) of finite alphabet codewords.**

I. Introduction

One of the unique features of Cell-Free (CF) Massive MIMO (MaMIMO) networks is that all the user terminals (UTs) are served by all the access points (APs) in a given area. This leads to the problem of pilot contamination, where the number of UTs exceeds the length of pilot sequences. To address this issue, Semi-Blind approaches have been explored as a viable solution to mitigate the effects of pilot contamination, as detailed in the work of Gholami et al. [1].

In the context of Bayesian inference, the Semi-Blind approach is modeled as a bilinear inference problem, where the APs must concurrently estimate both the channel state information (CSI) and the user signals. [2] have demonstrated that transitioning from the deterministic domain to the Bayesian domain results in significant gains. The challenge in addressing the bilinear inference problem primarily arises from obtaining the closed-form solution of the marginal likelihood of the observed data given the channel coefficient.

A. Prior Works

Message-passing algorithms play a critical role in the inference problems. Notably, one of the most potent message-passing algorithms is Expectation Propagation (EP), proposed by Minka [3]. It transforms the global inference problem into local inference problems and approximates complicated factors with simple factors (e.g., Gaussian distributions) to make high-dimensional integrals tractable. Belief Propagation (BP) can be viewed as a special case of EP without factor approximation. The close relation between EP and Bethe Free Energy (BFE) was studied in [4], which shows that they share the same fixed points. Given that the primary concern with EP is the uncertainty of convergence, a more robust solution can be achieved by directly analyzing the optimization of the BFE. A loop-free EP was proposed in [5] based on approximate BFE optimization. In [6], the authors proposed a Bilinear EP algorithm for non-coherent channel detection based on the codebook.

B. Main Contributions

In this paper, we consider a semi-blind system with orthogonal pilot sequences (though the same sequence may be reused by several users, leading to pilot contamination). We assume a structured prior distribution for the data that can encompass various channel coding approaches or simply exploit finite alphabet symbols. To estimate the channel and data jointly in this semi-blind system, we propose a vector-level EP where we treat the channel coefficients and data (sub-) sequences for each user as atomic variables. Since orthogonal pilots are used, we first combine the received pilot signals with the channel priors to form augmented channel priors, thereby reducing the number of factors. If the data prior corresponds to the use of codebooks with a certain structure, it can be exploited to reduce the computational load of the proposed method.

II. System Model

We examine the uplink semi-blind signal model

$$\mathbf{Y} = \begin{bmatrix}\mathbf{Y}_p & \mathbf{Y}_d\end{bmatrix} = \mathbf{H}\begin{bmatrix}\mathbf{X}_p^T & \mathbf{X}^T\end{bmatrix} + \begin{bmatrix}\mathbf{V}_p & \mathbf{V}_d\end{bmatrix} \in \mathbb{C}^{M\times(P+L)},$$

where $\mathbf{H}$ represents the uplink channel matrix from UTs to APs, which is unknown and modeled as an independent and identically distributed (i.i.d.) random matrix of size $M \times K$. Each column follows the distribution $\mathbf{h}_k \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Xi}_{\mathbf{h}_k})$. The input signal is composed of a pilot part $\mathbf{X}_p \in \mathbb{C}^{P\times K}$ and a data part $\mathbf{X} \in \mathbb{C}^{L\times K}$. Each column of $\mathbf{X}_p$ or $\mathbf{X}$ represents the pilot sequence $\boldsymbol{x}_{p,k}$ or data sequence $\boldsymbol{x}_k$ transmitted by user k. It's assumed that orthogonal pilot sequences are used, and each pilot sequence has power $\sigma_x^2 P$.

We define $\mathbf{V}_p$ and $\mathbf{V}_d$ as the Additive White Gaussian Noise (AWGN) present at the APs. Each element within these noise matrices is assumed to be independently drawn from a Gaussian distribution, specifically $\mathcal{CN}(0, \sigma_v^2)$.

A. Benefit of using Orthogonal Pilots

We correlate the received signal $\mathbf{Y}_p$ with the g-th pilot sequence

$$\mathbf{Y}_p \overline{\boldsymbol{x}}_{p,g}^* = \boldsymbol{y}_{p,g} = \sigma_x^2 P \mathbf{H}_g \mathbf{1}_{K_g} + \boldsymbol{v}_g = \sigma_x^2 P \mathbf{A}_{p,g}^H \mathbf{h}_g + \boldsymbol{v}_g, \quad (1)$$

where $\overline{\boldsymbol{x}}_{p,g}^*$ denotes the conjugated g-th pilot sequence, and each column of $\mathbf{H}_g$ denotes the channel coefficients from a user using the g-th pilot sequence. We use set G_g to denote the group of users using the g-th pilot. We use K_g to represent the size of G_g. The equivalent noise $\boldsymbol{v}_g = \mathbf{V}_p \overline{\boldsymbol{x}}_{p,g}^*$ is still i.i.d. zero mean Gaussian noise with covariance $\mathbf{C}_{\boldsymbol{v}_g} = \sigma_x^2 P \sigma_v^2 \mathbf{I}$. The last equal sign in (1) is obtained by vectorizing $\mathbf{H}_g$, where $\mathbf{A}_{p,g} = \mathbf{1}_{K_g} \otimes \mathbf{I}_M$ and $\mathbf{h}_g = \text{vec}(\mathbf{H}_g)$. Since orthogonal pilots are used, we know $\mathbb{E}[\boldsymbol{y}_{p,g}\boldsymbol{y}_{p,g'}^H] = \mathbf{0}$ if $g \neq g'$. The corresponding equivalent likelihood model for the received pilot is

$$p(\mathbf{Y}_p|\mathbf{H}) = \prod_g p(\boldsymbol{y}_{p,g}|\mathbf{h}_g) = \prod_g \mathcal{CN}(\boldsymbol{y}_{p,g}|\sigma_x^2 P \mathbf{A}_{p,g}^H \mathbf{h}_g, \mathbf{C}_{\boldsymbol{v}_g}).$$

B. Structured User Data Prior Distribution

We assume that the symbols in the data sequences are discrete and that data sequences are independent between the users. However, the symbols in the data sequence of one user are not independent. We assume that for all users, the code structure is identically modeled as

$$p(\boldsymbol{x}_k) \propto \prod_\alpha f_\alpha(\boldsymbol{x}_{k,\alpha}). \tag{2}$$

We partition the vector $\boldsymbol{x}_k$ into subvectors $\boldsymbol{x}_{k,\beta}$ such that:

1) : For all β, α, either $\boldsymbol{x}_{k,\beta} \subseteq \boldsymbol{x}_{k,\alpha}$ or $\boldsymbol{x}_{k,\beta} \cap \boldsymbol{x}_{k,\alpha} = \phi$.

2) : For all β, β', $\boldsymbol{x}_{k,\beta} \cap \boldsymbol{x}_{k,\beta'} = \phi$.

3) : $\cup_\beta \boldsymbol{x}_{k,\beta} = \boldsymbol{x}_k$.

In the above three statements, we abuse the notations and use $\boldsymbol{x}_{k,\beta}$, $\boldsymbol{x}_{k,\beta'}$, $\boldsymbol{x}_{k,\alpha}$, and $\boldsymbol{x}_k$ to denote the sets of entries contained by those vectors respectively. We note that α denotes a factor level partition while β denotes a variable level partition in the symbol time domain. For example, the probabilistic model $p(\boldsymbol{x}_{\beta_1}, \boldsymbol{x}_{\beta_2}, \boldsymbol{x}_{\beta_3}) = p(\boldsymbol{x}_{\beta_3}|\boldsymbol{x}_{\beta_1}, \boldsymbol{x}_{\beta_2}) p(\boldsymbol{x}_{\beta_1}, \boldsymbol{x}_{\beta_2})$, where $\boldsymbol{x}_{\alpha_1} = [\boldsymbol{x}_{\beta_1}\boldsymbol{x}_{\beta_2}\boldsymbol{x}_{\beta_3}]^T$, $\boldsymbol{x}_{\alpha_2} = [\boldsymbol{x}_{\beta_1}\boldsymbol{x}_{\beta_2}]^T$, can be used to describe codes with the parity-check component $\boldsymbol{x}_{\beta_3}$.

C. Probablistic Model and Factorization

To handle the bilinear mixing, we introduce an auxiliary variable $\mathbf{Z}_{k,\beta} = \mathbf{h}_k \boldsymbol{x}_{k,\beta}^T$ and consequently its vectorization $\mathbf{z}_{k,\beta} = (\boldsymbol{x}_{k,\beta} \otimes \mathbf{I}_M)\mathbf{h}_k$ where $\mathbf{z}_{k,\beta} := \text{vec}(\mathbf{Z}_{k,\beta})$. The joint probabilistic model of the uplink transmission can be factorized as follows:

$$p(\boldsymbol{x}_{\{k\}}, \mathbf{h}_{\{k\}}, \mathbf{z}_{\{k,\beta\}}, \mathbf{Y}) \propto \prod_\beta p(\boldsymbol{y}_{d,\beta}|\mathbf{z}_{\{k\},\beta}) \prod_g p(\mathbf{h}_g|\boldsymbol{y}_{p,g})$$
$$\prod_k \prod_\beta p(\mathbf{z}_{k,\beta}|\mathbf{h}_k, \boldsymbol{x}_{k,\beta}) \cdot \prod_k \prod_\alpha f_\alpha(\boldsymbol{x}_{k,\alpha}),$$

where we denote $\boldsymbol{y}_d = \text{vec}(\mathbf{Y}_d)$, and $p(\mathbf{h}_g|\boldsymbol{y}_{p,g}) \propto p(\boldsymbol{y}_{p,g}|\mathbf{h}_g) \prod_{k \in G_g} p(\mathbf{h}_k)$. The partitioned $\mathbf{Y}_{d,\beta}$ is composed of the columns in $\mathbf{Y}_d$ corresponding to the same symbol time intervals as $\boldsymbol{x}_{k,\beta}$, and we denote its vectorization as $\boldsymbol{y}_{d,\beta} = \text{vec}(\mathbf{Y}_{d,\beta})$.

For simplicity, we denote the factors as:

$$\Psi_{0,g} := p(\mathbf{h}_g|\boldsymbol{y}_{p,g}); \ \Psi_{1,k,\beta} := p(\mathbf{z}_{k,\beta}|\mathbf{h}_k, \boldsymbol{x}_{k,\beta})$$
$$\Psi_{2,\beta} := p(\boldsymbol{y}_{d,\beta}|\sum_k \mathbf{z}_{k,\beta}); \ \Psi_{3,k,\alpha} := p(\boldsymbol{x}_{k,\alpha})$$

We conclude the factorization in factor graph fig. 1.

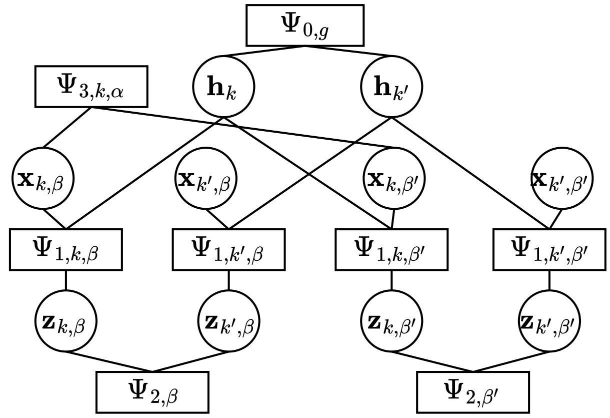

Fig. 1. Part of the factor graph, showing users k, k' and partition variables $\beta, \beta' \in \alpha$ (the data prior is shown only for user k).

III. Expectation Propagation

Expectation Propagation is a method to approximate the factored pdf $p(\boldsymbol{\theta})$ by another pdf $b(\boldsymbol{\theta})$ with factors of the desired family [6],

$$p(\boldsymbol{\theta}) \propto \prod_a \Psi_a(\boldsymbol{\theta}_a) \simeq b(\boldsymbol{\theta}) \propto \prod_a q_a(\boldsymbol{\theta}_a),$$

where we assume all the factors q_a are fully factorized [7], i.e., q_a can be factored at variable level $q_a(\boldsymbol{\theta}_a) \propto \prod_{\boldsymbol{\theta}_i \subseteq \boldsymbol{\theta}_a} m_{\Psi_a;\boldsymbol{\theta}_i}(\boldsymbol{\theta}_i)$. Here, $\boldsymbol{\theta}_i$ denotes a variable-level partition of the whole $\boldsymbol{\theta}$, i.e., for all variable-level partitions indices i, j and factor index a, if $i \neq j$, we have $\boldsymbol{\theta}_i \cap \boldsymbol{\theta}_j = \phi$, and the intersection $\boldsymbol{\theta}_i \cap \boldsymbol{\theta}_a$ is either $\boldsymbol{\theta}_i$ or ϕ. With the fully-factorizable assumption, EP can be interpreted as message passing by iteratively updating the factor to variable message $m_{\Psi_a;\boldsymbol{\theta}_i}$ and variable to factor message $m_{\boldsymbol{\theta}_i;\Psi_a}$: [8]

$$\begin{aligned} m_{\boldsymbol{\theta}_i;\Psi_a}(\boldsymbol{\theta}_i) &= \prod_{a' \in N(i)/\{a\}} m_{\Psi_{a'};\boldsymbol{\theta}_i}(\boldsymbol{\theta}_i), \\ m_{\Psi_a;\boldsymbol{\theta}_i}(\boldsymbol{\theta}_i) &= \frac{\text{proj}(b_{\Psi_a}(\boldsymbol{\theta}_i))}{m_{\boldsymbol{\theta}_i;\Psi_a}(\boldsymbol{\theta}_i)}, \end{aligned} \tag{3}$$

where $N(i)$ denotes the direct neighbor of the variable $\boldsymbol{\theta}_i$, and the operation $\text{proj}(\cdot)$ projects a distribution to a family F by optimizing a Kullback-Leibler divergence $\text{proj}(p) = \arg\min_{q \in F} KLD(p\|q)$. Here, $b_{\Psi_a}(\boldsymbol{\theta}_i)$ is the marginalization of the belief (approximated posterior) $b_{\Psi_a}(\boldsymbol{\theta}_a)$ at Ψ_a:

$$b_{\Psi_a}(\boldsymbol{\theta}_i) = \int b_{\Psi_a}(\boldsymbol{\theta}_a) d\boldsymbol{\theta}_{\bar{i}} = \int \Psi_a(\boldsymbol{\theta}_a) \prod_{i' \in N(a)} m_{\boldsymbol{\theta}_{i'};\Psi_a}(\boldsymbol{\theta}_{i'}) d\boldsymbol{\theta}_{\bar{i}},$$

where we use $\boldsymbol{\theta}_{\bar{i}}$ to denote all elements in $\boldsymbol{\theta}_a$ except the ones contained in $\boldsymbol{\theta}_i$ and $N(a)$ to denote the variables neighboring factor a. The messages appearing in this paper are all normalized to 1.

BP can be interpreted as a special case of EP by omitting the projection step in (3).

In the following, if a message distribution $m_{\boldsymbol{\theta}_i;\Psi_a}(\boldsymbol{\theta}_i)$ (or $m_{\Psi_a;\boldsymbol{\theta}_i}(\boldsymbol{\theta}_i)$) is Gaussian, we denote its mean and covariance matrix as $\boldsymbol{\mu}_{\boldsymbol{\theta}_i;\Psi_a}$ and $\mathbf{C}_{\boldsymbol{\theta}_i;\Psi_a}$ (or $\boldsymbol{\mu}_{\Psi_a;\boldsymbol{\theta}_i}$ and $\mathbf{C}_{\Psi_a;\boldsymbol{\theta}_i}$) respectively.

IV. MESSAGE FROM BILINEAR FACTOR NODE $\Psi_{1,k,\beta}$

We first look at the bilinear factor node $\Psi_{1,k,\beta}$. It is connected to three variable nodes. The extrinsic distributions to this node are calculated via line 10-12 in Algorithm 1. Following the EP rule, we need to compute the belief of each variable, approximate it to a certain desired distribution family, and compute the feedback message to each of the connected variable nodes. Rewrite the factor

$$\Psi_{1,k,\beta} = \delta(\mathbf{Z}_{k,\beta} - \mathbf{h}_k \boldsymbol{x}_{k,\beta}^T) = \delta(\mathbf{z}_{k,\beta} - \mathbf{A}_{k,\beta}\mathbf{h}_k),$$

where $\mathbf{A}_{k,\beta} = \boldsymbol{x}_{k,\beta} \otimes \mathbf{I}_M$. We denote the joint belief (normalized distribution) at factor node $\Psi_{1,k}$ as

$$\begin{aligned} &b_{\Psi_{1,k,\beta}}(\mathbf{z}_{k,\beta}, \mathbf{h}_k, \boldsymbol{x}_{k,\beta}) \propto \delta(\mathbf{z}_{k,\beta} - \mathbf{A}_{k,\beta}\mathbf{h}_k) \\ &\cdot m_{\mathbf{z}_{k,\beta};\Psi_{1,k,\beta}}(\mathbf{z}_{k,\beta}) m_{\mathbf{h}_k;\Psi_{1,k,\beta}}(\mathbf{h}_k) m_{\boldsymbol{x}_{k,\beta};\Psi_{1,k,\beta}}(\boldsymbol{x}_{k,\beta}), \end{aligned} \tag{4}$$

which can be interpreted as the joint posterior distribution. Since the channel coefficients follow a continuous distribution while the data symbols follow a discrete distribution, the marginalization of the channel coefficients results in an integral operation. In contrast, the marginalization of the data symbols involves only a summation over finite support. Therefore, in the following, we use EP for $\mathbf{h}_k$ and $\mathbf{z}_{k,\beta}$ but BP for $\boldsymbol{x}_{k,\beta}$.

A. Message from $\Psi_{1,k,\beta}$ to $\mathbf{h}_k$

The normalized marginal belief of $\mathbf{h}_k$ at factor $\Psi_{1,k,\beta}$ can be obtained via

$$\begin{aligned} b_{\Psi_{1,k,\beta}}(\mathbf{h}_k) &= \sum_{\boldsymbol{x}_{k,\beta}} \int b_{\Psi_{1,k,\beta}}(\mathbf{z}_{k,\beta}, \mathbf{h}_k, \boldsymbol{x}_{k,\beta}) d\mathbf{z}_{k,\beta} \\ &= \sum_{\boldsymbol{x}_{k,\beta}} \mathcal{CN}(\mathbf{h}_k | \boldsymbol{\mu}_{\widehat{\mathbf{h}}_k|\boldsymbol{x}_{k,\beta}}, \mathbf{C}_{\widehat{\mathbf{h}}_k|\boldsymbol{x}_{k,\beta}}) \omega(\boldsymbol{x}_{k,\beta}), \end{aligned} \tag{5}$$

where

$$\begin{aligned} &\mathbf{C}_{\widehat{\mathbf{h}}_k|\boldsymbol{x}_{k,\beta}} = \left(\mathbf{A}_{k,\beta}^H \mathbf{C}_{\mathbf{z}_{k,\beta};\Psi_{1,k,\beta}}^{-1} \mathbf{A}_{k,\beta} + \mathbf{C}_{\mathbf{h}_k;\Psi_{1,k,\beta}}^{-1}\right)^{-1}; \\ &\boldsymbol{\mu}_{\widehat{\mathbf{h}}_k|\boldsymbol{x}_{k,\beta}} = \mathbf{C}_{\widehat{\mathbf{h}}_k|\boldsymbol{x}_{k,\beta}} \\ &\cdot \left(\mathbf{A}_{k,\beta}^H \mathbf{C}_{\mathbf{z}_{k,\beta};\Psi_{1,k,\beta}}^{-1} \boldsymbol{\mu}_{\mathbf{z}_{k,\beta};\Psi_{1,k,\beta}} + \mathbf{C}_{\mathbf{h}_k;\Psi_{1,k,\beta}}^{-1} \boldsymbol{\mu}_{\mathbf{h}_k;\Psi_{1,k,\beta}}\right) \\ &\omega(\boldsymbol{x}_{k,\beta}) = \frac{1}{Z_b} m_{\boldsymbol{x}_{k,\beta};\Psi_{1,k,\beta}}(\boldsymbol{x}_{k,\beta}) \nu(\boldsymbol{x}_{k,\beta}) \\ &Z_b = \sum_{\boldsymbol{x}_{k,\beta}} m_{\boldsymbol{x}_{k,\beta};\Psi_{1,k,\beta}}(\boldsymbol{x}_{k,\beta}) \nu(\boldsymbol{x}_{k,\beta}) \\ &\nu(\boldsymbol{x}_{k,\beta}) = \mathcal{CN}(\mathbf{0} | \boldsymbol{\mu}_{\mathbf{z}_{k,\beta};\Psi_{1,k,\beta}} - \mathbf{A}_{k,\beta} \boldsymbol{\mu}_{\mathbf{h}_k;\Psi_{1,k,\beta}}, \\ &\mathbf{C}_{\mathbf{z}_{k,\beta};\Psi_{1,k,\beta}} + \mathbf{A}_{k,\beta} \mathbf{C}_{\mathbf{h}_k;\Psi_{1,k,\beta}} \mathbf{A}_{k,\beta}^H). \end{aligned} \tag{6}$$

The result in (5)-(6) can be derived by noticing the matrix identity

$$\mathbf{D}(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{C}^{-1}(\mathbf{D}\mathbf{A}^{-1}\mathbf{B} + \mathbf{C}^{-1})^{-1}\mathbf{D}\mathbf{A}^{-1}.$$

From the derivation, it is worth noticing that the ratio difference between the LHS and RHS of (4) is Z_b. The mean and covariance matrix of the marginal posterior $b_{\Psi_{1,k,\beta}}(\mathbf{h}_k)$ are given by

$$\begin{aligned} &\boldsymbol{\mu}_{\widehat{\mathbf{h}}_k} = \sum_{\boldsymbol{x}_{k,\beta}} \boldsymbol{\mu}_{\widehat{\mathbf{h}}_k|\boldsymbol{x}_{k,\beta}} \omega(\boldsymbol{x}_{k,\beta}); \\ &\mathbf{C}_{\widehat{\mathbf{h}}_k} = \sum_{\boldsymbol{x}_{k,\beta}} (\mathbf{C}_{\widehat{\mathbf{h}}_k|\boldsymbol{x}_{k,\beta}} + \boldsymbol{\mu}_{\widehat{\mathbf{h}}_k|\boldsymbol{x}_{k,\beta}} \boldsymbol{\mu}_{\widehat{\mathbf{h}}_k|\boldsymbol{x}_{k,\beta}}^H) \omega(\boldsymbol{x}_{k,\beta}) - \boldsymbol{\mu}_{\widehat{\mathbf{h}}_k} \boldsymbol{\mu}_{\widehat{\mathbf{h}}_k}^H. \end{aligned} \tag{7}$$

From this point, it is clear that the marginalized belief in (5) is a Gaussian mixture model, whose mean and covariance matrix are calculated by (7). The approximated marginal belief can be derived by projecting the marginal belief (5) into a Gaussian with mean and covariance obtained in (7). Consequently, the message from $\Psi_{1,k,\beta}$ to $\mathbf{h}_k$ is updated by

$$m_{\Psi_{1,k,\beta};\mathbf{h}_k}(\mathbf{h}_k) = \mathcal{CN}(\mathbf{h}_k | \boldsymbol{\mu}_{\Psi_{1,k,\beta};\mathbf{h}_k}, \mathbf{C}_{\Psi_{1,k,\beta};\mathbf{h}_k}), \tag{8}$$

where

$$\begin{aligned} &\mathbf{C}_{\Psi_{1,k,\beta};\mathbf{h}_k} = (\mathbf{C}_{\widehat{\mathbf{h}}_k}^{-1} - \mathbf{C}_{\mathbf{h}_k;\Psi_{1,k,\beta}}^{-1})^{-1} \\ &\boldsymbol{\mu}_{\Psi_{1,k,\beta};\mathbf{h}_k} = \mathbf{C}_{\Psi_{1,k,\beta};\mathbf{h}_k} (\mathbf{C}_{\widehat{\mathbf{h}}_k}^{-1} \boldsymbol{\mu}_{\widehat{\mathbf{h}}_k} - \mathbf{C}_{\mathbf{h}_k;\Psi_{1,k,\beta}}^{-1} \boldsymbol{\mu}_{\mathbf{h}_k;\Psi_{1,k,\beta}}). \end{aligned}$$

B. Message from $\Psi_{1,k,\beta}$ to $\mathbf{z}_{k,\beta}$

The message from $\Psi_{1,k,\beta}$ to $\mathbf{z}_{k,\beta}$ entails from the marginal belief $b_{\Psi_{1,k,\beta}}(\mathbf{z}_{k,\beta})$. This marginal belief can be obtained by

$$\begin{aligned} b_{\Psi_{1,k,\beta}}(\mathbf{z}_{k,\beta}) &= \sum_{\boldsymbol{x}_{k,\beta}} \int b_{\Psi_{1,k,\beta}}(\mathbf{z}_{k,\beta}, \mathbf{h}_k, \boldsymbol{x}_{k,\beta}) d\mathbf{h}_k \\ &= \sum_{\boldsymbol{x}_{k,\beta}} \mathcal{CN}(\mathbf{z}_{k,\beta} | \boldsymbol{\mu}_{\widehat{\mathbf{z}}_{k,\beta}|\boldsymbol{x}_{k,\beta}}, \mathbf{C}_{\widehat{\mathbf{z}}_{k,\beta}|\boldsymbol{x}_{k,\beta}}) \omega(\boldsymbol{x}_{k,\beta}), \end{aligned} \tag{9}$$

where

$$\begin{aligned} \mathbf{C}_{\widehat{\mathbf{z}}_{k,\beta}|\boldsymbol{x}_{k,\beta}} &= \mathbf{A}_{k,\beta} \mathbf{C}_{\widehat{\mathbf{h}}_k|\boldsymbol{x}_{k,\beta}} \mathbf{A}_{k,\beta}^H \\ \boldsymbol{\mu}_{\widehat{\mathbf{z}}_{k,\beta}|\boldsymbol{x}_{k,\beta}} &= \mathbf{A}_{k,\beta} \boldsymbol{\mu}_{\widehat{\mathbf{h}}_k|\boldsymbol{x}_{k,\beta}}. \end{aligned}$$

This belief (9) is also a Gaussian mixture model. The mean and covariance of the belief (approximated marginal posterior) $b_{\Psi_{1,k,\beta}}(\mathbf{h}_k)$ can be obtained as the moments of $\boldsymbol{x}_{k,\beta}$,

$$\begin{aligned} \boldsymbol{\mu}_{\widehat{\mathbf{z}}_{k,\beta}} &= \sum_{\boldsymbol{x}_{k,\beta}} \boldsymbol{\mu}_{\widehat{\mathbf{z}}_{k,\beta}|\boldsymbol{x}_{k,\beta}} \omega(\boldsymbol{x}_{k,\beta}); \\ \mathbf{C}_{\widehat{\mathbf{z}}_{k,\beta}} &= \sum_{\boldsymbol{x}_{k,\beta}} (\mathbf{C}_{\widehat{\mathbf{z}}_{k,\beta}|\boldsymbol{x}_{k,\beta}} + \boldsymbol{\mu}_{\widehat{\mathbf{z}}_{k,\beta}|\boldsymbol{x}_{k,\beta}} \boldsymbol{\mu}_{\widehat{\mathbf{z}}_{k,\beta}|\boldsymbol{x}_{k,\beta}}^H) \omega(\boldsymbol{x}_{k,\beta}) \\ &\quad - \boldsymbol{\mu}_{\widehat{\mathbf{z}}_{k,\beta}} \boldsymbol{\mu}_{\widehat{\mathbf{z}}_{k,\beta}}^H \end{aligned}$$

The message from $\Psi_{1,k,\beta}$ to $\mathbf{z}_{k,\beta}$ is computed as the quotient between the approximated belief $\mathcal{CN}(\mathbf{z}_{k,\beta} | \boldsymbol{\mu}_{\widehat{\mathbf{z}}_{k,\beta}}, \mathbf{C}_{\widehat{\mathbf{z}}_{k,\beta}})$ and $m_{\mathbf{z}_{k,\beta};\Psi_{1,k,\beta}}(\mathbf{z}_{k,\beta})$. Thus, we write this message as

$$m_{\Psi_{1,k,\beta};\mathbf{z}_{k,\beta}}(\mathbf{z}_{k,\beta}) = \mathcal{CN}(\mathbf{z}_{k,\beta} | \boldsymbol{\mu}_{\Psi_{1,k,\beta};\mathbf{z}_{k,\beta}}, \mathbf{C}_{\Psi_{1,k,\beta};\mathbf{z}_{k,\beta}}), \tag{10}$$

where

$$\begin{aligned} &\mathbf{C}_{\Psi_{1,k,\beta};\mathbf{z}_{k,\beta}} = \mathbf{C}_{\mathbf{z}_{k,\beta};\Psi_{1,k,\beta}} (\mathbf{C}_{\mathbf{z}_{k,\beta};\Psi_{1,k,\beta}} - \mathbf{C}_{\widehat{\mathbf{z}}_{k,\beta}})^{-1} \mathbf{C}_{\widehat{\mathbf{z}}_{k,\beta}} \\ &\boldsymbol{\mu}_{\Psi_{1,k,\beta};\mathbf{z}_{k,\beta}} = \mathbf{C}_{\mathbf{z}_{k,\beta};\Psi_{1,k,\beta}} (\mathbf{C}_{\mathbf{z}_{k,\beta};\Psi_{1,k,\beta}} - \mathbf{C}_{\widehat{\mathbf{z}}_{k,\beta}})^{-1} \boldsymbol{\mu}_{\widehat{\mathbf{z}}_{k,\beta}} \\ &- \mathbf{C}_{\widehat{\mathbf{z}}_{k,\beta}} (\mathbf{C}_{\mathbf{z}_{k,\beta};\Psi_{1,k,\beta}} - \mathbf{C}_{\widehat{\mathbf{z}}_{k,\beta}})^{-1} \boldsymbol{\mu}_{\mathbf{z}_{k,\beta};\Psi_{1,k,\beta}}. \end{aligned}$$

C. Message from $\Psi_{1,k,\beta}$ to $\boldsymbol{x}_{k,\beta}$

The marginal belief of $\boldsymbol{x}_{k,\beta}$ can be directly obtained by

$$b_{\Psi_{1,k,\beta}}(\boldsymbol{x}_{k,\beta}) = \iint b_{\Psi_{1,k,\beta}}(\mathbf{z}_{k,\beta}, \mathbf{h}_k, \boldsymbol{x}_{k,\beta}) d\mathbf{z}_{k,\beta} d\mathbf{h}_k = \omega(\boldsymbol{x}_{k,\beta}).$$

Since we use BP for estimating $\boldsymbol{x}_{k,\beta}$, there is no need to calculate the approximated belief. Thus, the message from $\Psi_{1,k,\beta}$ to $\boldsymbol{x}_{k,\beta}$ is

$$m_{\Psi_{1,k,\beta};\boldsymbol{x}_{k,\beta}}(\boldsymbol{x}_{k,\beta}) \propto \frac{\omega(\boldsymbol{x}_{k,\beta})}{m_{\boldsymbol{x}_{k,\beta};\Psi_{1,k,\beta}}(\boldsymbol{x}_{k,\beta})} = \nu(\boldsymbol{x}_{k,\beta}), \tag{11}$$

where $\nu(\boldsymbol{x}_{k,\beta})$ is defined in (6).

V. Message from the Pilots and Prior of $\mathbf{H}$

The combination of pilot measurement and channel prior is captured by $p(\mathbf{h}_g|\boldsymbol{y}_{p,g})$, which is denoted as $\Psi_{0,g}$ and can be viewed as the equivalent prior of channels $\forall k \in G_g, \mathbf{h}_k$. The extrinsic to this factor is computed as in line 4 in Algorithm 1. The belief of $\mathbf{h}_g$ at the $\Psi_{0,g}$ is

$$b_{\Psi_{0,g}}(\mathbf{h}_g) \propto p(\boldsymbol{y}_{p,g}|\mathbf{h}_g) \prod_{k\in G_g} p(\mathbf{h}_k) m_{\mathbf{h}_k;\Psi_{0,g}}(\mathbf{h}_k). \quad (12)$$

All the factors appearing in (12) are Gaussian pdfs. This means that $b_{\Psi_{0,g}}(\mathbf{h}_g)$ is always Gaussian. The marginalization of a Gaussian will also produce a Gaussian. Therefore, a Gaussian projection of $b_{\Psi_{0,g}}(\mathbf{h}_k)$ has no effect and results in the same $b_{\Psi_{0,g}}(\mathbf{h}_k)$. The message from $\Psi_{0,g}$ to $\mathbf{h}_k$ is

$$\begin{aligned} m_{\Psi_{0,g};\mathbf{h}_k}(\mathbf{h}_k) &\propto \frac{\int b_{\Psi_{0,g}}(\mathbf{h}_g) d\mathbf{h}_{\overline{k}}}{m_{\mathbf{h}_k;\Psi_{0,g}}(\mathbf{h}_k)} \\ &= p(\mathbf{h}_k) \frac{\int b_{\Psi_{0,g}}(\mathbf{h}_g) d\mathbf{h}_{\overline{k}}}{p(\mathbf{h}_k) m_{\mathbf{h}_k;\Psi_{0,g}}(\mathbf{h}_k)}. \end{aligned} \quad (13)$$

It has been shown in [9] that the quotient appearing in (13) can be interpreted as a Component-Wise-Conditionally-Unbiased (CWCU) LMMSE estimate for $\mathbf{h}_g$ with hypothetical prior $\prod_{k\in G_g} p(\mathbf{h}_k) m_{\mathbf{h}_k;\Psi_{0,g}}(\mathbf{h}_k)$ and likelihood $p(\boldsymbol{y}_{p,g}|\mathbf{h}_g)$. For simplicity, we denote the hypothetical prior as

$$\begin{aligned} q_{\mathbf{h}_k|\boldsymbol{y}_d}(\mathbf{h}_k) &= \mathcal{CN}(\mathbf{h}_k|\boldsymbol{\mu}_{\mathbf{h}_k|\boldsymbol{y}_d}, \mathbf{C}_{\mathbf{h}_k|\boldsymbol{y}_d}) \propto p(\mathbf{h}_k) m_{\mathbf{h}_k;\Psi_{0,g}}(\mathbf{h}_k) \\ &= \mathcal{CN}(\mathbf{h}_k|\mathbf{0}, \boldsymbol{\Xi}_{\mathbf{h}_k}) \mathcal{CN}(\mathbf{h}_k|\boldsymbol{\mu}_{\mathbf{h}_k;\Psi_{0,g}}, \mathbf{C}_{\mathbf{h}_k;\Psi_{0,g}}), \end{aligned}$$

where

$$\begin{aligned} \mathbf{C}_{\mathbf{h}_k|\boldsymbol{y}_d} &= (\boldsymbol{\Xi}_{\mathbf{h}_k}^{-1} + \mathbf{C}_{\mathbf{h}_k;\Psi_{0,g}}^{-1})^{-1} \\ \boldsymbol{\mu}_{\mathbf{h}_k|\boldsymbol{y}_d} &= \mathbf{C}_{\mathbf{h}_k|\boldsymbol{y}_d} (\mathbf{C}_{\mathbf{h}_k;\Psi_{0,g}}^{-1} \boldsymbol{\mu}_{\mathbf{h}_k;\Psi_{0,g}}). \end{aligned}$$

The vector-level CWCU LMMSE result is given by

$$\frac{\int p(\boldsymbol{y}_{p,g}|\mathbf{h}_g) \prod_{k'} q_{\mathbf{h}_{k'}|\boldsymbol{y}_d}(\mathbf{h}_{k'}) d\mathbf{h}_{\overline{k}}}{q_{\mathbf{h}_k|\boldsymbol{y}_d}(\mathbf{h}_k)} = \mathcal{CN}(\mathbf{h}_k|\boldsymbol{\mu}_{\mathbf{h}_k,CL}, \mathbf{C}_{\mathbf{h}_k,CL}), \quad (14)$$

where the subscript CL stands for CWCU LMMSE. The mean and covariance matrix are computed as

$$\begin{aligned} \boldsymbol{\mu}_{\mathbf{h}_k,CL} &= \frac{1}{\sigma_x^2 P} \boldsymbol{y}_{p,g} - \sum_{k'\in G_g/\{k\}} \boldsymbol{\mu}_{h_{k'}|\boldsymbol{y}_d} \\ \mathbf{C}_{\mathbf{h}_k,CL} &= \frac{\sigma_v^2}{\sigma_x^2 P} \mathbf{I} + \sum_{k'\in G_g/\{k\}} \mathbf{C}_{\mathbf{h}_{k'}|\boldsymbol{y}_d}. \end{aligned} \quad (15)$$

Finally, by using Gaussian reproduction lemma [8], we obtain the message from $\Psi_{0,g}$ to $\mathbf{h}_k$,

$$\begin{aligned} m_{\Psi_{0,g};\mathbf{h}_k}(\mathbf{h}_k) &:= \mathcal{CN}(\mathbf{h}_k|\boldsymbol{\mu}_{\Psi_{0,g};\mathbf{h}_k}, \mathbf{C}_{\Psi_{0,g};\mathbf{h}_k}) \\ &\propto p(\mathbf{h}_k) \mathcal{CN}(\mathbf{h}_k|\boldsymbol{\mu}_{\mathbf{h}_k,CL}, \mathbf{C}_{\mathbf{h}_k,CL}), \end{aligned} \quad (16)$$

where

$$\begin{aligned} \mathbf{C}_{\Psi_{0,g};\mathbf{h}_k} &= \left[\boldsymbol{\Xi}_{\mathbf{h}_k}^{-1} + \left(\frac{\sigma_v^2}{\sigma_x^2 P} \mathbf{I} + \sum_{k'\in G_g/\{k\}} \mathbf{C}_{\mathbf{h}_{k'}|\boldsymbol{y}_d} \right)^{-1} \right]^{-1} \\ \boldsymbol{\mu}_{\Psi_{0,g};\mathbf{h}_k} &= \boldsymbol{\Xi}_{\mathbf{h}_k} \left(\frac{\sigma_v^2}{\sigma_x^2 P} \mathbf{I} + \sum_{k'\neq k} \mathbf{C}_{\mathbf{h}_{k'}|\boldsymbol{y}_d} + \boldsymbol{\Xi}_{\mathbf{h}_k} \right)^{-1} \\ &\cdot \left(\frac{1}{\sigma_x^2 P} \boldsymbol{y}_{p,g} - \sum_{k'\in G_g/\{k\}} \boldsymbol{\mu}_{\mathbf{h}_{k'}|\boldsymbol{y}_d} \right). \end{aligned} \quad (17)$$

VI. Message from the data observation $\mathbf{Y}_d$

The node $\Psi_{2,\beta}$ represents the likelihood $p(\boldsymbol{y}_{d,\beta}|\sum_k \mathbf{z}_{k,\beta})$. The extrinsic to this factor node can be computed by line 8 in Algorithm 1. Thus, the belief at $\Psi_{2,\beta}$ is

$$b_{\Psi_{2,\beta}}(\mathbf{z}_{1,\beta}, \ldots, \mathbf{z}_{K,\beta}) = p(\boldsymbol{y}_{d,\beta}|\sum_k \mathbf{z}_{k,\beta}) \prod_k m_{\mathbf{z}_{k,\beta};\Psi_{2,\beta}}(\mathbf{z}_{k,\beta}).$$

Since all the factors in $b_{\Psi_{2,\beta}}(\mathbf{z}_{1,\beta}, \ldots, \mathbf{z}_{K,\beta})$ are Gaussian, the belief $b_{\Psi_{2,\beta}}$ is also Gaussian. Therefore, Gaussian projection will once again leave this belief unchanged. We can write the message from $\Psi_{2,\beta}$ to $\mathbf{z}_{k,\beta}$ as

$$m_{\Psi_{2,\beta};\mathbf{z}_{k,\beta}}(\mathbf{z}_{k,\beta}) \propto \frac{\int b_{\Psi_{2,\beta}}(\mathbf{z}_{1,\beta}, \ldots, \mathbf{z}_{K,\beta}) d\mathbf{z}_{\overline{k},\beta}}{m_{\mathbf{z}_{k,\beta};\Psi_{2,\beta}}(\mathbf{z}_{k,\beta})}. \quad (18)$$

We can identify (18) as a CWCU LMMSE estimate of $\mathbf{z}_{k,\beta}$ with hypothetical prior $m_{\mathbf{z}_{k,\beta};\Psi_{2,\beta}}(\mathbf{z}_{k,\beta})$ and measurement $p(\boldsymbol{y}_{d,\beta}|\sum_k \mathbf{z}_{k,\beta})$. Thus, we write the feedback message as

$$m_{\Psi_{2,\beta};\mathbf{z}_{k,\beta}}(\mathbf{z}_{k,\beta}) = \mathcal{CN}(\mathbf{z}_{k,\beta}|\boldsymbol{\mu}_{\Psi_{2,\beta};\mathbf{z}_{k,\beta}}, \mathbf{C}_{\Psi_{2,\beta};\mathbf{z}_{k,\beta}}), \quad (19)$$

where

$$\begin{aligned} \boldsymbol{\mu}_{\Psi_{2,\beta};\mathbf{z}_{k,\beta}} &= \boldsymbol{y}_{d,\beta} - \sum_{k'\neq k} \boldsymbol{\mu}_{\mathbf{z}_{k',\beta};\Psi_{2,\beta}} \\ \mathbf{C}_{\Psi_{2,\beta};\mathbf{z}_{k,\beta}} &= \sigma_v^2 \mathbf{I} + \sum_{k'\neq k} \mathbf{C}_{\mathbf{z}_{k',\beta};\Psi_{2,\beta}}. \end{aligned}$$

VII. Message from the Data Prior

Finally, we look at the factor node $\Psi_{3,k,\alpha}$ which represents the data prior $p(\boldsymbol{x}_{k,\alpha})$. The extrinsic to this factor is computed as in line 6 in Algorithm 1. Denote the belief at this factor node as

$$b_{\Psi_{3,k,\alpha}}(\boldsymbol{x}_{k,\alpha}) = f_\alpha(\boldsymbol{x}_{k,\alpha}) \prod_{\beta\in N(\alpha)} m_{\boldsymbol{x}_{k,\beta};\Psi_{3,k,\alpha}}(\boldsymbol{x}_{k,\beta}), \quad (20)$$

where $N(\alpha)$ denotes the set of variable nodes neighboring the factor node $\Psi_{3,k,\alpha}$. Since BP is used, there is no need for approximation. The feedback message can be obtained directly as

$$m_{\Psi_{3,k,\alpha};\boldsymbol{x}_{k,\beta}}(\boldsymbol{x}_{k,\beta}) = \sum_{\boldsymbol{x}_{k,\overline{\beta}}} f_\alpha(\boldsymbol{x}_{x,\alpha}) \prod_{\beta'\in N(\alpha)/\{\beta\}} m_{\boldsymbol{x}_{k,\beta'};\Psi_{3,k,\alpha}}(\boldsymbol{x}_{k,\beta'}). \quad (21)$$

We conclude the algorithm in Algorithm 1.

Algorithm 1 Proposed Method in one iteration

Require: $\boldsymbol{y}_{p,g}$, $\boldsymbol{y}_d$, $p(\boldsymbol{x}_k)$, $p(\mathbf{h}_k)$, $p(\boldsymbol{y}_{d,\beta}|\mathbf{z}_{\{k\},\beta})$

Initialize: $m_{\Psi_{1,k,\beta};\mathbf{h}_k}(\mathbf{h}_k)$, $m_{\Psi_{1,k,\beta};\boldsymbol{x}_{k,\beta}}(\boldsymbol{x}_{k,\beta})$,

$m_{\Psi_{1,k,\beta};\mathbf{z}_{k,\beta}}(\mathbf{z}_{k,\beta})$, $m_{\Psi_{3,k,\alpha};\boldsymbol{x}_{k,\beta}}(\boldsymbol{x}_{k,\beta})$

repeat for all $\alpha, \beta, k = 1:K, g = 1:P$

  $m_{\mathbf{h}_k;\Psi_{0,g}}(\mathbf{h}_k) = \prod_\beta m_{\Psi_{1,k,\beta};\mathbf{h}_k}(\mathbf{h}_k)$

  compute $m_{\Psi_{0,g};\mathbf{h}_k}(\mathbf{h}_k)$ according to (16)

  $m_{\boldsymbol{x}_{k,\beta};\Psi_{3,k,\alpha}}(\boldsymbol{x}_{k,\beta})$
  $= m_{\Psi_{1,k,\beta};\boldsymbol{x}_{k,\beta}}(\boldsymbol{x}_{k,\beta}) \prod_{\alpha' \in N(\beta)/\{\alpha\}} m_{\Psi_{3,k,\alpha'};\boldsymbol{x}_{k,\beta}}(\boldsymbol{x}_{k,\beta})$

  compute $m_{\Psi_{3,k,\alpha};\boldsymbol{x}_{k,\beta}}(\boldsymbol{x}_{k,\beta})$ according to (21)

  $m_{\mathbf{z}_{k,\beta};\Psi_{2,\beta}}(\mathbf{z}_{k,\beta}) = m_{\Psi_{1,k,\beta};\mathbf{z}_{k,\beta}}(\mathbf{z}_{k,\beta})$

  compute $m_{\Psi_{2,\beta};\mathbf{z}_{k,\beta}}(\mathbf{z}_{k,\beta})$ according to (19)

  $m_{\mathbf{h}_k;\Psi_{1,k,\beta}}(\mathbf{h}_k) = m_{\Psi_{0,g};\mathbf{h}_k}(\mathbf{h}_k) \prod_{\beta' \neq \beta} m_{\Psi_{1,k,\beta'};\mathbf{h}_k}(\mathbf{h}_k)$

  $m_{\boldsymbol{x}_{k,\beta};\Psi_{1,k,\beta}}(\boldsymbol{x}_{k,\beta}) = \prod_{\alpha \in N(\beta)} m_{\Psi_{3,k,\alpha};\boldsymbol{x}_{k,\beta}}(\boldsymbol{x}_{k,\beta})$

  $m_{\mathbf{z}_{k,\beta};\Psi_{1,k,\beta}}(\mathbf{z}_{k,\beta}) = m_{\Psi_{2,\beta};\mathbf{z}_{k,\beta}}(\mathbf{z}_{k,\beta})$

  compute $m_{\Psi_{1,k,\beta};\mathbf{z}_{k,\beta}}(\mathbf{z}_{k,\beta})$ according to (10).

  compute $m_{\Psi_{1,k,\beta};\mathbf{h}_k}(\mathbf{h}_k)$ according to (8)

  compute $m_{\Psi_{1,k,\beta};\boldsymbol{x}_{k,\beta}}(\boldsymbol{x}_{k,\beta})$ according to (11)

until Convergence

VIII. Simulation Results

In this section, we will 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. We use the same expression for the large-scale fading model as in [10], $\sigma^2_{H_{mk}}[\text{dB}] = -30.5 - 36.7\log_{10}(d_{mk})$, where d_{mk} is the distance between AP m and UT k. To induce pilot contamination, the default pilot sequence length is $P = 6$. The transmitted data sequence spans a length of $L = 10$. We consider an extreme case where the factorization of the data prior contains only one factor $p(\boldsymbol{x}_k)$. Namely, we assume the data of each user is drawn uniformly from a codebook containing 20 randomly generated codes (i.e. $p(\boldsymbol{x}_k)$ is discrete, uniform, with support of size 20). The six distinct pilot sequences are designed to be mutually orthogonal. Transmission power is set at $\sigma^2_x = 14(\text{dBm})$ for each UT, while the noise is set to $-96(\text{dBm})$.

We maintain consistent positions for all APs and UTs for different realizations and conduct simulations across 50 unique scenarios with varying $\mathbf{H}$, $\mathbf{V}$, and data signals. During the simulation, whenever the covariance matrices of (8) and (10) contain negative eigenvalue, we reset that eigenvalue to some large value to indicate that the Gaussian approximation for the Gaussian mixture model during that iteration is not correct. Since EP (and BP) have the same fixed point as the BFE, resetting the eigenvalue won't change the fixed point. The metric for evaluating performance is the normalized mean squared error (NMSE) of the channel estimation. It is calculated as $\frac{\sum \text{tr}(\widetilde{\mathbf{H}}\widetilde{\mathbf{H}}^H)}{\sum \text{tr}(\mathbf{H}\mathbf{H}^H)}$, where $\widetilde{\mathbf{H}}$ represents the estimation error. The simulation results are concluded in fig. 2. For comparison, we also plot the results of the variable level EP (VL-EP) algorithm [11], which assumes Gaussian inputs. In the Genie-Aided scenario, we assume the data to be known.

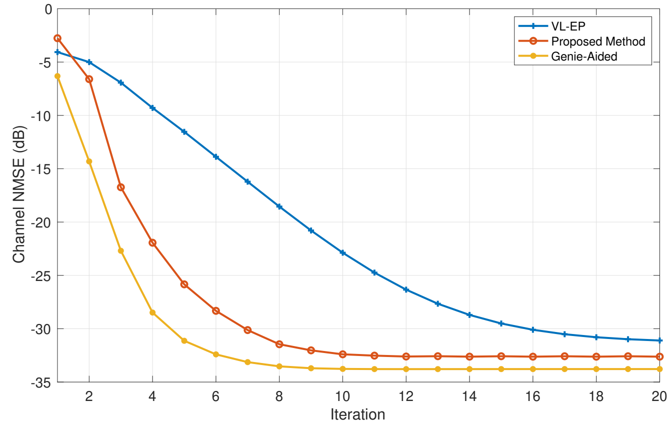

Fig. 2. NMSE [dB] versus Iteration

IX. Conclusions

This paper focuses on the joint estimation of the channel and data during uplink transmission in CF MaMIMO systems. By right multiplying the conjugated pilot sequence to $\mathbf{Y}_p$, we decompose the pilot measurements into P equivalent pilot observations. Due to the orthogonality, any two equivalent pilot observations contain different sets of $\mathbf{h}_k$, and the equivalent noise is also uncorrelated. This indicates that the P equivalent pilot observations are mutually independent. After that, we apply an EP algorithm to estimate the bilinear channel and data. By assuming that there is a structure in the codes, we treat the data (sub-) sequences as atomic variables and derive a vector-based EP. The simulation results verify the effectiveness of our proposed method.

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 *ISIT*. IEEE, 2021.

[2] Z. Zhao and D. Slock, "Semi-Blind Sparse Channel Learning in Cell-Free Massive MIMO-a CRB Analysis," in *Int'l. Conf. on Comm.*, 2023.

[3] T. P. Minka, "A Family of Algorithms for Approximate Bayesian Inference," Ph.D. dissertation, Massachusetts Institute of Technology, 2001.

[4] 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.

[5] H. Jiang, X. Yuan, and Q. Guo, "Hybrid Vector Message Passing for Generalized Bilinear Factorization," *arXiv preprint arXiv:2401.03626*, 2024.

[6] K.-H. Ngo, M. Guillaud, A. Decurninge, S. Yang, and P. Schniter, "Multi-User Detection Based on Expectation Propagation for the Non-Coherent SIMO Multiple Access Channel," *IEEE Transactions on Wireless Communications*, 2020.

[7] T. Minka *et al.*, "Divergence Measures and Message Passing," Technical report, Microsoft Research, Tech. Rep., 2005.

[8] Q. Zou, H. Zhang, C.-K. Wen, S. Jin, and R. Yu, "Concise Derivation for Generalized Approximate Message Passing Using Expectation Propagation," *IEEE Signal Processing Letters*, 2018.

[9] Z. Zhao, F. Xiao, and D. Slock, "Approximate Message Passing for Not So Large niid Generalized Linear Models," in *IEEE Int'l. Workshop on Signal Processing Advances in Wireless Comm. (SPAWC)*, 2023.

[10] 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.

[11] 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.