ALTERNATING CONSTRAINED MINIMIZATION BASED APPROXIMATE MESSAGE PASSING

Christo Kurisummoottil Thomas[1] *and Dirk Slock*[2]

[1]Wireless@VT, Bradley Department of Electrical and Computer Engineering,
Virginia Tech, Arlington, VA, USA,
[2]Communication Systems Department, EURECOM, France
Emails: christokt@vt.edu, Dirk.Slock@eurecom.fr

ABSTRACT

Generalized Approximate Message Passing (GAMP) allows for Bayesian inference in linear models with non-identically independently distributed (n.i.i.d.) priors and n.i.i.d. measurements of the linear mixture outputs. It represents an efficient technique for approximate inference, which becomes accurate when both rows and columns of the measurement matrix can be treated as sets of independent vectors and both dimensions become large. It has been shown that the fixed points of GAMP correspond to the extrema of a large system limit of the Bethe Free Energy (LSL-BFE), which represents a meaningful approximation optimization criterion regardless of whether the measurement matrix exhibits the independence properties. However, the convergence of (G)AMP can be notoriously problematic for certain measurement matrices, and the only sure fixes so far are damping (by a difficult-to-determine amount) or perform a double ADMM. In this paper, we revisit the GAMP algorithm (as e.g. for sparse Bayesian learning (SBL)) by more rigorously applying an alternating constrained minimization strategy to an appropriately reparameterized LSL BFE. This guarantees convergence, at least to a local optimum. We furthermore introduce a natural extension of the BFE to integrate the estimation of (the SBL) hyperparameters via Variational Bayes, leading to Variational AMBGAMP or VAMBGAMP.

1. INTRODUCTION

In the Gaussian noise case, a sparse signal vector x can be recovered using the signal model: $\boldsymbol{y} = \mathbf{A}\, x + \boldsymbol{v}$, where $\boldsymbol{y}$ is the observed data, $\mathbf{A}$ is the known measurement or sensing matrix of dimension $M \times N$, typically with $M < N$. In the sparse model, x contains only K nonzero entries, with $K < M < N$. Sparse Bayesian Learning (SBL) is a Bayesian inference algorithm proposed by [1] and [2]. SBL is based on a hierarchical prior on the sparse coefficients x, inducing sparsity by choosing priors for the hyperparameters that make most coefficients tend towards zero. In [3], the authors provide an overview of various sparse signal recovery algorithms, including Basis Pursuit, LASSO, and SBL. They demonstrate the superior recovery performance of SBL compared to conventional methods. However, the Linear Minimum Mean Squared Error (LMMSE) estimation step in SBL at each iteration involves matrix inversion, which makes it computationally complex, even for moderately large datasets. This complexity motivates the use of approximate inference methods.

Belief Propagation (BP)-based Sparse Bayesian Learning (SBL) algorithms, as discussed in [4], offer improved computational efficiency. For a more detailed overview of various approximate inference methods for SBL, refer to [5]. While BP has demonstrated empirical success [6], more rigorous work is needed to characterize its convergence behavior in loopy networks. Several studies have investigated the convergence analysis of Gaussian BP (GaBP) and can be found in [7–10]. [11] provides a convergence condition for GaBP, which requires the underlying distribution to be walk-summable. Their convergence analysis is based on the Gaussian Markov random field (GMRF) decomposition, which expresses the underlying distribution in terms of pairwise connections between variables.

The Approximate Message Passing (AMP) algorithm has been introduced to reduce the complexity of GaBP further, from $2MN$ to $M+N$ messages. Generalized AMP (GAMP) allows non-Gaussian priors and more general measurement processes. But convergence of (G)AMP can be problematic for some matrices $\mathbf{A}$. Many variations have been introduced to help (G)AMP converge, such as 1) adding the Alternating Direction Method of Multipliers (ADMM), 2) exploiting part of the singular value decomposition of the measurement matrix in Vector AMP (VAMP) (but which does not allow to handle n.i.i.d. priors conveniently), 3) sequential updating in Swept AMP (SwAMP), which can work in most cases, 4) by introducing damping, but with typically difficult to determine damping requirements.

The AMP algorithm and its variations have many potential applications in (machine learning-aided) wireless communications systems:

- multi-user detection [12],
- channel estimation [13],
- joint detection and channel estimation [14],
- compressive sensing [15],
- reduced complexity Linear Minimum Mean Squared Error (LMMSE) receiver or transmitter computation [13].

1.1. Contributions of this paper

- We propose the AMBGAMP version of GAMP that guarantees convergence by utilizing an alternating constrained minimization strategy applied to an augmented Lagrangian of the constrained Large System Limit (LSL) of the Bethe Free Energy (BFE), inspired by but different from [16], [17]. The algorithm differs from [18], [19] and involves a gradient update with line search, fixed-point iterations, and an ADMM-style Lagrange multiplier update.
- The Kullback-Leibler divergences (KLDs) appearing in the LSL-BFE actually are exactly of the form required for applying Variational Bayesian (VB) posterior updates, not only for the main primal variables x and z but also for hyperparameters involved in their priors. We consider here the precisions (inverse variances) in Gaussian priors as hyperparameters, for the case of SBL. The resulting algorithm that involves optimization of the posteriors for x and the auxiliary variables (including the hyperparameters) is called variational AMBGAMP-SBL (VAMBGAMP-SBL).
- We also indicate that asymptotically, under an i.i.d. element model for $\mathbf{A}$, the variance computations in AMBGAMP are exact. This allows to analyze the steady-state MSE as a function of system dimensions and prior pdfs $p(x)$, $p(\boldsymbol{y}|z)$ for x and z. In particular in the Gaussian case, this allows to analyze the performance for SBL.
- Gaussian simulation results validate the convergence to the LMMSE solution for different measurement matrices, including i.i.d. and low-rank cases.

2. SYSTEM MODEL

The data model considered in (EM-GAMP-)SBL is essentially a linear mixing model represented by

$$\boldsymbol{z} = \mathbf{A}\,\boldsymbol{x}\,,\ \boldsymbol{p_{x,\alpha}(x,\alpha)}\,,\ \boldsymbol{p_{y,\gamma|z}(y,\gamma|z)} \qquad (1)$$

with (possibly non) identically independently distributed (n.i.i.d.) prior $p_{x,\boldsymbol{\alpha}}(x,\boldsymbol{\alpha}) = \prod_{i=1}^{N} p_{x_i|\alpha_i}(x_i|\alpha_i)\,p_{\alpha_i}(\alpha_i)$ and n.i.i.d. measurements $p_{\boldsymbol{y},\gamma|z}(\boldsymbol{y},\gamma|z) = \prod_{k=1}^{M} p_{y_k|z_k}(y_k|z_k,\gamma_k)\,p_{\gamma_k}(\gamma_k)$. The measurement noise precision vector (inverse variance hyperparameter) is $\boldsymbol{\gamma} = [\gamma_1,\cdots,\gamma_M]^T$. In SBL, we introduce a Gaussian prior with unknown precisions, denoted as α_i, collected in the vector $\boldsymbol{\alpha} = [\alpha_1,\cdots\alpha_N]^T$. Hence, we write

$$p_{x_i,\alpha_i}(x_i,\alpha_i) = p_{x_i|\alpha_i}(x_i|\alpha_i)p_{\alpha_i}(\alpha_i) = \mathcal{N}(x_i;0,\alpha_i^{-1})\mathcal{G}(\alpha_i;a,b) \qquad (2)$$

where the Gamma distribution for α_i, $\mathcal{G}(\alpha_i;a,b) = \frac{\alpha_i^{a-1}e^{-b\alpha_i}b^{\alpha_i}}{\Gamma(\alpha_i)}$, with known a, b. Note that the case of an uninformative prior corresponds to $a = 1$, $b = 0$. Each γ_i is also assumed to have a Gamma prior distribution, $p_{\gamma_k}(\gamma_k) = \mathcal{G}(\gamma_k;c,d)$, where also c, d are known. In Bayesian estimation, we are interested in the posterior, which is given by

$$\boldsymbol{p_{x,z|y}(x,z,\alpha,\gamma|y)} = \frac{e^{-\sum\limits_{i=1}^{N} f(x_i,\alpha_i) - \sum\limits_{k=1}^{M} f(z_k,\gamma_k)}}{Z(\boldsymbol{y})}\,\mathbb{1}_{\{\boldsymbol{z}=\mathbf{A}\boldsymbol{x}\}}, \qquad (3)$$

where we define the negative log-likelihoods as $f_{x_i,\alpha_i}(x_i,\alpha_i) = -\ln p_{x_i,\alpha_i}(x_i,\alpha_i)$, $f_{z_k,\gamma_k}(z_k,\gamma_k) = -\ln p_{y_k,\gamma_k|z_k}(y_k,\gamma_k|z_k)$, where we omit the dependence of $f_{z_k,\gamma_k}(z_k,\gamma_k)$ on $\boldsymbol{y}$. The problem in Bayesian estimation is the computation of the normalization constant $Z(\boldsymbol{y})$ and of the posterior means and variances. Belief propagation is a message passing technique that allows to compute the posterior marginals. However, due to loops in the factor graph, loopy belief propagation may have convergence issues and is furthermore still relatively complex. GAMP is an approximate belief propagation technique which is motivated by asymptotic considerations in which the rows and columns of the measurement matrix $\mathbf{A}$ are considered as random and independent. In which case GAMP can actually produce the correct posterior marginals. In any case, GAMP computes a separable approximate posterior of the form

$$\begin{aligned}\boldsymbol{q_{x,\alpha,z,\gamma}(x,\alpha,z,\gamma)} &= \boldsymbol{q_x(x)\,q_\alpha(\alpha)\,q_z(z)q_\gamma(\gamma)}\\ &= \textstyle\prod_{i=1}^{N} q_{x_i}(x_i)q_{\alpha_i}(\alpha_i)\ \prod_{k=1}^{M} q_{z_k}(z_k)q_{\gamma_k}(\gamma_k),\end{aligned} \qquad (4)$$

in which the dependence on $\boldsymbol{y}$ has been omitted. The GAMP algorithm [20], [16] appears in the table as Algorithm 1. We only consider here Sum-Product GAMP (for MMSE estimation, as opposed to Max-Sum GAMP for MAP estimation). We emphasize that, apart from the hyperparameter updates, the algorithms specified in the tables in this paper, hold for arbitrary Generalized Linear Models. Only the indicated hyperparameter updates are specific for SBL. Algorithm 1 reflects the standard Expectation Maximization (EM) approach for hyperparameter estimation.

Algorithm 1 EM-GAMP-SBL

```
Require: $\boldsymbol{y}, \mathbf{A}, \mathbf{S} = \mathbf{A.A}, f_{x,\boldsymbol{\alpha}}(x,\boldsymbol{\alpha}), f_{z,\boldsymbol{\gamma}}(z,\boldsymbol{\gamma})$
Initialize: $t = 0, \widehat{x}^t, \boldsymbol{\tau}_x^t, \mathbf{s}^{t-1} = \mathbf{0}, \widehat{\boldsymbol{\alpha}}^{t-1}, \widehat{\boldsymbol{\gamma}}^{t-1}$
repeat
    [Output node update]
    $\boldsymbol{\tau}_p^t = \mathbf{S}\,\boldsymbol{\tau}_x^t$
    $\boldsymbol{p}^t = \mathbf{A}\,\widehat{x}^t - \mathbf{s}^{t-1}.\boldsymbol{\tau}_p^t$
    $\widehat{\boldsymbol{z}}^t = \mathbb{E}(\boldsymbol{z}|\boldsymbol{p}^t, \boldsymbol{\tau}_p^t, \widehat{\boldsymbol{\gamma}}^{t-1})$
    $\boldsymbol{\tau}_z^t = \mathrm{var}(\boldsymbol{z}|\boldsymbol{p}^t, \boldsymbol{\tau}_p^t, \widehat{\boldsymbol{\gamma}}^{t-1})$
    $\mathbf{s}^t = (\widehat{z}^t - \boldsymbol{p}^t)./\boldsymbol{\tau}_p^t$
    $\boldsymbol{\tau}_s^t = (\mathbf{1} - \boldsymbol{\tau}_z^t./\boldsymbol{\tau}_p^t)./\boldsymbol{\tau}_p^t$
    [Input node update]
    $\boldsymbol{\tau}_r^t = \mathbf{1}./(\mathbf{S}^T\boldsymbol{\tau}_s^t)$
    $\mathbf{r}^t = \widehat{x}^t + \boldsymbol{\tau}_r^t.\mathbf{A}^T\mathbf{s}^t$
    $\widehat{\boldsymbol{x}}^{t+1} = \mathbb{E}(\boldsymbol{x}|\mathbf{r}^t, \boldsymbol{\tau}_r^t, \widehat{\boldsymbol{\alpha}}^{t-1})$
    $\boldsymbol{\tau}_x^{t+1} = \mathrm{var}(\boldsymbol{x}|\mathbf{r}^t, \boldsymbol{\tau}_r^t, \widehat{\boldsymbol{\alpha}}^{t-1})$
    [Hyperparameters update]
    $\widehat{\alpha}_i^t = \frac{2a+1}{\mathbb{E}(x_i^2)+2b}, \forall i.$
    $\widehat{\gamma}_k^t = \frac{2c+1}{\mathbb{E}(|y_k-z_k|^2)+2d}, \forall k.$
until Convergence
```

3. PROPOSED VAMBGAMP-SBL

The abbreviation AMB stands for ACM-LSL-BFE, which denotes Alternating Constrained Minimization of the LSL of the BFE. AMBGAMP employs most of the same updates as GAMP, but GAMP does not rigorously perform alternating minimization (block coordinate descent), particularly in the presence of constraints. Previous work [21] has demonstrated that any fixed point of the GAMP algorithm is a critical point of the following constrained minimization of a LSL of the BFE (see also [16] and references therein), here extended to include the hyperparameters:

$$\begin{aligned}&\min_{q_x,q_z,\boldsymbol{\tau}_p,q_\alpha,q_\gamma} J_{LSL-BFE}(q_x,q_z,\boldsymbol{\tau}_p,q_{\boldsymbol{\alpha}},q_{\boldsymbol{\gamma}})\\ &\quad s.t.\ \mathbb{E}(\boldsymbol{z}|\boldsymbol{q_z}) = \mathbf{A}\ \mathbb{E}(\boldsymbol{x}|\boldsymbol{q_x})\\ &\qquad\ \ \boldsymbol{\tau_p} = \mathbf{S}\,\mathrm{var}(\boldsymbol{x}|\boldsymbol{q_x}),\end{aligned} \qquad (5)$$

where the LSL BFE is given by

$$\begin{aligned}&J_{LBFE}(q_x,q_z,\boldsymbol{\tau}_p,q_{\boldsymbol{\alpha}},q_{\boldsymbol{\gamma}}) = D(q_x q_{\boldsymbol{\alpha}}||e^{-f_{x,\alpha}}) + D(q_z q_{\boldsymbol{\gamma}}||e^{-f_{z,\gamma}})\\ &+H_G(q_z,\boldsymbol{\tau}_p),\ \text{with } H_G(q_z,\boldsymbol{\tau}_p) = \tfrac{1}{2}\sum_{k=1}^{M}\left[\frac{\mathrm{var}(z_k|q_{z_k})}{\tau_{p_k}} + \ln(2\pi\tau_{p_k})\right]\end{aligned} \qquad (6)$$

and where $D(q||p) = \mathbb{E}_q(\ln(\frac{q}{p}))$ is the KLD and $H_G(q_z,\boldsymbol{\tau}_p)$ is a sum of a KLD and an entropy of Gaussians with identical means but different variances. The LSL BFE optimization problem (6) can be reformulated with the following augmented Lagrangian

$$\begin{aligned}&\min_{q_x,q_z,q_{\boldsymbol{\alpha}},q_{\boldsymbol{\gamma}},\boldsymbol{\tau}_p,\boldsymbol{u}}\ \max_{\mathbf{s},\boldsymbol{\tau}_s}\ L(q_x,q_z,\boldsymbol{\tau}_p,\boldsymbol{u},\mathbf{s},\boldsymbol{\tau}_s,q_{\boldsymbol{\alpha}},q_{\boldsymbol{\gamma}}) \text{ with}\\ &L = D(q_x q_{\boldsymbol{\alpha}}||e^{-f_{x,\alpha}}) + D(q_z q_{\boldsymbol{\gamma}}||e^{-f_{z,\gamma}}) + H_G(q_z,\boldsymbol{\tau}_p)\\ &+\mathbf{s}^T(\mathbb{E}(\boldsymbol{z}|\boldsymbol{q_z}) - \mathbf{A}\ \mathbb{E}(\boldsymbol{x}|\boldsymbol{q_x})) - \tfrac{1}{2}\boldsymbol{\tau_s}^T(\boldsymbol{\tau_p} - \mathbf{S}\,\mathrm{var}(\boldsymbol{x}|\boldsymbol{q_x}))\\ &+\tfrac{1}{2}\|\,\mathbb{E}(\boldsymbol{x}|\boldsymbol{q_x}) - \boldsymbol{u}\|^2_{\boldsymbol{\tau_r}} + \tfrac{1}{2}\|\,\mathbb{E}(\boldsymbol{z}|\boldsymbol{q_z}) - \mathbf{A}\,\boldsymbol{u}\|^2_{\boldsymbol{\tau_p}},\end{aligned} \qquad (7)$$

where $\mathbf{s}$, $\boldsymbol{\tau}_s$ are Lagrange multipliers, and $\boldsymbol{\tau}_r = \mathbf{1}./(\mathbf{S}^T\boldsymbol{\tau}_s)$ is just a short-hand notation for a quantity that depends on $\boldsymbol{\tau}_s$. We also use the notations: $\|\boldsymbol{u}\|^2_{\boldsymbol{\tau}} = \sum_i u_i^2/\tau_i$, element-wise multiplication as in $\mathbf{s}.\boldsymbol{\tau}$, element-wise division as in $\mathbf{1}./\boldsymbol{\tau}$ and $\mathbf{1}$ is a vector of ones. In [18], [19], a careful updating schedule was considered with partial optimization steps on subsets of primal and dual variables. However, that approach is not guaranteed to converge in general. Here we continue to consider an alternating optimization approach in which the schedule is less critical and some of the optimizations are reduced to gradient updates. The resulting algorithm can be considered an extended and generalized version of the ADMM algorithm (extended: there are more than two primal variable groups, generalized: the quadratic augmentation term does not exactly correspond to the linear (mean) constraint). We propose the following updating order

$$\{\boldsymbol{u}\} \to \{q_z\} \to \{\mathbf{s}\} \to \{\boldsymbol{\tau}_p,\boldsymbol{\tau}_s\} \to \{q_x\} \to \{q_{\boldsymbol{\alpha}},q_{\boldsymbol{\gamma}}\}\,. \qquad (8)$$

In other words, at iteration t we have the following sequence

$$\{\boldsymbol{u}^t\} = \arg\min_{\boldsymbol{u}} L(q_x^{t-1},q_z^t,\boldsymbol{\tau}_p^{t-1},\boldsymbol{u},\mathbf{s}^{t-1},\boldsymbol{\tau}_s^{t-1},\boldsymbol{\tau}_s^t,q_{\boldsymbol{\alpha}}^{t-1},q_{\boldsymbol{\gamma}}^{t-1}) \qquad (9)$$

$$\{q_z^t\} = \arg\min_{q_z} L(q_x^{t-1},q_z,\boldsymbol{\tau}_p^{t-1},\boldsymbol{u}^t,\mathbf{s}^{t-1},\boldsymbol{\tau}_s^{t-1},\boldsymbol{\tau}_s^t,q_{\boldsymbol{\alpha}}^{t-1},q_{\boldsymbol{\gamma}}^{t-1}) \qquad (10)$$

$$\{\mathbf{s}^t\} = \arg\max_{\mathbf{s}} L(q_x^{t-1},q_z^t,\boldsymbol{\tau}_p^{t-1},\boldsymbol{u}^t,\mathbf{s},\boldsymbol{\tau}_s^{t-1},\boldsymbol{\tau}_s^t,q_{\boldsymbol{\alpha}}^{t-1},q_{\boldsymbol{\gamma}}^{t-1}) \qquad (11)$$

$$\{\boldsymbol{\tau}_p^t,\boldsymbol{\tau}_s^t\} = \arg\min_{\boldsymbol{\tau}_p}\max_{\boldsymbol{\tau}_s} L(q_x^{t-1},q_z^t,\boldsymbol{\tau}_p,\boldsymbol{u}^t,\mathbf{s}^t,\boldsymbol{\tau}_s,\boldsymbol{\tau}_s^t,q_{\boldsymbol{\alpha}}^{t-1},q_{\boldsymbol{\gamma}}^{t-1}) \qquad (12)$$

$$\{q_x^t\} = \arg\min_{q_x} L(q_x,q_z^t,\boldsymbol{\tau}_p^t,\boldsymbol{u}^t,\mathbf{s}^t,\boldsymbol{\tau}_s^t,\boldsymbol{\tau}_s^t,q_{\boldsymbol{\alpha}}^{t-1},q_{\boldsymbol{\gamma}}^{t-1}) \qquad (13)$$

$$\{q_{\boldsymbol{\alpha}}^t,q_{\boldsymbol{\gamma}}^t\} = \arg\min_{q_{\boldsymbol{\alpha}},q_{\boldsymbol{\gamma}}} L(q_x^t,q_z^t,\boldsymbol{\tau}_p^t,\boldsymbol{u}^t,\mathbf{s}^t,\boldsymbol{\tau}_s^t,q_{\boldsymbol{\alpha}},q_{\boldsymbol{\gamma}}). \qquad (14)$$

The result appears in the Algorithm 2 table.

3.1. Update of $\boldsymbol{u}$

To update $\boldsymbol{u}$, we use a gradient descent method with line search to optimize the step-size. From (7), (9), we get the relvant cost function

$$\begin{aligned}&L(q_x^{t-1}, q_z^{t-1}, \boldsymbol{\tau}_p^{t-1}, \boldsymbol{u}, \mathbf{s}^{t-1}, \boldsymbol{\tau}_s^{t-1}, q_{\boldsymbol{\alpha}}^{t-1}, q_{\boldsymbol{\gamma}}^{t-1})\\ &= \tfrac{1}{2}\|\widehat{x}^{t-1} - \boldsymbol{u}\|^2_{\boldsymbol{\tau}_r^{t-1}} + \tfrac{1}{2}\|\widehat{z}^{t-1} - \mathbf{A}\,\boldsymbol{u}\|^2_{\boldsymbol{\tau}_p^{t-1}} + const.\end{aligned} \tag{15}$$

where $const.$ denotes constants w.r.t. $\boldsymbol{u}$. The minimizing update can be obtained as

$$\boldsymbol{u}^t = \boldsymbol{u}^{t-1} - \eta^t\,\mathbf{g}^t \tag{16}$$

with gradient $\mathbf{g}^t = \mathbf{g}^t(\boldsymbol{u}^{t-1})$ where

$$\begin{aligned}&\mathbf{g}^t(\boldsymbol{u}) = \nabla_{\boldsymbol{u}} L(q_x^{t-1}, q_z^{t-1}, \boldsymbol{\tau}_p^{t-1}, \boldsymbol{u}, \mathbf{s}^{t-1}, \boldsymbol{\tau}_s^{t-1}, q_{\boldsymbol{\alpha}}^{t-1}, q_{\boldsymbol{\gamma}}^{t-1})\\ &= -\mathbf{A}^T((\widehat{z}^{t-1} - \mathbf{A}\boldsymbol{u})./\boldsymbol{\tau}_p^{t-1}) - (\widehat{x}^{t-1} - \boldsymbol{u})./\boldsymbol{\tau}_r^{t-1}\\ &= \mathbf{g}^t(\mathbf{0}) + \mathcal{H}^t\,\boldsymbol{u},\ \mathcal{H}^t = D(1./\boldsymbol{\tau}_r^{t-1}) + \mathbf{A}^T D(1./\boldsymbol{\tau}_p^{t-1})\mathbf{A}\end{aligned} \tag{17}$$

where $D(1./\boldsymbol{\tau})$ denotes a diagonal matrix with diagonal elements $\mathbf{1}./\boldsymbol{\tau}$. The step-size η^t gets optimized for maximum descent :

$$\begin{aligned}&\frac{\partial L(q_x^{t-1}, q_z^{t-1}, \boldsymbol{\tau}_p^{t-1}, \boldsymbol{u}^t, \mathbf{s}^{t-1}, \boldsymbol{\tau}_s^{t-1}, q_{\boldsymbol{\alpha}}^{t-1}, q_{\boldsymbol{\gamma}}^{t-1})}{\partial \eta^t} = 0\\ &\Rightarrow\ \eta^t = \|\mathbf{g}^t\|^2/\mathbf{g}^{t\,T}\mathcal{H}^t\mathbf{g}^t\,.\end{aligned} \tag{18}$$

3.2. Update of $\{q_z\}$

The relevant terms in the augmented Lagrangian are

$$\begin{aligned}&L(q_x^{t-1}, q_z, \boldsymbol{\tau}_p^{t-1}, \boldsymbol{u}^t, \mathbf{s}^{t-1}, \boldsymbol{\tau}_s^{t-1}, q_{\boldsymbol{\alpha}}^{t-1}, q_{\boldsymbol{\gamma}}^{t-1})\\ &= \boldsymbol{D(q_z q_\gamma || e^{-f_{z,\gamma}}) + \tfrac{1}{2}}\mathrm{var}\boldsymbol{(z|q_z)./\tau_p^{t-1}}\\ &\boldsymbol{+}\mathbf{s}^{\boldsymbol{t-1\,T}}\,\mathbb{E}\boldsymbol{(z|q_z) + \tfrac{1}{2}\|\,}\mathbb{E}\boldsymbol{(z|q_z) - \mathbf{A}\,u^t\|^2_{\tau_p^{t-1}} + const.}\\ &\boldsymbol{= D(q_z q_\gamma || e^{-f_{z,\gamma}}) + \tfrac{1}{2}\,}\mathbb{E}\boldsymbol{(z^T z|q_z)./\tau_p^{t-1}}\\ &\boldsymbol{-(}\mathbb{E}\boldsymbol{(z|q_z))^T((\mathbf{A}\,u^t)./\tau_p^{t-1} - }\mathbf{s}^{\boldsymbol{t-1}}\boldsymbol{) + const.}\\ &\boldsymbol{= D(q_z q_\gamma || e^{-f_{z,\gamma}}) + \tfrac{1}{2}\,}\mathbb{E}\boldsymbol{(\|z - p^t\|^2_{\tau_p^{t-1}}|q_z) + const.}\end{aligned} \tag{19}$$

where $const.$ denotes constants w.r.t. z and $\boldsymbol{p}^t = \mathbf{A}\,\boldsymbol{u}^t - \mathbf{s}^{t-1}.\boldsymbol{\tau}_p^{t-1}$. The Lagrangian in (19) is separable. We get per component

$$\begin{aligned}&\min_{q_{z_k}} D(q_{z_k} q_{\gamma_k} || g^t_{z_k}/Z^t_{z_k})\ \Rightarrow\ q^t_{z_k} = \tilde{g}^t_{z_k}/Z^t_{z_k}\ \text{with}\\ &\tilde{g}^t_{z_k} = e^{\mathbb{E}_{q^{t-1}_{\gamma_k}} \ln g^t_{z_k}},\ Z^t_{z_k} = \textstyle\int \tilde{g}^t_{z_k}\,dz_k\,,\ -\ln\tilde{g}^t_{z_k}\\ &= \mathbb{E}_{q^{t-1}_{\gamma_k}} f_{z_k,\gamma_k}(z_k,\gamma_k) + \tfrac{1}{2\tau^{t-1}_{p_k}}[(z_k - p^t_k)^2 - (p^t_k)^2]\end{aligned} \tag{20}$$

where the KLD optimization is as in standard Variational Bayes (VB), and $p_k^t = p_k^t(s_k^{t-1})$. Note that the partition function $Z^t_{z_k}$ acts as cumulant generating function:

$$\begin{aligned}&-\frac{\partial \ln Z^t_{z_k}}{\partial s_k} = \mathbb{E}(z_k|q^t_{z_k}) = \mathbb{E}(z_k|p_k^t, \tau^{t-1}_{p_k}, \widehat{\gamma}_k^{t-1}) = \widehat{z}_k^t\\ &\frac{\partial^2 \ln Z^t_{z_k}}{\partial s_k^2} = \mathrm{var}(z_k|p_k^t, \tau^{t-1}_{p_k}, \widehat{\gamma}_k^{t-1}) = \tau^t_{z_k}\\ &-\frac{\partial^3 \ln Z^t_{z_k}}{\partial s_k^3} = \mathbb{E}((z_k - \mathbb{E}\,z_k)^3|q^t_{z_k}).\end{aligned} \tag{21}$$

In the Gaussian prior case, we get a Gaussian posterior $q^t_{z_k}$ with

$$\mathbf{1}./\boldsymbol{\tau}_z^t = \mathbf{1}./\boldsymbol{\tau}_p^{t-1} + \boldsymbol{\gamma}^{t-1},\ \widehat{z}^t = \boldsymbol{\tau}_z^t.(\boldsymbol{y}.\boldsymbol{\gamma}^{t-1} + \boldsymbol{p}^t./\boldsymbol{\tau}_p^{t-1})\,. \tag{22}$$

3.3. Update of $\{\mathbf{s}\}$ (ADMM style)

Although the quadratic augmentation terms in the Lagrangian do not correspond exactly to a weighted quadratic version of the linear mean constraint, due to the introduction of the auxiliary variable $\boldsymbol{u}$ which streamlines the derivation of the updates of q_x and q_z, nevertheless an ADMM style update of the mean constraint Lagrange multiplier $\mathbf{s}$ is possible. Indeed, the terms in (19) that contains $\mathbf{s}$ or $\widehat{z}$ are

$$\widehat{z}^T((\frac{1}{2}\widehat{z} - \boldsymbol{p}^t)./\boldsymbol{\tau}_p^{t-1}) = \widehat{z}^T(\mathbf{s}^{t-1} + (\frac{1}{2}\widehat{z} - \mathbf{A}\boldsymbol{u}^t)./\boldsymbol{\tau}_p^{t-1}) \tag{23}$$

Taking the gradient w.r.t. $\widehat{z}$ (as part of the optimization over q_z) leads to the RHS of

$$\mathbf{s}^t = \mathbf{s}^{t-1} + (\widehat{z}^t - \mathbf{A}\boldsymbol{u}^t)./\boldsymbol{\tau}_p^{t-1}\,. \tag{24}$$

Hence, if we use this update for $\mathbf{s}$, then (23) reduces to $\widehat{z}^T\mathbf{s}^t$, as if the quadratic augmentation terms have disappeared! This if the main characteristic of the Lagrange multiplier update in ADMM, which corresponds to a gradient ascent with a particular choice of (diagonal) step-size.

3.4. Update of $\{\boldsymbol{\tau}_p, \boldsymbol{\tau}_s\}$

In [18], [19], the carefully chosen updating schedule made the quadratic augmentation terms inactive when $\{\boldsymbol{\tau}_p, \boldsymbol{\tau}_s\}$. Here these terms only become inactive at convergence. Nevertheless, these terms only play an active role for the means and not for the variances. Hence we shall ignore them here. Hence, the terms of interest in (7) for (12) are

$$\begin{aligned}&L(q_x^{t-1}, q_z^t, \boldsymbol{\tau}_p, \boldsymbol{u}^t, \mathbf{s}^t, \boldsymbol{\tau}_s, q_{\boldsymbol{\alpha}}^{t-1}, q_{\boldsymbol{\gamma}}^{t-1})\\ &= H_G(q_z^t, \boldsymbol{\tau}_p) - \tfrac{1}{2}\boldsymbol{\tau}_s^T(\boldsymbol{\tau}_p - \mathbf{S}\,\boldsymbol{\tau}_x^{t-1}) + const. = const. +\\ &\tfrac{1}{2}\sum_{k=1}^{M}\left[\frac{\tau^t_{z_k}}{\tau_{p_k}} + \ln(2\pi\,\tau_{p_k})\right] - \frac{1}{2}\sum_{k=1}^{M}\tau_{s_k}(\tau_{p_k} - \mathbf{S}_{k,:}\,\boldsymbol{\tau}_x^{t-1})\end{aligned} \tag{25}$$

where $const.$ denotes constants w.r.t. $\{\boldsymbol{\tau}_p, \boldsymbol{\tau}_s\}$. Deriving w.r.t. $\{\boldsymbol{\tau}_p, \boldsymbol{\tau}_s\}$ yields the feasibility conditions

$$\frac{\partial L}{\partial \tau_{p_k}} = 0 \Rightarrow\ \tau^t_{s_k} = \frac{1}{\tau^{t-1}_{p_k}}(1 - \frac{\tau^t_{z_k}}{\tau^{t-1}_{p_k}}). \tag{26}$$

$$\frac{\partial L}{\partial \tau_{s_k}} = 0\ \Rightarrow\ \tau^t_{p_k} = \mathbf{S}_{k,:}\,\boldsymbol{\tau}_x^{t-1} \tag{27}$$

which we run as a fixed-point sub-algorithm. To guarantee non-negativity, the update of $\tau_{\mathbf{s}}$ should come first.

3.5. Update of q_x

For the update of q_x, consider the relevant terms in the augmented Lagrangian (and remember that $\boldsymbol{\tau}_r^t = \mathbf{1}./(\mathbf{S}^T\boldsymbol{\tau}_s^t)$ or $\mathbf{1}./\boldsymbol{\tau}_r^t = \mathbf{S}^T\boldsymbol{\tau}_s^t$)

$$\begin{aligned}&L(q_x, q_z^t, \boldsymbol{\tau}_p^t, \boldsymbol{u}^t, \mathbf{s}^t, \boldsymbol{\tau}_s^t, q_{\boldsymbol{\alpha}}^{t-1}, q_{\boldsymbol{\gamma}}^{t-1})\\ &\boldsymbol{= D(q_x q_\alpha || e^{-f_{x,\alpha}}) - }\mathbf{s}^{\boldsymbol{t\,T}}\mathbf{A}\ \mathbb{E}\boldsymbol{(x|q_x)}\\ &\boldsymbol{+\tfrac{1}{2}\tau_s^{t\,T}}\mathbf{S}\,\mathrm{var}\boldsymbol{(x|q_x) + \tfrac{1}{2}\|\,}\mathbb{E}\boldsymbol{(x|q_x) - u^t\|^2_{\tau_r^t} + const.}\\ &\boldsymbol{= D(q_x q_\alpha || e^{-f_{x,\alpha}}) + \tfrac{1}{2}\,(1./\tau_r^t)^T\ }\mathbb{E}\boldsymbol{(x.x|q_x)}\\ &\boldsymbol{-}\mathbf{s}^{\boldsymbol{t\,T}}\mathbf{A}\ \mathbb{E}\boldsymbol{(x|q_x) - (u^t./\tau_r^t))^T\,}\mathbb{E}\boldsymbol{(x|q_x) + const.}\\ &\boldsymbol{= D(q_x q_\alpha || e^{-f_{x,\alpha}}) + \tfrac{1}{2}\,(1./\tau_r^t)^T\ }\mathbb{E}\boldsymbol{(x.x|q_x)}\\ &\boldsymbol{-(u^t + \tau_r^t.}\mathbf{A}^{\boldsymbol{T}}\mathbf{s}^{\boldsymbol{t}}\boldsymbol{)^T(}\mathbb{E}\boldsymbol{(x|q_x)./\tau_r^t) + const.}\\ &\boldsymbol{= D(q_x q_\alpha || e^{-f_{x,\alpha}}) + \tfrac{1}{2}\ }\mathbb{E}\boldsymbol{(\|x - }\mathbf{r}^{\boldsymbol{t}}\boldsymbol{\|^2_{\tau_r^t}|q_x) + const.}\end{aligned} \tag{28}$$

where $const.$ denotes constants w.r.t. x, and $\mathbf{r}^t = \boldsymbol{u}^t + \boldsymbol{\tau}_r^t.\mathbf{A}^T\mathbf{s}^t$. This cost function is again separable. We get per component a VB update

$$\begin{aligned}&\min_{q_{x_i}} D(q_{x_i} q^{t-1}_{\alpha_i} || g^t_{x_i}/Z^t_{x_i})\ \Rightarrow\ q^t_{x_i} = \tilde{g}^t_{x_i}/Z^t_{x_i}\ \text{with}\\ &\tilde{g}^t_{x_i} = e^{\mathbb{E}_{q^{t-1}_{\alpha_i}} \ln g^t_{x_i}},\ Z^t_{x_i} = \textstyle\int \tilde{g}^t_{x_i}\,dx_i\,,\ -\ln\tilde{g}^t_{x_i}\\ &= \mathbb{E}_{q^{t-1}_{\alpha_i}}(f_{x_i,\alpha_i}(x_i,\alpha_i)) + \tfrac{1}{2\tau^t_{r_i}}[(x_i - r_i^t)^2 - (r_i^t)^2]\,.\end{aligned} \tag{29}$$

The partition function $Z^t_{x_i}$ acts as cumulant generating function:

$$\begin{aligned}&\tau^t_{r_i}\,\frac{\partial \ln Z^t_{x_i}}{\partial r_i} = \mathbb{E}(x_i|q^t_{x_i}) = \mathbb{E}(x_i|r_i^t, \tau^t_{r_i}, \widehat{\alpha}_i^{t-1}) = \widehat{x}_i^t\\ &(\tau^t_{r_i})^2\,\frac{\partial^2 \ln Z^t_{x_i}}{\partial r_i^2} = \mathrm{var}(x_i|r_i^t, \tau^t_{r_i}, \widehat{\alpha}_i^{t-1}) = \tau^t_{x_i}\,.\end{aligned} \tag{30}$$

Simplifications occur again in the Gaussian scenario.

Algorithm 2 VAMBGAMP(-SBL)

```
Require: $\boldsymbol{y}, \mathbf{A}, \mathbf{S} = \mathbf{A}.\mathbf{A}, f_{x,\boldsymbol{\alpha}}(x, \boldsymbol{\alpha}), f_{z,\boldsymbol{\gamma}}(z, \boldsymbol{\gamma})$
Initialize: $t = 0, \widehat{x}^0, \boldsymbol{\tau}_x^0, \boldsymbol{u}^0, \boldsymbol{\tau}_p^0, \mathbf{s}^0 = \mathbf{0}, \widehat{\boldsymbol{\gamma}}^0, \widehat{\boldsymbol{\alpha}}^0$
repeat    (t=1,2,...)
    [Output node update]
    $\boldsymbol{u}^t = \boldsymbol{u}^{t-1} - \eta^t\, \mathbf{g}^t$, with $\mathbf{g}^t, \eta^t$ from (17), (18)
    $\boldsymbol{p}^t = \mathbf{A}\,\boldsymbol{u}^t - \mathbf{s}^{t-1}.\boldsymbol{\tau}_p^{t-1}$
    $\widehat{\boldsymbol{z}}^t = \mathbb{E}(\boldsymbol{z}|\boldsymbol{p}^t, \boldsymbol{\tau}_{\boldsymbol{p}}^{t-1}, \widehat{\boldsymbol{\gamma}}^{t-1})$, Gaussian case: $\widehat{z}^t = \boldsymbol{\tau}_z^t.(\boldsymbol{y}.\widehat{\boldsymbol{\gamma}}^{t-1} + \boldsymbol{p}^t./\boldsymbol{\tau}_p^{t-1})$
    $\boldsymbol{\tau}_{\boldsymbol{z}}^t = \text{var}(\boldsymbol{z}|\boldsymbol{p}^t, \boldsymbol{\tau}_{\boldsymbol{p}}^{t-1}, \widehat{\boldsymbol{\gamma}}^{t-1})$, Gaussian case: $\mathbf{1}./\boldsymbol{\tau}_z^t = \mathbf{1}./\boldsymbol{\tau}_p^{t-1} + \widehat{\boldsymbol{\gamma}}^{t-1}$
    $\mathbf{s}^t = \mathbf{s}^{t-1} + (z^t - \mathbf{A}\boldsymbol{u}^t)./\boldsymbol{\tau}_p^{t-1}$
    [Variance matching]
    $\boldsymbol{\tau}_s^t = (\mathbf{1} - \boldsymbol{\tau}_z^t./\boldsymbol{\tau}_p^{t-1})./\boldsymbol{\tau}_p^{t-1}$
    $\boldsymbol{\tau}_p^t = \mathbf{S}\,\boldsymbol{\tau}_x^{t-1}$
    $\boldsymbol{\tau}_r^t = \mathbf{1}./(\mathbf{S}^T \boldsymbol{\tau}_s^t)$
    [Input node update]
    $\mathbf{r}^t = \boldsymbol{u}^t + \boldsymbol{\tau}_r^t.\mathbf{A}^T \mathbf{s}^t$
    $\widehat{\boldsymbol{x}}^t = \mathbb{E}(\boldsymbol{x}|\mathbf{r}^t, \boldsymbol{\tau}_{\boldsymbol{r}}^t, \widehat{\boldsymbol{\alpha}}^{t-1})$, Gaussian case: $\widehat{x}^t = \mathbf{r}^t./(\mathbf{1} + \widehat{\boldsymbol{\alpha}}^{t-1}.\boldsymbol{\tau}_r^t)$
    $\boldsymbol{\tau}_{\boldsymbol{x}}^t = \text{var}(\boldsymbol{x}|\mathbf{r}^t, \boldsymbol{\tau}_{\boldsymbol{r}}^t, \widehat{\boldsymbol{\alpha}}^{t-1})$, Gaussian case: $\mathbf{1}./\boldsymbol{\tau}_x^t = \mathbf{1}./\boldsymbol{\tau}_r^t + \widehat{\boldsymbol{\alpha}}^{t-1}$
    [Hyperparameters update]
    $\widehat{\alpha}_i^t = \frac{2a+1}{\mathbb{E}_{q_{x_i}}(x_i^2)+2b}, \forall i$
    $\widehat{\gamma}_k^t = \frac{2c+1}{\mathbb{E}_{q_{z_k}}(|y_k - z_k|^2)+2d}, \forall k$
until Convergence
```

3.6. Update of Hyperparameter α

The relevant part of the augmented Lagrangian simply leads to a KLD term

$$q_{\boldsymbol{\alpha}}^t = \arg\min_{q_{\boldsymbol{\alpha}}} D(q_x^t q_{\boldsymbol{\alpha}} || e^{-f_{x,\boldsymbol{\alpha}}}) + const. \tag{31}$$

which gets minimized alternatingly as in VB [22]

$$\ln q_{\alpha_i}^t = (\frac{1}{2} + a - 1)\ln\alpha_i - (\frac{1}{2}\mathbb{E}_{q_{x_i}^t}(x_i^2) + b)\,\alpha_i + const. \tag{32}$$

This means that the posterior of α_i is a Gamma distribution: $q_{\alpha_i}^t = \mathcal{G}(\alpha_i; \widehat{a}, \widehat{b}^t)$, with $\widehat{a} = a + 1/2$ and $\widehat{b}^t = \mathbb{E}_{q_{x_i}^t}(x_i^2)/2 + b$, with mean $\frac{\widehat{a}}{\widehat{b}^t}$ or hence:

$$\widehat{\alpha}_i^t = \frac{2a+1}{\mathbb{E}_{q_{x_i}^t}(x_i^2) + 2b}. \tag{33}$$

3.7. Update of Hyperparameter γ

The relevant part of the augmented Lagrangian is again a KLD term:

$$q_{\boldsymbol{\gamma}}^t = \arg\min_{q_{\boldsymbol{\gamma}}(\boldsymbol{\gamma})} D(q_z^t q_{\boldsymbol{\gamma}} || e^{-f_{z,\boldsymbol{\gamma}}}) + const. \tag{34}$$

which gets minimized alternatingly as in VB [22]

$$\ln q_{\gamma_k}^t = (\frac{1}{2} + c - 1)\ln\gamma_i - (\frac{\mathbb{E}_{q_{z_k}^t}(|y_k - z_k|^2)}{2} + d)\gamma_i + const. \tag{35}$$

which means again that the posterior of γ_k is a Gamma distribution $q_{\gamma_k}^t = \mathcal{G}(\gamma_k; \widehat{c}, \widehat{d}^t)$, with $\widehat{c} = c + 1/2$ and $\widehat{d}^t = \mathbb{E}_{q_{\gamma_k}^t}(|y_k - z_k|^2)/2 + d$, and with mean $\frac{\widehat{c}}{\widehat{d}^t}$ or:

$$\gamma_k^t = \frac{2c+1}{\mathbb{E}_{q_{z_k}^t}(|y_k - z_k|^2) + 2d}. \tag{36}$$

For the case in which all noise variances are assumed to be equal, the update of $\gamma_0 = \gamma_k, \forall k$ can be shown to be [22]:

$$\gamma_0^t = \frac{2c + M}{\mathbb{E}_{q_z^t}(\|\boldsymbol{y} - z\|^2) + 2d}. \tag{37}$$

4. VAMBGAMP-SBL LARGE SYSTEM ANALYSIS

In GAMP, as opposed to AMP, we may not have (simple) analytical updates for means and variances. As a result, the take on large system analysis (LSA) for GAMP is from a different angle. If both the rows or the columns of $\mathbf{A}$ are now modeled as independent, then given that also the priors on x and z are independent (factorized), the true posteriors for x and z will become factorized and will equal the approximate posteriors $q_x(x)$, $q_z(z)$. So multiplication with $\mathbf{A}$ or $\mathbf{A}^T$ acts like scrambling in CDMA communications, that renders the individual outputs independent. Furthermore, the marginal posteriors are the product of the respective prior and extrinsic distributions that correspond to information coming through $\mathbf{A}$ or $\mathbf{A}^T$, the random nature of which will lead to Gaussian extrinsic distributions by the central limit theorem. In other words, in the LSA, in which the dimensions of x and z (the two dimensions of $\mathbf{A}$) tend to infinity at a constant ratio, the approximate posteriors handled in GAMP become asymptotically exact. As a result, the variance information propagated by GAMP corresponds asymptotically to the exact MSE of the (MMSE) estimates propagated by GAMP. The existing GAMP steady-state analysis results are valid, assuming that the algorithm has converged to such a steady state. Such steady-state analysis appears in [20] (particularly in the extended arxiv version), or in [23].

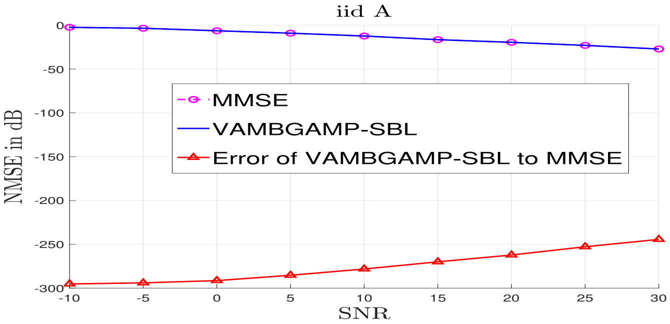

Fig. 1. NMSE curves for i.i.d. $\mathbf{A}$, with $M = 150, N = 250$. GAMP has similar performance as VAMBGAMP-SBL, hence omitted.

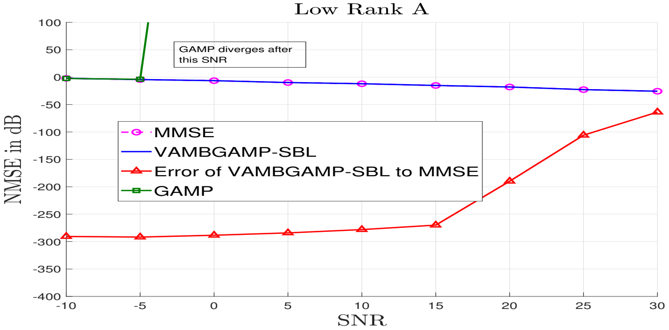

Fig. 2. NMSE curves for low rank $\mathbf{A}$, with $M = 150, N = 250$.

In the Gaussian case of SBL, MMSE estimation becomes LMMSE, for which we have investigated LSA in [24] using large random matrix theory. It can be checked that the LSA of the general GAMP case above reduces to these same results in the Gaussian case.

5. SIMULATION RESULTS

The figures show Normalized MSE $\boldsymbol{\tau}_x$ (2 top curves) and normalized MSE difference between x estimated by AMBAMP and LMMSE (2 bottom curves) with variance profile $\sigma_{x_i}^2 = 0.93^{i-1}$, $i = 1, \ldots, N$, for $M = 150, N = 250$. Fig. 1 is for i.i.d. Gaussian $\mathbf{A}$ whereas Fig. 2 is for low rank $\mathbf{A}$ in which the smallest half of the singular values in an i.i.d. $\mathbf{A}$ are set to zero. These simulations show that the VAMBGAMP-SBL algorithm continues to work in an unrealistically severe scenario, in which AMP diverges.

6. CONCLUDING REMARKS

We propose a convergent version of GAMP, VAMBGAMP, which applies alternating minimization to an augmented Lagrangian of a large system limit of the Bethe free Energy (BFE). One quadratic subproblem is minimized by a gradient descent with line search to maintain AMP style complexity. Some variance parameter updates are replaced by fixed-point updates. And an ADMM style update is used for the Lagrange multiplier associated to the mean constraint. Additionally, we have observed that the KLDs appearing in the BFE naturally lead to variational Bayesian inference for the hyperparameters.

Acknowledgements EURECOM's research is partially supported by its industrial members: ORANGE, BMW, SAP, iABG, Norton LifeLock, and by the French ANR project CellFree6G.

7. REFERENCES

[1] M. E. Tipping, “Sparse Bayesian Learning and the Relevance Vector Machine,” *J. Mach. Learn. Res.*, vol. 1, 2001.

[2] D. P. Wipf and B. D. Rao, “Sparse Bayesian Learning for Basis Selection ,” *IEEE Trans. on Sig. Proc.*, vol. 52, no. 8, Aug. 2004.

[3] R. Giri and Bhaskar D. Rao, “Type I and type II bayesian methods for sparse signal recovery using scale mixtures,” *IEEE Trans. on Sig Process.*, vol. 64, no. 13, 2018.

[4] X. Tan and J. Li, “Computationally Efficient Sparse Bayesian Learning via Belief Propagation ,” *IEEE Trans. on Sig. Proc.*, vol. 58, no. 4, Apr. 2013.

[5] C. K. Thomas and D. Slock, “Low Complexity Static and Dynamic Sparse Bayesian Learning Combining BP, VB and EP Message Passing,” in *Asilomar Conf. on Sig., Sys., and Comp.*, CA, USA, 2019.

[6] K. P. Murphy et al., “Loopy belief propagation for approximate inference: an empirical study. ,” in *In 15th Conf. Uncert. in Art. Intell. (UAI)*, Stockholm, Sweden, 1999.

[7] J. Du et al., “Convergence Analysis of Distributed Inference with Vector-Valued Gaussian Belief Propagation,” *Jrnl. of Mach. Learn. Res.*, April 2018.

[8] J. Du et al., “Convergence Analysis of the Information Matrix in Gaussian Belief Propagation ,” in *IEEE Intl. Conf. on Acoustics, Speech, and Sig. Process.*, New Orleans, LA, USA, 2017.

[9] Q. Su and Y. Wu, “Convergence Analysis of the Variance in Gaussian Belief Propagation,” *IEEE Trans. on Sig. Process.*, vol. 62, no. 19, Oct. 2014.

[10] B. Cseke and T. Heskes, “Properties of Bethe Free Energies and Message Passing in Gaussian Models ,” *Jrnl. of Art. Intell. Res.*, May 2011.

[11] D. M. Malioutov, J. K. Johnson, and A. S. Willsky, “Walk-Sums and Belief Propagation in Gaussian Graphical Models ,” *Jrnl. of Mach. Learn. Res.*, Oct. 2006.

[12] L. Liu, E.G. Larsson, W. Yu, P. Popovski, C. Stefanovic, and E. de Carvalho, “Sparse Signal Processing for Grant-Free Massive Connectivity,” *IEEE Sig. Proc. Magazine*, Sept. 2018.

[13] C. K. Thomas and Dirk Slock, “Variational Bayesian Learning for Channel Estimation and Transceiver Determination,” in *Info. Theo. and Appl. Wkshp*, San Diego, USA, February 2018.

[14] Khac-Hoang Ngo, Maxime Guillaud, Alexis Decurninge, Sheng Yang, and Philip Schniter, “Multi-user detection based on expectation propagation for the non-coherent simo multiple access channel,” *IEEE Transactions on Wireless Communications*, vol. 19, no. 9, pp. 6145–6161, 2020.

[15] C. K. Thomas and D. Slock, “Low Complexity Static and Dynamic Sparse Bayesian Learning Combining BP, VB and EP Message Passing,” in *52nd IEEE Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, USA, Nov. 2019.

[16] S. Rangan, P. Schniter, A. Fletcher, and V. Cevher, “Fixed Points of Generalized Approximate Message Passing with Arbitrary Matrices,” *IEEE Trans. Info. Theory*, Dec. 2016.

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

[18] D. Slock”, “Convergent Approximate Message Passing by Alternating Constrained Minimization of Bethe Free Energy,” in *IEEE 7th Forum on Research and Technologies for Society and Industry Innov. (RTSI)*, Paris, France, Aug. 2022.

[19] D. Slock”, “Convergent Approximate Message Passing,” in *IEEE Int’l Mediterr. Conf. Comm’s and Netw’ing (MeditCom)*, Athens, Greece, Sep. 2022.

[20] S. Rangan, “Generalized Approximate Message Passing for Estimation with Random Linear Mixing,” in *Proc. IEEE Int. Symp. Inf. Theory*, Saint Petersburg, Russia, 2011, extended version: arxiv1010.5141.

[21] F. Krzakala, A. Manoel, E. W. Tramel, and L. Zdeborova, “Variational Free Energies for Compressed Sensing,” in *IEEE Intl. Sympo. Info. Theo.*, Honolulu, HI, USA, Jun. 2014.

[22] C. K. Thomas and D. Slock, “SAVE - space alternating variational estimation for sparse Bayesian learning,” in *Data Science Workshop*, Jun. 2018.

[23] A. Javanmard and A. Montanari, “State Evolution for General Approximate Message Passing Algorithms, with Applications to Spatial Coupling,” *Inf. Inference*, vol. 2, no. 2, 2013, doi: 10.1093/imaiai/iat004.

[24] C.K. Thomas and D. Slock, “Posterior Variance Predictions in Sparse Bayesian Learning under Approximate Inference Techniques,” in *Asilomar Conf. on Sig., Sys., and Comp.*, 2020.