Gaussian PDF Divisions 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 popular Message Passing algorithm. In contrast to Belief Propagation (BP), it projects beliefs to the exponential family at every update. The update for a posterior factor gets obtained by dividing the projected tilted pdf by the other approximate posterior factors. This division can easily pose problems such as negative variances. This may happen if the factor to be updated, which in many instances is a prior, is very spread out such as a super Gaussian pdf or a Gaussian mixture model. Upon closer inspection however, it turns out that the posterior and extrinsic variances are computed in EP using different pdfs. We propose two solutions to remedy the problem, both based on using the same pdf to compute extrinsic and posterior variances. A third solution is proposed based on revisiting the EP optimization criterion. Furthermore, in a Generalized Linear Model where the signal is composed of multiple sub-signals of constant magnitude, we propose employing the Von Mises distribution for messages in order to circumvent the issue of non-proper distributions. We had encountered this problem in semiblind channel estimation in which we exploit the finite alphabet for the unknown symbols.**

I. Introduction

Sparse signal recovery is a fundamental problem in signal processing with a wide range of applications. Many of these problems can be framed as the task of estimating a latent vector $\boldsymbol{x}$ based on a correlated observation vector $\boldsymbol{y}$ [1]. In the Bayesian framework, the complexity of Canonical Methods such as MMSE and MAP experiences exponential growth as the dimension of the problem grows.

By exploiting the structure of the models, graphical model based methods prove to be effective. With a given factored joint probabilistic model, a factor graph can be obtained by first writing down all the factors and variables involved in the probabilistic model and then connect the variables to all the factors that contain it as a parameter. Each link between variable and factor node has messages of two directions. If the factor graph is tree-structured, Belief Propagation (BP) transforms the global inference problem into a local inference problem as outlined by [2]. The marginal posterior (or belief at variable/factor node) is obtained as a product of all the messages to the node in question (and the factor itself if the node is a factor node). To update the messages between variable and factor nodes, marginalized belief is first computed (integrate out all variables except the one contained in the variable node). Finally, the message from the node in question is the quotient of the marginal belief and the message on the same link but with reversed direction (i.e., to the node).

Loopy Belief Propagation (LBP) extends BP by directly employing BP on a factorization scheme for $p(\boldsymbol{x}|\boldsymbol{y})$ that may involve loops [3]. In comparison to BP, LBP can be considered as an approximation method. A limitation of (L)BP is that the (iterative) updating scheme leads to pdfs that correspond to the product of a large number of messages, leading to high complexity. To address this issue, Expectation Propagation (EP) was introduced [4]. EP has been shown to share a similar updating scheme as (L)BP, but for computational efficiency, the messages in (L)BP are approximated as members of the family of exponential distributions [4].

Expectation Propagation (EP) is a popular Message Passing algorithm. In contrast to BP, it projects beliefs to the exponential family at every update to ensure that the messages passed nodes belong to the exponential family as well. However, due to the projection, the division step can easily pose problems as is known in the statistics literature. For instance, if the exponential family corresponds to Gaussians, the variance obtained in the division of Gaussians may sometimes be negative. This may happen if the is very spread out such as a super Gaussian pdf or a Gaussian mixture model. Solutions to this problem remain elusive. This may seem counter intuitive because this would imply that the use of the prior next to the other pdfs, which we can jointly call the extrinsic, would result in a larger (posterior) mean squared error than the extrinsic variance. Upon closer inspection however, it turns out that the posterior and extrinsic variances are computed in EP using different pdfs. We propose two solutions to remedy the problem, both based on using the same pdf to compute extrinsic and posterior variances. A third solution is proposed based on revisiting the EP optimization criterion. Furthermore, in a Generalized Linear Model where the signal is composed of multiple sub-signals of constant magnitude, we propose employing the Von Mises distribution for messages in order to circumvent the issue of non-proper distributions. We had encountered this problem in semiblind channel estimation in which we exploit the finite alphabet for the unknown symbols.

II. EP Negative Gaussian Quotient Variance Issue

Consider a factored Joint Probabilistic Model

$$p(\boldsymbol{x}) \propto \prod_{\alpha} f_\alpha(\boldsymbol{x}_\alpha) \tag{1}$$

where $\boldsymbol{x}_\alpha$ is a sub-vector of $\boldsymbol{x}$ (all entries in $\boldsymbol{x}_\alpha$ are also entries of $\boldsymbol{x}$). The marginal distribution of $\forall i,\, x_i$ is often hard to obtain due to the intractable integrals. EP can be understood as an iterative method which convert the global inference problem into local ones. Following [4], EP can be concluded as the following iteration

$$b_{f_\alpha}(x_i) \propto m_{x_i \to f_\alpha}(x_i) \int f_\alpha(\boldsymbol{x}_\alpha) \prod_{j \in N(f_\alpha)/\{i\}} m_{x_j \to f_\alpha}(x_j)\, dx_j \tag{2}$$

$$m_{f_\alpha \to x_i}(x_i) = \frac{Proj_{\mathcal{F}}\{b_{f_\alpha}(x_i)\}}{m_{x_i \to f_\alpha}(x_i)} \tag{3}$$

$$m_{x_i \to f_\alpha}(x_i) = \prod_{\beta \in N(x_i)/\{\alpha\}} m_{f_\beta \to x_i}(x_i), \tag{4}$$

where $N(f_\alpha)$ and $N(x_i)$ denote the neighborhood set around f_α and x_i respectively. The $Proj_{\mathcal{F}}\{b(x)\}$ projects $b(x)$ into another distribution within a family $\mathcal{F}$ by KL-Divergence

$\arg\min_{\widehat{b}\in\mathcal{F}} KLD(b\|\widehat{b})$. In most cases $\mathcal{F}$ is assumed to be Gaussian family, and the projection simply means moment matching of first and second order moments. It is unnecessary for the messages $m_{\cdot\to\cdot}$ to be proper distributions. However, the belief b_{f_α} must satisfy the definition of a distribution. Actually, we can view b_{f_α} as an approximated posterior. That is why we call it belief at factor node. Furthermore, we can define $\forall\alpha \in N(x_i)$, $b_{x_i} = m_{f_\alpha\to x_i} m_{x_i\to f_\alpha}$ as a belief at variable node.

Revisited Generalized Vector Approximate Message Passing (reGVAMP) is an application of EP with a Generalized Linear Model (GLM) [5]

$$p(\boldsymbol{x},\mathbf{z}|\boldsymbol{y}) = \left(\prod_{m=1}^{M} f_{z_m}(z_m)\right) f_\delta(\mathbf{z},\boldsymbol{x}) \left(\prod_{n=1}^{N} f_{x_n}(x_x)\right)$$

where $f_{z_m}(z_m) = p(y_m|z_m)$, $f_{x_n}(x_x) = p(x_n)$, $f_\delta(\mathbf{z},\boldsymbol{x}) = \delta(\mathbf{z} - \mathbf{A}\boldsymbol{x})$. Due to the factorization, if we consider the belief b_{f_δ} derived from (2), it is natural to treat $m_{f_{x_i}\to x_i}$ as approximated priors. In reGVAMP, the projection family $\mathcal{F}$ is defined to be Gaussian family. Therefore, the division in (3) may lead to a non-distribution function (or Gaussian with negative variance) if some moments are constrained. In the Gaussian case, (co)variances are required to be positive (definite). Indefiniteness problems can occur when f_{x_i} is more widely spread than a Guassian, e.g. super Gaussian pdfs or Gaussian Mixture Model (GMM), includes discrete pmfs (e.g. finite alphabet for detection).

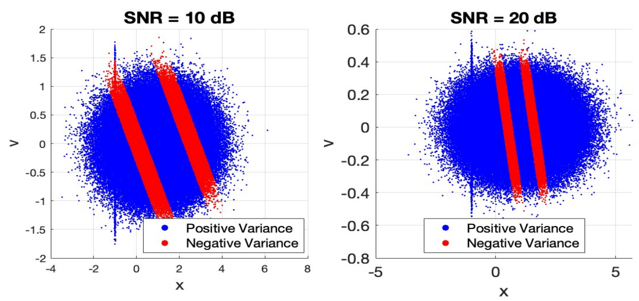

Fig. 1. Negative prior variance issue in reVAMP with GMM prior.

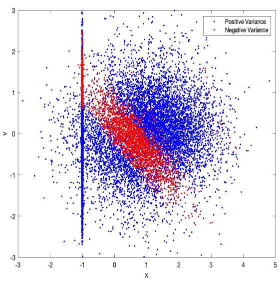

Fig. 2. Negative prior variance in global MMSE solution, SNR=10dB.

A. EP Negative Gaussian Quotient Variance : Proposed Solutions

The detailed derivation of reGVAMP can be found in [5]. We consider here the computation of $b_{f_{x_i}}$ and $m_{f_{x_i}\to x_i}$. Since the posterior can be computed as $b_{f_{x_i}} \propto m_{x_i\to f_{x_i}} \cdot f_{x_i}$, we shall call $m_{x_i\to f_{x_i}}$ "extrinsic". Furthermore, we denote the mean and variance of $b_{f_{x_i}}$ as $\mu_{\widehat{x}_i}$ and $\tau_{\widehat{x}_i}$; the mean and variance of $m_{x_i\to f_{x_i}}$ as μ_{r_i} and τ_{r_i}; the mean and variance of $m_{f_{x_i}\to x_i}$ as μ_{x_i} and τ_{x_i}. Since in the later context of this section we only focus on one symbol, we omit the subscript indices. The computation of the belief $b_{f_{x_i}}$ leads to a simple measurement model

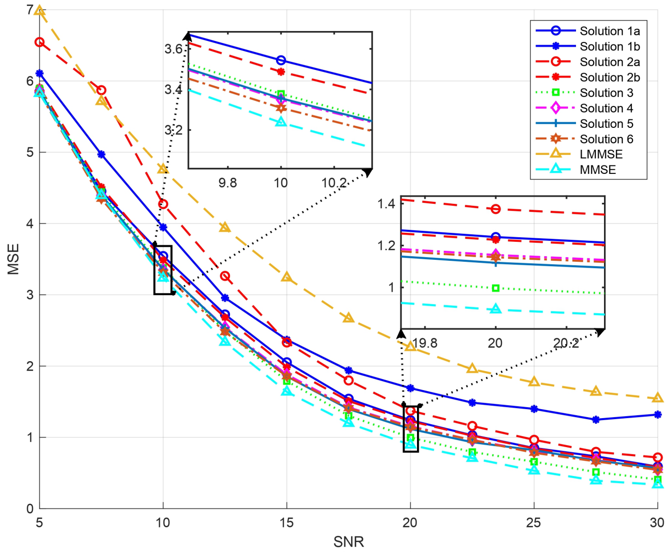

Fig. 3. NMSE for $\boldsymbol{x}$ as a function of SNR for various reVAMP solutions with GMM prior, and the LMMSE and MMSE solutions.

$$\mu_r = x + v_r, \text{ with } x \sim f_{x_i}, \text{ var}(v_r) = \tau_r. \quad (5)$$

Solution 1: posterior variance averaging : $\tau_{\widehat{x}} = \mathbb{E}_{\mu_r}[\tau_{\widehat{x}}(\mu_r)]$

$$\tau_r = \mathbb{E}_{x,\mu_r}(x-\mu_r)^2 = \tau_{\widehat{x}} + \mathbb{E}_{\mu_r}(\mu_{\widehat{x}} - \mu_r)^2 \geq \tau_{\widehat{x}} \quad (6)$$

where we exploited the orthogonality property of MMSE estimation. Note that Solution 1 coincides with noise unbiasing Advantage: no effect in Gaussian case. Disadvantage: carrying out $\mathbb{E}_{\mu_r}$ may be a complex operation. Actually, for the prior mean, with the usual update formula based on the Gaussian quotient as in [6], we consider two variations. Solution 1a: take the standard MMSE estimate $\mu_{\widehat{x}}$ for the posterior mean; Solution 1b: replace $\mu_{\widehat{x}}$ by μ_r, which also leads to μ_r as prior mean.

Solution 2: recompute the extrinsic variance as $\mathbb{E}_{x|\mu_r}(x-\mu_r)^2$.

$$\mathbb{E}_{x|\mu_r}(x-\mu_r)^2 = \tau_{\widehat{x}}(\mu_r) + (\mu_{\widehat{x}}(\mu_r) - \mu_r)^2 \geq \tau_{\widehat{x}}(\mu_r)\,. \quad (7)$$

Note that (6) can be obtained from (7) by taking $\mathbb{E}_{\mu_r}\{.\}$. Advantage: easy to apply. Disadvantage: this solution modifies the Gaussian case. Again, for the prior mean we consider the two variations. Solution 2a: maintain $\mu_{\widehat{x}}$ for the posterior mean; Solution 2b: replace $\mu_{\widehat{x}}$ by μ_r, which also leads to μ_r as prior mean. This second choice is more consistent with the prior variance calculation in Solution 2. Note that Noise Unbiasing [6] corresponds to a combination of Solution 1a for the prior mean and something similar to Solution 2 for the prior variance.

Solution 3: reconsider EP optimization.

Consider the projection operation of b_{f_x}. In the projection, we minimize

$$\begin{aligned}&\mathrm{KLD}(b_{f_x}||\widehat{b}_{f_x})\\&=\tfrac{\ln\tau_{\widehat{x}}(\mu_r)}{2}+\tfrac{1}{2\tau_{\widehat{x}}(\mu_r)}\int b_{f_x}(x)\,(x-\mu_{\widehat{x}}(\mu_r))^2dx+C.\end{aligned}\tag{8}$$

Minimization of the KLD w.r.t. $\mu_{\widehat{x}}$ leads to $\mu_{\widehat{x}} = \int x\, b_{f_x}(x)\, dx$. Define $\xi_x = \tau_x^{-1}$. Due to the division operation in (3), we have $\tau_{\widehat{x}}^{-1} = \tau_x^{-1} + \tau_r^{-1}$. Therefore, apart from a constant and a scale factor we rewrite (8) as a function of ξ_x:

$$\mathrm{KLD}(\xi_x) \propto -\ln(\xi_x+\tau_r^{-1}) + (\xi_x+\tau_r^{-1})\,\tau_{\widehat{x}}(\mu_r) \tag{9}$$

which needs to be minimized w.r.t. the prior precision $\xi_x \geq 0$, and where $\tau_{\widehat{x}}(\mu_r) = \int b_{f_x}(x)\,(x-\mu_{\widehat{x}}(\mu_r))^2dx$. $KLD(\xi_x)$ exhibits an extremum at $\xi_x^o = \tau_x^{-1} - \tau_r^{-1}$, which can be negative, and is increasing for $\xi_x > \xi_x^o$. Hence, under the constraint $\xi_x \geq 0$, the optimum is attained at $\xi_x = \max(0, \tau_x^{-1} - \tau_r^{-1})$. When $\xi_x = 0$, the prior mean m_x is unimportant. Solution 3 can be extended to the matrix case by clipping the eigenvalues of the quotient precision matrix to be nonnegative.

III. Prior with Constant Magnitude

For a class of signal recovery problems where the signal has a hierachical structure and can be decomposed into multiple sub-signals with constant maginitude, the Von Mises distribution can be used to approximate the messages [7]. Assume that the observation of the GLM is slightly modified by

$$p(\check{\boldsymbol{x}},\mathbf{z}|\boldsymbol{y}) = \left(\prod_{m=1} f_{z_m}(z_m)\right) f_\delta(\check{\boldsymbol{x}},\mathbf{z}) \left(\prod_{n,k} f_{x_{nk}}(x_{nk})\right)$$

with $f_\delta(\check{\boldsymbol{x}},\mathbf{z}) = \delta(\mathbf{z}-\mathbf{A}\check{\boldsymbol{x}})$, $f_{z_m} = p(y_m|z_m)$, $f_{x_{nk}} = p(x_{nk})$, where $\forall n \in [1,N]$, $\check{x}_n = \sum_{k=1}^K x_{nk}$, and the random variable x_{nk} is i.i.d. with constant magnitude, i.e., $|x_{nk}|^2 = \sigma_{x_{nk}}^2$. For simplicity, we also denote $\check{\boldsymbol{x}} = \sum_k \boldsymbol{x}_k$. It is easy to see all PSK and 4^K-QAM satisfy this assumption. To avoid the negative variances in the messages, we use Von Mises distribution to approximate the message $m_{f_{x_{nk}}\to x_{nk}}$. We will approximate the prior of x_{nk} as

$$x_{nk} = \sigma_{x_{nk}} e^{i\theta_{nk}}, \tag{10}$$

and denote the corresponding prior for θ_{nk} as $p_{\theta_{nk}}(\theta_{nk})$. Therefore, the approximated prior $m_{f_{x_{nk}}\to x_{nk}}$ is implicitly given by

$$\theta_{nk} \sim m_{f_{\theta_{nk}}\to\theta_{nk}}(\theta_{nk}) := \mathcal{VM}(\theta_{nk}|\mu_{\theta_{nk}},\kappa_{\theta_{nk}}). \tag{11}$$

The approximated prior mean and variance of x_{nk} is obtained as the circular mean and variance of $m_{f_{\theta_{nk}}\to\theta_{nk}}$, which are

$$\mu_{x_{nk}} = \frac{I_1(\kappa_{\theta_{nk}})}{I_0(\kappa_{\theta_{nk}})}\sigma_x e^{i\mu_{\theta_{nk}}},\ \tau_{x_{nk}} = \sigma_{x_{nk}}^2\left(1-\left[\frac{I_1(\kappa_{\theta_{nk}})}{I_0(\kappa_{\theta_{nk}})}\right]^2\right).$$

Based on the previous discussion, we use approximated priors instead of true prior in the extrinsic terms in $m_{\delta\to z_m} = m_{z_m\to f_{z_m}}$ and $m_{\delta\to x_{nk}} = m_{x_{nk}\to f_{x_{nk}}}$. We assume $m_{f_{z_m}\to z_m}$ to be Gaussian, just like in [5]. Due to CLT, the extrinsic $m_{\delta\to z_m}$ can be calculated as

$$m_{\delta\to z_m}(z_m) = \int\prod_{m'\neq m}\delta(z_m - \mathbf{A}_{m,:}\check{\boldsymbol{x}})\prod_{n,k}\mathcal{N}(x_{nk}|\mu_{x_{nk}},\sigma_{\widehat{x}_{nk}})dx_{nk}.$$

Following [5], we can compute the update of message $m_{f_{z_m}\to z_m}$ which is a Gaussian with mean and variance μ_{z_m} and τ_{z_m}. Likewise, apply Central Limit Theory (CLT) approximation and the the extrinsic $m_{\delta\to x_{nk}}$ becomes

$$m_{\delta\to x_{nk}}(x_{nk}) = \int\mathcal{N}(\mathbf{A}\boldsymbol{x}|\boldsymbol{\mu}_{\mathbf{z}},\mathbf{C}_{\mathbf{z}})\prod_{(n',k')\neq(n,k)} m_{x_{n'k'}\to\delta}(x_{n'k'})d\boldsymbol{x}_{n'k'}.$$

Denote the extrinsic as $m_{\delta\to x_{nk}}(x_{nk}) = \mathcal{N}(x_{nk}|\mu_{r_{nk}},\tau_{r_{nk}})$ which can be obtained as

$$\begin{aligned}\boldsymbol{\mu}_{\boldsymbol{u}_{nk}} &= \sum_{k'\neq k}\boldsymbol{\mu}_{\boldsymbol{x}_{k'}} + \sum_{n'\neq n}\boldsymbol{e}_{n'}\mu_{x_{n'k}}\\ \boldsymbol{\tau}_{\boldsymbol{u}_{nk}} &= \sum_{k'\neq k}\boldsymbol{\tau}_{\boldsymbol{x}_{k'}} + \sum_{n'\neq n}\boldsymbol{e}_{n'}\tau_{x_{n'k}}\\ \mathbf{C}_{\boldsymbol{u}_{nk}} &= Diag(\boldsymbol{\tau}_{\boldsymbol{u}_{nk}})\\ \tau_{r_{nk}} &= \left[\boldsymbol{a}_n^H\left(\mathbf{C}_{\mathbf{z}} + \mathbf{A}\mathbf{C}_{\boldsymbol{u}_{nk}}\mathbf{A}^H\right)^{-1}\boldsymbol{a}_n\right]^{-1}\\ r_{nk} &= \tau_{r_{nk}}\boldsymbol{a}_n^H\left(\mathbf{C}_{\mathbf{z}} + \mathbf{A}\mathbf{C}_{\boldsymbol{u}_{nk}}\mathbf{A}^H\right)^{-1}(\boldsymbol{\mu}_{\mathbf{z}} - \mathbf{A}\boldsymbol{\mu}_{\boldsymbol{u}_{nk}})\end{aligned}\tag{12}$$

where $\boldsymbol{e}_n$ is a unit vector with 1 at the n-th entry. From this point, we can compute the posterior mean and variance of x_{nk} based on $\mathcal{N}(x_{nk}|\mu_{r_{nk}},\tau_{r_{nk}})p_{x_{nk}}(x_{nk})$, which are denoted as $\mu_{\widehat{x}_{nk}}$ and $\tau_{\widehat{x}_{nk}}$. Using the definition of (10), we project the extrinsic of x_{nk} onto a circle to obtain pdf of θ_{nk}

$$m_{\delta\to\theta_{nk}}(\theta_{nk}) = \int m_{\delta\to x_{nk}}(x_{nk})\delta(x_{nk} - \sigma_{x_{nk}}e^{i\theta_{nk}})dx_{nk} \tag{13}$$

We define $\mu_{r_{nk}} = w_{nk}e^{i\rho_{nk}}$, $\kappa_{r_{nk}} = \frac{2w_{nk}\sigma_{x_{nk}}}{\tau_{r_{nk}}}$, and (13) becomes

$$m_{\theta_{nk}}(\theta_{nk}) \propto e^{\kappa_{r_{nk}}\cos(\theta_{nk}-\rho_{nk})}. \tag{14}$$

The belief of x_{nk} can be obtained via (2) as $b_{f_{x_{nk}}}(x_{nk}) = f_{x_{nk}}\cdot m_{\delta\to x_{nk}}$. Denote the belief mean and variance as $\mu_{\widehat{x}_{nk}}$ and $\tau_{\widehat{x}_{nk}}$. Based on the posterior mean and variance of x_{nk}, and following a procedure analogous to that described in equations (13)–(14), we transform the approximated belief into θ_{nk} representation. The corresponding posterior parameters $\kappa_{\widehat{\theta}_{nk}}$ and $\mu_{\widehat{\theta}_{nk}}$ are

$$\begin{aligned}\kappa_{\widehat{\theta}_{nk}} &= \frac{2|\mu_{\widehat{x}_{nk}}|\sigma_{x_{nk}}}{\tau_{\widehat{x}_{nk}}},\\ \mu_{\widehat{\theta}_{nk}} &= \arctan\frac{\Im(\mu_{\widehat{x}_{nk}})}{\Re(\mu_{\widehat{x}_{nk}})} + \begin{cases}0,\ \Re(\mu_{\widehat{x}_{nk}}) > 0\\ \pi,\ \Re(\mu_{\widehat{x}_{nk}}) < 0\end{cases}\end{aligned}\tag{15}$$

The approximated prior in (11) is then updated by

$$\begin{aligned}m_{f_{\theta_{nk}}\to\theta_{nk}}(\theta_{nk}) &= \frac{\mathcal{VM}(\theta_{nk}|\mu_{\widehat{\theta}_{nk}},\kappa_{\widehat{\theta}_{nk}})}{m_{\theta_{nk}}(\theta_{nk})}\\ &\propto \mathcal{VM}(\theta_{nk}|\mu_{\theta_{nk}},\kappa_{\theta_{nk}})\end{aligned}\tag{16}$$

where the updated $\mu_{\theta_{nk}}$, $\kappa_{\theta_{nk}}$ are

$$\begin{aligned}\kappa_{\theta_{nk}} &= \Big[(\kappa_{\widehat{\theta}_{nk}}\cos\mu_{\theta_{nk}} - \kappa_{r_{nk}}\cos\rho_{nk})^2\\ &\quad + (\kappa_{\widehat{\theta}_{nk}}\sin\mu_{\theta_{nk}} - \kappa_{r_{nk}}\sin\rho_{nk})^2\Big]^{\frac{1}{2}}\\ \mu_{\theta_{nk}} &= \arctan\frac{\kappa_{\widehat{\theta}_{nk}}\sin\mu_{\theta_{nk}} - \kappa_{r_{nk}}\sin\rho_{nk}}{\kappa_{\widehat{\theta}_{nk}}\cos\mu_{\theta_{nk}} - \kappa_{r_{nk}}\cos\rho_{nk}}\\ &\quad + \begin{cases}0,\ \text{denominator} > 0;\\ \pi,\ \text{denominator} < 0\end{cases}\end{aligned}\tag{17}$$

IV. Simulation Results

We consider a real linear model with $\mathbf{A}$ of size 10×10 and i.i.d. entries $\mathcal{N}(0, 0.1)$, and additive white Gaussian noise. One of the prior models for which the overall MMSE solution can be computed analytically is GMM. Hence we consider an i.i.d. prior

$$p(x_i) = 0.5\mathcal{N}(x_i; 1, 1) + 0.5\mathcal{N}(x_i; -1, 0.01) \quad (18)$$

which has zero mean. The results involve 10,000 realizations of $\boldsymbol{x}$ and $\boldsymbol{v}$. In Fig. 1 we consider a (reGVAMP) scalar MMSE problem as in (5). With the prior as in (18), we explore in the (x, v) plane where the issue of a negative quotient variance (in red) occurs (posterior variance larger than extrinsic variance). In Fig. 2, we explore the same issue, at SNR=10dB, for the global MMSE and extrinsic variances. The results are similar though not identical to what appears in reGVAMP. In any case, the negative quotient variance issue is clearly inherent in some Bayesian estimation scenarios. In Fig. 3, we plot the MSE of $\boldsymbol{x}$ as a function of SNR for reGVAMP EP at convergence, with various solutions for the negative quotient issue. The equivalent Gaussian priors are initialized by direct Gaussian matching of the prior (corresponding to LMMSE). Solutions 1 to 3 are as described in the previous section. In Solution 4 we don't update the prior in case of a negative prior variance update, and in Solution 5 we use the posterior mean and variance for their updated prior counterparts. Solution 6 corresponds to Noise Unbiasing. Finally, we also plot the MSE of LMMSE and global MMSE estimation. We can note that at each SNR, the best solution is very close to the global MMSE solution, which is far from LMMSE. However, at sufficiently high SNR, Solution 3 is the best, whereas for lower SNR it's Solution 6. Note that given the results in Figures 1 and 2, the negative prior variance issue is not just a transient issue during iterations, but can very well persist at convergence! For the Von-Mises Approximation, we consider a complex linear model with 4-QAM signal input and AWGN channel. The transmitted data has power $\sigma_x^2 = 1$. The measurement matrix has i.i.d. Gaussian entries, each of which is zero mean and has power $\sigma_{h_{mn}}^2 = 1/N$. The noise is zero mean Gaussian, whose power is tuned to mean the SNR requirement. To verify our proposed method, we plot the Symbol Error Rate (SER) vs SNR. We simulate different pairs of $M, N \in \{4, 8\}$ to show the performance in compressed sensing. The simulation results are plotted in Fig. 4.

V. Concluding Remarks

In this paper, we recalled the EP approach and reGVAMP, an application of EP to the GLM. We pointed out moment constraint violation issues that can occur in the EP pdf division step and elaborated on existing and new solutions for the case of negative quotient variances when using Gaussian pdfs. We also investigate the Von Mises message approximation, where the negative quotient variances is avoided completely. An open issue that arises from the simulations is to find a solution that would be optimal at all SNR.

We may refer the reader also to [8] where we present AM-BGAMP, a provably convergent fast version of reGVAMP. GAMP does not build equivalent Gaussian priors explicitly but nevertheless also has a quotient variance issue. [8] and this paper build upon the works of [1], [9], [10], [5]. Whereas GAMP is based on asymptotics of an i.i.d. (sign) model for $\mathbf{A}$, more general models are considered in [11], [12], [13], [14].

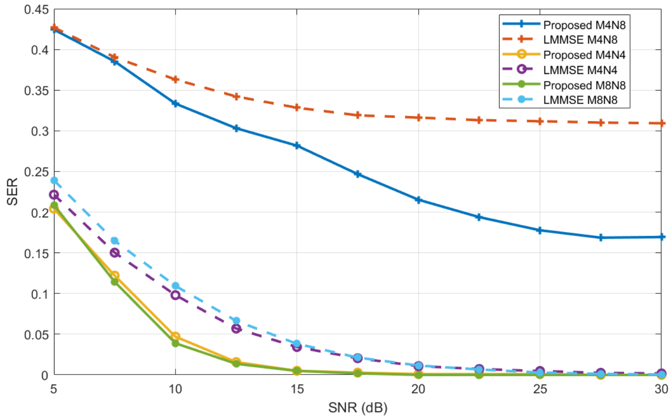

Fig. 4. Using Von Mises distribution as Message

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] 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.

[2] 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.

[3] K. Murphy, Y. Weiss, and M. I. Jordan, "Loopy Belief Propagation for Approximate Inference: An Empirical Study," *arXiv preprint arXiv:1301.6725*, 2013.

[4] T. Minka *et al.*, "Divergence Measures and Message Passing," Citeseer, Tech. Rep., 2005.

[5] Z. Zhao, F. Xiao, and D. Slock, "Vector approximate message passing for not so large N.I.I.D. generalized I/O linear models," in *IEEE Int'l Conf. Acoustics, Speech and Sig. Proc. (ICASSP)*, Seoul, 2024.

[6] R. F. Fischer, C. Sippel, and N. Goertz, "VAMP with Vector-Valued Diagonalization," in *IEEE Int'l Conf. Acoustics, Speech and Signal Processing (ICASSP)*, 2020.

[7] Z. Zhao and D. Slock, "Hierarchical Expectation Propagation for Semi-Blind Channel Estimation in Cell-Free Networks," in *ICASSP, International Conference on Acoustics, Speech and Signal Processing*, IEEE, Ed., 2025.

[8] C. Kurisummoottil Thomas, Z. Zhao, and D. Slock, "Towards Convergent Approximate Message Passing by Alternating Constrained Minimization of Bethe Free Energy," in *IEEE Information Theory Workshop (ITW)*, Saint Malo, France, 2023.

[9] S. Rangan, A. Fletcher, P. Schniter, and U. Kamilov, "Inference for Generalized Linear Models via Alternating Directions and Bethe Free Energy Minimization," *IEEE Trans. Info. Theory*, Jan. 2017.

[10] Z. Zhao and D. Slock, "Bethe Free Energy and Extrinsics in Approximate Message Passing," in *IEEE Asilomar Conf. Signals, Systems and Computers*, 2023.

[11] S. Rangan, P. Schniter, and A. K. Fletcher, "Vector Approximate Message Passing," *IEEE Trans. On Info. Theo.*, Oct. 2019.

[12] C. K. Thomas and D. Slock, "SAVE - Space alternating variational estimation for sparse Bayesian learning," in *IEEE Data Science Workshop*, June 2018.

[13] Z. Zhao and D. Slock, "Variance Predictions in VAMP/UAMP with Right Rotationally Invariant Measurement Matrices for niid Generalized Linear Models," in *European Sig. Proc. Conf. (EUSIPCO)*, Helsinki, Finland, 2023.

[14] ——, "Improved Variance Predictions in Approximate Message Passing," in *IEEE Int'l Workshop Machine Learning and Sig. Proc. (MLSP)*, Rome, Italy, 2023.