IMPROVED VARIANCE PREDICTIONS IN APPROXIMATE MESSAGE PASSING

Zilu Zhao, Dirk Slock

Communication Systems Department, EURECOM, France
zilu.zhao@eurecom.fr, dirk.slock@eurecom.fr

ABSTRACT

In the Generalized Linear Model (GLM), the unknowns may be non-identically independent distributed (niid), as for instance in the Sparse Bayesian Learning (SBL) problem. The Generalized Approximate Message Passing (GAMP) algorithm performs computationally efficient belief propagation for Bayesian inference. The GAMP algorithms predicts the posterior variances correctly in the case of measurement matrices with (n)iid entries. In order to cover more ill-conditioned measurement matrices, the (right) rotationally invariant (RRI) model was introduced in which the (right) singular vectors are Haar distributed, leading to Vector AMP VAMP however assumes iid priors and posteriors. Here we introduce a convergent version of AMB (AMBAMP) applied to Unitarily transformed data, with a variance correction based on Haar Large System Analysis (LSA). The recently introduced reVAMP perspective shows that the resulting AMBUAMP algorithm has an underlying multivariate Gaussian posterior approximation, that does not get computed but that allows the LSA. The individual variance predictions are exact asymptotically in the RRI setting, as illustrated by a Gaussian Mixture Model example.

***Index Terms*—** Approximate Message Passing, AMP, VAMP, reVAMP, AMB UAMP, Haar Large System Analysis

1. INTRODUCTION

The recovery of sparse signal vectors is a fundamental problem in signal processing and has wide-ranging applications, including compressed sensing, image and speech processing, and machine learning. Sparse Bayesian Learning (SBL) [1] is one popular method for sparse signal recovery. In spite of the a priori non sparsifying Gaussian prior, when the prior variances need to be estimated also, this hierarchical Gaussian setting becomes sparsifying. In the Gaussian noise case, the signal model for the recovery of a sparse signal vector $\boldsymbol{x}$ can be formulated as, $\boldsymbol{y} = \mathbf{A}\boldsymbol{x} + \boldsymbol{v}$, where $\boldsymbol{y}$ are the observations or data, $\mathbf{A}$ is called the known measurement or sensing matrix of dimension $M \times N$ with $M < N$ in the compressed sensing case. For a sparse model, $\boldsymbol{x}$ contains only K non-zero (or significant) entries, with $K < M < N$.

However, SBL can be computationally complex, especially when dealing with high-dimensional data. This complexity arises from the need to perform matrix inversion in each iteration (for the hyperparameter estimation).

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

To address this issue, the Vector Approximate Message Passing (VAMP) algorithm has been proposed [4]. It generates the factor graph by splitting one vector variable node $\boldsymbol{x}$ into two variable instances $\boldsymbol{x}_1 = \boldsymbol{x}_2 = \boldsymbol{x}$. An Expectation-Propagation (EP) like message passing algorithm [5] is then applied to the factor graph with vector valued messages. VAMP has been shown to perform well under Right Rotationally Invariant (RRI) $\mathbf{A}$, and the (variance) state evolution of VAMP has been rigorously established [4].

1.1. Prior Work

In previous research [4], the optimality of VAMP has been analyzed using the replica method [6]. The replica method gives a system of equations that describes the fixed point of the VAMP state evolution and the optimal (sum) Mean Squared Error (MSE). VAMP (implicitly) assumes identically independently distributed (iid) priors and posteriors. AMP provides individual MSEs, but with convergence issues. Unitary Transformed AMP (UTAMP) [7] applies AMP to a transformed linear model to make AMP more robust to $\mathbf{A}$.

In [8] we have introduced a convergent version of AMP, AMBGAMP, which works with arbitrary priors to accommodate e.g. LASSO and many other compressed sensing problem formulations. The variance predictions in AMP are correct in the Large System Limit (LSL) for a matrix $\mathbf{A}$ considered random with non iid (niid) elements (i.e. independent but not necessarily identically distributed, e.g. an $\mathbf{A}$ with deterministic magnitudes but iid signs). In [9] we have introduced UAMP which assumes a RRI model for $\mathbf{A}$. Such a model

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

covers a much wider class of matrices since only the right singular vectors are considered Haar distributed [10] (the row subspace of $\mathbf{A}$ is uniformly oriented in space), whereas left singular vectors and singular values are deterministic. The RRI model requires a variance correction to be correct in the LSL, based on Haar Large System Analysis (LSA). However, UAMP in [9] is limited to Gaussian priors.

1.2. Main Contribution

In this paper, we extend UAMP to handle arbitrary priors, leading to the Generalized Linear Model (GLM) (here still with Gaussian noise though). We introduce the Haar LSA based variance corrections to the convergent AMBAMP applied to a unitarily tranformed data model, resulting in the AMBUAMP algorithm. Since the LSA is based on Gaussian signal models, we invoke the recently introduced reVAMP perspective [11] to show that these AMP algorithms with arbitrary priors implicitly construct a multivariate Gaussian posterior approximation, on which the LSA can be based, and which is consistent with the marginal posterior first and second-order moments constructed by the algorithm. We show that without correction term the optimal MSE is not reached, but the Haar LSA allows to distill the proper correction term which only depends on quantities already appearing in the algorithm. To illustrate the proper functioning of the AMBUAMP algorithm, we work out the details for and simulate the case of Gaussian Mixture Model (GMM) priors.

2. UNITARILY TRANSFORMED LINEAR MODEL

The data model considered in AMP is essentially a linear mixing model

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

with (possibly) non identically independently distributed (niid) prior $p_{\boldsymbol{x}}(\boldsymbol{x}) = \prod_{i=1}^{N} p_{x_i}(x_i)$ and iid measurement noise $p_{\boldsymbol{v}}(\boldsymbol{v}) = \prod_{i=1}^{N} p_v(v_i)$, $\boldsymbol{v} \sim \mathcal{N}(0, \sigma_v^2\mathbf{I})$. $\mathbf{A}$ is the $M \times N$ measurement matrix and we assume $M < N$. In order to handle possibly ill-conditioned $\mathbf{A}$, consider the economy singular value decomposition (SVD) of $\mathbf{A}$:

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\,\mathbf{V}^T \tag{2}$$

where $\mathbf{U} \in \mathbb{R}^{M\times M}$ is an orthogonal matrix, $\boldsymbol{\Sigma} \in \mathbb{R}^{M\times M}$ is diagonal and $\mathbf{V} \in \mathbb{R}^{M\times N}$ is semi-orthogonal ($\mathbf{V}^T\mathbf{V} = \mathbf{I}$). As in the computationally efficient version of VAMP [7], [4], we can take advantage of the simple white Gaussian noise model to transform the linear model as:

$$\mathbf{U}^T\boldsymbol{y} = \boldsymbol{\Sigma}\,\mathbf{V}^T\boldsymbol{x} + \mathbf{U}^T\boldsymbol{v}\,. \tag{3}$$

Since we assume $\boldsymbol{v}$ to be iid Gaussian, we still have $\mathbf{U}^T\boldsymbol{v} \sim \mathcal{N}(0, \sigma_v^2\mathbf{I})$. For simplicity, let's denote $\boldsymbol{y}' = \mathbf{U}^T\boldsymbol{y}$, $\boldsymbol{v}' = \mathbf{U}^T\boldsymbol{v}$, $\boldsymbol{\sigma}_v^2 = \sigma_v^2\mathbf{1}$, $\mathbf{A}' = \boldsymbol{\Sigma}\mathbf{V}^T$, $\mathbf{S}' = \mathbf{A}'.\mathbf{A}'$ and $\boldsymbol{\lambda} = \boldsymbol{\Sigma}^2\mathbf{1}$, leading to the transformed linear model

$$\boldsymbol{y}' = \mathbf{A}'\,\boldsymbol{x} + \boldsymbol{v}' \tag{4}$$

where $\mathbf{z} = \mathbf{A}'\,\boldsymbol{x}$ is the result of the linear mixture.

3. PROPOSED AMBUAMP

The abbreviation AMB stands for ACM-LSL-BFE, which stands for Alternating Constrained Minimization of the LSL of the Bethe Free Energy (BFE). AMBGAMP employs most of the same updates as GAMP, but GAMP does not apply a strict alternating minimization (block coordinate descent) principle, particularly in the presence of constraints. AMBGAMP has been derived in [8] of which we recall the key steps. We shall apply here AMBGAMP to the unitarily transformed data (4), which will lead to AMBUAMP after a correction of the variances. GAMP corresponds to the constrained minimization of a LSL of the BFE (see also [12] and references therein):

$$\begin{aligned} &\min_{q_{\boldsymbol{x}},q_{\mathbf{z}},\boldsymbol{\tau}_p} J_{BFE}(q_{\boldsymbol{x}}, q_{\mathbf{z}}, \boldsymbol{\tau}_p)\\ &s.t.\ \mathbb{E}(\mathbf{z}|q_{\mathbf{z}}) = \mathbf{A}'\ \mathbb{E}(\boldsymbol{x}|q_{\boldsymbol{x}})\\ &\quad\boldsymbol{\tau}_p = \mathbf{S}'\,\mathrm{var}(\boldsymbol{x}|q_{\boldsymbol{x}}), \end{aligned} \tag{5}$$

where the LSL BFE is given by

$$J_{BFE}(q_{\boldsymbol{x}}, q_{\mathbf{z}}, \boldsymbol{\tau}_p) = D(q_{\boldsymbol{x}}||e^{-f_{\boldsymbol{x}}}) + D(q_{\mathbf{z}}||e^{-f_{\mathbf{z}}}) + H_G(q_{\mathbf{z}}, \boldsymbol{\tau}_p)$$
$$\text{with } H_G(q_{\mathbf{z}}, \boldsymbol{\tau}_p) = \frac{1}{2}\sum_{k=1}^{M}\left[\frac{\mathrm{var}(z_k|q_{z_k})}{\tau_{p_k}} + \ln(2\pi\tau_{p_k})\right] \tag{6}$$

and where $D(q||p) = \mathbb{E}_q(\ln(\frac{q}{p}))$ is the Kullback-Leibler Divergence (KLD) and $H_G(q_{\mathbf{z}}, \boldsymbol{\tau}_p)$ 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

$$\begin{aligned} &\min_{q_{\boldsymbol{x}},q_{\mathbf{z}},\boldsymbol{\tau}_p,\boldsymbol{u}}\ \max_{\mathbf{s},\boldsymbol{\tau}_s} L(q_{\boldsymbol{x}}, q_{\mathbf{z}}, \boldsymbol{\tau}_p, \boldsymbol{u}, \mathbf{s}, \boldsymbol{\tau}_s) \text{ with}\\ &L = D(q_{\boldsymbol{x}}||e^{-f_{\boldsymbol{x}}}) + D(q_{\mathbf{z}}||e^{-f_{\mathbf{z}}}) + H_G(q_{\mathbf{z}}, \boldsymbol{\tau}_p)\\ &+\mathbf{s}^T(\mathbb{E}(\mathbf{z}|q_{\mathbf{z}}) - \mathbf{A}'\ \mathbb{E}(\boldsymbol{x}|q_{\boldsymbol{x}})) - \tfrac{1}{2}\boldsymbol{\tau}_s^T(\boldsymbol{\tau}_p - \mathbf{S}'\,\mathrm{var}(\boldsymbol{x}|q_{\boldsymbol{x}}))\\ &+\tfrac{1}{2}\|\,\mathbb{E}(\boldsymbol{x}|q_{\boldsymbol{x}}) - \boldsymbol{u}\|_{\boldsymbol{\tau}_r}^2 + \tfrac{1}{2}\|\,\mathbb{E}(\mathbf{z}|q_{\mathbf{z}}) - \mathbf{A}'\,\boldsymbol{u}\|_{\boldsymbol{\tau}_p}^2, \end{aligned} \tag{7}$$

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: $\|\boldsymbol{u}\|_{\boldsymbol{\tau}}^2 = \sum_i u_i^2/\tau_i$, element-wise multiplication as in $\mathbf{s}.\boldsymbol{\tau}$ and element-wise division as in $\mathbf{1}./\boldsymbol{\tau}$, and $\mathbf{1}$ is a vector of ones. In [8] we apply an alternating optimization strategy, combined with a gradient update with line search for the auxiliary quantity $\boldsymbol{u}$, an ADMM update of the Lagrange multipliers $\mathbf{s}$, and fixed point iterations for $\boldsymbol{\tau}_p$ and $\boldsymbol{\tau}_s$. We detail two key steps.

3.1. Update of $\boldsymbol{u}$

To update $\boldsymbol{u}$, we use a gradient descent method with line search optimized step-size. From (7) we get at iteration t

$$\begin{aligned} &L(q_{\boldsymbol{x}}^{t-1}, q_{\mathbf{z}}^{t-1}, \boldsymbol{\tau}_p^{t-1}, \boldsymbol{u}, \mathbf{s}^{t-1}, \boldsymbol{\tau}_s^{t-1})\\ &= \tfrac{1}{2}\|\widehat{\boldsymbol{x}}^{t-1} - \boldsymbol{u}\|_{\boldsymbol{\tau}_r^{t-1}}^2 + \tfrac{1}{2}\|\widehat{\mathbf{z}}^{t-1} - \mathbf{A}'\,\boldsymbol{u}\|_{\boldsymbol{\tau}_p^{t-1}}^2 + const. \end{aligned} \tag{8}$$

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

$$\boldsymbol{u}^t = \boldsymbol{u}^{t-1} - \eta^t\,\mathbf{g}^t \tag{9}$$

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

$$\begin{aligned}
&\mathbf{g}^t(\boldsymbol{u}) = \nabla_{\boldsymbol{u}} L(q_{\boldsymbol{x}}^{t-1}, q_{\mathbf{z}}^{t-1}, \boldsymbol{\tau}_p^{t-1}, \boldsymbol{u}, \mathbf{s}^{t-1}, \boldsymbol{\tau}_s^{t-1}) \\
&= -\mathbf{A}'^T((\widehat{\mathbf{z}}^{t-1} - \mathbf{A}'\boldsymbol{u})./\boldsymbol{\tau}_p^{t-1}) - (\widehat{\boldsymbol{x}}^{t-1} - \boldsymbol{u})./\boldsymbol{\tau}_r^{t-1} \\
&= \mathbf{g}^t(\mathbf{0}) + \mathcal{H}^t \boldsymbol{u},\ \mathcal{H}^t = \mathbf{D}(\mathbf{1}./\boldsymbol{\tau}_r^{t-1}) + \mathbf{A}'^T \mathbf{D}(\mathbf{1}./\boldsymbol{\tau}_p^{t-1})\mathbf{A}'
\end{aligned} \tag{10}$$

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

$$\begin{aligned}
&\frac{\partial L(q_{\boldsymbol{x}}^{t-1}, q_{\mathbf{z}}^{t-1}, \boldsymbol{\tau}_p^{t-1}, \boldsymbol{u}^t, \mathbf{s}^{t-1}, \boldsymbol{\tau}_s^{t-1})}{\partial \eta^t} = 0 \\
&\Rightarrow\ \eta^t = \|\mathbf{g}^t\|^2 / {\mathbf{g}^t}^T \mathcal{H}^t \mathbf{g}^t\,.
\end{aligned} \tag{11}$$

3.2. Update of $q_{\boldsymbol{x}}$

For the update of $q_{\boldsymbol{x}}$, consider the relevant terms in the augmented Lagrangian (and remember that $\mathbf{1}./\boldsymbol{\tau}_r^{t-1} = \mathbf{S}'^T \boldsymbol{\tau}_s^{t-1}$)

$$\begin{aligned}
&L(q_{\boldsymbol{x}}, q_{\mathbf{z}}^{t-1}, \boldsymbol{\tau}_p^{t-1}, \boldsymbol{u}^t, \mathbf{s}^{t-1}, \boldsymbol{\tau}_s^{t-1}) \\
&= D(q_{\boldsymbol{x}} || e^{-f_{\boldsymbol{x}}}) - {\mathbf{s}^{t-1}}^T \mathbf{A}'\ \mathbb{E}(\boldsymbol{x}|q_{\boldsymbol{x}}) \\
&+ \tfrac{1}{2} {\boldsymbol{\tau}_s^{t-1}}^T \mathbf{S}'\ \mathrm{var}(\boldsymbol{x}|q_{\boldsymbol{x}}) + \tfrac{1}{2} \|\,\mathbb{E}(\boldsymbol{x}|q_{\boldsymbol{x}}) - \boldsymbol{u}^t\|^2_{\boldsymbol{\tau}_r^{t-1}} + const. \\
&= D(q_{\boldsymbol{x}} || e^{-f_{\boldsymbol{x}}}) + \tfrac{1}{2}\,(\mathbf{1}./\boldsymbol{\tau}_r^{t-1})^T\ \mathbb{E}(\boldsymbol{x}.\boldsymbol{x}|q_{\boldsymbol{x}}) \\
&- {\mathbf{s}^{t-1}}^T \mathbf{A}'\ \mathbb{E}(\boldsymbol{x}|q_{\boldsymbol{x}}) - (\boldsymbol{u}^t./\boldsymbol{\tau}_r^{t-1}))^T\ \mathbb{E}(\boldsymbol{x}|q_{\boldsymbol{x}}) + const. \\
&= D(q_{\boldsymbol{x}} || e^{-f_{\boldsymbol{x}}}) + \tfrac{1}{2}\,(\mathbf{1}./\boldsymbol{\tau}_r^{t-1})^T\ \mathbb{E}(\boldsymbol{x}.\boldsymbol{x}|q_{\boldsymbol{x}}) \\
&- (\boldsymbol{u}^t + \boldsymbol{\tau}_r^{t-1}.\mathbf{A}'^T \mathbf{s}^{t-1})^T (\mathbb{E}(\boldsymbol{x}|q_{\boldsymbol{x}})./\boldsymbol{\tau}_r^{t-1}) + const. \\
&= D(q_{\boldsymbol{x}} || e^{-f_{\boldsymbol{x}}}) + \tfrac{1}{2}\ \mathbb{E}(\|\boldsymbol{x} - \mathbf{r}^t\|^2_{\boldsymbol{\tau}_r^t} | q_{\boldsymbol{x}}) + const.
\end{aligned} \tag{12}$$

where $const.$ denotes constants w.r.t. $\boldsymbol{x}$, and $\mathbf{r}^t = \boldsymbol{u}^t + \boldsymbol{\tau}_r^{t-1}.\mathbf{A}'^T \mathbf{s}^{t-1}$. The Lagrangian in (12) is separable. We get per component

$$\begin{aligned}
&\min_{q_{x_i}} D(q_{x_i} || g_{x_i}^t / Z_{x_i}^t)\ \Rightarrow\ q_{x_i}^t = g_{x_i}^t / Z_{x_i}^t,\ Z_{x_i}^t = \int g_{x_i}^t(x_i)\, dx_i, \\
&-\ln g_{x_i}^t(x_i) = f_{x_i}(x_i) + \tfrac{1}{2\tau_{r_i}^t}[(x_k - r_i^t)^2 - {r_i^t}^2]\,.
\end{aligned} \tag{13}$$

The partition function $Z_{x_i}^t$ acts as cumulant generating function:

$$\begin{aligned}
&\tau_{r_i}^t \frac{\partial \ln Z_{x_i}^t}{\partial r_i^t} = \mathbb{E}(x_i | q_{x_i}^t) = \mathbb{E}(x_i | r_i^t, \tau_{r_i}^t) = \widehat{x}_i^t \\
&(\tau_{r_i}^t)^2 \frac{\partial^2 \ln Z_{x_i}^t}{\partial {r_i^t}^2} = \mathrm{var}(x_i | r_i^t, \tau_{r_i}^t) = \tau_{x_i}^t\,.
\end{aligned} \tag{14}$$

In the Gaussian prior case, we get a Gaussian posterior $q_{\boldsymbol{x}}^t$ with

$$\mathbf{1}./\boldsymbol{\tau}_x^t = \mathbf{1}./\boldsymbol{\tau}_r^{t-1} + \mathbf{1}./\boldsymbol{\sigma}_x^2,\ \widehat{\boldsymbol{x}}^t = \boldsymbol{\tau}_x^t.(\mathbf{r}^t./\boldsymbol{\tau}_r^{t-1})\,. \tag{15}$$

The AMBUAMP algorithm appears in the Algorithm 1 table. The variance correction term appears in red, where the introduction of the nonnegative part $\lfloor . \rfloor_+$ may avoid transient problems.

4. REVAMP INTERPRETATION

In [11], we introduced the reVAMP algorithm, which corresponds to a revisited VAMP approach to MMSE estimation in the GLM where the only asymptotic approximation that is made is to invoke the Central Limit Theorem (CLT) to approximate the extrinsic pdfs by Gaussians. A MMSE estimate corresponds to the mean of the posterior pdf, and decouples

Algorithm 1 AMBUAMP

Require: $\boldsymbol{y}', \mathbf{A}', \mathbf{S}' = \mathbf{A}'.\mathbf{A}', f_{\boldsymbol{x}}(\boldsymbol{x}), \sigma_v^2$

Initialize: $t = 0$, $\boldsymbol{u}^0 = \mathbf{0}$, $\widehat{\boldsymbol{x}}^0 = \mathbf{0}$, $\widehat{\mathbf{z}}^0 = \mathbf{0}$, $\mathbf{s}^0 = \mathbf{0}$, $\boldsymbol{\tau}_r^0 = \mathbf{1}$, $\boldsymbol{\tau}_p^0 = \mathbf{1}$

repeat (t=1,2,...)

- $\boldsymbol{u}^t = \boldsymbol{u}^{t-1} - \eta^t\, \mathbf{g}^t$, with $\mathbf{g}^t$, η^t from (10), (11)
- [Input node update]
- $\mathbf{r}^t = \boldsymbol{u}^t + \boldsymbol{\tau}_r^{t-1}.(\mathbf{A}'^T \mathbf{s}^{t-1})$
- $\widehat{\boldsymbol{x}}^t = \mathbb{E}(\boldsymbol{x}|\mathbf{r}^t, \boldsymbol{\tau}_r^{t-1})$, Gaussian $p_{\boldsymbol{x}}$: $\mathbf{1}./\boldsymbol{\tau}_x^t = \mathbf{1}./\boldsymbol{\tau}_r^{t-1} + \mathbf{1}./\boldsymbol{\sigma}_x^2$
- $\boldsymbol{\tau}_x^t = \mathrm{var}(\boldsymbol{x}|\mathbf{r}^t, \boldsymbol{\tau}_r^{t-1})$, Gaussian $p_{\boldsymbol{x}}$: $\widehat{\boldsymbol{x}}^t = \boldsymbol{\tau}_x^t.(\mathbf{r}^t./\boldsymbol{\tau}_r^{t-1})$
- $\boldsymbol{\tau}_p^t = \mathbf{S}'\,\boldsymbol{\tau}_x^t$
- [Output node update]
- $\boldsymbol{p}^t = \mathbf{A}'\,\boldsymbol{u}^t - \mathbf{s}^{t-1}.\boldsymbol{\tau}_p^t$
- $\mathbf{1}./\boldsymbol{\tau}_z^t = \mathbf{1}./\boldsymbol{\tau}_p^t + \mathbf{1}./\boldsymbol{\sigma}_v^2$
- $\widehat{\mathbf{z}}^t = \boldsymbol{\tau}_z^t.(\boldsymbol{y}'./\boldsymbol{\sigma}_v^2 + \boldsymbol{p}^t./\boldsymbol{\tau}_p^t)$
- $\mathbf{s}^t = \mathbf{s}^{t-1} + (\widehat{\mathbf{z}}^t - \mathbf{A}'\boldsymbol{u}^t)./\boldsymbol{\tau}_p^t$
- $\boldsymbol{\tau}_s^t = \mathbf{1}./(\boldsymbol{\tau}_p^t + \left\lfloor 1 - \frac{1}{N^2}\left(\mathbf{1}^T \boldsymbol{\tau}_x^t\right)\left(\boldsymbol{\lambda}^T \boldsymbol{\tau}_s^{t-1}\right)\right\rfloor_+ \boldsymbol{\sigma}_v^2)$
- $\boldsymbol{\tau}_r^t = \mathbf{1}./(\mathbf{S}'^T \boldsymbol{\tau}_s^t)$

until Convergence

between the components of a vector. So, consider the joint pdf

$$p(\boldsymbol{x}, \boldsymbol{y}') = p(\boldsymbol{y}'|\boldsymbol{x}) \prod_{i=1}^N p_{x_i}(x_i)\,. \tag{16}$$

The true posterior for x_i can be written as

$$\begin{aligned}
&p(x_i|\boldsymbol{y}') = p_{x_i}(x_i)\ \left(\int p(\boldsymbol{y}'|\boldsymbol{x}) \textstyle\prod_{j\neq i}^N p_{x_j}(x_j) dx_j\right) / Z_i(\boldsymbol{y}') \\
&\approx\ p_{x_i}(x_i)\, \mathcal{N}\,(x_i; r_i, \tau_{r_i})
\end{aligned} \tag{17}$$

where $\left(\int p(\boldsymbol{y}'|\boldsymbol{x}) \prod_{j\neq i}^N p_{x_j}(x_j) dx_j\right) / Z_i(\boldsymbol{y}')$ is the extrinsic for x_i and $Z_i(\boldsymbol{y}')$ is a normalization factor. For a very large class of models for $\mathbf{A}'$ and $\boldsymbol{x}$, it is clear that the CLT will allow to approximate the extrinsic by a Gaussian distribution, as indicated in (17). As explored in [11], the Gaussian extrinsic approximation can be built from the Component-Wise Conditionally Unbiased MMSE estimate of x_i. Now, from the non-Gaussian approximate posterior in (17), with Gaussian extrinsic but true prior, we can find the MMSE estimate $\widehat{x}_i$ and associated MMSE τ_{x_i}. These quantities allow us to build a Gaussian posterior approximation $q_{x_i}(x_i) = \mathcal{N}\,(x_i; \widehat{x}_i, \tau_{x_i})$ which minimizes the KLD to the non-Gaussian posterior. The (marginal) Gaussian posterior approximation $q_{x_i}(x_i)$ in turn allows us to find a Gaussian prior approximation $q_i(x_i)$ by Gaussian division

$$\begin{aligned}
&q_i(x_i) = \mathcal{N}\left(x_i; a_i, \sigma_{x_i}^2\right) \propto \mathcal{N}\left(x_i; \widehat{x}_i, \tau_{x_i}\right) / \mathcal{N}\left(x_i; r_i, \tau_{r_i}\right), \\
&1/\sigma_{x_i}^2 = 1/\tau_{x_i} - 1/\tau_{r_i}\ ,\ a_i = \sigma_{x_i}^2 (\widehat{x}_i/\tau_{x_i} - r_i/\tau_{r_i})\,.
\end{aligned} \tag{18}$$

Note that in the Gaussian prior case, the $\boldsymbol{\sigma}_x^2$ are of course the variances of the true prior. The joint distribution $\prod_{i=1}^N q_i(x_i)$ equals $\mathcal{N}(\boldsymbol{x}; \boldsymbol{a}, \boldsymbol{\sigma}_x^2)$. Now, since also $p(\boldsymbol{y}|\boldsymbol{x})$ is Gaussian, the (independent) Gaussian priors in turn induce a multivariate Gaussian posterior approximation

$$q_{\boldsymbol{x}}(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}; \mathbf{m}, \mathbf{C}_m) \propto p(\boldsymbol{y}|\boldsymbol{x}) \mathcal{N}(\boldsymbol{x}; \boldsymbol{a}, \boldsymbol{\sigma}_x^2) \tag{19}$$

where

$$\begin{aligned}
\mathbf{C}_m &= \left[\mathbf{A}'^T \mathbf{C}_{\boldsymbol{vv}}^{-1} \mathbf{A}' + \mathbf{D}_x^{-1}\right]^{-1} \\
\mathbf{m} &= \mathbf{C}_m \left[\mathbf{A}'^T \mathbf{C}_{\boldsymbol{vv}}^{-1} \boldsymbol{y}' + \mathbf{D}_x^{-1} \boldsymbol{a}\right],
\end{aligned} \tag{20}$$

$\mathbf{D}_x$ is a short hand notation for $D(\boldsymbol{\sigma}_x^2)$ and $\mathbf{C}_{\boldsymbol{vv}} = \sigma_v^2\mathbf{I}$. Furthermore, if we define $\boldsymbol{\tau}_m = \text{diag}(\mathbf{C}_m)$, then we have at convergence of the reVAMP algorithm $\mathbf{m} = \widehat{\boldsymbol{x}}$ and $\boldsymbol{\tau}_m = \boldsymbol{\tau}_x$, and $q_{x_i}(x_i)$ is a marginal of $q_{\boldsymbol{x}}(\boldsymbol{x})$. This reVAMP perspective on the (AMB)(U)AMP algorithm shows that asymptotically, the MMSE estimates can be associated to an underlying Gaussian linear model which will facilitate the large random matrix analysis for the variances.

5. HAAR LARGE SYSTEM ANALYSIS

Following [4], we model $\mathbf{A}$ as a right rotationally invariant matrix. That means that in the economy SVD of $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$, $\mathbf{U}$ and $\overline{\boldsymbol{\Sigma}}$ remain deterministic but $\mathbf{V}$ is considered drawn as M columns of a Haar distributed $N \times N$ random matrix (i.e. the columns of $\mathbf{V}$ form an orthonormal basis for a uniformly randomly oriented M-dimensional subspace of the N-dimensional space). The analysis of the large system primarily relies on the deterministic equivalent proposed in [13], which states

Lemma 1. *Let $\boldsymbol{P}$ be any Hermitian matrix with bounded spectral norm and let $\mathbf{V} \in \mathbb{R}^{N\times M}$ be $M < N$ columns of a Haar distributed (unitary) random matrix. Let $\mathbf{B}$ be a nonnegative definite matrix with $||\mathbf{B}|| < \infty$ ($||\mathbf{B}||$ represents the spectral norm) and $\mathbf{D}$ be any diagonal matrix with positive entries. Then the following convergence result holds almost surely,*

$$\frac{1}{N}tr\left[\mathbf{B}\left(\mathbf{V}\boldsymbol{P}\mathbf{V}^T + \mathbf{D}\right)^{-1}\right] - \frac{1}{N}tr\left[\mathbf{B}(\overline{e}\,\mathbf{I} + \mathbf{D})^{-1}\right] \xrightarrow{a.s.} 0. \tag{21}$$

The scalar $\overline{e}$ can be obtained as the unique solution (fixed point) of the following system of equations,

$$\begin{aligned} \overline{e} &= \frac{1}{N}tr\left[\boldsymbol{P}\left(e\boldsymbol{P} + (1 - e\,\overline{e})\mathbf{I}\right)^{-1}\right], \\ e &= \frac{1}{N}tr\left[\mathbf{B}(\overline{e}\mathbf{I} + \mathbf{D})^{-1}\right]. \end{aligned} \tag{22}$$

The MMSE solution for (1) is given by

$$\begin{aligned} \widehat{\boldsymbol{x}}_{\text{MMSE}} &= \left(\tfrac{1}{\sigma_v^2}\mathbf{A}^T\mathbf{A} + D\left(1./\boldsymbol{\sigma}_x^2\right)\right)^{-1}\tfrac{1}{\sigma_v^2}\mathbf{A}^T\boldsymbol{y}, \\ \mathbf{C}_{\text{MMSE}} &= \left(\tfrac{1}{\sigma_v^2}\mathbf{A}^T\mathbf{A} + D\left(1./\boldsymbol{\sigma}_x^2\right)\right)^{-1}. \end{aligned} \tag{23}$$

Thus, the MSE is

$$\text{MSE} = \frac{1}{N}\text{tr}[\mathbf{C}_{\text{MMSE}}]. \tag{24}$$

By using Lemma 1, we obtain the large system approximation

$$\text{MSE} \xrightarrow{\text{a.s.}} \frac{1}{N}\text{tr}\left[(\overline{e}^0\mathbf{I} + D(1./\boldsymbol{\sigma}_x^2))^{-1}\right] \tag{25}$$

with

$$\begin{aligned} \overline{e}^0 &= \tfrac{1}{N}\text{tr}\left[\tfrac{1}{\sigma_v^2}\overline{\boldsymbol{\Sigma}}^2\left(\tfrac{e^0}{\sigma_v^2}\overline{\boldsymbol{\Sigma}}^2 + (1 - e^0\overline{e}^0)\mathbf{I}\right)^{-1}\right], \\ e^0 &= \tfrac{1}{N}\text{tr}\left[(\overline{e}^0\mathbf{I} + D(1./\boldsymbol{\sigma}_x^2))^{-1}\right]. \end{aligned} \tag{26}$$

6. LARGE SYSTEM ANALYSIS OF AMBUAMP

We will first prove that under the assumption that $\overline{\mathbf{V}}$ is Haar-distributed, the averaged $\boldsymbol{\tau}_x$, namely $\frac{1}{N}\mathbf{1}^T\boldsymbol{\tau}_x$, does not match the optimal MSE defined in (24). Then in the next section, we will propose a correction term such that $\boldsymbol{\tau}_x$ matches the optimal MSE. The steady state of variances in UAMP can be summarized as follows

$$\begin{aligned} \mathbf{1}./\boldsymbol{\tau}_s^\infty &= \boldsymbol{\sigma}_v^2 + \mathbf{S}'\boldsymbol{\tau}_x^\infty \\ \mathbf{1}./\boldsymbol{\tau}_x^\infty &= \mathbf{1}./\boldsymbol{\sigma}_x^2 + \mathbf{S}'^T\boldsymbol{\tau}_s^\infty. \end{aligned} \tag{27}$$

With the large system assumptions, as N tend to infinity, we approximate $\overline{\mathbf{V}}^T\mathbf{D}_N\overline{\mathbf{V}}$ and $\overline{\mathbf{V}}\mathbf{D}_M\overline{\mathbf{V}}^T$ to $\frac{1}{N}\text{tr}(\mathbf{D}_N)\mathbf{I}$ and $\frac{1}{N}\text{tr}(\mathbf{D}_M)\mathbf{I}$ respectively. Thus, we have

$$\begin{aligned} \mathbf{S}'\boldsymbol{\tau}_x^\infty &= \text{diag}\left[\overline{\boldsymbol{\Sigma}}\overline{\mathbf{V}}^T D(\boldsymbol{\tau}_x^\infty)\overline{\mathbf{V}}\overline{\boldsymbol{\Sigma}}\right] = \frac{1}{N}\mathbf{1}^T\boldsymbol{\tau}_x^\infty\ \boldsymbol{\lambda}, \\ \mathbf{S}'^T\boldsymbol{\tau}_s^\infty &= \text{diag}\left[\overline{\mathbf{V}}\overline{\boldsymbol{\Sigma}}D(\boldsymbol{\tau}_s^\infty)\overline{\boldsymbol{\Sigma}}\overline{\mathbf{V}}^T\right] = \frac{1}{N}\boldsymbol{\lambda}^T\boldsymbol{\tau}_s^\infty\ \mathbf{1}. \end{aligned} \tag{28}$$

Now we show the following.

Lemma 2. *In AMP with equivalent measurement matrix $\mathbf{A}'$, the variance prediction $\frac{1}{N}\mathbf{1}^T\boldsymbol{\tau}_x^\infty$ does not match the optimal MSE in (25).*

Proof. If the noise is iid, the MSE remains unchanged under a unitary transformation. Therefore, equations (25) and (26) remain the same in this transformed system. We will prove by contradiction.
Suppose that $\boldsymbol{\tau}_x^\infty$ matches the optimal MSE, we then have

$$\begin{aligned} \frac{1}{N}\text{tr}[D(\boldsymbol{\tau}_x^\infty)] &= \frac{1}{N}\text{tr}\left[\left(D(\mathbf{S}'^T\boldsymbol{\tau}_s^\infty) + D(1./\boldsymbol{\sigma}_x^2)\right)^{-1}\right] \\ &= e^0 = \frac{1}{N}\text{tr}\left[(\overline{e}^0\mathbf{I} + D(1./\boldsymbol{\sigma}_x^2))^{-1}\right], \end{aligned} \tag{29}$$

which implies

$$\begin{aligned} \overline{e}^0 &= \tfrac{1}{N}\boldsymbol{\lambda}^T\boldsymbol{\tau}_s^\infty = \tfrac{1}{N}\text{tr}\left[\overline{\boldsymbol{\Sigma}}^2 D(\boldsymbol{\tau}_s^\infty)\right] \\ &= \tfrac{1}{N}\text{tr}\left[\overline{\boldsymbol{\Sigma}}^2\left(\tfrac{1}{N}\text{tr}[D(\boldsymbol{\tau}_x^\infty)]D(\boldsymbol{\lambda}) + \sigma_v^2\mathbf{I}\right)^{-1}\right] \\ &= \tfrac{1}{N}\text{tr}\left[\overline{\boldsymbol{\Sigma}}^2\left(e^0\overline{\boldsymbol{\Sigma}}^2 + \sigma_v^2\mathbf{I}\right)^{-1}\right] \end{aligned} \tag{30}$$

One can observe that $\overline{e}^0$ in (30) only equals $\overline{e}^0$ in (26) if $e^0\overline{e}^0 = 0$. □

6.1. Correction Term For τ_s

Considering the asymptotic MSE expression in (25), and the second equations in (27), (28), we can still write

$$\begin{aligned} \tfrac{1}{N}\text{tr}[D(\boldsymbol{\tau}_x^\infty)] &= \tfrac{1}{N}\text{tr}\left[\left(D(\mathbf{S}'^T\boldsymbol{\tau}_s^\infty) + D(1./\boldsymbol{\sigma}_x^2)\right)^{-1}\right] \\ &= \tfrac{1}{N}\text{tr}\left[\left(\tfrac{1}{N}\boldsymbol{\lambda}^T\boldsymbol{\tau}_s^\infty\mathbf{I} + D(1./\boldsymbol{\sigma}_x^2)\right)^{-1}\right] \end{aligned} \tag{31}$$

Introduce

$$e_c = \frac{1}{N}\text{tr}[D(\boldsymbol{\tau}_x^\infty)]\,, \overline{e}_c = \frac{1}{N}\boldsymbol{\lambda}^T\boldsymbol{\tau}_s^\infty. \tag{32}$$

Now compare (31),(32) with (25),(26), then we require $\overline{e}_c$ to be of the form

$$\overline{e}_c = \frac{1}{N}\text{tr}\left[\overline{\boldsymbol{\Sigma}}^2\left(e_c\overline{\boldsymbol{\Sigma}}^2 + (1 - e_c\overline{e}_c)\sigma_v^2\mathbf{I}\right)^{-1}\right]. \tag{33}$$

From the definition of $\overline{e}_c$ in (32), we have

$$\overline{e}_c = \frac{1}{N}\boldsymbol{\lambda}^T\boldsymbol{\tau}_s^\infty = \frac{1}{N}\text{tr}\left[\overline{\boldsymbol{\Sigma}}^2 D(\boldsymbol{\tau}_s^\infty)\right]. \tag{34}$$

Comparing (34) with (33), we want to design the update scheme of τ_s such that at steady state,

$$D(\boldsymbol{\tau}_s^{\infty}) = \left(e_c\overline{\boldsymbol{\Sigma}}^2 + (1-e_c\overline{e}_c)\sigma_v^2\mathbf{I}\right)^{-1}. \tag{35}$$

From the definition of e_c in (32), we have

$$e_c\overline{\boldsymbol{\Sigma}}^2 = \frac{1}{N}\text{tr}[D(\boldsymbol{\tau}_x)]\overline{\boldsymbol{\Sigma}}^2 = \frac{1}{N}\mathbf{1}^T\boldsymbol{\tau}_x\,\overline{\boldsymbol{\Sigma}}^2 = \frac{1}{N}\mathbf{1}^T\boldsymbol{\tau}_x\,D(\boldsymbol{\lambda}). \tag{36}$$

Under the large system approximation (28), we obtain

$$e_c\overline{\boldsymbol{\Sigma}}^2 = D(\frac{1}{N}\mathbf{1}^T\boldsymbol{\tau}_x\,\boldsymbol{\lambda}) = D(\mathbf{S}'\boldsymbol{\tau}_x^{\infty}). \tag{37}$$

Substituting (32) and (37) into (35), we get

$$D(\boldsymbol{\tau}_s^{\infty}) = \left[D(\mathbf{S}'\boldsymbol{\tau}_x^{\infty}) + \sigma_v^2\mathbf{I} - \frac{1}{N^2}\left(\mathbf{1}^T\boldsymbol{\tau}_x^{\infty}\right)\left(\boldsymbol{\lambda}^T\boldsymbol{\tau}_s^{\infty}\right)\sigma_v^2\mathbf{I}\right]^{-1}.$$

Therefore, we propose a simple correction for the update of $\boldsymbol{\tau}_s^t$ in the algorithm

$$\boldsymbol{\tau}_s^t = \mathbf{1}./\left[\boldsymbol{\sigma}_v^2 - \frac{1}{N^2}\left(\mathbf{1}^T\boldsymbol{\tau}_x^t\right)\left(\boldsymbol{\lambda}^T\boldsymbol{\tau}_s^{t-1}\right)\boldsymbol{\sigma}_v^2 + \boldsymbol{\tau}_p^t\right] \tag{38}$$

One can verify by Lemma 1 that with this correction, $(e^c, \overline{e}^c)$ converge to a fixed point of (26) and hence $\boldsymbol{\tau}_x$ will converge to the optimal MSE.

7. CONVERGENCE ANALYSIS

Since the variance updates have been modified, the convergence of the variance subsystem needs to be reanalyzed. We provide here a sketch analysis following [8], where we analyze the contractivity via the infinity norm of the Jacobian. From the variance updates

$$\begin{aligned}&\mathbf{1}./\boldsymbol{\tau}_x^t = \mathbf{1}./\boldsymbol{\sigma}_x^2 + \mathbf{S}^T\boldsymbol{\tau}_s^{t-1},\\ &\mathbf{1}./\boldsymbol{\tau}_s^t = \sigma_v^2\mathbf{1} + \mathbf{S}\boldsymbol{\tau}_x^t - \tfrac{\sigma_v^2}{N^2}(\mathbf{1}^T\boldsymbol{\tau}_x^t)(\boldsymbol{\lambda}^T\boldsymbol{\tau}_s^{t-1}),\end{aligned} \tag{39}$$

we obtain the following Jacobians

$$\begin{aligned}\mathbf{J}_{xs}^t &= \tfrac{\partial\boldsymbol{\tau}_x^t}{\partial\boldsymbol{\tau}_s^{t-1,T}} = -\mathbf{D}_x^{t,2}\mathbf{S}^T;\\ \mathbf{J}_{sx}^t &= \tfrac{\partial\boldsymbol{\tau}_s^t}{\partial\boldsymbol{\tau}_x^{t,T}} = -\mathbf{D}_s^{t,2}(\mathbf{S} - \tfrac{\sigma_v^2}{N^2}(\boldsymbol{\lambda}^T\boldsymbol{\tau}_s^{t-1})\mathbf{1}_M\mathbf{1}_N^T);\\ \mathbf{J}_{ss}^t &= \tfrac{\partial\boldsymbol{\tau}_s^t}{\partial\boldsymbol{\tau}_s^{t-1,T}} = -\mathbf{D}_s^{t,2}(\tfrac{\sigma_v^2}{N^2}(\mathbf{1}^T\boldsymbol{\tau}_x^t)\mathbf{1}_N\boldsymbol{\lambda}^T).\end{aligned} \tag{40}$$

Following the chain rule, we write with the gradient of $\boldsymbol{\tau}_s^t$

$$\begin{aligned}\boldsymbol{G}^t &= -(\mathbf{D}_s^t)^{-1}\tfrac{d\boldsymbol{\tau}_s^t}{d\boldsymbol{\tau}_s^{t-1}}\mathbf{D}_s^{t-1} = -(\mathbf{D}_s^t)^{-1}(\mathbf{J}_{sx}^t\mathbf{J}_{xs}^t + \mathbf{J}_{ss}^t)\mathbf{D}_s^{t-1}\\ &= -\mathbf{D}_s^t\left(\mathbf{S} - \tfrac{\sigma_v^2}{N}(\boldsymbol{\lambda}^T\boldsymbol{\tau}_s^{t-1})\mathbf{1}_M\mathbf{1}_N^T\right)\mathbf{D}_x^{t,2}\mathbf{S}^T\mathbf{D}_s^{t-1}\\ &\quad -\mathbf{D}_s^t\left(\tfrac{\sigma_v^2}{N^2}(\mathbf{1}^T\boldsymbol{\tau}_x^t)\mathbf{1}_N\boldsymbol{\lambda}^T\right)\mathbf{D}_s^{t-1}.\end{aligned} \tag{41}$$

The optimal test vector for the matrix infinity norm will again be $\mathbf{1}$. We then have

$$\begin{aligned}\boldsymbol{G}^t\mathbf{1} &= \mathbf{D}_s^t\left(\mathbf{S} - \tfrac{\sigma_v^2}{N}(\boldsymbol{\lambda}^T\boldsymbol{\tau}_s^{t-1})\mathbf{1}_M\mathbf{1}_N^T\right)\mathbf{D}_x^t\mathbf{D}_x^t\mathbf{S}^T\boldsymbol{\tau}_s^{t-1}\\ &\quad +\mathbf{D}_s^t\left(\tfrac{\sigma_v^2}{N^2}(\mathbf{1}^T\boldsymbol{\tau}_x^t)\mathbf{1}_N\boldsymbol{\lambda}^T\right)\boldsymbol{\tau}_s^{t-1}.\end{aligned} \tag{42}$$

From (39), we see that $\mathbf{D}_x^{t-1}\mathbf{S}^T\boldsymbol{\tau}_s^{t-1} \preceq \mathbf{1}$, leading to

$$\begin{aligned}\boldsymbol{G}^t\mathbf{1} &\preceq \mathbf{D}_s^t\left(\mathbf{S} - \tfrac{\sigma_v^2}{N}(\boldsymbol{\lambda}^T\boldsymbol{\tau}_s^{t-1})\mathbf{1}_M\mathbf{1}_N^T\right)\boldsymbol{\tau}_x^t\\ &\quad +\mathbf{D}_s^t\left(\tfrac{\sigma_v^2}{N^2}(\mathbf{1}^T\boldsymbol{\tau}_x^t)\mathbf{1}_N\boldsymbol{\lambda}^T\right)\boldsymbol{\tau}_s^{t-1} = \mathbf{D}_s^t\mathbf{S}\boldsymbol{\tau}_x^t\end{aligned} \tag{43}$$

From (39), $\mathbf{D}_s^t\mathbf{S}\boldsymbol{\tau}_x^t \preceq \mathbf{1}$, and hence, $\|\boldsymbol{G}^t\|_\infty < 1$.

8. GAUSSIAN MIXTURE MODEL PRIOR CASE

8.1. True MMSE Solution

Assume that the prior distribution of each x_i is given by:

$$p_{x_i}(x_i) = \textstyle\sum_{n=1}^L \alpha_n\mathcal{N}(x_i;\mu_{ni},\sigma_{ni}^2);\ \sum_{n=L}^3\alpha_n = 1. \tag{44}$$

The true MMSE estimation mean and covariance matrix for a Gaussian mixture are derived in [11]

$$\begin{aligned}&\widehat{\boldsymbol{x}}_{\text{MMSE}} = \mathbb{E}[\boldsymbol{x}|\boldsymbol{y}],\\ &\mathbf{C}_{\text{MMSE}} = \mathbb{E}[\boldsymbol{x}\boldsymbol{x}^T|\boldsymbol{y}] - \mathbb{E}[\boldsymbol{x}|\boldsymbol{y}]\,\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}]^T.\end{aligned} \tag{45}$$

8.2. AMBUAMP Scalar MMSE Calculations

The computation of the posterior (13) in UAMP with a Gaussian mixture prior (44) can be derived analogously to (45), see [11]. Define Z_{ni} :

$$\begin{aligned}Z_{ni} &= \textstyle\int\alpha_n\mathcal{N}(x;\mu_{ni},\sigma_{ni}^2)\mathcal{N}(x;r_i,\tau_{r_i})dx\\ &= \tfrac{\alpha_n}{\sqrt{2\pi(\sigma_{ni}^2+\tau_{r_i})}}e^{-\frac{(\mu_{ni}-r_i)^2}{2(\sigma_{ni}^2+\tau_{r_i})}}.\end{aligned} \tag{46}$$

Then we have

$$\begin{aligned}\widehat{x}_i &= \tfrac{\int x\sum_{n=1}^3\alpha_n\mathcal{N}(x;\mu_{ni},\sigma_{ni}^2)\mathcal{N}(x;r_i,\tau_{r_i})dx}{\int\sum_{n=1}^3\alpha_n\mathcal{N}(x;\mu_{ni},\sigma_{ni}^2)\mathcal{N}(x;r_i,\tau_{r_i})dx}\\ &= \tfrac{\sum_{n=1}^3\frac{\mu_{ni}\tau_{r_i}+r_i\sigma_{ni}^2}{\sigma_{ni}^2+\tau_{r_i}}Z_{ni}}{\sum_{n=1}^3 Z_{ni}},\end{aligned} \tag{47}$$

$$\begin{aligned}\tau_{x_i} &= \tfrac{\int x^2\sum_{n=1}^3\alpha_n\mathcal{N}(x;\mu_{ni},\sigma_{ni}^2)\mathcal{N}(x;r_i,\tau_{r_i})dx}{\int\sum_{n=1}^3\alpha_n\mathcal{N}(x;\mu_{ni},\sigma_{ni}^2)\mathcal{N}(x;r_i,\tau_{r_i})dx} - \widehat{x}_i^2\\ &= \tfrac{\sum_{n=1}^3\left[(\frac{\mu_{ni}\tau_{r_i}+r_i\sigma_{ni}^2}{\sigma_{ni}^2+\tau_{r_i}})^2+\frac{\tau_{r_i}\sigma_{ni}^2}{\sigma_{ni}^2+\tau_{r_i}}\right]Z_{ni}}{\sum_{n=1}^3 Z_{ni}} - \widehat{x}_i^2.\end{aligned} \tag{48}$$

9. SIMULATION RESULTS

To verify our findings, we simulate the scenarios where the measurement matrices $\mathbf{A} \in \mathbb{R}^{5\times 10}$ are ill-conditioned. We set the condition number of the measurement matrix to 100 and the diagonal entries in the singular matrix of $\mathbf{A}$ form a geometry series. We fix the SNR to 20dB. Each element of the signal vector $\boldsymbol{x}$ is independently drawn from a Gaussian mixture model

$$p_{x_i}(x_i) = 0.5\mathcal{N}(x_i;0,4\times 0.25^{i-1}) + 0.5\mathcal{N}(x_i;0,0.5^{i-1}) \tag{49}$$

To evaluate the averaged posterior variance prediction error, we compare the normalized square of the sum difference, which is defined as

$$\frac{(\text{tr}[\mathbf{C}_{\text{MMSE}}] - \mathbf{1}^T\boldsymbol{\tau}_x)^2}{\text{tr}[\mathbf{C}_{\text{MMSE}}]^2}. \tag{50}$$

Similarly, we use the normalized sum of squares to evaluate the individual prediction error of the posterior variances, which is defined as

$$\frac{(\mathrm{diag}[\mathbf{C}_{\mathrm{MMSE}}]-\boldsymbol{\tau}_x)^T(\mathrm{diag}[\mathbf{C}_{\mathrm{MMSE}}]-\boldsymbol{\tau}_x)}{\mathrm{tr}[\mathbf{C}_{\mathrm{MMSE}}]^2}. \tag{51}$$

The simulation results appear in Fig. 1. It is an average of 20 realizations of different measurement matrices $\mathbf{A}$. The correction term is seen to lead to improved variance predictions even at these small dimensions.

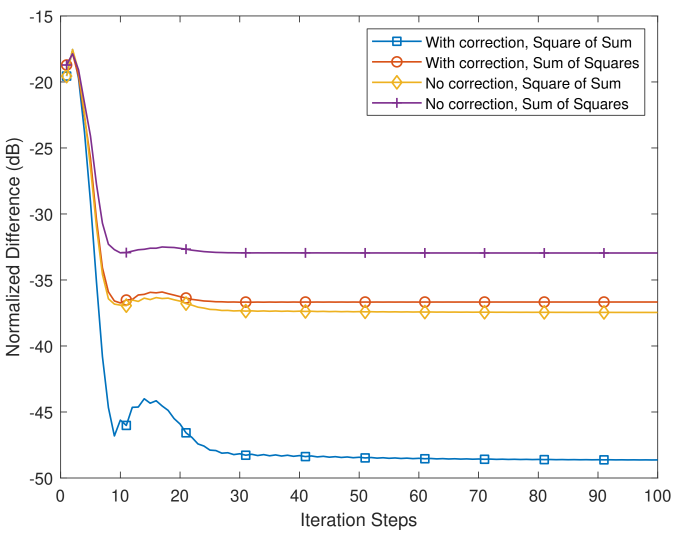

Fig. 1. MSE with or without correction.

10. CONCLUDING REMARKS

In this work, we have investigated the recovery of a sparse signal vector with a non-identically independently distributed arbitrary prior. This occurs typically in a compressed sensing problem but the approach is applicable to MMSE estimation in any GLM with Gaussian noise. The reVAMP algorithm we have introduced in [11] also solves this problem without any convergence or variance approximation issues, only involving the asymptotic Gaussian extrinsic approximation. However, the AMBUAMP proposed here addresses the high complexity issues in reVAMP for high dimensions. It combines the convergent AMP in [8] with the Haar LSA based variance correction in [9]. The Haar LSA is based on the underlying approximate multivariate Gaussian posterior approximation elucidated in [11]. The Haar LSA is formulated for the sum MSE. However, the expressions in (28) indicate that asymptotically also the individual component MSEs become correct, based on the same sum MSE variance correction. We have illustrated AMBUAMP for a GMM prior, leading to a GMM posterior. The efficient computation of MMSE estimates with Gaussian extrinsics and arbitrary prior deserves further work however.

11. REFERENCES

[1] Michael E Tipping, “Sparse Bayesian Learning and the Relevance Vector Machine,” *Journal of Machine Learning Research*, vol. 1, no. Jun, pp. 211–244, 2001.

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

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

[4] Philip Schniter, Sundeep Rangan, and Alyson K Fletcher, “Vector Approximate Message Passing for the Generalized Linear Model,” in *Asilomar Conf. Signals, Systems and Computers*, 2016.

[5] T. P. Minka, “Expectation Propagation for Approximate Bayesian Inference ,” in *in Proc. of Conf. on Uncert. in Artif. Intell. (UAI)*, 2001.

[6] Antonia M Tulino, Giuseppe Caire, Sergio Verdú, and Shlomo Shamai, “Support Recovery with Sparsely Sampled Free Random Matrices,” *IEEE Transactions on Information Theory*, vol. 59, no. 7, pp. 4243–4271, 2013.

[7] Qinghua Guo and Jiangtao Xi, “Approximate Message Passing with Unitary Transformation,” *arXiv preprint arXiv:1504.04799*, 2015.

[8] Christo Kurisummoottil Thomas, Zilu Zhao, and Dirk Slock, “Towards Convergent Approximate Message Passing by Alternating Constrained Minimization of Bethe Free Energy,” in *Proc. IEEE Information Theory Workshop (ITW)*, Apr. 2023.

[9] Zilu Zhao and Dirk Slock, “Variance Predictions in VAMP/UAMP with Right Rotationally Invariant Measurement Matrices for niid Generalized Linear Models,” in *Proc. European Signal Processing Conf. (EUSIPCO)*, Sept. 2023.

[10] Romain Couillet, Jakob Hoydis, and Mérouane Debbah, “Random Beamforming Over Quasi-Static and Fading Channels: A Deterministic Equivalent Approach,” *IEEE Trans. Information Theory*, Oct. 2012.

[11] Zilu Zhao and Fangqing Xiao Dirk 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.

[12] 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.

[13] Dirk TM Slock, “Nonlinear MMSE using Linear MMSE Bricks and Application to Compressed Sensing and Adaptive Kalman Filtering,” Int’l Conf. Acoustics, Speech, and Signal Processing (ICASSP), Expert to Non Expert (ETON) Primer, 2020.