Bilinear Generalized Vector Approximate Message Passing

Mohamed Akrout, Anis Housseini, Faouzi Bellili, *Member, IEEE*, and Amine Mezghani, *Member, IEEE*

E2-390 E.I.T.C, 75 Chancellor's Circle Winnipeg, MB, Canada, R3T 5V6.
Emails: {akroutm, housseia}@myumanitoba.ca, {faouzi.bellili, amine.mezghani}@umanitoba.ca.

***Abstract*—We introduce the bilinear generalized vector approximate message passing (BiG-VAMP) algorithm which jointly recovers two matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ from their noisy product through a probabilistic observation model. BiG-VAMP provides computationally efficient approximate implementations of both max-sum and sum-product loopy belief propagation (BP). We show how the proposed BiG-VAMP algorithm recovers different types of structured matrices and overcomes the fundamental limitations of other state-of-the-art approaches to the bilinear recovery problem, such as BiG-AMP, BAd-VAMP and LowRAMP. In essence, BiG-VAMP applies to a broader class of practical applications which involve a general form of structured matrices. For the sake of theoretical performance prediction, we also conduct a state evolution (SE) analysis of the proposed algorithm and show its consistency with the asymptotic empirical mean-squared error (MSE). Numerical results on various applications such as matrix factorization, dictionary learning, and matrix completion demonstrate unambiguously the effectiveness of the proposed BiG-VAMP algorithm and its superiority over state-of-the-art algorithms. Using the developed SE framework, we also examine (as one example) the phase transition diagrams of the matrix completion problem, thereby unveiling a low detectability region corresponding to the low signal-to-noise ratio (SNR) regime.**

***Index Terms*—Bayesian inference, approximate message passing, bilinear structured matrix recovery, inference algorithms, matrix factorization, dictionary learning, matrix completion.**

I. Introduction

A. Background and related work

We consider an observation matrix, $\boldsymbol{Y} \in \mathbb{R}^{N\times M}$, obtained from the following generalized bilinear model:

$$p_{\mathbf{Y}|\mathbf{Z}}(\boldsymbol{Y}|\boldsymbol{Z}) \;=\; \prod_{i=1}^{N}\prod_{j=1}^{M} p_{\mathsf{y}_{ij}|\mathsf{z}_{ij}}(y_{ij}|z_{ij}) \quad \text{with} \quad \boldsymbol{Z} = \boldsymbol{U}\boldsymbol{V}^{\mathsf{T}}, \tag{1}$$

where $\boldsymbol{U}$ and $\boldsymbol{V}$ are two unknown matrices in $\mathbb{R}^{N\times r}$ and $\mathbb{R}^{M\times r}$, respectively. The goal is to recover $\boldsymbol{U}$ and $\boldsymbol{V}$ based on the knowledge of $\boldsymbol{Y}$ and the model in (1). The latter applies to a myriad of problems ranging from noisy dictionary learning [1], matrix completion [2], and sparse PCA [3], to matrix factorization [4], low-rank matrix reconstruction [5], and subgraph estimation [6], just to name a few. A special relevant case of the generalized bilinear model in (1) is:

$$\boldsymbol{Y} \;=\; \phi\Big(\boldsymbol{U}\boldsymbol{V}^{\top} + \boldsymbol{W}\Big), \tag{2}$$

in which $\boldsymbol{W} \in \mathbb{R}^{N\times M}$ is an additive white Gaussian noise matrix whose entries are assumed to be mutually independent with mean zero and variance γ_w^{-1}, i.e., $w_{i,j} \;\sim\; \mathcal{N}(w_{ij}; 0, \gamma_w^{-1})$. While convex relaxation of the bilinear recovery problem under the observation model in (2) is possible in some cases using the augmented Lagrange multiplier method (ALMM) [7], different non-convex formulations have been investigated in some special cases over the last decade based on:

1) the alternating direction method of multipliers (ADMM) [8],
2) the variational sparse Bayesian learning (VSBL) method [9],
3) the approximate message passing (AMP) paradigm which is discussed hereafter in some depth in order to put our contribution in a proper perspective.

In fact, AMP-based computational information/data processing have attracted a lot of interest in different fields since the early introduction of the AMP algorithm in [10] within the compressed sensing (CS) framework. More specifically, in a typical CS problem, one is interested in recovering an unknown sparse vector, $\boldsymbol{x}\in\mathbb{R}^N$, from its noisy linear measurements/observations:

$$\boldsymbol{y} \;=\; \boldsymbol{A}\boldsymbol{x} \;+\; \boldsymbol{w}, \tag{3}$$

wherein $\boldsymbol{A} \in \mathbb{R}^{M\times N}$ (with $M \ll N$) is a known sensing matrix. AMP strikes a proper balance between reconstruction performance and computational complexity as compared to traditional convex optimization-based and iterative soft thresholding algorithms [11]. The AMP algorithm was later extended in [12] to generalized linear models of the form:

$$p_{\mathbf{y}|\mathbf{z}}(\boldsymbol{y}|\boldsymbol{z}) = \prod_{m=1}^{M} p_{\mathsf{y}_m|\mathsf{z}_m}(y_m|z_m) \quad \text{with} \quad \boldsymbol{z} = \boldsymbol{A}\boldsymbol{x}. \tag{4}$$

Aside from handling nonlinear transformations, the advantage of the generalized AMP (GAMP) algorithm over its AMP predecessor lies in its ability to accommodate statistical priors on the sparse vector $\boldsymbol{x}$. From another perspective, the performance of both AMP and GAMP can be rigorously tracked by a set of scalar update equations, known as state evolution (SE) [13], [14]

The authors are with the Department of Electrical and Computer Engineering at the University of Manitoba, Winnipeg, MB, Canada. This work was supported by the Discovery Grants Program of the Natural Sciences and Engineering Research Council of Canada (NSERC).

in statistics or the cavity method in statistical physics [15]. One limitation of AMP and GAMP, however, is that they often diverge if the sensing matrix, $\boldsymbol{A}$, is ill-conditioned and/or has a non-zero mean. It is precisely in this context that the vector AMP (VAMP) algorithm has been recently introduced and rigorously analyzed in [16]–[18][1]. Although, there is no theoretical guarantees that VAMP will always converge, there is strong empirical evidence that it is more resilient to mean-perturbed or badly conditioned sensing matrices $\boldsymbol{A}$, provided that the latter is right-orthogonally invariant [18].

To date, extensions of the original AMP algorithm to the bilinear recovery problem were already made in [5] and [20] within the context of low-rank matrix reconstruction. In this context, the so-called LowRAMP algorithm introduced in [20] extends the method in [5] to non-Gaussian likelihoods and establishes the associated SE analysis. It does so by simplifying the BP messages in [5] via second-order Taylor series approximations which become more accurate in the limit of large N. However, the involved approximations hold in the per-measurement low-SNR regime only $\big[$i.e., SNR $= \mathcal{O}(\frac{1}{N})\big]$ and, hence, LowRAMP is not suitable for other types of measurement processes such as quantization and matrix sub-sampling. Moreover, the "low-rank" assumption which is critical in both [5] and [20] precludes a large number of practical applications which involve a *general-rank* decomposition of structured matrices rather than a *low-rank* decomposition of unstructured matrices.

GAMP itself was also extended to the bilinear case in [21] thereby leading to the so-called bilinear generalized AMP[2] (BiG-AMP) algorithm. Although being completely oblivious to the *low-rank* assumption, BiG-AMP inherits all the GAMP-related convergence issues and was developed for separable priors on both $\boldsymbol{U}$ and $\boldsymbol{V}$ matrices. To accommodate a larger class of $\boldsymbol{V}$ matrices in (1), Sarkar *et al.* made an attempt in [23] to generalize the VAMP framework to bilinear recovery problems and the algorithm introduced therein was called *Bilinear Adaptive VAMP* (BAd-VAMP). In essence, BAd-VAMP is a ping-pong-like approach which reconstructs both $\boldsymbol{U}$ and $\boldsymbol{V}$ matrices by alternating between $i)$ the expectation-maximization (EM) algorithm [24] to find the maximimum-likelihood (ML) estimate of $\boldsymbol{U}$ and $ii)$ the VAMP algorithm [18] to find the minimum mean-squared error (MMSE) estimate of $\boldsymbol{V}$. Unfortunately, due to the use of the EM algorithm, BAd-VAMP does not accommodate general priors on the matrix $\boldsymbol{U}$. For instance, trying to enforce a binary or sparsity prior on $\boldsymbol{U}$ renders the E-step of the EM algorithm computationally prohibitive. To overcome all the aforementioned limitations, this paper introduces a new algorithm along with its state evolution analysis to solve a broader class of the bilinear recovery problems as modelled in (1).

B. Contributions

This paper builds upon the prior work in [5] and provides a broader solution to the bilinear recovery problem under different structured matrices beyond the "low-rank" assumption. Our approach for bilinear recovery does not alternate between the EM and VAMP algorithms, but rather relies entirely on message passing and it is dubbed *bilinear generalized VAMP* (BiG-VAMP). The BiG-VAMP approach that we propose in this work enables the use of arbitrary priors on both $\boldsymbol{U}$ and $\boldsymbol{V}$, thereby allowing the exploitation of other matrix structures such as finite-alphabet, binarity, sparsity, constant-modulus, assignment, etc. The proposed BiG-VAMP algorithm is suitable to a broader class of bilinear recovery problems that $i)$ cover more general prior distributions on the unknown matrices (unlike BiG-AMP) and $ii)$ does not rely on the use of the EM algorithm with automated hyperparameter tuning to estimate $\boldsymbol{U}$ (unlike BAd-VAMP). The key differences between BiG-VAMP and LowRAMP are, however, as follows:

1) BiG-VAMP provides a systematic way for handling nonlinear outputs without the "low-SNR" assumption which does not hold, e.g., in the matrix completion problem. Moreover, BiG-VAMP allows maximum *a posteriori* bilinear reconstruction under *non-differentiable* output functions such as quantization, perceptron activation, selection, and phase-retrieval, etc.
2) BiG-VAMP applies to a broader class of practical applications which involve a *general-rank* decomposition of structured matrices in addition to a *low-rank* decomposition of unstructured matrices.

Much like BiG-AMP, BiG-VAMP is also applicable to maximum *a posteriori* (MAP) and MMSE inference problems alike, as will be explained later on. Furthermore, it comes with theoretical performance guarantees, established in Section IV, that validate its superiority against state-of-the-art BiG-AMP, BAd-VAMP, and LowRAMP algorithms.

Notation

We use Sans Serif font (e.g., x) for random variables and Serif font (e.g., x) for its realizations. We use boldface lowercase letters for vectors (e.g., $\boldsymbol{\mathsf{x}}$ and $\boldsymbol{x}$) and boldface uppercase letters for matrices (e.g., $\boldsymbol{\mathsf{X}}$ and $\boldsymbol{X}$). Vectors are in column-wise orientation by default. Given any matrix $\boldsymbol{X}$, we use $\boldsymbol{x}_i$ and x_{ij} to denote its ith column and ijth entry, respectively. We also denote the ith component of a vector $\boldsymbol{x}$ as $[\boldsymbol{x}]_i$ or x_i. The operator diag$(\boldsymbol{X})$ stacks the diagonal elements of $\boldsymbol{X}$ in a vector while $\boldsymbol{I}$ stands for the identity matrix. The operator tr$(\boldsymbol{X})$ returns the sum of the diagonal elements of $\boldsymbol{X}$. We also use $p_{\mathsf{x}}(x;\boldsymbol{\theta})$, $p_{\boldsymbol{\mathsf{x}}}(\boldsymbol{x};\boldsymbol{\theta})$, and $p_{\boldsymbol{\mathsf{X}}}(\boldsymbol{X};\boldsymbol{\theta})$ to denote the pdf of random variables/vectors/matrices; as being parameterized by a set of parameters $\boldsymbol{\theta}$. Moreover, $\mathcal{N}(\boldsymbol{x};\widehat{\boldsymbol{x}},\boldsymbol{R})$ stands for the multivariate Gaussian pdf of any random vector $\boldsymbol{\mathsf{x}}$ with mean $\widehat{\boldsymbol{x}}$ and covariance matrix $\boldsymbol{R}$. We use $\sim$ and $\propto$ as short-hand motations for "distributed according to" and "proportional to", respectively. We also use $\mathbb{E}[\boldsymbol{\mathsf{x}}|d(\boldsymbol{x})]$ to denote the expectation of $\boldsymbol{\mathsf{x}} \sim d(\boldsymbol{x})$ and $\delta(\boldsymbol{x})$ refers to the Dirac delta distribution. Moreover, $\langle\boldsymbol{x}\rangle$ and $\langle\boldsymbol{X}\rangle$ return the (empirical) average values of vectors and matrices, i.e., $\langle\boldsymbol{x}\rangle \triangleq \frac{1}{N}\sum_{i=1}^{N} x_i$ for $\boldsymbol{x} \in \mathbb{R}^N$ and $\langle\boldsymbol{X}\rangle \triangleq \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} x_{ij}$ for $\boldsymbol{X} \in \mathbb{R}^{N\times M}$. Finally, the symbol $\odot$ denotes the Hadamard (i.e., elementwise) product between any two matrices.

[1] It is worth mentioning here that VAMP was independently derived by two research groups in [16] and [19] but under two different names, i.e., VAMP and orthogonal AMP (OAMP), respectively.

[2] The reader is also referred to the parametric version of BiG-AMP in [22].

II. Background on the low-rank matrix reconstruction

In this section, we briefly review the main results of the prior work on *low-rank* matrix reconstruction in [5] at the detail needed for a comprehensive exposition of BiG-VAMP. We emphasize, however, the fact that borrowing such results does not restrict the proposed BiG-VAMP algorithm to the bilinear *low-rank* matrix recovery as is the case in [5].

Consider the bilinear recovery of two random independent matrices $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_N]^\top \in \mathbb{R}^{N\times r}$ and $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_M]^\top \in \mathbb{R}^{M\times r}$ from a linear noisy observation $\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{V}^\top + \boldsymbol{W} \in \mathbb{R}^{N\times M}$. Given some common priors, $p_{\mathbf{u}}(\boldsymbol{u}_i)$ and $p_{\mathbf{v}}(\boldsymbol{v}_i)$, on the vectors $\boldsymbol{u}_i$ and $\boldsymbol{v}_j$, respectively, the goal of BP is to approximate their joint posterior distribution:

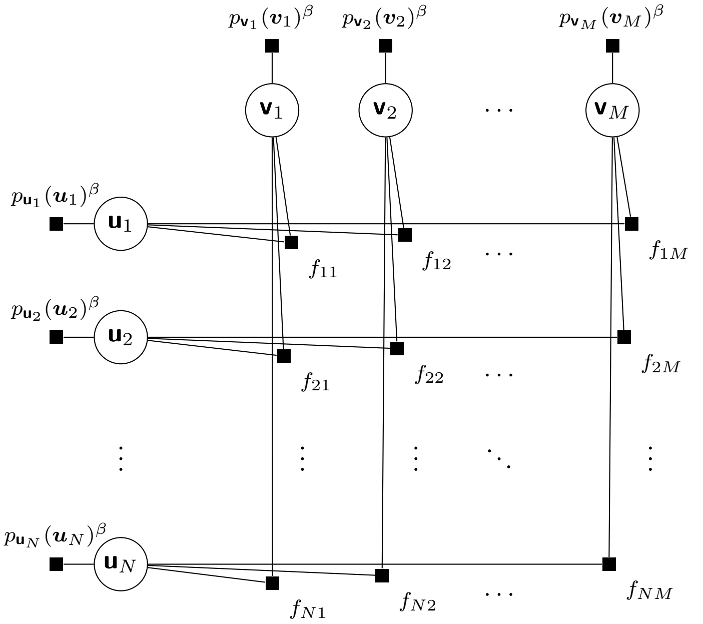
$$\begin{aligned} & p_{\mathbf{u},\mathbf{v}|\mathbf{Y}}(\boldsymbol{u}_i, \boldsymbol{v}_j|\boldsymbol{Y}; \gamma_w^{-1}, \beta) \\ & \qquad \propto \prod_{i=1}^{N}\prod_{j=1}^{M} p_{\mathsf{y}_{ij}|\mathbf{u},\mathbf{v}}\left(y_{ij}|\boldsymbol{u}_i, \boldsymbol{v}_j; \gamma_w^{-1}\right)^\beta\, p_{\mathbf{u}}(\boldsymbol{u}_i)^\beta\, p_{\mathbf{v}}(\boldsymbol{v}_j)^\beta, \end{aligned} \tag{5}$$

wherein β is a parameter introduced here to treat the MMSE ($\beta = 1$) and MAP ($\beta = +\infty$) inference problems in a unified framework. We also assume the priors, $p_{\mathbf{U}}(\boldsymbol{U})$ and $p_{\mathbf{V}}(\boldsymbol{V})$, on the unknown matrices of $\mathbf{U}$ and $\mathbf{V}$ to be row-wise separable, i.e., $p_{\mathbf{U}}(\boldsymbol{U}) = \prod_{i=1}^{N} p_{\mathbf{u}}(\boldsymbol{u}_i)$ and $p_{\mathbf{V}}(\boldsymbol{V}) = \prod_{j=1}^{M} p_{\mathbf{v}}(\boldsymbol{v}_j)$. Fig. 1 depicts the factor graph associated to (5) with variable nodes, $\mathbf{u}_i$ and $\mathbf{v}_j$, their prior factor nodes, $p_{\mathbf{u}}(\boldsymbol{u}_i)^\beta$ and $p_{\mathbf{v}}(\boldsymbol{v}_j)^\beta$, and the labels, f_{ij}, which we use as a shorthand notations for the factor nodes:

$$\begin{aligned} f(\boldsymbol{u}_i, \boldsymbol{v}_j) &\triangleq p_{\mathsf{y}_{ij}|\mathbf{u},\mathbf{v}}\left(y_{ij}|\boldsymbol{u}_i, \boldsymbol{v}_j, \gamma_w^{-1}\right)^\beta, & (6)\\ &= \mathcal{N}\left(y_{ij}; \boldsymbol{u}_i^\top \boldsymbol{v}_j, \beta^{-1}\gamma_w^{-1}\right). & (7) \end{aligned}$$

By taking $\beta = 1$ (i.e., minimum mean square error estimation), $p_{\mathbf{u},\mathbf{v}|\mathsf{y}_{ij}}(\boldsymbol{u}_i, \boldsymbol{v}_j|y_{ij}; \beta)$ reduces to the true joint posterior $p_{\mathbf{u},\mathbf{v}|\mathsf{y}_{ij}}(\boldsymbol{u}_i, \boldsymbol{v}_j|y_{ij})$. In the limit $\beta \to \infty$ (i.e., maximum a posteriori estimation), however, it concentrates on the maxima of $p_{\mathbf{u},\mathbf{v}|\mathsf{y}_{ij}}(\boldsymbol{u}_i, \boldsymbol{v}_j|y_{ij})$. Using the message derivation rules of loopy BP, the four messages defined in Fig. 1, , are expressed as follows (where t stands for the iteration index):

$$\begin{aligned} \mu_{(i,j)\to i,t}(\boldsymbol{u}_i) &\propto \int f(\boldsymbol{u}_i, \boldsymbol{v}_j)\, \nu_{j\to(i,j),t-1}(\boldsymbol{v}_j)\, \mathrm{d}\boldsymbol{v}_j, & (8)\\ \mu_{i\to(i,j),t+1}(\boldsymbol{u}_i) &\propto p_{\mathbf{u}}(\boldsymbol{u}_i)^\beta \prod_{l\neq j} \mu_{(i,l)\to i,t}(\boldsymbol{u}_i), & (9)\\ \nu_{(i,j)\to j,t}(\boldsymbol{v}_j) &\propto \int f(\boldsymbol{u}_i, \boldsymbol{v}_j)\, \mu_{i\to(i,j),t-1}(\boldsymbol{u}_i)\, \mathrm{d}\boldsymbol{u}_i, & (10)\\ \nu_{j\to(i,j),t+1}(\boldsymbol{v}_j) &\propto p_{\mathbf{v}}(\boldsymbol{v}_j)^\beta \prod_{k\neq i} \nu_{(k,j)\to j,t}(\boldsymbol{v}_j). & (11) \end{aligned}$$


$p_{\mathbf{v}_1}(\boldsymbol{v}_1)^\beta$
$p_{\mathbf{v}_2}(\boldsymbol{v}_2)^\beta$
$p_{\mathbf{v}_M}(\boldsymbol{v}_M)^\beta$
$\mathbf{v}_1$
$\mathbf{v}_2$
$\cdots$
$\mathbf{v}_M$
$p_{\mathbf{u}_1}(\boldsymbol{u}_1)^\beta$
$\mathbf{u}_1$
f_{11}
f_{12}
f_{1M}
$p_{\mathbf{u}_2}(\boldsymbol{u}_2)^\beta$
$\mathbf{u}_2$
f_{21}
f_{22}
f_{2M}
$p_{\mathbf{u}_N}(\boldsymbol{u}_N)^\beta$
$\mathbf{u}_N$
f_{N1}
f_{N2}
f_{NM}

$\mu_{(i,j)\to i,t}$	message from node f_{ij} to node $\mathbf{u}_i$
$\nu_{(i,j)\to j,t}$	message from node f_{ij} to node $\mathbf{v}_j$
$\mu_{i\to(i,j),t}$	message from node $\mathbf{u}_i$ to node f_{ij}
$\nu_{j\to(i,j),t}$	message from node $\mathbf{v}_j$ to node f_{ij}

Fig. 1: Factor graph associated to (5). The circles represent variable nodes and the squares represent factor nodes.

Assuming $\mathbf{v}_l \sim \nu_{l\to(i,l),t}(\boldsymbol{v}_l)$ with mean $\boldsymbol{v}_{l\to(i,l),t}$ and covariance matrix $\beta^{-1}\boldsymbol{R}_{\mathbf{v},l\to(i,l),t}$, it was shown in [5] by virtue of the central limit theorem (CLT) that the product of incoming messages, $\prod_{l\neq j}\mu_{(i,l)\to i,t}(\boldsymbol{u}_i)$, to any variable node $\mathbf{u}_i$ from all factor nodes $\{f_{il}\}_{l\neq j}$ can be approximated by a Gaussian random variable with mean $\boldsymbol{b}_{\mathbf{u},i\to(i,j),t}$ and precision $\beta\boldsymbol{\Lambda}_{\mathbf{u},i\to(i,j),t}$:

$$\boldsymbol{b}_{\mathbf{u},i\to(i,j),t} = \gamma_w \sum_{l\neq j} y_{i,l}\, \widehat{\boldsymbol{v}}_{l\to(i,l),t}, \tag{12a}$$

$$\begin{aligned} \boldsymbol{\Lambda}_{\mathbf{u},i\to(i,j),t} = \gamma_w \sum_{l\neq j} \Big(& \widehat{\boldsymbol{v}}_{l\to(i,l),t}\, \widehat{\boldsymbol{v}}_{l\to(i,l),t}^\top + \beta^{-1}\boldsymbol{R}_{\mathbf{v},l\to(i,l),t} \\ & - \gamma_w\, y_{i,l}^2\, \boldsymbol{R}_{\mathbf{v},l\to(i,l),t}\Big). \end{aligned} \tag{12b}$$

In other words, by dropping the normalization factor that does not depend on $\boldsymbol{u}_i$, we have:

$$\begin{aligned} & \prod_{l\neq j} \mu_{(i,l)\to i,t}(\boldsymbol{u}_i) \\ & \qquad \propto \exp\left(-\frac{\beta}{2}\boldsymbol{u}_i^\top \boldsymbol{\Lambda}_{\mathbf{u},i\to(i,j),t}\boldsymbol{u}_i + \beta\, \boldsymbol{u}_i^\top \boldsymbol{b}_{\mathbf{u},i\to(i,j),t}\right). \end{aligned} \tag{13}$$

The inherent symmetry among the variable nodes $\mathbf{u}_i$ and $\mathbf{v}_j$ yields an equivalent Gaussian approximation for $\prod_{k\neq i}\nu_{(k,j)\to j,t}(\boldsymbol{v}_j)$ under the density of $\mathbf{u}_k \sim \mu_{k\to(k,j),t}(\boldsymbol{u}_k)$ with mean $\boldsymbol{u}_{k\to(k,j),t}$ and covariance matrix $\beta^{-1}\boldsymbol{R}_{\mathbf{u},k\to(k,j),t}$. That is to say:

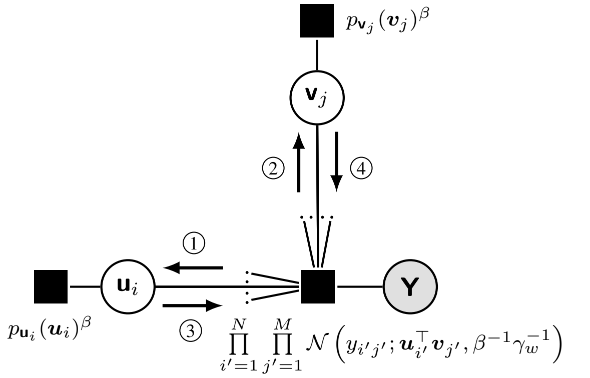
$$\begin{aligned} & \prod_{k\neq i} \nu_{(k,j)\to j,t}(\boldsymbol{v}_j) \\ & \qquad \propto \exp\left(-\frac{\beta}{2}\boldsymbol{v}_j^\top \boldsymbol{\Lambda}_{\mathbf{v},j\to(i,j),t}\boldsymbol{v}_j + \beta\, \boldsymbol{v}_j^\top \boldsymbol{b}_{\mathbf{v},j\to(i,j),t}\right), \end{aligned} \tag{14}$$

with

$$\boldsymbol{b}_{\mathbf{v},j\to(i,j),t} \;=\; \gamma_w \sum_{k\neq i} y_{k,j}\,\widehat{\boldsymbol{u}}_{k\to(k,j),t}, \tag{15a}$$

$$\boldsymbol{\Lambda}_{\mathbf{v},j\to(i,j),t} \;=\; \gamma_w \sum_{k\neq i}\Big(\widehat{\boldsymbol{u}}_{k\to(k,j),t}\,\widehat{\boldsymbol{u}}_{k\to(k,j),t}^{\top} + \beta^{-1}\boldsymbol{R}_{\mathbf{u},k\to(k,j),t} - \gamma_w\, y_{k,j}^2\,\boldsymbol{R}_{\mathbf{u},k\to(k,j),t}\Big), \tag{15b}$$

Pictorially, the messages given in (13) and (14) are shown in Fig. 2 as messages ① and ② sent by all factor nodes $\{f_{ij'}\}_{j'=1,j'\neq j}^{M}$ and $\{f_{i'j}\}_{i'=1,i'\neq i}^{N}$ to $\boldsymbol{u}_i$ and $\boldsymbol{v}_j$, respectively. Moreover, the mean values, $\widehat{\boldsymbol{u}}_{i\to(i,j),t+1}$ and $\widehat{\boldsymbol{v}}_{j\to(i,j),t+1}$, as well as the covariance matrices, $\beta^{-1}\boldsymbol{R}_{\mathbf{u},i\to(i,j),t+1}$ and $\beta^{-1}\boldsymbol{R}_{\mathbf{v},j\to(i,j),t+1}$, of messages ③ and ④, respectively, are given by:


$p_{\mathbf{v}_j}(\boldsymbol{v}_j)^\beta$
$\mathbf{v}_j$
②
④
①
$\mathbf{u}_i$
$\mathbf{Y}$
$p_{\mathbf{u}_i}(\boldsymbol{u}_i)^\beta$
③
$\prod_{i'=1}^{N}\prod_{j'=1}^{M}\mathcal{N}\left(y_{i'j'};\boldsymbol{u}_{i'}^\top\boldsymbol{v}_{j'},\beta^{-1}\gamma_w^{-1}\right)$

①	$\mathcal{N}(\boldsymbol{u}_i;\boldsymbol{\Lambda}_{\mathbf{u},i\to(i,j),t}^{-1}\boldsymbol{b}_{\mathbf{u},i\to(i,j),t},\beta^{-1}\boldsymbol{\Lambda}_{\mathbf{u},i\to(i,j),t}^{-1})$
②	$\mathcal{N}(\boldsymbol{v}_j;\boldsymbol{\Lambda}_{\mathbf{v},j\to(i,j),t}^{-1}\boldsymbol{b}_{\mathbf{v},j\to(i,j),t},\beta^{-1}\boldsymbol{\Lambda}_{\mathbf{v},j\to(i,j),t}^{-1})$

Fig. 2: Explicit messages resulting from the CLT approximations shown here for a single cell of the entire factor graph depicted in Fig. 1.

$$\widehat{\boldsymbol{u}}_{i\to(i,j),t+1} \;=\; \mathsf{f}_{\mathsf{u}}(\boldsymbol{b}_{\mathbf{u},i\to(i,j),t},\boldsymbol{\Lambda}_{\mathbf{u},i\to(i,j),t}^{-1}), \tag{16}$$

$$\widehat{\boldsymbol{v}}_{j\to(i,j),t+1} \;=\; \mathsf{f}_{\mathsf{v}}(\boldsymbol{b}_{\mathbf{v},j\to(i,j),t},\boldsymbol{\Lambda}_{\mathbf{v},j\to(i,j),t}^{-1}), \tag{17}$$

$$\boldsymbol{R}_{\mathbf{u},i\to(i,j),t+1} \;=\; \nabla_{\boldsymbol{b}_{\mathbf{u},i\to(i,j),t}}\mathsf{f}_{\mathsf{u}}(\boldsymbol{b}_{\mathbf{u},i\to(i,j),t},\boldsymbol{\Lambda}_{\mathbf{u},i\to(i,j),t}^{-1})^{\mathsf{T}}, \tag{18}$$

$$\boldsymbol{R}_{\mathbf{v},j\to(i,j),t+1} \;=\; \nabla_{\boldsymbol{b}_{\mathbf{v},j\to(i,j),t}}\mathsf{f}_{\mathsf{v}}(\boldsymbol{b}_{\mathbf{v},j\to(i,j),t},\boldsymbol{\Lambda}_{\mathbf{v},j\to(i,j),t}^{-1})^{\mathsf{T}}, \tag{19}$$

where

$$\mathsf{f}_{\mathsf{u}}(\boldsymbol{b},\boldsymbol{\Lambda}^{-1}) \;=\; \frac{\displaystyle\int \boldsymbol{u}\,p_{\mathbf{u}}(\boldsymbol{u})^{\beta}\,\mathcal{N}\big(\boldsymbol{u};\boldsymbol{\Lambda}^{-1}\boldsymbol{b},\beta^{-1}\boldsymbol{\Lambda}^{-1}\big)d\boldsymbol{u}}{\displaystyle\int p_{\mathbf{u}}(\boldsymbol{u})^{\beta}\,\mathcal{N}\big(\boldsymbol{u};\boldsymbol{\Lambda}^{-1}\boldsymbol{b},\beta^{-1}\boldsymbol{\Lambda}^{-1}\big)d\boldsymbol{u}}, \tag{20}$$

$$\mathsf{f}_{\mathsf{v}}(\boldsymbol{b},\boldsymbol{\Lambda}^{-1}) \;=\; \frac{\displaystyle\int \boldsymbol{v}\,p_{\mathbf{v}}(\boldsymbol{v})^{\beta}\,\mathcal{N}\big(\boldsymbol{v};\boldsymbol{\Lambda}^{-1}\boldsymbol{b},\beta^{-1}\boldsymbol{\Lambda}^{-1}\big)d\boldsymbol{v}}{\displaystyle\int p_{\mathbf{v}}(\boldsymbol{v})^{\beta}\,\mathcal{N}\big(\boldsymbol{v};\boldsymbol{\Lambda}^{-1}\boldsymbol{b},\beta^{-1}\boldsymbol{\Lambda}^{-1}\big)d\boldsymbol{v}}, \tag{21}$$

and the nabla operator, $\nabla_{\boldsymbol{x}}$, with respect to any $n-$dimensional vector, $\boldsymbol{x} = [x_1, x_2, \cdots, x_n]^{\mathsf{T}}$, is given by:

$$\nabla_{\boldsymbol{x}} \;=\; \left[\frac{\partial}{\partial x_1},\frac{\partial}{\partial x_2},\cdots,\frac{\partial}{\partial x_n}\right]^{\mathsf{T}}. \tag{22}$$

To reduce the complexity in the number of computed messages, the $(i,j)-$dependent quantities in (12) and (15), are replaced by the following $(i,j)-$independent (i.e., broadcast) ones:

$$\boldsymbol{b}_{\mathbf{u},i,t} \;=\; \gamma_w \sum_{l} y_{i,l}\,\widehat{\boldsymbol{v}}_{l\to(i,l),t}, \tag{23a}$$

$$\boldsymbol{\Lambda}_{\mathbf{u},i,t} \;=\; \gamma_w \sum_{l}\Big(\widehat{\boldsymbol{v}}_{l\to(i,l),t}\,\widehat{\boldsymbol{v}}_{l\to(i,l),t}^{\top} \;+\; \big(\beta^{-1}-\gamma_w\; y_{i,l}^2\big)\,\boldsymbol{R}_{\mathbf{v}_l,t}\Big). \tag{23b}$$

$$\boldsymbol{b}_{\mathbf{v},j,t} \;=\; \gamma_w \sum_{k} y_{j,k}\,\widehat{\boldsymbol{u}}_{k\to(k,j),t}, \tag{24a}$$

$$\boldsymbol{\Lambda}_{\mathbf{v},j,t} \;=\; \gamma_w \sum_{k}\Big(\widehat{\boldsymbol{u}}_{k\to(k,j),t}\,\widehat{\boldsymbol{u}}_{k\to(k,j),t}^{\top} + \big(\beta^{-1}-\gamma_w y_{k,j}^2\big)\,\boldsymbol{R}_{\mathbf{u}_k,t}\Big). \tag{24b}$$

after approximating the covariance matrices, $\beta^{-1}\boldsymbol{R}_{\mathbf{u},i\to(i,j),t}$ and $\beta^{-1}\boldsymbol{R}_{\mathbf{v},j\to(i,j),t}$, involved in (12) and (15) by broadcast covariances, $\beta^{-1}\boldsymbol{R}_{\mathbf{u}_i,t}$ and $\beta^{-1}\boldsymbol{R}_{\mathbf{v}_j,t}$, respectively, with a vanishing error order $O(M^{-1})$ [5]. By recalling (16)-(19), the underlying broadcast means and the associated brodcacst covariances are thus givens by:

$$\widehat{\boldsymbol{u}}_{i,t+1} \;=\; \mathsf{f}_{\mathsf{u}}(\boldsymbol{b}_{\mathbf{u},i,t},\boldsymbol{\Lambda}_{\mathbf{u},i,t}^{-1}) \tag{25}$$

$$\boldsymbol{R}_{\mathbf{u}_i,t+1} \;=\; \nabla_{\boldsymbol{b}_{\mathbf{u},i,t}}\mathsf{f}_{\mathsf{u}}(\boldsymbol{b}_{\mathbf{u},i,t},\boldsymbol{\Lambda}_{\mathbf{u},i,t}^{-1})^{\mathsf{T}}, \tag{26}$$

$$\widehat{\boldsymbol{v}}_{j,t+1} \;=\; \mathsf{f}_{\mathsf{v}}(\boldsymbol{b}_{\mathbf{v},j,t},\boldsymbol{\Lambda}_{\mathbf{v},j,t}^{-1}) \tag{27}$$

$$\boldsymbol{R}_{\mathbf{v}_j,t+1} \;=\; \nabla_{\boldsymbol{b}_{\mathbf{v},j,t}}\mathsf{f}_{\mathsf{v}}(\boldsymbol{b}_{\mathbf{v},j,t},\boldsymbol{\Lambda}_{\mathbf{v},j,t}^{-1})^{\mathsf{T}}. \tag{28}$$

Moreover, the (i,j) posterior means, $\widehat{\boldsymbol{u}}_{i\to(i,j),t}$ and $\widehat{\boldsymbol{v}}_{j\to(i,j),t}$, are related to the their broadcast versions, $\widehat{\boldsymbol{u}}_{i,t}$ and $\widehat{\boldsymbol{v}}_{j,t}$, through small Osanger correction terms of order $O(M^{-1/2})$ as follows:

$$\widehat{\boldsymbol{u}}_{i\to(i,j),t} \;\approx\; \widehat{\boldsymbol{u}}_{i,t} \;-\; \underbrace{\gamma_w\,y_{ij}\,\boldsymbol{R}_{\boldsymbol{u}_i,t}\,\widehat{\boldsymbol{v}}_{j,t-1}}_{\text{Osanger correction term on }\boldsymbol{u}_i}\,, \tag{29a}$$

$$\widehat{\boldsymbol{v}}_{j\to(i,j),t} \;\approx\; \widehat{\boldsymbol{v}}_{j,t} \;-\; \underbrace{\gamma_w\,y_{ij}\,\boldsymbol{R}_{\boldsymbol{v}_j,t}\,\widehat{\boldsymbol{u}}_{i,t-1}}_{\text{Osanger correction term on }\boldsymbol{v}_j}\,. \tag{29b}$$

Yet, the correction terms in (29) are taken into account during the calculation of $\boldsymbol{b}_{\mathbf{u},i,t}$ and $\boldsymbol{b}_{\mathbf{v},j,t}$ only. For the computation of $\boldsymbol{\Lambda}_{\mathbf{u},i,t}$ and $\boldsymbol{\Lambda}_{\mathbf{v},j,t}$, however, one replaces $\widehat{\boldsymbol{u}}_{i\to(i,j),t}$ by $\widehat{\boldsymbol{u}}_{i,t}$ and $\widehat{\boldsymbol{v}}_{j\to(i,j),t}$ by $\widehat{\boldsymbol{v}}_{j,t}$ after ignoring terms of vanishing order as M and N grow large. The algorithm in [5] also relies on the following low-SNR approximation

$$y_{k,j}^2 \;\approx\; \mathbb{E}[y_{k,j}^2] \;\approx \gamma_w^{-1}. \tag{30}$$

which is used in both (23b) and (24b). We emphasize, however, the fact that the low-SNR regime is mainly conceivable in presence of very-low-rank structures with a fully observed matrix $\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{V}^{\top} + \boldsymbol{W}$.

In conclusion, the algorithmic steps of the bilinear recovery technique introduced in [5] are summarized in Algorithm 1. We refer the reader to the supplementary materials of [5] for further details. The limitation of this method, however, lies in the intractable multi-dimensional integrals in (20) and (21) which can be evaluated only in some specific priors, $p_{\mathbf{u}}(\cdot)$ and $p_{\mathbf{v}}(\cdot)$,

Algorithm 1 AMP-based structured matrix reconstruction [5]

```
Require : Matrix $\boldsymbol{Y} \in \mathbb{R}^{N\times M}$; noise precision $\gamma_w$; temperature parameter $\beta$; two denoising functions, $\mathbf{f}_{\mathsf{u}}(.)$ and $\mathbf{f}_{\mathsf{v}}(.)$, given in (20) and (21); number of iterations $T_{\text{max}}$.
Initialize
▷ posterior means and covariances
$\widehat{\boldsymbol{U}}_0, \widehat{\boldsymbol{V}}_0, \widehat{\boldsymbol{U}}_1, \widehat{\boldsymbol{V}}_1, \{\boldsymbol{R}_{\boldsymbol{u}_i,1}\}_{i=1}^N, \{\boldsymbol{R}_{\boldsymbol{v}_j,1}\}_{j=1}^M$
for $t = 1, \dots, T_{\text{max}}$ do
        ▷ Compute means and precisions related to the messages in eqs. (13) and (14)
    $\boldsymbol{B}_{\boldsymbol{U},t} = \gamma_w \left(\boldsymbol{Y}\widehat{\boldsymbol{V}}_t - \gamma_w \widehat{\boldsymbol{U}}_{t-1} \sum_{j=1}^M \boldsymbol{R}_{\boldsymbol{v}_j,t}\right)$
    $\boldsymbol{\Lambda}_{\boldsymbol{U},t} = \gamma_w \left(\widehat{\boldsymbol{V}}_t^\top \widehat{\boldsymbol{V}}_t + (\frac{1}{\beta} - 1) \sum_{j=1}^M \boldsymbol{R}_{\boldsymbol{v}_j,t}\right)$
    $\boldsymbol{B}_{\boldsymbol{V},t} = \gamma_w \left(\boldsymbol{Y}\widehat{\boldsymbol{U}}_t - \gamma_w \widehat{\boldsymbol{V}}_{t-1} \sum_{i=1}^N \boldsymbol{R}_{\boldsymbol{u}_i,t}\right)$
    $\boldsymbol{\Lambda}_{\boldsymbol{V},t} = \gamma_w \left(\widehat{\boldsymbol{U}}_t^\top \widehat{\boldsymbol{U}}_t + (\frac{1}{\beta} - 1) \sum_{i=1}^N \boldsymbol{R}_{\boldsymbol{u}_i,t}\right)$
        ▷ Update the posterior means, $\widehat{\boldsymbol{U}} = [\widehat{\boldsymbol{u}}_1, \dots, \widehat{\boldsymbol{u}}_N]^\mathsf{T}$ and $\widehat{\boldsymbol{V}} = [\widehat{\boldsymbol{v}}_1, \dots, \widehat{\boldsymbol{v}}_M]^\mathsf{T}$, of the matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ and the corresponding row-wise scaled covariance matrices
    $\forall i: \ \widehat{\boldsymbol{u}}_{i,t+1} = \mathbf{f}_{\mathsf{u}}(\boldsymbol{b}_{\mathbf{u},i,t}, \boldsymbol{\Lambda}_{\boldsymbol{U},t}^{-1})$   with   $\boldsymbol{B}_{\boldsymbol{U},t} = [\boldsymbol{b}_{\mathbf{u},1,t}, \dots, \boldsymbol{b}_{\mathbf{u},N,t}]^\mathsf{T}$
    $\forall i: \ \boldsymbol{R}_{\boldsymbol{u}_i,t+1} = \nabla_{\boldsymbol{b}_{\mathbf{u}}} \mathbf{f}_{\mathsf{u}}(\boldsymbol{b}_{\mathbf{u},i,t}, \boldsymbol{\Lambda}_{\boldsymbol{U},t}^{-1})^\mathsf{T}$
    $\forall j: \ \widehat{\boldsymbol{v}}_{j,t+1} = \mathbf{f}_{\mathsf{v}}(\boldsymbol{b}_{\mathbf{v},j,t}, \boldsymbol{\Lambda}_{\boldsymbol{V},t}^{-1})$   with   $\boldsymbol{B}_{\boldsymbol{V},t} = [\boldsymbol{b}_{\mathbf{v},1,t}, \dots, \boldsymbol{b}_{\mathbf{v},M,t}]^\mathsf{T}$
    $\forall j: \ \boldsymbol{R}_{\boldsymbol{v}_j,t+1} = \nabla_{\boldsymbol{b}_{\mathbf{v}}} \mathbf{f}_{\mathsf{v}}(\boldsymbol{b}_{\mathbf{v},j,t}, \boldsymbol{\Lambda}_{\boldsymbol{V},t}^{-1})^\mathsf{T}$
end for
Return $\widehat{\boldsymbol{U}}_{T_{\text{max}}+1}, \widehat{\boldsymbol{V}}_{T_{\text{max}}+1}$
```

such as Gaussian and community[3] priors. It is impossible, for instance, to consider a binary prior on $\mathbf{u}_i$ and/or $\mathbf{v}_j$ since the underlying integrals become combinatorial sums over 2^r terms. In this context, the fundamental novelties brought by the proposed BiG-VAMP algorithm consist in its combined abilities to handle:

- A broader class of practical applications which involve a *high-rank* decomposition of structured matrices in addition to a *low-rank* decomposition of (possibly) unstructured matrices,
- General priors on both $\boldsymbol{U}$ and $\boldsymbol{V}$ matrices owing to appropriate Gaussian approximation of the extrinsic information exchanged between its constituent blocks.
- General separable output distributions, $p_{\mathbf{Y}|\mathbf{Z}}(\boldsymbol{Y}|\boldsymbol{Z})$, in (1).

III. The BiG-VAMP algorithm

Before delving into the derivation details, we first introduce BiG-VAMP which runs iteratively according to the algorithmic steps of Algorithm 2. There, t stands for the iteration index and subscripts p and e are used to distinguish "posterior" and "extrinsic" variables, respectively. As a visual reminder, we also use the hat symbol "$\widehat{\ }$" to refer to mean values. Moreover, Algorithm 2 updates the means and precisions of all messages simultaneously (i.e., for all i and j at the same time). For instance, at each iteration t, the entire matrix $\boldsymbol{B}_{\boldsymbol{U},t} \triangleq [\boldsymbol{b}_{\mathbf{u},1,t}, \boldsymbol{b}_{\mathbf{u},2,t}, \dots, \boldsymbol{b}_{\mathbf{u},N,t}]^\mathsf{T}$ is updated where $\{\boldsymbol{b}_{\mathbf{u},i,t}\}_{i=1}^N$ is the t^{th} update of the message pertaining to the $\{i^{th}\}_{i=1}^N$ variable node $\{\mathbf{u}_i\}_{i=1}^N$. For better illustration, the block diagram of Algorithm 2 is depicted in Fig. 4 whereby we show its different constituent blocks, namely the different denoisers as they interact with the so-called bi-LMMSE module through the extrinsic information (cf. Section III-A for more details). In the sequel, we first briefly discuss the "low rank" assumption used in [5] and [20] which is no longer needed for the derivation of all BiG-VAMP messages. Then, we describe the Gaussian approximation of the extrinsic information as a key means to handle general priors on both $\boldsymbol{U}$ and $\boldsymbol{V}$ matrices in bilinear models. Finally, we extend the results to the generalized bilinear models.

A. Bilinear Vector Approximate Message Passing (Bi-VAMP) with general rank

In this case, the data matrix, $\boldsymbol{Y}$, is obtained from the bilinear observation model in (2) with $\phi(.)$ being the identity, i.e., $\phi(x) = x, \ \forall x \in \mathbb{R}$. That is to say:

$$\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{V}^\mathsf{T} + \boldsymbol{W}, \tag{31}$$

where the noise components, w_{ij}, are mutually independent and Gaussian distributed with zero mean and variance γ_w^{-1}. All the algorithmic steps of Bi-VAMP which will be explained in this section are summarized in Algorithm 2 after excluding the update equations pertaining to the "generalized output step" (i.e., lines 26-32), while replacing $\gamma_{\boldsymbol{Z}_\mathsf{e}^+}$ by γ_w and $\widehat{\boldsymbol{Z}}_\mathsf{e}^+$ by $\boldsymbol{Y}$.

To sidestep the problem of computing the intractable integrals in (20) and (21) during the evaluation of the mean and variance of messages ③ and ④ in Fig. 2, we proceed as follows. We rewrite the original posterior factorization in (5) by splitting $\mathbf{u}_i$ (resp. $\mathbf{v}_j$) into two identical variables with equality constraints in between, i.e., $\mathbf{u}_i^+ = \mathbf{u}_i^-$ (resp. $\mathbf{v}_j^+ = \mathbf{v}_j^-$), thereby yielding the following equivalent factorization:

$$\begin{aligned} & p_{\mathbf{u}_i^+,\mathbf{u}_i^-,\mathbf{v}_j^+,\mathbf{v}_j^-|\mathbf{Y}}\left(\boldsymbol{u}_i^+, \boldsymbol{u}_i^-, \boldsymbol{v}_j^+, \boldsymbol{v}_j^- | \boldsymbol{Y}; \gamma_w^{-1}, \beta\right) \\ & \qquad \propto \prod_{i=1}^N \prod_{j=1}^M p_{\mathsf{y}_{ij}|\mathbf{u}_i^-,\mathbf{v}_j^-}\left(y_{ij}|\boldsymbol{u}_i^-, \boldsymbol{v}_j^-; \gamma_w^{-1}\right)^\beta \\ & \qquad\qquad \times \delta(\boldsymbol{u}_i^- - \boldsymbol{u}_i^+) p_{\mathbf{u}}(\boldsymbol{u}_i^+)^\beta \\ & \qquad\qquad\qquad \times \delta(\boldsymbol{v}_j^- - \boldsymbol{v}_j^+)\, p_{\mathbf{v}}(\boldsymbol{v}_j^+)^\beta. \end{aligned} \tag{32}$$

The new factorization in (32) transforms the original factor graph in Fig. 2 into the new one depicted in Fig. 3. To handle the newly introduced equality constraints, the new variables $\mathbf{u}_i^+$, $\mathbf{u}_i^-$, $\mathbf{v}_j^+$, and $\mathbf{v}_j^-$ are rather regarded as processing nodes which exchange scalar messages/beliefs in the form of component-wise (i.e., decoupled) Gaussian densities. The beliefs provided by $\mathbf{u}_i^-$ and $\mathbf{v}_i^-$ on $\mathbf{u}_i^+$ and $\mathbf{v}_i^+$ (and vice versa) are known as the *extrinsic information* and are modelled by the Gaussian messages ①', ②', ③' and ④' in Fig. 3 whose parameters will be calculated later in this section.

The decoupling in the messages on each side of the equality nodes results in two simple types of denoising functions, namely $\mathbf{f}_{\mathsf{u}}(\cdot,\cdot)$ $\big[$resp. $\mathbf{f}_{\mathsf{v}}(\cdot,\cdot)\big]$ and $\mathbf{g}_{\mathsf{u}}(\cdot,\cdot)$ $\big[$resp. $\mathbf{g}_{\mathsf{v}}(\cdot,\cdot)\big]$ to recover $\boldsymbol{u}_i^-$ (resp. $\boldsymbol{v}_j^-$) and $\boldsymbol{u}_i^+$ (resp. $\boldsymbol{v}_j^+$). Such decoupled message passing is made possible by ignoring the off-diagonal elements of the error covariance matrices calculated on the side of $\mathbf{u}_i^-$ and $\mathbf{v}_j^-$ nodes. Although the integrals of the denoising functions, $\mathbf{f}_{\mathsf{u}}(\cdot,\cdot)$ and $\mathbf{f}_{\mathsf{v}}(\cdot,\cdot)$, in (20) and (21) are still required for Bi-VAMP, they are now evaluated in closed form after replacing the actual priors, $p_{\mathbf{u}}(\cdot)$ and $p_{\mathbf{v}}(\cdot)$, with the extrinsic Gaussian messages ③' and ④', respectively. Messages ① and ② are updated in lines 4–7 of Algorithm 2 in a matrix form (i.e., $\boldsymbol{B}_{\boldsymbol{U},t}$, $\boldsymbol{B}_{\boldsymbol{V},t}$, $\boldsymbol{\Lambda}_{\boldsymbol{U},t}$, and $\boldsymbol{\Lambda}_{\boldsymbol{V},t}$)

[3] A community prior on a random vector, $\mathbf{x}$, in the presence of r communities (i.e., clusters) is given by $p_{\mathbf{x}}(\boldsymbol{x}) = \frac{1}{r}\sum_{l=1}^r \delta(\boldsymbol{x} - \boldsymbol{e}_l)$, where $\boldsymbol{e}_l$ is the l^{th} canonical basis vector in $\mathbb{R}^r$.

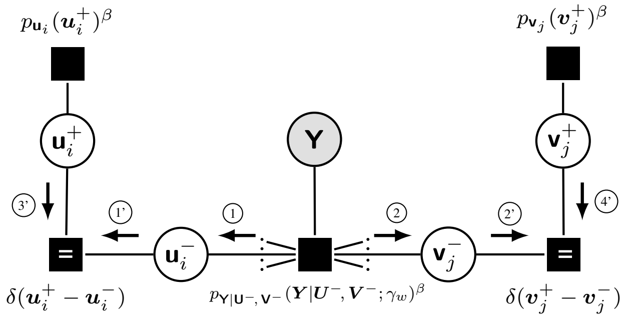
Algorithm 2 BiG-VAMP

```
Require : Matrix $\boldsymbol{Y} \in \mathbb{R}^{N\times M}$; temperature parameter $\beta$; precision tolerance ($\xi = 10^{-6}$); maximum number of iterations ($T_{\max}$); two denoisers $\mathbf{g}_{\mathsf{u}}(\cdot)$ and $\mathbf{g}_{\mathsf{v}}(\cdot)$ from (37) and (38); (noise precision $\gamma_w$ only for Bi-VAMP, i.e., under bilinear observation models)
1: Initialize
2: $t \leftarrow 1$
    ▷ posterior means, covariances and precisions
    $\widehat{\boldsymbol{U}}^{-}_{\mathsf{p},0}, \widehat{\boldsymbol{V}}^{-}_{\mathsf{p},0}, \widehat{\boldsymbol{U}}^{-}_{\mathsf{p},1}, \widehat{\boldsymbol{V}}^{-}_{\mathsf{p},1}, \widehat{\boldsymbol{Z}}^{-}_{\mathsf{p},1}, \boldsymbol{R}_{\boldsymbol{U}_{\mathsf{p}}^{-},1}, \boldsymbol{R}_{\boldsymbol{V}_{\mathsf{p}}^{-},1}, \gamma_{\boldsymbol{Z}_{\mathsf{p}}^{-},1}$
    $\widehat{\boldsymbol{U}}^{+}_{\mathsf{p},1}, \widehat{\boldsymbol{V}}^{+}_{\mathsf{p},1}, \widehat{\boldsymbol{Z}}^{+}_{\mathsf{p},1}, \gamma_{\boldsymbol{U}_{\mathsf{p}}^{-},1}, \gamma_{\boldsymbol{V}_{\mathsf{p}}^{-},1}, \gamma_{\boldsymbol{Z}_{\mathsf{p}}^{+},1}$
    ▷ extrinsic means and precisions
    $\widehat{\boldsymbol{U}}^{-}_{\mathsf{e},1}, \widehat{\boldsymbol{V}}^{-}_{\mathsf{e},1}, \widehat{\boldsymbol{Z}}^{-}_{\mathsf{e},1}, \gamma_{\boldsymbol{U}_{\mathsf{e}}^{-},1}, \gamma_{\boldsymbol{V}_{\mathsf{e}}^{-},1}, \gamma_{\boldsymbol{Z}_{\mathsf{e}}^{-},1}$
    $\widehat{\boldsymbol{U}}^{+}_{\mathsf{e},1}, \widehat{\boldsymbol{V}}^{+}_{\mathsf{e},1}, \widehat{\boldsymbol{Z}}^{+}_{\mathsf{e},1}, \gamma_{\boldsymbol{U}_{\mathsf{e}}^{+},1}, \gamma_{\boldsymbol{V}_{\mathsf{e}}^{+},1}, \gamma_{\boldsymbol{Z}_{\mathsf{e}}^{+},1}$
    ▷ means and precisions in eqs. (13), (12), (14) and (15)
    $\boldsymbol{B}_{\boldsymbol{U},1}, \boldsymbol{\Lambda}_{\boldsymbol{U},1}, \boldsymbol{B}_{\boldsymbol{V},1}, \boldsymbol{\Lambda}_{\boldsymbol{V},1}$
3: repeat
        I. Approximate Bi-LMMSE step
        ▷ Compute the approximated message in (29a)
4:      $\boldsymbol{B}_{\boldsymbol{U},t} = \gamma_{\boldsymbol{Z}_{\mathsf{e}}^{+},t}\Big(\widehat{\boldsymbol{Z}}^{+}_{\mathsf{e},t}\,\widehat{\boldsymbol{V}}^{-}_{\mathsf{p},t} - M\,\gamma_{\boldsymbol{Z}_{\mathsf{e}}^{+},t}\,\widehat{\boldsymbol{U}}^{-}_{\mathsf{p},t-1}\,\boldsymbol{R}_{\boldsymbol{V}_{\mathsf{p}}^{-},t}\,\langle\widehat{\boldsymbol{Z}}^{+}_{\mathsf{e},t}\odot\widehat{\boldsymbol{Z}}^{+}_{\mathsf{e},t}\rangle\Big)$
5:      $\boldsymbol{\Lambda}_{\boldsymbol{U},t} = \gamma_{\boldsymbol{Z}_{\mathsf{e}}^{+},t}\Big(\widehat{\boldsymbol{V}}^{-\top}_{\mathsf{p},t}\,\widehat{\boldsymbol{V}}^{-}_{\mathsf{p},t} + \frac{M}{\beta}\,\boldsymbol{R}_{\boldsymbol{V}_{\mathsf{p}}^{-},t} - M\,\gamma_{\boldsymbol{Z}_{\mathsf{e}}^{+},t}\,\boldsymbol{R}_{\boldsymbol{V}_{\mathsf{p}}^{-},t}\,\langle\widehat{\boldsymbol{Z}}^{+}_{\mathsf{e},t}\odot\widehat{\boldsymbol{Z}}^{+}_{\mathsf{e},t}\rangle\Big)$
        ▷ Compute the approximated message in (29b)
6:      $\boldsymbol{B}_{\boldsymbol{V},t} = \gamma_{\boldsymbol{Z}_{\mathsf{e}}^{+},t}\Big(\widehat{\boldsymbol{Z}}^{+}_{\mathsf{e},t}\,\widehat{\boldsymbol{U}}^{-}_{\mathsf{p},t} - N\,\gamma_{\boldsymbol{Z}_{\mathsf{e}}^{+},t}\,\widehat{\boldsymbol{V}}^{-}_{\mathsf{p},t-1}\,\boldsymbol{R}_{\boldsymbol{U}_{\mathsf{p}}^{-},t}\,\langle\widehat{\boldsymbol{Z}}^{+}_{\mathsf{e},t}\odot\widehat{\boldsymbol{Z}}^{+}_{\mathsf{e},t}\rangle\Big)$
7:      $\boldsymbol{\Lambda}_{\boldsymbol{V},t} = \gamma_{\boldsymbol{Z}_{\mathsf{e}}^{+},t}\Big(\widehat{\boldsymbol{U}}^{-\top}_{\mathsf{p},t}\,\widehat{\boldsymbol{U}}^{-}_{\mathsf{p},t} + \frac{N}{\beta}\,\boldsymbol{R}_{\boldsymbol{U}_{\mathsf{p}}^{-},t} - N\,\gamma_{\boldsymbol{Z}_{\mathsf{e}}^{+},t}\,\boldsymbol{R}_{\boldsymbol{U}_{\mathsf{p}}^{-},t}\,\langle\widehat{\boldsymbol{Z}}^{+}_{\mathsf{e},t}\odot\widehat{\boldsymbol{Z}}^{+}_{\mathsf{e},t}\rangle\Big)$
        ▷ Update the posterior statistics $\widehat{\boldsymbol{U}}^{-}_{\mathsf{p},t}, \boldsymbol{R}_{\boldsymbol{U}_{\mathsf{p}}^{-},t}, \widehat{\boldsymbol{U}}^{-}_{\mathsf{p},t}, \boldsymbol{R}_{\boldsymbol{V}_{\mathsf{p}}^{-},t}$
8:      $\boldsymbol{R}_{\boldsymbol{U}_{\mathsf{p}}^{-},t+1} = (\gamma_{\boldsymbol{U}_{\mathsf{e}}^{+},t}\,\boldsymbol{I} + \boldsymbol{\Lambda}_{\boldsymbol{U},t})^{-1}$
9:      $\widehat{\boldsymbol{U}}^{-}_{\mathsf{p},t+1} = (\boldsymbol{B}_{\boldsymbol{U},t} + \gamma_{\boldsymbol{U}_{\mathsf{e}}^{+},t}\widehat{\boldsymbol{U}}^{+}_{\mathsf{e},t})\,\boldsymbol{R}_{\boldsymbol{U}_{\mathsf{p}}^{-},t+1}$
10:     $\boldsymbol{R}_{\boldsymbol{V}_{\mathsf{p}}^{-},t+1} = (\gamma_{\boldsymbol{V}_{\mathsf{e}}^{+},t}\,\boldsymbol{I} + \boldsymbol{\Lambda}_{\boldsymbol{V},t})^{-1}$
11:     $\widehat{\boldsymbol{V}}^{-}_{\mathsf{p},t+1} = (\boldsymbol{B}_{\boldsymbol{V},t} + \gamma_{\boldsymbol{V}_{\mathsf{e}}^{+},t}\widehat{\boldsymbol{V}}^{+}_{\mathsf{e},t})\,\boldsymbol{R}_{\boldsymbol{V}_{\mathsf{p}}^{-},t+1}$
        II. Denoising step
        ▷ Update the extrinsic statistics $\widehat{\boldsymbol{U}}^{-}_{\mathsf{e},t+1}, \gamma_{\boldsymbol{U}_{\mathsf{e}}^{-},t+1}, \widehat{\boldsymbol{V}}^{-}_{\mathsf{e},t+1}, \gamma_{\boldsymbol{V}_{\mathsf{e}}^{-},t+1}$
12:     $\gamma_{\boldsymbol{U}_{\mathsf{p}}^{-},t+1} = \big(\tfrac{1}{r}\mathrm{Tr}(\boldsymbol{R}_{\boldsymbol{U}_{\mathsf{p}}^{-},t+1})\big)^{-1}$
13:     $\gamma_{\boldsymbol{U}_{\mathsf{e}}^{-},t+1} = \gamma_{\boldsymbol{U}_{\mathsf{p}}^{-},t+1} - \gamma_{\boldsymbol{U}_{\mathsf{e}}^{+},t}$
14:     $\widehat{\boldsymbol{U}}^{-}_{\mathsf{e},t+1} = \gamma^{-1}_{\boldsymbol{U}_{\mathsf{e}}^{-},t+1}\,(\gamma_{\boldsymbol{U}_{\mathsf{p}}^{-},t+1}\widehat{\boldsymbol{U}}^{-}_{\mathsf{p},t+1} - \gamma_{\boldsymbol{U}_{\mathsf{e}}^{+},t}\widehat{\boldsymbol{U}}^{+}_{\mathsf{e},t})$
15:     $\gamma_{\boldsymbol{V}_{\mathsf{p}}^{-},t+1} = \big(\tfrac{1}{r}\mathrm{Tr}(\boldsymbol{R}_{\boldsymbol{V}_{\mathsf{p}}^{-},t+1})\big)^{-1}$
16:     $\gamma_{\boldsymbol{V}_{\mathsf{e}}^{-},t+1} = \gamma_{\boldsymbol{V}_{\mathsf{p}}^{-},t+1} - \gamma_{\boldsymbol{V}_{\mathsf{e}}^{+},t}$
17:     $\widehat{\boldsymbol{V}}^{-}_{\mathsf{e},t+1} = \gamma^{-1}_{\boldsymbol{V}_{\mathsf{e}}^{-},t+1}\,(\gamma_{\boldsymbol{V}_{\mathsf{p}}^{-},t+1}\widehat{\boldsymbol{V}}^{-}_{\mathsf{p},t+1} - \gamma_{\boldsymbol{V}_{\mathsf{e}}^{+},t}\widehat{\boldsymbol{V}}^{+}_{\mathsf{e},t})$
        ▷ Denoising the rows, $\widehat{\boldsymbol{u}}^{-}_{i,\mathsf{p},t+1}$, of $\widehat{\boldsymbol{U}}^{-}_{\mathsf{p},t+1}$ and the columns $\widehat{\boldsymbol{v}}^{-}_{j,\mathsf{p},t+1}$ of $\widehat{\boldsymbol{V}}^{-}_{\mathsf{p},t+1}$
18:     $\forall i$ : update the $i$th row of $\widehat{\boldsymbol{U}}^{+}_{\mathsf{p},t+1}$ as $\widehat{\boldsymbol{u}}^{+}_{i,\mathsf{p},t+1} = \mathbf{g}_{\mathsf{u}}(\widehat{\boldsymbol{u}}^{-}_{i,\mathsf{e},t+1}, \gamma^{-1}_{\boldsymbol{U}_{\mathsf{e}}^{-},t+1})$,
19:     $\forall j$ : update the $j$th column of $\widehat{\boldsymbol{V}}^{+}_{\mathsf{p},t+1}$ as $\widehat{\boldsymbol{v}}^{+}_{j,\mathsf{p},t+1} = \mathbf{g}_{\mathsf{v}}(\widehat{\boldsymbol{v}}^{-}_{j,\mathsf{e},t+1}, \gamma^{-1}_{\boldsymbol{V}_{\mathsf{e}}^{-},t+1})$,
20:     $\gamma_{\boldsymbol{U}_{\mathsf{p}}^{+},t+1} = \gamma_{\boldsymbol{U}_{\mathsf{e}}^{-},t+1}\left(\frac{1}{N}\sum_{i=1}^{N}\langle\mathbf{g}'_{\mathsf{u}}(\widehat{\boldsymbol{u}}^{-}_{i,\mathsf{e},t+1}, \gamma^{-1}_{\boldsymbol{U}_{\mathsf{e}}^{-},t+1})\rangle\right)^{-1}$
21:     $\gamma_{\boldsymbol{V}_{\mathsf{p}}^{+},t+1} = \gamma_{\boldsymbol{V}_{\mathsf{e}}^{-},t+1}\left(\frac{1}{M}\sum_{j=1}^{M}\langle\mathbf{g}'_{\mathsf{v}}(\widehat{\boldsymbol{v}}^{-}_{j,\mathsf{e},t+1}, \gamma^{-1}_{\boldsymbol{U}_{\mathsf{e}}^{-},t+1})\rangle\right)^{-1}$
        ▷ update the extrinsic statistics $\widehat{\boldsymbol{U}}^{+}_{\mathsf{e},t+1}, \gamma_{\boldsymbol{U}^{+}_{\mathsf{e},t+1}}, \widehat{\boldsymbol{V}}^{+}_{\mathsf{e},t+1}, \gamma_{\boldsymbol{V}_{\mathsf{e}}^{+},t+1}$
22:     $\gamma_{\boldsymbol{U}_{\mathsf{e}}^{+},t+1} = \gamma_{\boldsymbol{U}_{\mathsf{p}}^{+},t+1} - \gamma_{\boldsymbol{U}_{\mathsf{e}}^{-},t+1}$
23:     $\gamma_{\boldsymbol{V}_{\mathsf{e}}^{+},t+1} = \gamma_{\boldsymbol{V}_{\mathsf{p}}^{+},t+1} - \gamma_{\boldsymbol{V}_{\mathsf{e}}^{-},t+1}$
24:     $\widehat{\boldsymbol{U}}^{+}_{\mathsf{e},t+1} = \gamma^{-1}_{\boldsymbol{U}_{\mathsf{e}}^{+},t+1}\,(\gamma_{\boldsymbol{U}_{\mathsf{p}}^{+},t+1}\,\widehat{\boldsymbol{U}}^{+}_{\mathsf{p},t+1} - \gamma_{\boldsymbol{U}_{\mathsf{e}}^{-},t+1}\,\widehat{\boldsymbol{U}}^{-}_{\mathsf{e},t+1})$
25:     $\widehat{\boldsymbol{V}}^{+}_{\mathsf{e},t+1} = \gamma^{-1}_{\boldsymbol{V}_{\mathsf{e}}^{+},t+1}\,(\gamma_{\boldsymbol{V}_{\mathsf{p}}^{+},t+1}\,\widehat{\boldsymbol{V}}^{+}_{\mathsf{p},t+1} - \gamma_{\boldsymbol{V}_{\mathsf{e}}^{-},t+1}\widehat{\boldsymbol{V}}^{-}_{\mathsf{e},t+1})$
        III. Generalized output step
        ▷ Compute the posterior statistics $\widehat{\boldsymbol{Z}}^{-}_{\mathsf{p},t+1}$ and $\gamma_{\boldsymbol{Z}_{\mathsf{p}}^{-},t+1}$
26:     $\widehat{\boldsymbol{Z}}^{-}_{\mathsf{p},t+1} = \widehat{\boldsymbol{U}}^{-}_{\mathsf{p},t+1}\,\widehat{\boldsymbol{V}}^{-\top}_{\mathsf{p},t+1} + \frac{\gamma_{\boldsymbol{Z}_{\mathsf{e}}^{+},t}}{\beta}\,\widehat{\boldsymbol{Z}}^{+}_{\mathsf{e},t}\,\mathrm{Tr}(\boldsymbol{R}_{\boldsymbol{U}_{\mathsf{p}}^{-},t+1}\,\boldsymbol{R}^{\top}_{\boldsymbol{V}_{\mathsf{p}}^{-},t+1})$
27:     $\gamma_{\boldsymbol{Z}_{\mathsf{p}}^{-},t+1} = \gamma_{\boldsymbol{Z}_{\mathsf{e}}^{+},t} + MN\Big(\mathrm{Tr}\big(\frac{M\,N}{\beta}\,\boldsymbol{R}_{\boldsymbol{U}_{\mathsf{p}}^{-},t+1}\,\boldsymbol{R}^{\top}_{\boldsymbol{V}_{\mathsf{p}}^{-},t+1}$
                $+ N\,\boldsymbol{R}_{\boldsymbol{U}_{\mathsf{p}}^{-},t+1}\,\widehat{\boldsymbol{V}}^{-\top}_{\mathsf{p},t+1}\,\widehat{\boldsymbol{V}}^{-}_{\mathsf{p},t+1}$
                $+ `M\,\boldsymbol{R}_{\boldsymbol{V}_{\mathsf{p}}^{-},t+1}\,\widehat{\boldsymbol{U}}^{-\top}_{\mathsf{p},t+1}\,\widehat{\boldsymbol{U}}^{-}_{\mathsf{p},t+1})\Big)^{-1}$
        ▷ Compute the extrinsic statistics $\widehat{\boldsymbol{Z}}^{-}_{\mathsf{e},t+1}$ and $\gamma_{\boldsymbol{Z}_{\mathsf{e}}^{-},t+1}$
28:     $\gamma_{\boldsymbol{Z}_{\mathsf{e}}^{-},t+1} = \gamma_{\boldsymbol{Z}_{\mathsf{p}}^{-},t+1} - \gamma_{\boldsymbol{Z}_{\mathsf{e}}^{+},t}$
29:     $\widehat{\boldsymbol{Z}}^{-}_{\mathsf{e},t+1} = \gamma^{-1}_{\boldsymbol{Z}_{\mathsf{e}}^{-},t+1}\,(\widehat{\boldsymbol{Z}}^{-}_{\mathsf{p},t}\,\gamma_{\boldsymbol{Z}_{\mathsf{p}}^{-},t+1} - \widehat{\boldsymbol{Z}}^{+}_{\mathsf{e},t}\,\gamma_{\boldsymbol{Z}_{\mathsf{e}}^{+},t})$
        ▷ Compute the posterior and extrinsic statistics $\widehat{\boldsymbol{Z}}^{+}_{\mathsf{p},t+1}, \gamma_{\boldsymbol{Z}_{\mathsf{p}}^{+},t+1}, \widehat{\boldsymbol{Z}}^{+}_{\mathsf{e},t+1}, \gamma_{\boldsymbol{Z}_{\mathsf{e}}^{+},t+1}$
30:     Compute $\widehat{\boldsymbol{Z}}^{+}_{\mathsf{p},t+1}$ using (43) and $\gamma_{\boldsymbol{Z}_{\mathsf{p}}^{+},t+1} = \frac{1}{MN}\sum_i\sum_j\gamma_{\mathbf{z}^{+}_{ij,\mathsf{p}},t+1}$ using (44)
31:     $\gamma_{\boldsymbol{Z}_{\mathsf{e}}^{+},t+1} = \gamma_{\boldsymbol{Z}_{\mathsf{p}}^{+},t+1} - \gamma_{\boldsymbol{Z}_{\mathsf{e}}^{-},t+1}$
32:     $\widehat{\boldsymbol{Z}}^{+}_{\mathsf{e},t+1} = \gamma_{\boldsymbol{Z}_{\mathsf{e}}^{+},t+1}\,(\widehat{\boldsymbol{Z}}^{+}_{\mathsf{p},t+1}\,\gamma_{\boldsymbol{Z}_{\mathsf{p}}^{+},t+1} - \widehat{\boldsymbol{Z}}^{-}_{\mathsf{e},t+1}\,\gamma_{\boldsymbol{Z}_{\mathsf{e}}^{-},t+1})$
33:     $t \leftarrow t+1$
34: until $\Big(\|\widehat{\boldsymbol{U}}^{+}_{\mathsf{p},t+1} - \widehat{\boldsymbol{U}}^{+}_{\mathsf{p},t}\|^2_{\mathrm{F}} + \|\widehat{\boldsymbol{V}}^{+}_{\mathsf{p},t+1} - \widehat{\boldsymbol{V}}^{+}_{\mathsf{p},t}\|^2_{\mathrm{F}}\Big)$
            $\leq \xi\Big(\|\widehat{\boldsymbol{U}}^{+}_{\mathsf{p},t}\|^2_{\mathrm{F}} + \|\widehat{\boldsymbol{V}}^{+}_{\mathsf{p},t}\|^2_{\mathrm{F}}\Big)$ or $\Big(t > T_{\max}\Big)$
35: return $\widehat{\boldsymbol{U}}^{+}_{\mathsf{p},T_{\max}+1}, \widehat{\boldsymbol{V}}^{+}_{\mathsf{p},T_{\max}+1}$
```

as discussed in Section II. Note, however, that Bi-VAMP avoids the approximation in (30) which is valid for the very-low-rank structure only. To that end, Bi-VAMP incorporates explicitly the contribution of all $y_{k,j}^2$ in (23b) and (24b) during the computation of the broadcast precision matrices $\boldsymbol{\Lambda}_{\boldsymbol{U},t}$ and $\boldsymbol{\Lambda}_{\boldsymbol{V},t}$, respectively.

The posterior estimates of $\boldsymbol{u}_i^-$ and $\boldsymbol{v}_j^-$ and their error covariance matrices are updated in closed forms as shown in lines 8–11 of Algorithm 2. The resulting regularized matrix inverse structure therein suggests that the updates in lines 4–11 are in essence performing approximate bi-LMMSE recovery of $\boldsymbol{U}$ and $\boldsymbol{V}$ under Gaussian prior information. Based on the i.i.d. assumption, we further reduce the posterior covariance matrices to common scalar variances obtained by simply averaging their diagonal entries (see lines 12 and 15 in Algorithm 2) while ignoring the off-diagonal part. As a result, the posterior covariance matrices are given by the scaled identities, $\gamma^{-1}_{\boldsymbol{U}_{\mathsf{p}}^{-},t+1}\boldsymbol{I}$ and $\gamma^{-1}_{\boldsymbol{V}_{\mathsf{p}}^{-},t+1}\boldsymbol{I}$. Such approximation of messages by their means and scalar variances is a common practice in the message passing paradigm, also known as expectation propagation (EP) principle[4]. After computing the posterior estimates and the associated common precision (i.e., $\gamma_{\boldsymbol{U}_{\mathsf{p}}^{-},t+1}$ and $\gamma_{\boldsymbol{V}_{\mathsf{p}}^{-},t+1}$), each processing node $\mathbf{u}_i^-$ (resp. $\mathbf{v}_j^-$) subtracts the contribution of its incoming extrinsic message ③' (resp. ④') before returning the extrinsic messages ①' (resp. ②') to the other side of the equality node. Formally speaking, this amounts to approximating the posterior messages by Gaussian distributions with the same means and variances $\big($i.e., $\mathcal{N}(\boldsymbol{u}_i^+; \widehat{\boldsymbol{u}}^{-}_{i,\mathsf{p},t+1}, \beta^{-1}\gamma^{-1}_{\boldsymbol{U}_{\mathsf{p}}^{-},t+1}\boldsymbol{I})$ and $\mathcal{N}(\boldsymbol{v}_j^+; \widehat{\boldsymbol{v}}^{-}_{j,\mathsf{p},t+1}, \beta^{-1}\gamma^{-1}_{\boldsymbol{V}_{\mathsf{p}}^{-},t+1}\boldsymbol{I})\big)$ and extracting the extrinsic

[4] In error control codes literature, the concept of EP is also know as the turbo principle.


$p_{\mathsf{u}_i}(\boldsymbol{u}_i^+)^\beta$
$p_{\mathsf{v}_j}(\boldsymbol{v}_j^+)^\beta$
$\mathbf{u}_i^+$
Y
$\mathbf{v}_j^+$
3’
1’
1
2
2’
4’
=
$\mathbf{u}_i^-$
$\mathbf{v}_j^-$
=
$\delta(\boldsymbol{u}_i^+ - \boldsymbol{u}_i^-)$
$p_{\mathsf{Y}|\mathsf{U}^-,\mathsf{V}^-}(\boldsymbol{Y}|\boldsymbol{U}^-, \boldsymbol{V}^-; \gamma_w)^\beta$
$\delta(\boldsymbol{v}_j^+ - \boldsymbol{v}_j^-)$

$\textcircled{1}$	$\mathcal{N}(\boldsymbol{u}_i^-; \boldsymbol{\Lambda}^{-1}_{\mathbf{u}_i^-,t}\boldsymbol{b}_{\mathbf{u}_i^-,t}, \beta^{-1}\boldsymbol{\Lambda}^{-1}_{\mathbf{u}_i^-,t})$
$\textcircled{1'}$	$\mathcal{N}(\boldsymbol{u}_i^+; \widehat{\boldsymbol{u}}^-_{i,\mathsf{e},t+1}, \beta^{-1}\gamma^{-1}_{\mathbf{u}^-_{i,\mathsf{e}},t+1}\boldsymbol{I})$
$\textcircled{2}$	$\mathcal{N}(\boldsymbol{v}_j^-; \boldsymbol{\Lambda}^{-1}_{\mathbf{v}_i^-,t}\boldsymbol{b}_{\mathbf{v}_j^-,t}, \beta^{-1}\boldsymbol{\Lambda}^{-1}_{\mathbf{v}_j^-,t})$
$\textcircled{2'}$	$\mathcal{N}(\boldsymbol{v}_j^+; \widehat{\boldsymbol{v}}^-_{j,\mathsf{e},t+1}, \beta^{-1}\gamma^{-1}_{\mathbf{v}^-_{j,\mathsf{e}},t+1}\boldsymbol{I})$
$\textcircled{3'}$	$\mathcal{N}(\boldsymbol{u}_i^-; \widehat{\boldsymbol{u}}^+_{i,\mathsf{e},t}, \beta^{-1}\gamma^{-1}_{\mathbf{u}^+_{i,\mathsf{e}},t}\boldsymbol{I})$
$\textcircled{4'}$	$\mathcal{N}(\boldsymbol{v}_j^-; \widehat{\boldsymbol{v}}^+_{j,\mathsf{e},t}, \beta^{-1}\gamma^{-1}_{\mathbf{v}^+_{j,\mathsf{e}},t}\boldsymbol{I})$

Fig. 3: Factor graph under generalized priors on $\boldsymbol{U}$ and $\boldsymbol{V}$ along with the Gaussian approximations for the extrinsic information (as reflected by the index e) that handles the equality constraints.

messages as follows:

$$\mathcal{N}(\boldsymbol{u}_i^+; \widehat{\boldsymbol{u}}^-_{i,\mathsf{e},t+1}, \beta^{-1}\gamma^{-1}_{\boldsymbol{U}_\mathsf{e}^-,t+1}\boldsymbol{I}) \propto \frac{\mathcal{N}(\boldsymbol{u}_i^+; \widehat{\boldsymbol{u}}^-_{i,\mathsf{p},t+1}, \frac{1}{\beta}\gamma^{-1}_{\boldsymbol{U}_\mathsf{p}^-,t+1}\boldsymbol{I})}{\mathcal{N}(\boldsymbol{u}_i^+; \widehat{\boldsymbol{u}}^+_{i,\mathsf{e},t}, \frac{1}{\beta}\gamma^{-1}_{\boldsymbol{U}_\mathsf{e}^+,t}\boldsymbol{I})},$$

$$\mathcal{N}(\boldsymbol{v}_j^+; \widehat{\boldsymbol{v}}^-_{j,\mathsf{e},t+1}, \beta^{-1}\gamma^{-1}_{\boldsymbol{V}_\mathsf{e}^-,t+1}\boldsymbol{I}) \propto \frac{\mathcal{N}(\boldsymbol{v}_j^+; \widehat{\boldsymbol{v}}^-_{j,\mathsf{p},t+1}, \frac{1}{\beta}\gamma^{-1}_{\boldsymbol{V}_\mathsf{p}^-,t+1}\boldsymbol{I})}{\mathcal{N}(\boldsymbol{v}_j^+; \widehat{\boldsymbol{v}}^+_{j,\mathsf{e},t}, \frac{1}{\beta}\gamma^{-1}_{\boldsymbol{V}_\mathsf{e}^+,t}\boldsymbol{I})}.$$

By doing so, the extrinsic/posterior means and precisions are related as follows:

$$\gamma_{\boldsymbol{U}_\mathsf{e}^-,t+1} = \gamma_{\boldsymbol{U}_\mathsf{p}^-,t+1} - \gamma_{\boldsymbol{U}_\mathsf{e}^+,t}, \tag{33}$$

$$\gamma_{\boldsymbol{V}_\mathsf{e}^-,t+1} = \gamma_{\boldsymbol{V}_\mathsf{p}^-,t+1} - \gamma_{\boldsymbol{V}_\mathsf{e}^+,t}, \tag{34}$$

$$\widehat{\boldsymbol{u}}^-_{i,\mathsf{e},t+1} = \gamma^{-1}_{\boldsymbol{U}_\mathsf{e}^-,t+1}\big(\gamma_{\boldsymbol{U}_\mathsf{p}^-,t+1}\widehat{\boldsymbol{u}}^-_{i,\mathsf{p},t+1} + \gamma_{\boldsymbol{U}_\mathsf{e}^+,t}\widehat{\boldsymbol{u}}^+_{i,\mathsf{e},t}\big), \tag{35}$$

$$\widehat{\boldsymbol{v}}^-_{j,\mathsf{e},t+1} = \gamma^{-1}_{\boldsymbol{V}_\mathsf{e}^-,t+1}\big(\gamma_{\boldsymbol{V}_\mathsf{p}^-,t+1}\widehat{\boldsymbol{v}}^-_{j,\mathsf{p},t+1} + \gamma_{\boldsymbol{V}_\mathsf{e}^+,t}\widehat{\boldsymbol{v}}^+_{j,\mathsf{e},t}\big). \tag{36}$$

These extrinsic precisions and means are updated in lines 13, 14, 16, and 17 of Algorithm 2. Given this extrinsic information, the denoising functions, $\mathrm{g}_\mathsf{u}(\cdot,\cdot)$, and $\mathrm{g}_\mathsf{v}(\cdot,\cdot)$, used to estimate $\boldsymbol{u}_i^+$ and $\boldsymbol{v}_j^+$, along with their respective divergences, $\mathrm{g}'_\mathsf{u}(\cdot,\cdot)$ and $\mathrm{g}'_\mathsf{v}(\cdot,\cdot)$ in lines 18–21 of Algorithm 2 are given by:

$$\mathrm{g}_\mathsf{u}(\widehat{\boldsymbol{u}}, \gamma_{\boldsymbol{U}}^{-1}) = \frac{\displaystyle\int \boldsymbol{u}\, p_\mathbf{u}(\boldsymbol{u})^\beta\, \mathcal{N}(\boldsymbol{u}; \widehat{\boldsymbol{u}}, \beta^{-1}\gamma_{\boldsymbol{U}}^{-1}\boldsymbol{I})\, d\boldsymbol{u}}{\displaystyle\int p_\mathbf{u}(\boldsymbol{u})^\beta\, \mathcal{N}(\boldsymbol{u}; \widehat{\boldsymbol{u}}, \beta^{-1}\gamma_{\boldsymbol{U}}^{-1}\boldsymbol{I})\, d\boldsymbol{u}}, \tag{37}$$

$$\mathrm{g}_\mathsf{v}(\widehat{\boldsymbol{v}}, \gamma_{\boldsymbol{V}}^{-1}) = \frac{\displaystyle\int \boldsymbol{v}\, p_\mathbf{v}(\boldsymbol{v})^\beta\, \mathcal{N}(\boldsymbol{v}; \widehat{\boldsymbol{v}}, \beta^{-1}\gamma_{\boldsymbol{V}}^{-1}\boldsymbol{I})\, d\boldsymbol{v}}{\displaystyle\int p_\mathbf{v}(\boldsymbol{v})^\beta\, \mathcal{N}(\boldsymbol{v}; \widehat{\boldsymbol{v}}, \beta^{-1}\gamma_{\boldsymbol{V}}^{-1}\boldsymbol{I})\, d\boldsymbol{v}}, \tag{38}$$

$$\big[\mathrm{g}'_\mathsf{u}(\widehat{\boldsymbol{u}}, \gamma_{\boldsymbol{U}}^{-1})\big]_\ell = \frac{\partial\big[\mathrm{g}_\mathsf{u}(\widehat{\boldsymbol{u}}, \gamma_{\boldsymbol{U}}^{-1})\big]_\ell}{\partial\big[\widehat{\boldsymbol{u}}\big]_\ell}, \; \ell = 1, \ldots, r, \tag{39}$$

$$\big[\mathrm{g}'_\mathsf{v}(\widehat{\boldsymbol{v}}, \gamma_{\boldsymbol{V}}^{-1})\big]_\ell = \frac{\partial\big[\mathrm{g}_\mathsf{v}(\widehat{\boldsymbol{v}}, \gamma_{\boldsymbol{V}}^{-1})\big]_\ell}{\partial\big[\widehat{\boldsymbol{v}}\big]_\ell}, \; \ell = 1, \ldots, r. \tag{40}$$

In essence, $\widehat{\boldsymbol{u}}^+_{i,\mathsf{p},t+1} = \mathrm{g}_\mathsf{u}(\widehat{\boldsymbol{u}}^-_{i,\mathsf{e},t+1}, \gamma^{-1}_{\boldsymbol{U}_\mathsf{e}^-,t+1})$ and $\widehat{\boldsymbol{v}}^+_{j,\mathsf{p},t+1} = \mathrm{g}_\mathsf{v}(\widehat{\boldsymbol{v}}^-_{j,\mathsf{e},t+1}, \gamma^{-1}_{\boldsymbol{V}_\mathsf{e}^-,t+1})$ are the posterior means of $\mathbf{u}_i^+$ and $\mathbf{v}_j^+$, respectively. In addition, their common posterior precisions updated in lines 20 and 21 of Algorithm 2 are given by:

$$\gamma_{\boldsymbol{U}_\mathsf{p}^+,t+1} = \gamma_{\boldsymbol{U}_\mathsf{e}^-,t+1}\left(\frac{1}{N}\sum_{i=1}^{N}\big\langle \mathrm{g}'_\mathsf{u}(\widehat{\boldsymbol{u}}^-_{i,\mathsf{e},t+1}, \gamma^{-1}_{\boldsymbol{U}_\mathsf{e}^-,t+1})\big\rangle\right)^{-1}, \tag{41}$$

$$\gamma_{\boldsymbol{V}_\mathsf{p}^+,t+1} = \gamma_{\boldsymbol{V}_\mathsf{e}^-,t+1}\left(\frac{1}{M}\sum_{j=1}^{M}\big\langle \mathrm{g}'_\mathsf{v}(\widehat{\boldsymbol{v}}^-_{j,\mathsf{e},t+1}, \gamma^{-1}_{\boldsymbol{U}_\mathsf{e}^-,t+1})\big\rangle\right)^{-1} \tag{42}$$

Notice that unlike the multi-dimensional integrals in (20) and (21) which restrict the existing low-rank matrix recovery algorithms in [5] and [20] to the case of Gaussian and community priors, all the one-dimensional integrals involved in (37) and (38) can be found analytically for almost all statistical priors of practical interest. As one example, the intractable posterior mean of $\mathbf{u}_i$ in (20) with a binary prior becomes straightforwardly equal to $\mathrm{g}_\mathsf{u}(\widehat{\boldsymbol{u}}^-_{i,\mathsf{e},t+1}, \gamma^{-1}_{\boldsymbol{U}_\mathsf{e}^-,t+1}) = \tanh(\gamma_{\boldsymbol{U}_\mathsf{e}^-,t+1}\widehat{\boldsymbol{u}}^-_{i,\mathsf{e},t+1})$ instead of summing over 2^r terms. Finally, the extrinsic precisions and means for the messages $\textcircled{3'}$ and $\textcircled{4'}$ are updated analogously to (33)-(36) in lines 22–25 of Algorithm 2.

B. From Bi-VAMP to BiG-VAMP

In this section, we extend the Bi-VAMP algorithm introduced in Section III-A to the generalized bilinear model given in (1) so as to complete the derivation of BiG-VAMP. To that end, we introduce the intermediate random matrix, $\mathbf{Z} \triangleq \mathbf{U}\mathbf{V}^\top$, whose ijth entry is given by $\mathsf{z}_{ij} = \mathbf{u}_i^\top\mathbf{v}_j$. We again resort to the expectation propagation (aka turbo) principle to approximate the posterior messages $\textcircled{5'}$ and $\textcircled{6'}$ in Fig. 5 by Gaussian distributions, $\mathcal{N}(z_{ij}^-; \widehat{z}^-_{ij,\mathsf{e},t+1}, \beta^{-1}\gamma^{-1}_{\boldsymbol{Z}_\mathsf{e}^-,t+1})$ and $\mathcal{N}(z_{ij}^+; \widehat{z}^+_{ij,\mathsf{e},t}, \beta^{-1}\gamma^{-1}_{\boldsymbol{Z}_\mathsf{e}^+,t})$, respectively, whose means and variances are calculated in the sequel. To start with, by defining $\mathsf{z}_{ij}^- \triangleq \mathbf{u}_i^{-\top}\mathbf{v}_j^-$, the posterior mean and variance, $\widehat{z}^+_{ij,\mathsf{p}}$ and $\gamma^{-1}_{\mathsf{z}^+_{ij,\mathsf{p}}}$, of $\mathsf{z}_{ij}^+ \triangleq \mathbf{u}_i^{+\top}\mathbf{v}_j^+$ under the scalar likelihood, $p_{\mathsf{y}_{ij}|\mathsf{z}_{ij}^+}(y_{ij}|z_{ij}^+)$, are obtained as follows:

$$\widehat{z}^+_{ij,\mathsf{p},t+1} = g_\mathsf{z}(y_{ij}, \widehat{z}^-_{ij,\mathsf{e},t+1}, \gamma^{-1}_{\boldsymbol{Z}_\mathsf{e}^-,t+1}), \tag{43}$$

$$\gamma^{-1}_{\mathsf{z}^+_{ij,\mathsf{p}},t+1} = \gamma^{-1}_{\boldsymbol{Z}_\mathsf{e}^-,t+1}\frac{\partial g_\mathsf{z}\big(y_{ij}, \widehat{z}^-_{ij,\mathsf{e},t+1}, \gamma^{-1}_{\boldsymbol{Z}_\mathsf{e}^-,t+1}\big)}{\partial\widehat{z}^-_{ij,\mathsf{e},t+1}}, \tag{44}$$

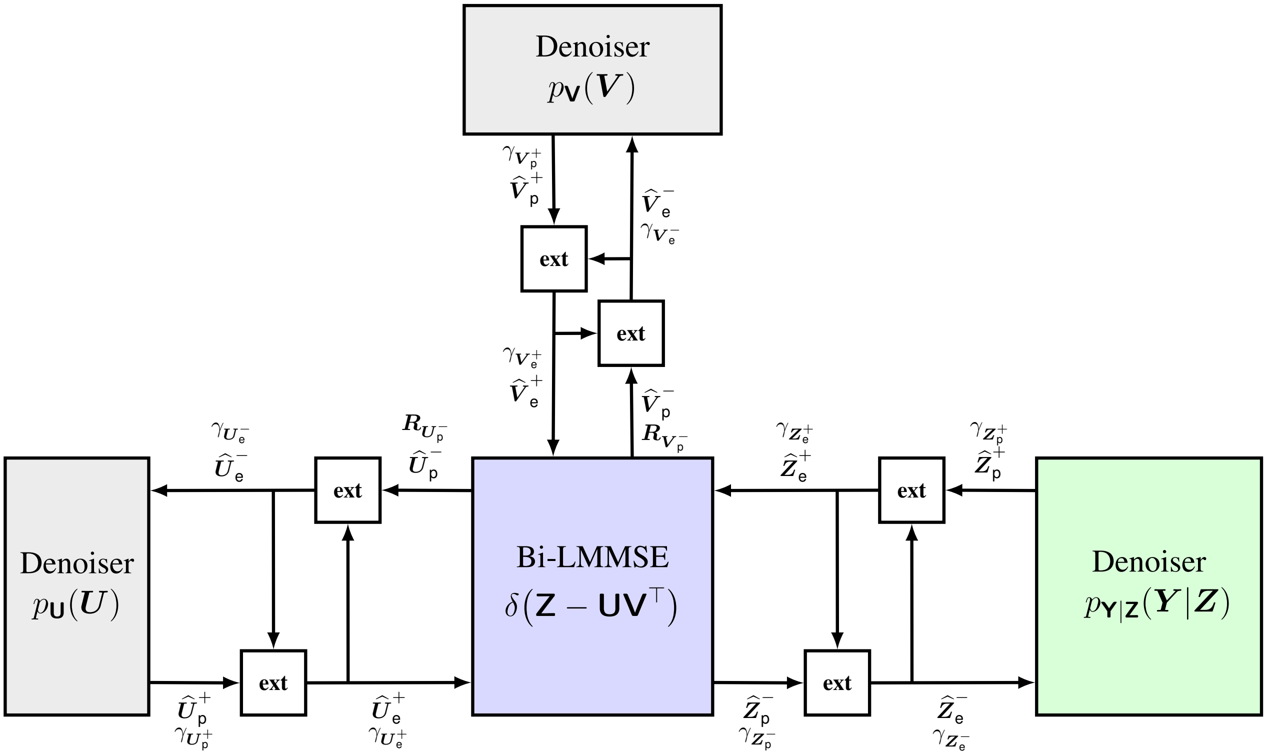

Denoiser
$p_{\mathbf{V}}(\boldsymbol{V})$
$\gamma_{\boldsymbol{V}_{\mathsf{p}}^{+}}$
$\widehat{\boldsymbol{V}}_{\mathsf{p}}^{+}$
$\widehat{\boldsymbol{V}}_{\mathsf{e}}^{-}$
$\gamma_{\boldsymbol{V}_{\mathsf{e}}^{-}}$
ext
ext
$\gamma_{\boldsymbol{V}_{\mathsf{e}}^{+}}$
$\widehat{\boldsymbol{V}}_{\mathsf{e}}^{+}$
$\widehat{\boldsymbol{V}}_{\mathsf{p}}^{-}$
$\boldsymbol{R}_{\boldsymbol{V}_{\mathsf{p}}^{-}}$
$\gamma_{\boldsymbol{U}_{\mathsf{e}}^{-}}$
$\widehat{\boldsymbol{U}}_{\mathsf{e}}^{-}$
$\boldsymbol{R}_{\boldsymbol{U}_{\mathsf{p}}^{-}}$
$\widehat{\boldsymbol{U}}_{\mathsf{p}}^{-}$
$\gamma_{\boldsymbol{Z}_{\mathsf{e}}^{+}}$
$\widehat{\boldsymbol{Z}}_{\mathsf{e}}^{+}$
$\gamma_{\boldsymbol{Z}_{\mathsf{p}}^{+}}$
$\widehat{\boldsymbol{Z}}_{\mathsf{p}}^{+}$
ext
ext
Denoiser
$p_{\mathbf{U}}(\boldsymbol{U})$
Bi-LMMSE
$\delta(\mathbf{Z}-\mathbf{U}\mathbf{V}^{\top})$
Denoiser
$p_{\mathbf{Y}|\mathbf{Z}}(\boldsymbol{Y}|\boldsymbol{Z})$
ext
ext
$\widehat{\boldsymbol{U}}_{\mathsf{p}}^{+}$
$\gamma_{\boldsymbol{U}_{\mathsf{p}}^{+}}$
$\widehat{\boldsymbol{U}}_{\mathsf{e}}^{+}$
$\gamma_{\boldsymbol{U}_{\mathsf{e}}^{+}}$
$\widehat{\boldsymbol{Z}}_{\mathsf{p}}^{-}$
$\gamma_{\boldsymbol{Z}_{\mathsf{p}}^{-}}$
$\widehat{\boldsymbol{Z}}_{\mathsf{e}}^{-}$
$\gamma_{\boldsymbol{Z}_{\mathsf{e}}^{-}}$

Fig. 4: Block diagram of BiG-VAMP with its four modules: two denoising modules (MMSE or MAP) incorporating the prior information, $p_{\mathbf{U}}(\cdot)$ and $p_{\mathbf{V}}(\cdot)$, the approximate bi-LMMSE module, and one output denoising module (MMSE or MAP) handling the observation model $p_{\mathbf{Y}|\mathbf{Z}}(\boldsymbol{Y}|\boldsymbol{Z})$. The four modules exchange extrinsic information/messages through the $\boxed{\textbf{ext}}$ blocks.

.

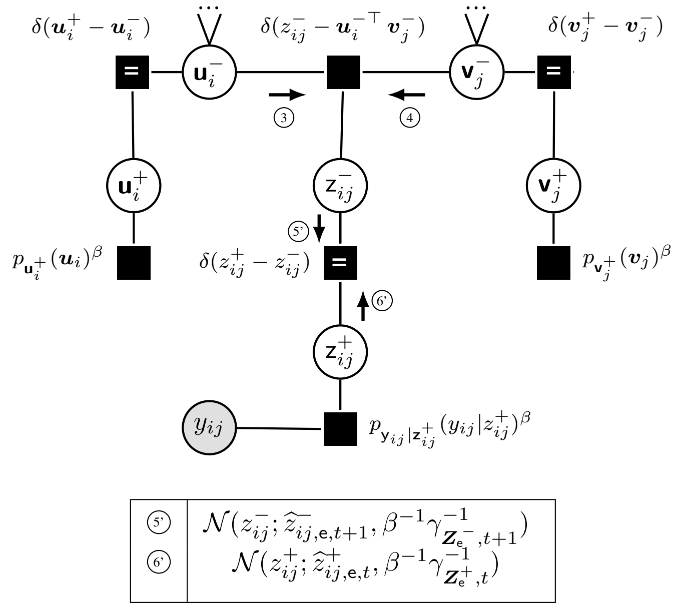

$\delta(\boldsymbol{u}_i^{+}-\boldsymbol{u}_i^{-})$
$\delta(z_{ij}^{-}-\boldsymbol{u}_i^{-\top}\boldsymbol{v}_j^{-})$
$\delta(\boldsymbol{v}_j^{+}-\boldsymbol{v}_j^{-})$
=
$\mathbf{u}_i^{-}$
$\mathbf{v}_j^{-}$
=
③
④
$\mathbf{u}_i^{+}$
z_{ij}^{-}
$\mathbf{v}_j^{+}$
5'
$p_{\mathbf{u}_i^{+}}(\boldsymbol{u}_i)^{\beta}$
$\delta(z_{ij}^{+}-z_{ij}^{-})$
=
$p_{\mathbf{v}_j^{+}}(\boldsymbol{v}_j)^{\beta}$
6'
z_{ij}^{+}
y_{ij}
$p_{\mathsf{y}_{ij}|\mathsf{z}_{ij}^{+}}(y_{ij}|z_{ij}^{+})^{\beta}$
5' $\mathcal{N}(z_{ij}^{-};\widehat{z}_{ij,\mathsf{e},t+1}^{-},\beta^{-1}\gamma_{\boldsymbol{Z}_{\mathsf{e}}^{-},t+1}^{-1})$
6' $\mathcal{N}(z_{ij}^{+};\widehat{z}_{ij,\mathsf{e},t}^{+},\beta^{-1}\gamma_{\boldsymbol{Z}_{\mathsf{e}}^{+},t}^{-1})$

Fig. 5: Factor graph for the generalized bilinear signal recovery problem

.

with $g_{\mathsf{z}}(y,\widehat{z},\gamma_z^{-1})$ being the following scalar denoising function:

$$g_{\mathsf{z}}(y,\widehat{z},\gamma_z^{-1}) \;=\; \frac{\int_{-\infty}^{+\infty} z\,\mathcal{N}(z;\widehat{z},\beta^{-1}\gamma_z^{-1})\,p_{\mathsf{y}|\mathsf{z}}(y|z)^{\beta}\;dz}{\int_{-\infty}^{+\infty}\;\mathcal{N}(z;\widehat{z},\beta^{-1}\gamma_z^{-1})\,p_{\mathsf{y}|\mathsf{z}}(y|z)^{\beta}\;dz}. \quad (45)$$

Now, we detail the derivation of the scalar extrinsic message ⑤' which is approximated by the Gaussian density[5] $\mathcal{N}(z_{ij}^{+};\widehat{z}_{ij,\mathsf{e},t+1}^{-},\beta^{-1}\gamma_{\boldsymbol{Z}_{\mathsf{e}}^{-},t+1}^{-1})$ with a common variance for all nodes. We start with the derivation of the individual variance, $\beta^{-1}\gamma_{z_{ij,\mathsf{e}}^{-},t+1}^{-1}$, of z_{ij}^{-}:

$$\beta^{-1}\gamma_{\mathsf{z}_{ij,\mathsf{e}}^{-},t+1}^{-1} \;=\; \mathbb{E}\Big[\big(\mathbf{u}_i^{-\top}\mathbf{v}_j^{-}-\mathbb{E}[\mathbf{u}_i^{-\top}\mathbf{v}_j^{-}]\big)^2\Big]. \quad (46)$$

In (46), the expectation is taken with respect to the densities on $\mathbf{u}_i^{-}$ and $\mathbf{v}_j^{-}$ while assuming them to be independent.Those densities are given by messages ③ and ④, at iteration at iteration t, whose first- and second-order statistics were already evaluated in (16)-(19). Therefore, it follows from (46) that:

$$\begin{aligned}
&\beta^{-1}\gamma_{\mathsf{z}_{ij,\mathsf{e}}^{-},t+1}^{-1}\\
&= \mathbb{E}\big[(\mathbf{u}_i^{-\top}\mathbf{v}_j^{-})^2\big] \;-\; \big(\widehat{\boldsymbol{u}}_{i\to(i,j),t+1}^{\top}\widehat{\boldsymbol{v}}_{j\to(i,j),t+1}\big)^2,\\
&= \mathbb{E}\Big[\mathrm{Tr}\big(\mathbf{u}_i^{-}\,\mathbf{u}_i^{-\top}\mathbf{v}_j^{-}\,\mathbf{v}_j^{-\top}\big)\Big] \;-\; \big(\widehat{\boldsymbol{u}}_{i\to(i,j),t+1}^{\top}\widehat{\boldsymbol{v}}_{j\to(i,j),t+1}\big)^2,\\
&= \mathrm{Tr}\Big(\mathbb{E}\big[\mathbf{u}_i^{-}\,\mathbf{u}_i^{-\top}\big]\,\mathbb{E}\big[\mathbf{v}_j^{-}\,\mathbf{v}_j^{-\top}\big]\Big) \;-\; \big(\widehat{\boldsymbol{u}}_{i\to(i,j),t+1}^{\top}\widehat{\boldsymbol{v}}_{j\to(i,j),t+1}\big)^2.
\end{aligned} \quad (47)$$

[5] Note that by virtue of the CLT $\mathsf{z}_{ij}^{-}\triangleq\mathbf{u}_i^{-\top}\mathbf{v}_j^{-}$ (up to an appropriate scaling) converges to a Gaussian random variable in the large system limit. Hence, finding its first- and second-order statistics is enough to completely specify its distribution. Approximating its extrinsic information by a Gaussian density is thus equivalent to performing exact message passing.

Using the covriance identity, $\mathrm{cov}(\mathbf{x},\mathbf{x}) = \mathbb{E}[\mathbf{x}\,\mathbf{x}^\top] - \mathbb{E}[\mathbf{x}]\,\mathbb{E}[\mathbf{x}]^\top$, for any random vector $\mathbf{x}$, (47) is equivalent to:

$$\begin{aligned}
&\beta^{-1}\,\gamma^{-1}_{z^-_{ij,\mathsf{e}},t+1}\\
&\quad= \mathrm{Tr}\Big(\Big[\mathrm{cov}(\mathbf{u}^-_i,\mathbf{u}^-_i) + \widehat{\boldsymbol{u}}_{i\to(i,j),t+1}\,\widehat{\boldsymbol{u}}^\top_{i\to(i,j),t+1}\Big]\\
&\qquad\qquad \times \Big[\mathrm{cov}(\mathbf{v}^-_j,\mathbf{v}^-_j) + \widehat{\boldsymbol{v}}_{j\to(i,j),t+1}\,\widehat{\boldsymbol{v}}^\top_{j\to(i,j),t+1}\Big]\Big)\\
&\qquad\qquad\qquad - \Big(\widehat{\boldsymbol{u}}^\top_{i\to(i,j),t+1}\widehat{\boldsymbol{v}}_{j\to(i,j),t+1}\Big)^2. \qquad (48)
\end{aligned}$$

By recalling the fact that $\mathrm{cov}(\mathbf{u}^-_i,\mathbf{u}^-_i) \triangleq \beta^{-1}\boldsymbol{R}_{\boldsymbol{u}_i,t+1} = \beta^{-1}\boldsymbol{R}_{\boldsymbol{U}^-_\mathsf{p},t+1}$ and $\mathrm{cov}(\mathbf{v}^-_j,\mathbf{v}^-_j) \triangleq \beta^{-1}\boldsymbol{R}_{\boldsymbol{v}_j,t+1} = \beta^{-1}\boldsymbol{R}_{\boldsymbol{V}^-_\mathsf{p},t+1}$, with common $\boldsymbol{R}_{\boldsymbol{U}^-_\mathsf{p},t+1}$ and $\boldsymbol{R}_{\boldsymbol{V}^-_\mathsf{p},t+1}$ from line 8 and 10 in Algorithm 2, it follows that:

$$\begin{aligned}
\gamma^{-1}_{z^-_{ij,\mathsf{e}},t+1} = \mathrm{Tr}\Big(&\beta^{-1}\boldsymbol{R}_{\boldsymbol{U}^-_\mathsf{p},t+1}\boldsymbol{R}_{\boldsymbol{V}^-_\mathsf{p},t+1}\\
&+ \boldsymbol{R}_{\boldsymbol{V}^-_\mathsf{p},t+1}\widehat{\boldsymbol{u}}_{i\to(i,j),t+1}\,\widehat{\boldsymbol{u}}^\top_{i\to(i,j),t+1}\\
&+ \boldsymbol{R}_{\boldsymbol{U}^-_\mathsf{p},t+1}\widehat{\boldsymbol{v}}_{j\to(i,j),t+1}\,\widehat{\boldsymbol{v}}^\top_{j\to(i,j),t+1}\Big). \qquad (49)
\end{aligned}$$

In (49), we further replace $\widehat{\boldsymbol{u}}_{i\to(i,j),t+1}$ and $\widehat{\boldsymbol{v}}_{j\to(i,j),t+1}$ by their broadcast versions $\widehat{\boldsymbol{u}}^-_{i,\mathsf{p},t+1}$ and $\widehat{\boldsymbol{v}}^-_{j,\mathsf{p},t+1}$, respectively, while incurring a negligible error after ignoring terms of vanishing order as M and N grow large:

$$\begin{aligned}
\gamma^{-1}_{z^-_{ij,\mathsf{e}},t+1} = \mathrm{Tr}\Big[&\beta^{-1}\boldsymbol{R}_{\boldsymbol{U}^-_\mathsf{p},t+1}\boldsymbol{R}_{\boldsymbol{V}^-_\mathsf{p},t+1}\\
&+ \boldsymbol{R}_{\boldsymbol{V}^-_\mathsf{p},t+1}\widehat{\boldsymbol{u}}^-_{i,\mathsf{p},t+1}\,\widehat{\boldsymbol{u}}^{-\top}_{i,\mathsf{p},t+1}\\
&+ \boldsymbol{R}_{\boldsymbol{U}^-_\mathsf{p},t+1}\widehat{\boldsymbol{v}}^-_{j,\mathsf{p},t+1}\,\widehat{\boldsymbol{v}}^{-\top}_{j,\mathsf{p},t+1}\Big]. \qquad (50)
\end{aligned}$$

As usually done in approximate message passing practices, we combine the individual variances, $\gamma^{-1}_{z^-_{ij,\mathsf{e}},t+1}$, into one common variance, $\gamma^{-1}_{\boldsymbol{Z}^-_\mathsf{e},t+1}$, for all nodes:

$$\begin{aligned}
\gamma^{-1}_{\boldsymbol{Z}^-_\mathsf{e},t+1} &= \frac{1}{MN}\sum_{i=1}^{N}\sum_{j=1}^{M}\gamma^{-1}_{z^-_{ij,\mathsf{e}},t+1}\\
&\approx \mathrm{Tr}\Big(\beta^{-1}\boldsymbol{R}_{\boldsymbol{U}^-_\mathsf{p},t+1}\boldsymbol{R}_{\boldsymbol{V}^-_\mathsf{p},t+1}\\
&\qquad + \boldsymbol{R}_{\boldsymbol{V}^-_\mathsf{p},t+1}\Big[\tfrac{1}{N}\textstyle\sum_{i=1}^{N}\widehat{\boldsymbol{u}}_{i,\mathsf{p},t+1}\,\widehat{\boldsymbol{u}}^\top_{i,\mathsf{p},t+1}\Big]\\
&\qquad + \boldsymbol{R}_{\boldsymbol{U}^-_\mathsf{p},t+1}\Big[\tfrac{1}{M}\textstyle\sum_{j=1}^{M}\widehat{\boldsymbol{v}}_{j,\mathsf{p},t+1}\,\widehat{\boldsymbol{v}}^\top_{j,\mathsf{p},t+1}\Big]\Big),
\end{aligned}$$

where terms of vanishing order are ignored as M and N grow large. This is used to find the posterior precision, $\gamma_{\boldsymbol{Z}^-_\mathsf{p}}$, as follows:

$$\gamma_{\boldsymbol{Z}^-_\mathsf{p},t+1} = \gamma_{\boldsymbol{Z}^+_\mathsf{e},t} + \gamma_{\boldsymbol{Z}^-_\mathsf{e},t+1}. \qquad (51)$$

Moreover, the posterior mean, $\widehat{z}^-_{ij,\mathsf{p},t+1}$ is evaluated as follows:

$$\begin{aligned}
&\widehat{z}^-_{ij,\mathsf{p},t+1}\\
&\quad= \gamma^{-1}_{\boldsymbol{Z}^-_\mathsf{p},t+1}\Big(\gamma_{\boldsymbol{Z}^+_\mathsf{e},t}\widehat{z}^+_{ij,\mathsf{e},t} + \gamma_{z^-_{ij,\mathsf{e}},t+1}\,\widehat{\boldsymbol{u}}^\top_{i\to(i,j),t+1}\widehat{\boldsymbol{v}}_{j\to(i,j),t+1}\Big). \qquad (52)
\end{aligned}$$

Recall here the Osanger correction terms in (29a) and (29b) which lead to the following approximations of $\widehat{\boldsymbol{u}}_{i\to(i,j),t+1}$ and $\widehat{\boldsymbol{v}}_{j\to(i,j),t+1}$:

$$\widehat{\boldsymbol{u}}_{i\to(i,j),t+1} \approx \widehat{\boldsymbol{u}}^-_{i,\mathsf{p},t+1} - \gamma_{\boldsymbol{Z}^+_\mathsf{e},t}\,\widehat{z}^+_{ij,\mathsf{e},t}\,\boldsymbol{R}_{\boldsymbol{U}^-_\mathsf{p},t+1}\,\widehat{\boldsymbol{v}}^-_{j,\mathsf{p},t}, \qquad (53a)$$

$$\widehat{\boldsymbol{v}}_{j\to(i,j),t+1} \approx \widehat{\boldsymbol{v}}^-_{j,\mathsf{p},t+1} - \gamma_{\boldsymbol{Z}^+_\mathsf{e},t}\,\widehat{z}^+_{ij,\mathsf{e},t}\,\boldsymbol{R}_{\boldsymbol{V}^-_\mathsf{p},t+1}\,\widehat{\boldsymbol{u}}^-_{i,\mathsf{p},t}. \qquad (53b)$$

These approximated messages are injected back into (52) thereby yielding the approximate posterior mean in (54) on the top of the next page. The approximation in (54c) is a result of dropping the last term which has a vanishing order as $M \to \infty$. The approximation in (54d) follows from the observation that one makes an error of vanishing order (as both M and N grow large) due to the fact that $\widehat{\boldsymbol{u}}^-_{i,\mathsf{p},t}\widehat{\boldsymbol{u}}^{-\top}_{i,\mathsf{p},t+1} \approx \widehat{\boldsymbol{u}}^-_{i,\mathsf{p},t+1}\widehat{\boldsymbol{u}}^{-\top}_{i,\mathsf{p},t+1}$ and $\widehat{\boldsymbol{v}}^-_{j,\mathsf{p},t+1}\widehat{\boldsymbol{v}}^{-\top}_{j,\mathsf{p},t} \approx \widehat{\boldsymbol{v}}^-_{j,\mathsf{p},t+1}\widehat{\boldsymbol{v}}^{-\top}_{j,\mathsf{p},t+1}$. Then using (50) in (54d) leads to (54e) in which we further replace $\gamma_{z^-_{ij,\mathsf{e}},t+1}$ by $\gamma_{\boldsymbol{Z}^-_\mathsf{e},t+1}$ thereby yielding:

$$\begin{aligned}
\widehat{z}^-_{ij,\mathsf{p},t+1} = \frac{\gamma_{\boldsymbol{Z}^-_\mathsf{e},t+1}}{\gamma_{\boldsymbol{Z}^-_\mathsf{p},t+1}}\Big(&\widehat{\boldsymbol{u}}^{-\top}_{i,\mathsf{p},t+1}\widehat{\boldsymbol{v}}^-_{j,\mathsf{p},t+1}\\
&+ \beta^{-1}\gamma_{\boldsymbol{Z}^+_\mathsf{e},t}\widehat{z}^+_{ij,\mathsf{e},t}\,\mathrm{Tr}\big(\boldsymbol{R}_{\boldsymbol{U}^-_\mathsf{p},t+1}\,\boldsymbol{R}_{\boldsymbol{V}^-_\mathsf{p},t+1}\big)\Big). \qquad (55)
\end{aligned}$$

After plugging the expression of $\gamma_{\boldsymbol{Z}^-_\mathsf{e},t+1}$ as obtained from (51), that is $\gamma_{\boldsymbol{Z}^-_\mathsf{e},t+1} = \gamma_{\boldsymbol{Z}^-_\mathsf{p},t+1} - \gamma_{\boldsymbol{Z}^+_\mathsf{e},t}$, (55) becomes:

$$\begin{aligned}
\widehat{z}^-_{ij,\mathsf{p},t+1} = \frac{\gamma_{\boldsymbol{Z}^-_\mathsf{p},t+1} - \gamma_{\boldsymbol{Z}^+_\mathsf{e},t}}{\gamma_{\boldsymbol{Z}^-_\mathsf{p},t+1}}\Big(&\widehat{\boldsymbol{u}}^{-\top}_{i,\mathsf{p},t+1}\widehat{\boldsymbol{v}}^-_{j,\mathsf{p},t+1}\\
&+ \beta^{-1}\gamma_{\boldsymbol{Z}^+_\mathsf{e},t}\widehat{z}^+_{ij,\mathsf{e},t}\,\mathrm{Tr}\big(\boldsymbol{R}_{\boldsymbol{U}^-_\mathsf{p},t+1}\,\boldsymbol{R}_{\boldsymbol{V}^-_\mathsf{p},t+1}\big)\Big). \qquad (56)
\end{aligned}$$

Finally, we assume that $\gamma_{\boldsymbol{Z}^-_\mathsf{p},t+1} \gg \gamma_{\boldsymbol{Z}^+_\mathsf{e},t}$ which follows from the observation that the information on $z_{ij} = \boldsymbol{u}_i^\top\boldsymbol{v}_j$ that is brought by the strong structure on both $\boldsymbol{U}$ and $\boldsymbol{V}$ overwhelms the information brought by the observation $\boldsymbol{Y}$. This leads to:

$$\begin{aligned}
\widehat{z}^-_{ij,\mathsf{p},t+1} \approx\ &\widehat{\boldsymbol{u}}^{-\top}_{i,\mathsf{p},t+1}\widehat{\boldsymbol{v}}^-_{j,\mathsf{p},t+1}\\
&+ \beta^{-1}\gamma_{\boldsymbol{Z}^+_\mathsf{e},t}\widehat{z}^+_{ij,\mathsf{e},t}\,\mathrm{Tr}\big(\boldsymbol{R}_{\boldsymbol{U}^-_\mathsf{p},t+1}\,\boldsymbol{R}_{\boldsymbol{V}^-_\mathsf{p},t+1}\big). \qquad (57)
\end{aligned}$$

In summary, (57) and (51) are the element-wise expressions of (58a) and (58b), respectively:

$$\widehat{\boldsymbol{Z}}^-_{\mathsf{p},t+1} = \widehat{\boldsymbol{U}}^-_{\mathsf{e},t+1}\,\widehat{\boldsymbol{V}}^{-\top}_{\mathsf{e},t+1} + \frac{\gamma_{\boldsymbol{Z}^+_\mathsf{e},t}}{\beta}\,\widehat{\boldsymbol{Z}}^+_{\mathsf{e},t}\,\mathrm{Tr}\left(\boldsymbol{R}_{\boldsymbol{U}^-_\mathsf{p},t+1}\,\boldsymbol{R}_{\boldsymbol{V}^-_\mathsf{p},t+1}\right), \qquad (58a)$$

$$\begin{aligned}
\gamma_{\boldsymbol{Z}^-_\mathsf{p},t+1} = \gamma_{\boldsymbol{Z}^+_\mathsf{e},t} + MN\,\mathrm{Tr}\Big(&\frac{M\,N}{\beta}\,\boldsymbol{R}_{\boldsymbol{U}^-_\mathsf{p},t+1}\,\boldsymbol{R}_{\boldsymbol{V}^-_\mathsf{p},t+1}\\
&+ N\,\boldsymbol{R}_{\boldsymbol{U}^-_\mathsf{p},t+1}\,\widehat{\boldsymbol{V}}^{-\top}_{\mathsf{p},t+1}\,\widehat{\boldsymbol{V}}^-_{\mathsf{p},t+1}\\
&+ M\,\boldsymbol{R}_{\boldsymbol{V}^-_\mathsf{p},t+1}\,\widehat{\boldsymbol{U}}^{-\top}_{\mathsf{p},t+1}\,\widehat{\boldsymbol{U}}^-_{\mathsf{p},t+1}\Big)^{-1}, \qquad (58b)
\end{aligned}$$

which correspond to lines 26–27 in Algorithm 2. The extrinsic values, $\widehat{z}^-_{ij,\mathsf{e},t+1}$ and $\gamma^{-1}_{\boldsymbol{Z}^-_\mathsf{e},t+1}$, for message $\textcircled{5'}$ in Fig. 5 are then easily evaluated. Finally, the extrinsic mean and variance, $\widehat{z}^+_{ij,\mathsf{e},t+1}$ and $\gamma^{-1}_{\boldsymbol{Z}^+_\mathsf{e},t+1}$, of message $\textcircled{6'}$ can be calculated from the posterior mean, $\widehat{z}^+_{ij,\mathsf{p},t+1}$ and common precision, $\gamma_{\boldsymbol{Z}^+_\mathsf{p},t+1}$, as specified in lines 31–32 of Algorithm 2.

IV. State Evolution

Our main goal is to understand the behavior of the proposed BiG-VAMP algorithm in the asymptotic regime for a certain class of $\boldsymbol{U}$ and $\boldsymbol{V}$ matrices, i.e., when they both have zero mean i.i.d. priors. For simplicity, we will focus on MMSE estimation, i.e., $\beta = 1$. In our case, the asymptotic regime refers to the case where

$$\widehat{z}_{ij,\mathsf{p},t+1}^{-} \approx \gamma_{\boldsymbol{Z}_{\mathsf{p}}^{-},t+1}^{-1}\left[\gamma_{\boldsymbol{Z}_{\mathsf{e}}^{+},t}\widehat{z}_{ij,\mathsf{e},t}^{+} + \gamma_{z_{ij,\mathsf{e}}^{-},t+1}\Big(\widehat{\boldsymbol{u}}_{i,\mathsf{p},t+1}^{-} - \gamma_{\boldsymbol{Z}_{\mathsf{e}}^{+},t}\,\widehat{z}_{ij,\mathsf{e},t}^{+}\,\boldsymbol{R}_{\boldsymbol{U}_{\mathsf{p}}^{-},t+1}\,\widehat{\boldsymbol{v}}_{j,\mathsf{p},t}^{-}\Big)^{\top}\Big(\widehat{\boldsymbol{v}}_{j,\mathsf{p},t+1}^{-} - \gamma_{\boldsymbol{Z}_{\mathsf{e}}^{+},t}\,\widehat{z}_{ij,\mathsf{e},t}^{+}\,\boldsymbol{R}_{\boldsymbol{V}_{\mathsf{p}}^{-},t+1}\,\widehat{\boldsymbol{u}}_{i,\mathsf{p},t}^{-}\Big)\right], \tag{54a}$$

$$\begin{aligned} = \gamma_{\boldsymbol{Z}_{\mathsf{p}}^{-},t+1}^{-1}\Big[\gamma_{\boldsymbol{Z}_{\mathsf{e}}^{+},t}\widehat{z}_{ij,\mathsf{e},t}^{+} + \gamma_{z_{ij,\mathsf{e}}^{-},t+1}\Big(&\widehat{\boldsymbol{u}}_{i,\mathsf{p},t+1}^{-\top}\,\widehat{\boldsymbol{v}}_{j,\mathsf{p},t+1}^{-} - \gamma_{\boldsymbol{Z}_{\mathsf{e}}^{+},t}\,\widehat{z}_{ij,\mathsf{e},t}^{+}\,\widehat{\boldsymbol{u}}_{i,\mathsf{p},t+1}^{-\top}\,\boldsymbol{R}_{\boldsymbol{V}_{\mathsf{p}}^{-},t+1}\,\widehat{\boldsymbol{u}}_{i,\mathsf{p},t}^{-} \\ &- \gamma_{\boldsymbol{Z}_{\mathsf{e}}^{+},t}\,\widehat{z}_{ij,\mathsf{e},t}^{+}\,\widehat{\boldsymbol{v}}_{j,\mathsf{p},t}^{-\top}\,\boldsymbol{R}_{\boldsymbol{U}_{\mathsf{p}}^{-},t+1}\,\widehat{\boldsymbol{v}}_{j,\mathsf{p},t+1}^{-} + (\gamma_{\boldsymbol{Z}_{\mathsf{e}}^{+},t}\,\widehat{z}_{ij,\mathsf{e},t}^{+})^{2}\,\widehat{\boldsymbol{v}}_{j,\mathsf{p},t}^{-\top}\boldsymbol{R}_{\boldsymbol{U}_{\mathsf{p}}^{-},t+1}\,\boldsymbol{R}_{\boldsymbol{V}_{\mathsf{p}}^{-},t+1}\,\widehat{\boldsymbol{u}}_{i,\mathsf{p},t}^{-}\Big)\Big], \end{aligned} \tag{54b}$$

$$\begin{aligned} \approx \gamma_{\boldsymbol{Z}_{\mathsf{p}}^{-},t+1}^{-1}\Big[\gamma_{\boldsymbol{Z}_{\mathsf{e}}^{+},t}\widehat{z}_{ij,\mathsf{e},t}^{+} + \gamma_{z_{ij,\mathsf{e}}^{-},t+1}\Big(&\widehat{\boldsymbol{u}}_{i,\mathsf{p},t+1}^{-\top}\,\widehat{\boldsymbol{v}}_{j,\mathsf{p},t+1}^{-} - \gamma_{\boldsymbol{Z}_{\mathsf{e}}^{+},t}\,\widehat{z}_{ij,\mathsf{e},t}^{+}\,\mathrm{Tr}\big(\boldsymbol{R}_{\boldsymbol{V}_{\mathsf{p}}^{-},t+1}\,\widehat{\boldsymbol{u}}_{i,\mathsf{p},t}^{-}\widehat{\boldsymbol{u}}_{i,\mathsf{p},t+1}^{-\top}\big) \\ &- \gamma_{\boldsymbol{Z}_{\mathsf{e}}^{+},t}\,\widehat{z}_{ij,\mathsf{e},t}^{+}\,\mathrm{Tr}\big(\boldsymbol{R}_{\boldsymbol{U}_{\mathsf{p}}^{-},t+1}\,\widehat{\boldsymbol{v}}_{j,\mathsf{p},t+1}^{-}\widehat{\boldsymbol{v}}_{j,\mathsf{p},t}^{-\top}\big)\Big)\Big], \end{aligned} \tag{54c}$$

$$\begin{aligned} \approx \gamma_{\boldsymbol{Z}_{\mathsf{p}}^{-},t+1}^{-1}\Big[\gamma_{\boldsymbol{Z}_{\mathsf{e}}^{+},t}\widehat{z}_{ij,\mathsf{e},t}^{+} + \gamma_{z_{ij,\mathsf{e}}^{-},t+1}\Big(&\widehat{\boldsymbol{u}}_{i,\mathsf{p},t+1}^{-\top}\,\widehat{\boldsymbol{v}}_{j,\mathsf{p},t+1}^{-} - \gamma_{\boldsymbol{Z}_{\mathsf{e}}^{+},t}\,\widehat{z}_{ij,\mathsf{e},t}^{+}\,\mathrm{Tr}\big(\boldsymbol{R}_{\boldsymbol{V}_{\mathsf{p}}^{-},t+1}\,\widehat{\boldsymbol{u}}_{i,\mathsf{p},t+1}^{-}\widehat{\boldsymbol{u}}_{i,\mathsf{p},t+1}^{-\top}\big) \\ &- \gamma_{\boldsymbol{Z}_{\mathsf{e}}^{+},t}\,\widehat{z}_{ij,\mathsf{e},t}^{+}\,\mathrm{Tr}\big(\boldsymbol{R}_{\boldsymbol{U}_{\mathsf{p}}^{-},t+1}\,\widehat{\boldsymbol{v}}_{j,\mathsf{p},t+1}^{-}\widehat{\boldsymbol{v}}_{j,\mathsf{p},t+1}^{-\top}\big)\Big)\Big], \end{aligned} \tag{54d}$$

$$\begin{aligned} = \gamma_{\boldsymbol{Z}_{\mathsf{p}}^{-},t+1}^{-1}\Big[\gamma_{\boldsymbol{Z}_{\mathsf{e}}^{+},t}\widehat{z}_{ij,\mathsf{e},t}^{+} + \gamma_{z_{ij,\mathsf{e}}^{-},t+1}&\Big(\widehat{\boldsymbol{u}}_{i,\mathsf{p},t+1}^{-\top}\,\widehat{\boldsymbol{v}}_{j,\mathsf{p},t+1}^{-} - \gamma_{\boldsymbol{Z}_{\mathsf{e}}^{+},t}\,\gamma_{z_{ij,\mathsf{e}}^{-},t+1}^{-1}\widehat{z}_{ij,\mathsf{e},t}^{+}\Big) \\ &+ \beta^{-1}\gamma_{\boldsymbol{Z}_{\mathsf{e}}^{+},t}\gamma_{z_{ij,\mathsf{e}}^{-},t+1}\,\widehat{z}_{ij,\mathsf{e},t}^{+}\,\mathrm{Tr}\big(\boldsymbol{R}_{\boldsymbol{U}_{\mathsf{p}}^{-},t+1}\,\boldsymbol{R}_{\boldsymbol{V}_{\mathsf{p}}^{-},t+1}\big)\Big]. \end{aligned} \tag{54e}$$

$r, N, M \to +\infty$ with $\frac{r}{N} \to \beta_u = \mathcal{O}(1)$ and $\frac{r}{M} \to \beta_v = \mathcal{O}(1)$ for some fixed ratios $\beta_u \leq 1$ and $\beta_v \leq 1$.

In approximate message passing practices, a state evolution ansatz is based on the following concentration of measure for the precision variables in the asymptotic regime:

$$\lim_{M,N\to\infty}\Big(\gamma_{\boldsymbol{U}_{\mathsf{p}}^{+},t}, \gamma_{\boldsymbol{U}_{\mathsf{e}}^{+},t}\Big) = \Big(\bar{\gamma}_{\boldsymbol{U}_{\mathsf{p}}^{+},t}, \bar{\gamma}_{\boldsymbol{U}_{\mathsf{e}}^{+},t}\Big), \tag{59a}$$

$$\lim_{M,N\to\infty}\Big(\gamma_{\boldsymbol{U}_{\mathsf{p}}^{-},t}, \gamma_{\boldsymbol{U}_{\mathsf{e}}^{-},t}\Big) = \Big(\bar{\gamma}_{\boldsymbol{U}_{\mathsf{p}}^{-},t}, \bar{\gamma}_{\boldsymbol{U}_{\mathsf{e}}^{-},t}\Big), \tag{59b}$$

$$\lim_{M,N\to\infty}\Big(\gamma_{\boldsymbol{V}_{\mathsf{p}}^{+},t}, \gamma_{\boldsymbol{V}_{\mathsf{e}}^{+},t}\Big) = \Big(\bar{\gamma}_{\boldsymbol{V}_{\mathsf{p}}^{+},t}, \bar{\gamma}_{\boldsymbol{V}_{\mathsf{e}}^{+},t}\Big), \tag{59c}$$

$$\lim_{M,N\to\infty}\Big(\gamma_{\boldsymbol{V}_{\mathsf{p}}^{-},t}, \gamma_{\boldsymbol{V}_{\mathsf{e}}^{-},t}\Big) = \Big(\bar{\gamma}_{\boldsymbol{V}_{\mathsf{p}}^{-},t}, \bar{\gamma}_{\boldsymbol{V}_{\mathsf{e}}^{-},t}\Big), \tag{59d}$$

$$\lim_{M,N\to\infty}\Big(\gamma_{\boldsymbol{Z}_{\mathsf{p}}^{+},t}, \gamma_{\boldsymbol{Z}_{\mathsf{e}}^{+},t}\Big) = \Big(\bar{\gamma}_{\boldsymbol{Z}_{\mathsf{p}}^{+},t}, \bar{\gamma}_{\boldsymbol{Z}_{\mathsf{e}}^{+},t}\Big), \tag{59e}$$

$$\lim_{M,N\to\infty}\Big(\gamma_{\boldsymbol{Z}_{\mathsf{p}}^{-},t}, \gamma_{\boldsymbol{Z}_{\mathsf{e}}^{-},t}\Big) = \Big(\bar{\gamma}_{\boldsymbol{Z}_{\mathsf{p}}^{-},t}, \bar{\gamma}_{\boldsymbol{Z}_{\mathsf{e}}^{-},t}\Big). \tag{59f}$$

A. State Evolution of Bi-VAMP

Recall here that Bi-VAMP applies to the bilinear observation model in which the data matrix, $\boldsymbol{Y}$, is obtained according to (2) while taking $\phi(.)$ to be the identity, i.e., $\phi(x) = x$, $\forall\, x \in \mathbb{R}$. That is to say:

$$\boldsymbol{Y} \;=\; \boldsymbol{U}\boldsymbol{V}^{\mathsf{T}} + \boldsymbol{W}. \tag{60}$$

in which $\boldsymbol{W} \in \mathbb{R}^{N\times M}$ is the additive white Gaussian noise matrix whose entries are assumed to be mutually independent with mean zero and variance $\sqrt{MN}\gamma_w^{-1}$, i.e., $w_{i,j} \sim \mathcal{N}(w_{ij}; 0, \sqrt{MN}\gamma_w^{-1})$. Note here that by letting M and N grow unboundedly one must scale the noise variance by $\sqrt{MN}$ to maintain the same signal-to-noise ratio irrespectively of the values of M and N. We emphasize the fact that this scaling is, however, required for the purpose of SE analysis only. For ease of notation, we will also drop the iteration index, t, and reintroduce it in the final state evolution recursion. Recall here that the algorithmic steps of Bi-VAMP are summarized in Algorithm 2 excluding the update equations pertaining to the "generalized output step" (i.e., lines 26 to 32) while replacing $\widehat{\boldsymbol{Z}}_{\mathsf{e}}^{+}$ by $\boldsymbol{Y}$ and $\gamma_{\boldsymbol{Z}_{\mathsf{e}}^{+}}$ by $\gamma_w/\sqrt{MN}$ (after taking into account the aforementioned appropriate scaling by $\sqrt{MN}$). Consequently, from the update equations in lines 5, 7, 8, and 10 of Algorithm 2, it follows that in the large system limit the component-wise MSEs of the bi-LMMSE denoisers are given by (61) and (62) displayed on the top of this page. Moreover, we assume that for large enough M and N the element-wise errors in the matrix updates, $\widehat{\boldsymbol{U}}_{\mathsf{p}}^{-}$ and $\widehat{\boldsymbol{V}}_{\mathsf{p}}^{-}$, are i.i.d. with zero mean thereby leading to:

$$\boldsymbol{R}_{\boldsymbol{U}_{\mathsf{p}}^{-}} \;\approx\; \bar{\gamma}_{\boldsymbol{U}_{\mathsf{p}}^{-}}^{-1}\boldsymbol{I}_r, \tag{63a}$$

$$\boldsymbol{R}_{\boldsymbol{V}_{\mathsf{p}}^{-}} \;\approx\; \bar{\gamma}_{\boldsymbol{V}_{\mathsf{p}}^{-}}^{-1}\boldsymbol{I}_r, \tag{63b}$$

For large enough M and N, we also make use of the following approximation:

$$\langle \boldsymbol{Y} \odot \boldsymbol{Y} \rangle \;\approx\; \bar{\sigma}_u^2\bar{\sigma}_v^2 r \,+\, \sqrt{MN}\gamma_w^{-1}, \tag{64}$$

which follows from the observation that $\sigma_u^2 \triangleq \langle \boldsymbol{U} \odot \boldsymbol{U} \rangle \approx \bar{\sigma}_u^2 \triangleq \mathbb{E}[\mathsf{u}_{i,\ell}^2 | p_{\mathsf{u}}(u)]$ and $\sigma_v^2 \triangleq \langle \boldsymbol{V} \odot \boldsymbol{V} \rangle \approx \bar{\sigma}_v^2 \triangleq \mathbb{E}[\mathsf{v}_{j\ell}^2 | p_{\mathsf{v}}(v)]$ $\forall i, j, \ell$. Here, $p_{\mathsf{u}}(u)$ [resp., $p_{\mathsf{v}}(v)$] is a common prior on the entries of the matrix $\mathbf{U}$ [resp., $\mathbf{V}$]. By the same virtue, we also approximate

$$\mathcal{E}_{u^-} = \lim_{r\to\infty}\frac{1}{r}\,\mathrm{Tr}\left(\boldsymbol{R}_{\boldsymbol{U}_{\mathsf{p}}^-}\right) = \lim_{M,N\to\infty}\frac{1}{r}\,\mathrm{Tr}\left(\left[\gamma_{\boldsymbol{U}_{\mathsf{e}}^+}\boldsymbol{I} + \frac{\gamma_w}{\sqrt{MN}}\left(\widehat{\boldsymbol{V}}_{\mathsf{p}}^{-\top}\widehat{\boldsymbol{V}}_{\mathsf{p}}^{-} + \frac{M}{\beta}\boldsymbol{R}_{\boldsymbol{V}_{\mathsf{p}}^-} - \frac{M\gamma_w}{\sqrt{MN}}\boldsymbol{R}_{\boldsymbol{V}_{\mathsf{p}}^-}\langle\boldsymbol{Y}\odot\boldsymbol{Y}\rangle\right)\right]^{-1}\right), \quad (61)$$

$$\mathcal{E}_{v^-} = \lim_{r\to\infty}\frac{1}{r}\,\mathrm{Tr}\left(\boldsymbol{R}_{\boldsymbol{V}_{\mathsf{p}}^-}\right) = \lim_{M,N\to\infty}\frac{1}{r}\,\mathrm{Tr}\left(\left[\gamma_{\boldsymbol{V}_{\mathsf{e}}^+}\boldsymbol{I} + \frac{\gamma_w}{\sqrt{MN}}\left(\widehat{\boldsymbol{U}}_{\mathsf{p}}^{-\top}\widehat{\boldsymbol{U}}_{\mathsf{p}}^{-} + \frac{N}{\beta}\boldsymbol{R}_{\boldsymbol{U}_{\mathsf{p}}^-} - \frac{N\gamma_w}{\sqrt{MN}}\boldsymbol{R}_{\boldsymbol{U}_{\mathsf{p}}^-}\langle\boldsymbol{Y}\odot\boldsymbol{Y}\rangle\right)\right]^{-1}\right). \quad (62)$$

$\gamma_{\boldsymbol{U}_{\mathsf{e}}^+}$ and $\gamma_{\boldsymbol{V}_{\mathsf{e}}^+}$ by $\bar{\gamma}_{\boldsymbol{U}_{\mathsf{e}}^+}$ and $\bar{\gamma}_{\boldsymbol{V}_{\mathsf{e}}^+}$, respectively. After using these approximations in (61) and (62), it follows that:

$$\mathcal{E}_{u^-}(\widetilde{\gamma}_{\boldsymbol{U}_{\mathsf{e}}^+}) = \lim_{r\to\infty}\frac{1}{r}\,\mathrm{Tr}\left(\left[\widetilde{\gamma}_{\boldsymbol{U}_{\mathsf{e}}^+}\boldsymbol{I}_r + \frac{\gamma_w}{\sqrt{MN}}\widehat{\boldsymbol{V}}_{\mathsf{p}}^{-\top}\widehat{\boldsymbol{V}}_{\mathsf{p}}^{-}\right)^{-1}\right), \quad (65)$$

$$\mathcal{E}_{v^-}(\widetilde{\gamma}_{\boldsymbol{V}_{\mathsf{e}}^+}) = \lim_{r\to\infty}\frac{1}{r}\,\mathrm{Tr}\left(\left[\widetilde{\gamma}_{\boldsymbol{V}_{\mathsf{e}}^+}\boldsymbol{I}_r + \frac{\gamma_w}{\sqrt{MN}}\widehat{\boldsymbol{U}}_{\mathsf{p}}^{-\top}\widehat{\boldsymbol{U}}_{\mathsf{p}}^{-}\right]^{-1}\right), \quad (66)$$

where

$$\begin{aligned}\widetilde{\gamma}_{\boldsymbol{U}_{\mathsf{e}}^+} &= \bar{\gamma}_{\boldsymbol{U}_{\mathsf{e}}^+} + \frac{1}{\beta}\sqrt{\frac{\beta_u}{\beta_v}}\gamma_w\bar{\gamma}^{-1}_{\boldsymbol{V}_{\mathsf{p}}^-} \\ &\quad - \sqrt{\frac{\beta_u}{\beta_v}}\gamma_w\bar{\gamma}^{-1}_{\boldsymbol{V}_{\mathsf{p}}^-}\left(\sqrt{\beta_u\beta_v}\bar{\sigma}_u^2\bar{\sigma}_v^2\gamma_w + 1\right), \quad (67)\\ \widetilde{\gamma}_{\boldsymbol{V}_{\mathsf{e}}^+} &= \bar{\gamma}_{\boldsymbol{V}_{\mathsf{e}}^+} + \frac{1}{\beta}\sqrt{\frac{\beta_v}{\beta_u}}\gamma_w\bar{\gamma}^{-1}_{\boldsymbol{U}_{\mathsf{p}}^-} \\ &\quad - \sqrt{\frac{\beta_v}{\beta_u}}\gamma_w\bar{\gamma}^{-1}_{\boldsymbol{U}_{\mathsf{p}}^-}\left(\sqrt{\beta_u\beta_v}\bar{\sigma}_u^2\bar{\sigma}_v^2\gamma_w + 1\right), \quad (68)\end{aligned}$$

To find the limits in (65) and (66), we define the following two matrices:

$$\boldsymbol{H}_u = \frac{1}{\sqrt{(\bar{\sigma}_u^2 - \bar{\gamma}^{-1}_{\boldsymbol{U}_{\mathsf{p}}^-})N}}\widehat{\boldsymbol{U}}_{\mathsf{p}}^{-}. \quad (69a)$$

$$\boldsymbol{H}_v = \frac{1}{\sqrt{(\bar{\sigma}_v^2 - \bar{\gamma}^{-1}_{\boldsymbol{V}_{\mathsf{p}}^-})M}}\widehat{\boldsymbol{V}}_{\mathsf{p}}^{-}. \quad (69b)$$

Under the matched conditions[6], the entries of $\boldsymbol{H}_u$ (resp., $\boldsymbol{H}_v$) are i.i.d. with zero mean and variance $\frac{1}{N}$ (resp., $\frac{1}{M}$). For ease of notation, we also define the two quantities:

$$\alpha_u \triangleq \sqrt{\frac{\beta_v}{\beta_u}}\frac{\gamma_w(\bar{\sigma}_u^2 - \bar{\gamma}^{-1}_{\boldsymbol{U}_{\mathsf{p}}^-})}{\widetilde{\gamma}_{\boldsymbol{V}_{\mathsf{e}}^+}}, \quad (70a)$$

$$\alpha_v \triangleq \sqrt{\frac{\beta_u}{\beta_v}}\frac{\gamma_w(\bar{\sigma}_v^2 - \bar{\gamma}^{-1}_{\boldsymbol{V}_{\mathsf{p}}^-})}{\widetilde{\gamma}_{\boldsymbol{U}_{\mathsf{e}}^+}}. \quad (70b)$$

Now, using a well-known result in random matrix theory (cf. eq. (1.16) in [25]) — we show that:

$$\begin{aligned}\mathcal{E}_{u^-}(\widetilde{\gamma}_{\boldsymbol{U}_{\mathsf{e}}^+}) &= \lim_{r\to\infty}\frac{1}{r}\,\mathrm{Tr}\left(\left[\widetilde{\gamma}_{\boldsymbol{U}_{\mathsf{e}}^+}\boldsymbol{I} + \frac{\gamma_w}{\sqrt{MN}}\widehat{\boldsymbol{V}}_{\mathsf{p}}^{-\top}\widehat{\boldsymbol{V}}_{\mathsf{p}}^{-}\right]^{-1}\right),\\ &= \lim_{r\to\infty}\frac{\widetilde{\gamma}^{-1}_{\boldsymbol{U}_{\mathsf{e}}^+}}{r}\,\mathrm{Tr}\left(\left[\boldsymbol{I} + \sqrt{\frac{\beta_u}{\beta_v}}\gamma_w\widetilde{\gamma}^{-1}_{\boldsymbol{U}_{\mathsf{e}}^+}(\bar{\sigma}_v^2 - \bar{\gamma}^{-1}_{\boldsymbol{V}_{\mathsf{p}}^-})\right.\right.\\ &\qquad\qquad \left.\left.\boldsymbol{H}_v^\top\boldsymbol{H}_v\right]^{-1}\right),\\ &= \widetilde{\gamma}^{-1}_{\boldsymbol{U}_{\mathsf{e}}^+}\left(1 - \frac{\mathcal{F}(\alpha_v,\beta_v)}{4\beta_v\alpha_v}\right),\end{aligned} \quad (71)$$

wherein the function $\mathcal{F}(.,.)$ is defined as:

$$\mathcal{F}(x,z) = \left(\sqrt{x(1+\sqrt{z})^2+1} - \sqrt{x(1-\sqrt{z})^2+1}\right)^2, \quad (72)$$

Similarly we show that:

$$\begin{aligned}\mathcal{E}_{v^-}(\widetilde{\gamma}_{\boldsymbol{V}_{\mathsf{e}}^+}) &= \lim_{r\to\infty}\frac{1}{r}\,\mathrm{Tr}\left(\left[\widetilde{\gamma}_{\boldsymbol{V}_{\mathsf{e}}^+}\boldsymbol{I} + \frac{\gamma_w}{\sqrt{MN}}\widehat{\boldsymbol{U}}_{\mathsf{p}}^{-\top}\widehat{\boldsymbol{U}}_{\mathsf{p}}^{-}\right]^{-1}\right),\\ &= \lim_{r\to\infty}\frac{\widetilde{\gamma}^{-1}_{\boldsymbol{V}_{\mathsf{e}}^+}}{r}\,\mathrm{Tr}\left(\left[\boldsymbol{I} + \sqrt{\frac{\beta_v}{\beta_u}}\gamma_w\widetilde{\gamma}^{-1}_{\boldsymbol{V}_{\mathsf{e}}^+}(\bar{\sigma}_u^2 - \gamma^{-1}_{\boldsymbol{U}_{\mathsf{p}}^-})\right.\right.\\ &\qquad\qquad \left.\left.\boldsymbol{H}_u^\top\boldsymbol{H}_u\right]^{-1}\right),\\ &= \widetilde{\gamma}^{-1}_{\boldsymbol{V}_{\mathsf{e}}^+}\left(1 - \frac{\mathcal{F}(\alpha_u,\beta_u)}{4\beta_u\alpha_u}\right).\end{aligned} \quad (73)$$

Recall here that as $r, N, M \to +\infty$ we have $\frac{r}{N} \to \beta_u$ and $\frac{r}{M} \to \beta_v$. Now, the output variances of the MMSE denoisers of $\boldsymbol{U}$ and $\boldsymbol{V}$ matrices are obtained from lines 20 and 21 of Algorithm 2 as follows:

$$\mathcal{E}_{u^+}(\gamma_{\boldsymbol{U}_{\mathsf{e}}^-}) \triangleq \frac{1}{\gamma_{\boldsymbol{U}_{\mathsf{p}}^+}} = \frac{1}{N\gamma_{\boldsymbol{U}_{\mathsf{e}}^-}}\sum_{i=1}^{N}\langle\mathbf{g}'_{\mathsf{u}}(\widehat{\boldsymbol{u}}_i^-, \gamma^{-1}_{\boldsymbol{U}_{\mathsf{e}}^-})\rangle, \quad (74)$$

$$\mathcal{E}_{v^+}(\gamma_{\boldsymbol{V}_{\mathsf{e}}^-}) \triangleq \frac{1}{\gamma_{\boldsymbol{V}_{\mathsf{p}}^+}} = \frac{1}{M\gamma_{\boldsymbol{V}_{\mathsf{e}}^-}}\sum_{j=1}^{M}\langle\mathbf{g}'_{\mathsf{v}}(\widehat{\boldsymbol{v}}_j^-, \gamma^{-1}_{\boldsymbol{U}_{\mathsf{e}}^-})\rangle. \quad (75)$$

In the large system limits, the empirical averages involved in (74) and (75) can be approximated by the following statistical averages in which we use the fact that $\gamma_{\boldsymbol{U}_{\mathsf{p}}^+} \longrightarrow \bar{\gamma}_{\boldsymbol{U}_{\mathsf{p}}^+}$ and $\gamma_{\boldsymbol{V}_{\mathsf{p}}^+} \longrightarrow \bar{\gamma}_{\boldsymbol{V}_{\mathsf{p}}^+}$ as M and N grow large:

$$\mathcal{E}_{u^+}(\bar{\gamma}_{\boldsymbol{U}_{\mathsf{e}}^-}) \triangleq \frac{1}{\bar{\gamma}_{\boldsymbol{U}_{\mathsf{e}}^-}}\mathbb{E}\left[g'_{\mathsf{u}}(\widehat{u}_{\mathsf{e}}^-, \bar{\gamma}^{-1}_{\boldsymbol{U}_{\mathsf{e}}^-})\middle| p_{\mathsf{u}}(u)p_{\widehat{\mathsf{u}}_{\mathsf{e}}^-|\mathsf{u}}(\widehat{u}_{\mathsf{e}}^-|u)\right] \quad (76)$$

[6] That is to say the true variance of the MMSE estimation error in the entries of $\widehat{\boldsymbol{V}}_{\mathsf{p}}^-$ (resp., $\widehat{\boldsymbol{U}}_{\mathsf{p}}^-$) is equal the one predicted by the algorithm, i.e., $\bar{\gamma}^{-1}_{\boldsymbol{V}_{\mathsf{p}}^-}$ (resp., $\bar{\gamma}^{-1}_{\boldsymbol{U}_{\mathsf{p}}^-}$). The matched condition assumption is common in previous works on state evolution analysis and it holds true if the algorithm at hand is optimum.

$$\mathcal{E}_{v^+}(\bar{\gamma}_{\boldsymbol{V}_{\mathsf{e}}^-}) \ \triangleq \ \frac{1}{\bar{\gamma}_{\boldsymbol{V}_{\mathsf{e}}^-}}\mathbb{E}\left[g_{\mathsf{v}}'(\widehat{v}_{\mathsf{e}}^-, \bar{\gamma}_{\boldsymbol{V}_{\mathsf{e}}^-}^{-1})\Big|p_{\mathsf{v}}(v)p_{\widehat{\mathsf{v}}_{\mathsf{e}}^-|\mathsf{v}}(\widehat{v}_{\mathsf{e}}^-|v)\right] \quad (77)$$

In (76) and (77), $p_{\widehat{\mathsf{u}}_{\mathsf{e}}^-|\mathsf{u}}(\widehat{u}_{\mathsf{e}}^-|u)$ and $p_{\widehat{\mathsf{v}}_{\mathsf{e}}^-|\mathsf{v}}(\widehat{v}_{\mathsf{e}}^-|v)$ correspond to the scalar models $\widehat{\mathsf{u}}_{\mathsf{e}}^- = \mathsf{u} + \mathsf{w}_u$ and $\widehat{\mathsf{v}}_{\mathsf{e}}^- = \mathsf{v} + \mathsf{w}_v$ where under the matched conditions, we have $\mathsf{w}_u \sim \mathcal{N}(w_u; 0, \bar{\gamma}_{\boldsymbol{U}_{\mathsf{e}}^-}^{-1})$ and $\mathsf{w}_v \sim \mathcal{N}(w_v; 0, \bar{\gamma}_{\boldsymbol{V}_{\mathsf{e}}^-}^{-1})$. Under all the aforementioned assumptions and matched conditions, we can now describe in Algorithm 3 our main result, which is the SE recursion equations for Bi-VAMP.

Algorithm 3 Bi-VAMP State Evolution

```
Require : Noise precision $\gamma_w$; set $\beta = 1$; $\beta_u$ and $\beta_v$; number of iterations $T_{\max}$.
Initialization : extrinsic precisions $\bar{\gamma}_{\boldsymbol{U}_{\mathsf{e}}^+,1}, \bar{\gamma}_{\boldsymbol{V}_{\mathsf{e}}^+,1}, \bar{\gamma}_{\boldsymbol{U}_{\mathsf{p}}^-,1}, \bar{\gamma}_{\boldsymbol{V}_{\mathsf{p}}^-,1}$
for $t = 1, \ldots, T_{\max}$ do
    ▷ compute the effective inverse noise variance for the Bi-LMMSE block
    $\widetilde{\gamma}_{\boldsymbol{U}_{\mathsf{e}}^+,t+1} = \bar{\gamma}_{\boldsymbol{U}_{\mathsf{e}}^+,t} - \sqrt{\frac{\beta_u}{\beta_v}}\gamma_w\bar{\gamma}_{\boldsymbol{V}_{\mathsf{p}}^-,t}^{-1}\left(\sqrt{\beta_u\beta_v}\bar{\sigma}_u^2\bar{\sigma}_v^2\gamma_w + 1\right) + \frac{1}{\beta}\sqrt{\frac{\beta_u}{\beta_v}}\gamma_w\bar{\gamma}_{\boldsymbol{V}_{\mathsf{p}}^-,t}^{-1}$
    $\widetilde{\gamma}_{\boldsymbol{V}_{\mathsf{e}}^+,t+1} = \bar{\gamma}_{\boldsymbol{V}_{\mathsf{e}}^+,t} - \sqrt{\frac{\beta_v}{\beta_u}}\gamma_w\bar{\gamma}_{\boldsymbol{U}_{\mathsf{p}}^-,t}^{-1}\left(\sqrt{\beta_u\beta_v}\bar{\sigma}_u^2\bar{\sigma}_v^2\gamma_w + 1\right) + \frac{1}{\beta}\sqrt{\frac{\beta_v}{\beta_u}}\gamma_w\bar{\gamma}_{\boldsymbol{U}_{\mathsf{p}}^-,t}^{-1}$
    ▷ compute the analytical posterior and extrinsic precision of $\mathbf{u}^-$
    $\bar{\gamma}_{\boldsymbol{U}_{\mathsf{p}}^-,t+1} = \frac{1}{\mathcal{E}_{u^-}\left(\widetilde{\gamma}_{\boldsymbol{U}_{\mathsf{e}}^+,t+1}\right)}$
    $\bar{\gamma}_{\boldsymbol{U}_{\mathsf{e}}^-,t+1} = \bar{\gamma}_{\boldsymbol{U}_{\mathsf{p}}^-,t+1} - \bar{\gamma}_{\boldsymbol{U}_{\mathsf{e}}^+,t}$
    ▷ compute the analytical posterior and extrinsic precision of $\mathbf{v}^-$
    $\bar{\gamma}_{\boldsymbol{V}_{\mathsf{p}}^-,t+1} = \frac{1}{\mathcal{E}_{v^-}\left(\widetilde{\gamma}_{\boldsymbol{V}_{\mathsf{e}}^+,t+1}\right)}$
    $\bar{\gamma}_{\boldsymbol{V}_{\mathsf{e}}^-,t+1} = \bar{\gamma}_{\boldsymbol{V}_{\mathsf{p}}^-,t+1} - \bar{\gamma}_{\boldsymbol{V}_{\mathsf{e}}^+,t}$
    ▷ compute the analytical posterior and extrinsic precision of $\mathbf{u}^+$
    $\bar{\gamma}_{\boldsymbol{U}_{\mathsf{p}}^+,t+1} = \frac{1}{\mathcal{E}_{u^+}(\bar{\gamma}_{\boldsymbol{U}_{\mathsf{e}}^-,t+1})}$
    $\bar{\gamma}_{\boldsymbol{U}_{\mathsf{e}}^+,t+1} = \bar{\gamma}_{\boldsymbol{U}_{\mathsf{p}}^+,t+1} - \bar{\gamma}_{\boldsymbol{U}_{\mathsf{e},t+1}^-}$
    ▷ compute the analytical posterior and extrinsic precision of $\mathbf{v}^+$
    $\bar{\gamma}_{\boldsymbol{V}_{\mathsf{p}}^+,t+1} = \frac{1}{\mathcal{E}_{v^+}(\bar{\gamma}_{\boldsymbol{V}_{\mathsf{e}}^-,t+1})}$
    $\bar{\gamma}_{\boldsymbol{V}_{\mathsf{e}}^+,t+1} = \bar{\gamma}_{\boldsymbol{V}_{\mathsf{p}}^+,t+1} - \bar{\gamma}_{\boldsymbol{V}_{\mathsf{e}}^-,t+1}$
end for
Return $\bar{\gamma}_{\boldsymbol{U}_{\mathsf{p}}^+,T_{\max}+1}, \bar{\gamma}_{\boldsymbol{V}_{\mathsf{p}}^+,T_{\max}+1}$
```

Note here that at convergence, one must have equality between the posterior variances of the same variable:

$$\mathcal{E}_{u^+}(\bar{\gamma}_{\boldsymbol{U}_{\mathsf{e}}^-,\infty}) \ \triangleq \ \bar{\gamma}_{\boldsymbol{U}_{\mathsf{p}}^+,\infty}^{-1} \ = \ \mathcal{E}_{u^-}(\widetilde{\gamma}_{\boldsymbol{U}_{\mathsf{e}}^+,\infty}) \ \triangleq \ \bar{\gamma}_{\boldsymbol{U}_{\mathsf{p}}^-,\infty}^{-1} \quad (78a)$$

$$\mathcal{E}_{v^+}(\bar{\gamma}_{\boldsymbol{V}_{\mathsf{e}}^-,\infty}) \ \triangleq \ \bar{\gamma}_{\boldsymbol{V}_{\mathsf{p}}^+,\infty}^{-1} \ = \ \mathcal{E}_{v^-}(\widetilde{\gamma}_{\boldsymbol{V}_{\mathsf{e}}^+,\infty}) \ \triangleq \ \bar{\gamma}_{\boldsymbol{V}_{\mathsf{p}}^-,\infty}^{-1} \quad (78b)$$

B. Extension to the Generalized Bilinear Model: State Evolution of BiG-VAMP

In this case, the observation matrix, $\boldsymbol{Y}$, is obtained from the generalized bilinear model in (1). To account for the residual error of the output denoiser (cf. Fig. 4), all the previous SE equations summarized in Algorithm 3 remain the same except for replacing the noise variance, γ_w^{-1}, by the extrinsic error variance, $\gamma_{\boldsymbol{Z}_{\mathsf{e}}^+}^{-1}$, whose SE update equation will be characterized in the following. Again, to maintain the same energy per each entry in the matrix $\mathbf{U}\mathbf{V}^{\mathsf{T}}$, as we grow M and N, we redefine $\mathbf{Z}$ as $\mathbf{Z} \triangleq \frac{1}{\sqrt[4]{MN}}\mathbf{U}\mathbf{V}^{\mathsf{T}}$.

Recall form line 27 in Algorithm 2 that the posterior variance, $\mathcal{E}_{z^-}(\gamma_{\boldsymbol{Z}_{\mathsf{e}}^+}) \triangleq \gamma_{\boldsymbol{Z}_{\mathsf{p}}^-}^{-1}$, of each z_{ij} estimate provided by the Bi-LMMSE module is given by (after taking into account the effect of the above scaling by $\frac{1}{\sqrt[4]{MN}}$):

$$\mathcal{E}_{z^-}(\gamma_{\boldsymbol{Z}_{\mathsf{e}}^+}) = \left(\gamma_{\boldsymbol{Z}_{\mathsf{e}}^+} + \sqrt{MN}\left[\mathrm{Tr}\left(\frac{1}{\beta}\boldsymbol{R}_{\boldsymbol{U}_{\mathsf{p}}^-}\boldsymbol{R}_{\boldsymbol{V}_{\mathsf{p}}^-}^{\top} + \frac{1}{M}\boldsymbol{R}_{\boldsymbol{U}_{\mathsf{p}}^-}\widehat{\boldsymbol{V}}_{\mathsf{p}}^{-\top}\widehat{\boldsymbol{V}}_{\mathsf{p}}^- + \frac{1}{N}\boldsymbol{R}_{\boldsymbol{V}_{\mathsf{p}}^-}\widehat{\boldsymbol{U}}_{\mathsf{p}}^{-\top}\widehat{\boldsymbol{U}}_{\mathsf{p}}^-\right)\right]^{-1}\right)^{-1}. \quad (79)$$

For large enough M and N, by plugging (63a), (63b), and (69) in (79), it follows that:

$$\mathcal{E}_{z^-}(\bar{\gamma}_{\boldsymbol{Z}_{\mathsf{e}}^+}) = \bar{\gamma}_{\boldsymbol{Z}_{\mathsf{p}}^-}^{-1} = \left(\bar{\gamma}_{\boldsymbol{Z}_{\mathsf{e}}^+} + \frac{1}{\sqrt{\beta_u\beta_v}}\left[\frac{1}{\beta}\,\bar{\gamma}_{\boldsymbol{U}_{\mathsf{p}}^-}^{-1}\bar{\gamma}_{\boldsymbol{V}_{\mathsf{p}}^-}^{-1} + \frac{(\bar{\sigma}_v^2 - \bar{\gamma}_{\boldsymbol{V}_{\mathsf{p}}^-}^{-1})}{\bar{\gamma}_{\boldsymbol{U}_{\mathsf{p}}^-}} + \frac{(\bar{\sigma}_u^2 - \bar{\gamma}_{\boldsymbol{U}_{\mathsf{p}}^-}^{-1})}{\bar{\gamma}_{\boldsymbol{V}_{\mathsf{p}}^-}}\right]^{-1}\right)^{-1}. \quad (80)$$

Now, the common component-wise variance of the output denoiser for $\boldsymbol{Z}$ is obtained from line 30 of Algorithm 2 as follows:

$$\mathcal{E}_{z^+}(\gamma_{\boldsymbol{Z}_{\mathsf{e}}^-}) \triangleq \frac{1}{\gamma_{\boldsymbol{Z}_{\mathsf{p}}^+}} = \frac{1}{MN\gamma_{\boldsymbol{Z}_{\mathsf{e}}^-}}\sum_{i=1}^{N}\sum_{j=1}^{M} g_{\mathsf{z}}'(y_{ij}, \widehat{z}_{ij,\mathsf{e}}^-, \gamma_{\boldsymbol{Z}_{\mathsf{e}}^-}^{-1}). \quad (81)$$

In the large system limits, the empirical average involved in (81) can be approximated by the following statistical average in which we use the fact that $\gamma_{\boldsymbol{Z}_{\mathsf{p}}^+} \longrightarrow \bar{\gamma}_{\boldsymbol{Z}_{\mathsf{p}}^+}$ as M and N grow large:

$$\mathcal{E}_{z^+}(\bar{\gamma}_{\boldsymbol{Z}_{\mathsf{e}}^-}) \triangleq \frac{1}{\bar{\gamma}_{\boldsymbol{Z}_{\mathsf{e}}^-}}\mathbb{E}\left[g_{\mathsf{z}}'(y, \widehat{z}_{\mathsf{e}}^-, \bar{\gamma}_{\boldsymbol{Z}_{\mathsf{e}}^-}^{-1})\Big|p_{\mathsf{z}}(z)p_{\mathsf{y}|\mathsf{z}}(y|z)p_{\widehat{\mathsf{z}}_{\mathsf{e}}^-|\mathsf{z}}(\widehat{z}_{\mathsf{e}}^-|z)\right]. \quad (82)$$

In (82), $p_{\widehat{\mathsf{z}}_{\mathsf{e}}^-|\mathsf{z}}(\widehat{z}_{\mathsf{e}}^-|z)$ corresponds to the scalar model $\widehat{\mathsf{z}}_{\mathsf{e}}^- = \mathsf{z} + \mathsf{w}_z$ where under the matched conditions, we have $\mathsf{w}_z \sim \mathcal{N}(w_z; 0, \bar{\gamma}_{\boldsymbol{Z}_{\mathsf{e}}^-}^{-1})$. Moreover, since we have $\mathsf{z} = \frac{1}{\sqrt[4]{MN}}\mathbf{u}^{\mathsf{T}}\mathbf{v}$, then owing to the central limit theorem we have $p_{\mathsf{z}}(z) = \mathcal{N}(z; 0, \sqrt{\beta_u\beta_v}\bar{\sigma}_u^2\bar{\sigma}_v^2)$. Under all the assumptions and matched conditions that we stated above, we can now describe in Algorithm 4 our main result, which is the SE recursion for BiG-VAMP. Note that at convergence, on top of the equalities in (78), we also have:

$$\mathcal{E}_{z^+}(\bar{\gamma}_{\boldsymbol{Z}_{\mathsf{e}}^-,\infty}) \ \triangleq \ \bar{\gamma}_{\boldsymbol{Z}_{\mathsf{p}}^+,\infty}^{-1} \ = \ \mathcal{E}_{z^-}(\widetilde{\gamma}_{\boldsymbol{Z}_{\mathsf{e}}^+,\infty}) \ \triangleq \ \bar{\gamma}_{\boldsymbol{Z}_{\mathsf{p}}^-,\infty}^{-1}. \quad (83)$$

V. Simulation Results

In this section, we assess the performance behavior of the proposed BiG-Vamp algorithm and benchmark it against BiG-AMP [21], BAd-VAMP [23] and LowRAMP [20] algorithms for different applications, namely:

- matrix factorization,
- dictionary learning,
- matrix completion.

Algorithm 4 BiG-VAMP State Evolution

```
**Require** : set $\beta = 1$; $\beta_u$ and $\beta_v$; number of iterations $T_{\text{max}}$.
**Initialization** : extrinsic precisions $\bar{\gamma}_{\boldsymbol{Z}_{\text{e}}^{+},1}, \bar{\gamma}_{\boldsymbol{U}_{\text{e}}^{+},1}, \bar{\gamma}_{\boldsymbol{V}_{\text{e}}^{+},1}, \bar{\gamma}_{\boldsymbol{U}_{\text{p}}^{-},1}, \bar{\gamma}_{\boldsymbol{V}_{\text{p}}^{-},1}$
for $t = 1, \ldots, T_{\text{max}}$ do
    ▷ compute the effective inverse noise variance for the Bi-LMMSE block
    $\widetilde{\gamma}_{\boldsymbol{U}_{\text{e}}^{+},t+1} = \bar{\gamma}_{\boldsymbol{U}_{\text{e}}^{+},t} + \frac{1}{\beta}\sqrt{\frac{\beta_u}{\beta_v}}\bar{\gamma}_{\boldsymbol{Z}_{\text{e}}^{+},t}\bar{\gamma}_{\boldsymbol{V}_{\text{p}}^{-},t}^{-1} - \sqrt{\frac{\beta_u}{\beta_v}}\bar{\gamma}_{\boldsymbol{Z}_{\text{e}}^{+},t}\bar{\gamma}_{\boldsymbol{V}_{\text{p}}^{-},t}^{-1}\left(\sqrt{\beta_u\beta_v}\bar{\sigma}_u^2\bar{\sigma}_v^2\bar{\gamma}_{\boldsymbol{Z}_{\text{e}}^{+},t} + 1\right)$
    $\widetilde{\gamma}_{\boldsymbol{V}_{\text{e}}^{+},t+1} = \bar{\gamma}_{\boldsymbol{V}_{\text{e}}^{+},t} + \frac{1}{\beta}\sqrt{\frac{\beta_v}{\beta_u}}\bar{\gamma}_{\boldsymbol{Z}_{\text{e}}^{+},t}\bar{\gamma}_{\boldsymbol{U}_{\text{p}}^{-},t}^{-1} - \sqrt{\frac{\beta_v}{\beta_u}}\bar{\gamma}_{\boldsymbol{Z}_{\text{e}}^{+},t}\bar{\gamma}_{\boldsymbol{U}_{\text{p}}^{-},t}^{-1}\left(\sqrt{\beta_u\beta_v}\bar{\sigma}_u^2\bar{\sigma}_v^2\bar{\gamma}_{\boldsymbol{Z}_{\text{e}}^{+},t} + 1\right)$
    ▷ compute the analytical posterior and extrinsic precision of $\mathbf{u}^-$
    $\bar{\gamma}_{\boldsymbol{U}_{\text{p}}^{-},t+1} = \frac{1}{\mathcal{E}_{u^-}\left(\widetilde{\gamma}_{\boldsymbol{U}_{\text{e}}^{+},t+1}\right)}$
    $\bar{\gamma}_{\boldsymbol{U}_{\text{e}}^{-},t+1} = \bar{\gamma}_{\boldsymbol{U}_{\text{p}}^{-},t+1} - \bar{\gamma}_{\boldsymbol{U}_{\text{e}}^{+},t}$
    ▷ compute the analytical posterior and extrinsic precision of $\mathbf{v}^-$
    $\bar{\gamma}_{\boldsymbol{V}_{\text{p}}^{-},t+1} = \frac{1}{\mathcal{E}_{v^-}\left(\widetilde{\gamma}_{\boldsymbol{V}_{\text{e}}^{+},t+1}\right)}$
    $\bar{\gamma}_{\boldsymbol{V}_{\text{e}}^{-},t+1} = \bar{\gamma}_{\boldsymbol{V}_{\text{p}}^{-},t+1} - \bar{\gamma}_{\boldsymbol{V}_{\text{e}}^{+},t}$
    ▷ compute the analytical posterior and extrinsic precision of $\mathbf{z}^-$
    $\bar{\gamma}_{\boldsymbol{Z}_{\text{p}}^{-},t+1} = \frac{1}{\mathcal{E}_{z^-}\left(\widetilde{\gamma}_{\boldsymbol{Z}_{\text{e}}^{+},t}\right)}$
    $\bar{\gamma}_{\boldsymbol{Z}_{\text{e}}^{-},t+1} = \bar{\gamma}_{\boldsymbol{Z}_{\text{p}}^{-},t+1} - \bar{\gamma}_{\boldsymbol{Z}_{\text{e}}^{+},t}$
    ▷ compute the analytical posterior and extrinsic precision of $\mathbf{u}^+$
    $\bar{\gamma}_{\boldsymbol{U}_{\text{p}}^{+},t+1} = \frac{1}{\mathcal{E}_{u^+}(\bar{\gamma}_{\boldsymbol{U}_{\text{e}}^{-},t+1})}$
    $\bar{\gamma}_{\boldsymbol{U}_{\text{e}}^{+},t+1} = \bar{\gamma}_{\boldsymbol{U}_{\text{p}}^{+},t+1} - \bar{\gamma}_{\boldsymbol{U}_{\text{e},t+1}^{-}}$
    ▷ compute the analytical posterior and extrinsic precision of $\mathbf{v}^+$
    $\bar{\gamma}_{\boldsymbol{V}_{\text{p}}^{+},t+1} = \frac{1}{\mathcal{E}_{v^+}(\bar{\gamma}_{\boldsymbol{V}_{\text{e}}^{-},t+1})}$
    $\bar{\gamma}_{\boldsymbol{V}_{\text{e}}^{+},t+1} = \bar{\gamma}_{\boldsymbol{V}_{\text{p}}^{+},t+1} - \bar{\gamma}_{\boldsymbol{V}_{\text{e}}^{-},t+1}$
    ▷ compute the analytical posterior and extrinsic precision of $\mathbf{z}^+$
    $\bar{\gamma}_{\boldsymbol{Z}_{\text{p}}^{+},t+1} = \frac{1}{\mathcal{E}_{z^+}(\bar{\gamma}_{\boldsymbol{Z}_{\text{e}}^{-},t+1})}$
    $\bar{\gamma}_{\boldsymbol{Z}_{\text{e}}^{+},t+1} = \bar{\gamma}_{\boldsymbol{Z}_{\text{p}}^{+},t+1} - \bar{\gamma}_{\boldsymbol{Z}_{\text{e}}^{-},t+1}$
end for
Return $\bar{\gamma}_{\boldsymbol{U}_{\text{p}}^{+},T_{\text{max}}+1}, \bar{\gamma}_{\boldsymbol{V}_{\text{p}}^{+},T_{\text{max}}+1}, \bar{\gamma}_{\boldsymbol{Z}_{\text{p}}^{+},T_{\text{max}}+1}$
```

In all simulations, we set $T_{\text{max}} = 1000$ and the precision tolerance to $\xi = 10^{-6}$ and we perform $N_{\text{MC}} = 100$ Monte-Carlo trials for different values of the SNR:

$$\text{SNR} = 10 \log_{10}\left(\frac{\|\boldsymbol{Z}\|_{\text{F}}^2}{\|\boldsymbol{W}\|_{\text{F}}^2}\right),$$

where $\|.\|_{\text{F}}$ is the Frobenius norm. We also use the normalized root MSE (NRMSE) as performance measure which defined as follows:

$$\text{NRMSE} = \frac{1}{N_{\text{MC}}}\sum_{\ell=1}^{N_{\text{MC}}}\frac{\|\boldsymbol{Z}_\ell - \widehat{\boldsymbol{Z}}_\ell\|_{\text{F}}}{\|\boldsymbol{Z}_\ell\|_{\text{F}}}$$

where $\boldsymbol{Z}_\ell$ is ℓth realization of $\mathbf{Z}$ and $\widehat{\boldsymbol{Z}}_\ell$ is its reconstruction during the ℓth Monte-Carlo trial. As per BiG-VAMP's initialization setting, all the initial means and covariances were set to the all-zero vector and the identity matrix, respectively. The results disclosed in the sequel demonstrate that BiG-VAMP yields considerable improvements in reconstruction performance and robustness as compared to state-of-the-art algorithms, especially in presence of discrete priors on either $\boldsymbol{U}$ or $\boldsymbol{V}$.

A. Noisy dictionary learning

Here, we apply the proposed BiG-VAMP algorithm to the well-known dictionary learning problem wherein the goal is to find, from a noisy observation $\boldsymbol{Y}$, a dictionary matrix $\boldsymbol{U}$ and a sparse matrix $\boldsymbol{V}$. For that purpose, we use Gaussian and Bernoulli-Gaussian priors on $\boldsymbol{U}$ and $\boldsymbol{V}$ matrices, respectively. We depict the NRMSE on the estimated $\widehat{\boldsymbol{Z}}$ in Fig. 6 wherin we bechchmark the proposed algorithm against BiG-AMP and BAd-VAMP. Note that BAd-VAMP was simulated with its defaults parameters from [23], i.e., $\tau_{1,\text{max}} = 1$ and $\tau_{2,\text{max}} = 0$, $\zeta = 0.8$ and $\gamma_{\text{min}} = 10^{-6}$. In Fig. 6, it is seen that BiG-AMP and BAd-VAMP[7] exhibit the same performance as BiG-VAMP which is hardly surprising since all algorithms are optimally exploiting the considered priors on $\boldsymbol{U}$ and $\boldsymbol{V}$ and do not suffer from any convergence issues.

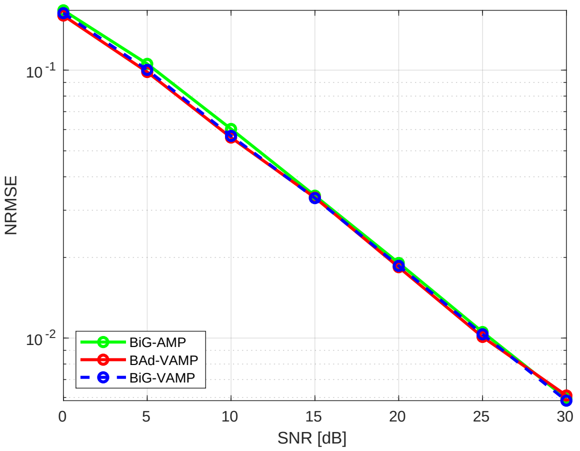


Fig. 6: NRMSE of BiG-VAMP and BAd-VAMP vs. the SNR for the dictionary learning problem: Gaussian prior on $\boldsymbol{U}$, Bernoulli-Gaussian prior on $\boldsymbol{V}$ with a sparsity level of 95%, $N = 1000$, $M = 1000$, and $r = 20$.

Next, we consider a binary prior on the unknown dictionary $\boldsymbol{U}$. Note that in this case the underlying *dictionary learning* problem is also known as the $\mathbb{Z}/2$ synchronization problem in the mathematical literature [28] or blind detection problem in the communication literature [29]. In this context, we again compare BiG-VAMP to BiG-AMP and we consider both low rank (i.e., $r = 5$) and moderately high rank (i.e., $r = 25$) structures as illustrated in Figs. 7 and 8, respectively.

We do not include BAd-VAMP in the subsequent simulations since it is not designed to handlenon-Gaussian priors on $\boldsymbol{U}$ This is in fact due to the inherent limitation imposed by the combinatorial maximization step in the EM algorithm that BAd-VAMP uses to update $\boldsymbol{U}$ at every iteration.

Figs. 7 and 8 show order of-magnitude difference in the NRMSE performance of BiG-VAMP and BiG-AMP. While BiG-VAMP outperforms BiG-AMP under both low-rank and high-rank structures, the gap between the two algorithms is not inherently related to the rank value, but rather to the inablity of BiG-AMP to converge under discrete priors. Moreover, it is seen from both figures that the empirical NRMSE of BiG-VAMP is

[7] Note here that the Gaussian prior on $\boldsymbol{U}$ was incorporated during the maximization step of the EM algorithm inside BAd-VAMP. In this special case, BAd-VAMP is actually performing MAP estimation of $\boldsymbol{U}$ (instead of maximum likelihood estimation).

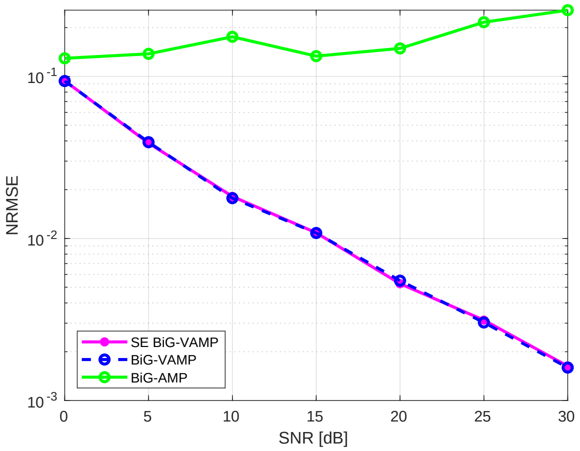


Fig. 7: NRMSE of BiG-VAMP and BiG-AMP vs. the SNR for the dictionary learning problem: binary prior on $\boldsymbol{U}$, Bernoulli-Gaussian prior on $\boldsymbol{V}$ with a sparsity level of 95%, $N = 100$, $M = 100$ and $r = 5$.

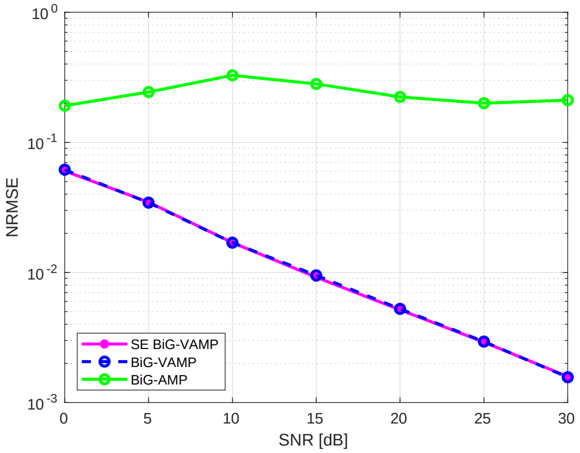


Fig. 8: NRMSE of BiG-VAMP and BiG-AMP vs. the SNR for the dictionary learning problem: binary prior on $\boldsymbol{U}$, Bernoulli-Gaussian prior on $\boldsymbol{V}$ with a sparsity level of 95%, $N = 500$, $M = 500$ and $r = 25$.

accurately predicted by the analytical state evolution recursion thereby corroborating the theoretical analysis we conducted in Section IV.

B. Matrix factorization

In this case, we compare BiG-VAMP to BIG-AMP [26] for the case where $\boldsymbol{U}$ is a binary matrix and $\boldsymbol{V}$ is Gaussian-distributed. Fig. 9 depicts the NRMSE of both algorithms and reveals that BiG-AMP's performance again deteriorates considerably in presence of a discrete prior either on $\boldsymbol{U}$ or $\boldsymbol{V}$ (due to the symmetry of the problem). BiG-VAMP, however, finds an accurate factorisation of $\boldsymbol{Z}$ over the entire SNR range and its empirical NRMSE is again theoretically predicted by the established state evolution analysis. This endows the proposed algorithm with offline design guidelines when applied to different engineering problems in practice.

A close inspection of the NRMSE time evolution, as shown in Fig. 10, illustrates qualitatively a transiently chaotic trajectory that is similar to those of fluid parcels in turbulent flows studied in [**?**].Such a chaotic behaviour may be a generic feature of algorithms searching for solutions in hard optimization problems with applications as diverse as protein folding and Sudoku [**?**].

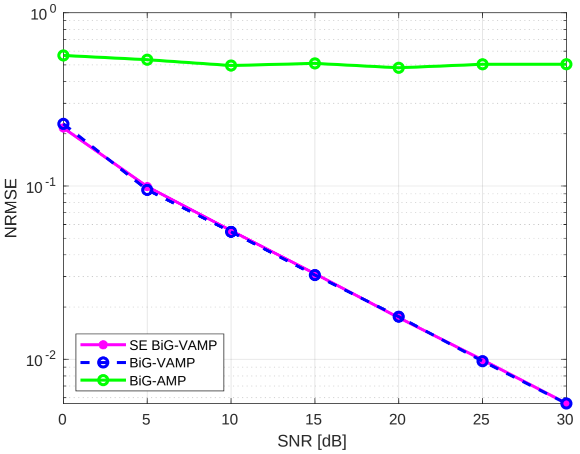


Fig. 9: NRMSE of BiG-VAMP and BiG-AMP vs. the SNR for the matrix factorization problem: binary prior on $\boldsymbol{U}$, Gaussian prior on $\boldsymbol{V}$, $N = 1000$, $M = 200$ and $r = 30$.

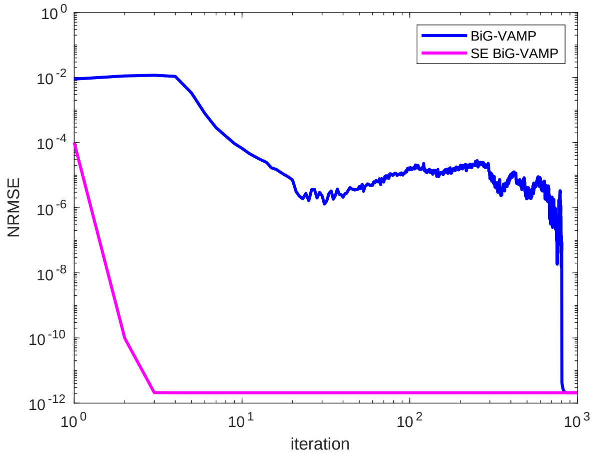


Fig. 10: NRMSE of BiG-VAMP and its SE for the matrix factorization problem at SNR = 10 dB: binary prior on $\boldsymbol{U}$, Gaussian prior on $\boldsymbol{V}$, $N = 1000$, $M = 500$ and $r = 10$.

C. Matrix completion

Unlike the previous two applications (i.e., matrix factorization and dictionary learning) wherein the observation model was linear, we now turn our attention to a popular generalized bilnear recovery problem, namely the matrix completion problem. In this context, Fig. 11 compares BiG-VAMP to BIG-AMP under the nonlinear observation model in (2) by taking $\phi(\cdot)$ to be a random selection with a rate of 20%. There, it is also seen that BiG-VAMP outperforms by far BiG-AMP which is in principle able to deal with any non-linearity in the observation model. But its deficiency here stems from the fact that it diverges for the considered challenging problem.

For the case where the only problem structure available at hand is the low rank approximation, we also benchmark BiG-VAMP against LowRAMP [20] whose MATLAB code is publicly available from [30]. The results are depicted in Fig. 12 and as

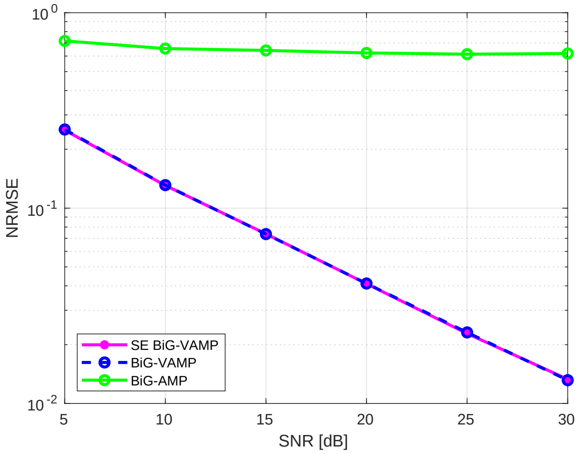


Fig. 11: NRMSE of BiG-VAMP and BiG-AMP vs. the SNR for the matrix completion problem with a selection rate of 20%: binary prior on $\boldsymbol{U}$, Gaussian prior on $\boldsymbol{V}$, $N = 1000$, $M = 500$ and $r = 30$.

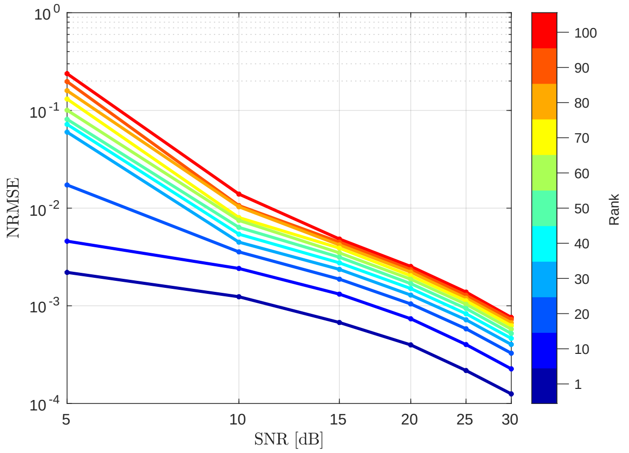


Fig. 13: Phase transition diagram BiG-VAMP for the matrix completion problem with $N = 1000$, $M = 500$, $1 \leq r \leq 100$, a selection rate of 20%, binary matrix $\mathbf{U}$ and Gaussian matrix $\boldsymbol{V}$.

seen there LowRAMP is not able to correctly recover $\boldsymbol{U}$ and $\boldsymbol{V}$ while BiG-VAMP exhibits the same performance as BiG-VAMP. Under the considered non linear selection function, $\phi(\cdot)$, the performance degradation of LowRAMP is mainly due to the fact the second-order Taylor series approximation of the output channel (4), around $z_{ij} = 0$, is not accurate high SNR values (which is a crucial condition to solve the matrix completion problem).

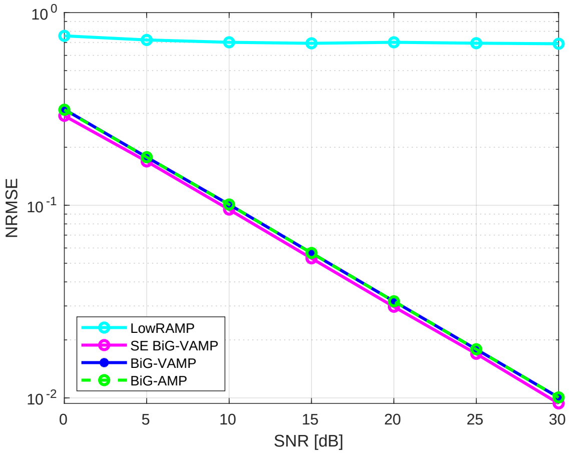


Fig. 12: NRMSE of BiG-VAMP, BiG-AMP, and LowRAMP vs. the SNR for the matrix completion with the selection rate of 10%: Gaussian prior on $\boldsymbol{U}$, Gaussian prior on $\mathbf{V}$, $N = 1000$, $M = 500$ and $r = 3$.

Fig. 13 shows the phase transition diagrams in the NRMSE for both low- and high-SNR regimes at a fixed selection rate of 20%. In the large-rank limit, we observe a low detectability regime when the SNR is in $[0, 10]$ dB. This result is consistent with the non negligible uncertainty of estimators [31] to conduct inference in the low-SNR regime for noisy matrix completion.

VI. Conclusion

In this work, we introduced a new algorithm, dubbed BiG-VAMP, to solve the generalized bilinear recovery problem based on the approximate message passing paradigm while treating both the MMSE and MAP inference problems in a unified framework. We described how BiG-VAMP provides a broader solution to the bilinear recovery problem under different structured matrices beyond the "low rank" structure heavily investigated in the existing literature. In particular, our numerical results for applications in matrix-factorization, dictionary learning, and matrix completion demonstrated that BiG-VAMP exhibits the best reconstruction performance under non Gaussian priors as compared to existing state-of-the-art algorithms such as BiG-AMP, BAd-VAMP, and LowRAMP. Additionally, we derived the state evolution equations of BiG-VAMP and characterized its phrase transition for the matrix completion problem.

Acknowledgment

This work was supported by the Discovery Grants (DG) program from the Natural Science and Engineering Council of Canada (NSERC).

References

[1] R. Gribonval, R. Jenatton, and F. Bach, "Sparse and spurious: dictionary learning with noise and outliers," *IEEE Transactions on Information Theory*, vol. 61, no. 11, pp. 6298–6319, 2015.

[2] A. Montanari, Y. Eldar, and G. Kutyniok, "Graphical models concepts in compressed sensing," *Compressed Sensing: Theory and Applications*, pp. 394–438, 2012.

[3] H. Zou, T. Hastie, and R. Tibshirani, "Sparse principal component analysis," *Journal of computational and graphical statistics*, vol. 15, no. 2, pp. 265–286, 2006.

[4] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, no. 8, pp. 30–37, 2009.

[5] R. Matsushita and T. Tanaka, "Low-rank matrix reconstruction and clustering via approximate message passing," in *Advances in Neural Information Processing Systems*, 2013, pp. 917–925.

[6] L. Wang, Z. Zhang, and D. Dunson, "Symmetric bilinear regression for signal subgraph estimation," *IEEE Transactions on Signal Processing*, vol. 67, no. 7, pp. 1929–1940, 2019.

[7] Z. Lin, M. Chen, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," *arXiv preprint arXiv:1009.5055*, 2010.

[8] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.

[9] S. D. Babacan, M. Luessi, R. Molina, and A. K. Katsaggelos, "Sparse bayesian methods for low-rank matrix estimation," *IEEE Transactions on Signal Processing*, vol. 60, no. 8, pp. 3964–3977, 2012.

[10] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proceedings of the National Academy of Sciences*, vol. 106, no. 45, pp. 18 914–18 919, 2009.

[11] Y. C. Eldar and G. Kutyniok, *Compressed sensing: theory and applications*. Cambridge university press, 2012.

[12] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *2011 IEEE International Symposium on Information Theory Proceedings*, 2011, pp. 2168–2172.

[13] M. Bayati and A. 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.

[14] J. Barbier, F. Krzakala, N. Macris, L. Miolane, and L. Zdeborová, "Optimal errors and phase transitions in high-dimensional generalized linear models," *Proceedings of the National Academy of Sciences*, vol. 116, no. 12, pp. 5451–5460, 2019.

[15] M. Mezard and A. Montanari, *Information, physics, and computation*. Oxford University Press, 2009.

[16] S. Rangan, P. Schniter, and A. K. Fletcher, "Vector approximate message passing," in *2017 IEEE International Symposium on Information Theory (ISIT)*, 2017, pp. 1588–1592.

[17] P. Schniter, S. Rangan, and A. 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.

[18] S. Rangan, P. Schniter, and A. K. Fletcher, "Vector approximate message passing," *IEEE Transactions on Information Theory*, vol. 65, no. 10, pp. 6664–6684, 2019.

[19] J. Ma and L. Ping, "Orthogonal amp," *IEEE Access*, vol. 5, pp. 2020–2033, 2017.

[20] T. Lesieur, F. Krzakala, and L. Zdeborová, "Phase transitions in sparse pca," in *2015 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2015, pp. 1635–1639.

[21] J. T. Parker, P. Schniter, and V. Cevher, "Bilinear generalized approximate message passing—part i: Derivation," *IEEE Transactions on Signal Processing*, vol. 62, no. 22, pp. 5839–5853, 2014.

[22] J. T. Parker and P. Schniter, "Parametric bilinear generalized approximate message passing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 4, pp. 795–808, 2016.

[23] S. Sarkar, A. K. Fletcher, S. Rangan, and P. Schniter, "Bilinear recovery using adaptive vector-amp," *IEEE Transactions on Signal Processing*, vol. 67, no. 13, pp. 3383–3396, 2019.

[24] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, pp. 1–22, 1977.

[25] A. M. Tulino, S. Verdú *et al.*, "Random matrix theory and wireless communications," *Foundations and Trends® in Communications and Information Theory*, vol. 1, no. 1, pp. 1–182, 2004.

[26] P. Schniter, "BiGAMP." [Online]. Available: https://sourceforge.net/projects/gampmatlab/

[27] S. Sarkar, "Bilinear Adaptive Vector Approximate Message Passing," Jul. 2019. [Online]. Available: https://github.com/sbrsarkar/BAdVAMP

[28] A. Perry, A. Wein, A. Bandeira, and A. Moitra, "Message-Passing Algorithms for Synchronization Problems over Compact Group," *Communications on Pure and Applied Mathematics*, 10 2016.

[29] A. Mezghani and A. L. Swindlehurst, "Blind Estimation of Sparse Broadband Massive MIMO Channels With Ideal and One-bit ADCs," *IEEE Transactions on Signal Processing*, vol. 66, no. 11, pp. 2972–2983, 2018.

[30] F. Krzakala, "Low rank Matrix Factorization with AMP," Aug. 2015. [Online]. Available: https://github.com/krzakala/LowRAMP

[31] Y. Chen, J. Fan, C. Ma, and Y. Yan, "Inference and uncertainty quantification for noisy matrix completion," *Proceedings of the National Academy of Sciences*, vol. 116, no. 46, pp. 22 931–22 937, 2019.