

# Expectations in Expectation Propagation

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

**Abstract**—Expectation Propagation (EP) is a widely used message-passing algorithm that decomposes a global inference problem into multiple local ones. It approximates marginal distributions (beliefs) using intermediate functions (messages). While beliefs must be proper probability distributions that integrate to one, messages may have infinite integral values. In Gaussian-projected EP, such messages take a Gaussian form and appear as if they have “negative” variances. Although allowed within the EP framework, these negative-variance messages can impede algorithmic progress.

In this paper, we investigate EP in linear models and analyze the relationship between the corresponding beliefs. Based on the analysis, we propose both non-persistent and persistent approaches that prevent the algorithm from being blocked by messages with infinite integral values.

Furthermore, by examining the relationship between the EP messages in linear models, we develop an additional approach that avoids the occurrence of messages with infinite integral values.

## I. INTRODUCTION

Signal recovery is a fundamental problem in signal processing, with a wide range of applications. In the Bayesian framework, however, canonical inference methods such as MMSE and MAP suffer from exponential computational complexity as the problem dimension increases.

Graphical-model-based iterative methods have proven effective by exploiting structural properties of the underlying models [1]. Expectation Propagation (EP), introduced in [2], transforms a global inference problem into multiple local inference problems. EP can be interpreted as an iterative procedure for finding the stationary point of the constrained Bethe Free Energy (BFE) [3], [4]. Within this framework, EP approximates marginal distributions (beliefs) using functions known as messages, which correspond one-to-one with the Lagrange multipliers in the constrained BFE formulation. In EP and constrained BFE minimization, beliefs are required to be proper distributions—that is, they must be normalizable and integrate to one—whereas messages are allowed to have infinite integral values.

Although messages are not required to have finite integrals, messages with infinite integral values may produce beliefs that also fail to integrate to a finite value and may even prevent the algorithm from progressing.

### A. Main Contribution

The revisited Vector Approximate Message Passing (reVAMP) algorithm [5] is an instance of EP applied to linear models, whose joint probability density function factorizes into a likelihood term and symbol-wise priors. In this paper, we analyze the messages and beliefs in reVAMP and show that the belief at the likelihood factor node is guaranteed to have a finite integral value whenever the beliefs at the prior nodes have finite integral values.

Based on this observation, we propose both non-persistent and persistent approaches that prevent EP/reVAMP from being blocked by messages with negative variances, while still

allowing such messages to exist when they do not hinder the algorithm.

Furthermore, we show that the messages to the prior factors always have finite integral values whenever the messages to the likelihood factor have finite integral values. In addition, we investigate the Kullback–Leibler divergence (KLD) minimization step in EP and propose Analytic Continuation reVAMP (ACreVAMP), which prevents the occurrence of messages with negative variances.

### B. Notations

We use the notation  $[H]_{i,j}$  to denote the entry in the  $i$ -th row and  $j$ -th column  $H$ . Following MATLAB convention, we denote  $[H]_{:,i}$  as the  $i$ -th column of  $H$ , and for simplicity we write  $\mathbf{h}_i$  as shorthand for  $[H]_{:,i}$ .

Similarly,  $[\theta]_i$  denotes the  $i$ -th entry of the vector  $\theta$ , and we write  $\theta_i$  as shorthand for  $[\theta]_i$ .

We define the operator  $\text{Diag}(\theta)$ , which maps a vector  $\theta$  to a diagonal matrix  $\Theta = \text{Diag}(\theta)$  such that  $[\Theta]_{i,i} = \theta_i$ . Conversely, the operator  $\text{diag}(\Theta)$  extracts the diagonal entries of an arbitrary matrix  $\Theta$  into a vector  $\theta = \text{diag}(\Theta)$ , where  $[\theta]_i = [\Theta]_{i,i}$ .

Finally, we refer to a non-negative function as a proper distribution if it integrates to one.

## II. EXPECTATION PROPAGATION

Consider a factored joint PDF with non-negative factors (denoted as  $\forall \alpha : f_\alpha$ ):

$$p(\theta) \propto \prod_{\alpha} f_{\alpha}(\theta_{\alpha}), \quad (1)$$

where for each  $\alpha$ , the vector  $\theta_{\alpha}$  denotes a sub-vector of  $\theta$ . The goal of EP is to approximate the marginal statistics (e.g. mean and variance) of the distribution in (1).

A corresponding factor graph for (1) is constructed by introducing a factor node for every factor  $f_{\alpha}$  and a variable node for every variable  $\theta_i$  in the vector  $\theta$ . The graph is then completed by connecting each factor node to every variable node that appears as a parameter in that factor (the connecting line is called an edge).

Each edge carries two types of messages: a variable-to-factor message and a factor-to-variable message. Furthermore, each factor and variable node forms a belief that is uniquely determined by the incoming messages. These beliefs serve as approximations to the marginals of  $p(\theta)$ .

*Remark.* In EP, messages are not required to be proper distributions and may integrate to infinity, whereas beliefs must be proper distributions and must always normalize to one. For simplicity, throughout this paper we normalize even the messages to one whenever they have finite integral value.

EP iteratively updates all messages using the following rules:

- Belief at the factor node,  $\forall \alpha$  :

$$b_{f_\alpha}(\boldsymbol{\theta}_\alpha) \propto f_\alpha(\boldsymbol{\theta}_\alpha) \prod_{i \in N(\alpha)} \Delta_i^{f_\alpha \leftarrow \theta}(\theta_i). \quad (2)$$

- Message from factor to variable,  $\forall \alpha, i \in N(\alpha)$  :

$$\Delta_i^{f_\alpha \rightarrow \theta}(\theta_i) \propto \frac{\text{proj}[\int b_{f_\alpha}(\boldsymbol{\theta}_\alpha) \prod_{i' \in N(\alpha)/\{i\}} d\theta_{i'}]}{\Delta_i^{f_\alpha \leftarrow \theta}(\theta_i)}. \quad (3)$$

- Belief at the variable node,  $\forall i$  :

$$b_{\theta_i}(\theta_i) \propto \prod_{\alpha \in N(i)} \Delta_i^{f_\alpha \rightarrow \theta}(\theta_i). \quad (4)$$

- Message from variable to factor  $\forall i, \alpha \in N(i)$  :

$$\Delta_i^{f_\alpha \leftarrow \theta}(\theta_i) \propto \frac{b_{\theta_i}(\theta_i)}{\Delta_i^{f_\alpha \rightarrow \theta}(\theta_i)}. \quad (5)$$

In (2)–(5),  $N(i)$  denotes the set of factor indices whose corresponding factors are directly connected to variable  $\theta_i$  (i.e., in the neighborhood of  $\theta_i$ ). Similarly,  $N(\alpha)$  denote the set of variable indices connected to factor  $f_\alpha$ .

In (3), the operator  $\text{proj}[\cdot]$  projects an arbitrary distribution onto a desired family by minimizing the Kullback–Leibler divergence (KLD). In this paper, we consider only the Gaussian family  $\mathcal{G}$ . Thus,

$$\text{proj}[p(\theta)] \triangleq \arg \min_{q(\theta) \in \mathcal{G}} \text{KLD}[p(\theta) \| q(\theta)]. \quad (6)$$

If  $\mathcal{G}$  is Gaussian, minimizing (6) is equivalent to matching the mean and variance of  $p(\theta)$  and  $q(\theta)$ . We use the term "extrinsic of factor  $f_\alpha$ " to denote the product of incoming messages to  $f_\alpha$ , such as the message product in (2).

### A. Problem Formulation

Observe the message in (3). The division in this update may yield a function with an infinite integral value even when both the numerator and denominator are proper distributions. Consequently, the resulting message may appear to be a Gaussian with a "negative" variance. Although such messages are allowed within EP, they can hinder the algorithm because the beliefs in (2) and (4) are required to remain Gaussian.

In the remainder of this paper, we examine this negative-variance issue in linear models and propose several approaches to prevent the resulting blocking behavior.

### III. USEFUL RELATIONS

In this section, we list several useful relations that will be used throughout the paper.

#### A. Matrix Inversion Lemma

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

#### B. Matrix Determinant Lemma

$$\det(\mathbf{A} + \mathbf{BCD}) = \det(\mathbf{A}) \det(\mathbf{C}) \det(\mathbf{C}^{-1} + \mathbf{DA}^{-1}\mathbf{B}). \quad (7)$$

### C. Multivariate Gaussian Reproduction Lemma

$$\begin{aligned} \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}_1, \mathbf{C}_1) \mathcal{N}(\mathbf{A}\mathbf{x} | \boldsymbol{\mu}_2, \mathbf{C}_2) \\ = \mathcal{N}(\boldsymbol{\mu}_2 | \mathbf{A}\boldsymbol{\mu}_1, \mathbf{A}\mathbf{C}_1\mathbf{A}^T + \mathbf{C}_2) \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}_3, \mathbf{C}_3), \end{aligned} \quad (8)$$

where

$$\begin{aligned} \mathbf{C}_3 &= (\mathbf{C}_1^{-1} + \mathbf{A}^T \mathbf{C}_2^{-1} \mathbf{A})^{-1}; \\ \boldsymbol{\mu}_3 &= \mathbf{C}_3(\mathbf{C}_1^{-1} \boldsymbol{\mu}_1 + \mathbf{A}^T \mathbf{C}_2^{-1} \boldsymbol{\mu}_2). \end{aligned} \quad (9)$$

### IV. SYSTEM MODEL

We consider the linear measurement model:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}, \quad (10)$$

where  $\mathbf{A} \in \mathbb{R}^{M \times N}$ , the noise  $\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_v)$ , and the signal vector  $\mathbf{x}$  has independent entries, i.e.,  $\forall n \in [1, N] : x_n \sim p(x_n)$ . The joint PDF associated with the system model (10) factorizes as

$$p(\mathbf{x}, \mathbf{y}) = f_{\mathbf{y}}(\mathbf{x}) \prod_n f_{x_n}(x_n), \quad (11)$$

with factors

$$\begin{aligned} f_{\mathbf{y}}(\mathbf{x}) &\triangleq p(\mathbf{y} | \mathbf{x}) = \mathcal{N}(\mathbf{y} | \mathbf{A}\mathbf{x}, \mathbf{C}_v) \\ f_{x_n}(x_n) &\triangleq p(x_n). \end{aligned} \quad (12)$$

### V. BRIEF INTRODUCTION TO REVAMP

The explicit forms of the messages in reVAMP are summarized in Table I, where we denote

$$\mathcal{UN}(\boldsymbol{\theta} | \boldsymbol{\mu}, \mathbf{C}) \triangleq \exp -\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\mu})^T \mathbf{C}^{-1}(\boldsymbol{\theta} - \boldsymbol{\mu}). \quad (13)$$

Message and Beliefs	Notation	Explicit Form (proportional)
$x_n \rightarrow f_{\mathbf{y}}$	$\Delta_n^{f_{\mathbf{y}} \leftarrow x}(x_n, t)$	$\mathcal{UN}(x_n   \mu_{p_n}(t), \tau_{p_n}(t))$
$f_{x_n} \rightarrow x_n$	$\Delta_n^{f_{x_n} \rightarrow x}(x_n, t)$	$\mathcal{UN}(x_n   \mu_{r_n}(t), \tau_{r_n}(t))$
$f_{\mathbf{y}} \rightarrow x_n$	$\Delta_n^{f_{\mathbf{y}} \rightarrow x}(x_n, t)$	$\mathcal{UN}(x_n   \mu_{r_n}(t), \tau_{r_n}(t))$
$x_n \rightarrow f_{x_n}$	$\Delta_n^{f_{x_n} \leftarrow x}(x_n, t)$	$\mathcal{UN}(x_n   \mu_{p_n}(t), \tau_{p_n}(t))$
Belief at $f_{\mathbf{y}}$	$b_{f_{\mathbf{y}}}(\mathbf{x}, t)$	$f_{\mathbf{y}}(\mathbf{x}) \prod_n \Delta_n^{f_{\mathbf{y}} \leftarrow x}(x_n, t)$
Belief at $f_{x_n}$	$b_{f_{x_n}}(x_n, t + 1)$	$f_{x_n}(x_n) \Delta_n^{f_{x_n} \leftarrow x}(x_n, t)$

TABLE I  
EXPLICIT FORMS OF MESSAGES AND BELIEFS

The factor graph corresponding to the joint PDF (11) is illustrated in Fig. 1.

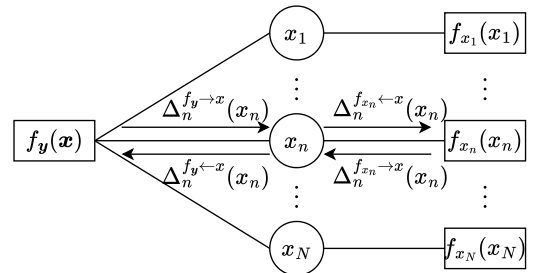


Fig. 1. Sparse Signal Recovery

To make the discussion concise, we introduce the following vector and matrix notations:

$$\boldsymbol{\mu}_{\mathbf{p}}(t) = [\mu_{p_1}(t) \cdots \mu_{p_N}(t)]^T, \quad \boldsymbol{\tau}_{\mathbf{p}}(t) = [\tau_{p_1}(t) \cdots \tau_{p_N}(t)]^T,$$

and define

$$\mathbf{C}_{\mathbf{p}}(t) = \text{Diag}(\boldsymbol{\tau}_{\mathbf{p}}(t)).$$

Similarly, we stack  $\mu_{r_n}(t)$  and  $\tau_{r_n}(t)$  into vectors  $\boldsymbol{\mu}_{\mathbf{r}}(t)$  and  $\boldsymbol{\tau}_{\mathbf{r}}(t)$ , respectively.

Let  $\mathbf{a}_n$  denote the  $n$ -th column of  $\mathbf{A}$ , and let the subscript  $\bar{n}$  indicate that the  $n$ -th column or entry is removed. For example,

$$\mathbf{A}_{\bar{n}} \triangleq [\mathbf{a}_1 \quad \cdots \quad \mathbf{a}_{n-1} \quad \mathbf{a}_{n+1} \quad \cdots \quad \mathbf{a}_N].$$

Furthermore, we define

$$\begin{aligned} \boldsymbol{\mu}_{\mathbf{p}_{\bar{n}}}(t) &\triangleq [\mu_{p_1}(t) \cdots \mu_{p_{n-1}}(t) \mu_{p_{n+1}}(t) \cdots \mu_{p_N}(t)]^T \\ \boldsymbol{\tau}_{\mathbf{p}_{\bar{n}}}(t) &\triangleq [\tau_{p_1}(t) \cdots \tau_{p_{n-1}}(t) \tau_{p_{n+1}}(t) \cdots \tau_{p_N}(t)]^T \\ \mathbf{C}_{\mathbf{p}_{\bar{n}}}(t+1) &\triangleq \text{Diag}[\boldsymbol{\tau}_{\mathbf{p}_{\bar{n}}}(t+1)]. \end{aligned} \quad (14)$$

In addition to the above, we use the shorthand  $\forall n$  to denote  $\forall n \in \{1, \dots, N\}$ .

Throughout this paper, unless explicitly stated otherwise (for example, in the proof of Lemma 1), we distinguish between the index  $n$  and the time-dependent index  $n(t)$ . The index  $n$  refers to an arbitrary entry, whereas  $n(t)$  can be viewed as a selection function denoting the entry selected at time  $t$ . The two indices are independent of each other;  $n$  has no relation to  $n(t)$ , and vice versa. To avoid ambiguity, we explicitly use the quantifiers  $\forall$  and  $\exists$  whenever needed.

Next, we briefly describe the ideal reVAMP procedure, assuming that no negative message variances occur.

#### A. Ideal reVAMP: Message from $f_{\mathbf{y}}$ to $\mathbf{x}$

The belief at the likelihood factor can be written as

$$\begin{aligned} b_{f_{\mathbf{y}}}(\mathbf{x}, t) &\propto f_{\mathbf{y}}(\mathbf{x}) \prod_n \Delta_n^{f_{\mathbf{y}} \leftarrow x}(x_n, t) \\ &\propto \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}_{\hat{\mathbf{x}}|\mathbf{y}}(t), \mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t)), \end{aligned} \quad (15)$$

where

$$\begin{aligned} \mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t) &= (\mathbf{A}^T \mathbf{C}_{\mathbf{v}}^{-1} \mathbf{A} + \mathbf{C}_{\mathbf{p}}(t))^{-1} \\ \boldsymbol{\mu}_{\hat{\mathbf{x}}|\mathbf{y}}(t) &= \mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t) (\mathbf{A}^T \mathbf{C}_{\mathbf{v}}^{-1} \mathbf{y} + \mathbf{C}_{\mathbf{p}}(t)^{-1} \boldsymbol{\mu}_{\mathbf{p}}(t)). \end{aligned} \quad (16)$$

Since (15) is already Gaussian, its marginals are also Gaussian, and the KLD projection has no effect. Denote the  $n$ -th diagonal entry of  $\mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t)$  by  $\tau_{\hat{x}_n|\mathbf{y}}(t) = e_n^T \mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t) e_n$ , and the  $n$ -th entry of  $\boldsymbol{\mu}_{\hat{\mathbf{x}}|\mathbf{y}}(t)$  by  $\mu_{\hat{x}_n|\mathbf{y}}(t) = e_n^T \boldsymbol{\mu}_{\hat{\mathbf{x}}|\mathbf{y}}(t)$ . Then the message from  $f_{\mathbf{y}}$  to  $x_n$  is

$$\begin{aligned} \forall n: \Delta_n^{f_{\mathbf{y}} \rightarrow x}(x_n, t) &\propto \frac{\text{proj}(\int b_{f_{\mathbf{y}}}(\mathbf{x}, t) \prod_{n' \neq n} dx_{n'})}{\Delta_n^{f_{\mathbf{y}} \leftarrow x}(x_n, t)} \\ &\propto \mathcal{UN}(x_n | \mu_{r_n}(t), \tau_{r_n}(t)), \end{aligned} \quad (17)$$

where  $\forall n$ :

$$\begin{aligned} \tau_{r_n}(t) &= \left( \frac{1}{\tau_{\hat{x}_n|\mathbf{y}}(t)} - \frac{1}{\tau_{p_n}(t)} \right)^{-1} \\ \mu_{r_n}(t) &= \tau_{r_n}(t) \left( \frac{\mu_{\hat{x}_n|\mathbf{y}}(t)}{\tau_{\hat{x}_n|\mathbf{y}}(t)} - \frac{\mu_{p_n}(t)}{\tau_{p_n}(t)} \right). \end{aligned} \quad (18)$$

As we can see, the updates of  $\tau_{r_n}(t)$  and  $\mu_{r_n}(t)$  for all  $n$  depend only on the values of  $\boldsymbol{\mu}_{\mathbf{p}}(t)$  and  $\boldsymbol{\tau}_{\mathbf{p}}(t)$ . Using the matrix inversion lemma, one can verify that (18) is equivalent to the following expressions (i.e., both (18) and (18') yield the same results):

$$\begin{aligned} \forall n: \tau_{r_n}(t) &= F_{\tau_{r_n}}[\boldsymbol{\mu}_{\mathbf{p}_{\bar{n}}}(t), \boldsymbol{\tau}_{\mathbf{p}_{\bar{n}}}(t)] \\ \forall n: \mu_{r_n}(t) &= F_{\mu_{r_n}}[\boldsymbol{\mu}_{\mathbf{p}_{\bar{n}}}(t), \boldsymbol{\tau}_{\mathbf{p}_{\bar{n}}}(t)], \end{aligned} \quad (18')$$

where  $\forall n$ :

$$\begin{aligned} F_{\tau_{r_n}}[\boldsymbol{\mu}_{\mathbf{p}_{\bar{n}}}(t), \boldsymbol{\tau}_{\mathbf{p}_{\bar{n}}}(t)] &\triangleq [\mathbf{a}_n^T (\mathbf{C}_{\mathbf{v}} + \mathbf{A}_{\bar{n}} \mathbf{C}_{\mathbf{p}_{\bar{n}}}(t) \mathbf{A}_{\bar{n}})^{-1} \mathbf{a}_n]^{-1} \\ F_{\mu_{r_n}}[\boldsymbol{\mu}_{\mathbf{p}_{\bar{n}}}(t), \boldsymbol{\tau}_{\mathbf{p}_{\bar{n}}}(t)] &\triangleq [\mathbf{a}_n^T (\mathbf{C}_{\mathbf{v}} + \mathbf{A}_{\bar{n}} \mathbf{C}_{\mathbf{p}_{\bar{n}}}(t) \mathbf{A}_{\bar{n}})^{-1} \mathbf{a}_n]^{-1} \\ &\quad \cdot \mathbf{a}_n^T (\mathbf{C}_{\mathbf{v}} + \mathbf{A}_{\bar{n}} \mathbf{C}_{\mathbf{p}_{\bar{n}}}(t) \mathbf{A}_{\bar{n}})^{-1} (\mathbf{y} - \mathbf{A}_{\bar{n}} \mathbf{p}_{\bar{n}}(t)). \end{aligned} \quad (19)$$

#### B. Ideal reVAMP: Message from $f_{x_{n(t)}}$ to $x_{n(t)}$

Since we employ sequential updates, suppose that at computation time  $t$ , we are updating the pair  $(\mu_{p_{n(t)}}(t+1), \tau_{p_{n(t)}}(t+1))$ , where  $n(t)$  is a selection function that chooses an index in  $\{1, \dots, N\}$  based on time  $t$ .

According to the EP rules, the belief at the prior factor  $f_{x_{n(t)}}$  is given by

$$b_{f_{x_{n(t)}}}(x_{n(t)}, t+1) \propto f_{x_{n(t)}}(x_{n(t)}) \Delta_{n(t)}^{f_x \leftarrow x}(x_{n(t)}, t). \quad (20)$$

To obtain the mean and variance of  $b_{f_{x_{n(t)}}}$ , a non-analytic (numerical) evaluation may be required, since  $f_{x_{n(t)}}(x_{n(t)}) = p(x_{n(t)})$  defined in (12) can be an arbitrary prior. Nevertheless, this evaluation is only one-dimensional. We define the mean and variance of belief in (20) as:

$$\begin{aligned} \mu_{\hat{x}_{n(t)}|\boldsymbol{\mu}_{\mathbf{r}}}(t+1) &\triangleq \mathbb{E}_{b_{f_{x_{n(t)}}}(t+1)}[x_{n(t)}] \\ \tau_{\hat{x}_{n(t)}|\boldsymbol{\mu}_{\mathbf{r}}}(t+1) &\triangleq \text{var}_{b_{f_{x_{n(t)}}}(t+1)}[x_{n(t)}]. \end{aligned} \quad (21)$$

Again following the EP rules, the outgoing message from the prior factor is

$$\begin{aligned} \Delta_{n(t)}^{f_x \rightarrow x}(x_{n(t)}, t+1) &\propto \frac{\text{proj}[b_{f_{x_{n(t)}}}(x_{n(t)}, t+1)]}{\Delta_{n(t)}^{f_x \leftarrow x}(x_{n(t)}, t)} \\ &\propto \mathcal{UN}(x_{n(t)} | \mu_{p_{n(t)}}(t+1), \tau_{p_{n(t)}}(t+1)), \end{aligned} \quad (22)$$

where

$$\begin{aligned} \tau_{p_{n(t)}}(t+1) &= \left( \frac{1}{\tau_{\hat{x}_{n(t)}|\boldsymbol{\mu}_{\mathbf{r}}}(t+1)} - \frac{1}{\tau_{r_n}(t)} \right)^{-1}; \\ \mu_{p_{n(t)}}(t+1) &= \tau_{p_{n(t)}}(t+1) \left( \frac{\mu_{\hat{x}_{n(t)}|\boldsymbol{\mu}_{\mathbf{r}}}(t+1)}{\tau_{\hat{x}_{n(t)}|\boldsymbol{\mu}_{\mathbf{r}}}(t+1)} - \frac{\mu_{r_n}(t)}{\tau_{r_n}(t)} \right). \end{aligned} \quad (23)$$

For all other indices  $\forall n' \neq n(t)$ , the parameters remain unchanged:

$$\forall n' \neq n(t): \mu_{p_{n'}}(t+1) = \mu_{p_{n'}}(t), \tau_{p_{n'}}(t+1) = \tau_{p_{n'}}(t) \quad (24)$$

This closes the reVAMP update loop. By iteratively updating the pairs  $(\boldsymbol{\mu}_{\mathbf{p}}, \boldsymbol{\tau}_{\mathbf{p}})$  and  $(\boldsymbol{\mu}_{\mathbf{r}}, \boldsymbol{\tau}_{\mathbf{r}})$  until convergence, we obtain an approximation of the posterior mean:  $\mathbb{E}[\mathbf{x}|\mathbf{y}]$  as  $\mathbb{E}[\mathbf{x}|\mathbf{y}] \simeq \boldsymbol{\mu}_{\hat{\mathbf{x}}|\mathbf{y}}(\infty) \simeq \boldsymbol{\mu}_{\hat{\mathbf{x}}|\boldsymbol{\mu}_{\mathbf{r}}}(\infty)$ .

### C. Low-Rank Update

It is worth noting that the update of  $\mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}$  requires only  $O(N^2)$  multiplications if only one diagonal entry of  $\mathbf{C}_{\mathbf{p}}$  is updated. Since the pair  $(\mu_{p_n(t)}, \tau_{p_n(t)})$  is updated at time  $t$ , we define

$$\frac{1}{\Delta\tau_{p_n(t)}(t+1)} \triangleq \frac{1}{\tau_{p_n(t)}(t+1)} - \frac{1}{\tau_{p_n(t)}(t)}. \quad (25)$$

Then the rank-one update for  $\mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}$  is

$$\begin{aligned} \mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t+1) &= \mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t) - [\mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t)]_{:,n(t)} \\ &\cdot (\Delta\tau_{p_n(t)}(t+1) + \tau_{\hat{x}_{n(t)}|\mathbf{y}}(t))^{-1} [\mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t)]_{:,n(t)}^T. \end{aligned} \quad (26)$$

Note that in Section V-A, it may appear that we compute the parameters  $(\mu_{r_n(t)}, \tau_{r_n(t)})$  for all  $n$  in a parallel fashion. However, this is effectively equivalent to computing only  $(\mu_{r_{n(t)}(t)}, \tau_{r_{n(t)}(t)})$ , because the parameters corresponding to the other indices are not used in Section V-B. Nevertheless, expressing the update of  $(\mu_{r_n(t)}, \tau_{r_n(t)})$  in parallel form simplifies the discussions in Section VI and VII. Moreover, the low-rank update of the belief covariance still requires  $O(N^2)$  operations, whereas computing all  $\forall n : (\mu_{r_n(t)}, \tau_{r_n(t)})$  incurs only  $O(N)$  complexity.

Although EP permits messages that are not proper distributions—such as Gaussian-like messages with negative variances—these messages may induce beliefs that are no longer proper distributions. This is not allowed in EP and may prevent the algorithm from progressing. For example, if  $\tau_{r_{n(t)}(t)}$  is negative, the belief in (20) may fail to be a proper distribution, and the expectation operations in (21) may become ill-defined.

## VI. KEEPING NEGATIVE VARIANCE

In EP, the messages in Table I are not required to be proper distributions; in particular, their variances may take negative values. In contrast, the beliefs  $b_{f_{\mathbf{y}}}$  and  $\forall n : b_{f_{x_n}}$  must be proper distributions. From the previous discussion, evaluating  $b_{f_{\mathbf{y}}}$  in (15) involves only Gaussian PDF manipulations, whereas obtaining the statistics of  $\forall n : b_{f_{x_n}}$  in (20) requires integrating with respect to an arbitrary prior. Therefore, we propose to allow negative values of  $\tau_{p_n}$  and apply special treatment only when a given  $\tau_{p_n}$  would lead to a negative value of some  $\tau_{r_n}$ .

**Lemma 1.** *If, in reVAMP, sequential updates are used and  $\tau_{p_n(t)}(t+1)$  is updated via (23), then*

$$\frac{\det[\mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t)]}{\det[\mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t+1)]} = \frac{\tau_{\hat{x}_{n(t)}|\mathbf{y}}(t)}{\tau_{\hat{x}_{n(t)}|\mu_{\mathbf{r}}}(t+1)}. \quad (27)$$

*Proof.* Recall from Section V that  $n(t)$  denotes the index of the entry in  $\mu_{\mathbf{p}}$  and  $\tau_{\mathbf{p}}$  that is updated at computation time  $t$ . For brevity, in this proof (from (28) to (34)), we use the shorthand  $n = n(t)$ . Define

$$\begin{aligned} \mathbf{C}_{\hat{\mathbf{x}}_{\bar{n}}|\mathbf{y}}(t+1)^{-1} &\triangleq \mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t+1)^{-1} - \mathbf{e}_n \tau_{p_n}(t+1)^{-1} \mathbf{e}_n^T \\ &= \mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t)^{-1} - \mathbf{e}_n \tau_{p_n}(t)^{-1} \mathbf{e}_n^T. \end{aligned} \quad (28)$$

Using (7) and (28), we obtain

$$\begin{aligned} \det(\mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t+1)^{-1}) &= \det(\mathbf{C}_{\hat{\mathbf{x}}_{\bar{n}}|\mathbf{y}}(t+1)^{-1}) \\ &\cdot \frac{\tau_{p_n}(t+1) + \mathbf{e}_n^T \mathbf{C}_{\hat{\mathbf{x}}_{\bar{n}}|\mathbf{y}}(t+1) \mathbf{e}_n}{\tau_{p_n}(t+1)}. \end{aligned} \quad (29)$$

Since, under sequential updates at computation time  $t$ , only  $\tau_{p_n}(t+1)$  is updated while all other entries remain unchanged, we have

$$\begin{aligned} \text{Sequential Update} &\Rightarrow \forall n' \neq n(t) : \tau_{p_{n'}}(t+1) = \tau_{p_{n'}}(t+1) \\ &\Rightarrow \tau_{p_{\bar{n}}}(t+1) = \tau_{p_{\bar{n}}}(t) \\ &\Rightarrow \mathbf{C}_{p_{\bar{n}}}(t+1) = \mathbf{C}_{p_{\bar{n}}}(t). \end{aligned} \quad (30)$$

From (30), and using the matrix inversion lemma, we obtain

$$\mathbf{e}_n^T \mathbf{C}_{\hat{\mathbf{x}}_{\bar{n}}|\mathbf{y}}(t+1) \mathbf{e}_n = \tau_{r_n}(t). \quad (31)$$

Substituting the update expression for  $\tau_{p_n}(t+1)$  in (23) together with (31) into (29) yields

$$\det(\mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t+1)^{-1}) = \det(\mathbf{C}_{\hat{\mathbf{x}}_{\bar{n}}|\mathbf{y}}(t+1)^{-1}) \frac{\tau_{r_n}(t)}{\tau_{\hat{x}_n|\mu_{\mathbf{r}}}(t+1)}. \quad (32)$$

Next, from the second line of (28) and the matrix determinant lemma (7), we have

$$\begin{aligned} \det(\mathbf{C}_{\hat{\mathbf{x}}_{\bar{n}}|\mathbf{y}}(t+1)^{-1}) &= \det(\mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t)^{-1}) \frac{\tau_{p_n}(t) - \tau_{\hat{x}_n|\mathbf{y}}(t)}{\tau_{p_n}(t)} \\ &= \det(\mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t)^{-1}) \tau_{\hat{x}_n|\mathbf{y}}(t) \left( \frac{1}{\tau_{\hat{x}_n|\mathbf{y}}(t)} - \frac{1}{\tau_{p_n}(t)} \right) \end{aligned} \quad (33)$$

Substituting the right-hand side of the first line of (18) into (33), and then inserting the resulting expression into (32), we obtain

$$\begin{aligned} \det(\mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t+1)^{-1}) &= \det(\mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t)^{-1}) \frac{\tau_{\hat{x}_n|\mathbf{y}}(t)}{\tau_{\hat{x}_n|\mu_{\mathbf{r}}}(t+1)} \\ &\Leftrightarrow \frac{\det[\mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t)]}{\det[\mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t+1)]} = \frac{\tau_{\hat{x}_n(t)|\mathbf{y}}(t)}{\tau_{\hat{x}_n(t)|\mu_{\mathbf{r}}}(t+1)}. \end{aligned} \quad (34)$$

□

**Theorem 1.** *In reVAMP as described in Section V, if  $\forall t : b_{f_{x_n(t)}}(x_{n(t)}, t+1)$  in (20) are proper distributions (normalize to one) with respect to  $x_{n(t)}$ , and if  $\tau_{\mathbf{p}}(1)$  is initialized with positive entries, then the belief covariance matrices  $\forall t : \mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t)$  in (16) remain positive definite.*

*Proof.* We use mathematical induction together with Sylvester's criterion.

Since we initialize  $\forall n : \tau_{p_n}(1)$  with positive numbers, the matrix  $\mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(1)$  at the beginning of the first iteration is positive definite.

We can always reorder the entries of  $\mathbf{x}$  and the corresponding columns of  $\mathbf{A}$  such that the updated  $\tau_{p_n}(t+1)$  appears in the last diagonal entry of  $\mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t+1)^{-1}$ . Thus, without loss of generality, assume that  $N = n(t)$ , and that  $\mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t)$  is positive definite (induction hypothesis). We will show that  $\mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t+1)$  is also positive definite.

By Sylvester's criterion, a symmetric matrix is positive definite if and only if all its leading principal minors (the determinants of its upper-left  $k \times k$  submatrices for all  $k$ ) are positive. The only difference between  $\mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t)^{-1}$  and  $\mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t+1)^{-1}$  is the

last diagonal entry. Hence, the first through  $(N-1)$ -th leading principal minors of  $\mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t+1)^{-1}$  and  $\mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t)^{-1}$  coincide and are therefore positive. It remains to show that the  $N$ -th order leading principal minor, i.e.,  $\det(\mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t+1)^{-1})$ , is also positive.

From Lemma 1, with the assumption that  $N = n(t)$ , we have

$$\det(\mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t+1)^{-1}) = \det(\mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t)^{-1}) \frac{\tau_{\hat{\mathbf{x}}_N|\mathbf{y}}(t)}{\tau_{\hat{\mathbf{x}}_N|\boldsymbol{\mu}_r}(t+1)}. \quad (35)$$

By the induction hypothesis,  $\det(\mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t)^{-1}) > 0$  and  $\tau_{\hat{\mathbf{x}}_N|\mathbf{y}}(t) > 0$ . Moreover, from (21) and the assumption that  $b_{f_{x_n(t)}}(x_n(t), t+1)$  is a proper distribution, we have  $\tau_{\hat{\mathbf{x}}_N|\boldsymbol{\mu}_r}(t+1) > 0$ . Hence,  $\det(\mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t+1)^{-1}) > 0$ . Since all leading principal minors of  $\mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t+1)^{-1}$  are positive, the matrix is positive definite.  $\square$

**Theorem 2.** *If the initial values  $\forall n : \tau_{p_n}(1)$  are all positive, then the covariance matrix  $\forall t : \mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t)$  remains positive definite when using any mixture of the following two update strategies for  $\forall t : \tau_{p_n(t)}(t+1)$ :*

- $\tau_{p_n(t)}(t+1)$  is updated via (23)
- $\tau_{p_n(t)}(t+1) := \tau_{p_n(t)}(t)$

*Proof.* We again use mathematical induction. As in Theorem 1,  $\mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(1)$  is positive definite by assumption. We need to show that  $\mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t+1)$  is positive definite given that  $\mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t)$  is positive definite.

If  $\tau_{p_n(t)}(t+1)$  is updated via (23), we can follow the same proof as in Theorem 1 using Lemma 1 to conclude that  $\mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t+1)$  is positive definite.

If instead  $\tau_{p_n(t)}(t+1)$  is set to  $\tau_{p_n(t)}(t+1) := \tau_{p_n(t)}(t)$ , then  $\mathbf{C}_{\mathbf{p}}(t+1) = \mathbf{C}_{\mathbf{p}}(t)$ , since for entries with indices  $\forall n' \neq n(t)$ , the update rule (24) implies that  $\tau_{p_{n'}(t+1)} = \tau_{p_{n'}(t)}$ . Consequently,  $\mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t+1) = \mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t)$  which is positive definite by the induction hypothesis.  $\square$

*Remark.* Theorem 1 implies that, at any stationary point, the belief  $b_{f_{\mathbf{y}}}(\mathbf{x}, \infty)$  automatically becomes a proper distribution if  $\forall n : b_{f_{x_n}}(x_n, \infty)$  are proper distributions. During the iterations, however,

#### A. Non-Persistent Approach

According to Theorem 1, we must ensure that, at every time  $t$ , the selection function  $n(t)$  chooses an index for which the corresponding  $b_{f_{x_n(t)}}(x_n(t), t+1)$ , obtained from (20) is a proper distribution. Furthermore, since EP can be interpreted as an iterative method for finding stationary points of the constrained BFE, we know that, at convergence, all beliefs must be proper distributions.

Motivated by these two considerations, we propose to check the values  $\forall n, \tau_{r_n}(t+1)$  when using (23) to compute  $\tau_{p_n(t)}(t+1)$  from  $\tau_{p_n(t)}(t)$ , before actually applying the update to  $\tau_{p_n(t)}(t)$ .

1) : During each iteration  $t$  of this non-persistent EP scheme, we first follow (15)-(21).

2) : Next, we compute the unchecked updates in the same manner as in (23), but denote them with a dot to distinguish them from the checked values:

$$\begin{aligned} \dot{\tau}_{p_n(t)}(t+1) &= \left( \frac{1}{\tau_{\hat{\mathbf{x}}_n(t)|\boldsymbol{\mu}_r}(t+1)} - \frac{1}{\tau_{r_n(t)}(t)} \right)^{-1}; \\ \dot{\mu}_{p_n(t)}(t+1) &= \dot{\tau}_{p_n(t)}(t+1) \left( \frac{\mu_{\hat{\mathbf{x}}_n(t)|\boldsymbol{\mu}_r}(t+1)}{\tau_{\hat{\mathbf{x}}_n(t)|\boldsymbol{\mu}_r}(t+1)} - \frac{\mu_{r_n(t)}(t)}{\tau_{r_n(t)}(t)} \right). \end{aligned} \quad (36)$$

Meanwhile  $\forall n' \neq n(t)$ , we set

$$\dot{\mu}_{p_{n'}}(t+1) = \mu_{p_{n'}}(t), \quad \dot{\tau}_{p_{n'}}(t+1) = \tau_{p_{n'}}(t) \quad (37)$$

As a result,

$$\begin{aligned} \dot{\boldsymbol{\tau}}_{\mathbf{p}}(t+1) &= \boldsymbol{\tau}_{\mathbf{p}}(t) + \mathbf{e}_{n(t)}[\dot{\tau}_{p_n(t)}(t+1) - \tau_{p_n(t)}(t)] \\ \dot{\mathbf{C}}_{\mathbf{p}}(t+1) &= \mathbf{C}_{\mathbf{p}}(t) + \mathbf{e}_{n(t)}[\dot{\tau}_{p_n(t)}(t+1) - \tau_{p_n(t)}(t)]\mathbf{e}_{n(t)}^T \\ \dot{\boldsymbol{\mu}}_{\mathbf{p}}(t+1) &= \boldsymbol{\mu}_{\mathbf{p}}(t) + \mathbf{e}_{n(t)}[\dot{\mu}_{p_n(t)}(t+1) - \mu_{p_n(t)}(t)]. \end{aligned}$$

3) : We then check whether this unchecked update would lead to beliefs that are not proper distributions. To this end, we first compute the extrinsic messages of the prior factors. Based on (18) and (18'), we obtain  $\forall n$ :

$$\begin{aligned} \dot{\tau}_{r_n}(t+1) &= F_{\tau_{r_n}}[\dot{\boldsymbol{\mu}}_{\mathbf{p}_{\bar{n}}}(t+1), \dot{\boldsymbol{\tau}}_{\mathbf{p}_{\bar{n}}}(t+1)] \\ \dot{\mu}_{r_n}(t+1) &= F_{\mu_{r_n}}[\dot{\boldsymbol{\mu}}_{\mathbf{p}_{\bar{n}}}(t+1), \dot{\boldsymbol{\tau}}_{\mathbf{p}_{\bar{n}}}(t+1)], \end{aligned} \quad (38)$$

where, for all  $n$  and all  $n' \neq n$ , the vectors  $\dot{\boldsymbol{\mu}}_{\mathbf{p}_{\bar{n}}}(t+1)$  and  $\dot{\boldsymbol{\tau}}_{\mathbf{p}_{\bar{n}}}(t+1)$  stack the elements  $\mu_{p_{n'}}(t+1)$  and  $\tau_{p_{n'}}(t+1)$ , respectively, following the same convention as in (14).

Then, according to (20), the beliefs resulting from the unchecked update  $\dot{\mu}_{p_n(t)}(t+1)$  and  $\dot{\tau}_{p_n(t)}(t+1)$  are

$$\forall n : \dot{b}_{f_{x_n}}(x_n, t+2) \propto f_{x_n}(x_n) \mathcal{UN}(x_n | \dot{\mu}_{r_n}(t+1), \dot{\tau}_{r_n}(t+1)). \quad (39)$$

4) : Finally, the actual (checked) updates are defined as

$$\begin{aligned} \tau_{p_n(t)}(t+1) &= \begin{cases} \dot{\tau}_{p_n(t)}(t+1) & \forall n : \int \dot{b}_{f_{x_n}}(x_n, t+2) dx_n < \infty \\ \tau_{p_n(t)}(t) & \text{otherwise} \end{cases} \\ \mu_{p_n(t)}(t+1) &= \begin{cases} \dot{\mu}_{p_n(t)}(t+1) & \forall n : \int \dot{b}_{f_{x_n}}(x_n, t+2) dx_n < \infty \\ \mu_{p_n(t)}(t) & \text{otherwise,} \end{cases} \end{aligned} \quad (40)$$

where we use the notation  $\int df(\theta) < \infty$  to denote that the function  $f(\theta)$  has finite integral value. If evaluating the integrals  $\forall n : \int \dot{b}_{f_{x_n}}(x_n, t+2) dx_n$  in (40) is too costly, we may instead check the signs of  $\forall n : \dot{\tau}_{r_n}(t+1)$  in (38). This leads to a relaxed non-persistent approach:

- If  $\forall n : \dot{\tau}_{r_n}(t+1) > 0$ , then  $\tau_{p_n(t)}(t+1)$  and  $\mu_{p_n(t)}(t+1)$  are updated to  $\dot{\tau}_{p_n(t)}(t+1)$  and  $\dot{\mu}_{p_n(t)}(t+1)$ .
- If  $\exists n : \tau_{r_n}(t+1) \leq 0$ , then  $\tau_{p_n(t)}(t+1)$  and  $\mu_{p_n(t)}(t+1)$  remain unchanged as  $\tau_{p_n(t)}(t)$  and  $\mu_{p_n(t)}(t)$ .

We refer to the scheme based on (40) as the strict non-persistent approach, and to the sign-based rule above as the relaxed non-persistent approach.

5) : Afterward, we proceed to iteration  $t + 1$  and repeat from Section VI-A1.

### B. Persistent Approach

In this approach, we do not perform any look-ahead. At computation time  $t$ , we proceed from (15) to (19) following the standard EP/reVAMP steps to obtain  $\forall n : \mu_{r_n}(t)$  and  $\tau_{r_n}(t)$ . After that, before updating  $\tau_{p_{n(t)}}(t + 1)$  and  $\mu_{p_{n(t)}}(t + 1)$ , we check whether  $b_{f_{x_{n(t)}}}(x_{n(t)}(t), t + 1)$ , obtained from (20), is a proper distribution (integrates to finite value).

- If the belief  $b_{f_{x_{n(t)}}}(x_{n(t)}(t), t + 1)$  is a proper distribution, then  $\tau_{p_{n(t)}}(t + 1)$  and  $\mu_{p_{n(t)}}(t + 1)$  are updated by following (20)-(24).
- Otherwise if the belief  $b_{f_{x_{n(t)}}}(x_{n(t)}(t), t + 1)$  is not a proper distribution, then  $\tau_{p_{n(t)}}(t + 1)$  and  $\mu_{p_{n(t)}}(t + 1)$  remain unchanged, i.e.,

$$\forall n \in \{1, \dots, N\} : \mu_{p_n}(t + 1) = \mu_{p_n}(t), \tau_{p_n}(t + 1) = \tau_{p_n}(t). \quad (41)$$

If checking whether  $b_{f_{x_{n(t)}}}(x_{n(t)}(t), t + 1)$  is a proper distribution via (20) is computationally demanding, we can relax the condition and instead check the sign of  $\tau_{r_{n(t)}}$ . If  $\tau_{r_{n(t)}}$  is positive, we proceed with the standard update in (20)-(24); otherwise, we apply the no-update rule in (41).

### VII. AVOIDING NEGATIVE MESSAGE VARIANCES

By matrix inversion lemma (analog to (31)), one can show that the messages  $\mathcal{UN}(x|\mu_r(t), \text{Diag}[\tau_r(t)])$  are proper distributions if  $\forall n, t : \tau_{p_n}(t)$  are positive. Therefore, in this section we focus on the message from  $f_{x_n}$  to  $x_n$  and on enforcing  $\forall n, t : \tau_{p_n}(t) > 0$ . W.l.o.g., in the following of this section, we assume the Ansatz that  $\forall n, t : \tau_{r_n}(t) > 0$ .

To gain further insight into (20), we abstract a simplified one-dimensional measurement model whose posterior PDF shares the same factorization structure as that in (20):

$$\mu_r = x + v_r, \quad (42)$$

where  $v_r \sim \mathcal{N}(0, \tau_r)$ , and  $x \sim p(x)$  is an arbitrary distribution. We denote the posterior PDF, mean and variance of (42) as  $b_{f_x}(x) = p(x|\mu_r)$ ,  $\mu_{\hat{x}|\mu_r}$  and  $\tau_{\hat{x}|\mu_r}$ . According to the EP rule, negative message variances occur when the posterior variance  $\tau_{\hat{x}|\mu_r}$  is larger than  $\tau_r$ .

#### A. A Simple Example of Negative Message Variances

Suppose  $p(x)$  is BPSK. Negative message variances arise when we observe an incorrect measurement with high precision. A simple example is  $\mu_r = 0$  and  $\tau_r$  arbitrarily small, because in the BPSK case with  $\mu_r = 0$ , the posterior variance  $\tau_{\hat{x}|\mu_r}$  is independent of  $\tau_r$ .

#### B. Reconsidering the KLD Projection

Based on EP rule and (22), the message parameterized by the pair  $(\mu_p, \tau_p)$  is updated by

$$(\mu_{\tilde{x}|\mu_r}, \tau_{\tilde{x}|\mu_r}) = \underset{\mu_{\tilde{x}|\mu_r}, \tau_{\tilde{x}|\mu_r}}{\text{argmin}} \text{KLD}[b_{f_x}(x) \|\mathcal{N}(x|\mu_{\tilde{x}|\mu_r}, \tau_{\tilde{x}|\mu_r})];$$

$$\mathcal{UN}(x|\mu_p, \tau_p) \propto \frac{\mathcal{N}(x|\mu_{\tilde{x}|\mu_r}, \tau_{\tilde{x}|\mu_r})}{\mathcal{N}(x|\mu_r, \tau_r)}.$$

(43)

Consider the second line of (43) and impose the constraint  $\tau_p > 0$ . Then, we have

$$\mathcal{N}(x|\mu_{\tilde{x}|\mu_r}(p, \tau_p), \tau_{\tilde{x}|\mu_r}(p, \tau_p)) \propto \mathcal{N}(x|\mu_r, \tau_r) \mathcal{N}(x|\mu_p, \tau_p),$$

where we explicitly emphasize the dependence of  $(\mu_{\tilde{x}|\mu_r}, \tau_{\tilde{x}|\mu_r})$  on the unknown parameters  $(\mu_p, \tau_p)$ :

$$\begin{aligned} \tau_{\tilde{x}|\mu_r}(\mu_p, \tau_p) &= \frac{\tau_p \tau_r}{\tau_p + \tau_r}; \\ \mu_{\tilde{x}|\mu_r}(\mu_p, \tau_p) &= \frac{\tau_p \mu_r + \tau_r \mu_p}{\tau_p + \tau_r}. \end{aligned} \quad (44)$$

Substituting (44) into the first line of (43) and imposing the constraint  $\tau_p > 0$  yields the following constrained optimization problem:

$$\begin{aligned} \min_{\mu_p, \tau_p} \quad & \text{KLD}[b_{f_x}(x) \|\mathcal{N}(x|\mu_{\tilde{x}|\mu_r}(\mu_p, \tau_p), \tau_{\tilde{x}|\mu_r}(\mu_p, \tau_p))], \\ \text{s.t.} \quad & \tau_p > 0. \end{aligned} \quad (45)$$

Note that (45) may yield a different solution from (43) due to the additional constraint on  $\tau_p$ .

Expanding the KLD, (45) can be rewritten as

$$\begin{aligned} \min_{\mu_p, \tau_p} \quad & L(\mu_p, \tau_p) \\ \text{s.t.} \quad & \tau_p > 0, \end{aligned} \quad (46)$$

where

$$L(\mu_p, \tau_p) \triangleq \frac{\ln \tau_{\tilde{x}|\mu_r}(\mu_p, \tau_p)}{2} + \frac{\tau_{\hat{x}|\mu_r} + [\mu_{\hat{x}|\mu_r} - \mu_{\tilde{x}|\mu_r}(\mu_p, \tau_p)]^2}{2\tau_{\tilde{x}|\mu_r}(\mu_p, \tau_p)}. \quad (47)$$

Taking the partial derivative of the objective function with respect to  $\mu_p$  and setting it to zero yields

$$\frac{\partial L(\mu_p, \tau_p)}{\partial \mu_p} = 0 \Rightarrow \mu_{\tilde{x}|\mu_r}(\mu_p, \tau_p) = \mu_{\hat{x}|\mu_r}. \quad (48)$$

Next, compute the partial derivative with respect to  $\tau_p$  and substitute (48) into the result:

$$\left. \frac{\partial L(\mu_p, \tau_p)}{\partial \tau_p} \right|_{\mu_{\tilde{x}|\mu_r}(\mu_p, \tau_p) = \mu_{\hat{x}|\mu_r}} = \frac{\xi_p^2 (\xi_{\hat{p}} - \xi_p)}{\frac{2}{\tau_{\hat{x}|\mu_r}} (\xi_r + \xi_p)}, \quad (49)$$

where we define  $\xi_p \triangleq \frac{1}{\tau_p}$ ,  $\xi_r \triangleq \frac{1}{\tau_r}$ , and  $\xi_{\hat{p}} \triangleq \frac{1}{\tau_{\hat{x}|\mu_r}} - \frac{1}{\tau_r}$ . From these definitions, we have  $\xi_{\hat{p}} > \xi_r$  and  $\xi_r > 0$ . Thus, we only need to consider two cases:  $\xi_{\hat{p}} > 0$  and  $\xi_{\hat{p}} \leq 0$ .

1) If  $\xi_{\hat{p}} > 0$ : In this case, the optimum is attained at  $\xi_p = \xi_{\hat{p}}$ . Therefore, the update rule in (23) remains unchanged.

2) If  $\xi_{\hat{p}} \leq 0$ : In this case, (49) is monotonically increasing in  $\xi_p > 0$ , and therefore the minimum is achieved as  $\xi_p \rightarrow 0$ , which corresponds to  $\tau_p \rightarrow \infty$ . Instead of computing the corresponding mean, we inspect the Gaussian natural parameters and define  $\nu_p = \frac{\mu_p}{\tau_p} = \xi_p \mu_p$ . Substituting  $\tau_p \rightarrow \infty$  into (48), we obtain  $\nu_p \rightarrow \frac{\mu_{\hat{x}|\mu_r} - \mu_r}{\tau_r}$ .

### C. Analytic Continuation of ReVAMP

We now revisit the discussion in Section V. To extend the message domain of reVAMP,  $\forall n$  we propose to use Gaussian natural parameters

$$\xi_{p_n}(t) = \tau_{p_n}(t)^{-1}; \nu_{p_n}(t) = \frac{\mu_{p_n}(t)}{\tau_{p_n}(t)} \quad (50)$$

instead of the  $(\mu_{p_n}(t), \tau_{p_n}(t))$  representation of the messages. Furthermore, we define

$$\begin{aligned} \xi_{\mathbf{p}}(t) &= \text{diag}[\mathbf{C}_{\mathbf{p}}(t)^{-1}] = [\xi_{p_1}(t) \quad \dots \quad \xi_{p_N}(t)]^T \\ \Xi_{\mathbf{p}}(t) &= \mathbf{C}_{\mathbf{p}}(t)^{-1} = \text{Diag}[\xi_{\mathbf{p}}(t)]; \\ \nu_{\mathbf{p}}(t) &= \mathbf{C}_{\mathbf{p}}(t)^{-1} \boldsymbol{\mu}_{\mathbf{p}}(t). \end{aligned} \quad (51)$$

The extended reVAMP algorithm with analytic continuation (ACreVAMP) is summarized as follows.

1) *ACreVAMP Message from  $f_{\mathbf{y}}$  to  $\mathbf{x}$* : To incorporate infinite variances, we rewrite the belief mean and covariance in (15) as

$$\begin{aligned} \mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t) &= (\mathbf{A}^T \mathbf{C}_{\mathbf{v}}^{-1} \mathbf{A} + \Xi_{\mathbf{p}}(t))^{-1} \\ \boldsymbol{\mu}_{\hat{\mathbf{x}}|\mathbf{y}}(t) &= \mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t) (\mathbf{A}^T \mathbf{C}_{\mathbf{v}}^{-1} \mathbf{y} + \nu_{\mathbf{p}}(t)). \end{aligned} \quad (52)$$

Similar to Section V, we denote,  $\forall n$ , the  $n$ -th diagonal entry in  $\mathbf{C}_{\mathbf{x}|\mathbf{y}}$  as  $\tau_{\hat{x}_n|\mathbf{y}}$  and  $\forall n$ , the  $n$ -th entry of  $\boldsymbol{\mu}_{\hat{\mathbf{x}}|\mathbf{y}}$  as  $\mu_{\hat{x}_n|\mathbf{y}}$ . Extending the EP rule to allow infinite  $\tau_{p_n}(t)$ , the message update in (17)-(18) becomes,  $\forall n$ ,

$$\begin{aligned} \tau_{r_n}(t) &= \left( \frac{1}{\tau_{\hat{x}_n|\mathbf{y}}(t)} - \xi_{p_n}(t) \right)^{-1} \\ \mu_{r_n}(t) &= \tau_{r_n}(t) \left( \frac{\mu_{\hat{x}_n|\mathbf{y}}(t)}{\tau_{\hat{x}_n|\mathbf{y}}(t)} - \nu_{p_n}(t) \right). \end{aligned} \quad (53)$$

2) *ACreVAMP Message from  $f_{x_n}$  to  $x_n$* : Unlike the methods in Section VI, AcreVAMP can be carried out in both parallel and sequential forms. In this paper, we focus on the sequential update to maintain compatibility with the previous discussions. As discussed earlier,  $\forall n, t$ :  $\tau_{r_n}(t)$  is positive if  $\forall n, t$ :  $\tau_{p_n}(t)$  is positive. Therefore, the computation of the belief mean  $\mu_{\hat{x}_n(t)|\boldsymbol{\mu}_{\mathbf{r}}}(t+1)$  and variance  $\tau_{\hat{x}_n(t)|\boldsymbol{\mu}_{\mathbf{r}}}(t+1)$  remains the same as in (20)-(21).

However, based on the discussion in Section VII-B, the parameters  $\nu_{p_n(t)}(t+1)$  and  $\xi_{p_n(t)}(t+1)$  are updated as:

$$\begin{aligned} \nu_{p_n(t)}(t+1) &= \begin{cases} \nu_{\hat{p}_n(t)}(t+1) & \xi_{\hat{p}_n(t)}(t+1) > 0 \\ \frac{\mu_{\hat{x}_n(t)|\boldsymbol{\mu}_{\mathbf{r}}}(t+1) - \mu_{r_n(t)}(t)}{\tau_{r_n(t)}(t)} & \text{otherwise} \end{cases} \\ \xi_{p_n(t)}(t+1) &= \begin{cases} \xi_{\hat{p}_n(t)}(t+1) & \xi_{\hat{p}_n(t)}(t+1) > 0 \\ 0 & \text{otherwise,} \end{cases} \end{aligned}$$

where

$$\begin{aligned} \xi_{\hat{p}_n(t)}(t+1) &= \frac{1}{\tau_{\hat{x}_n(t)|\boldsymbol{\mu}_{\mathbf{r}}}(t+1)} - \frac{1}{\tau_{r_n(t)}(t)} \\ \nu_{\hat{p}_n(t)}(t+1) &= \frac{\mu_{\hat{x}_n(t)|\boldsymbol{\mu}_{\mathbf{r}}}(t+1)}{\tau_{\hat{x}_n(t)|\boldsymbol{\mu}_{\mathbf{r}}}(t+1)} - \frac{\mu_{r_n(t)}(t)}{\tau_{r_n(t)}(t)}. \end{aligned} \quad (54)$$

3) *Low Rank Update*: If sequential update is used, we can still reduce the complexity by exploiting matrix inversion lemma. Define  $\Delta\xi_{p_n(t)}(t+1) \triangleq \xi_{p_n(t)}(t+1) - \xi_{p_n(t)}(t)$ . Analogous to (26), the belief covariance matrix is updated as

$$\mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t+1) = \begin{cases} \mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t) + \Delta\mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t+1), & \Delta\xi_{p_n(t)}(t+1) \neq 0 \\ \mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t), & \text{otherwise} \end{cases} \quad (55)$$

where, for the non-trivial cases  $\Delta\xi_{p_n(t)}(t+1) \neq 0$

$$\begin{aligned} \Delta\mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t+1) &= -[\mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t)]_{:,n(t)} \cdot \\ &\cdot (\Delta\xi_{p_n(t)}(t+1)^{-1} + \tau_{\hat{x}_n(t)|\mathbf{y}}(t))^{-1} [\mathbf{C}_{\hat{\mathbf{x}}|\mathbf{y}}(t)]_{:,n(t)}^T \end{aligned} \quad (56)$$

This closes the loop of AcreVAMP.

## VIII. NUMERICAL SIMULATIONS

In our simulations, we consider two scenarios: sparse signal recovery and binary phase-shift keying (BPSK) signal recovery. In both cases, i.i.d. white Gaussian noise is assumed.

### A. Sparse Signal Recovery Setup

To verify the proposed approach in sparse signal recovery, we consider a measurement matrix  $\mathbf{A} \in \mathbb{R}^{M \times N}$  of size  $M \times N = 8 \times 10$ . Each entry of  $\mathbf{A}$  is i.i.d. and drawn from a Gaussian distribution, i.e., for all  $m, n$ ,  $[\mathbf{A}]_{m,n} \sim \mathcal{N}(0, 1/N)$ . The signal vector  $\mathbf{x}$  contains independent but non-identically distributed Gaussian-mixture entries with exponentially decaying amplitudes:

$$\forall n : p(x_n) = 0.5\mathcal{N}(x_n | -1 \times 3.2^{1-n}, 0.1 \times 3.2^{2-2n}) + 0.5\mathcal{N}(x_n | 1 \times 3.2^{1-n}, 0.1 \times 3.2^{2-2n}). \quad (57)$$

### B. BPSK Signal Recovery Setup

In the BPSK signal recovery scenario, we consider systems of size  $M \times N = 20 \times 10$ . Similarly, the entries of the measurement matrix  $\mathbf{A}$  are i.i.d. Gaussian:  $\forall m, n : [\mathbf{A}]_{m,n} \sim \mathcal{N}(0, 1/N)$ . BPSK can be viewed as a Gaussian mixture whose components have (ideally) zero variance. In our simulations, we approximate this by the following i.i.d. signal prior:

$$\forall n : p(x_n) = 0.5\mathcal{N}(x_n | -1, 0.01) + 0.5\mathcal{N}(x_n | 1, 0.01).$$

### C. Simulation Results

In both scenarios, we vary the noise power and evaluate the algorithms at SNR levels  $\forall i = \{1, \dots, 11\} : SNR = 5(i-1)$  dB. For each SNR level, 500 independent problem instances are generated. We denote by  $\mathcal{I}_l$  the set of indices corresponding to all instances simulated at  $SNR = l$ .

We use the normalized mean squared error (NMSE) relative to the MMSE solution as the performance metric. At SNR level  $SNR = l$ , the NMSE is defined as

$$NMSE_l = \frac{\sum_{i \in \mathcal{I}_l} (\hat{x}_i - \hat{\mathbf{x}}_{\text{MMSE},i})^T (\hat{x}_i - \hat{\mathbf{x}}_{\text{MMSE},i})}{\sum_{i \in \mathcal{I}_l} \hat{\mathbf{x}}_{\text{MMSE},i}^T \hat{\mathbf{x}}_{\text{MMSE},i}}. \quad (58)$$

The exact MMSE result is computed by brute-force evaluation with complexity  $\mathcal{O}(2^N N^3)$ .

We compare the proposed methods against LMMSE, the widely used clipping approach [6], and Noise-Unbiasing [7].

The clipping approach is conceptually similar to the discussion in Section VII, in that it aims to avoid negative message variances. However, unlike the proposed ACreVAMP, if  $\tau_{p_n(t)}(t+1)$  obtained from (23) is non-positive, the clipping method resets the message variance to  $\tau_{p_n(t)}(t+1) = \infty$  and the message mean to  $p_n(t)(t+1) = 0$ .

The sparse signal recovery results are shown in Fig. 2, while the BPSK signal recovery results are shown in Fig. 3.

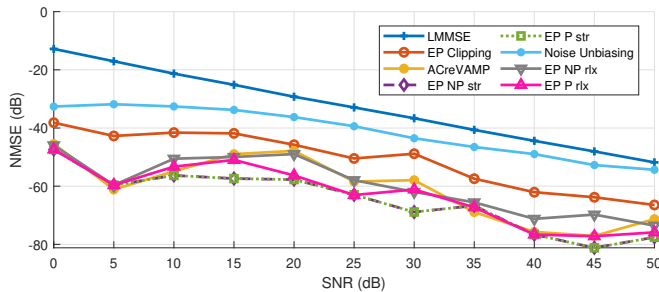


Fig. 2. Sparse Signal Recovery

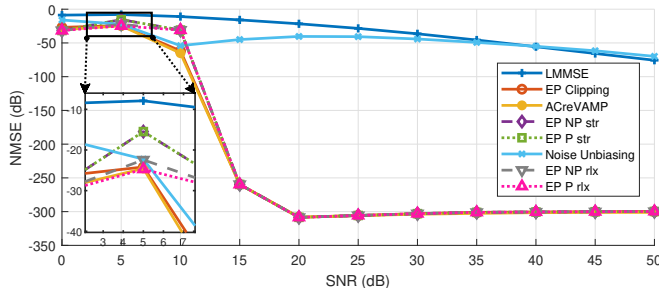


Fig. 3. BPSK Signal Recovery

In the plots, "EP P" and "EP NP" denote persistent and non-persistent approaches proposed in Section VI. The labels "str" and "rlx" indicate whether strict checking or relaxed checking, as discussed in Section VI, is used.

From the two plots, we observe that when strict checking is employed, the persistent and non-persistent methods proposed in Section VI share the same stationary points, as indicated by the overlapping curves. In contrast, under relaxed checking, the persistent and non-persistent approaches converge to different solutions.

As the signal  $x$  transitions from continuous-valued to essentially discrete, the proposed ACreVAMP method in Section VII outperforms both the non-persistent and persistent approaches, particularly at low SNR.

## IX. CONCLUDING REMARKS

In this paper, we reviewed the EP algorithm and reVAMP, an application of EP to linear models. EP and reVAMP approximate marginal distributions—referred to as beliefs—through the exchange of messages. While the beliefs must be proper probability distributions normalized to one, the messages themselves may have infinite integral values and therefore need not be proper distributions.

In our setting, due to the use of Gaussian projection, all messages take a Gaussian form. Consequently, when a message has an infinite integral value, it appears as though it has a

"negative" variance. Such messages can sometimes prevent the EP/reVAMP algorithm from proceeding. Depending on how these "negative-variance" messages are handled, we proposed three approaches. The non-persistent and persistent EP variants allow the existence of negative message variances, whereas ACreVAMP suppresses or prevents messages with negative variances.

Simulation results demonstrate that the proposed approaches outperform standard benchmark methods. For continuous-valued signal recovery, allowing negative-variance messages leads to superior performance, while for discrete signal recovery, preventing or suppressing negative variances proves more advantageous.

## X. ACKNOWLEDGEMENTS

EURECOM's research is partially supported by its industrial members: ORANGE, BMW, SAP, iABG, Norton LifeLock, by the Franco-German projects CellFree6G and 5G-OPERA, the French PEPR-5G projects PERSEUS and YACARI, the EU H2030 project CONVERGE, and by a Huawei France funded Chair towards Future Wireless Networks.

## REFERENCES

- [1] M. J. Wainwright, M. I. Jordan *et al.*, "Graphical Models, Exponential Families, and Variational Inference," *Foundations and Trends® in Machine Learning*, vol. 1, no. 1–2, 2008.
- [2] T. Minka *et al.*, "Divergence Measures and Message Passing," Citeseer, Tech. Rep., 2005.
- [3] T. Heskes, M. Opper, W. Wiegand, O. Winther, and O. Zoeter, "Approximate Inference Techniques with Expectation Constraints," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2005, no. 11, 2005.
- [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*, vol. 20, no. 7, 2021.
- [5] Z. Zhao, F. Xiao, and D. Slock, "Approximate Message Passing for Not So Large IID Generalized Linear Models," in *Int'l Workshop on Signal Processing Advances in Wireless Comm's (SPAWC)*, 2023.
- [6] A. Karataev, C. Forsch, and L. Cottatellucci, "Bilinear expectation propagation for distributed semi-blind joint channel estimation and data detection in cell-free massive mimo," *IEEE Open Journal of Signal Processing*, vol. 5, pp. 284–293, 2024.
- [7] R. F. Fischer, C. Sippel, and N. Goertz, "Vamp with vector-valued diagonalization," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 9110–9114.
- [8] K. Murphy, Y. Weiss, and M. I. Jordan, "Loopy Belief Propagation for Approximate Inference: An Empirical Study," *arXiv preprint arXiv:1301.6725*, 2013.
- [9] Q. Zou and H. Yang, "A Concise Tutorial on Approximate Message Passing," *arXiv preprint arXiv:2201.07487*, 2022.
- [10] S. Rangan, P. Schniter, E. Riegler, A. K. Fletcher, and V. Cevher, "Fixed Points of Generalized Approximate Message Passing with Arbitrary Matrices," *IEEE Transactions on Information Theory*, vol. 62, no. 12, 2016.
- [11] M. Triki and D. Slock, "Component-Wise Conditionally Unbiased Bayesian Parameter Estimation: General Concept and Applications to Kalman Filtering and LMMSE Channel Estimation," in *IEEE Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, USA, 2005.
- [12] M. Huemer and O. Lang, "CWCU LMMSE Estimation: Prerequisites and Properties," *arXiv preprint arXiv:1412.1567*, 2014.
- [13] C. Sippel and R. F. Fischer, "Variants of VAMP for Signal Recovery in Wireless Sensor Networks," in *2022 IEEE International Mediterranean Conference on Communications and Networking (MeditCom)*, 2022.