Bilinear Hybrid Expectation Maximization and Expectation Propagation for Semi-Blind Channel Estimation

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

***Abstract*—This paper discusses channel estimation during uplink transmission in Cell-Free (CF) Massive MIMO (MaMIMO) systems. We model the problem as a semi-blind estimation problem with independent and identically distributed (i.i.d.) Gaussian input.**
Two hybrid Expectation Maximization (EM) and Expectation Propagation (EP) algorithms are proposed to improve convergence behavior. The first algorithm, EM-EP, adopts a vector-level EP approach by treating the per-user channel coefficients and data sequence as EP variables. To make the algorithm tractable, we use the central limit theorem (CLT) to approximate the interference terms and employ EM to construct a majorizer function for the likelihood of the received data, leading to majorization minimization.
To further enhance convergence behavior, we propose a matrix-level loop-free EM-EP algorithm. In this algorithm, we treat the channel coefficients and data sequences corresponding to users using the same pilot as EP variables. This method is an alternating minimization algorithm, ensuring convergence.
Our simulations verify the effectiveness of the two proposed algorithms.

I. Introduction

A critical challenge in Cell-Free (CF) Massive MIMO (MaMIMO) networks is pilot contamination, where the number of user terminals (UTs) in a given area exceeds the length of the pilot sequence. To address this issue, semi-blind approaches have been explored to mitigate the effects of pilot contamination [1].

1) System Model: We examine an uplink semi-blind signal model described as follows:

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

where $\boldsymbol{H}$ represents the channel matrix, unknown and modeled as an independent and identically distributed (i.i.d.) random matrix of size $M \times K$. Each column follows the distribution $\mathbf{h}_k \sim \mathcal{N}(\mathbf{0}, \Xi_{\mathbf{h}_k})$. The input signal is composed of a pilot part $\mathbf{X}_p \in \mathcal{C}^{P\times K}$ and a data part $\mathbf{X} \in \mathcal{C}^{T\times K}$. All data symbols in $\mathbf{X}$ are assumed to be independently drawn from a Gaussian distribution with power σ_x^2.

We define $\mathbf{V}_p$ and $\mathbf{V}$ as the Additive White Gaussian Noise (AWGN) at the access points (APs). Each element within these noise matrices is assumed to be independently drawn from a Gaussian distribution with power σ_v^2.

2) Orthogonal Pilot: When orthogonal pilot is used, we can correlate $\mathbf{Y}_p$ with each pilot sequence $\boldsymbol{x}_{p,g}^*$ to obtain an equivalent observation $\boldsymbol{y}_g$ such that

$$\boldsymbol{y}_{p,g} = \mathbf{Y}_p\boldsymbol{x}_{p,g}^* = P\sigma_x^2\sum_{k\in G_g}\mathbf{h}_k + \boldsymbol{v}_{p,g}, \tag{1}$$

where $\boldsymbol{v}_{p,g} = \mathbf{V}_p\boldsymbol{x}_{p,g}^* \sim \mathcal{N}(\boldsymbol{v}_{p,g}|\mathbf{0}, \sigma_x^2\sigma_v^2P\mathbf{I})$ and G_g denotes the group of users using the g-th pilot sequence. Since one UT k can only send one pilot sequence, the user groups $G_g \cap G_{g'} = \phi$ if $g \neq g'$. Due to orthogonal pilots, we also have $\mathbb{E}[\boldsymbol{y}_{p,g}\boldsymbol{y}_{p,g'}] = \mathbf{0}$ if $g \neq g'$.

A. Prior Works

1) Expectation-propagation: Expectation Propagation (EP) approximates a factored joint probability density function (pdf) with complicated factors by transforming it into a new pdf with simpler factors [2]. Specifically, EP aims to approximate a joint pdf $p(\boldsymbol{\theta})$ with complex factors $p_\alpha(\boldsymbol{\theta}_\alpha)$ by another pdf $b(\boldsymbol{\theta})$ with simpler factors $q_\alpha(\boldsymbol{\theta}_\alpha)$:

$$p(\boldsymbol{\theta}) = \prod_\alpha p_\alpha(\boldsymbol{\theta}_\alpha) \simeq b(\boldsymbol{\theta}) = \prod_\alpha q_\alpha(\boldsymbol{\theta}_\alpha). \tag{2}$$

Each approximating factor is obtained by iteratively minimizing the Kullback-Leibler divergence (KLD)

$$\arg\min_{q_\alpha\in Q} \mathrm{KLD}\left[\frac{p_\alpha(\boldsymbol{\theta}_\alpha)\prod_{\beta\neq\alpha}q_\beta(\boldsymbol{\theta}_\beta)}{Z_{p_\alpha}}\middle\|\frac{q_\alpha(\boldsymbol{\theta}_\alpha)\prod_{\beta\neq\alpha}q_\beta(\boldsymbol{\theta}_\beta)}{Z_b}\right],$$

where Z_b and Z_{p_α} are normalization factors, and Q denotes the family of simplified distributions, often assumed to be an exponential family or Gaussian family. When using the exponential family, minimizing the KLD is equivalent to matching the moments of the left and right arguments.

In [3], the authors propose a simplified EP algorithm, the Variable Level Expectation Propagation (VL-EP) algorithm, using posterior as extrinsic information.

Given that the primary concern with EP is the uncertainty of convergence, a more robust solution is studied, and a variation of EP based on loop-free factorization was proposed in [4].

2) Expectation-maximization: Expectation-maximization can be viewed as a special case of the Majorization-Minimization (MM) algorithm. In [5], the authors compare two variants of the EM approach within the context of semi-blind channel estimation. This is essentially a comparison between Bayesian and deterministic approaches. The authors of [6] proposed a space-alternating generalized expectation maximization (SAGE) algorithm to handle the lack of prior information. To tackle the challenges posed by systems with a more generalized prior distribution for input signals, [7] explored the combination of EM-based on a Gaussian mixture prior.

B. Main Contribution

In this paper, we develop two EP-based methods for semi-blind channel estimation. To make the algorithms tractable, we adopt the idea of EM and construct a majorization function to approximate the original negative log-likelihood, followed by performing majorization minimization.

The first method we propose is EM-EP, where we factorize the joint pdf of the system model at the vector level. Our simulation results suggest that EM-EP converges faster than VL-EP by using the correct extrinsic information.

To further improve convergence behavior, we propose a matrix-level loop-free EM-EP (LF-EM-EP). LF-EM-EP can be viewed as a Majorization-Minimization (MM) algorithm with Central Limit Theorem (CLT) approximations. It guarantees convergence because it can be shown to be an alternating minimization algorithm.

II. EM-EP Derivation

In this section, we treat the channel coefficients of the k-th user $\mathbf{h}_k$ and the data sequence of the k-th user $\boldsymbol{x}_k$ as EP variables.

The joint probability of the underlying system model is

$$p(\mathbf{Y}_p, \mathbf{Y}_d, \mathbf{h}_{\{k\}}, \boldsymbol{x}_{\{k\}}) = p(\boldsymbol{y}_{p,\{g\}}, \mathbf{Y}_d, \mathbf{h}_{\{k\}}, \boldsymbol{x}_{\{k\}}) \\ = p(\mathbf{Y}_d|\mathbf{h}_{\{k\}}, \boldsymbol{x}_{\{k\}}) \prod_g p(\boldsymbol{y}_{p,g}, \boldsymbol{H}_g) \prod_k p(\boldsymbol{x}_k), \tag{3}$$

where $\{\cdot\}$ denotes iteration over all indices, and the first equality holds due to the orthogonal pilot.

According to EP, we approximate the joint pdf (3) by $b(\boldsymbol{H}, \mathbf{X})$:

$$p(\boldsymbol{y}_{p,\{g\}}, \mathbf{Y}_d, \mathbf{h}_{\{k\}}, \boldsymbol{x}_{\{k\}}) \simeq b(\boldsymbol{H}, \mathbf{X}) \\ :\propto q_{\boldsymbol{H},\mathbf{X}}(\mathbf{h}_{\{k\}}, \boldsymbol{x}_{\{k\}}) \prod_g q_{\boldsymbol{H}_g}(\mathbf{h}_{\{k \in G_g\}}) \prod_k p(\boldsymbol{x}_k) \\ = \prod_k \mu_{q_{\boldsymbol{H},\mathbf{X}};\mathbf{h}_k}(\mathbf{h}_k) \mu_{q_{\boldsymbol{H},\mathbf{X}};\boldsymbol{x}_k}(\boldsymbol{x}_k) \mu_{q_{\boldsymbol{H}};\mathbf{h}_k}(\mathbf{h}_k) p(\boldsymbol{x}_k), \tag{4}$$

where we further assume that $q_{\boldsymbol{H},\mathbf{X}}$ and $q_{\boldsymbol{H}_g}$ can be factorized at variable level $\mathbf{h}_k$, $\boldsymbol{x}_k$,

$$q_{\boldsymbol{H},\mathbf{X}}(\boldsymbol{H}, \mathbf{X}) = \prod_k \mu_{q_{\boldsymbol{H},\mathbf{X}};\mathbf{h}_k}(\mathbf{h}_k) \mu_{q_{\boldsymbol{H},\mathbf{X}};\boldsymbol{x}_k}(\boldsymbol{x}_k) \\ q_{\boldsymbol{H}_g}(\boldsymbol{H}_g) = \prod_{k \in G_g} \mu_{q_{\boldsymbol{H}};\mathbf{h}_k}(\mathbf{h}_k). \tag{5}$$

In EP algorithm, we refine the approximated factors $q_{\boldsymbol{H},\mathbf{X}}$ and $q_{\boldsymbol{H}_g}$ iteratively. Now, we examine the refinement for the bilinear factor $q_{\boldsymbol{H},\mathbf{X}}$.

The KLD objective function for refining $q_{\boldsymbol{H},\mathbf{X}}$ is

$$\text{KLD}[b_{q_{\boldsymbol{H},\mathbf{X}}}(\boldsymbol{H}, \mathbf{X}) \| b(\boldsymbol{H}, \mathbf{X})], \tag{6}$$

where

$$b_{q_{\boldsymbol{H},\mathbf{X}}}(\boldsymbol{H}, \mathbf{X}) \propto p(\mathbf{Y}_d|\boldsymbol{H}, \mathbf{X}) \prod_g q_{\boldsymbol{H}_g}(\mathbf{h}_{\{k \in G_g\}}) \prod_k p(\boldsymbol{x}_{d,k})$$

is the belief with true factor $p(\mathbf{Y}_d|\boldsymbol{H}, \mathbf{X})$. Substituting $b(\boldsymbol{H}, \mathbf{X})$ with (4) and ignoring the terms that are irrelevant to $q_{\boldsymbol{H},\mathbf{X}}$, the KLD objective function becomes:

$$\text{KLD}[b_{q_{\boldsymbol{H},\mathbf{X}}}(\boldsymbol{H}, \mathbf{X}) \| b(\boldsymbol{H}, \mathbf{X})] \tag{7} \\ = -\sum_k \int b_{q_{\boldsymbol{H},\mathbf{X}}}(\boldsymbol{H}, \mathbf{X}) \ln \mu_{q_{\boldsymbol{H},\mathbf{X}};\mathbf{h}_k}(\mathbf{h}_k) \mu_{q_{\boldsymbol{H}};\mathbf{h}_k}(\mathbf{h}_k) d\boldsymbol{H} d\mathbf{X} \\ -\sum_k \int b_{q_{\boldsymbol{H},\mathbf{X}}}(\boldsymbol{H}, \mathbf{X}) \ln \mu_{q_{\boldsymbol{H},\mathbf{X}};\boldsymbol{x}_k}(\boldsymbol{x}_k) p(\boldsymbol{x}_k) d\boldsymbol{H} d\mathbf{X} + \ln Z_{b_{\boldsymbol{H},\mathbf{X}}},$$

where $Z_{b_{\boldsymbol{H},\mathbf{X}}}$ is the normalization factor corresponding to (4), i.e., $Z_{b_{\boldsymbol{H},\mathbf{X}}} = Z_{b_{\mathbf{h}_k}} Z_{b_{\boldsymbol{x}_k}}$, with

$$Z_{b_{\mathbf{h}_k}} = \int \mu_{q_{\boldsymbol{H},\mathbf{X}};\mathbf{h}_k}(\mathbf{h}_k) \mu_{q_{\boldsymbol{H}};\mathbf{h}_k}(\mathbf{h}_k) d\mathbf{h}_k \\ Z_{b_{\boldsymbol{x}_k}} = \int \mu_{q_{\boldsymbol{H},\mathbf{X}};\boldsymbol{x}_k}(\boldsymbol{x}_k) p(\boldsymbol{x}_k) d\boldsymbol{x}_k. \tag{8}$$

At this point, we observe that all the variable-level factors $\mu_{q_{\boldsymbol{H},\mathbf{X}};\mathbf{h}_k}(\mathbf{h}_k)$ and $\mu_{q_{\boldsymbol{H},\mathbf{X}};\boldsymbol{x}_k}(\boldsymbol{x}_k)$ of $q_{\boldsymbol{H},\mathbf{X}}$ can be decoupled. Therefore, refining the factor $q_{\boldsymbol{H},\mathbf{X}}$ based on (7) is equivalent to refining all the variable-level factors $\mu_{q_{\boldsymbol{H},\mathbf{X}};\mathbf{h}_k}(\mathbf{h}_k)$ and $\mu_{q_{\boldsymbol{H},\mathbf{X}};\boldsymbol{x}_k}(\boldsymbol{x}_k)$ in parallel.

To refine $\mu_{q_{\boldsymbol{H},\mathbf{X}};\mathbf{h}_k}(\mathbf{h}_k)$, we omit the terms in (7) that do not contain $\mu_{q_{\boldsymbol{H},\mathbf{X}};\mathbf{h}_k}(\mathbf{h}_k)$. The resulting objective function for updating $\mu_{q_{\boldsymbol{H},\mathbf{X}};\mathbf{h}_k}(\mathbf{h}_k)$ is:

$$\text{KLD}[b_{q_{\boldsymbol{H},\mathbf{X}}}(\boldsymbol{H}, \mathbf{X}) \| b(\boldsymbol{H}, \mathbf{X})] \\ = \text{KLD}\left[b_{q_{\boldsymbol{H},\mathbf{X}}}(\mathbf{h}_k) \| \frac{\mu_{q_{\boldsymbol{H},\mathbf{X}};\mathbf{h}_k}(\mathbf{h}_k) \mu_{q_{\boldsymbol{H}};\mathbf{h}_k}(\mathbf{h}_k)}{Z_{b_{\mathbf{h}_k}}} \right] + c, \tag{9}$$

where marginalized belief is denoted as $b_{q_{\boldsymbol{H},\mathbf{X}}}(\mathbf{h}_k) = \int b_{q_{\boldsymbol{H},\mathbf{X}}}(\boldsymbol{H}, \mathbf{X}) d\mathbf{h}_{\overline{k}} d\mathbf{X}$, with $\mathbf{h}_{\overline{k}}$ denoting all the channel coefficients $\mathbf{h}_i$ with $i \neq k$. The marginal belief $b_{q_{\boldsymbol{H},\mathbf{X}}}$ is computed as

$$b_{q_{\boldsymbol{H},\mathbf{X}}}(\mathbf{h}_k) \propto \int \mu_{q_{\boldsymbol{H}};\mathbf{h}_i}(\mathbf{h}_i) p(\boldsymbol{x}_i) p(\mathbf{Y}_d|\mathbf{h}_k \boldsymbol{x}_k^T + \sum_{i \neq k} \mathbf{h}_i \boldsymbol{x}_i^T) \\ \cdot \prod_{i \neq k} \mu_{q_{\boldsymbol{H}};\mathbf{h}_i}(\mathbf{h}_i) p(\boldsymbol{x}_i) d\mathbf{h}_{\overline{k}} d\mathbf{X}. \tag{10}$$

Due to CLT, we can approximate the interference term $\sum_{i \neq k} \mathbf{h}_i \boldsymbol{x}_i^T$ as a zero-mean Gaussian, where the component variables $\mathbf{h}_i$ and $\boldsymbol{x}_i$ follow $\mathbf{h}_i \sim \mu_{q_{\boldsymbol{H}};\mathbf{h}_i}(\mathbf{h}_i)$ and $\boldsymbol{x}_i \sim p(\boldsymbol{x}_i)$. We define a zero mean Gaussian random matrix $\mathbf{Z}_k \simeq \sum_{i \neq k} \mathbf{h}_i \boldsymbol{x}_i^T$ and the vector $\mathbf{z}_k = \text{vec}(\mathbf{Z}_k)$ to approximate the interference:

$$\mathbf{z}_k \sim \mathcal{N}(\mathbf{z}_k | \mathbf{0}, \mathbf{I}_T \otimes \mathbf{C}_{\mathbf{z}_k}), \tag{11}$$

where $\mathbf{I}_T \otimes \mathbf{C}_{\mathbf{z}_k}$ is the covariance matrix of the vectorized interference $\text{vec}(\sum_{i \neq k} \mathbf{h}_i \boldsymbol{x}_i^T)$, i.e.,

$$\mathbf{C}_{\mathbf{z}_k} = \sigma_x^2 \sum_{i \neq k} (\mathbf{C}_{q_{\boldsymbol{H}};\mathbf{h}_i} + \mathbf{m}_{q_{\boldsymbol{H}};\mathbf{h}_i} \mathbf{m}_{q_{\boldsymbol{H}};\mathbf{h}_i}^H) \tag{12}$$

With this CLT approximation, we define the noise plus interference term $\widetilde{\boldsymbol{v}}_k = \mathbf{z}_k + \text{vec}(\mathbf{V}) = \mathcal{N}(\widetilde{\boldsymbol{v}}_k | \mathbf{0}, \mathbf{I}_T \otimes \mathbf{C}_{\widetilde{\boldsymbol{v}}_k})$, where

$$\mathbf{C}_{\widetilde{\boldsymbol{v}}_k} = \sigma_v^2 \mathbf{I} + \sigma_x^2 \sum_{i \neq k} (\mathbf{C}_{q_{\boldsymbol{H}};\mathbf{h}_i} + \mathbf{m}_{q_{\boldsymbol{H}};\mathbf{h}_i} \mathbf{m}_{q_{\boldsymbol{H}};\mathbf{h}_i}^H).$$

We can then simplify (10) as

$$b_{q_{\boldsymbol{H},\mathbf{X}}}(\mathbf{h}_k) \simeq \int \mu_{q_{\boldsymbol{H}};\mathbf{h}_k}(\mathbf{h}_k) p(\boldsymbol{x}_k) p(\mathbf{Y}_d|\mathbf{h}_k \boldsymbol{x}_k^T, \mathbf{C}_{\widetilde{\boldsymbol{v}}_k}) d\boldsymbol{x}_k \\ = \mu_{q_{\boldsymbol{H}};\mathbf{h}_k}(\mathbf{h}_k) p(\mathbf{Y}_d|\mathbf{h}_k, \mathbf{C}_{\widetilde{\boldsymbol{v}}_k}) \tag{13}$$

where $p(\mathbf{Y}_d|\mathbf{h}_k \boldsymbol{x}_k^T, \mathbf{C}_{\widetilde{\boldsymbol{v}}_k})$ is defined as

$$p(\mathbf{Y}_d|\mathbf{h}_k \boldsymbol{x}_k^T, \mathbf{C}_{\mathbf{z}_k}) = \int p(\mathbf{Y}_d|\mathbf{h}_k \boldsymbol{x}_k^T + \mathbf{Z}_k) \mathcal{N}(\mathbf{z}_k | \mathbf{0}, \mathbf{I}_T \otimes \mathbf{C}_{\mathbf{z}_k}) d\mathbf{z}_k \\ = \mathcal{N}\left(\boldsymbol{y}_d | \text{vec}(\mathbf{h}_k \boldsymbol{x}_k^T), \mathbf{I}_T \otimes \mathbf{C}_{\widetilde{\boldsymbol{v}}_k}\right).$$

The likelihood $p(\mathbf{Y}_d|\mathbf{h}_k, \mathbf{C}_{\widetilde{\boldsymbol{v}}_k})$ is not tractable. To obtain the posterior mean and covariance of $b_{q_{\boldsymbol{H},\mathbf{X}}}$, we consider an EM approach and approximate $b_{q_{\boldsymbol{H},\mathbf{X}}}$ with a majorization function. The negated log-likelihood of $p(\mathbf{Y}_d|\mathbf{h}_k, \mathbf{C}_{\widetilde{\boldsymbol{v}}_k})$ is

$$l_{\mathbf{h}_k}(\mathbf{h}_k) = -\ln p(\mathbf{Y}_d|\mathbf{h}_k, \mathbf{C}_{\widetilde{\boldsymbol{v}}_k}). \tag{14}$$

Since we have $p(\boldsymbol{x}_k) = \mathcal{N}(\boldsymbol{x}_k|\mathbf{0}, \sigma_x^2\mathbf{I})$ and $\widetilde{\boldsymbol{v}}_k$ has a block diagonal covariance matrix, the marginalized log-likelihood function is circularly symmetric, i.e., $\forall\varphi, l_{\mathbf{h}_k}(\mathbf{h}_k) = l_{\mathbf{h}_k}(\mathbf{h}_k e^{j\varphi})$.

We construct a majorization function,

$$\begin{aligned} &u_{\mathbf{h}_k}(\mathbf{h}_k|\widehat{\mathbf{h}}_k) \qquad (15)\\ &= l_{\mathbf{h}_k}(\widehat{\mathbf{h}}_k) - \mathbb{E}_{\boldsymbol{x}_k|\mathbf{Y}_d,\widehat{\mathbf{h}}_k,\mathbf{C}_{\widetilde{\boldsymbol{v}}_k}}\left[\ln\left(\frac{p(\mathbf{Y}_d, \boldsymbol{x}_k|\mathbf{h}_k, \mathbf{C}_{\widetilde{\boldsymbol{v}}_k})}{p(\mathbf{Y}_d, \boldsymbol{x}_k|\widehat{\mathbf{h}}_k, \mathbf{C}_{\widetilde{\boldsymbol{v}}_k})}\right)\right], \end{aligned}$$

where $\widehat{\mathbf{h}}_k$ equals the posterior estimate $\mathbf{m}_{\widehat{\mathbf{h}}_k}$ from previous iteration. One can verify by Jensen's inequality that $u_{\mathbf{h}_k}(\mathbf{h}_k|\widehat{\mathbf{h}}_k) \geq l_{\mathbf{h}_k}(\mathbf{h}_k)$, where equality is reached at $\mathbf{h}_k = \widehat{\mathbf{h}}_k$. This implies the following inequality for the log-belief: $-\ln(b_{q_{\boldsymbol{H},\mathbf{X}}}) \leq u_{\mathbf{h}_k}(\mathbf{h}_k|\widehat{\mathbf{h}}_k) - \ln(\mu_{q_{\boldsymbol{H}};\mathbf{h}_k}(\mathbf{h}_k))$. To obtain a tractable algorithm, we approximate the belief $b_{q_{\boldsymbol{H},\mathbf{X}}}(\mathbf{h}_k)$ by the minorization function $\exp[-u_{\mathbf{h}_k}(\mathbf{h}_k|\widehat{\mathbf{h}}_k) + \ln(\mu_{q_{\boldsymbol{H}};\mathbf{h}_k}(\mathbf{h}_k))]$ which is proportional to a Gaussian.

By omitting constant terms in (15) that do not involve $\mathbf{h}_k$, we derive an effective majorization function of $l_{\mathbf{h}_k}(\mathbf{h}_k)$

$$\begin{aligned} \overline{u}_{\mathbf{h}_k}(\mathbf{h}_k|\widehat{\mathbf{h}}_k) &= -\mathbb{E}_{\boldsymbol{x}_k|\mathbf{Y}_d,\widehat{\mathbf{h}}_k,\mathbf{C}_{\widetilde{\boldsymbol{v}}_k}}\left[\ln\left(p(\mathbf{Y}_d|\boldsymbol{x}_k, \mathbf{h}_k, \mathbf{C}_{\widetilde{\boldsymbol{v}}_k})\right)\right]\\ &= \|\boldsymbol{y}_d - (\widehat{\boldsymbol{x}}_k \otimes \mathbf{I}_M)\mathbf{h}_k\|^2_{(\mathbf{I}_T\otimes\mathbf{C}_{\widetilde{\boldsymbol{v}}_k})^{-1}} + \mathrm{tr}[\mathbf{C}_{\widehat{\boldsymbol{x}}}]\|\mathbf{h}_k\|^2_{\mathbf{C}^{-1}_{\widetilde{\boldsymbol{v}}_k}}, \end{aligned}$$

where $\widehat{\boldsymbol{x}}_k$ and $\mathbf{C}_{\widehat{\boldsymbol{x}}}$ are the mean and covariance matrix of the posterior $p(\boldsymbol{x}_k|\mathbf{Y}_k, \widehat{\mathbf{h}}_k, \mathbf{C}_{\mathbf{z}_k})$. These statistics can be obtained via LMMSE as

$$\begin{aligned} \tau_{\boldsymbol{x}_k} &= \left(\widehat{\mathbf{h}}_k^H \mathbf{C}^{-1}_{\widetilde{\boldsymbol{v}}_k}\widehat{\mathbf{h}}_k + \sigma_x^{-2}\right)^{-1}; \qquad (16)\\ \widehat{\boldsymbol{x}}_k^T &= \tau_{\boldsymbol{x}_k}\widehat{\mathbf{h}}_k^H\mathbf{C}^{-1}_{\widetilde{\boldsymbol{v}}_k}\mathbf{Y}_d;\ \mathbf{C}_{\widehat{\boldsymbol{x}}_k} = \tau_{\boldsymbol{x}_k}\mathbf{I}_T. \end{aligned}$$

By combining $\overline{u}(\mathbf{h}_k|\widehat{\mathbf{h}}_k)$ with the non-constant terms in $-\ln(\mu_{q_{\boldsymbol{H}};\mathbf{h}_k}(\mathbf{h}_k))$, we obtain an effective majorization function for $-\ln(b_{q_{\boldsymbol{H},\mathbf{X}}}(\mathbf{h}_k))$,

$$\begin{aligned} &-\ln(b_{q_{\boldsymbol{H},\mathbf{X}}}(\mathbf{h}_k)) \leq \overline{u}(\mathbf{h}_k|\widehat{\mathbf{h}}_k) - \ln(\mu_{q_{\boldsymbol{H}};\mathbf{h}_k}(\mathbf{h}_k)) + c\\ &= \|\boldsymbol{y}_d - (\widehat{\boldsymbol{x}}_k\otimes\mathbf{I}_M)\mathbf{h}_k\|^2_{(\mathbf{I}_T\otimes\mathbf{C}_{\widetilde{\boldsymbol{v}}_k})^{-1}} + \mathrm{tr}[\mathbf{C}_{\widehat{\boldsymbol{x}}}]\|\mathbf{h}_k\|^2_{\mathbf{C}^{-1}_{\widetilde{\boldsymbol{v}}_k}} \qquad (17)\\ &+ \|\mathbf{h}_k - \mathbf{m}_{q_{\boldsymbol{H}};\mathbf{h}_k}\|^2_{\mathbf{C}^{-1}_{q_{\boldsymbol{H}};\mathbf{h}_k}} + c. \end{aligned}$$

The exponential $\exp[-\overline{u}(\mathbf{h}_k|\widehat{\mathbf{h}}_k) + \ln(\mu_{q_{\boldsymbol{H}};\mathbf{h}_k}(\mathbf{h}_k))]$ can be identified as Gaussian with mean and covariance

$$\begin{aligned} \mathbf{C}_{b_{q_{\boldsymbol{H},\mathbf{X}}};\widehat{\mathbf{h}}_k} &= \left[(\widehat{\boldsymbol{x}}_k^H\widehat{\boldsymbol{x}}_k + T\tau_{\widehat{\boldsymbol{x}}_k})\mathbf{C}^{-1}_{\widetilde{\boldsymbol{v}}_k} + \mathbf{C}^{-1}_{q_{\boldsymbol{H}};\mathbf{h}_k}\right]^{-1}\\ \mathbf{m}_{b_{q_{\boldsymbol{H},\mathbf{X}}};\widehat{\mathbf{h}}_k} &= \mathbf{C}_{b_{q_{\boldsymbol{H},\mathbf{X}}};\widehat{\mathbf{h}}_k}\left(\mathbf{C}^{-1}_{\widetilde{\boldsymbol{v}}_k}\mathbf{Y}_d\widehat{\boldsymbol{x}}_k^* + \mathbf{C}^{-1}_{q_{\boldsymbol{H}};\mathbf{h}_k}\mathbf{m}_{q_{\boldsymbol{H}};\mathbf{h}_k}\right) \end{aligned}$$

Optimizing the KLD (9) and substituting $b_{q_{\boldsymbol{H},\mathbf{X}}(\mathbf{h}_k)} = \mathcal{N}(\mathbf{h}_k|\mathbf{m}_{b_{q_{\boldsymbol{H},\mathbf{X}}};\widehat{\mathbf{h}}_k}, \mathbf{C}_{b_{q_{\boldsymbol{H},\mathbf{X}}};\widehat{\mathbf{h}}_k})$, we obtain the update $\mu_{q_{\boldsymbol{H},\mathbf{X}};\mathbf{h}_k}(\mathbf{h}_k) = \frac{\mathcal{N}(\mathbf{h}_k|\mathbf{m}_{b_{q_{\boldsymbol{H},\mathbf{X}}};\widehat{\mathbf{h}}_k}, \mathbf{C}_{b_{q_{\boldsymbol{H},\mathbf{X}}};\widehat{\mathbf{h}}_k})}{\mu_{q_{\boldsymbol{H}};\mathbf{h}_k}(\mathbf{h}_k)}$ which is a Gaussian with mean and covariance matrix given by

$$\begin{aligned} \mathbf{C}_{q_{\boldsymbol{H},\mathbf{X}};\mathbf{h}_k} &= \frac{\mathbf{C}_{\widetilde{\boldsymbol{v}}_k}}{\widehat{\boldsymbol{x}}_k^H\widehat{\boldsymbol{x}}_k + T\tau_{\widehat{\boldsymbol{x}}_k}};\\ \mathbf{m}_{q_{\boldsymbol{H},\mathbf{X}};\mathbf{h}_k} &= \frac{\mathbf{Y}_d\widehat{\boldsymbol{x}}_k^*}{\widehat{\boldsymbol{x}}_k^H\widehat{\boldsymbol{x}}_k + T\tau_{\widehat{\boldsymbol{x}}_k}}. \end{aligned}$$

To refine the approximate $q_{\boldsymbol{H}_g}(\boldsymbol{H}_g) = \prod_{k\in G_g}\mu_{q_{\boldsymbol{H}};\mathbf{h}_k}(\mathbf{h}_k)$, we optimize the KLD objective function

$$\arg\min_{q_{\boldsymbol{H}_g}} \mathrm{KLD}[b_{q_{\boldsymbol{H}_g}}(\boldsymbol{H}, \mathbf{X})\|b(\boldsymbol{H}, \mathbf{X})], \qquad (18)$$

where the belief

$$b_{q_{\boldsymbol{H}_g}}(\boldsymbol{H}, \mathbf{X}) = p(\boldsymbol{y}_{p,g}, \boldsymbol{H}_g)q_{\boldsymbol{H},\mathbf{X}}(\boldsymbol{H}, \mathbf{X})\prod_{j\neq g}q_{\boldsymbol{H}_j}(\boldsymbol{H}_j)\prod_k p(\boldsymbol{x}_k).$$

Substituting $q_{\boldsymbol{H},\mathbf{X}}$, $q_{\boldsymbol{H}_j}$ with (5), we observe that optimizing (18) is equivalent to updating the all the variable-level factors $\forall k \in G_g, \mu_{q_{\boldsymbol{H}};\mathbf{h}_k}$ in parallel with the objective function:

$$\begin{aligned} &\mathrm{KLD}[b_{q_{\boldsymbol{H}_g}}(\boldsymbol{H}, \mathbf{X})\|b(\boldsymbol{H}, \mathbf{X})] \qquad (19)\\ &= \sum_{k\in G_g}\mathrm{KLD}\left[b_{q_{\boldsymbol{H}_g}}(\boldsymbol{H}_g)\|\frac{\mu_{q_{\boldsymbol{H},\mathbf{X}};\mathbf{h}_k}(\mathbf{h}_k)\mu_{q_{\boldsymbol{H}};\mathbf{h}_k}(\mathbf{h}_k)}{Z_{b_{\mathbf{h}_k}}}\right] + c, \end{aligned}$$

where $b_{q_{\boldsymbol{H}_g}}(\boldsymbol{H}_g)$ denotes the marginalized belief $b_{q_{\boldsymbol{H}_g}}(\boldsymbol{H}, \mathbf{X})$ and is calculated by

$$\begin{aligned} b_{q_{\boldsymbol{H}_g}}(\boldsymbol{H}_g) &\propto \int b_{q_{\boldsymbol{H}_g}}(\boldsymbol{H}, \mathbf{X})d\mathbf{X}d\boldsymbol{H}_{\overline{g}}\\ &= p(\boldsymbol{y}_{p,g}|\boldsymbol{H}_g)\prod_{k\in G_g}p(\mathbf{h}_k)\mu_{q_{\boldsymbol{H},\mathbf{X}};\mathbf{h}_k}(\mathbf{h}_k). \qquad (20) \end{aligned}$$

Since the right argument of (19) contains the k-th channel coefficient only, we further marginalize (20) to $b_{q_{\boldsymbol{H}_g}}(\mathbf{h}_k) = \int b_{q_{\boldsymbol{H}_g}}(\boldsymbol{H}_g)d\mathbf{h}_{\overline{k}}$. According to the property of KLD, the variable-level factor is finally updated by $\mu_{q_{\boldsymbol{H}};\mathbf{h}_k}(\mathbf{h}_k) = \frac{b_{q_{\boldsymbol{H}_g}}(\mathbf{h}_k)}{\mu_{q_{\boldsymbol{H},\mathbf{X}};\mathbf{h}_k}(\mathbf{h}_k)} \propto \mathcal{N}(\mathbf{h}_k|\mathbf{m}_{q_{\boldsymbol{H}};\mathbf{h}_k}, \mathbf{C}_{q_{\boldsymbol{H}};\mathbf{h}_k})$, where

$$\begin{aligned} \mathbf{C}_{q_{\boldsymbol{H}};\mathbf{h}_k} &= \left[\left(\sum_{i\in G_g/\{k\}}(\mathbf{C}^{-1}_{q_{\boldsymbol{H},\mathbf{X}};\mathbf{h}_i} + \boldsymbol{\Xi}^{-1}_{\mathbf{h}_i})^{-1} + \frac{\sigma_v^2}{\sigma_x^2 P}\mathbf{I}\right)^{-1} + \boldsymbol{\Xi}^{-1}_{\mathbf{h}_k}\right]^{-1}\\ \mathbf{m}_{q_{\boldsymbol{H}};\mathbf{h}_k} &= \boldsymbol{\Xi}_{\mathbf{h}_k}\left[\boldsymbol{\Xi}_{\mathbf{h}_k} + \frac{\sigma_v^2}{\sigma_x^2 P}\mathbf{I} + \sum_{i\in G_g/\{k\}}(\mathbf{C}^{-1}_{q_{\boldsymbol{H},\mathbf{X}};\mathbf{h}_i} + \boldsymbol{\Xi}^{-1}_{\mathbf{h}_i})^{-1}\right]^{-1}\\ &\cdot\left[\frac{\boldsymbol{y}_{p,g}}{P\sigma_x^2} - \sum_{i\in G_g/\{k\}}\boldsymbol{\Xi}_{\mathbf{h}_i}(\boldsymbol{\Xi}_{\mathbf{h}_i} + \mathbf{C}_{q_{\boldsymbol{H},\mathbf{X}};\mathbf{h}_i})^{-1}\mathbf{m}_{q_{\boldsymbol{H},\mathbf{X}};\mathbf{h}_i}\right]. \qquad (21) \end{aligned}$$

The approximated belief of $\mathbf{h}_k$ is obtained by integrating (4):

$$\begin{aligned} b(\mathbf{h}_k) &= \int b(\boldsymbol{H}, \mathbf{X})d\mathbf{h}_{\overline{k}}d\mathbf{X} = \mu_{q_{\boldsymbol{H}};\mathbf{h}_k}(\mathbf{h}_k)\mu_{q_{\boldsymbol{H},\mathbf{X}};\mathbf{h}_k}(\mathbf{h}_k)\\ &= \mathcal{N}(\mathbf{h}_k|\mathbf{m}_{\widehat{\mathbf{h}}_k}, \mathbf{C}_{\widehat{\mathbf{h}}_k}), \end{aligned}$$

where

$$\begin{aligned} \mathbf{C}_{\widehat{\mathbf{h}}_k} &= (\mathbf{C}^{-1}_{q_{\boldsymbol{H}};\mathbf{h}_k} + \mathbf{C}^{-1}_{q_{\boldsymbol{H},\mathbf{X}}})^{-1}\\ \mathbf{m}_{\widehat{\mathbf{h}}_k} &= \mathbf{C}_{\widehat{\mathbf{h}}_k}(\mathbf{C}^{-1}_{q_{\boldsymbol{H}};\mathbf{h}_k}\mathbf{m}_{q_{\boldsymbol{H}};\mathbf{h}_k} + \mathbf{C}^{-1}_{q_{\boldsymbol{H},\mathbf{X}}}\mathbf{m}_{q_{\boldsymbol{H},\mathbf{X}}}). \qquad (22) \end{aligned}$$

Algorithm 1 proposes a suggested update order. The complexity per iteration is $O(KM^3)$ dominated by the matrix inversion in line 13 of Algorithm 1.

III. LF-EM-EP Derivation

The above-discussed method is derived from a loopy factorization scheme. To improve the convergence behavior, we treat matrices $\boldsymbol{H}_g$ and $\mathbf{X}_g$ as the EP variables, where the columns of $\boldsymbol{H}_g$ and $\mathbf{X}_g$ are composed of $\mathbf{h}_{k\in G_g}$ and $\boldsymbol{x}_{k\in G_g}$. The resulting factorization is obtained by:

Algorithm 1 Hybrid EM-EP

Require: $\mathbf{Y}_p$, $\mathbf{Y}_d$, $\mathbf{X}_p$, $\boldsymbol{\Xi}_{\mathbf{h}_k}$, σ_x^2, σ_v^2

Initialize: $\mathbf{C}_{\widehat{\mathbf{h}}_k}$, $\mathbf{m}_{\widehat{\mathbf{h}}_k}$, $\mathbf{C}_{q_{\boldsymbol{H},\mathbf{X}};\mathbf{h}_k}$, $\mathbf{m}_{q_{\boldsymbol{H},\mathbf{X}};\mathbf{h}_k}$

repeat

 [Update based on Pilot Observation $\forall k$]

 Compute $\mathbf{m}_{q_{\boldsymbol{H}};\mathbf{h}_k}$ and $\mathbf{C}_{q_{\boldsymbol{H}};\mathbf{h}_k}$ based on (21)

 [Update based on Data Observation $\forall k$]

 $\widehat{\mathbf{h}}_k = \mathbf{m}_{\widehat{\mathbf{h}}_k}$

 $\mathbf{C}_{\widetilde{\boldsymbol{v}}_k} = \sigma_v^2 \mathbf{I} + \sigma_x^2 \sum_{i\neq k}(\mathbf{C}_{q_{\boldsymbol{H}};\mathbf{h}_i} + \mathbf{m}_{q_{\boldsymbol{H}};\mathbf{h}_i}\mathbf{m}_{q_{\boldsymbol{H}};\mathbf{h}_i}^H)$

 $\tau_{\boldsymbol{x}_k} = \left(\widehat{\mathbf{h}}_k^H \mathbf{C}_{\widetilde{\boldsymbol{v}}_k}^{-1} \widehat{\mathbf{h}}_k + \sigma_x^{-2}\right)^{-1}$

 $\widehat{\boldsymbol{x}}_k^T = \tau_{\boldsymbol{x}_k} \widehat{\mathbf{h}}_k^H \mathbf{C}_{\widetilde{\boldsymbol{v}}_k}^{-1} \mathbf{Y}_d$

 $\mathbf{C}_{q_{\boldsymbol{H},\mathbf{X}};\mathbf{h}_k} = \frac{\mathbf{C}_{\widetilde{\boldsymbol{v}}_k}}{\widehat{\boldsymbol{x}}_k^H \widehat{\boldsymbol{x}}_k + T\tau_{\widehat{\boldsymbol{x}}_k}}$

 $\mathbf{m}_{q_{\boldsymbol{H},\mathbf{X}};\mathbf{h}_k} = \frac{\mathbf{Y}_d \widehat{\boldsymbol{x}}_k^*}{\widehat{\boldsymbol{x}}_k^H \widehat{\boldsymbol{x}}_k + T\tau_{\widehat{\boldsymbol{x}}_k}}$

 [Posterior Update $\forall k$]

 $\mathbf{C}_{\widehat{\mathbf{h}}_k} = (\mathbf{C}_{q_{\boldsymbol{H}};\mathbf{h}_k}^{-1} + \mathbf{C}_{q_{\boldsymbol{H},\mathbf{X}}}^{-1})^{-1}$

 $\mathbf{m}_{\widehat{\mathbf{h}}_k} = \mathbf{C}_{\widehat{\mathbf{h}}_k}(\mathbf{C}_{q_{\boldsymbol{H}};\mathbf{h}_k}^{-1}\mathbf{m}_{q_{\boldsymbol{H}};\mathbf{h}_k} + \mathbf{C}_{q_{\boldsymbol{H},\mathbf{X}}}^{-1}\mathbf{m}_{q_{\boldsymbol{H},\mathbf{X}}})$

until Convergence

$$\begin{aligned} p(\mathbf{Y}_d, \boldsymbol{y}_{p,g}, \boldsymbol{H}, \mathbf{X}) &= p(\mathbf{Y}_d|\boldsymbol{H}, \mathbf{X}) \prod_g p(\boldsymbol{y}_{p,g}, \boldsymbol{H}_g) p(\mathbf{X}_g) \\ &\simeq q'_{\boldsymbol{H},\mathbf{X}}(\boldsymbol{H}, \mathbf{X}) \prod_g p(\boldsymbol{y}_{p,g}, \boldsymbol{H}_g) p(\mathbf{X}_g), \end{aligned}$$

where $q'_{\boldsymbol{H},\mathbf{X}}(\boldsymbol{H},\mathbf{X}) = \prod_g \mu_{q_{\boldsymbol{H},\mathbf{X}};\boldsymbol{H}_g}(\boldsymbol{H}_g)\mu_{q_{\boldsymbol{H},\mathbf{X}};\mathbf{X}_g}(\mathbf{X}_g)$.
We can rewrite (1) as:

$$\boldsymbol{y}_{p,g} = P\sigma_x^2(\mathbf{1}_{K_g}^T \otimes \mathbf{I}_M)\mathbf{h}_{G_g} + \boldsymbol{v}_{p,g}, \tag{23}$$

where K_g denotes the number of users using the g-th pilot, $\mathbf{1}_{K_g}$ denotes an all-one column vector with K_g entries, and $\mathbf{h}_{G_g} = \text{vec}(\boldsymbol{H}_g)$. For simplicity, we exploit the notations and define the block prior covariance matrix $\boldsymbol{\Xi}_{G_g} = \text{diag}(\boldsymbol{\Xi}_{\mathbf{h}_{k_1}}, \dots, \mathbf{X}_{\mathbf{h}_{k_{K_g}}})$, where $k_1, \dots, k_{K_g}$ are the users using the g-th pilot sequence. We calculate the equivalent prior $p(\boldsymbol{y}_{p,g}, \boldsymbol{H}_g) \propto \mathcal{N}(\mathbf{h}_{G_g}|\mathbf{m}_{\mathbf{h}_{G_g}}, \mathbf{C}_{\mathbf{h}_{G_g}})$, where the mean and covariance matrix are obtained by:

$$\begin{aligned} \mathbf{C}_{\mathbf{h}_{G_g}} &= \boldsymbol{\Xi}_{\mathbf{h}_{G_g}} - \boldsymbol{\Xi}_{\mathbf{h}_{G_g}} \left[\mathbf{1}_{K_g}\mathbf{1}_{K_g}^T \otimes \left(\frac{\sigma_v^2 \mathbf{I}_M}{\sigma_x^2 P} + \sum_{k\in G_g} \boldsymbol{\Xi}_{\mathbf{h}_k}\right)^{-1}\right] \boldsymbol{\Xi}_{\mathbf{h}_{G_g}} \\ \mathbf{m}_{\mathbf{h}_{G_g}} &= \mathbf{C}_{\mathbf{h}_{G_g}}(\mathbf{1}_{K_g} \otimes \mathbf{I}_M)\frac{\boldsymbol{y}_{p,g}}{\sigma_v^2}. \end{aligned} \tag{24}$$

We will use the equivalent prior $p(\boldsymbol{y}_{p,g}, \boldsymbol{H}_g)$ and $p(\mathbf{X}_g)$ to derive an EM-EP algorithm for estimating $\boldsymbol{H}$.
The observed data symbols can be represented as

$$\mathbf{Y}_d = \boldsymbol{H}_g \mathbf{X}_g^T + \sum_{j\neq g} \boldsymbol{H}_j \mathbf{X}_j^T + \mathbf{V}. \tag{25}$$

As in the vector EM-EP, we treat $\boldsymbol{H}_j\mathbf{X}_j^T$ as interference. Due to CLT, we approximate the sum of interference $\sum_{j\neq g} \text{vec}(\boldsymbol{H}_j\mathbf{X}_j^T)$ as Gaussian. We observe:

$$\begin{aligned} \sum_{j\neq g} \text{vec}(\boldsymbol{H}_j\mathbf{X}_j^T) &= \sum_{i\notin G_g} \text{vec}(\mathbf{h}_i \boldsymbol{x}_i^T) = \sum_{i\notin G_g} (\boldsymbol{x}_i \otimes \mathbf{1}_M)(1 \otimes \mathbf{h}_i) \\ &= \sum_{i\notin G_g} (\boldsymbol{x}_i \otimes \mathbf{h}_i). \end{aligned} \tag{26}$$

Define the approximated sum of interference as $\mathbf{Z}_{G_g} \simeq \sum_{j\neq g} \boldsymbol{H}_g \mathbf{X}_g^T$ and its vectorization $\mathbf{z}_{G_g} = \text{vec}(\mathbf{Z}_{G_g})$. We compute the mean and covariance of (26) and match the values to the mean and covariance of $\mathbf{Z}_g$. Therefore, the approximated interference follows $\mathbf{z}_{G_g} \sim \mathcal{N}(\mathbf{z}_{G_g}|\mathbf{0}, \mathbf{I}_T \otimes \mathbf{C}_{\mathbf{z}_{G_g}})$, where

$$\mathbf{C}_{\mathbf{z}_{G_g}} = \sigma_x^2 \sum_{i\notin G_g} (\mathbf{C}_{\mathbf{h}_i} + \mathbf{m}_{\mathbf{h}_i}\mathbf{m}_{\mathbf{h}_i}^H), \tag{27}$$

with $\mathbf{C}_{\mathbf{h}_i} = (\mathbf{1}_{K_g}^T \otimes \mathbf{I}_M)\mathbf{C}_{\mathbf{h}_{G_g}}(\mathbf{1}_{K_g} \otimes \mathbf{I}_M)$ and $\mathbf{m}_{\mathbf{h}_i} = (\mathbf{1}_{K_g}^T \otimes \mathbf{I}_M)\mathbf{m}_{\mathbf{h}_{G_g}}$. We combine the noise and interference $\widetilde{\mathbf{V}}_{G_g} = \mathbf{V} + \mathbf{Z}_{G_g}$. Its vectorization $\widetilde{\boldsymbol{v}}_{G_g} = \text{vec}(\widetilde{\mathbf{V}}_{G_g})$ can be verified to be a Gaussian $\widetilde{\boldsymbol{v}}_{G_g} \sim \mathcal{N}(\widetilde{\boldsymbol{v}}_{G_g}|\mathbf{0}, \mathbf{I}_T \otimes \mathbf{C}_{\widetilde{\boldsymbol{v}}_{G_g}})$, where

$$\mathbf{C}_{\widetilde{\boldsymbol{v}}_{G_g}} = \sigma_x^2 \sum_{i\notin G_g} (\mathbf{C}_{\mathbf{h}_i} + \mathbf{m}_{\mathbf{h}_i}\mathbf{m}_{\mathbf{h}_i}^H) + \sigma_v^2 \mathbf{I}_M. \tag{28}$$

Following the same steps from (14) to (17) we have

$$\begin{aligned} -\ln[p(\mathbf{Y}_d, \boldsymbol{y}_{p,g}, \boldsymbol{H}_g|\mathbf{C}_{\widetilde{\boldsymbol{v}}_{G_g}})] &\leq \overline{u}_{\boldsymbol{H}_g}(\boldsymbol{H}_g|\widehat{\boldsymbol{H}}_g) \\ &\quad - \ln[p(\boldsymbol{y}_{p,g}, \boldsymbol{H}_g)] + c, \end{aligned} \tag{29}$$

where $\widehat{\boldsymbol{H}}_g$ equals to the posterior mean of $\boldsymbol{H}_g$ from the previous iteration, and the approximate marginalized pdf $p(\mathbf{Y}_d, \boldsymbol{y}_{p,g}, \boldsymbol{H}_g|\mathbf{C}_{\widetilde{\boldsymbol{v}}_{G_g}}) \simeq \int p(\mathbf{Y}_d, \boldsymbol{y}_{p,g}, \boldsymbol{H}, \mathbf{X}) d\mathbf{X} d\boldsymbol{H}_{\overline{g}}$ is exact when the CLT approximation $\mathbf{Z}_{G_g}$ is exact, i.e., when $K \rightarrow +\infty$. The term $\overline{u}_{\boldsymbol{H}_g}(\boldsymbol{H}_g|\widehat{\boldsymbol{H}}_g)$ is the effective majorization function of the negative log-likelihood $-\ln[p(\mathbf{Y}_d|\boldsymbol{H}_g, \mathbf{C}_{\widetilde{\boldsymbol{v}}_{G_g}})]$, and is defined as

$$\begin{aligned} \overline{u}_{\boldsymbol{H}_g}(\boldsymbol{H}_g|\widehat{\boldsymbol{H}}_g) &= -\mathbb{E}_{\mathbf{X}_g|\mathbf{Y}_d, \widehat{\boldsymbol{H}}_g, \mathbf{C}_{\widetilde{\boldsymbol{v}}_{G_g}}}[\ln(p(\mathbf{Y}_d|\mathbf{X}_g, \boldsymbol{H}_g, \mathbf{C}_{\widetilde{\boldsymbol{v}}_{G_g}}))] \\ &= \|\boldsymbol{y}_d - (\widehat{\mathbf{X}}_g \otimes \mathbf{I}_M)\mathbf{h}_{G_g}\|^2_{\mathbf{I}_T \otimes \mathbf{C}_{\widetilde{\boldsymbol{v}}_{G_g}}^{-1}} + \|\mathbf{h}_{G_g}\|_{T\mathbf{C}_{\widehat{\boldsymbol{x}}_{G_g}} \otimes \mathbf{C}_{\widetilde{\boldsymbol{v}}_{G_g}}^{-1}}. \end{aligned}$$

where $\widehat{\boldsymbol{x}}_{G_g}$ and $\mathbf{C}_{\widehat{\boldsymbol{x}}_{G_g}}$ are the LMMSE results from the model:

$$\boldsymbol{y}_d = (\mathbf{I}_T \otimes \widehat{\boldsymbol{H}}_g)\boldsymbol{x}_{G_g} + \widetilde{\boldsymbol{v}}_{G_g}, \tag{30}$$

with $\boldsymbol{x}_{G_g} = \text{vec}(\mathbf{X}_g^T)$ such that $p(\mathbf{X}_g|\mathbf{Y}_d, \widehat{\boldsymbol{H}}_g, \mathbf{C}_{\widetilde{\boldsymbol{v}}_{G_g}}) = \mathcal{N}(\boldsymbol{x}_{G_g}|\widehat{\boldsymbol{x}}_{G_g}, \mathbf{I}_T \otimes \mathbf{C}_{\widehat{\boldsymbol{x}}_{G_g}})$,

$$\begin{aligned} \mathbf{C}_{\widehat{\boldsymbol{x}}_{G_g}} &= (\widehat{\boldsymbol{H}}_g^H \mathbf{C}_{\widetilde{\boldsymbol{v}}_{G_g}}^{-1} \widehat{\boldsymbol{H}}_g + \sigma_x^{-2}\mathbf{I}_{K_g})^{-1} \\ \widehat{\boldsymbol{x}}_{G_g} &= [\mathbf{I}_T \otimes (\mathbf{C}_{\widehat{\boldsymbol{x}}_{G_g}} \widehat{\boldsymbol{H}}_g^H \mathbf{C}_{\widetilde{\boldsymbol{v}}_{G_g}}^{-1})]\boldsymbol{y}_d \Leftrightarrow \widehat{\mathbf{X}}_g^T = \mathbf{C}_{\widehat{\boldsymbol{x}}_{G_g}} \widehat{\boldsymbol{H}}_g^H \mathbf{C}_{\widetilde{\boldsymbol{v}}_{G_g}}^{-1} \mathbf{Y}_d \end{aligned}$$

The matrix mean $\widehat{\mathbf{X}}_g$ is obtained by $\text{vec}^{-1}_{K_g\times T}(\widehat{\boldsymbol{x}}_{G_g})^T$, where the $\text{vec}^{-1}_{A\times B}(\boldsymbol{\theta})$ operator maps a vector $\boldsymbol{\theta}$ to a matrix $\boldsymbol{\Theta} \in \mathcal{C}^{A\times B}$ such that $\text{vec}(\boldsymbol{\Theta}) = \boldsymbol{\theta}$.
The majorization function of the negative log-joint pdf (29) can be expanded as:

$$\begin{aligned} &-\ln[p(\mathbf{Y}_d, \boldsymbol{y}_{p,g}, \boldsymbol{H}_g|\mathbf{C}_{\widetilde{\boldsymbol{v}}_{G_g}})] \leq c + \|\mathbf{h}_{G_g} - \mathbf{m}_{\mathbf{h}_{G_g}}\|^2_{\mathbf{C}_{\mathbf{h}_{G_g}}^{-1}} \\ &+ \|\boldsymbol{y}_d - (\widehat{\mathbf{X}}_g \otimes \mathbf{I}_M)\mathbf{h}_{G_g}\|^2_{\mathbf{I}_T \otimes \mathbf{C}_{\widetilde{\boldsymbol{v}}_{G_g}}^{-1}} + \|\mathbf{h}_{G_g}\|_{T\mathbf{C}^*_{\widehat{\boldsymbol{x}}_{G_g}} \otimes \mathbf{C}_{\widetilde{\boldsymbol{v}}_{G_g}}^{-1}}. \end{aligned}$$

Optimizing the majorization function, we obtain the posterior mean and covariance matrix

$$\begin{aligned} \mathbf{C}_{\widehat{\mathbf{h}}_{G_g}} &= \left[(\widehat{\mathbf{X}}_g^H \widehat{\mathbf{X}}_g + T\mathbf{C}^*_{\widehat{\boldsymbol{x}}_{G_g}}) \otimes \mathbf{C}_{\widetilde{\boldsymbol{v}}_{G_g}}^{-1} + \mathbf{C}_{\mathbf{h}_{G_g}}^{-1}\right]^{-1} \\ \mathbf{m}_{\widehat{\mathbf{h}}_{G_g}} &= \mathbf{C}_{\widehat{\mathbf{h}}_{G_g}} \left[(\widehat{\mathbf{X}}_g^H \otimes \mathbf{C}_{\widetilde{\boldsymbol{v}}_{G_g}}^{-1})\boldsymbol{y}_d + \mathbf{C}_{\mathbf{h}_{G_g}}^{-1}\mathbf{m}_{\mathbf{h}_{G_g}}\right]. \end{aligned} \tag{31}$$

Algorithm 2 Loop-Free EM-EP

Require: $\mathbf{Y}_p$, $\mathbf{Y}_d$, $\mathbf{X}_p$, $\Xi_{\mathbf{h}_k}$, σ_x^2, σ_v^2

1: Initialize: $\mathbf{m}_{\widehat{\mathbf{h}}_{G_g}}$
2: [Update based on Pilot Observation $\forall g$]
3: Compute $\mathbf{m}_{\mathbf{h}_{G_g}}$ and $\mathbf{C}_{\mathbf{h}_{G_g}}$ based on (24)
4: **repeat**
5: [Update based on Data Observation $\forall g$]
6: $\widehat{\boldsymbol{H}}_g = \text{vec}^{-1}_{M,K_g}(\mathbf{m}_{\widehat{\mathbf{h}}_{G_g}})$
7: $\mathbf{C}_{\widetilde{\boldsymbol{v}}_{G_g}} = \sigma_x^2 \sum_{i \notin G_g} (\mathbf{C}_{\mathbf{h}_i} + \mathbf{m}_{\mathbf{h}_i}\mathbf{m}_{\mathbf{h}_i}^H) + \sigma_v^2 \mathbf{I}_M$
8: $\mathbf{C}_{\widehat{\boldsymbol{x}}_{G_g}} = (\widehat{\boldsymbol{H}}_g^H \mathbf{C}^{-1}_{\widetilde{\boldsymbol{v}}_{G_g}} \widehat{\boldsymbol{H}}_g + \sigma_x^{-2}\mathbf{I}_{K_g})^{-1}$
9: $\widehat{\mathbf{X}}_g^T = \mathbf{C}_{\widehat{\boldsymbol{x}}_{G_g}} \widehat{\boldsymbol{H}}_g^H \mathbf{C}^{-1}_{\widetilde{\boldsymbol{v}}_{G_g}} \mathbf{Y}_d$
10: $\mathbf{C}_{\widehat{\mathbf{h}}_{G_g}} = \left[(\widehat{\mathbf{X}}_g^H \widehat{\mathbf{X}}_g + T\mathbf{C}^*_{\widehat{\boldsymbol{x}}_{G_g}}) \otimes \mathbf{C}^{-1}_{\widetilde{\boldsymbol{v}}_{G_g}} + \mathbf{C}^{-1}_{\mathbf{h}_{G_g}} \right]^{-1}$
11: $\mathbf{m}_{\widehat{\mathbf{h}}_{G_g}} = \mathbf{C}_{\widehat{\mathbf{h}}_{G_g}} \left[(\widehat{\mathbf{X}}_g^H \otimes \mathbf{C}^{-1}_{\widetilde{\boldsymbol{v}}_{G_g}}) \boldsymbol{y}_d + \mathbf{C}^{-1}_{\mathbf{h}_{G_g}} \mathbf{m}_{\mathbf{h}_{G_g}} \right]$
12: **until** Convergence

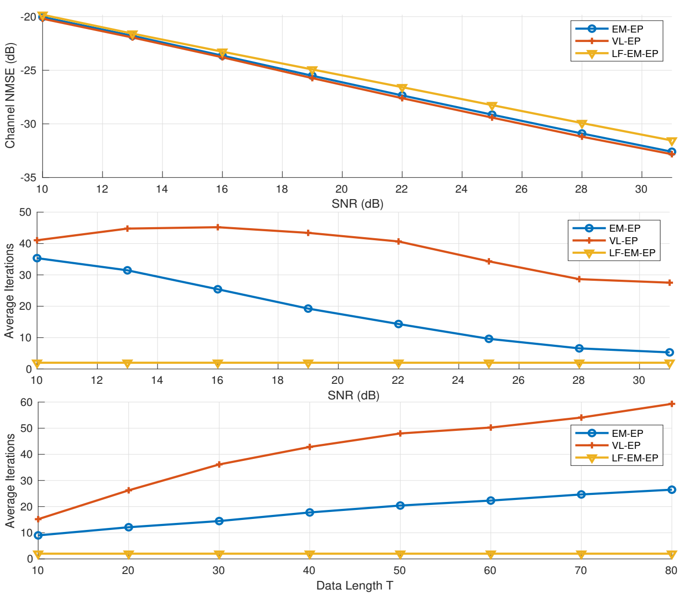

Fig. 1. Convergence speed and NMSE comparison, top and mid graph: fixed data length $T = 40$, bottom graph: fixed SNR $20dB$

We propose a possible update order in Algorithm 2. This method can be viewed as an MM algorithm with CLT approximations. Therefore, it will always converge. The complexity of LF-EM-EP is $O[P(K_{g_{\max}}M)^3]$ due to the matrix inversion in line 10 of Algorithm 2. If we assume the ratio between K and P to be fixed, LF-EM-EP has the same order of complexity as EM-EP.

IV. Simulation Results

In this section, we will verify the algorithm using numerical simulations. We consider a 400 m $\times$ 400 m 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 $i, j \in \{0, \dots, 3\}$. The UTs are uniformly randomly distributed over this area. We use the same expression for the large-scale fading model as in [8],

$$\sigma^2_{H_{ik}}(\text{dB}) = -30.5 - 36.7 \log_{10}(d_{ik}), \tag{32}$$

where d_{ik} is the distance between AP i and UT k. We set the pilot sequence length as $P = 6$ to induce pilot contamination. The 6 distinct pilot sequences are designed to be mutually orthogonal. The noise power is set to be -96 dBm. The transmission power for all the UTs is identical, σ_x^2. We modify the transmission power σ_x^2 to tune the SNR $= \frac{\mathbb{E}\,\text{tr}[\boldsymbol{H}\mathbf{X}_d\mathbf{X}_d^H\boldsymbol{H}^H]}{\mathbb{E}\,\text{tr}[\mathbf{V}_d\mathbf{V}_d^H]} = \frac{\sigma_x^2 \sum_k \text{tr}[\Xi_{\mathbf{h}_k}]}{M\sigma_v^2}$ from 10 dB to 31 dB.
For each SNR level, we maintain consistent positions for all APs and UTs and conduct simulations across 50 unique scenarios with varying $\boldsymbol{H}$, $\mathbf{V}_d$, $\mathbf{V}_p$ and $\mathbf{X}$. One metric for evaluating performance is the normalized mean squared error (NMSE) of the channel estimation. It is calculated as $\frac{\sum \text{tr}(\widetilde{\boldsymbol{H}}\widetilde{\boldsymbol{H}}^H)}{\sum \text{tr}(\boldsymbol{H}\boldsymbol{H}^H)}$, where $\widetilde{\boldsymbol{H}}$ represents the estimation error, and the summation extends over all 50 distinct realizations. Another performance metric is the average number of iterations to converge, where we observe the average number of iterations until the NMSE difference between two consecutive iterations is smaller than a threshold 10^{-5}.
The simulation results are concluded in Fig. 1. For comparison, we also plot the results of the VL-EP algorithm [3]. The simulation results show that Channel NMSE performances of the three algorithms are similar. However, the convergence speed of LF-EM-EP is much faster than the other two methods.

V. Conclusions

This paper focuses on the semi-blind channel estimation in uplink transmission of CF MaMIMO systems. We propose two EP-based algorithms. The first algorithm, EM-EP, can be viewed as a variation of VL-EP with correct extrinsic information. Simulation results show that EM-EP converges faster than VL-EP. To further improve convergence behavior, we propose LF-EM-EP, an alternating minimization algorithm. The simulation results demonstrate that LF-EM-EP converges even faster than EM-EP.

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] T. Minka *et al.*, "Divergence measures and message passing," Technical report, Microsoft Research, Tech. Rep., 2005.
[3] 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.
[4] H. Jiang, X. Yuan, and Q. Guo, "Hybrid Vector Message Passing for Generalized Bilinear Factorization," *arXiv preprint arXiv:2401.03626*, 2024.
[5] E. Nayebi and B. D. Rao, "Semi-blind Channel Estimation for Multiuser Massive MIMO Systems," *IEEE Transactions on Signal Processing*, 2018.
[6] K. Mawatwal, D. Sen, and R. Roy, "A Semi-Blind Channel Estimation Algorithm for Massive MIMO Systems," *IEEE Wireless Communications Letters*, 2017.
[7] E. Nayebi and B. D. Rao, "Semi-Blind Channel Estimation in Massive Mimo Systems with Different Priors on Data Symbols," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018.
[8] 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.