Extrinsics and Linearized Component-Wise Conditionally Unbiased MMSE Estimation as in GAMP

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

***Abstract*—Generalized Approximate Message Passing (GAMP) algorithms have demonstrated significant efficacy in signal recovery. GAMP has been derived by applying asymptotic approximations to Expectation Propagation (EP). EP algorithms start from a factored approximate posterior in an exponential family. They update a factor by fitting an exponential family pdf to a approximate posterior which is obtained by replacing one approximate factor by the original (prior) factor. The remaining factors form the approximate extrinsic. Hence extrinsics are obtained by marginalizing the product of all pdf factors except for the prior. A marginal posterior is then obtained by combining the extrinsic with the prior. Low complexity algorithms like GAMP in turn obtain the extrinsic from the posterior. In the Gaussian case, we reveal the intimate relation of extrinsics to Component-Wise Conditionally Unbiased Minimum Mean Squared Error (CWCU MMSE) estimation, whereas the posterior allows MMSE estimation. In the Gaussian case, MMSE estimation means Linear MMSE estimation, non-Gaussianity leads to nonlinear estimators. We rederive the revisited GVAMP algorithm as asymptotic alternating minimization of a Kullback-Leibler Divergence. We then explore the extrinsics in GAMP by asymptotic perturbations relating posterior beliefs and extrinsics.**

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. Belief Propagation (BP) transforms the global inference problem into a local inference problem as outlined by [2]. 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 projected into a suitable member of the family of exponential distributions [4].

A. Prior Work

In both [1] and [5], the authors unify EP and BP within the framework of minimizing variational free energy. They demonstrate the close relationship between the fixed points of various message-passing algorithms and the stationary points of Bethe Free Energy (BFE).

EP can serve as an inference method in the linear Gaussian model. However, the computational cost in terms of the message count is quadratic in the data size. Approximate Message Passing (AMP) [6] builds upon EP, but through the application of large system approximations (LSA), it effectively reduces the number of messages to the order of the data size, providing a more computationally efficient approach.

In [7], the authors investigated the fixed points of the Generalized AMP (GAMP) algorithm for generalized linear models (GLMs). They discovered that GAMP shares the same fixed point as the stationary points of the Large System Limit Bethe Free Energy (LSL BFE).

B. Main Contributions

We rederive the reGVAMP algorithm that we introduced in [8], [9], from the point of view of alternating minimization of a LSL version of a desirable KLD. The asymptotics here involve only the CLT for extrinsics. We then derive the GAMP algorithm by directly introducing LSL simplifications in the LBP algorithm. This leads us to relate extrinsic messages to posterior pdfs by first order Taylor series expansion based perturbations. We also apply LSL approximations to the variances of the various Gaussians involved, which in fact leads to a rederivation of a fundamental LSA theorem describing the deterministic limit of LMMSE posterior variances.

II. Generalized Linear Model (GLM)

We consider a GLM with

$$p(\boldsymbol{x})=\textstyle\prod_{i=1}^{N} p(x_i),\ \mathbf{z}=\mathbf{A}\boldsymbol{x},\ p(\boldsymbol{y}|\mathbf{z}) = \prod_{j=1}^{M} p(y_j|z_j), \quad (1)$$

where the ratio N/M is a constant for large system considerations. We interpret the linear mixing as a conditional probability

$$p(\mathbf{z}|\boldsymbol{x}) = \delta(\mathbf{z} - \mathbf{A}\boldsymbol{x}). \quad (2)$$

This generalized linear model is characterized by the following factored posterior pdf:

$$p(\boldsymbol{x}, \mathbf{z}|\boldsymbol{y}) \propto p(\boldsymbol{x}, \boldsymbol{y}, \mathbf{z}) = p(\boldsymbol{y}|\mathbf{z})\, \delta(\mathbf{z} - \mathbf{A}\boldsymbol{x})\, p(\boldsymbol{x}). \quad (3)$$

The problem in Bayesian estimation is the computation of the normalization constant $p(\boldsymbol{y})$ and of the posterior means and variances (if the interest is MMSE estimation).

III. reGVAMP

reGVAMP (revisited Generalized Vector AMP) is motivated by only a *single asymptotic approximation*: the asymptotic Gaussianity of extrinsics. The extrinsic pdf of a variable x_i is the conditional pdf $p(\boldsymbol{y}|x_i)$, in which x_i is treated as a deterministic variable (no prior information), but the other variables $\boldsymbol{x}_{\bar{i}}$ remain random and their prior pdf is exploited to eliminate them from the joint pdf. The randomness of $\boldsymbol{x}$ and $\mathbf{A}$ will quickly lead to Gaussianity of $p(\boldsymbol{y}|x_i)$ by the CLT (think of asymptotic Gaussianity of Maximum Likelihood estimates). reVAMP introduces both Gaussian and non-Gaussian marginal posteriors from Gaussian extrinsics and the true prior. This involves also the introduction of Gaussian approximations for the priors. Which in turn also leads to a multivariate Gaussian posterior approximation, which exhibits the posterior correlations between the variables. reGVAMP postulates a factored posterior approximation of the form

$$\begin{aligned} &q_{\boldsymbol{x},\mathbf{z}}(\boldsymbol{x},\mathbf{z}) = \textstyle\prod_i q_{x_i|\boldsymbol{y}}(x_i) \prod_j q_{z_j|\boldsymbol{y}}(z_j) \\ &= \textstyle\prod_i q_{x_i}(x_i) m_{x_i}(x_i) \prod_j q_{z_j}(z_j) m_{z_j}(z_j), \end{aligned} \tag{4}$$

where q_{x_i} and q_{z_j} are the Gaussian approximations for the priors while m_{x_i} and m_{z_j} are the Gaussian extrinsics for x_i and z_j.

A byproduct are non-Gaussian posterior marginals, e.g. of the form $m_i(x_i)\, p(x_i)$ where $p(x_i)$ is the true prior for x_i. Note that involving the true priors is something that could also be considered in Variational Bayes (VB) [1]. reVAMP attemps to optimize the better $\mathrm{KLD}(p,q)$ whereas VB optimizes $\mathrm{KLD}(q,p)$.

So, reGVAMP performs alternating minimization of the following KLD $\arg\min_{q_{\boldsymbol{x},\mathbf{z}|\boldsymbol{y}}} \mathrm{KLD}(p(\boldsymbol{x},\mathbf{z}|\boldsymbol{y}) \| q_{\boldsymbol{x},\mathbf{z}|\boldsymbol{y}}(\boldsymbol{x},\mathbf{z}))$, with the approximate posterior as in (4). The KLD becomes

$$\begin{aligned} &\mathrm{KLD}(p(\boldsymbol{x},\mathbf{z}|\boldsymbol{y}) \| q_{\boldsymbol{x}|\boldsymbol{y}}(\boldsymbol{x})] + \mathrm{KLD}[p(\boldsymbol{x},\mathbf{z}|\boldsymbol{y}) \| q_{\mathbf{z}|\boldsymbol{y}}(\mathbf{z})) + \mathrm{c}^{\mathrm{t}} \\ &= \textstyle\sum_i \mathrm{KLD}(p(\boldsymbol{x},\mathbf{z}|\boldsymbol{y}) \| q_{x_i|\boldsymbol{y}}(x_i)) \\ &\quad + \textstyle\sum_j \mathrm{KLD}(p(\boldsymbol{x},\mathbf{z}|\boldsymbol{y}) \| q_{z_j|\boldsymbol{y}}(z_j)) + \mathrm{c}^{\mathrm{t}} \\ &= \textstyle\sum_i \mathrm{KLD}(p(x_i|\boldsymbol{y}) \| q_{x_i|\boldsymbol{y}}(x_i)) \\ &\quad + \textstyle\sum_j \mathrm{KLD}(p(z_j|\boldsymbol{y}) \| q_{z_j|\boldsymbol{y}}(z_j)) + \mathrm{c}^{\mathrm{t}} \end{aligned} \tag{5}$$

where c^{t} denotes some constant. In the last equality, we marginalized out the irrelevant variables. The marginalized posteriors $p(x_i|\boldsymbol{y})$ and $p(z_j|\boldsymbol{y})$ are

$$p(x_i|\boldsymbol{y}) \propto \underbrace{p_{x_i}(x_i)}_{\text{prior}} \underbrace{\int p(\boldsymbol{y}|\mathbf{z}) p(\mathbf{z}|\boldsymbol{x}) \prod_{k\neq i} p_{x_k}(x_k) d\mathbf{z} d\boldsymbol{x}_{\bar{i}}}_{\text{extrinsic } p(\boldsymbol{y}|x_i)}, \tag{6}$$

$$\begin{aligned} &p(z_j|\boldsymbol{y}) \propto p(\boldsymbol{y}, z_j) = \textstyle\int p(\boldsymbol{y},\mathbf{z}) d\mathbf{z}_{\bar{j}} \\ &= \underbrace{p_{y_j|z_j}(z_j)}_{\text{prior}} \underbrace{\int \prod_{k\neq j} p_{y_k|z_k}(z_k) \delta(\mathbf{z} - \mathbf{A}\boldsymbol{x}) p(\boldsymbol{x}) d\boldsymbol{x} d\mathbf{z}_{\bar{j}}}_{\text{extrinsic } p(\boldsymbol{y}_{\bar{j}}, z_j)}. \end{aligned} \tag{7}$$

In order to see which probability the extrinsic for z corresponds to, consider the short hand notation

$$p(\mathbf{z}) = \int \delta(\mathbf{z} - \mathbf{A}\boldsymbol{x}) p(\boldsymbol{x}) d\boldsymbol{x} = p(\mathbf{z}_{\bar{j}}|z_j) p(z_j) \tag{8}$$

which depends only on the prior for $\boldsymbol{x}$. Therefore, in (7),

$$\begin{aligned} &\textstyle\int \prod_{k\neq j} p_{y_k|z_k}(z_k) \delta(\mathbf{z} - \mathbf{A}\boldsymbol{x}) p(\boldsymbol{x}) d\boldsymbol{x} \\ &= p_{\boldsymbol{y}_{\bar{j}}|\mathbf{z}_{\bar{j}}}(\mathbf{z}_{\bar{j}}) p(\mathbf{z}_{\bar{j}}|z_j) p(z_j) = p(\boldsymbol{y}_{\bar{j}}, \mathbf{z}_{\bar{j}}, z_j), \end{aligned} \tag{9}$$

Thus, we have

$$\begin{aligned} p(x_i|\boldsymbol{y}) &\simeq p_{x_i}(x_i) m_{x_i}(x_i), \\ p(z_j|\boldsymbol{y}) &\simeq p_{y_j|z_j}(z_j) m_{z_j}(z_j). \end{aligned} \tag{10}$$

Due to the CLT, the extrinsics can be approximated as Gaussian when system dimensions increase. The marginal KLDs become

$$\begin{aligned} &\arg\min_{q_{x_i|\boldsymbol{y}}} \mathrm{KLD}(p(x_i|\boldsymbol{y}) \| q_{x_i|\boldsymbol{y}}(x_i)) \\ &\simeq \arg\min_{q_{x_i}} \mathrm{KLD}(p_{x_i}(x_i) m_{x_i}(x_i) \| q_{x_i}(x_i) m_{x_i}(x_i)), \end{aligned} \tag{11}$$

$$\begin{aligned} &\arg\min_{q_{z_j|\boldsymbol{y}}} \mathrm{KLD}(p(z_j|\boldsymbol{y}) \| q_{z_j|\boldsymbol{y}}(z_j)) \\ &\simeq \arg\min_{q_{z_j}} \mathrm{KLD}(p_{y_j|z_j}(z_j) m_{z_j}(z_j) \| q_{z_j}(z_j) m_{z_j}(z_j)). \end{aligned} \tag{12}$$

A. reGVAMP from (Minka) EP

We can arrive at the same point (11),(12) by Minka-style EP. Approximate p by q at factor level, with
$p(\boldsymbol{x},\mathbf{z}|\boldsymbol{y}) = 1/Z_p \prod_i p_{x_i}(x_i) \prod_j p_{y_j|z_j}(z_j)\, \delta(\mathbf{z} - \mathbf{A}\boldsymbol{x})$,
$q(\boldsymbol{x},\mathbf{z}) = 1/Z_q \prod_i q_{x_i}(x_i) \prod_j q_{z_j}(z_j)\, m(\boldsymbol{x},\mathbf{z})$.
What is $m(\boldsymbol{x},\mathbf{z})$? The tilted pdf $\widetilde{p}_\delta = 1/Z_q \prod_i q_{x_i}(x_i) \prod_j q_{z_j}(z_j)\, \delta(\mathbf{z} - \mathbf{A}\boldsymbol{x})$ is already Gaussian, hence is unchanged after Gaussian projection. So we can take $m(\boldsymbol{x},\mathbf{z}) = \delta(\mathbf{z} - \mathbf{A}\boldsymbol{x})$ and we get $q(\boldsymbol{x},\mathbf{z}) = 1/Z_q \prod_i q_{x_i}(x_i) \prod_j q_{z_j}(z_j)\, \delta(\mathbf{z} - \mathbf{A}\boldsymbol{x})$. For the optimization of a factor $q_{x_i}(x_i)$, fit a Gaussian to the *tilted/target pdf*

$$\widetilde{p}_{x_i}(\boldsymbol{x},\mathbf{z}) = 1/Z_{\widetilde{p}_{x_i}}\, p_{x_i}(x_i) \prod_{k\neq i} q_{x_k}(x_k) \prod_j q_{z_j}(z_j) \delta(\mathbf{z} - \mathbf{A}\boldsymbol{x}) \tag{13}$$

We get:

$$\begin{aligned} &\mathrm{argmin}_{f(x_i)}\, KLD(\widetilde{p}_{x_i}, q) \\ &= \arg\min_{q_{x_i}(x_i)} KLD(p(x_k) \prod_{k\neq i} q_{x_k}(x_k) \prod_j q_{z_j}(z_j) \delta(\mathbf{z} - \mathbf{A}\boldsymbol{x}), \\ &\qquad \textstyle\prod_k q_{x_k}(x_k) \prod_j q_{z_j}(z_j) \delta(\mathbf{z} - \mathbf{A}\boldsymbol{x})) = \\ &\arg\min_{q_{x_i}(x_i)} \int p(x_i) \prod_{k\neq i} q_{x_k}(x_k) \prod_j p_{z_j}(z_j) \delta(\mathbf{z} - \mathbf{A}\boldsymbol{x}) \ln(\frac{p(x_i)}{p_{x_i}(x_i)}) d\boldsymbol{x} d\mathbf{z} \\ &= \arg\min_{q_{x_i}(x_i)} \int p(x_i) \ln(\frac{p(x_i)}{q_{x_i}(x_i)}) \\ &\quad \underbrace{[\int \prod_{k\neq i} q_{x_i}(x_k) \prod_j q_{z_j}(z_j) \delta(\mathbf{z} - \mathbf{A}\boldsymbol{x}) d\boldsymbol{x}_{\bar{k}} d\mathbf{z}]}_{= m_{x_i}(x_i) \text{ Gaussian extrinsic}} dx_i \\ &= \arg\min_{q_{x_i}(x_i)} \int p_{x_i}(x_i) m_{x_i}(x_i) \ln \frac{p_{x_i}(x_i) m_{x_i}(x_i)}{q_{x_i}(x_i) m_{x_i}(x_i)} dx_i \\ &= \arg\min_{q(x_i)/m_{x_i}(x_i)} \int p_{x_i}(x_i) m_{x_i}(x_i) \ln \frac{p_{x_i}(x_i) m_{x_i}(x_i)}{q(x_i)} dx_i \end{aligned} \tag{14}$$

Since $m_{x_i}(x_i)$ is Gaussian, it will suffice to fit a Gaussian in x_i, say $q(x_i)$, via

$$\begin{aligned}&\text{KLD}(p(x_i|\boldsymbol{y})||q(x_i)) = \text{KLD}(p_{x_i}(x_i)p(\boldsymbol{y}|x_i)/Z_i||q(x_i))\\ &\approx \text{KLD}(p_{x_i}(x_i)m_{x_i}(x_i)/Z_i||q(x_i))\\ &= \text{KLD}(p_{x_i}(x_i)m_{x_i}(x_i)/Z_i||q_{x_i}(x_i)m_{x_i}(x_i)/Z_i')\,.\end{aligned} \tag{15}$$

The reVAMP algorithm [8] approximates the posterior to Gaussian with the approximated Gaussian extrinsic:

$$p(x_i|\boldsymbol{y}) \approx \frac{p_{x_i}(x_i)m_{x_i}(x_i)}{Z_{x_i}(\boldsymbol{y})} \approx \mathcal{N}(x_i;\widehat{x}_i,\tau_{x_i}) = q(x_i)\,. \tag{16}$$

where $m_{x_i}(x_i) = \mathcal{N}(x_i; r_i, \tau_{r_i})$. The approximate Gaussian posterior $q(x_i)$ is obtained by moment matching with the better posterior approximation $p_{x_i}(x_i)\, m_{x_i}(x_i)/Z_{x_i}$.

We interpret the quotient of the approximated posterior and the approximate extrinsic as the approximated Gaussian prior.

$$\begin{aligned}&p_{x_i}(x_i) \approx q_{x_i}(x_i) = \mathcal{N}\left(x_i; m_{x_i}, \sigma_{x_i}^2\right) \propto \frac{\mathcal{N}(x_i;\widehat{x}_i,\tau_{x_i})}{\mathcal{N}(x_i;r_i,\tau_{r_i})},\\ &1/\sigma_{x_i}^2 = 1/\tau_{x_i} - 1/\tau_{r_i}\ ,\ m_{x_i} = \sigma_{x_i}^2(\widehat{x}_i/\tau_{x_i} - r_i/\tau_{r_i})\,.\end{aligned} \tag{17}$$

This Gaussian approximation $q_{x_i}(x_i)$ does not correspond to direct moment matching of the true prior $p_{x_i}(x_i)$. So, reGVAMP admits two points of view:

(1) minimize KLD(p, q) with $q = \prod_i q(x_i) \prod_j q(z_j)$

(2) do Minka EP with $q = \prod_i q_{x_i}(x_i) \prod_j q_{z_j}(z_j)\delta(\mathbf{z} - \mathbf{A}\boldsymbol{x})$

Both points of view lead to the same results!

The sense of the Gaussian prior approximations $q_{x_i}(x_i)$, $q_{z_j}(z_j)$ is that they are the equivalent Gaussian priors that, in the presence of the Gaussian extrinsics, produce the exact (nonlinear) MMSE estimate and variance that the original non-Gaussian prior would do! Direct Gaussian approximation of the priors is very suboptimal because that would only produce the correct LMMSE estimate and variance!!!

In the case the true priors are Gaussian, the two are the same of course.

Apart from the *improved marginal posteriors* $m_{x_i}(x_i)p_{x_i}(x_i)/Z_i'$ (and similar for the z_j), together with $\delta(\mathbf{z} - \mathbf{A}\boldsymbol{x})$, the Gaussian prior approximations $q_{x_i}(x_i)$, $q_{z_j}(z_j)$ in reGVAMP lead to an equivalent overall Gaussian linear model. This can be used for Large System Analysis (random $\mathbf{A}$ model) for the resulting posterior variances (MSEs), as obtained by GAMP. reVAMP does *alternating minimization of KLD*(p, q) which becomes iterative because an extrinsic $m_{x_i}(x_i)$ depends on the approximate Gaussian priors $\prod_{j\neq i} q_{x_j}(x_j)$, $\prod_j q_{z_j}(z_j)$. Since alternating minimization of a convex cost function converges, reVAMP can be expected to converge.

The Gaussian extrinsics approximations $p(x_i|\boldsymbol{y}) \approx m_i(x_i)$ are *asymptotically tight*. The Gaussian approximations that are not tight and that constitute the variational approximations are approximating marginal posteriors by Gaussian $q(x_i)$ or what follows from that, approximating priors $p_{x_i}(x_i)$ by Gaussian $q_{x_i}(x_i)$. Or the overall multivariate Gaussian posterior approximation is not tight also, but at least *captures full second-order moments*.

Hence in one point of view, re(G)VAMP minimized the desirable KLD(p, q), which becomes feasible thanks to asymptotic Gaussianity of the extrinsics like $p(\boldsymbol{y}|x_i) \approx m_{x_i}(x_i)$. However, re(G)VAMP can also be derived using EP, using a different formal posterior approximation.

IV. Relation to CWCU MMSE Estimator

The algorithm proposed by [8] can be interpreted as an iterative method of finding consistent extrinsic and posterior messages for the case of a AWGN $p(\boldsymbol{y}|\mathbf{z})$. [8] also shows the close relation between CWCU LMMSE estimation [10] and the extrinsic. In the following, we will interpret the extrinsic as CWCU LMMSE estimation based on the Gauss-Markov theorem.

Based on the discussion of the previous section, when deriving the extrinsic for $\mathbf{z}$ and $\boldsymbol{x}$, we find the system to be equivalent to a Gaussian linear model. Therefore, we can use the approximate prior and approximate likelihood as if they are the true prior and likelihood when deriving the extrinsics without large system approximations [11].

Consider jointly Gaussian $\boldsymbol{y}$ and x (scalar)

$$\begin{bmatrix} \boldsymbol{y} \\ x \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{m}_{\boldsymbol{y}} \\ m_x \end{bmatrix}, \begin{bmatrix} \mathbf{C}_{\boldsymbol{yy}} & \mathbf{C}_{\boldsymbol{y}x} \\ \mathbf{C}_{x\boldsymbol{y}} & C_{xx} \end{bmatrix}\right) \tag{18}$$

Then the extrinsic $p(\boldsymbol{y}|x)$ is Gaussian and based on Gauss-Markov theorem

$$\begin{aligned}&-2\ln p(\boldsymbol{y}|x) = c + (\boldsymbol{y} - \mathbf{m}_{\boldsymbol{y}|x})^T \mathbf{C}_{\boldsymbol{y}|x}^{-1}(\boldsymbol{y} - \mathbf{m}_{\boldsymbol{y}|x}),\ \text{with}\\ &\mathbf{m}_{\boldsymbol{y}|x} = \mathbf{m}_{\boldsymbol{y}} + \mathbf{C}_{\boldsymbol{y}x}C_{xx}^{-1}(\boldsymbol{x} - m_x),\\ &\mathbf{C}_{\boldsymbol{y}|x} = \mathbf{C}_{\boldsymbol{yy}} - \mathbf{C}_{\boldsymbol{y}x}\mathbf{C}_{xx}^{-1}\mathbf{C}_{x\boldsymbol{y}}\end{aligned} \tag{19}$$

Interpreting (19) as a pdf in x (which Fisher called fiducial statistics), we can rewrite this quadratic exponent as

$$\begin{aligned}&-2\ln p(\boldsymbol{y}|x) = c(\boldsymbol{y}) + (x - \widehat{x}_{CL})^2/\mathbf{C}_{\widetilde{x}_{CL}\widetilde{x}_{CL}},\\ &\widehat{x}_{CL} = m_x + d\,\mathbf{C}_{x\boldsymbol{y}}\mathbf{C}_{\boldsymbol{yy}}^{-1}(\boldsymbol{y} - \mathbf{m}_{\boldsymbol{y}}) = d\,\widehat{x}_L + (1-d)\,m_x\\ &C_{\widetilde{x}_{CL}\widetilde{x}_{CL}} = d\,C_{\widetilde{x}_L\widetilde{x}_L},\\ &\text{with}\\ &\widehat{x}_L = m_x + \mathbf{C}_{x\boldsymbol{y}}\mathbf{C}_{\boldsymbol{yy}}^{-1}(\boldsymbol{y} - \mathbf{m}_{\boldsymbol{y}}),\ C_{\widetilde{x}_L\widetilde{x}_L} = C_{xx} - \mathbf{C}_{x\boldsymbol{y}}\mathbf{C}_{\boldsymbol{yy}}^{-1}\mathbf{C}_{\boldsymbol{y}x}\\ &d = \frac{C_{xx}}{\mathbf{C}_{x\boldsymbol{y}}\mathbf{C}_{\boldsymbol{yy}}^{-1}\mathbf{C}_{\boldsymbol{y}x}} \geq 1,\end{aligned} \tag{20}$$

where $\widehat{x}_{CL}$, $\mathbf{C}_{\widetilde{x}_{CL}\widetilde{x}_{CL}}$ are the CWCU LMMSE estimate and error variance, and $\widehat{x}_L$, $\mathbf{C}_{\widetilde{x}_L\widetilde{x}_L}$ are the LMMSE (and hence MMSE since Gaussian) estimate and error variance.

Now we will investigate the vector case. Define the operation $Diag(\mathbf{C}) = \text{diag}[\text{diag}(\mathbf{C})]$, which returns a diagonal matrix from the vector diag$(\mathbf{C})$, composed of the diagonal elements of square matrix $\mathbf{C}$.

Interpreting the previous x as a component x_i of a vector $\boldsymbol{x}$, we can write

$$\begin{aligned}&\widehat{\boldsymbol{x}}_{CL} = \mathbf{m}_{\boldsymbol{x}} + \mathbf{D}\,\mathbf{C}_{\boldsymbol{xy}}\mathbf{C}_{\boldsymbol{yy}}^{-1}(\boldsymbol{y} - \mathbf{m}_{\boldsymbol{y}}) = \mathbf{D}\,\widehat{\boldsymbol{x}}_L + (\mathbf{I} - \mathbf{D})\,\mathbf{m}_{\boldsymbol{x}}\\ &\mathbf{C}_{\widetilde{\boldsymbol{x}}_{CL}\widetilde{\boldsymbol{x}}_{CL}} = \mathbf{C}_{\widetilde{\boldsymbol{x}}_L\widetilde{\boldsymbol{x}}_L} + (\mathbf{D} - \mathbf{I})\mathbf{C}_{\widehat{\boldsymbol{x}}_L\widehat{\boldsymbol{x}}_L}(\mathbf{D} - \mathbf{I})\\ &\text{with}\\ &\mathbf{D} = Diag(\mathbf{C}_{\boldsymbol{xx}})[Diag(\mathbf{C}_{\widehat{\boldsymbol{x}}_L\widehat{\boldsymbol{x}}_L})]^{-1},\ \mathbf{C}_{\widehat{\boldsymbol{x}}_L\widehat{\boldsymbol{x}}_L} = \mathbf{C}_{\boldsymbol{xy}}\mathbf{C}_{\boldsymbol{yy}}^{-1}\mathbf{C}_{\boldsymbol{yx}}\end{aligned} \tag{21}$$

where the expression for $\mathbf{C}_{\widetilde{\boldsymbol{x}}_{CL}\widetilde{\boldsymbol{x}}_{CL}}$ follows from $\widetilde{\boldsymbol{x}}_{CL} = \boldsymbol{x} - \widehat{\boldsymbol{x}}_{CL} = \widetilde{\boldsymbol{x}}_L - (\mathbf{D} - \mathbf{I})\,\mathbf{C}_{\boldsymbol{xy}}\mathbf{C}_{\boldsymbol{yy}}^{-1}(\boldsymbol{y} - \mathbf{m}_{\boldsymbol{y}})$ and the two terms in

this difference are decorrelated by the orthogonality property of LMMSE estimation.

Next, we'll show: $\mathbf{D} = \text{diag}(\boldsymbol{\tau}_{CL}./\boldsymbol{\tau}_L)$, where $\boldsymbol{\tau}_L = \text{diag}(\mathbf{C}_{\widetilde{\boldsymbol{x}}_L\widetilde{\boldsymbol{x}}_L})$ and $\boldsymbol{\tau}_{CL} = \text{diag}(\mathbf{C}_{\widetilde{\boldsymbol{x}}_{CL}\widetilde{\boldsymbol{x}}_{CL}})$, and "./" denotes element-wise division.

$$\begin{aligned}\mathbf{C}_{\widetilde{\boldsymbol{x}}_{CL}\widetilde{\boldsymbol{x}}_{CL}} &= \mathbf{C}_{\widetilde{\boldsymbol{x}}_L\widetilde{\boldsymbol{x}}_L} + (\mathbf{D}-\mathbf{I})\mathbf{C}_{\widehat{\boldsymbol{x}}_L\widehat{\boldsymbol{x}}_L}(\mathbf{D}-\mathbf{I}) \\ &= \mathbf{C}_{\boldsymbol{xx}} - \mathbf{C}_{\widehat{\boldsymbol{x}}_L\widehat{\boldsymbol{x}}_L}\mathbf{D} - \mathbf{D}\,\mathbf{C}_{\widehat{\boldsymbol{x}}_L\widehat{\boldsymbol{x}}_L} + \mathbf{D}\,\mathbf{C}_{\widehat{\boldsymbol{x}}_L\widehat{\boldsymbol{x}}_L}\mathbf{D}\end{aligned} \tag{22}$$

Calculate the diagonal elements

$$\begin{aligned}&\text{diag}(\boldsymbol{\tau}_{CL}) = Diag(\mathbf{C}_{\widetilde{\boldsymbol{x}}_{CL}\widetilde{\boldsymbol{x}}_{CL}}) = Diag(\mathbf{C}_{\boldsymbol{xx}}) \\ &+\mathbf{D}\,Diag(\mathbf{C}_{\widehat{\boldsymbol{x}}_L\widehat{\boldsymbol{x}}_L})\mathbf{D} - Diag(\mathbf{C}_{\widehat{\boldsymbol{x}}_L\widehat{\boldsymbol{x}}_L})\mathbf{D} - \mathbf{D}\,Diag(\mathbf{C}_{\widehat{\boldsymbol{x}}_L\widehat{\boldsymbol{x}}_L}) \\ &= Diag(\mathbf{C}_{\boldsymbol{xx}})[Diag(\mathbf{C}_{\widehat{\boldsymbol{x}}_L\widehat{\boldsymbol{x}}_L})]^{-1}Diag(\mathbf{C}_{\boldsymbol{xx}}) - Diag(\mathbf{C}_{\boldsymbol{xx}}),\end{aligned} \tag{23}$$

where we use $\mathbf{D} = Diag(\mathbf{C}_{\boldsymbol{xx}})\,[Diag(\mathbf{C}_{\widehat{\boldsymbol{x}}_L\widehat{\boldsymbol{x}}_L})]^{-1}$ in (21).

Now we want to show $\mathbf{D}\,\text{diag}(\boldsymbol{\tau}_L) = \text{diag}(\boldsymbol{\tau}_{CL})$:

$$\begin{aligned}&\mathbf{D}\,\text{diag}(\boldsymbol{\tau}_L) = \mathbf{D}\,Diag(\mathbf{C}_{\widetilde{\boldsymbol{x}}_L\widetilde{\boldsymbol{x}}_L}) \\ &= Diag(\mathbf{C}_{\boldsymbol{xx}})[Diag(\mathbf{C}_{\widehat{\boldsymbol{x}}_L\widehat{\boldsymbol{x}}_L})]^{-1}\cdot \\ &\cdot[Diag(\mathbf{C}_{\boldsymbol{xx}}) - Diag(\mathbf{C}_{\widehat{\boldsymbol{x}}_L\widehat{\boldsymbol{x}}_L})] = \text{diag}(\boldsymbol{\tau}_{CL})\end{aligned} \tag{24}$$

The extrinsic for $\boldsymbol{x}$ without large system approximations can be interpreted as CWCU MMSE estimation from the Gaussian model

$$\begin{bmatrix}\mathbf{m}_{\mathbf{z}} \\ \boldsymbol{x}\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}\mathbf{A}\mathbf{m}_{\boldsymbol{x}} \\ \mathbf{m}_{\boldsymbol{x}}\end{bmatrix}, \begin{bmatrix}\mathbf{A}\mathbf{D}_{\boldsymbol{\sigma}^2_{\boldsymbol{x}}}\mathbf{A}^T + \mathbf{D}_{\boldsymbol{\sigma}^2_{\mathbf{z}}} & \mathbf{A}\mathbf{D}_{\boldsymbol{\sigma}^2_{\boldsymbol{x}}} \\ \mathbf{D}_{\boldsymbol{\sigma}^2_{\boldsymbol{x}}}\mathbf{A}^T & \mathbf{D}_{\boldsymbol{\sigma}^2_{\boldsymbol{x}}}\end{bmatrix}\right). \tag{25}$$

The underlying equivalent Gaussian linear model is

$$\mathbf{m}_{\mathbf{z}} = \mathbf{A}\boldsymbol{x} + \boldsymbol{v}_{\boldsymbol{x}} \tag{26}$$

where $\boldsymbol{x} \sim \mathcal{N}(\mathbf{m}_{\boldsymbol{x}}, \mathbf{D}_{\boldsymbol{\sigma}^2_{\boldsymbol{x}}})$ and $\boldsymbol{v}_{\boldsymbol{x}} \sim \mathcal{N}(\mathbf{0}, \mathbf{D}_{\boldsymbol{\sigma}^2_{\mathbf{z}}})$.

Likewise, we can interpret the extrinsic for $\mathbf{z}$ as CWCU MMSE estimation from

$$\begin{bmatrix}\mathbf{A}\mathbf{m}_{\boldsymbol{x}} \\ \mathbf{z}\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}\mathbf{m}_{\mathbf{z}} \\ \mathbf{m}_{\mathbf{z}}\end{bmatrix}, \begin{bmatrix}\mathbf{D}_{\boldsymbol{\sigma}^2_{\mathbf{z}}} + \mathbf{A}\mathbf{D}_{\boldsymbol{\sigma}^2_{\boldsymbol{x}}}\mathbf{A}^T & \mathbf{D}_{\boldsymbol{\sigma}^2_{\mathbf{z}}} \\ \mathbf{D}_{\boldsymbol{\sigma}^2_{\mathbf{z}}} & \mathbf{D}_{\boldsymbol{\sigma}^2_{\mathbf{z}}}\end{bmatrix}\right). \tag{27}$$

The underlying equivalent Gaussian linear model is

$$\mathbf{A}\mathbf{m}_{\boldsymbol{x}} = \mathbf{z} + \boldsymbol{v}_{\mathbf{z}} \tag{28}$$

where $\mathbf{z} \sim \mathcal{N}(\mathbf{m}_{\mathbf{z}}, \mathbf{D}_{\boldsymbol{\sigma}^2_{\mathbf{z}}})$ and $\boldsymbol{v}_{\mathbf{z}} \sim \mathcal{N}(\mathbf{0}, \mathbf{A}\mathbf{D}_{\boldsymbol{\sigma}^2_{\boldsymbol{x}}}\mathbf{A}^T)$.

V. GAMP from LSL Belief Propagation

In reGVAMP, extrinsics in the GLM are built from the equivalent Gaussian linear model, which introduces equivalent Gaussian priors from Gaussian posterior approximations and Gaussian extrinsics.

GAMP exploits LSL simplifications of reGVAMP for a random $\mathbf{A}$ with i.i.d. signs which leads to

(i) Gaussianity of extrinsics (also in reGVAMP), and

(ii) independence of marginals (extra w.r.t. reGVAMP).

(ii) leads to the large system simplifications of the variances, avoiding covariance matrix inverses. But also posterior and extrinsic estimates $\widehat{\boldsymbol{x}}$, $\widehat{\mathbf{z}}$ and $\mathbf{r}$, $\boldsymbol{p}$ that are constructed by combining decoupled pieces of information. These estimates are non-linear MMSE and CWCU MMSE estimates in general. Extrinsics are not obtained as linear perturbations of corresponding MMSE estimates because those are not necessarily close to each other. Rather the interplay between $\boldsymbol{x}$ and $\mathbf{z}$ is exploited with perturbations due to the small

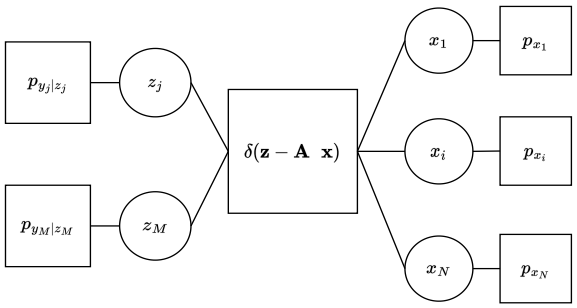

Fig. 1. Factor Graph for the GLM used by reGVAMP. Circles: variable nodes, squares: factor nodes.

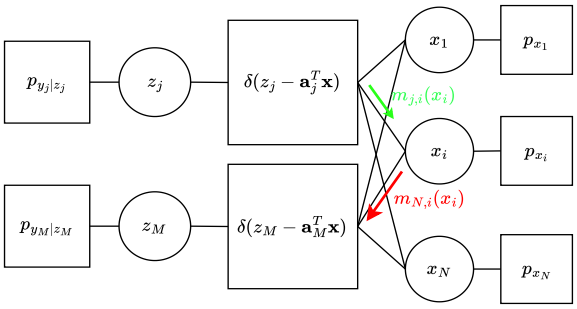

Fig. 2. Factor Graph for the GLM used by GVAMP.

effect of a single term in $\mathbf{A}$ in the LSL. In both reGVAMP and GAMP, we have:

Gaussian extrinsics: $\mathcal{N}(\boldsymbol{x};\mathbf{r},\boldsymbol{\tau}_r)$, $\mathcal{N}(\mathbf{z};\boldsymbol{p},\boldsymbol{\tau}_p)$, and

Posterior marginals proportional to: $p_{\boldsymbol{x}}(\boldsymbol{x})\,\mathcal{N}(\boldsymbol{x};\mathbf{r},\boldsymbol{\tau}_r)$, $p_{\boldsymbol{y}|\mathbf{z}}(\boldsymbol{y}|\mathbf{z})\,\mathcal{N}(\mathbf{z};\boldsymbol{p},\boldsymbol{\tau}_p)$ with Gaussian approximations $\mathcal{N}(\boldsymbol{x};\widehat{\boldsymbol{x}},\boldsymbol{\tau}_x)$, $\mathcal{N}(\mathbf{z};\widehat{\mathbf{z}},\boldsymbol{\tau}_z)$.

reGVAMP considers the joint pdf factorization into $M+N+1$ factors

$$p(\boldsymbol{x},\mathbf{z},\boldsymbol{y}) = \delta(\mathbf{z}-\mathbf{A}\boldsymbol{x})\prod_{i=1}^{N}p_{x_i}(x_i)\prod_{k=1}^{M}p_{y_k|z_k}(y_k|z_k) \tag{29}$$

where $\delta(\mathbf{z}-\mathbf{A}\boldsymbol{x}) = \prod_{k=1}^{M}\delta(z_k-\boldsymbol{a}_k^T\boldsymbol{x})$, $\mathbf{A}^T = [\boldsymbol{a}_1\cdots\boldsymbol{a}_M]$. The factor graph in Fig. 1 is without cycles. The factor graph considered determines the associated Belief or Expectation Propagation algorithms for minimizing the Bethe Free Energy [5]. GVAMP on the other hand considers the following joint pdf factorization into $2M+N$ factors

$$p(\boldsymbol{x},\mathbf{z},\boldsymbol{y}) = \prod_{i=1}^{N}p_{x_i}(x_i)\prod_{k=1}^{M}p_{y_k|z_k}(y_k|z_k)\,\delta(z_k-\boldsymbol{a}_k^T\boldsymbol{x}) \tag{30}$$

which leads to the factor graph in Fig. 2 which does contain cycles.

Message passing in the GLM scalar level factor graph of Fig. 2 alternates between the following message updates:

$$\begin{aligned}m_{k,n}(x_n) &\sim \int p(y_k|z_k)\,\delta(z_k-\mathbf{A}_{k,:}\boldsymbol{x})\prod_{m\neq n}m_{m,k}(x_m)\,dz_k d\boldsymbol{x}_{\overline{n}} \\ m_{n,k}(x_n) &\sim p_{x_n}(x_n)\prod_{i\neq k}m_{i,n}(x_n)\end{aligned} \tag{31}$$

where $\sim$ denotes equality up to a normalization factor. This results in:
marginal posteriors: $m_n(x_n) \sim p_{x_n}(x_n) \prod_i m_{i,n}(x_n)$,
extrinsic $z_k : \sim \int \delta(z_k - \mathbf{A}_{k,:}\boldsymbol{x}) \prod_n m_{n,k}(x_n)\, d\boldsymbol{x}$,
extrinsic $x_n : \sim \prod_i m_{i,n}(x_n)$.
Like reGVAMP, GAMP uses Gaussian approximations for extrinsics. This requires Gaussian models for the messages. GAMP applies Gaussian approximations in 2 steps: (middle expression = prior $\times$ Gaussian extrinsic)

$$m_{k,n}(x_n) \to \int p(y_k|z_k)\, \delta(z_k - \mathbf{A}_{k,:}\boldsymbol{x}) \prod_{m\neq n} q_{m,k}(x_m)\, dz_k d\boldsymbol{x}_{\overline{n}}$$
$$\to q_{k,n}(x_n) = \mathcal{N}(x_n; \widehat{x}_{k,n}, \tau^x_{k,n}) \quad (32)$$
$$m_{n,k}(x_n) \to p_{x_n}(x_n) \prod_{i\neq k} q_{i,n}(x_n)$$
$$\to q_{n,k}(x_n) = \mathcal{N}(x_n; \widehat{x}_{n,k}, \tau^x_{n,k}) \quad (33)$$

A. Output Node

We get for the incomplete extrinsic for z_k:
$\int \delta(z_k - \mathbf{A}_{k,:}\boldsymbol{x}) \prod_{m\neq n} q_{m,k}(x_m)\, d\boldsymbol{x}_{\overline{n}} \sim \mathcal{N}(z_k; p_{k,n} + \mathbf{A}_{k,n}x_n, \tau^p_{k,n})$
with $p_{k,n} = \mathbf{A}_{k,\overline{n}}\, \widehat{\boldsymbol{x}}_{\overline{n},k}, \tau^p_{k,n} = \mathbf{S}_{k,\overline{n}}\, \boldsymbol{\tau}^x_{\overline{n},k} \approx \mathbf{S}_{k,\overline{n}}\, \boldsymbol{\tau}^x_{\overline{n}}$
Define $p_k = \mathbf{A}_{k,:}\widehat{\boldsymbol{x}}_{:,k} \Rightarrow p_{k,n} = p_k - \mathbf{A}_{k,n}\widehat{x}_{n,k}$.
And $\tau^p_{k,n} = \tau^p_k - \mathbf{S}_{k,n}\tau^x_{n,k}$ where $\tau^p_k = \mathbf{S}_{k,:}\boldsymbol{\tau}_x$.
Neglecting terms of order $\mathbf{S}_{k,n}$, we get $\mathcal{N}(z_k; p_{k,n} + \mathbf{A}_{k,n}x_n, \tau^p_{k,n}) \approx \mathcal{N}(z_k; p_k + \mathbf{A}_{k,n}\widetilde{x}_n, \tau^p_k)$ with $\widetilde{x}_n = x_n - \widehat{x}_n$.
Then $m_{k,n}(x_n) \overset{\propto}{\sim} Z_z(p_k + \mathbf{A}_{k,n}\widetilde{x}_n, y_k, \tau^p_k)$ with

$$Z_z(p, y, \tau_p) = \int p_{y|z}(y|z)\, e^{-\frac{1}{2\tau_p}(z-p)^2} dz$$
$$\frac{\partial \ln Z_z}{\partial p} = \frac{Z'_z}{Z_z} = s = \frac{\widehat{z}-p}{\tau_p}, \widehat{z} = \frac{1}{Z_z}\int z\, p_{y|z}(y|z)\, e^{-\frac{1}{2\tau_p}(z-p)^2} dz$$
$$\frac{\partial^2 \ln Z_z}{\partial p^2} = -\tau_s = -\tau_s = \frac{Z''_z}{Z_z} - (\frac{Z'_z}{Z_z})^2 = -(1 - \tau_z/\tau_p)/\tau_p$$

Then up to second order in $\mathbf{A}_{k,n}\widetilde{x}_n$ (Laplacian approximation in MAP case, Gaussian moment matching in MMSE case), a single measurement extrinsic for x_n becomes: $\ln m_{k,n}(x_n)$

$$\approx \ln Z_z(p_k, y_k, \tau^p_k) + \frac{\partial \ln Z_z}{\partial p}\, \mathbf{A}_{k,n}\widetilde{x}_n + \frac{\partial^2 \ln Z_z}{2\,\partial p^2}\, \mathbf{A}^2_{k,n}\widetilde{x}^2_n$$
$$= c^t + [s_k\mathbf{A}_{k,n} + \mathbf{A}^2_{k,n}\tau^s_k\, \widehat{x}_n]\, x_n - \tfrac{1}{2}\, \tau^s_k\, \mathbf{A}^2_{k,n}\, x^2_n\,.$$

Now $\ln m_{n,k}(x_n) = c^t + \ln p_{x_n}(x_n) + \sum_{i\neq k} \ln m_{i,n}(x_n)$

$$= c^t + \ln p_{x_n}(x_n) - \tfrac{1}{2\tau^r_{n,k}}(x_n - r_{n,k})^2$$
with $\frac{1}{\tau^r_{n,k}} = \mathbf{S}^T_{k,n}\boldsymbol{\tau}^s_k \quad (\approx \mathbf{S}^T_{:,n}\boldsymbol{\tau}_s = \frac{1}{\tau^r_n})$
and $r_{n,k} = \tau^r_{n,k}(\mathbf{s}^T_{\overline{k}}\mathbf{A}_{\overline{k},n} + \mathbf{S}^T_{\overline{k},n}\boldsymbol{\tau}^s_{\overline{k}}\, \widehat{x}_n) = \widehat{x}_n + \tau^r_{n,k}\, \mathbf{s}^T_{\overline{k}}\mathbf{A}_{\overline{k},n}\,.$

B. Input Node

We now get for the approximate posterior
$m_n(x_n) = \frac{1}{Z_x(r_n,\tau^r_n)} p_{x_n}(x_n)\, e^{-\frac{1}{\tau^r_n}(\frac{x^2_n}{2} - x_n r_n)}$ with

$$Z_x(r, \tau_r) = \int p_x(x)\, e^{-\frac{1}{\tau_r}(\frac{x^2}{2} - x\, r)} dx$$
$$\tau_r\, \frac{\partial \ln Z_x}{\partial r} = \int x\, m(x) dx = \mathbb{E}(x|r, \tau_r) = \widehat{x} = \widehat{x}(r, \tau_r)$$
$$\tau^2_r\, \frac{\partial^2 \ln Z_x}{\partial r^2} = \tau_r\, \frac{\partial \widehat{x}}{\partial r} = \tau_x$$

Now, with $r_n = \widehat{x}_n + \tau^r_n\, \mathbf{s}^T\mathbf{A}_{:,n}$, we can write
$r_{n,k} \approx \widehat{x}_n + \tau^r_n\, \mathbf{s}^T_{\overline{k}}\mathbf{A}_{\overline{k},n} = r_n - \tau^r_n\, s_k\, \mathbf{A}_{k,n}$. We get similarly for the mean $\widehat{x}_{n,k}$ of $m_{n,k}(x_n)$:

$$\widehat{x}_{n,k} = \widehat{x}_n(r_{n,k}, \tau^r_n) = \widehat{x}_n(r_n - \tau^r_n\, s_k\, \mathbf{A}_{k,n}, \tau^r_n)$$
$$\approx \widehat{x}_n(r_n, \tau^r_n) - \tfrac{\partial}{\partial r_n}\, \widehat{x}_n(r_n, \tau^r_n)\, \tau^r_n\, s_k\, \mathbf{A}_{k,n} = \widehat{x}_n - \tau^x_n\, s_k\, \mathbf{A}_{k,n}$$

Plugging this in, we get

$$p_k = \mathbf{A}_{k,:}\, \widehat{\boldsymbol{x}}_{:,k} = \mathbf{A}_{k,:}\, \widehat{\boldsymbol{x}} - \mathbf{S}_{k,:}\boldsymbol{\tau}_x\, s_k = \mathbf{A}_{k,:}\, \widehat{\boldsymbol{x}} - \tau^p_k\, s_k$$

which completes the message passing. We may note that the variance derivations in the LSL of BP are equivalent to the large random matrix analysis of the MSE of LMMSE in the equivalent Gaussian linear model.

VI. Concluding Remarks

We rederived the reGVAMP algorithm from the point of view of alternating minimization of a LSL version of a desirable KLD. The asymptotics here involve only the CLT for extrinsics. We then derive the GAMP algorithm by directly introducing LSL simplifications in the LBP algorithm. This leads us to relate extrinsic messages to posterior pdfs by first order Taylor series expansion based perturbations. We also apply LSL approximations to the variances of the various Gaussians involved, which in fact leads to a rederivation of a fundamental LSA theorem describing the deterministic limit of LMMSE posterior variances.

VII. 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, 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] T. Heskes, M. Opper, W. Wiegerinck, O. Winther, and O. Zoeter, "Approximate Inference Techniques with Expectation Constraints," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2005, no. 11, 2005.
[6] Q. Zou and H. Yang, "A Concise Tutorial on Approximate Message Passing," *arXiv preprint arXiv:2201.07487*, 2022.
[7] 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.
[8] Z. Zhao, F. Xiao, and D. Slock, "Approximate Message Passing for Not So Large niid Generalized Linear Models," in *Proc. Int'l Workshop Signal Processing Advances in Wireless Comm's (SPAWC)*, Sept. 2023.
[9] ——, "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.
[10] M. Triki and D. T. 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.
[11] M. Huemer and O. Lang, "CWCU LMMSE Estimation: Prerequisites and Properties," *arXiv preprint arXiv:1412.1567*, 2014.