Optimizing Approximate Message Passing for Variable Measurement Noise

Stefan C. Birgmeier and Norbert Goertz
TU Wien
Institute of Telecommunications
Email: {stefan.birgmeier, norbert.goertz}@tuwien.ac.at

***Abstract*—The standard Approximate Message Passing (AMP) algorithm optionally considers i.i.d. measurement noise. The governing parameter is the noise variance. When the noise is independent, but not identically distributed, applying AMP with the noise variance parameter set to the average of the actual noise variance results in significantly degraded performance. We propose a modified AMP algorithm called AMP-VN which improves performance for known noise variances.**

I. Introduction

Signal acquisition in compressed sensing (CS) is modeled as [1]

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{w}, \tag{1}$$

where $\boldsymbol{A}$ is the measurement matrix of size $L \times N$, $L \ll N$ with entries from $\mathcal{N}(0, L^{-1})$. The unknown vector $\boldsymbol{x}$ is distributed according to $f_{\mathsf{x}}(\boldsymbol{x})$. In the simplest case, this distribution factors into N identical, one-dimensional probability density functions $f_{\mathsf{x}}(x)$. Often, $\boldsymbol{x}$ is assumed to be sparse in some sense, e.g. with $f_{\mathsf{x}}(x)$ being "Bernoulli-Gaussian":

$$f_{\mathsf{x}}(x) = (1-\gamma)\delta(x) + \gamma\mathcal{N}(0, \sigma_{\mathsf{x}}^2) \tag{2}$$

and $0 < \gamma \ll 1$. The aim is reconstruction of $\boldsymbol{x}$ using the measurements $\boldsymbol{y}$ and the sensing matrix $\boldsymbol{A}$. In this paper, the signal vector $\mathbf{x}$ is assumed to be i.i.d. with known probability density function. Furthermore, the noise variances $\sigma_{\mathsf{w},a}^2$ are known. The variant of the original AMP algorithm considered in this paper was introduced in [2] and is referred to as Algorithm 1. It assumes that the noise variance is independent and identical for all entries of $\boldsymbol{y}$. A thorough discussion of AMP's performance in this case can be found in [3].

Subsequently, the noise vector $\mathbf{w}$ is assumed to be distributed according to

$$\mathbf{w} \sim \prod_a \mathcal{N}(0, \sigma_{\mathsf{w},a}^2). \tag{3}$$

Furthermore, $|\mathcal{V}|$ different variances $\sigma_{\mathsf{w},a}^2$ are presumed to exist. The set $\mathcal{V}$ is a set of sets, i.e.

$$\mathcal{V} = \{\mathcal{W}_1, \mathcal{W}_2, \ldots, \mathcal{W}_K\}, \tag{4}$$

with a set $\mathcal{W}_k$ containing all random variables w_a with identical variance. The variance of variables in $\mathcal{W}_k$ shall be denoted $\sigma_{\mathsf{w},k}^2$. Equivalently, and slightly abusing notation, the set $\mathcal{W}_k$ contains all indices a for which the random variables w_a have identical variance $\sigma_{\mathsf{w},k}^2$.

II. Derivation

Subsequently, the derivation of Approximate Message Passing for Variable Noise Variance (AMP-VN) is presented. It starts with the Gaussian Message Passing formulation of AMP, which has its roots in the sum-product algorithm [4] applied to the complete factor-graph resulting from the dependencies between $\mathbf{y}$ and $\mathbf{x}$. A derivation of AMP's Gaussian message passing formulation can be found in [5, pp. 126] and shall not be repeated here. We follow the established convention of denoting messages from matrix factor nodes (indices a, b) to variables (indices i, j) as subscript $a \to i$ (and vice versa). Whenever sums or products over a, b or i, j occur, these cover the range $\{1, \ldots, L\}$ and $\{1, \ldots, N\}$ respectively. Similarly, terms v_a, v_b and v_i, v_j are entries of vectors in $\mathbb{R}^L$ and $\mathbb{R}^N$ respectively, while targeted messages $m_{i\to a}$ and $m_{a\to i}$ can be written as matrices in $\mathbb{R}^{N\times L}$ and $\mathbb{R}^{L\times N}$.

The functions $F(\boldsymbol{\mu}_{\mathsf{x}}^{(l)}, \sigma_{\mathsf{x}}^{2(l)})$ and $G(\boldsymbol{\mu}_{\mathsf{x}}^{(l)}, \sigma_{\mathsf{x}}^{2(l)})$ are applied component-wise and defined as [5]

$$F_i(\mu_{\mathsf{x}_i}^{(l)}, \sigma_{\mathsf{x}}^{2(l)}) = \frac{\int_{\mathsf{x}_i} x_i f_{\mathsf{x}_i}(x_i)\mathcal{N}_{\mathsf{x}_i}(\mu_{\mathsf{x},i}^{(l)}, \sigma_{\mathsf{x}}^{2(l)})\mathrm{d}x_i}{\int_{\mathsf{x}_i} f_{\mathsf{x}_i}(x_i)\mathcal{N}_{\mathsf{x}_i}(\mu_{\mathsf{x},i}^{(l)}, \sigma_{\mathsf{x}}^{2(l)})\mathrm{d}x_i}$$

$$G_i(\mu_{\mathsf{x}_i}^{(l)}, \sigma_{\mathsf{x}}^{2(l)}) = \frac{\int_{\mathsf{x}_i} x_i^2 f_{\mathsf{x}_i}(x_i)\mathcal{N}_{\mathsf{x}_i}(\mu_{\mathsf{x},i}^{(l)}, \sigma_{\mathsf{x}}^{2(l)})\mathrm{d}x_i}{\int_{\mathsf{x}_i} f_{\mathsf{x}_i}(x_i)\mathcal{N}_{\mathsf{x}_i}(\mu_{\mathsf{x},i}^{(l)}, \sigma_{\mathsf{x}}^{2(l)})\mathrm{d}x_i} - F_i(\ldots)^2.$$

Algorithm 1 AMP

$\boldsymbol{\mu}_{\mathsf{z}} \leftarrow \boldsymbol{y}$, $\sigma_{\mathsf{x}}^{2(l)} \leftarrow \|\boldsymbol{y}\|_2^2\, L^{-1}$
$\sigma_{\mathsf{w}}^2 \leftarrow L^{-1}\sum_a \sigma_{\mathsf{w},a}^2$
Set constants $t_{\max}$, ε, $\boldsymbol{A}$.
All other variables are initialized to zero.
repeat
  $\boldsymbol{\mu}_{\mathsf{x}}^{(l)} \leftarrow \boldsymbol{A}^T\boldsymbol{\mu}_{\mathsf{z}} + \boldsymbol{\mu}_{\mathsf{x}}$
  $\boldsymbol{\mu}_{\mathsf{x}}^{[t-1]} \leftarrow \boldsymbol{\mu}_{\mathsf{x}}$
  $\boldsymbol{\mu}_{\mathsf{x}} \leftarrow \boldsymbol{F}(\boldsymbol{\mu}_{\mathsf{x}}^{(l)}, \sigma_{\mathsf{x}}^{2(l)})$
  $\boldsymbol{\sigma}_{\mathsf{x}}^2 \leftarrow \boldsymbol{G}(\boldsymbol{\mu}_{\mathsf{x}}^{(l)}, \sigma_{\mathsf{x}}^{2(l)})$
  $v_a \leftarrow \frac{\mu_{\mathsf{z},a}}{L}\sum_i \frac{\partial F_i(\mu_{\mathsf{x},i}^{(l)}, \sigma_{\mathsf{x}}^{2(l)})}{\partial \mu_{\mathsf{x},i}^{(l)}}$
  $\boldsymbol{\mu}_{\mathsf{z}} \leftarrow \boldsymbol{y} - \boldsymbol{A}\boldsymbol{\mu}_{\mathsf{x}} + \boldsymbol{v}$
  $\sigma_{\mathsf{x}}^{2(l)} \leftarrow \sigma_{\mathsf{w}}^2 + \frac{1}{L}\sum_i \sigma_{\mathsf{x},i}^2$
until $t > t_{\max}$ **or** $t > 1$ **and** $\left\|\boldsymbol{\mu}_{\mathsf{x}} - \boldsymbol{\mu}_{\mathsf{x}}^{[t-1]}\right\|_2^2 < \varepsilon \left\|\boldsymbol{\mu}_{\mathsf{x}}\right\|_2^2$
$\hat{\boldsymbol{x}} = \boldsymbol{\mu}_{\mathsf{x}}$

ISBN 978-90-827970-1-5 © EURASIP 2018

We do not use the soft-thresholding heuristic in our simulations. A complete iteration is examined for the Gaussian messages' variances and subsequently for the expectations.

A. Message Passing for Variances

For the variances, an iteration according to the message passing rules (see [5]) is defined as

$$\sigma^{2(l)}_{\mathsf{x},i\to a} = \left(\sum_{b\neq a}\frac{1}{\sigma^2_{\mathsf{x},b\to i}}\right)^{-1} = \left(\sum_{b\neq a}\frac{A^2_{b,i}}{\sigma^2_{\mathsf{z},b\to i}+\sigma^2_{\mathsf{w},b}}\right)^{-1} \quad (5)$$

$$\sigma^2_{\mathsf{x},i\to a} = G(\mu^{(l)}_{\mathsf{x},i\to a},\sigma^{2(l)}_{\mathsf{x},i\to a}) \quad (6)$$

$$\sigma^2_{\mathsf{z},a\to i} = \sum_{j\neq i} A^2_{a,j}\sigma^2_{\mathsf{x},j\to a}. \quad (7)$$

Using the approximation (as common for AMP, see [5]) $A^2_{a,i}\approx L^{-1}$, it is possible to write

$$\sigma^{2(l)}_{\mathsf{x},i\to a} \approx L\left(\sum_{b\neq a}\frac{1}{\sigma^2_{\mathsf{z},b\to i}+\sigma^2_{\mathsf{w},b}}\right)^{-1} \quad (8)$$

$$\sigma^2_{\mathsf{z},a\to i} \approx \frac{1}{L}\sum_{j\neq i}\sigma^2_{\mathsf{x},j\to a}. \quad (9)$$

In a second iteration of approximations, assume that all $|\mathcal{W}_k|\gg 1$ and that all $\sigma^2_{\mathsf{z},a\to i}$ are approximately equal. The second assumption is justified if all $\sigma^2_{\mathsf{x},i\to a}$ are approximately equal $\forall a,i$. In this case,

$$\sigma^2_{\mathsf{z},a\to i} \approx \sigma^2_{\mathsf{z}} = \frac{N-1}{NL}\sum_i \sigma^2_{\mathsf{x},i} \quad (10)$$

$$\sigma^{2(l)}_{\mathsf{x},*\to a} = L\left(\sum_b \frac{1}{\sigma^2_{\mathsf{z}}+\sigma^2_{\mathsf{w},b}} - \frac{1}{\sigma^2_{\mathsf{z}}+\sigma^2_{\mathsf{w},a}}\right)^{-1} \quad (11)$$

$$\approx \sigma^{2(l)}_{\mathsf{x}} = L\left(\sum_a \frac{1}{\sigma^2_{\mathsf{z}}+\sigma^2_{\mathsf{w},a}}\right)^{-1}. \quad (12)$$

In the step from (11) to (12), $|\mathcal{W}_k|\gg 1$ is used. The larger $|\mathcal{W}_k|$, the more the $\sigma^2_{\mathsf{w},a}$ can differ in magnitude with the approximation still holding reasonably well. The omitted term is then small compared to the $|\mathcal{W}_k|-1$ identical terms remaining in the sum. Making use of these terms results in

$$\sigma^{2(l)}_{\mathsf{x}} = L\left(\sum_{\mathcal{W}_k\in\mathcal{V}}\sum_{b\in\mathcal{W}_k}\frac{1}{\sigma^2_{w,k}+\sigma^2_{\mathsf{z}}}\right)^{-1} \quad (13)$$

$$= L\left(\sum_{\mathcal{W}_k\in\mathcal{V}}\frac{|\mathcal{W}_k|}{\sigma^2_k}\right)^{-1}, \quad (14)$$

with $\sigma^2_k = \sigma^2_{\mathsf{w},k}+\sigma^2_{\mathsf{z}}$. Note that (14) only requires $|\mathcal{V}|+1$ divisions. Since it is required that $|\mathcal{W}_k|\gg 1$ and $\sum_k|\mathcal{W}_k| = L$, $|\mathcal{V}|$ needs to be small and thus the computational complexity is low.

B. Message Passing for Expectations

The message-passing iteration for the expectations is

$$\mu_{\mathsf{z},a\to i} = y_a - \sum_{j\neq i} A_{a,j}\mu_{\mathsf{x},j\to a} \quad (15)$$

$$\mu_{\mathsf{x},a\to i} = \frac{\mu_{\mathsf{z},a\to i}}{A_{a,i}} \approx L(A_{a,i}\mu_{\mathsf{z},a\to i}) \quad (16)$$

$$\mu^{(l)}_{\mathsf{x},i\to a} = \sigma^{2(l)}_{\mathsf{x},i\to a}\sum_{b\neq a}\frac{\mu_{\mathsf{x},b\to i}}{\sigma^2_{\mathsf{x},b\to i}} \quad (17)$$

$$\approx \sigma^{2(l)}_{\mathsf{x},i\to a}\sum_{b\neq a}\frac{L(A_{a,i}\mu_{\mathsf{z},a\to i})}{L(\sigma^2_{\mathsf{z}}+\sigma^2_{\mathsf{w},a})}, \quad (18)$$

which again uses $A^2_{a,i}\approx L^{-1}$. Writing the sum in (18) in terms of $\mathcal{W}_k$ results in

$$\mu^{(l)}_{\mathsf{x},i\to a} \approx \frac{\sigma^{2(l)}_{\mathsf{x}}}{L}\left(\sum_{\mathcal{W}_k\in\mathcal{V}}\sum_{b\in\mathcal{W}_k}\frac{\mu_{\mathsf{x},b\to i}}{\sigma^2_{\mathsf{w},b}+\sigma^2_{\mathsf{z}}} - \frac{\mu_{\mathsf{x},b\to i}}{\sigma^2_{\mathsf{w},a}+\sigma^2_{\mathsf{z}}}\right) \quad (19)$$

$$= \frac{\sigma^{2(l)}_{\mathsf{x}}}{L}\left(\sum_{\mathcal{W}_k\in\mathcal{V}}\frac{1}{\sigma^2_k}\sum_{b\in\mathcal{W}_k}\mu_{\mathsf{x},b\to i} - \frac{\mu_{\mathsf{x},a\to i}}{\sigma^2_{\mathsf{w},a}+\sigma^2_{\mathsf{z}}}\right) \quad (20)$$

$$\approx \sigma^{2(l)}_{\mathsf{x}}\left(\sum_{\mathcal{W}_k\in\mathcal{V}}\frac{1}{\sigma^2_k}\sum_{b\in\mathcal{W}_k}A_{b,i}\mu_{\mathsf{z},b\to i} - \frac{A_{a,i}\mu_{\mathsf{z},a\to i}}{\sigma^2_{\mathsf{w},a}+\sigma^2_{\mathsf{z}}}\right). \quad (21)$$

C. From Message Passing to AMP

While the variances have already lost their "targeted" message-passing character (cf. (10), (12)), expectations are still updated using message-passing rules, requiring computation of LN values in each iteration. Similarly to [5, p.108], the message (21) is dissected:

$$\mu^{(l)}_{\mathsf{x},i\to a} = \mu^{(l)}_{\mathsf{x},i} + \delta^{\mu(l)}_{\mathsf{x},i\to a} + O(N^{-1}), \quad (22)$$

using $\mu_{\mathsf{z},a\to i} = \mu_{\mathsf{z},a} + \delta^{\mu}_{\mathsf{z},a\to i} + O(N^{-1})$:

$$\begin{aligned}\mu^{(l)}_{\mathsf{x},i\to a} &= \sum_{\mathcal{W}_k\in\mathcal{V}}\frac{\sigma^{2(l)}_{\mathsf{x}}}{\sigma^2_k}\sum_{b\in\mathcal{W}_k}A_{b,i}(\mu_{\mathsf{z},b}+\delta^{\mu}_{\mathsf{z},b\to i}+O(N^{-1}))\\ &\quad - \frac{\sigma^{2(l)}_{\mathsf{x}}}{\sigma^2_k}A_{a,i}(\mu_{\mathsf{z},a}+\delta^{\mu}_{\mathsf{z},a\to i}+O(N^{-1}))\\ &= \underbrace{\sum_{\mathcal{W}_k\in\mathcal{V}}\frac{\sigma^{2(l)}_{\mathsf{x}}}{\sigma^2_k}\sum_{b\in\mathcal{W}_k}A_{b,i}(\mu_{\mathsf{z},b}+\delta^{\mu}_{\mathsf{z},b\to i})}_{\mu^{(l)}_{\mathsf{x},i}} \quad (23)\\ &\quad \underbrace{-\frac{\sigma^{2(l)}_{\mathsf{x}}}{\sigma^2_k}A_{a,i}\mu_{\mathsf{z},a}}_{\delta^{\mu(l)}_{\mathsf{x},i\to a}} + \underbrace{\dots}_{O(N^{-1})}. \quad (24)\end{aligned}$$

Terms of size $O(N^{-1})$ are neglected, which results in *Approximate* Message Passing. Furthermore, since $\delta^{\mu}_{\mathsf{z},a\to i} = A_{a,i}\mu_{\mathsf{x},i}$, (23) can be reformulated as

$$\mu^{(l)}_{\mathsf{x},i} = \sum_{\mathcal{W}_k\in\mathcal{V}}\frac{\sigma^{2(l)}_{\mathsf{x}}}{\sigma^2_k}\sum_{b\in\mathcal{W}_k}A_{b,i}\mu_{\mathsf{z},b} \quad (25)$$

ISBN 978-90-827970-1-5 © EURASIP 2018

$$+ \sum_{\mathcal{W}_k \in \mathcal{V}} \frac{\sigma_{\mathsf{x}}^{2(l)}}{\sigma_k^2} \sum_{b \in \mathcal{W}_k} \underbrace{A_{b,i}^2}_{\approx L^{-1}} \mu_{\mathsf{x},i} \tag{26}$$

$$= \sum_{\mathcal{W}_k \in \mathcal{V}} \frac{\sigma_{\mathsf{x}}^{2(l)}}{\sigma_k^2} \sum_{b \in \mathcal{W}_k} A_{b,i} \mu_{\mathsf{z},b} \tag{27}$$

$$+ \mu_{\mathsf{x},i} \frac{\sigma_{\mathsf{x}}^{2(l)}}{L} \sum_{\mathcal{W}_k \in \mathcal{V}} \frac{|\mathcal{W}_k|}{\sigma_k^2}. \tag{28}$$

Developing (28) by taking into account (14), $\sigma_{\mathsf{x}}^{2(l)}$ cancels down and thus

$$\mu_{\mathsf{x},i}^{(l)} = \sum_{\mathcal{W}_k \in \mathcal{V}} \frac{\sigma_{\mathsf{x}}^{2(l)}}{\sigma_k^2} \sum_{b \in \mathcal{W}_k} A_{b,i} \mu_{\mathsf{z},b} + \mu_{\mathsf{x},i} \tag{29}$$

$$\text{with } \mu'_{\mathsf{z},a} = \frac{\sigma_{\mathsf{x}}^{2(l)}}{\sigma_k^2} \mu_{\mathsf{z},a}, \quad a \in \mathcal{W}_k \tag{30}$$

$$\boldsymbol{\mu}_{\mathsf{x}}^{(l)} = \boldsymbol{A}^T \boldsymbol{z}' + \boldsymbol{\mu}_{\mathsf{x}}. \tag{31}$$

The expression (29) is similar to $\boldsymbol{\mu}_{\mathsf{x}}^{(l)} = \boldsymbol{A}^T \boldsymbol{z} + \boldsymbol{\mu}_{\mathsf{x}}$ in AMP and indeed identical for $|\mathcal{V}| = 1$. It remains to deploy the term $\delta_{\mathsf{x},i\to a}^{\mu(l)}$ from (24) in the expression for $\mu_{\mathsf{x},i\to a}$:

$$\mu_{\mathsf{x},i\to a} \approx F(\mu_{\mathsf{x},i}^{(l)}, \sigma_{\mathsf{x}}^{2(l)}) + \delta_{\mathsf{x},i\to a}^{\mu(l)} \frac{\partial F(\mu_{\mathsf{x},i}^{(l)}, \sigma_{\mathsf{x}}^{2(l)})}{\partial \mu_{\mathsf{x},i}^{(l)}} + O(N^{-1})$$

$$= \underbrace{F(\mu_{\mathsf{x},i}^{(l)}, \sigma_{\mathsf{x}}^{2(l)})}_{\mu_{\mathsf{x},i}} \tag{32}$$

$$\underbrace{-\mu'_{\mathsf{z},a} A_{a,i} \frac{\partial F(\mu_{\mathsf{x},i}^{(l)}, \sigma_{\mathsf{x}}^{2(l)})}{\partial \mu_{\mathsf{x},i}^{(l)}}}_{\delta_{\mathsf{x},i\to a}^{\mu}} + O(N^{-1}). \tag{33}$$

Finally, $\delta_{\mathsf{x},i\to a}^{\mu}$ appears in $\mu_{\mathsf{z},a}$:

$$\mu_{\mathsf{z},a} = y_a - \sum_j A_{a,j} (\mu_{\mathsf{x},j} + \delta_{\mathsf{x},j\to a}^{\mu}) \tag{34}$$

$$= y_a - \sum_j A_{a,j} \mu_{\mathsf{x},j} + \sum_j A_{a,j}^2 \mu'_{\mathsf{z},a} \frac{\partial F(\mu_{\mathsf{x},j}^{(l)}, \sigma_{\mathsf{x}}^{2(l)})}{\partial \mu_{\mathsf{x},j}^{(l)}}$$

$$= y_a - \sum_j A_{a,j} \mu_{\mathsf{x},j} + \frac{\mu'_{\mathsf{z},a}}{L} \sum_j \frac{\partial F(\mu_{\mathsf{x},j}^{(l)}, \sigma_{\mathsf{x}}^{2(l)})}{\partial \mu_{\mathsf{x},j}^{(l)}}. \tag{35}$$

III. State Evolution

The framework of State Evolution was first developed by Donoho, Maleki and Montanari in [6] and rigorously analyzed in [7]. Using the iterative nature of AMP, a simple recursive equation for the estimation of $\sigma_{\mathsf{x}}^{2(l)}$ and σ_{z}^2 is derived. Expressions for these variances can be identified in Algorithm 1:

$$\sigma_{\mathsf{z}}^2 = L^{-1} \sum_i \sigma_{\mathsf{x},i}^2 \tag{36}$$

$$\sigma_{\mathsf{x}}^{2(l)} = \sigma_{\mathsf{w}}^2 + \sigma_{\mathsf{z}}^2. \tag{37}$$

Furthermore, it can be shown [7] that the following representations are valid in the context of State Evolution:

$$\mu_{\mathsf{x}}^{(l)} \equiv x_0 + \sigma_{\mathsf{x}}^{(l)} \mathsf{u} \tag{38}$$

$$\sigma_{\mathsf{x},i}^2 \approx \sigma_{\mathsf{x}}^2 = \mathbb{E}\left\{ (F(x_0 + \sigma_{\mathsf{x}}^{(l)} \mathsf{u}) - x_0)^2 \right\}, \tag{39}$$

where $\mathsf{u} \sim \mathcal{N}(0,1)$ and x_0 is sampled from $f_{\mathsf{x}}(x)$. State Evolution for regular AMP can thus be written as

$$\sigma_{\mathsf{x},[t]}^{2(l)} = \sigma_{\mathsf{w}}^2 + \rho^{-1} \mathbb{E}\left\{ (F_{[t-1]}(x_0 + \sigma_{\mathsf{x},[t-1]}^{(l)} \mathsf{u}) - x_0)^2 \right\}, \tag{40}$$

where the index $[t]$ is used to identify the recursion iteration. In AMP-VN, the variance $\sigma_{\mathsf{x}}^{2(l)}$ is given by (14) contrary to AMP, where it is simply the sum of the average noise variance σ_{w}^2 and σ_{z}^2. Applying this change to State Evolution results in an iterative algorithm:

$$\sigma_{\mathsf{x},[t]}^{2(l)} = L \left(\sum_{\mathcal{W}_k \in \mathcal{V}} \frac{|\mathcal{W}_k|}{\sigma_{\mathsf{w},k}^2 + \sigma_{\mathsf{z},[t-1]}^2} \right)^{-1} \tag{41}$$

$$\sigma_{\mathsf{z},[t]}^2 = \rho^{-1} \mathbb{E}\left\{ \left(F(x_0 + \sigma_{\mathsf{x},[t]}^{(l)} \mathsf{u}) - x_0 \right)^2 \right\}. \tag{42}$$

For $|V| = 1$, i.e. identical noise variances, the adapted state evolution recursion (42), (41) reduces to (40).

IV. AMP-VN Algorithm and Results

The complete AMP-VN algorithm is shown in Algorithm 2. The transformation of $\mu_{\mathsf{z},a} \to \mu'_{\mathsf{z},a}$ can intuitively be explained as a weighting procedure where noisy entries are given less weight than noiseless samples. Note that the divisions involved in obtaining $\mu_{\mathsf{z},a}$ and $\sigma_{\mathsf{x}}^{2(l)}$ contain the same denominator. Compared to AMP, only $|\mathcal{V}| + 1$ additional divisions are necessary in each iteration; or $2|\mathcal{V}| + 1$ if one aims for higher numerical accuracy (which was done for our simulations).

The simulation results were obtained using an i.i.d. Bernoulli-Gaussian prior for x (cf. (2)) with $\gamma = 0.2$. For the average signal-to-noise ratio (SNR) of the measurements, the definition

$$\mathrm{SNR}_{\mathrm{dB}} = 10 \log_{10} \left(\frac{\|\boldsymbol{Ax}\|_2^2}{\|\boldsymbol{w}\|_2^2} \right) \tag{43}$$

Algorithm 2 AMP-VN

$\boldsymbol{\mu}_{\mathsf{z}} \leftarrow \boldsymbol{y}$, $\sigma_{\mathsf{z}}^2 \leftarrow \|\boldsymbol{y}\|_2^2 L^{-1}$
Set constants $\sigma_{\mathsf{w}_a}^2$, $t_{\max}$, ε, $\boldsymbol{A}$.
All other variables are initialized to zero.
repeat
 $\sigma_{\mathsf{x}}^{2(l)} \leftarrow L \left(\sum_{\mathcal{W}_k \in \mathcal{V}} \frac{|\mathcal{W}_k|}{\sigma_{\mathsf{w},k}^2 + \sigma_{\mathsf{z}}^2} \right)^{-1}$
 $\mu'_{\mathsf{z},a} \leftarrow \frac{\sigma_{\mathsf{x}}^{2(l)}}{\sigma_{\mathsf{w},a}^2 + \sigma_{\mathsf{z}}^2} \mu_{\mathsf{z},a}$
 $\boldsymbol{\mu}_{\mathsf{x}}^{(l)} \leftarrow \boldsymbol{A}^T \boldsymbol{\mu}'_{\mathsf{z}} + \boldsymbol{\mu}_{\mathsf{x}}$
 $\boldsymbol{\mu}_{\mathsf{x}}^{[t-1]} \leftarrow \boldsymbol{\mu}_{\mathsf{x}}$
 $\boldsymbol{\mu}_{\mathsf{x}} \leftarrow \boldsymbol{F}(\boldsymbol{\mu}_{\mathsf{x}}^{(l)}, \sigma_{\mathsf{x}}^{2(l)})$
 $\boldsymbol{\sigma}_{\mathsf{x}}^2 \leftarrow \boldsymbol{G}(\boldsymbol{\mu}_{\mathsf{x}}^{(l)}, \sigma_{\mathsf{x}}^{2(l)})$
 $v_a \leftarrow \frac{\mu'_{\mathsf{z},a}}{L} \sum_i \frac{\partial F_i(\mu_{\mathsf{x},i}^{(l)}, \sigma_{\mathsf{x}}^{2(l)})}{\partial \mu_{\mathsf{x},i}^{(l)}}$
 $\boldsymbol{\mu}_{\mathsf{z}} \leftarrow \boldsymbol{y} - \boldsymbol{A}\boldsymbol{\mu}_{\mathsf{x}} + \boldsymbol{v}$
 $\sigma_z^2 \leftarrow \frac{N-1}{NL} \sum_i \sigma_{\mathsf{x},i}^2$
until $t > t_{\max}$ **or** $t > 1$ **and** $\left\| \boldsymbol{\mu}_{\mathsf{x}} - \boldsymbol{\mu}_{\mathsf{x}}^{[t-1]} \right\|_2^2 < \varepsilon \|\boldsymbol{\mu}_{\mathsf{x}}\|_2^2$
$\hat{\boldsymbol{x}} = \boldsymbol{\mu}_{\mathsf{x}}$

was used. The undersampling factor is $\rho = \frac{L}{N}$, with the dimension $N = 1000$, abort threshold $\varepsilon = 10^{-4}$ and $t_{\max} = 100$ in all simulations. Several settings are compared. First, the performance of standard AMP for noisy $\boldsymbol{y}$ is examined, using an i.i.d. Gaussian prior for the noise $\boldsymbol{w}$. Secondly, standard AMP is explored for the case of independent, non-identically distributed noise. Finally, the behavior of AMP-VN is presented. Three different noise distributions are explored, in each of which one third of $\boldsymbol{y}$'s entries are more noisy than the others. The first two ("variable noise") use a noise variance which is 10 and 100 times larger than the remaining entries' variance. In the third case ("sparse noise"), every third entry is affected by noise while all other entries are noise-free.

These noise patterns appear in applications with burst-noise such as wireless communications channels, where AMP can be used for multi-user detection [8]. Another scheme employing AMP for decoding are Sparse Regression Codes (SPARCs) [9]. A SPARC codeword is the sum of several columns of a wide matrix $\boldsymbol{A}$, $\boldsymbol{c} = \boldsymbol{A}\boldsymbol{u}'$, where $\boldsymbol{u}'$ is a representation of the user data $\boldsymbol{u}$ (see [9] for details). A communications channel might apply different attenuations to the transmitted symbols c_a and add i.i.d. noise. This can be modeled as

$$\boldsymbol{c}' = \boldsymbol{H}\boldsymbol{A}\boldsymbol{u}' + \boldsymbol{w}, \tag{44}$$

where $\boldsymbol{H}$ is the (diagonal) matrix of channel coefficients and $\boldsymbol{w}$ is the noise. Equalization with $\boldsymbol{H}^{-1}$ leads to

$$\boldsymbol{c}' = \boldsymbol{H}^{-1}\boldsymbol{H}\boldsymbol{A}\boldsymbol{u}' + \boldsymbol{H}^{-1}\boldsymbol{w}. \tag{45}$$

The effective noise $\boldsymbol{H}^{-1}\boldsymbol{w}$ is then no longer i.i.d. Optimal decoding is possible with AMP-VN when $\boldsymbol{H}$ is known sufficiently well.

In this paper, the performance of AMP and AMP-VN is evaluated using the signal-to-distortion ratio (SDR):

$$\mathrm{SDR}_{\mathrm{dB}} = 10\log_{10}\left(\frac{\|\boldsymbol{x}\|_2^2}{\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2^2}\right). \tag{46}$$

Fig. 1 shows the phase transitions for AMP and AMP-VN. The behavior of AMP is examined for i.i.d. noise and sparse noise, showing that it is identical for both: it does not matter whether all samples of $\boldsymbol{y}$ are equally affected by noise or one third in particular. The similarity in performance is also visible in Fig. 4. The performance of AMP is visibly affected by the noise, resulting in a low recovery SDR even for regions where it converges.

The proposed AMP-VN algorithm makes use of the knowledge about individual noise variance. In Fig. 1, AMP-VN's phase transition curve can be seen to move towards lower SNR and lower subsampling factor ρ with increasing contrast between the $\sigma^2_{\mathrm{w},k}$. For "sparse noise" its performance does not deteriorate even in regions with low average SNR. For all three noise distributions, AMP-VN significantly outperforms regular AMP in terms of recovery SDR, as can be seen in Fig. 5. In case of low SNR but sparse noise, AMP-VN compares in performance to regular AMP with the noisy samples removed, as intuition suggests. This is shown in Fig. 2, where noiseless AMP is compared with AMP-VN at 0 dB SNR and sparse noise. One third of all entries of $\boldsymbol{y}$ are affected by noise. The phase transition for noiseless AMP happens at $\rho \approx 0.35$ while for noisy AMP-VN it occurs at $\rho \approx 0.55$. The scaled SDR curve takes into account that AMP-VN has 66.7% of AMP's noise-free samples at its disposal. Discarding noisy samples and using regular AMP for recovery is thus a valid strategy in the low-SNR regime. For higher SNR, not taking into account noisy samples incurs a performance penalty as shown in Fig. 2.

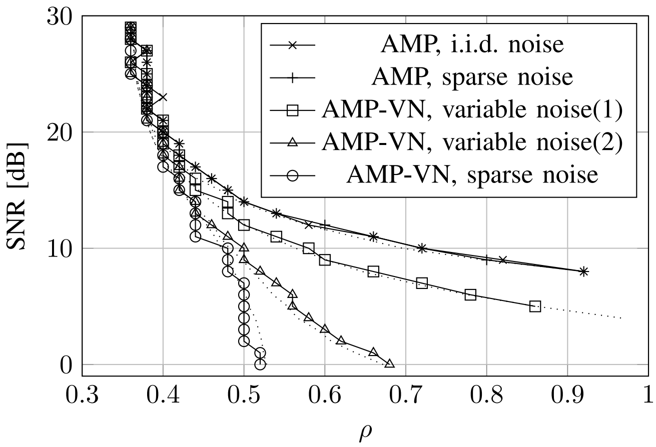

Fig. 1. Phase transition for Bernoulli-Gauss prior. The recovery SDR is larger than 10dB northeast of the curve. The dotted lines are the result of State Evolution estimation of the 10dB recovery SDR threshold of the associated curve. The SNR on the y-axis is the average measurement noise according to (43), ρ is the undersampling factor.

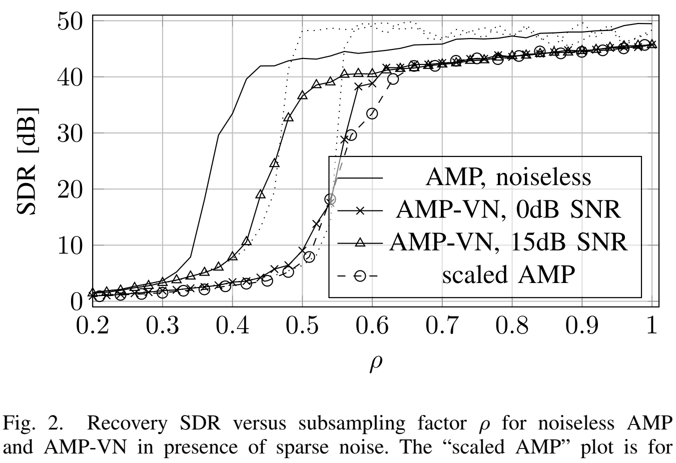

Fig. 2. Recovery SDR versus subsampling factor ρ for noiseless AMP and AMP-VN in presence of sparse noise. The "scaled AMP" plot is for comparison purposes only and explained above. The dotted curves are the result of State Evolution for AMP-VN at 15dB and 0dB SNR in presence of sparse noise. In the convergence region the predicted SDR deviates due to differing abortion criteria.

In Fig. 3, results from original and modified State Evolution iterations as defined in Section III can be seen. The modification to State Evolution does not change the behavior of the "denoiser" $F(\dots)$, which is the same for all cases. Behavior of original AMP is identical to the case of i.i.d. noise. The cutoff of AMP-VN is at lower values of σ_z^2 for non-i.i.d. noise. State

ISBN 978-90-827970-1-5 © EURASIP 2018

evolution also predicts the perfect recovery in case of sparse noise, which fits well with our other results (cf. Fig. 5).

V. Conclusions

The original AMP algorithm performs poorly for unevenly distributed measurement noise. An adapted algorithm is proposed which overcomes these limitations. A rigorous derivation as well as a State Evolution framework are provided. In the limiting case of sparse noise and low SNR, AMP-VN offers comparable performance to standard AMP after the removal of all noisy samples. For higher SNR, AMP-VN significantly outperforms AMP by using information owed to noisy samples.

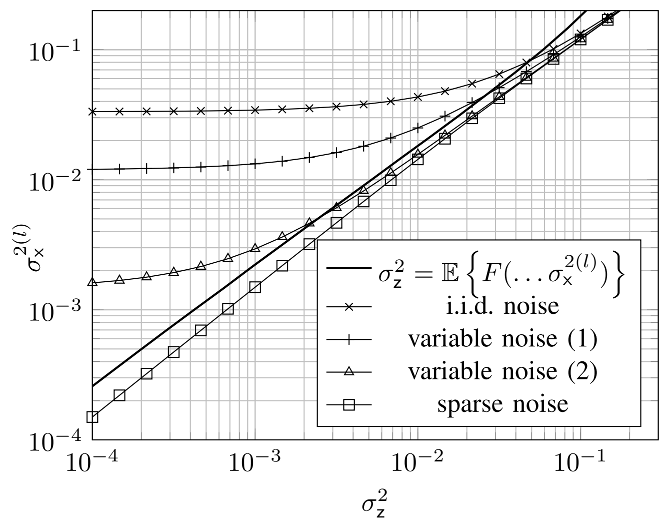

Fig. 3. "Exit-plot": the evolution of σ_z^2 and $\sigma_x^{2(l)}$, estimated with State Evolution for a subsampling factor $\rho = 0.6$ and SNR $= 10$dB.

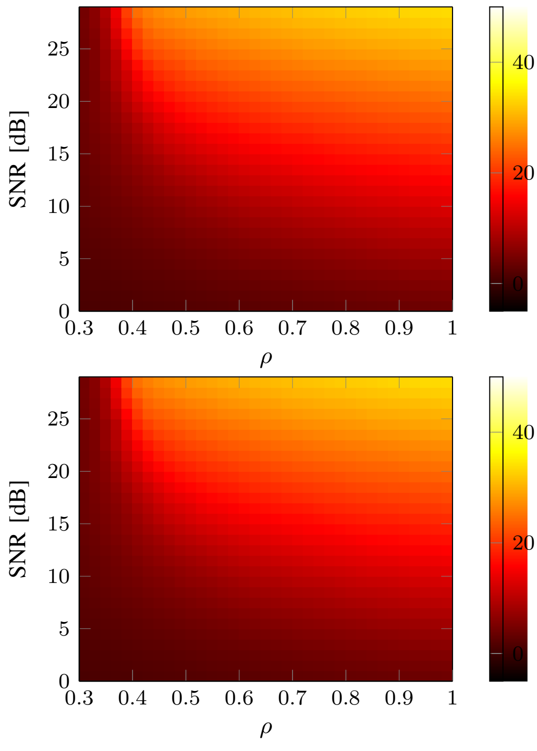

Fig. 4. Recovery SDR [dB] (color coded) vs. subsampling ratio ρ and noise SNR. Standard AMP with i.i.d. noise (top), with sparse noise (bottom).

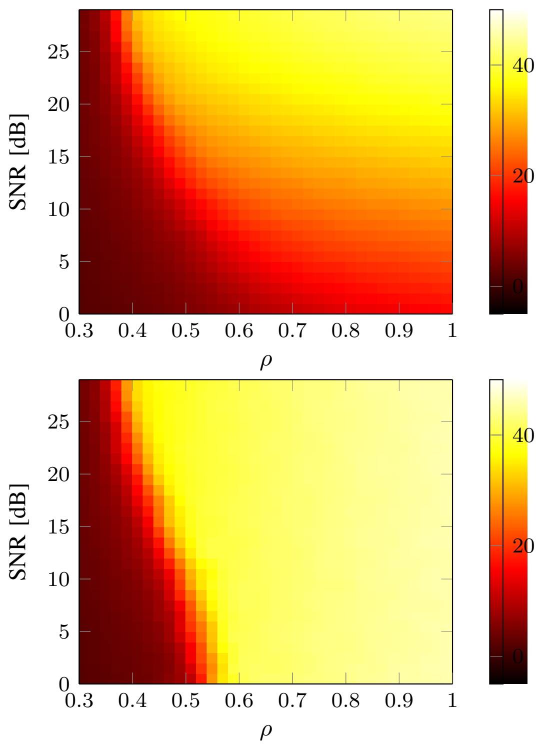

Fig. 5. Recovery SDR [dB] (color coded) vs. subsampling ratio ρ and noise SNR. AMP-VN with variable noise (variant 2, top) and sparse noise (bottom).

References

[1] Y. C. Eldar and G. Kutyniok, *Compressed sensing: theory and applications*. Cambridge University Press, 2012.

[2] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing: I. motivation and construction," in *2010 IEEE Information Theory Workshop on Information Theory (ITW 2010, Cairo)*, Jan 2010, pp. 1–5.

[3] ——, "The noise-sensitivity phase transition in compressed sensing," *IEEE Transactions on Information Theory*, vol. 57, no. 10, pp. 6920–6941, 2011.

[4] F. R. Kschischang, B. J. Frey, and H. A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 498–519, Feb 2001.

[5] A. Maleki, "Approximate message passing algorithms for compressed sensing," Ph.D. dissertation, Stanford University, September 2011.

[6] 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.

[7] 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, Feb 2011.

[8] S. Wu, L. Kuang, Z. Ni, J. Lu, D. Huang, and Q. Guo, "Low-complexity iterative detection for large-scale multiuser MIMO-OFDM systems using approximate message passing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 902–915, Oct 2014.

[9] C. Rush, A. Greig, and R. Venkataramanan, "Capacity-achieving sparse regression codes via approximate message passing decoding," in *2015 IEEE International Symposium on Information Theory (ISIT)*, June 2015, pp. 2016–2020.