Approximate Message Passing for Not So Large niid Generalized Linear Models

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

***Abstract*—Many signal processing problems involve a Generalized Linear Model (GLM), which is a linear model in which the unknowns may be non-identically independently distributed (n.i.i.d.). Vector Approximate Message Passing (VAMP) is a computationally efficient belief propagation technique used for Bayesian inference. However, the posterior variances obtained from (limited complexity) VAMP are only exact when an independent and identically distributed (i.i.d.) prior is assumed, due to the averaging operations involved. In many problems, it is desirable to not only get estimates of the unknowns but also correct posterior distributions. Whereas VAMP and esp. AMP is applicable to problems of high dimensions, in many applications the dimensions are not very high, allowing for more complex operations. Also, in finite dimensions, the asymptotic regime leading to correct variances under certain measurement matrix model assumptions does not hold. To address these challenges, we propose a revisited version of VAMP, called reVAMP, which provides both a multivariate Gaussian posterior approximation (including inter-parameter correlations) and accurate posterior marginals which only require the extrinsic distributions to become Gaussian.**

I. Introduction

The recovery of signal vectors is a fundamental problem in signal processing and finds applications in various domains, including image and speech processing, communications, machine learning, and localization. In many problems, such as compressed sensing, we are interested in recovering a random signal vector from noisy measurements through a linear measurement model, as mentioned in the abstract. In other problems, such as Direction of Arrival estimation, the measurement matrix may be a parametric dictionary. In those problems, the distribution of the random signals may involve the deterministic parameters of interest. Maximum Likelihood parameter estimation in such mixed deterministic-random problems can be carried out using Expectation Maximization but still requires the posterior distribution of the signal vector.

In the field of localization [1], we are sometimes more interested in estimating the parameter of the random vector $\boldsymbol{x}$, namely, calculating the posterior $\mathcal{E}\{\boldsymbol{\theta}|\boldsymbol{y}\}$ while the observations $p(\boldsymbol{y}|\boldsymbol{x})$ and priors $p(\boldsymbol{x}|\boldsymbol{\theta})$ are given. By using iterative methods such as Expectation Maximization, the estimation problem is transformed into the recovery of a signal vector.

Even in lower dimensions, the application of Bayesian estimation (e.g., Minimum Mean Squared Error (MMSE)) becomes challenging in a non-Gaussian scenario due to the intractability of the involved integrals.

To address this challenge, approximate inference methods have been developed, among which Approximate Message Passing (AMP) is a popular and efficient approach [2]. AMP has demonstrated effectiveness in recovering high-dimensional signals, and its dynamics can be fully characterized by a state evolution [3]. However, the convergence of AMP can be problematic when dealing with ill-conditioned measurement matrices $\mathbf{A}$.

The Vector Approximate Message Passing (VAMP) algorithm has been proposed to handle ill-conditioned $\mathbf{A}$ matrices [4]. It achieves this by splitting one variable node $\boldsymbol{x}$ into two variable nodes, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, both representing $\boldsymbol{x}$. Subsequently, an iterative Expectation-Propagation (EP)-like message passing algorithm is applied to the factor graph using vector-valued messages. VAMP has demonstrated favorable performance under right rotationally invariant $\mathbf{A}$, and its state evolution has been rigorously established [4].

A. Prior Work

In [4], the VAMP algorithm was introduced. VAMP approximates the posterior marginals as having identical variances to reduce computational complexity. The paper also proves that for log-concave priors, the variances of the extrinsic distributions are always positive.

In [5], an alternative method for implementing VAMP is presented, which avoids the need to approximate the extrinsic distribution covariances as multiples of the identity matrix. This approach considers prior distributions with non-log-concave probability density functions and introduces a correction term to ensure the non-negativity of the extrinsic variances.

In [6], Component-Wise Conditionally Unbiased (CWCU) MMSE estimation with zero prior mean is proposed. They establish a close relationship between Linear MMSE, Best Linear Unbiased Estimation (BLUE), and CWCU MMSE estimators.

B. Main Contribution

In VAMP and its variants, it is often assumed that the system dimension is high, which motivates the efforts to avoid expensive matrix inversions and the need for additional approximations to reduce complexity.

However, in many estimation problems, non-Gaussian distributions may require approximate Bayesian techniques even when the dimension is not very high. There is a specific interest in estimating posterior distributions, particularly variances. The original VAMP algorithm only provides averaged variances, which motivates the development of the revisited Approximate Message Passing (reVAMP) algorithm presented in this paper. This algorithm leverages the properties of multivariate Gaussian marginalization and adopts a similar Expectation-Propagation (EP)-like derivation approach as described in [4].

In reVAMP, each marginal extrinsic distribution is approximated as a Gaussian distribution using EP. This approximation is then used to calculate the marginal posterior of the signal vector entries through the sum-product rule. Additionally, a joint Gaussian approximation for the joint posterior is obtained as a byproduct. Furthermore, this paper explores the relationship between the CWCU estimator and the derivation of extrinsic in reVAMP, and extends the CWCU estimator by considering non-zero prior mean

C. Notations

The operations $\boldsymbol{x}.\boldsymbol{y}$ and $\boldsymbol{x}./\boldsymbol{y}$ represent the element-wise multiplication and division of two vectors, respectively. We use $D(\boldsymbol{\tau})$ to represent a diagonal matrix constructed from vector $\boldsymbol{\tau}$. We use the $\mathcal{N}(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma})$ to denote the Gaussian distribution function evaluated at $\boldsymbol{x}$ with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

II. EP-like Derivation

In the linear mixing data model:

$$\boldsymbol{y} = \mathbf{A}\boldsymbol{x} + \boldsymbol{v}, \quad p_{\boldsymbol{x}}(\boldsymbol{x}), \quad p_{\boldsymbol{v}}(\boldsymbol{v}), \tag{1}$$

where $\boldsymbol{y}$ is the observed data vector, $\mathbf{A} \in \mathbb{R}^{M\times N}$ is the measurement matrix, $\boldsymbol{x}$ is the signal vector, and $\boldsymbol{v}$ represents the measurement noise. The niid prior distribution of $\boldsymbol{x}$ is denoted as $p_{\boldsymbol{x}}(\boldsymbol{x}) = \prod_{i=1}^{N} p_{x_i}(x_i)$, and the noise is assumed to follow a zero-mean Gaussian distribution with covariance matrix $\mathbf{C}_{\boldsymbol{vv}} \in \mathbb{R}^{M\times M}$, given by $p_{\boldsymbol{v}}(\boldsymbol{v}) = \mathcal{N}(\boldsymbol{v};\mathbf{0},\mathbf{C}_{\boldsymbol{vv}})$.
In order to factorize the joint distribution, we express it as:

$$p(\boldsymbol{x},\boldsymbol{y}) = p(\boldsymbol{y}|\boldsymbol{x})\prod_{i=1}^{N} p_{x_i}(x_i). \tag{2}$$

This factorization can be viewed as a factor graph in Fig. 1 with $x_i, i \in \{1\dots N\}$ denoting variable nodes and the factors $p(\boldsymbol{y}|\boldsymbol{x})$ and $p_{x_i}(x_i), i \in \{1\dots N\}$ standing for factor nodes.

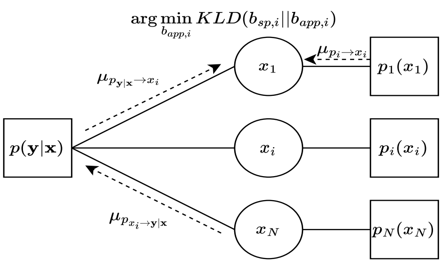

Fig. 1. Factor graph of reVAMP

By applying the sum-product rule [7], the messages passed from factor nodes a to variable node i is $\mu_{f_a\to x_i}(x_i) \propto \int p(\boldsymbol{x})\prod_{j\neq i}\mu_{x_j\to f_a}(x_j)dx_j$.
At variable node i, the received message $b_{sp,i}(x_i) \propto \prod_a \mu_{f_a\to x_i}$ is approximated to Gaussian belief $b_{app,i}(x_i) = \mathcal{N}(x_i;\widehat{x}_i,\tau_{x_i})$ by minimizing the Kullback–Leibler divergence (KLD). We formulate the approximation by $\arg\min_{b_{app,i}} \mathrm{KLD}(b_{sp,i}(x_i)||b_{app,i}(x_i))$.
Analog to the sum-product rule, the message sent from the variable node i to the factor node a can be represented by the distribution $\mu_{x_i\to f_a}(x_i) = b_{app,i}(x_i)/\mu_{f_a\to x_i}(x_i)$.
In the following, we will provide a detailed derivation for the case where the measurement noise is assumed to be Gaussian.

A. Extrinsic to variable nodes

Suppose that at each iteration, the message passed from variable nodes to the factor node $p(\mathbf{y}|\mathbf{x})$ is $q_i(x_i)$ for all $i = 1,\dots,N$. If for all $i \in 1,\dots,N$, $q_i(x_i)$ is initialized as a Gaussian distribution, the EP-like procedure guarantees that they will remain Gaussian. Without loss of generality, we define that $q_i(x_i) = \mathcal{N}(x_i;p_i,\tau_{p_i})$, where p_i, τ_{p_i} are the extrinsic (assumed prior) mean and variance of the i-th element.
The joint distribution $\prod_{i=1,\dots N} q_i(x_i)$ equals to $\mathcal{N}(\boldsymbol{x};\boldsymbol{p},\boldsymbol{\tau}_p)$, where $\boldsymbol{p} = \left[p_1\dots p_N\right]^T$ and $\boldsymbol{\tau}_p = \left[\tau_{p_1}\dots\tau_{p_N}\right]^T$. The posterior $p(\boldsymbol{x}|\boldsymbol{y})$ is approximated as $q(\boldsymbol{x}) \propto p(\boldsymbol{y}|\boldsymbol{x})\prod_{i=1\dots N} q_i(x_i)$. The extrinsic for any variable node i is the marginalization of (3) over x_i

$$q(\boldsymbol{x})/q_i(x_i) \propto \frac{p(\boldsymbol{y}|\boldsymbol{x})\mathcal{N}(\boldsymbol{x};\boldsymbol{p},\boldsymbol{\tau}_p)}{q_i(x_i)} \propto \frac{\mathcal{N}(\boldsymbol{x};\mathbf{m},\mathbf{C}_m)}{q_i(x_i)}, \tag{3}$$

where

$$\begin{aligned}\mathbf{C}_m &= \left[\mathbf{A}^T\mathbf{C}_{\boldsymbol{vv}}^{-1}\mathbf{A} + \mathbf{D}_p^{-1}\right]^{-1}\\ \mathbf{m} &= \mathbf{C}_m\left[\mathbf{A}^T\mathbf{C}_{\boldsymbol{vv}}^{-1}\boldsymbol{y} + \mathbf{D}_p^{-1}\boldsymbol{p}\right],\end{aligned} \tag{4}$$

and $\mathbf{D}_p$ is the short hand notation for $D(\boldsymbol{\tau}_p)$. Furthermore, we define

$$\boldsymbol{\tau}_m = \left[\tau_{m_1}\dots\tau_{m_N}\right]^T = \mathrm{diag}(\mathbf{C}_m). \tag{5}$$

Utilizing the property of multivariate Gaussian distribution, the marginal of (3) can be expressed as follows:

$$\begin{aligned}&\frac{\int_{\boldsymbol{x}_{\bar{i}}} q(\boldsymbol{x})d\boldsymbol{x}_{\bar{i}}}{q_i(x_i)} = \frac{\mathcal{N}(x_i;m_i,[\mathbf{C}_m]_{ii})}{q_i(x_i)}\\ &\propto \mathcal{N}\left(x_i;\frac{\tau_{p_i}m_i-\tau_{m_i}p_i}{\tau_{p_i}-\tau_{m_i}},\frac{\tau_{m_i}\tau_{p_i}}{\tau_{p_i}-\tau_{m_i}}\right) := \mathcal{N}\left(x_i;r_i,\tau_{r_i}\right),\end{aligned} \tag{6}$$

where $\boldsymbol{x}_{\bar{i}}$ represents a vector that is the same as $\boldsymbol{x}$ except that it excludes the i-th entry. The extrinsic from $p(\boldsymbol{y}|\boldsymbol{x})$ to variable node i is represented by the normal distribution $\mathcal{N}(x_i;r_i,\tau_{r_i})$.

B. Approximation

To approximate the belief at variable node i as a Gaussian distribution, we minimize the KLD

$$\begin{aligned}&\arg\min_q \mathrm{KLD}\left[\frac{p_i(x_i)\mathcal{N}(x_i;r_i,\tau_{r_i})}{Z_i(r_i,\tau_{r_i})}||b_{app,i}(x_i)\right]\\ &\Leftrightarrow \arg\min_{\widehat{x}_i,\tau_{x_i}} \mathrm{KLD}\left[\frac{p_i(x_i)\mathcal{N}(x_i;r_i,\tau_{r_i})}{Z_i(r_i,\tau_{r_i})}||\mathcal{N}(x_i;\widehat{x}_i,\tau_{x_i})\right],\end{aligned} \tag{7}$$

where $Z_i(r_i,\tau_{r_i})$ is the normalization factor given by

$$Z_i(r_i,\tau_{r_i}) = \int p_i(x)\mathcal{N}(x;r_i,\tau_{r_i})dx. \tag{8}$$

We define

$$\begin{aligned}g_i(r_i,\tau_{r_i}) &= \frac{\int x p_i(x)\mathcal{N}(x;r_i,\tau_{r_i})dx}{Z_i(r_i,\tau_{r_i})}\\ g_i'(r_i,\tau_{r_i}) &= \frac{\partial g_i(r_i,\tau_{r_i})}{\partial r_i}.\end{aligned} \tag{9}$$

Set the partial derivative of the KLD in (7) with respect to $\widehat{x}_i$ and τ_{x_i} to zero, we obtain

$$\begin{aligned}\widehat{x}_i &= \frac{\int x p_i(x)\mathcal{N}(x;r_i,\tau_{r_i})dx}{Z_i(r_i,\tau_{r_i})} = g(r_i,\tau_{r_i})\\ \tau_{x_i} &= \frac{\int (x-\widehat{x}_i)^2 p_i(x)\mathcal{N}(x;r_i,\tau_{r_i})dx}{Z_i(r_i,\tau_{r_i})} = \tau_{r_i}g'(r_i,\tau_{r_i}).\end{aligned} \tag{10}$$

It is worth noting that (10) is equivalent to

$$\begin{aligned}\widehat{x}_i &= \tau_{r_i}\frac{\partial \ln Z_i(r_i,\tau_{r_i})}{\partial r_i} + r_i,\\ \tau_{x_i} &= \tau_{r_i}^2\frac{\partial^2 \ln Z_i(r_i,\tau_{r_i})}{\partial r_i^2} + \tau_{r_i}.\end{aligned} \tag{11}$$

C. Pass the approximation to the factor node $p(\boldsymbol{y}|\boldsymbol{x})$

The message distribution (approximated prior) passed from variable node i to factor node $p(\boldsymbol{y}|\boldsymbol{x})$ is proportional to the quotient of two Gaussian probability density functions. Therefore, this message distribution is also Gaussian if $\tau_{r_i} \geq \tau_{x_i}$. Specifically, it is defined as

$$\mathcal{N}(x; p_i, \tau_{p_i}) \propto \frac{\mathcal{N}(x; \widehat{x}_i, \tau_{x_i})}{\mathcal{N}(x; r_i, \tau_{r_i})}. \tag{12}$$

From (12), p_i and τ_{p_i} are obtained by

$$\begin{aligned} \tau_{p_i} &= (\tfrac{1}{\tau_{x_i}} - \tfrac{1}{\tau_{r_i}})^{-1} = \tfrac{\tau_{r_i}\tau_{x_i}}{\tau_{r_i}-\tau_{x_i}} \\ p_i &= \tau_{p_i}\left(\tfrac{\widehat{x}_i}{\tau_{x_i}} - \tfrac{r_i}{\tau_{r_i}}\right) = \tfrac{\tau_{r_i}\widehat{x}_i - \tau_{x_i} r_i}{\tau_{x_i}\tau_{r_i}}. \end{aligned} \tag{13}$$

It is important to note that if the sequential updating method is used, the complexity of the matrix inverse operation in line 5 can be reduced by employing the matrix inverse lemma. Let's denote the resulting value of τ_{p_i} as $\tau_{p_i}^{new}$ during the update of the i-th element. We define $\Delta_p = \frac{1}{\tau_{p_i}^{new}} - \frac{1}{\tau_{p_i}}$. Moreover, we define $h_{\mathbf{C}}(\cdot)$ as the update of $\mathbf{C}_m$ with the new value of $\tau_{p_i}^{new}$ as follows:

$$\begin{aligned} &\mathbf{C}_m^{new} := h_{\mathbf{C}}(\mathbf{C}_m, \boldsymbol{e}_i, \Delta_p) = \left[\mathbf{C}_m^{-1} + \Delta_p \boldsymbol{e}_i \boldsymbol{e}_i^T\right]^{-1} \\ &= \mathbf{C}_m - \mathbf{C}_m \boldsymbol{e}_i \left(1/\Delta_p + \boldsymbol{e}_i^T \mathbf{C}_m \boldsymbol{e}_i\right)^{-1} \boldsymbol{e}_i^T \mathbf{C}_m. \end{aligned} \tag{14}$$

Here, $\boldsymbol{e}_i$ is a unit vector with only the i-th entry set to 1. To handle the cycles, we define $\boldsymbol{e}_0 = \boldsymbol{e}_N$.

The computation for updating $\mathbf{m}$ can also be simplified with the same technique. Define $\Omega_p = \frac{p_i^{new}}{\tau_{p_i}^{new}} - \frac{p_i}{\tau_{p_i}}$. We denote $h_{\mathbf{m}}(\cdot)$ as its update equation:

$$\begin{aligned} &\mathbf{m}^{new} := h_{\mathbf{m}}(\mathbf{m}, \mathbf{C}_m, \boldsymbol{e}_i, \Delta_p, \Omega_p) \\ &= \mathbf{C}_m^{new}(\mathbf{A}^T \mathbf{C}_{vv}^{-1} \boldsymbol{y} + \mathbf{D}_p \boldsymbol{p} + \Omega_p \boldsymbol{e}_i) \\ &= \mathbf{m} + \tfrac{\Omega_p - \Delta_p \boldsymbol{e}_i^T \mathbf{m}}{1 + \Delta_p \boldsymbol{e}_i^T \mathbf{C}_m \boldsymbol{e}_i} \mathbf{C}_m \boldsymbol{e}_i \end{aligned} \tag{15}$$

To summarize, we repeatedly compute the messages from the factor nodes to the variable nodes and then compute the message from the variable nodes back to the factor nodes until convergence. The final approximation for $p(\boldsymbol{x}|\boldsymbol{y})$ is given by $\mathcal{N}(\boldsymbol{x}; \mathbf{m}, \mathbf{C}_m)$. We have presented these steps in Algorithm 1. Additionally, note that these update steps can be performed in parallel which will generate an algorithm similar to VAMP but with individual variance updates. By exploiting matrix inverse lemma, the sequential update has the same complexity as the parallel update.

III. Relation of EP pdf Division and Component-Wise Conditionally Unbiased (CWCU) MMSE Estimation for Extrinsics

In the context of estimating the i-th entry of the signal vector $\boldsymbol{x}$, we can follow the approach of the CWCU estimator. This approach assumes that the i-th entry of $\boldsymbol{x}$ is deterministic while the other entries are random. With the approximated prior $q_i(x_i) = \mathcal{N}(x_i; p_i, \tau_{p_i})$, the Linear MMSE (LMMSE) for estimating signal $\boldsymbol{x}$ is

$$\begin{aligned} &\mathbf{m} = \boldsymbol{p} + (\mathbf{D}_p^{-1} + \mathbf{A}^T \mathbf{C}_{\boldsymbol{vv}}^{-1} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{C}_{\boldsymbol{vv}}^{-1} (\boldsymbol{y} - \mathbf{A}\boldsymbol{p}); \\ &\mathbf{C}_m = \left(\mathbf{D}_p^{-1} + \mathbf{A}^T \mathbf{C}_{\boldsymbol{vv}}^{-1} \mathbf{A}\right)^{-1}, \end{aligned} \tag{16}$$

Algorithm 1 reVAMP (Gaussian measurement noise sequential update)

Require: $\boldsymbol{y}$, $\mathbf{A}$, $p_{\boldsymbol{x}}(\boldsymbol{x})$, $p_{\boldsymbol{v}}(\boldsymbol{v})$, define $\boldsymbol{e}_0 := \boldsymbol{e}_N$

1: Initialize: $\boldsymbol{\tau}_p$, $\boldsymbol{p}$, $\Delta_p = 0$, $\mathbf{C}_m = \left(\mathbf{A}^T \mathbf{C}_{\boldsymbol{vv}}^{-1} \mathbf{A} + \mathbf{D}_p^{-1}\right)^{-1}$
2: **repeat**[For iteration step t]
3:     **repeat** [For each $i = 1 \ldots N$]
4:         [Update the posterior approximation]
5:         $\mathbf{C}_m^{t,i} = h_{\mathbf{C}}(\mathbf{C}_m^{t,i-1}, \boldsymbol{e}_{i-1}, \Delta_p^{t,i-1})$
6:         $\mathbf{m}^{t,i} = h_{\mathbf{m}}(\mathbf{m}^{t,i-1}, \mathbf{C}_m^{t,i-1}, \boldsymbol{e}_{i-1}, \Delta_p^{t,i-1}, \Omega_p^{t,i-1})$
7:         [Update the extrinsic]
8:         $\boldsymbol{\tau}_m^{t,i} = \mathrm{diag}(\mathbf{C}_m^{t,i})$
9:         $r_i^t = \frac{\tau_{p_i}^{t-1} m_i^t - \tau_{m_i}^t p_i^{t-1}}{\tau_{p_i}^{t-1} - \tau_{m_i}^t}$
10:        $\tau_{r_i}^t = \frac{\tau_{m_i}^t \tau_{p_i}^{t-1}}{\tau_{p_i}^{t-1} - \tau_{m_i}^t}$
11:        [Approximate the marginal posterior]
12:        $\widehat{x}_i^t = g_i(r_i^t, \tau_{r_i}^t)$
13:        $\tau_{x_i}^t = \tau_{r_i}^t g_i'(r_i^t, \tau_{r_i}^t)$
14:        [Propagate the approximation back]
15:        $p_i^t = \frac{\tau_{r_i}^t \widehat{x}_i^t - \tau_{x_i}^t r_i^t}{\tau_{r_i}^t - \tau_{x_i}^t}$
16:        $\tau_{p_i}^t = \frac{\tau_{r_i}^t \tau_{x_i}^t}{\tau_{r_i}^t - \tau_{x_i}^t}$
17:        $\Delta_p^{t,i} = \frac{1}{\tau_{p_i^t}} - \frac{1}{\tau_{p_i^{t-1}}}$
18:        $\Omega_p^{t,i} = \frac{p_i^t}{\tau_{p_i^t}} - \frac{p_i^{t-1}}{\tau_{p_i^{t-1}}}$
19:     **until** All i-s have been updated
20:     $[\mathbf{C}_m^{t+1,0}, \boldsymbol{p}^{t+1,0}, \Delta_p^{t+1,0}, \Omega_p^{t+1,0}] = [\mathbf{C}_m^{t,N}, \boldsymbol{p}^{t,N}, \Delta_p^{t,N}, \Omega_p^{t,N}]$
21: **until** Convergence

where $\boldsymbol{p} = \mathcal{E}\{\boldsymbol{x}\}$ denotes the prior mean and $\mathbf{D}_p = \mathbb{E}\{(\boldsymbol{x} - \boldsymbol{p})(\boldsymbol{x} - \boldsymbol{p})^T\} = D(\begin{bmatrix}\tau_{p_1} & \ldots & \tau_{p_N}\end{bmatrix}^T)$ is the prior covariance matrix. Based on (16), we define

$$\begin{aligned} &\boldsymbol{F} = \mathbf{C}_m^{-1} = (\mathbf{D}_p^{-1} + \mathbf{A}^T \mathbf{C}_{\boldsymbol{vv}}^{-1} \mathbf{A}); \\ &\boldsymbol{\tau}_m = \mathrm{diag}\left(\mathbf{C}_m\right). \end{aligned} \tag{17}$$

When considering only the i-th entry of the signal vector $\boldsymbol{x}$ to be deterministic (assume the prior variance to be $+\infty$), and treating the other entries as random variables, we can estimate the i-th entry of $\boldsymbol{x}$ and the associated error using the following equations:

$$\begin{aligned} r_i &= \boldsymbol{e}_i^T \boldsymbol{p} + \boldsymbol{e}_i^T (\boldsymbol{F} - \tfrac{1}{\tau_{p_i}} \boldsymbol{e}_i \boldsymbol{e}_i^T)^{-1} \mathbf{A}^T \mathbf{C}_{\boldsymbol{vv}}^{-1} (\boldsymbol{y} - \mathbf{A}\boldsymbol{p}); \\ \tau_{r_i} &= \boldsymbol{e}_i^T (\boldsymbol{F} - \tfrac{1}{\tau_{p_i}} \boldsymbol{e}_i \boldsymbol{e}_i^T)^{-1} \boldsymbol{e}_i. \end{aligned} \tag{18}$$

We will first show that the estimation step in (18) can be used to obtain the extrinsic from the LMMSE step. After that, we will demonstrate that (18) is CWCU estimation.

Note that $\boldsymbol{e}_i^T \boldsymbol{F}^{-1} \boldsymbol{e}_i = \boldsymbol{\tau}_{m_i}$. We apply matrix inverse lemma and the common term $\boldsymbol{e}_i^T (\boldsymbol{F} - \frac{1}{\tau_{p_i}} \boldsymbol{e}_i \boldsymbol{e}_i^T)^{-1}$ in τ_{r_i} and r_i can be simplified to

$$\begin{aligned} &\boldsymbol{e}_i^T (\boldsymbol{F} - \tfrac{1}{\tau_{p_i}} \boldsymbol{e}_i \boldsymbol{e}_i^T)^{-1} \\ &= \boldsymbol{e}_i^T \boldsymbol{F}^{-1} - \boldsymbol{e}_i^T \boldsymbol{F}^{-1} \boldsymbol{e}_i (\boldsymbol{e}_i^T \boldsymbol{F}^{-1} \boldsymbol{e}_i - \tau_{p_i})^{-1} \boldsymbol{e}_i^T \boldsymbol{F}^{-1} \\ &= (1 - \tfrac{\tau_{m_i}}{\tau_{m_i} - \tau_{p_i}}) \boldsymbol{e}^T \boldsymbol{F}^{-1} = \tfrac{\tau_{p_i}}{\tau_{p_i} - \tau_{m_i}} \boldsymbol{e}^T \boldsymbol{F}^{-1}. \end{aligned} \tag{19}$$

With this simplification, we obtain τ_{r_i} and r_i by

$$\begin{aligned} \tau_{r_i} &= \tfrac{\tau_{m_i}\tau_{p_i}}{\tau_{p_i} - \tau_{m_i}} \Rightarrow \tfrac{1}{\tau_{r_i}} = \tfrac{1}{\tau_{m_i}} - \tfrac{1}{\tau_{p_i}}, \\ r_i &= p_i + \tfrac{\tau_{r_i}}{\tau_{m_i}} \boldsymbol{e}_i^T \boldsymbol{F}^{-1} \mathbf{A}^T \mathbf{C}_{\boldsymbol{vv}}^{-1} (\boldsymbol{y} - \mathbf{A}\boldsymbol{p}). \end{aligned} \tag{20}$$

Observe the LMMSE estimate in (16), we have

$$m_i = p_i + \boldsymbol{e}_i^T \boldsymbol{F}^{-1} \mathbf{A}^T \mathbf{C}_{\boldsymbol{vv}}^{-1} (\boldsymbol{y} - \mathbf{A}\boldsymbol{p}). \tag{21}$$

Compare r_i in (20) and m_i in (21), we obtain the relation

$$\begin{aligned} & r_i = \tfrac{\tau_{r_i}}{\tau_{m_i}} m_i + (1 - \tfrac{\tau_{r_i}}{\tau_{m_i}}) p_i = \tfrac{\tau_{r_i}}{\tau_{m_i}} m_i - \tfrac{\tau_{r_i}}{\tau_{p_i}} p_i \\ & \Rightarrow \tfrac{r_i}{\tau_{r_i}} = \tfrac{m_i}{\tau_{m_i}} - \tfrac{p_i}{\tau_{p_i}}. \end{aligned} \tag{22}$$

From (20) and (22), we observe that the estimation of r_i and τ_{r_i} given by (18) matches the extrinsic obtained from (6).
In the following, we will demonstrate that equation (18) corresponds to CWCU estimation.
An important relationship to note is:

$$\begin{aligned} & \boldsymbol{F}^{-1} \mathbf{A}^T \mathbf{C}_{\boldsymbol{vv}}^{-1} \mathbf{A} = [(\mathbf{A}^T \mathbf{C}_{\boldsymbol{vv}}^{-1} \mathbf{A})^{-1} \mathbf{D}_p^{-1} + \mathbf{I}]^{-1} \\ & = \mathbf{I} - \boldsymbol{F}^{-1} \mathbf{D}_p^{-1}. \end{aligned} \tag{23}$$

If $\boldsymbol{p} = \mathbf{0}$, it can be shown that $\mathbf{r}$ corresponds to the CWCU estimator described in [6]. According to the result from [6], the CWCU estimator is given by:

$$\begin{aligned} & m_{i,u} = (\boldsymbol{e}_i^T [(\mathbf{A}^T \mathbf{C}_{\boldsymbol{vv}}^{-1} \mathbf{A})^{-1} \mathbf{D}_p^{-1} + \mathbf{I}]^{-1} \boldsymbol{e}_i)^{-1} m_i \\ & = (1 - \boldsymbol{e}_i^T \boldsymbol{F}^{-1} \mathbf{D}_p^{-1} \boldsymbol{e}_i)^{-1} m_i = \tfrac{\tau_{p_i}}{\tau_{p_i} - \tau_{m_i}} m_i = \tfrac{\tau_{r_i}}{\tau_{m_i}} m_i. \end{aligned} \tag{24}$$

If we compare r_i in (22) and $m_{i,u}$ in (24) when $\boldsymbol{p} = \mathbf{0}$, we observe that r_i is equivalent to the CWCU estimation. However, when $\boldsymbol{p} \neq \mathbf{0}$, we can split $\boldsymbol{p}$ as $\boldsymbol{p} = \sum_{i=1}^N \boldsymbol{e}_i \, p_i$. Subsequently, from (20), we obtain the following expression:

$$\begin{aligned} & r_i = p_i - \tfrac{\tau_{r_i}}{\tau_{m_i}} \boldsymbol{e}_i^T \boldsymbol{F}^{-1} \mathbf{A}^T \mathbf{C}_{\boldsymbol{vv}}^{-1} \mathbf{A} \boldsymbol{e}_i p_i \\ & - \tfrac{\tau_{r_i}}{\tau_{m_i}} \boldsymbol{e}_i^T \boldsymbol{F}^{-1} \mathbf{A}^T \mathbf{C}_{\boldsymbol{vv}}^{-1} \mathbf{A} \sum_{j=1, j\neq i}^N \boldsymbol{e}_j p_j \\ & + \tfrac{\tau_{r_i}}{\tau_{m_i}} \boldsymbol{e}_i^T \boldsymbol{F}^{-1} \mathbf{A}^T \mathbf{C}_{\boldsymbol{vv}}^{-1} \boldsymbol{y}. \end{aligned} \tag{25}$$

From (23) and (24), we can deduce that:

$$\begin{aligned} & \boldsymbol{e}_i^T \boldsymbol{F}^{-1} \mathbf{A}^T \mathbf{C}_{\boldsymbol{vv}}^{-1} \mathbf{A} \boldsymbol{e}_i \\ & = \boldsymbol{e}_i^T [(\mathbf{A}^T \mathbf{C}_{\boldsymbol{vv}}^{-1} \mathbf{A})^{-1} \mathbf{D}_p^{-1} + \mathbf{I}]^{-1} \boldsymbol{e}_i = \tfrac{\tau_{m_i}}{\tau_{r_i}}. \end{aligned} \tag{26}$$

Therefore, the estimate in (25) can be written as

$$r_i = \tfrac{\tau_{r_i}}{\tau_{m_i}} \boldsymbol{e}_i^T \boldsymbol{F}^{-1} \mathbf{A}^T \mathbf{C}_{\boldsymbol{vv}} (\boldsymbol{y} - \mathbf{A} \textstyle\sum_{j=1, j\neq i}^N \boldsymbol{e}_j p_j). \tag{27}$$

The conditional expectation is

$$\begin{aligned} & \mathbb{E}_{\boldsymbol{x},\boldsymbol{v}|x_i}[r_i] = \mathbb{E}_{\boldsymbol{x},\boldsymbol{v}|x_i}[\tfrac{\tau_{r_i}}{\tau_{m_i}} \boldsymbol{e}_i^T \boldsymbol{F}^{-1} \mathbf{A}^T \mathbf{C}_{\boldsymbol{vv}} (\mathbf{A} \textstyle\sum_{j=1}^N \boldsymbol{e}_j x_j \\ & + \boldsymbol{v} - \mathbf{A} \textstyle\sum_{j=1, j\neq i}^N \boldsymbol{e}_j p_j)] = x_i, \end{aligned} \tag{28}$$

which is indeed conditionally unbiased.
Next, we want to find out whether the estimation error corresponds to $\boldsymbol{\tau}_r$ given by (18). To analyze the estimation errors, we can represent the estimations in equation (18) as vectors:

$$\mathbf{r} = \begin{bmatrix} r_1 & \ldots & r_N \end{bmatrix}^T; \; \boldsymbol{\tau}_r = \begin{bmatrix} \tau_{r_1} & \ldots & \tau_{r_N} \end{bmatrix}^T. \tag{29}$$

From (22), the vector $\mathbf{r}$ can also be expressed as

$$\mathbf{r} = \boldsymbol{p} + \mathbf{D}_r \mathbf{D}_m^{-1} (\mathbf{m} - \boldsymbol{p}). \tag{30}$$

The estimation error correlation matrix is given by

$$\begin{aligned} & \mathbf{C}_{\mathbf{rr}} = \mathbb{E}[(\mathbf{r} - \boldsymbol{x})(\mathbf{r} - \boldsymbol{x})^T] \\ & = \mathbb{E}\Big[\big((\boldsymbol{p} - \boldsymbol{x}) + \mathbf{D}_r \mathbf{D}_m^{-1} (\mathbf{m} - \boldsymbol{p})\big)\big((\boldsymbol{p} - \boldsymbol{x}) + \mathbf{D}_r \mathbf{D}_m^{-1} (\mathbf{m} - \boldsymbol{p})\big)^T\Big] \end{aligned} \tag{31}$$

We observe that, according to (16), the term $\mathbf{m} - \boldsymbol{p}$ in (31) can be expressed as:

$$\mathbf{m} - \boldsymbol{p} = \boldsymbol{F}^{-1} \mathbf{A}^T \mathbf{C}_{\boldsymbol{vv}}^{-1} [\mathbf{A}(\boldsymbol{x} - \boldsymbol{p}) + v]. \tag{32}$$

By applying the matrix inverse lemma to $\boldsymbol{F}^{-1}$ and utilizing the relation given by (23), we can obtain the following expression:

$$\begin{aligned} & \mathbb{E}[(\mathbf{m} - \boldsymbol{p})(\boldsymbol{x} - \boldsymbol{p})^T] \\ & = \boldsymbol{F}^{-1} \mathbf{A}^T \mathbf{C}_{\boldsymbol{vv}}^{-1} \mathbf{A} \, \mathbb{E}[(\boldsymbol{x} - \boldsymbol{p})(\boldsymbol{x} - \boldsymbol{p})^T] \\ & = \boldsymbol{F}^{-1} \mathbf{A}^T \mathbf{C}_{\boldsymbol{vv}}^{-1} \mathbf{A} \mathbf{D}_p = \mathbf{D}_p - \boldsymbol{F}^{-1} = \mathbf{D}_p \mathbf{A}^T \mathbf{C}_{yy}^{-1} \mathbf{A} \mathbf{D}_p. \end{aligned} \tag{33}$$

Similarly, we have

$$\begin{aligned} & \mathbb{E}[(\mathbf{m} - \boldsymbol{p})(\mathbf{m} - \boldsymbol{p})^T] \\ & = \boldsymbol{F}^{-1} \mathbf{A}^T \mathbf{C}_{\boldsymbol{vv}}^{-1} [\mathbf{A} \mathbf{D}_p \mathbf{A}^T + \mathbf{C}_{\boldsymbol{vv}}] \mathbf{C}_{\boldsymbol{vv}}^{-1} \mathbf{A} \boldsymbol{F}^{-1} \\ & = \mathbf{D}_p \mathbf{A}^T \mathbf{C}_{\boldsymbol{yy}}^{-1} \mathbf{A} \mathbf{D}_p. \end{aligned} \tag{34}$$

Applying (34) and (33) into (31), it follows that

$$\begin{aligned} & \mathbf{C}_{\mathbf{rr}} = \mathbf{D}_p + \mathbf{D}_r \mathbf{D}_m^{-1} \mathbf{D}_p \mathbf{A}^T \mathbf{C}_{\boldsymbol{yy}}^{-1} \mathbf{A} \mathbf{D}_p \mathbf{D}_m^{-1} \mathbf{D}_r - \\ & \mathbf{D}_r \mathbf{D}_m^{-1} \mathbf{D}_p \mathbf{A}^T \mathbf{C}_{\boldsymbol{yy}}^{-1} \mathbf{A} \mathbf{D}_p - \mathbf{D}_p \mathbf{A}^T \mathbf{C}_{\boldsymbol{yy}}^{-1} \mathbf{A} \mathbf{D}_p \mathbf{D}_m^{-1} \mathbf{D}_r. \end{aligned} \tag{35}$$

We can express (20) in the form of diagonal matrices as follows:

$$\mathbf{I} - \mathbf{D}_r \mathbf{D}_m^{-1} = \mathbf{D}_r (\mathbf{D}_r^{-1} - \mathbf{D}_m^{-1}) = -\mathbf{D}_r \mathbf{D}_p^{-1}. \tag{36}$$

With this relation, we further simplify $\mathbf{C}_{\mathbf{rr}}$ by

$$\begin{aligned} & \mathbf{C}_{\mathbf{rr}} = \mathbf{C}_{\mathbf{rr}} + \mathbf{D}_p \mathbf{A}^T \mathbf{C}_{\boldsymbol{yy}}^{-1} \mathbf{A} \mathbf{D}_p - \mathbf{D}_p \mathbf{A}^T \mathbf{C}_{\boldsymbol{yy}}^{-1} \mathbf{A} \mathbf{D}_p \\ & = \mathbf{D}_p - \mathbf{D}_p \mathbf{A}^T \mathbf{C}_{\boldsymbol{yy}}^{-1} \mathbf{A} \mathbf{D}_p + \mathbf{D}_r \mathbf{A}^T \mathbf{C}_{\boldsymbol{yy}}^{-1} \mathbf{A} \mathbf{D}_r \end{aligned} \tag{37}$$

To establish a relationship between $\mathbf{C}_{\mathbf{rr}}$ and $\mathbf{D}_r$, we observe that $\mathbf{C}_m$ defined in (16) can be expressed by

$$\mathbf{C}_m = \mathbf{D}_p - \mathbf{D}_p \mathbf{A}^T \mathbf{C}_{\boldsymbol{yy}}^{-1} \mathbf{A} \mathbf{D}_p. \tag{38}$$

Combine (37) with (38),

$$\begin{aligned} & \mathbf{C}_{\mathbf{rr}} = \mathbf{C}_m + \mathbf{D}_r \mathbf{D}_p^{-1} (\mathbf{D}_p - \mathbf{C}_m) \mathbf{D}_p^{-1} \mathbf{D}_r \\ & \Rightarrow \mathrm{diag}(\mathbf{C}_{\mathbf{rr}}) = \mathbf{D}_m + \mathbf{D}_r (\mathbf{D}_p^{-1} - \mathbf{D}_p^{-2} \mathbf{D}_m) \mathbf{D}_r \\ & = \mathbf{D}_m + \mathbf{D}_r^2 \mathbf{D}_p^{-1} \mathbf{D}_m (\mathbf{D}_m^{-1} - \mathbf{D}_p^{-1}) = \mathbf{D}_r. \end{aligned} \tag{39}$$

Therefore, we can conclude that the extrinsic distribution represented by the parameters $(\mathbf{r}, \boldsymbol{\tau}_r)$ can also be interpreted as the mean and error of the CWCU estimation, which is used to recover the true signal $\boldsymbol{x}$ from observations $\boldsymbol{y}$ using an approximate prior distribution $\mathcal{N}(\boldsymbol{p}, \mathbf{D}_p)$.

IV. Simulation Results

A. MMSE for Gaussian mixture model

Assume that the prior distribution of each element x_n of vector $\boldsymbol{x} \in \mathcal{R}^{N \times 1}$ is given by:

$$p_{x_i}(x_i) = \textstyle\sum_{n=1}^3 \alpha_n \mathcal{N}(x_i; \mu_{ni}, \sigma_{ni}^2); \; \sum_{n=1}^2 \alpha_n = 1. \tag{40}$$

For each combination sequence $[n_1, \ldots, n_N] \in \{1, 2, 3\}^N$, we can define a bijective mapping $l : \{1, 2, 3\}^N \to \mathbb{N}$

$$l = \sum_{i=1}^N (n_i - 1) \cdot 3^{i-1}. \tag{41}$$

We denote its inverse mapping as $\boldsymbol{n}_l = [n_{l1}, \ldots, n_{lN}]$. For simplicity, we define

$$\begin{aligned} & c_l = \alpha_1^{i_l} \alpha_2^{j_l} \alpha_3^{k_l}, \boldsymbol{\mu}_l = [\mu_{n_{l1}1}, \ldots, \mu_{n_{lN}N}]^T, \\ & \boldsymbol{\Sigma}_l = \mathrm{diag}[\sigma_{n_{l1}1}^2, \ldots, \sigma_{n_{lN}N}^2], \end{aligned} \tag{42}$$

where i_l, j_l, k_l are the numbers of 1-s 2-s and 3-s in $\boldsymbol{n}_l$. In this case, the distribution of the vector $\boldsymbol{x}$ can be presented as

$$p(\boldsymbol{x}) = \prod_{i=1}^{N} p_{x_i}(x_i) = \sum_{l=1}^{3^N} c_l \ \mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l). \tag{43}$$

Due to the Bayesian law, the exact posterior first- and second-order moments are computed as

$$\begin{aligned} \mathbb{E}[\boldsymbol{x}] &= \int \boldsymbol{x} p(\boldsymbol{x}|\boldsymbol{y}) d\boldsymbol{x} = \frac{\int \boldsymbol{x} p(\boldsymbol{x}) p(\boldsymbol{y}|\boldsymbol{x}) d\boldsymbol{x}}{\int p(\boldsymbol{x}) p(\boldsymbol{y}|\boldsymbol{x}) d\boldsymbol{x}} \\ \mathbb{E}[\boldsymbol{x}\boldsymbol{x}^T] &= \int \boldsymbol{x}\boldsymbol{x}^T p(\boldsymbol{x}|\boldsymbol{y}) d\boldsymbol{x} = \frac{\int \boldsymbol{x}\boldsymbol{x}^T p(\boldsymbol{x}) p(\boldsymbol{y}|\boldsymbol{x}) d\boldsymbol{x}}{\int p(\boldsymbol{x}) p(\boldsymbol{y}|\boldsymbol{x}) d\boldsymbol{x}}, \end{aligned} \tag{44}$$

where

$$\begin{aligned} &\int p(\boldsymbol{y}|\boldsymbol{x}) \mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l) d\boldsymbol{x} = (2\pi)^{-\frac{M}{2}} |\mathbf{A}\boldsymbol{\Sigma}_l \mathbf{A}^T + \mathbf{C}_{\boldsymbol{vv}}|^{-\frac{1}{2}} \\ &\exp\left[-\tfrac{1}{2}(\boldsymbol{y} - \mathbf{A}\boldsymbol{\mu}_l)^T (\mathbf{A}\boldsymbol{\Sigma}_l \mathbf{A}^T + \mathbf{C}_{\boldsymbol{vv}})^{-1} (\boldsymbol{y} - \mathbf{A}\boldsymbol{\mu}_l)\right], \end{aligned} \tag{45}$$

$$\int p(\boldsymbol{x}) p(\boldsymbol{y}|\boldsymbol{x}) d\boldsymbol{x} = \sum_{l=1}^{3^N} c_l \int p(\boldsymbol{y}|\boldsymbol{x}) \mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l) d\boldsymbol{x} \tag{46}$$

$$\begin{aligned} &\int \boldsymbol{x} p(\boldsymbol{x}) p(\boldsymbol{y}|\boldsymbol{x}) d\boldsymbol{x} = \sum_{l=1}^{3^N} c_l \left[\boldsymbol{\mu}_l + (\boldsymbol{\Sigma}_l^{-1} + \mathbf{A}^T \mathbf{C}_{\boldsymbol{vv}}^{-1} \mathbf{A})^{-1} \mathbf{A}^T \right. \\ &\left. \mathbf{C}_{\boldsymbol{vv}}^{-1} (\boldsymbol{y} - \mathbf{A}\boldsymbol{\mu}_l)\right] \int p(\boldsymbol{y}|\boldsymbol{x}) \mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l) d\boldsymbol{x}, \end{aligned} \tag{47}$$

$$\begin{aligned} &\int \boldsymbol{x}\boldsymbol{x}^T p(\boldsymbol{x}) p(\boldsymbol{y}|\boldsymbol{x}) d\boldsymbol{x} = \sum_{l=1}^{3^N} c_l \left[(\boldsymbol{\Sigma}_l^{-1} + \mathbf{A}^T \mathbf{C}_{\boldsymbol{vv}}^{-1} \mathbf{A})^{-1} + \right. \\ &\left. \boldsymbol{\mu_l} \boldsymbol{\mu_l}^T\right] \int p(\boldsymbol{y}|\boldsymbol{x}) \mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l) d\boldsymbol{x}. \end{aligned} \tag{48}$$

The true MMSE estimation mean and covariance matrix for Gaussian mixture can therefore be expressed as

$$\begin{aligned} \widehat{\boldsymbol{x}}_{\mathrm{MMSE}} &= \mathbb{E}[\boldsymbol{x}], \\ \mathbf{C}_{\mathrm{MMSE}} &= \mathbb{E}[\boldsymbol{x}\boldsymbol{x}^T] - \mathbb{E}[\boldsymbol{x}]\, \mathbb{E}[\boldsymbol{x}]^T. \end{aligned} \tag{49}$$

B. reVAMP algorithm

The computation of the posterior (10) in reVAMP with a Gaussian mixture prior (40) can be derived analogously to (49). Thus we have

$$\widehat{x}_i = \frac{\int x \sum_{n=1}^{3} \alpha_n \mathcal{N}(x_i; \mu_{ni}, \sigma_{ni}^2) \mathcal{N}(x; r_i, \tau_{r_i}) dx}{\int \sum_{n=1}^{3} \alpha_n \mathcal{N}(x_i; \mu_{ni}, \sigma_{ni}^2) \mathcal{N}(x; r_i, \tau_{r_i}) dx} \tag{50}$$

$$\tau_{x_i} = \frac{\int x^2 \sum_{n=1}^{3} \alpha_n \mathcal{N}(x_i; \mu_{ni}, \sigma_{ni}^2) \mathcal{N}(x; r_i, \tau_{r_i}) dx}{\int \sum_{n=1}^{3} \alpha_n \mathcal{N}(x_i; \mu_{ni}, \sigma_{ni}^2) \mathcal{N}(x; r_i, \tau_{r_i}) dx} - \widehat{x}_i^2 \tag{51}$$

C. MATLAB simulation for Gaussian mixture model

In the simulation, the measurement matrix $\mathbf{A}$ has the dimension $M \times N := 10 \times 5$. Its entries are independently drawn from $\mathcal{N}(0, 1)$ Gaussian distribution. The prior distribution of signal vector $\boldsymbol{x}$ follows

$$\begin{aligned} &\forall i \in 1, \ldots, N, p_{x_i}(x_i) = 0.25\mathcal{N}(x_i; 0, 4 \cdot i) \\ &+0.5\mathcal{N}(x_i; 0, 1) + 0.25\mathcal{N}(x_i; 0, 0.25 \cdot i). \end{aligned} \tag{52}$$

The measurement noise is set to be a random vector following $\mathcal{N}(\mathbf{0}, \mathbf{I})$. To determine the performance of the reVAMP posterior estimation of the first- and second-order moment, we compare the KL divergent between the Gaussian distribution generated from MMSE solution and the Gaussian distribution given by reVAMP

$$\mathrm{KLD}[\mathcal{N}(\boldsymbol{x}; \widehat{\boldsymbol{x}}_{\mathrm{MMSE}}, \mathbf{C}_{\mathrm{MMSE}}) || \mathcal{N}(\boldsymbol{x}; \mathbf{m}, \mathbf{C}_m)]. \tag{53}$$

The simulation results can be found in fig. 2. The KL-Divergence shown in the figure is the average of 200 simulation results. From the figure, we see that sequential updates and parallel updates have the same steady state. However, when the sequential update is used, the algorithm converges faster than the parallel update method.

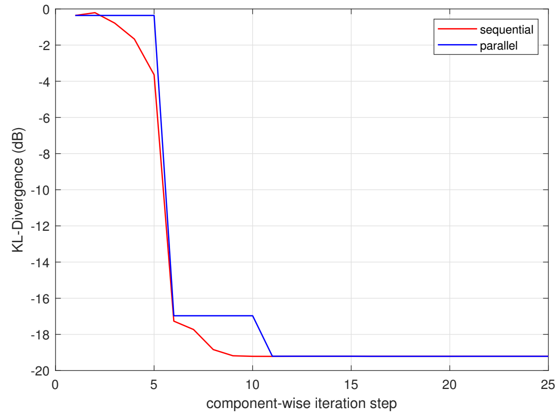

Fig. 2. Performance of sequential and parallel update scheme by comparing their KL divergence between Gaussian generated from MMSE result

V. Concluding Remarks

In this paper, we present an iterative method to calculate the posteriors in a linear mixing model, where the prior is considered to be independent and the measurement noise is Gaussian. Similar to VAMP, we apply EP-like approximations to the factor graph. However, since the system dimension is not high, the complexity of matrix inversion in each iteration is assumed affordable. Further research is required for the convergence analysis.

Acknowledgements EURECOM's research is partially supported by its industrial members: ORANGE, BMW, SAP, iABG, Norton LifeLock, and by the Franco-German projects CellFree6G and 5G-OPERA.

References

[1] Mahdi Triki, Dirk TM Slock, Vincent Rigal, and Pierrick François, "Mobile Terminal Positioning via Power Delay Profile Fingerprinting: Reproducible Validation Simulations," in *IEEE Vehicular Technology Conference*, 2006, pp. 1–5.

[2] Qiuyun Zou and Hongwen Yang, "A Concise Tutorial on Approximate Message Passing," *arXiv preprint arXiv:2201.07487*, 2022.

[3] Mohsen Bayati and Andrea Montanari, "The Dynamics of Message Passing on Dense Graphs, with Applications to Compressed Sensing," *IEEE Transactions on Information Theory*, vol. 57, no. 2, pp. 764–785, 2011.

[4] Philip Schniter, Sundeep Rangan, and Alyson K Fletcher, "Vector Approximate Message Passing for the Generalized Linear Model," in *2016 50th Asilomar Conference on Signals, Systems and Computers*. IEEE, 2016, pp. 1525–1529.

[5] Robert F.H. Fischer, Carmen Sippel, and Norbert Goertz, "Vamp with vector-valued diagonalization," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 9110–9114.

[6] Mahdi Triki and Dirk TM Slock, "Component-Wise Conditionally Unbiased Bayesian Parameter Estimation: General Concept and Applications to Kalman Filtering and LMMSE Channel Estimation," in *Conference Record of the Thirty-Ninth Asilomar Conference onSignals, Systems and Computers, 2005.* IEEE, 2005, pp. 670–674.

[7] Sundeep Rangan, Philip Schniter, and Alyson K. Fletcher, "Vector Approximate Message Passing," 2018.