Variance Predictions in VAMP/UAMP with Right Rotationally Invariant Measurement Matrices for niid Generalized Linear Models

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 and the measurements 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 algorithm 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 model was introduced in which the (right) singular vectors are Haar distributed. The associated extension of (G)AMP is the Vector (G)AMP ((G)VAMP) algorithm, which yields correct posterior variance predictions in the case of iid unknowns and measurements. However, due to averaging operations, these predictions become inexact in the case of niid unknowns and/or measurements. In this paper we apply Haar Large System Analysis (LSA) to characterize the variance prediction errors that can occur. We also introduce Unitary AMP (UAMP), which can continue to yield correct results with AMP style complexity.**

I. Introduction

The recovery of sparse signal vectors is a fundamental problem in signal processing with a wide range of applications, including compressive sensing, image and speech processing, and machine learning. The Gaussian signal model is commonly employed for the recovery of sparse signals due to its simplicity and effectiveness.

In the Gaussian case, the signal model for recovering a sparse signal vector $\boldsymbol{x}$ can be formulated as follows: $\boldsymbol{y} = \mathbf{A}\boldsymbol{x} + \boldsymbol{v}$, where $\boldsymbol{y}$ represents the observations or data. The matrix $\mathbf{A}$, known as the measurement or sensing matrix, has dimensions $M \times N$, where N/M is a constant typically greater than one. In the sparse model, the vector $\boldsymbol{x}$ contains only K non-zero (or significant) entries, where $K < M < N$.

Sparse Bayesian Learning (SBL) [1] is a well-known method for sparse signal recovery. However, SBL can be computationally demanding, particularly when dealing with high-dimensional data. This complexity is due to the requirement of performing matrix inversions within each iteration.

To address this challenge, approximation inference methods have been developed, and one popular and efficient approach is Approximate Message Passing (AMP) [2]. AMP is particularly effective for recovering high-dimensional signals, and its dynamics can be fully characterized through state evolution [3]. However, the convergence of AMP can become problematic when dealing with ill-conditioned measurement matrices $\mathbf{A}$.

To tackle this problem, the Vector AMP (VAMP) algorithm was introduced [4]. VAMP splits the variable node $\boldsymbol{x}$ into two variable nodes $\boldsymbol{x}_1 = \boldsymbol{x}_2 = \boldsymbol{x}$ within the factor graph. It utilizes an Expectation-Propagation (EP)-like message passing algorithm that iteratively operates on the factor graph using vector-valued messages. VAMP has demonstrated strong performance when dealing with right rotationally invariant (RRI) $\mathbf{A}$, and its state evolution has been rigorously established [4].

A. Prior Work

In previous research [4], the optimality of VAMP has been analyzed using the replica method [5]. The replica method yields a system of equations that describes the fixed point of the VAMP state evolution and the optimal (sum) mean squared error (MSE). While the (G)AMP method provides individual MSEs, its convergence is not guaranteed when dealing with ill-conditioned $\mathbf{A}$. To enhance the robustness of AMP to $\mathbf{A}$, the Unitary Transformation AMP (UTAMP) [6] transforms the linear model.

B. Main Contribution

In this paper, we propose a new method to perform large system analysis for the VAMP algorithm with niid Gaussian distributed prior $\boldsymbol{x}$ and RRI $\mathbf{A}$ by using the deterministic equivalent approach for Haar matrices [7]. We show that the optimal MSE and the VAMP posterior variances correspond to the fixed point of the same equations. By using large system analysis, it is also shown that with niid $\mathbf{A}$, (G)AMP also gives optimal individual MSE. A new method called Unitary AMP (UAMP) is proposed to provide the optimal individual MSEs as posterior variances. This method can be applied when $\mathbf{A}$ is RRI. We demonstrate that UAMP introduces a necessary correction term that allows it to reach the optimal MSE.

C. Notations

The symbol $\langle\cdot\rangle$ denotes the sample mean of a given set of data. The operations $\boldsymbol{x}.\boldsymbol{y}$ and $\boldsymbol{x}./\boldsymbol{y}$ represents the element-wise multiplication and division of two vectors. We use $D(\boldsymbol{\tau})$ to represent a diagonal matrix constructed from vector $\boldsymbol{\tau}$. We use $\mathbf{D}_M$ for a general diagonal matrix of dimension M.

II. Vector Approximate Message Passing

The data model considered in VAMP 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) niid prior $p_{\boldsymbol{x}}(\boldsymbol{x}) = \prod_{i=1}^{N} p_{x_i}(x_i)$ and iid measurements noise $p_{\boldsymbol{v}}(\boldsymbol{v}) = \prod_{i=1}^{N} p_v(v_i)$. Furthermore, we assume $M < N$. In [4], the authors proposed an expectation propagation [8] (EP)-like derivation to derive VAMP. Splitting $\boldsymbol{x}$ into two identical random vectors $\boldsymbol{x}_1 = \boldsymbol{x}_2$ gives an factorization

$$p(\boldsymbol{y}, \boldsymbol{x}_1, \boldsymbol{x}_2) = p(\boldsymbol{x}_1)p(y|\boldsymbol{x}_2)\delta(\boldsymbol{x}_1 - \boldsymbol{x}_2), \tag{2}$$

Algorithm 1 VAMP

Require: $\boldsymbol{y}$, $\mathbf{A}$, $p_{\boldsymbol{x}}(\boldsymbol{x})$, $p_{\boldsymbol{v}}(\boldsymbol{v})$
Initialize: $\mathbf{r}_1^0 = \mathbf{0}$, $\gamma_r^0 = 1$
repeat
  [Estimate $\boldsymbol{x}_1$]
  $\widehat{\boldsymbol{x}}_1^t = \mathbf{g}_1(\mathbf{r}_1^t, \gamma_r^t)$
  $\gamma_1^t = \gamma_r^t \langle \mathbf{g}_1'(\mathbf{r}_1^t, \gamma_r^t)\rangle$
  [Generate messages for $\boldsymbol{x}_2$]
  $\mathbf{r}_2^t = (\gamma_r^t \widehat{\boldsymbol{x}}_1^t - \gamma_1^t \mathbf{r}_1)/(\gamma_r^t - \gamma_1^t)$
  $\gamma_p^t = \gamma_1^t \gamma_r^t/(\gamma_r^t - \gamma_1^t)$
  [Estimate $\boldsymbol{x}_2$]
  $\widehat{\boldsymbol{x}}_2^t = \mathbf{g}_2(\mathbf{r}_2^t, \gamma_p^t)$
  $\gamma_2^t = \gamma_p^t \langle \mathbf{g}_2'(\mathbf{r}_2^t, \gamma_p^t)\rangle$
  [Generate messages for $\boldsymbol{x}_1$]
  $\mathbf{r}_1^{t+1} = (\gamma_p^t \widehat{\boldsymbol{x}}_2^t - \gamma_2^t \mathbf{r}_2)/(\gamma_p^t - \gamma_2^t)$
  $\gamma_r^{t+1} = \gamma_2^t \gamma_p^t/(\gamma_p^t - \gamma_2^t)$
until Convergence

where $\delta(\cdot)$ is the Dirac delta distribution. The VAMP is illustrated in Algorithm 1.

If we consider the Gaussian case, with niid Gaussian $p_{\boldsymbol{x}}(\boldsymbol{x})$ and iid Gaussian $p_{\boldsymbol{v}}(\boldsymbol{v})$, we have the following distributions:

$$\boldsymbol{x} \sim \mathcal{N}\left(\mathbf{0}, D(\boldsymbol{\sigma}_x^2)\right), \, \boldsymbol{v} \sim \mathcal{N}(\mathbf{0}, \sigma_v^2 \mathbf{I}), \tag{3}$$

where $D(\cdot)$ denotes a diagonal matrix constructed from the vector argument, and $\boldsymbol{\sigma}_x^2 = \begin{bmatrix}\sigma_{x1}^2 & \ldots & \sigma_{xN}^2\end{bmatrix}^T$. With MMSE estimation, the function $\mathbf{g}_1(\mathbf{r}_1, \gamma_r)$ can be interpreted as the MMSE estimation given the prior $\boldsymbol{x}_1 \sim \mathcal{N}(\mathbf{r}_1, \gamma_r \mathbf{I})$ and likelihood $\mathcal{N}(\mathbf{0}; \boldsymbol{x}_1, D(\boldsymbol{\sigma}_x^2))$. Similarly, function $\mathbf{g}_2(\mathbf{r}_2, \gamma_p)$ can be interpreted as the MMSE estimation with prior $\boldsymbol{x}_2 \sim \mathcal{N}(\mathbf{r}_2, \gamma_p \mathbf{I})$ and likelihood $\mathcal{N}(\boldsymbol{y}; \mathbf{A}\, \boldsymbol{x}_2, \sigma_v^2 \mathbf{I})$.

Thus, $\mathbf{g}_1(\mathbf{r}_1, \gamma_r)$ and $\mathbf{g}_2(\mathbf{r}_2, \gamma_p)$ are

$$\begin{aligned}
&\mathbf{g}_1(\mathbf{r}_1, \gamma_r) = \boldsymbol{\sigma}_x^2 . \mathbf{r}_1 ./(\gamma_r \mathbf{1} + \boldsymbol{\sigma}_x^2), \\
&\mathbf{g}_2(\mathbf{r}_2, \gamma_p) = \left(\frac{1}{\sigma_v^2}\mathbf{A}^T\mathbf{A} + \frac{1}{\gamma_p}\mathbf{I}\right)^{-1} \left(\frac{1}{\sigma_v^2}\mathbf{A}^T \boldsymbol{y} + \frac{1}{\gamma_p}\mathbf{r}_2\right).
\end{aligned} \tag{4}$$

Note that posterior variances of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ can be calculated from the derivative of the MMSE estimators w.r.t. their first arguments. However, in order to reduce computational complexity, the VAMP algorithm approximates the posterior variances as multiples of identity matrices. This results

$$\begin{aligned}
\gamma_1 &= \frac{1}{N}\sum_{i=1}^{N} \frac{\sigma_{xi}^2 \gamma_r}{\sigma_{xi}^2 + \gamma_r}, \\
\gamma_2 &= \frac{1}{N}\mathrm{tr}\left[\left(\frac{1}{\sigma_v^2}\mathbf{A}^T\mathbf{A} + \frac{1}{\gamma_p}\mathbf{I}\right)^{-1}\right].
\end{aligned} \tag{5}$$

III. Large System Analysis For VAMP .

Following the approach described in [4], we model $\mathbf{A}$ as a RRI matrix. We obtain the economy singular value decomposition (SVD) of $\mathbf{A}$ as follows:

$$\mathbf{A} = \mathbf{U}\overline{\boldsymbol{\Sigma}}\,\overline{\mathbf{V}}^T, \tag{6}$$

where $\mathbf{U} \in \mathbb{R}^{M\times M}$ is an orthogonal deterministic matrix, $\overline{\boldsymbol{\Sigma}} = D(\boldsymbol{\sigma}) \in \mathbb{R}^{M\times M}$ is a diagonal deterministic matrix with $\boldsymbol{\sigma} = \begin{bmatrix}\sigma_1 & \ldots & \sigma_M\end{bmatrix}^T$ and $\overline{\mathbf{V}} \in \mathbb{R}^{M\times N}$ is obtained by selecting M columns from a Haar-distributed $N \times N$ random matrix. The analysis of the large system primarily relies on the deterministic equivalent proposed in [9], 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,*

$$\frac{1}{N}tr\left[\mathbf{B}\left(\mathbf{V}\boldsymbol{P}\mathbf{V}^T + \mathbf{D}\right)^{-1}\right] - \frac{1}{N}tr\left[\mathbf{B}(\overline{e}\,\mathbf{I} + \mathbf{D})^{-1}\right] \xrightarrow{a.s.} 0. \tag{7}$$

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

$$\begin{aligned}
\overline{e} &= \frac{1}{N}tr\left[\boldsymbol{P}\left(e\boldsymbol{P} + (1 - e\,\overline{e})\mathbf{I}\right)^{-1}\right], \\
e &= \frac{1}{N}tr\left[\mathbf{B}(\overline{e}\mathbf{I} + \mathbf{D})^{-1}\right].
\end{aligned} \tag{8}$$

The MMSE solution for (1) is given by

$$\begin{aligned}
\widehat{\boldsymbol{x}}_{\mathrm{MMSE}} &= \left(\frac{1}{\sigma_v^2}\mathbf{A}^T\mathbf{A} + D\left(1./\boldsymbol{\sigma}_x^2\right)\right)^{-1}\frac{1}{\sigma_v^2}\mathbf{A}^T\boldsymbol{y}, \\
\mathbf{C}_{\mathrm{MMSE}} &= \left(\frac{1}{\sigma_v^2}\mathbf{A}^T\mathbf{A} + D\left(1./\boldsymbol{\sigma}_x^2\right)\right)^{-1}.
\end{aligned} \tag{9}$$

Thus, the MSE is

$$\mathrm{MSE} = \frac{1}{N}\mathrm{tr}[\mathbf{C}_{\mathrm{MMSE}}]. \tag{10}$$

By using Lemma 1, we obtain the large system approximation

$$\mathrm{MSE} \xrightarrow{\text{a.s.}} \frac{1}{N}\mathrm{tr}\left[(\overline{e}^0\mathbf{I} + D(1./\boldsymbol{\sigma}_x^2))^{-1}\right] \tag{11}$$

with

$$\begin{aligned}
\overline{e}^0 &= \frac{1}{N}\mathrm{tr}\left[\frac{1}{\sigma_v^2}\overline{\boldsymbol{\Sigma}}^2\left(\frac{e^0}{\sigma_v^2}\overline{\boldsymbol{\Sigma}}^2 + (1 - e^0\overline{e}^0)\mathbf{I}\right)^{-1}\right], \\
e^0 &= \frac{1}{N}\mathrm{tr}\left[(\overline{e}^0\mathbf{I} + D(1./\boldsymbol{\sigma}_x^2))^{-1}\right].
\end{aligned} \tag{12}$$

Now we try to derive the steady state of niid Gaussian prior VAMP. At convergence, we have

$$\begin{aligned}
\gamma_1^\infty &= \frac{1}{N}\sum_{i=1}^{N}\frac{\sigma_{xi}^2\gamma_r^\infty}{\sigma_{xi}^2 + \gamma_r^\infty} \\
\gamma_p^\infty &= \frac{\gamma_1^\infty\gamma_r^\infty}{\gamma_r^\infty - \gamma_1^\infty} \\
\gamma_2^\infty &= \frac{1}{N}\mathrm{tr}\left[\left(\frac{1}{\sigma_v^2}\mathbf{A}^T\mathbf{A} + \frac{1}{\gamma_p^\infty}\right)^{-1}\right] \\
\gamma_r^\infty &= \frac{\gamma_2^\infty\gamma_p^\infty}{\gamma_p^\infty - \gamma_2^\infty}.
\end{aligned} \tag{13}$$

We first examine the extrinsic γ_p^∞ and γ_r^∞. From (13), we have

$$\frac{1}{\gamma_1^\infty} = \frac{1}{\gamma_p^\infty} - \frac{1}{\gamma_r^\infty} = \frac{1}{\gamma_2^\infty}. \tag{14}$$

Apply Lemma 1 to find the deterministic equivalent

$$\gamma_2^{\infty} = \frac{\gamma_p^{\infty}}{\overline{e}^{\infty}\gamma_p^{\infty} + 1}; \tag{15}$$

with

$$\begin{aligned}\overline{e}^{\infty} &= \frac{1}{N}\text{tr}\left[\frac{1}{\sigma_v^2}\overline{\mathbf{\Sigma}}^2\left(\frac{e^{\infty}}{\sigma_v^2}\overline{\mathbf{\Sigma}}^2 + (1 - e^{\infty}\overline{e}^{\infty})\mathbf{I}\right)^{-1}\right], \\ e^{\infty} &= \gamma_2^{\infty} = \frac{\gamma_p^{\infty}}{\overline{e}^{\infty}\gamma_p^{\infty} + 1}.\end{aligned} \tag{16}$$

Combining equations (13), (15), and (16), we can conclude that

$$e^{\infty} = \gamma_1^{\infty} = \frac{1}{N}\text{tr}\left[(\overline{e}^{\infty}\mathbf{I} + D(1./\boldsymbol{\sigma}_x^2))^{-1}\right]. \tag{17}$$

The system of equations describing the relation between $(\overline{e}^0, e^0)$ is equivalent to the system of equations corresponding to $(\overline{e}^{\infty}, e^{\infty})$. This implies that if the system of equations in (12) has a unique solution, then applying VAMP to linear systems with niid Gaussian prior results in the optimal MSE. In the case where the prior distribution of $\boldsymbol{x}$ is iid Gaussian, $\boldsymbol{x} \sim \mathcal{N}(\mathbf{0}, \sigma_x^2\mathbf{I})$, we examine the message generated for the node $\boldsymbol{x}_2$ as follows:

$$\mathbf{r}_2^t = \mathbf{0};\ \gamma_p^t = \sigma_x^2. \tag{18}$$

This implies that regardless of the matrix $\mathbf{A}$ and the initialization, $\widehat{\boldsymbol{x}}_2$ always returns the MMSE estimate of $\boldsymbol{x}$, and γ_2 always provides the MSE.

IV. Unitary AMP

The VAMP algorithm can only provide the optimal (sum) MSE. However, we are currently exploring a method to obtain the optimal element-wise MSE of $\boldsymbol{x}$. To achieve this, we employ a similar method as proposed in [6], where we transform the linear model by left-multiplying the matrix $\mathbf{U}^T$:

$$\mathbf{U}^T\boldsymbol{y} = \overline{\mathbf{\Sigma}}\overline{\mathbf{V}}^T\boldsymbol{x} + \mathbf{U}^T\boldsymbol{v}, \tag{19}$$

where $\mathbf{U}$, $\overline{\mathbf{\Sigma}}$, and $\overline{\mathbf{V}}$ are obtained by applying economy SVD on $\mathbf{A}$. Since we assume $\boldsymbol{v}$ to be iid Gaussian, the equivalent noise remains invariant under orthogonal transformation. Thus, we have $\mathbf{U}^T\boldsymbol{v} \sim \mathcal{N}(0, \sigma_v^2\mathbf{I})$.

For simplicity, let's denote $\boldsymbol{y}' = \mathbf{U}^T\boldsymbol{y}$, $\boldsymbol{v}' = \mathbf{U}^T\boldsymbol{v}$, $\boldsymbol{\sigma}_v^2 = \sigma_v^2\mathbf{1}$, $\mathbf{A}' = \overline{\mathbf{\Sigma}}\mathbf{V}^T$, $\mathbf{S}' = \mathbf{A}'.\mathbf{A}'$ and $\boldsymbol{\lambda} = \overline{\mathbf{\Sigma}}^2\mathbf{1}$.

To state the AMP algorithm more easily, we also reformulate (1) as

$$\mathbf{z} = \mathbf{A}'\boldsymbol{x},\ \boldsymbol{y}' = \mathbf{z} + \boldsymbol{v}'. \tag{20}$$

Now, we will utilize the AMP algorithm [10] [11] for this model and summarize it in Algorithm 2. In this algorithm, we introduce $\boldsymbol{\tau}_x$ and $\boldsymbol{\tau}_z$ as the posterior variances of $\boldsymbol{x}$ and $\mathbf{z}$, respectively, similar to γ_1 and γ_2 in Algorithm 1.

V. Large System Analysis of UAMP

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 (10). Then in the next section, we will propose a correction term such that $\boldsymbol{\tau}_x$ matches the

Algorithm 2 UAMP

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

Initialize: $t = 0$, $\widehat{\boldsymbol{x}}^t$, $\boldsymbol{\tau}_x^t$, $\mathbf{s}^{t-1} = \mathbf{0}$

repeat

  [Output node update]

  $\boldsymbol{\tau}_p^t = \mathbf{S}'\,\boldsymbol{\tau}_x^t$

  $\boldsymbol{p}^t = \mathbf{A}'\,\widehat{\boldsymbol{x}}^t - \mathbf{s}^{t-1}.\boldsymbol{\tau}_p^t$

  $\widehat{\mathbf{z}}^t = \boldsymbol{p}^t.\boldsymbol{\sigma}_v^2./(\boldsymbol{\sigma}_v^2 + \boldsymbol{\tau}_p^t) + \boldsymbol{y}.\boldsymbol{\tau}_p^t./(\boldsymbol{\sigma}_v^2 + \boldsymbol{\tau}_p^t)$

  $\boldsymbol{\tau}_z^t = \boldsymbol{\sigma}_v^2.\boldsymbol{\tau}_p^t./(\boldsymbol{\sigma}_v^2 + \boldsymbol{\tau}_p^t)$

  $\mathbf{s}^t = (\widehat{\mathbf{z}}^t - \boldsymbol{p}^t)./\boldsymbol{\tau}_p^t$

  $\boldsymbol{\tau}_s^t = \mathbf{1}./(\boldsymbol{\sigma}_v^2 + \boldsymbol{\tau}_p^t)$

  [Input node update]

  $\boldsymbol{\tau}_r^t = \mathbf{1}./(\mathbf{S}'^T\boldsymbol{\tau}_s^t)$

  $\mathbf{r}^t = \widehat{\boldsymbol{x}}^t + \boldsymbol{\tau}_r^t.\mathbf{A}'^T\mathbf{s}^t$

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

  $\boldsymbol{\tau}_x^{t+1} = \boldsymbol{\tau}_r^t.\boldsymbol{\sigma}_x^2./(\boldsymbol{\sigma}_x^2 + \boldsymbol{\tau}_r^t)$

until Convergence

optimal MSE. The steady state of variances in UAMP can be summarized as follows

$$\begin{aligned}\mathbf{1}./\boldsymbol{\tau}_s^{\infty} &= \boldsymbol{\sigma}_v^2 + \mathbf{S}'\boldsymbol{\tau}_x^{\infty} \\ \mathbf{1}./\boldsymbol{\tau}_x^{\infty} &= \mathbf{1}./\boldsymbol{\sigma}_x^2 + \mathbf{S}'^T\boldsymbol{\tau}_s^{\infty}.\end{aligned} \tag{21}$$

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

$$\begin{aligned}\mathbf{S}'\boldsymbol{\tau}_x^{\infty} &= \text{diag}\left[\overline{\mathbf{\Sigma}}\overline{\mathbf{V}}^T D(\boldsymbol{\tau}_x^{\infty})\overline{\mathbf{V}}\overline{\mathbf{\Sigma}}\right] = \frac{1}{N}\mathbf{1}^T\boldsymbol{\tau}_x^{\infty}\,\boldsymbol{\lambda}, \\ \mathbf{S}'^T\boldsymbol{\tau}_s^{\infty} &= \text{diag}\left[\overline{\mathbf{V}}\overline{\mathbf{\Sigma}}D(\boldsymbol{\tau}_s^{\infty})\overline{\mathbf{\Sigma}}\overline{\mathbf{V}}^T\right] = \frac{1}{N}\boldsymbol{\lambda}^T\boldsymbol{\tau}_s^{\infty}\,\mathbf{1}.\end{aligned} \tag{22}$$

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 (11).*

Proof. If the noise is iid, the MSE remains unchanged under a unitary transformation. Therefore, equations (11) and (12) 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

$$\begin{aligned}\frac{1}{N}\text{tr}[D(\boldsymbol{\tau}_x^{\infty})] &= \frac{1}{N}\text{tr}\left[\left(D(\mathbf{S}'^T\boldsymbol{\tau}_s^{\infty}) + D(1./\boldsymbol{\sigma}_x^2)\right)^{-1}\right] \\ &= e^0 = \frac{1}{N}\text{tr}\left[(\overline{e}^0\mathbf{I} + D(1./\boldsymbol{\sigma}_x^2))^{-1}\right],\end{aligned} \tag{23}$$

which implies

$$\begin{aligned}\overline{e}^0 &= \frac{1}{N}\boldsymbol{\lambda}^T\boldsymbol{\tau}_s^{\infty} = \frac{1}{N}\text{tr}\left[\overline{\mathbf{\Sigma}}^2 D(\boldsymbol{\tau}_s^{\infty})\right] \\ &= \frac{1}{N}\text{tr}\left[\overline{\mathbf{\Sigma}}^2\left(\frac{1}{N}\text{tr}[D(\boldsymbol{\tau}_x^{\infty})]D(\boldsymbol{\lambda}) + \sigma_v^2\mathbf{I}\right)^{-1}\right] \\ &= \frac{1}{N}\text{tr}\left[\overline{\mathbf{\Sigma}}^2\left(e^0\overline{\mathbf{\Sigma}}^2 + \sigma_v^2\mathbf{I}\right)^{-1}\right]\end{aligned} \tag{24}$$

One can observe that $\overline{e}^0$ in (24) only equals $\overline{e}^0$ in (12) if $e^0\overline{e}^0 = 0$. □

VI. Correction Term For τ_s

Considering the asymptotic MSE expression in (11), and the second equations in (21), (22), we can still write

$$\begin{aligned}\frac{1}{N}\text{tr}[D(\boldsymbol{\tau}_x^\infty)] &= \frac{1}{N}\text{tr}\left[\left(D(\mathbf{S}'^T\boldsymbol{\tau}_s^\infty)+D(1./\boldsymbol{\sigma}_x^2)\right)^{-1}\right]\\ &= \frac{1}{N}\text{tr}\left[\left(\frac{1}{N}\boldsymbol{\lambda}^T\boldsymbol{\tau}_s^\infty\mathbf{I}+D(1./\boldsymbol{\sigma}_x^2)\right)^{-1}\right]\end{aligned} \quad (25)$$

Introduce

$$e_c = \frac{1}{N}\text{tr}[D(\boldsymbol{\tau}_x^\infty)]\,, \overline{e}_c = \frac{1}{N}\boldsymbol{\lambda}^T\boldsymbol{\tau}_s^\infty. \quad (26)$$

Now compare (25),(26) with (11),(12), then we require $\overline{e}_c$ to be of the form

$$\overline{e}_c = \frac{1}{N}\text{tr}\left[\overline{\boldsymbol{\Sigma}}^2\left(e_c\overline{\boldsymbol{\Sigma}}^2+(1-e_c\overline{e}_c)\sigma_v^2\mathbf{I}\right)^{-1}\right]. \quad (27)$$

From the definition of $\overline{e}_c$ in (26), we have

$$\overline{e}_c = \frac{1}{N}\boldsymbol{\lambda}^T\boldsymbol{\tau}_s^\infty = \frac{1}{N}\text{tr}\left[\overline{\boldsymbol{\Sigma}}^2 D(\boldsymbol{\tau}_s^\infty)\right]. \quad (28)$$

Comparing (28) with (27), we want to design the update scheme of τ_s such that at steady state,

$$D(\boldsymbol{\tau}_s^\infty) = \left(e_c\overline{\boldsymbol{\Sigma}}^2+(1-e_c\overline{e}_c)\sigma_v^2\mathbf{I}\right)^{-1}. \quad (29)$$

From the definition of e_c in (26), we have

$$e_c\overline{\boldsymbol{\Sigma}}^2 = \frac{1}{N}\text{tr}[D(\boldsymbol{\tau}_x)]\overline{\boldsymbol{\Sigma}}^2 = \frac{1}{N}\mathbf{1}^T\boldsymbol{\tau}_x\,\overline{\boldsymbol{\Sigma}}^2 = \frac{1}{N}\mathbf{1}^T\boldsymbol{\tau}_x\,D(\boldsymbol{\lambda}). \quad (30)$$

Under the large system approximation (22), we obtain

$$e_c\overline{\boldsymbol{\Sigma}}^2 = D(\frac{1}{N}\mathbf{1}^T\boldsymbol{\tau}_x\,\boldsymbol{\lambda}) = D(\mathbf{S}'\boldsymbol{\tau}_x^\infty). \quad (31)$$

Substituting (26) and (31) into (29), 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 line 9 of Algorithm 2

$$\boldsymbol{\tau}_s^t = \mathbf{1}./\left[\boldsymbol{\sigma}_v^2-\frac{1}{N^2}\left(\mathbf{1}^T\boldsymbol{\tau}_x^t\right)\left(\boldsymbol{\lambda}^T\boldsymbol{\tau}_s^{t-1}\right)\boldsymbol{\sigma}_v^2+\boldsymbol{\tau}_p^t\right] \quad (32)$$

One can verify by Lemma 1 that with this correction, $(e^c,\overline{e}^c)$ converge to a fixed point of (12) and hence $\boldsymbol{\tau}_x$ will converge to the optimal MSE.

VII. Relation to AMP

The original AMP algorithm can be derived by substituting all instances of $\mathbf{A}'$ with $\mathbf{A}$ and all occurrences of $\mathbf{S}'$ with $\mathbf{S} = \mathbf{A}.\mathbf{A}$ in Algorithm 2. We will now demonstrate that in a more general scenario with niid matrix $\mathbf{A}$ and niid noise signals $\boldsymbol{v}$, $\boldsymbol{\tau}_x^\infty$ provides the optimal individual MSE. To analyze the steady state of AMP, we rely on the theorem presented in [12].

Theorem 1. *Let $\boldsymbol{Q}_N \in \mathbb{C}^{N\times N}$ be a Hermitian deterministic matrix and $\mathbf{A}_N = \mathbf{X}_N\mathbf{D}\mathbf{X}_N^H = \sum_{i=1}^M d_i\boldsymbol{x}_i\boldsymbol{x}_i^H$, with diagonal $\mathbf{D}$ and $\mathbf{X}_N$ containing M independent columns $\boldsymbol{x}_i$ with covariance matrix $\boldsymbol{\Theta}_i$. Also, assume that $\boldsymbol{Q}_N$, $\boldsymbol{\Theta}_i$ and $\mathbf{D}'_N$ have uniformly bounded spectral norms. Then, as $M, N \to \infty$ at constant ratio*

$$\begin{aligned}&\frac{1}{N}tr\left[\boldsymbol{Q}_N(\mathbf{A}_N+\mathbf{D}'_N)^{-1}\right]-\frac{1}{N}tr\left[\boldsymbol{Q}_N\mathbf{T}\right]\xrightarrow{a.s.}0,\ \textit{with}\\ &\mathbf{T} = \left(\sum_{i=1}^M\frac{d_i\boldsymbol{\Theta}_i}{1+e_i}+\mathbf{D}'_N\right)^{-1},\ \textit{where}\\ &e_k = tr\left[d_k\boldsymbol{\Theta}_k\left(\sum_{i=1}^M\frac{d_i\boldsymbol{\Theta}_i}{1+e_i}+\mathbf{D}'_N\right)^{-1}\right].\end{aligned} \quad (33)$$

We now show that AMP with niid $\mathbf{A}$ leads to correct variance predictions.

Proof. We assume the columns of $\mathbf{A}^T = [\boldsymbol{a}_1\ldots\boldsymbol{a}_M]$ to be zero mean and independent with diagonal covariance matrix $\mathbb{E}\left(\boldsymbol{a}_i\boldsymbol{a}_i^T\right) = \boldsymbol{\Theta}_i$. The optimal MSE is given by (10). By applying Theorem 1 with an arbitrary diagonal weighting matrix $\boldsymbol{Q}_N$, we obtain the weighted optimal MSE (WMSE):

$$\begin{aligned}\text{WMSE}(\boldsymbol{Q}_N) &= \frac{1}{N}\text{tr}\left[\boldsymbol{Q}_N\left(\mathbf{A}^TD(\mathbf{1}./\boldsymbol{\sigma}_v^2)\mathbf{A}+D(\mathbf{1}./\boldsymbol{\sigma}_x^2)\right)^{-1}\right]\\ &\xrightarrow{\text{a.s.}}\frac{1}{N}\text{tr}\left[\boldsymbol{Q}_N\left(\sum_{i=1}^M\frac{\boldsymbol{\Theta}_i}{\sigma_{v,i}^2(1+e_i)}+D(\mathbf{1}./\boldsymbol{\sigma}_x^2)\right)^{-1}\right]\end{aligned} \quad (34)$$

with

$$e_k = \text{tr}\left[\frac{\boldsymbol{\Theta}_k}{\sigma_{v,k}^2}\left(\sum_{i=1}^M\frac{\boldsymbol{\Theta}_i}{\sigma_{v,i}^2(1+e_i)}+D(\mathbf{1}./\boldsymbol{\sigma}_x^2)\right)^{-1}\right]. \quad (35)$$

On the other hand, the steady-state equations of variances in AMP can also be represented as

$$\begin{aligned}\mathbf{1}./\boldsymbol{\tau}_s^\infty &= \boldsymbol{\sigma}_v^2+\mathbf{S}\boldsymbol{\tau}_x^\infty\\ \mathbf{1}./\boldsymbol{\tau}_x^\infty &= \mathbf{1}./\boldsymbol{\sigma}_x^2+\mathbf{S}^T\boldsymbol{\tau}_s^\infty.\end{aligned} \quad (36)$$

With large $\mathbf{A}$, $\mathbf{S}\boldsymbol{\tau}_x^\infty$ and $\mathbf{S}^T\boldsymbol{\tau}_s^\infty$ converge to their expected values

$$\mathbb{E}\left[\mathbf{S}\boldsymbol{\tau}_x^\infty\right]_i = \mathbb{E}\left[\mathbf{A}D(\boldsymbol{\tau}_x^\infty)\mathbf{A}^T\right]_{ii} = \text{tr}[\boldsymbol{\Theta}_iD(\boldsymbol{\tau}_x^\infty)]; \quad (37)$$

$$\mathbb{E}\,D\left(\mathbf{S}^T\boldsymbol{\tau}_s^\infty\right) = \mathbb{E}\text{ diag}\left(\mathbf{A}^TD(\boldsymbol{\tau}_s^\infty)\mathbf{A}\right) = \sum_{k=1}^M\tau_{s,k}^\infty\boldsymbol{\Theta}_k. \quad (38)$$

Therefore, the weighted mean of the posterior $\boldsymbol{\tau}_x^\infty$ with the same weighting matrix $\boldsymbol{Q}_N$ becomes

$$\frac{1}{N}\text{tr}[\boldsymbol{Q}_ND(\boldsymbol{\tau}_x^\infty)] = \frac{1}{N}\text{tr}\left[\boldsymbol{Q}_N\left(D(\mathbf{1}./\boldsymbol{\sigma}_x^2)+\sum_{k=1}^M\tau_{s,k}^\infty\boldsymbol{\Theta}_k\right)^{-1}\right]. \quad (39)$$

From (36) and (38), we obtain

$$\tau_{s,k}^\infty = \frac{1}{\sigma_{v,k}+\text{tr}[\boldsymbol{\Theta}_kD(\boldsymbol{\tau}_x^\infty)]}. \quad (40)$$

Define $e'_k = \frac{\text{tr}[\boldsymbol{\Theta}_kD(\boldsymbol{\tau}_x^\infty)]}{\sigma_{v,k}^2}$ and substituting (40) into (39), we obtain

$$\frac{1}{N}\text{tr}[\boldsymbol{Q}_N\mathbf{D}(\boldsymbol{\tau}_x^{\infty})]$$
$$= \frac{1}{N}\text{tr}\left[\boldsymbol{Q}_N\left(D(\mathbf{1}./\boldsymbol{\sigma}_x^2)+\sum_{i=1}^{M}\frac{\boldsymbol{\Theta}_i}{\sigma_{v,i}(1+e_i')}\right)^{-1}\right]; \quad (41)$$
$$e_k' = \text{tr}\left[\frac{\boldsymbol{\Theta}_k}{\sigma_{v,k}^2}\left[D(\mathbf{1}./\boldsymbol{\sigma}_x^2)+\sum_{i=1}^{M}\frac{\boldsymbol{\Theta}_i}{\sigma_{v,i}(1+e_i')}\right]^{-1}\right\}.$$

We can now observe that for any diagonal weighting matrix $\boldsymbol{Q}_N$, the weighted mean of $\boldsymbol{\tau}_x^{\infty}$ is equal to the weighted optimal MSE. □

VIII. Relation to Unitary Transformation AMP

In [6], the authors also proposed an AMP method based on a unitarily transformed model. However, due to the complexity limitations, they used the large system approximation in every iteration. UAMP can be modified to UTAMP by applying (22) to line 4 and line 12 in Algorithm 2, which then become resp.

$$\boldsymbol{\tau}_p^t = \frac{1}{N}(\mathbf{1}^T\boldsymbol{\tau}_x^t)\boldsymbol{\lambda}\,,\ \boldsymbol{\tau}_r^t = \frac{1}{\frac{1}{N}(\boldsymbol{\lambda}^T\boldsymbol{\tau}_s^t)}\mathbf{1}. \quad (42)$$

The same proof as in section V can be used here to show that averaged $\boldsymbol{\tau}_x^{\infty}$ does not converge to the optimal MSE. However, we can apply the same correction stated in (32).

IX. Simulation Results

In our simulation, we considered a sparse signal recovery problem with niid Gaussian priors. The system dimensions were set to $M \times N = 512 \times 1024$, and the signal-to-noise ratio (SNR) was fixed at 0dB. The noise was assumed to be white Gaussian with unit variance. To ensure a specific condition number for the measurement matrix $\mathbf{A}$, we followed the approach described in [4]. We set the condition number as $\frac{s_1}{s_M} = 1000$, where s_1 and s_M represent the largest and smallest singular values of $\mathbf{A}$, respectively. The sequence of singular values $\mathbf{s} = [s_1, \ldots, s_M]^T$ was generated as a geometric sequence. The measurement matrix $\mathbf{A}$ was constructed using the economy SVD: $\mathbf{A} = \mathbf{U}D(\mathbf{s})\overline{\mathbf{V}}^T$, where the singular vector matrices $\mathbf{U}$ and $\overline{\mathbf{V}}$ were drawn from a Haar distribution. In order to compare the individual MSE of different methods, we define

$$r_{diff} = \frac{(\boldsymbol{\tau}_x - \boldsymbol{\tau}_{\text{MMSE}})^T(\boldsymbol{\tau}_x - \boldsymbol{\tau}_{\text{MMSE}})}{\boldsymbol{\tau}_{\text{MMSE}}^T\boldsymbol{\tau}_{\text{MMSE}}}, \quad (43)$$

where $\boldsymbol{\tau}_{\text{MMSE}} = \text{diag}(\mathbf{C}_{MMSE})$. In Figure 1, we present the results of our comparison between two cases: one where the variances of $\boldsymbol{\sigma}_x$ are set to be $\mathbf{1}$, and the other where the variances follow an exponential decay with a base of 0.991.

X. Concluding Remarks

In this study, we have investigated the recovery of a sparse signal vector with niid priors. The VAMP algorithm assumes the measurement matrix $\mathbf{A}$ to be RRI. Thus, we study the Haar large system analysis based on Lemma 1. We find out that at convergence, we have equality of the MSEs for the two VAMP nodes $\gamma_1^{\infty} = \gamma_2^{\infty}$. Furthermore, they both correspond to the optimal MSE. In the trivial case, in which the prior distribution of $\boldsymbol{x}$ is iid Gaussian, γ_2 will become the optimal MSE in the first iteration.

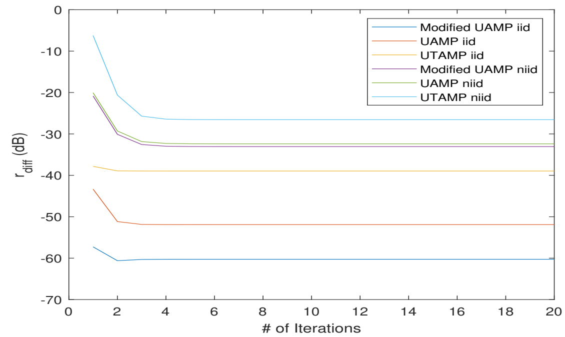

Fig. 1. Variance differences to LMMSE.

We have proposed UAMP. It is similar to UTAMP, but UAMP does not approximate the posterior variances of $\boldsymbol{x}$ to be equal (in which case VAMP becomes of high complexity). We then perform the new Haar based large system analysis to UAMP. In order for it to give the optimal MSE value, a correction is needed for the update of $\boldsymbol{\tau}_s^t$. The same correction terms could also be added in UTAMP, to give optimal (sum) MSE (whereas UAMP may provide correct element-wise MSEs).

Acknowledgements

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

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 *2016 50th Asilomar Conference on Signals, Systems and Computers*. IEEE, 2016, pp. 1525–1529.
[5] 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.
[6] Qinghua Guo and Jiangtao Xi, "Approximate Message Passing with Unitary Transformation," *arXiv preprint arXiv:1504.04799*, 2015.
[7] Romain Couillet, Jakob Hoydis, and Mérouane Debbah, "Random Beamforming Over Quasi-Static and Fading Channels: A Deterministic Equivalent Approach," *IEEE Transactions on Information Theory*, vol. 58, no. 10, pp. 6392–6425, 2012.
[8] 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.
[9] Dirk TM Slock, "Nonlinear MMSE using Linear MMSE Bricks and Application to Compressed Sensing and Adaptive Kalman Filtering," ICASSP 2020, 45th International Conference on Acoustics, Speech, and Signal Processing, Expert to Non Expert (ETON) Primer, 4-8 May 2020, Barcelona, Spain, 2020.
[10] Sundeep Rangan, "Generalized Approximate Message Passing for Estimation with Random Linear Mixing," in *2011 IEEE International Symposium on Information Theory Proceedings*. IEEE, 2011, pp. 2168–2172.
[11] Sundeep Rangan, Philip Schniter, Erwin Riegler, Alyson K Fletcher, and Volkan Cevher, "Fixed Points of Generalized Approximate Message Passing with Arbitrary Matrices," *IEEE Transactions on Information Theory*, vol. 62, no. 12, pp. 7464–7474, 2016.
[12] Sebastian Wagner, Romain Couillet, Mérouane Debbah, and Dirk TM Slock, "Large System Analysis of Linear Precoding in Correlated MISO Broadcast Channels under Limited Feedback," *IEEE Transactions on Information Theory*, vol. 58, no. 7, pp. 4509–4537, 2012.