ROBUST APPROXIMATE MESSAGE PASSING FOR NONZERO-MEAN SENSING MATRICES

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 efficiently recovers a sparse signal from a small number of noisy linear measurements. It requires the measurement matrix to be zero-mean, however. Even small deviations from this requirement cause it to diverge. In this paper, we show how mean-removal can be combined with standard Bayesian AMP to achieve signal recovery. Furthermore, a modified Bayesian AMP algorithm is presented, which achieves performance comparable to AMP in the zero-mean measurement matrix regime even for large mean. Simulation results and state evolution for both techniques are provided.

***Index Terms*—** Compressive Sensing, Approximate Message Passing.

1. INTRODUCTION

In compressive sensing (CS), a measurement signal $\boldsymbol{y} \in \mathbb{R}^L$ is acquired using linear measurements of a vector $\boldsymbol{x} \in \mathbb{R}^N$, with $L < N$. Often, Gaussian measurement noise $\boldsymbol{w}$ is considered:

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

The unknown signal $\boldsymbol{x}$ is assumed to be sparse and its entries independent and identically distributed (i.i.d.). The probability density function (pdf) of an entry x_i of x equals

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

where $\gamma \in (0 \dots 1)$ controls the sparsity. The noise $\boldsymbol{w}$ is also assumed i.i.d. and zero-mean in this paper. The entries of the measurement matrix $\boldsymbol{A}$ are often assumed to be i.i.d. and normal distributed with mean zero and variance L^{-1}. This construction allows for a variety of recovery algorithms like Iterative Soft Thresholding [1], Approximate Message Passing [2, 3, 4] and Generalized Approximate Message Passing (GAMP)[5].

It is not always possible to choose a measurement matrix satisfying these requirements, however. Popular applications of CS employ matrices with nonzero mean; one such application is the single pixel camera [6], where the sensing matrix is composed of ones and zeros. These indicate whether the light incident from a particular part of the image is accumulated in a measurement.

1.1. Relation to Prior Work

The problem of nonzero-mean measurement matrices has been investigated in the literature. In [7], the stability of standard AMP for nonzero-mean measurement matrices is analyzed using the Nishimori condition. Furthermore, a sequential, rather than parallel, update schedule is proposed for relaxed Belief Propagation (BP). This algorithm is computationally more expensive than AMP, which approximates BP. Our work focuses on solutions based on AMP.

In [8], mean removal techniques and an adaptive damping scheme are discussed in the context of GAMP, both for the Maximum a Posteriori (MAP) and the Minimum Mean Squared Error (MMSE) case. While their algorithm is still sensitive to the matrix mean it should be noted that it is also applicable to rank-deficient and column-correlated measurement matrices for the GAMP case, while this paper focuses on the case of nonzero-mean measurement matrices and classical MMSE Bayesian AMP. The new algorithms presented in this paper are much less sensitive to the matrix mean compared to original Bayesian AMP. A good characterization of (G)AMP with generic $\boldsymbol{A}$ can also be found in [9].

1.2. Motivation and Problem Statement

We consider the linear compressed sensing system

$$\boldsymbol{v} = \boldsymbol{B}\boldsymbol{x} + \boldsymbol{w} \,, \tag{3}$$

where $\boldsymbol{B} \in \mathbb{R}^{L\times N}$ with $L < N$ and the vectors $\boldsymbol{x}$ and $\boldsymbol{w}$ defined as before. It is assumed that the prior of x is known. The matrix $\boldsymbol{B}$ consists of i.i.d. entries for which

$$\mathbb{E}\left\{B_{a,i}\right\} = \mu_B \tag{4}$$

$$\mathbb{E}\left\{(B_{a,i} - \mathbb{E}\left\{B_{a,i}\right\})^2\right\} = L^{-1} \tag{5}$$

applies, where $B_{a,i}$ identifies the entry in the a^{th} row and i^{th} column of $\boldsymbol{B}$. Equivalently, the matrix $\boldsymbol{B}$ can be written as

$$\boldsymbol{B} = \mu_B \mathbf{1}_L \mathbf{1}_N^T + \boldsymbol{A} \,, \tag{6}$$

where $\mathbf{1}_M$ is the all-ones column vector of length M. Taking into account this decomposition, (3) can be written as

$$\boldsymbol{v} = (\mu_B \mathbf{1}_L \mathbf{1}_N^T + \boldsymbol{A})\boldsymbol{x} + \boldsymbol{w} \tag{7}$$

$$= \mu_B \mathbf{1}_L \mathbf{1}_N^T \boldsymbol{x} + \underbrace{\boldsymbol{A}\boldsymbol{x} + \boldsymbol{w}}_{\boldsymbol{y}} = \mu_B \mathbf{1}_L \mathbf{1}_N^T \boldsymbol{x} + \boldsymbol{y} \tag{8}$$

$$v_a = \sum_i \mu_B x_i + y_a = N\mu_B \bar{x} + y_a \,. \tag{9}$$

In (9), the value $\bar{x}$ is defined as the arithmetic mean of the entries x_i of $\boldsymbol{x}$. Note that even if $\mathbb{E}_{\mathsf{x}}\left\{\bar{\mathsf{x}}\right\} = 0$, its variance is

$$\sigma_{\bar{\mathsf{x}}}^2 = \mathbb{E}_{\mathsf{x}}\left\{\bar{\mathsf{x}}^2\right\} = \frac{1}{N^2}\sum_i\sum_j \mathbb{E}\left\{\mathsf{x}_i\mathsf{x}_j\right\} = \frac{1}{N}\sigma_{\mathsf{x}}^2 \,, \tag{10}$$

which is nonzero for finite N. Thus, for a particular measurement $\boldsymbol{v}$, the expression $N\mu_B\bar{x}$ has significant magnitude. The mean and variance of $\boldsymbol{y}$'s entries are

$$\mu_{\mathsf{y},a} = \mathbb{E}\left\{\mathsf{y}_a\right\} = \sum_i \mathbb{E}\left\{\mathsf{A}_{a,i}\mathsf{x}_i\right\} + \mathbb{E}\left\{w_a\right\} = 0 \tag{11}$$

$$\sigma^2_{\mathsf{y},a} = \mathbb{E}\Big\{\sum_{i,j} \mathsf{A}_{a,i}\mathsf{A}_{a,j}\mathsf{x}_i\mathsf{x}_j\Big\} + \mathbb{E}\left\{w_a^2\right\} \tag{12}$$

$$= \rho^{-1}\sigma^2_{\mathsf{x}} + \sigma^2_{\mathsf{w}}\,, \tag{13}$$

with $\rho = L/N$ the subsampling ratio. Since the mean $\mu_{\mathsf{y},a} = 0$, it is tempting to reformulate (3) as a CS system with zero-mean measurement matrix by setting

$$\boldsymbol{A} = \boldsymbol{B} - \mu_B \mathbf{1}_L \mathbf{1}_N^T \tag{14}$$

$$\boldsymbol{y} = \boldsymbol{v} - \bar{v}\mathbf{1}_L\,, \tag{15}$$

thus removing the effect of the term $N\mu_B\bar{x}$ in (9). It should be noted that for a particular instance of a CS problem, the arithmetic mean $\bar{y}$ of $\boldsymbol{y}$ is in general different from zero. Thus, performing mean-removal as per (15) introduces an error

$$\varepsilon = y_a - (v_a - \bar{v}) = -\mu_B \sum_i x_i + \bar{v} \tag{16}$$

$$= \frac{1}{L}\Big(\sum_i x_i \sum_a A_{a,i} + \sum_a w_a\Big) = \bar{y}\,. \tag{17}$$

In the noiseless case, the arithmetic mean of $\boldsymbol{y}$ is zero only if the columns of $\boldsymbol{A}$ have an arithmetic mean of zero. This can be used when the problem permits a custom design of the sensing matrix: choosing the matrix such that its columns have an arithmetic mean of zero makes it possible to compute $\boldsymbol{y}$ from $\boldsymbol{v}$ in the noiseless case. The same can be achieved in the presence of noise if the CS problem permits precise measurement of the arithmetic mean of $\boldsymbol{x}$. Then, the term $N\mu_B\bar{x}$ in (9) can be eliminated and $\boldsymbol{A}$ obtained by mean-removal as per (14). In all other cases, subtracting the arithmetic mean from $\boldsymbol{v}$ to obtain $\boldsymbol{y}$ results in an estimation error ε which is moreover correlated with $\boldsymbol{x}$.

2. STANDARD AMP WITH MEAN-REMOVAL

In this section, a simple way to stabilize AMP for the problem obtained by mean-removal as per (14), (15) shall be presented. The variance of the estimation error ε is

$$\sigma^2_\varepsilon = \frac{1}{L}\Big(\rho^{-1}\sigma^2_{\mathsf{x}} + \sigma^2_{\mathsf{w}}\Big) = \frac{1}{L}\sigma^2_{\mathsf{y},a}\,. \tag{18}$$

In the noiseless case, the average effective signal-to-noise ratio (SNR) is thus

$$\mathrm{SNR}^{\mathrm{eff}}_{\mathrm{dB}} = 10\log_{10}\left(\frac{\rho^{-1}\sigma^2_{\mathsf{x},i}}{(L\rho)^{-1}\sigma^2_{\mathsf{x},i}}\right) = 10\log_{10}(L)\,. \tag{19}$$

It is therefore possible to apply standard AMP to a modified noiseless CS problem, where the mean of $\boldsymbol{v}$ and $\boldsymbol{B}$ were removed, by setting the noise variance to $L^{-2}\left\|\boldsymbol{y}\right\|_2^2$. It shall be shown below that this stabilizes AMP for the modified problem and that AMP follows the convergence predicted by regular state evolution with the noise variance set accordingly. Achievable signal-to-distortion ratio (SDR) of the recovered signal is limited, especially for low-dimensional problems ($N \approx 10^3, \rho < 1 \implies \mathrm{SNR}^{\mathrm{eff}}_{\mathrm{dB}} < 30\mathrm{dB}$) and low-rate CS (e.g. $\rho = 10^{-3}, N = 10^5 \implies \mathrm{SNR}^{\mathrm{eff}}_{\mathrm{dB}} = 20\mathrm{dB}$). A suitably modified

Algorithm 1 MR-AMP

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

AMP algorithm is listed as Algorithm 1 and shall be referred to as MR-AMP (AMP after mean-removal). The function $F(\dots)$ is an appropriately chosen "denoiser" [3] for a particular prior of x while the function $G(\dots)$ computes the variance of F's estimate.

3. AMP WITH ITERATIVE ERROR CANCELLATION

As shown in (17), the estimation error ε depends on the true value of x. Furthermore, $\boldsymbol{y}$ can be computed without error when $\bar{x}$ is known, c.f. (9). For the remaining analysis, assume that $\bar{x} \neq 0$ and unknown, as well as $\sum_a A_{a,i} \neq 0$ and known μ_B. The dependence of $\boldsymbol{v}$ on $\boldsymbol{x}$ can be represented as a factor graph shown in Fig. 1. There are two new factors which are not present in a factor graph describing a classical CS problem, namely $f_{vy,a}$ and $f_{\bar{\mathsf{x}}}$. The factors $f_{vy,a}$ represent the dependencies between v_a, y_a and $\bar{\mathsf{x}}$ (and optionally the noise w_a). For zero-mean Gaussian noise with variance $\sigma^2_{\mathsf{w},a} > 0$ the factor reads

$$f_{vy,a}(\bar{x}, y_a) \propto \exp\left(-\frac{1}{2\sigma^2_{\mathsf{w},a}}\left(v_a - N\mu_B\bar{x} - y_a\right)^2\right). \tag{20}$$

The factor $f_{\bar{\mathsf{x}}}$ describes the relationship between $\bar{\mathsf{x}}$ and x, where $\bar{\mathsf{x}}$ is an auxiliary variable. It is possible to get rid of it by introducing dependencies between the factors $f_{vy,a}$ and x. Note that this results in adding $O(NL)$ edges, while with the auxiliary variable only $O(N+L)$ edges are necessary. The factor $f_{\bar{\mathsf{x}}}$ is given as

$$f_{\bar{\mathsf{x}}}(\bar{x}, \boldsymbol{x}) = \delta\Big(\bar{x} - \frac{1}{N}\sum_i x_i\Big)\,. \tag{21}$$

The message passing rules of the sum-product algorithm [10] (in AMP context [11]) can be used to derive expressions of messages along the newly introduced edges. Since there are four new types of edges, eight new messages can be defined. To simplify the resulting algorithm, only six messages are used. Following the message passing rules, one obtains

$$m_{vy\to\mathsf{y},a}(y_a) = \int_{\bar{\mathsf{x}}} f_{vy,a}(\bar{x}, y_a)\, m_{\bar{x}\to vy}(\bar{x})\,\mathrm{d}\bar{x} \tag{22}$$

$$m_{\mathsf{y},a\to vy}(y_a) = m_{f,a\to\mathsf{y},a}(y_a) \tag{23}$$

$$m_{vy\to\bar{\mathsf{x}}}(\bar{x}) = \int_{\bar{\mathsf{x}}} f_{vy,a}(\bar{x}, y_a)\, m_{\mathsf{y},a\to vy}(y_a)\,\mathrm{d}y_a \tag{24}$$

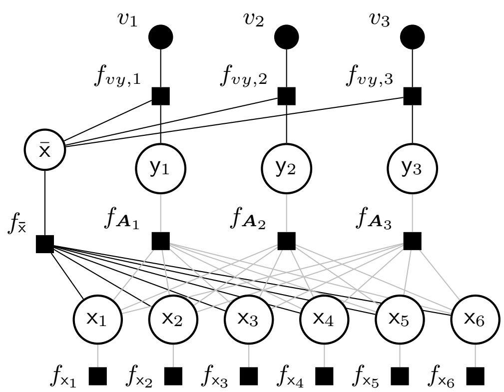

Fig. 1: Example of a graphical model describing a CS problem with a nonzero-mean sensing matrix $\boldsymbol{B}$. New edges have been emphasized. There are $N = 9$ unknown variables x_i and $L = 3$ known variables v_a.

$$m_{\bar{\mathsf{x}}\to vy,a}(\bar{x}) = m_{f,\bar{x}\to\bar{x}}(\bar{x}) \prod_{b\neq a} m_{vy,b\to\bar{\mathsf{x}}}(\bar{x}) \tag{25}$$

$$m_{f,\bar{x}\to\bar{\mathsf{x}}}(\bar{x}) = \int_{\mathbf{x}} f_{\bar{\mathsf{x}}}(\bar{x}, \boldsymbol{x}) \prod_i m_{\mathsf{x},i\to f,\bar{x}}(x_i) \mathrm{d}\boldsymbol{x} \tag{26}$$

$$m_{\mathsf{x},i\to f,\bar{x}}(x_i) = m_{f,\mathsf{x},i\to\mathsf{x}_i}(x_i) \prod_a m_{f,a\to\mathsf{x},i}(x_i)\,. \tag{27}$$

The algorithm can be initialized by starting with the last message, $m_{\mathsf{x},i\to f,\bar{x}}$, parameterized using the first two moments. During iterations, this is the tuple $(\mu_{\mathsf{x},i}, \sigma^2_{\mathsf{x},i})$ with $\mu_{\mathsf{x},i} = F(\dots)$ and $\sigma^2_{\mathsf{x},i} = G(\dots)$. For initialization, the prior mean and variance of x_i can be used. Subsequently, $m_{f,\bar{x}\to\bar{\mathsf{x}}}$ can be parameterized using mean and variance with

$$\mu_{f,\bar{\mathsf{x}}\to\bar{\mathsf{x}}} = N^{-1} \sum \mu_{\mathsf{x},i} \tag{28}$$

$$\sigma^2_{f,\bar{\mathsf{x}}\to\bar{\mathsf{x}}} = N^{-2} \sum \sigma^2_{\mathsf{x},i}\,. \tag{29}$$

During initialization, $m_{vy\to\bar{\mathsf{x}}}$ is unknown and thus, $m_{\bar{\mathsf{x}}\to vy,a}$ consists of the parameters $\mu_{f,\bar{\mathsf{x}}\to\bar{\mathsf{x}}}$ and $\sigma^2_{f,\bar{\mathsf{x}}\to\bar{\mathsf{x}}}$. Once $m_{vy,a\to\bar{\mathsf{x}}}$ is known, they can be mixed in according to (25), which is done below in (38), (40). With $m_{\bar{\mathsf{x}}\to vy,a}$ known, the message from $f_{vy,a}$ to y_a can be evaluated (c.f. (22)). The mean and variance of y_a thus compute as (c.f. (20))

$$\mu_{vy,a\to\mathsf{y},a} = v_a - N\mu_B \mu_{\bar{\mathsf{x}}\to vy,a} \tag{30}$$

$$\sigma^2_{vy,a\to\mathsf{y},a} = \sigma^2_{\mathsf{w},a} + N^2\mu_B^2 \sigma^2_{\bar{\mathsf{x}}\to vy,a}\,. \tag{31}$$

This can be shown by expressing (20) as a function of $\bar{\mathsf{x}}$, computing the product with the Gaussian message given by the parameters $(\mu_{\bar{\mathsf{x}}\to vy,a}, \sigma^2_{\bar{\mathsf{x}}\to vy,a})$ and finally evaluating the Gaussian integral (22) with respect to $\bar{\mathsf{x}}$. Subsequently, one iteration of AMP can be computed with $\mu_{vy,a\to y,a}$ substituted for y_a and $\sigma^2_{vy,a\to y,a}$ substituted for $\sigma^2_{w,a}$. The initial value of $\sigma^2_{\mathsf{z},a}$ can be computed as the empirical variance of $\mu_{vy,a\to y,a}$. After this iteration, the message $(\mu_{y,a\to vy,a}, \sigma^2_{y,a\to vy,a})$ is

$$\mu_{y,a\to vy,a} = \sum_i A_{a,i}\mu_{i\to a} \tag{32}$$

$$\approx \sum_i A_{a,i}\mu_{\mathsf{x},i} - \frac{\mu_{\mathsf{z},a}}{L} \sum_j \frac{\partial F}{\partial \mu^{(l)}} \tag{33}$$

$$\sigma^2_{y,a\to vy,a} = \sum_i A^2_{a,i}\sigma^2_{i\to a} \approx \frac{N}{L}\sigma^2_{\mathsf{x}} = \sigma^2_{y\to vy}\,. \tag{34}$$

In (32), $\mu_{i\to a}$ is the message from variable x_i to sensing factor node $f_{\boldsymbol{A},a}$. The approximation of $\mu_{i\to a}$ derived for AMP [11] has been used in the step from (32) to (33). Given v_a and the message from y_a to $f_{vy,a}$, the message from $f_{vy,a}$ to $\bar{x}$ can be computed:

$$N\mu_B\mu_{vy,a\to\bar{x}} = v_a - \mu_{y,a\to vy,a} \tag{35}$$

$$(N\mu_B)^2\sigma^2_{vy,a\to\bar{x}} = \sigma^2_{y\to vy} + \sigma^2_{\mathsf{w},a}\,. \tag{36}$$

At $\bar{x}$ there are now $L+1$ incoming messages, namely L from the nodes $f_{vy,a}$ and one from $f_{\bar{x}}$. Formally and following (25), the message from $\bar{x}$ to the factor nodes $f_{vy,a}$ is

$$\sigma^2_{\bar{x}\to vy,a} = \left(\frac{1}{\sigma^2_{f,\bar{x}\to\bar{x}}} + \sum_{b\neq a}\frac{1}{\sigma^2_{vy,b\to\bar{x}}}\right)^{-1} \tag{37}$$

$$\approx \left(\frac{1}{\sigma^2_{f,\bar{x}\to\bar{\mathsf{x}}}} + \frac{L-1}{\sigma^2_{vy\to\bar{\mathsf{x}}}}\right)^{-1} \tag{38}$$

$$\mu_{\bar{x}\to vy,a} = \sigma^2_{\bar{x}\to vy,a}\left(\frac{\mu_{f,\bar{x}\to\bar{x}}}{\sigma^2_{f,\bar{x}\to\bar{x}}} + \sum_{b\neq a}\frac{\mu_{vy,b\to\bar{x}}}{\sigma^2_{vy,b\to\bar{x}}}\right) \tag{39}$$

$$\approx \sigma^2_{\bar{\mathsf{x}}\to vy}\left(\frac{\mu_{f,\bar{\mathsf{x}}\to\bar{\mathsf{x}}}}{\sigma^2_{f,\bar{\mathsf{x}}\to\bar{\mathsf{x}}}} + \frac{1}{\sigma^2_{vy\to\bar{\mathsf{x}}}}\sum_b \mu_{vy,b\to\bar{\mathsf{x}}}\right). \tag{40}$$

In the step from (37) to (38) and (39) to (40), the independence of $\sigma^2_{y\to vy}$ from a was used (c.f. (34)). The resulting AMP algorithm for a nonzero-mean sensing matrix (MEAN-AMP) is listed as Algorithm 2.

The behavior of MEAN-AMP is intuitive: it estimates $\bar{x}$ from the current estimate of $\mathbf{x}$ and the residual $\mathbf{z}$. The estimate's variance influences the convergence of the algorithm through $\sigma^{2(l)}_{\mathsf{x}}$.

4. STATE EVOLUTION

It is possible to formulate a state evolution recursion by taking into account modifications to $\sigma^{2(l)}_{\mathsf{x}}$. State evolution [2] for standard AMP is defined by the recursion

$$\sigma^2_{\mathsf{x},[t]} = \mathbb{E}\left\{\left(F_{[t-1]}\left(x_0 + \sigma^{(l)}_{\mathsf{x},[t-1]}\mathsf{z}\right) - x_0\right)^2\right\} \tag{41}$$

$$\sigma^{2(l)}_{\mathsf{x},[t]} = \sigma^2_{\mathsf{w}} + \rho^{-1}\sigma^2_{\mathsf{x},[t]} = \sigma^2_{\mathsf{w}} + \sigma^2_{\tilde{\mathsf{z}}}\,, \tag{42}$$

where z is random variable distributed according to the standard normal distribution and x_0 is the true value of x. MEAN-AMP modifies the second line of the state evolution since it introduces an effective noise variance $\sigma^2_{vy\to y}$. Its value is lower bounded by $\sigma^2_{\mathsf{w}} + \rho^{-1}\sigma^2_{\mathsf{x},[t]}$ and achieves this bound if $\mu_B = 0$. Thus, for a mean-free sensing matrix $\boldsymbol{A}$, state evolution for MEAN-AMP reduces to state evolution for standard AMP. The recursion can be written as

$$\sigma^2_{\mathsf{x},[t]} = \mathbb{E}\left\{\left(F_{[t-1]}\left(x_0 + \sigma^{(l)}_{\mathsf{x},[t-1]}\mathsf{z}\right) - x_0\right)^2\right\} \tag{43}$$

$$\sigma^2_{\tau} = \sigma^2_{\mathsf{w}} + \rho^{-1}\sigma^2_{\mathsf{x},[t]} = \sigma^2_{\mathsf{w}} + \sigma^2_{\tilde{\mathsf{z}}} \tag{44}$$

$$\sigma^2_{\kappa} = \Big(\frac{1}{N\mu_B^2\sigma^2_{\mathsf{x},[t]}} + \frac{L-1}{\sigma^2_{\tau}}\Big)^{-1} \tag{45}$$

$$\sigma^{2(l)}_{\mathsf{x},[t]} = \sigma^2_{\tau} + \sigma^2_{\kappa}\,. \tag{46}$$

If the algorithm converges, $\sigma^2_{\mathsf{x},[t]} \to 0$, thus $\sigma^2_{\kappa} \to 0$ and $\sigma^{2(l)}_{\mathsf{x},[t]}$ assumes the value predicted by standard AMP state evolution.

Algorithm 2 MEAN-AMP

Set constants σ_{w}^2, $t_{\max}$, ϵ, $\boldsymbol{A} = \boldsymbol{B} - \mu_B \mathbf{1}_L \mathbf{1}_N^T$

$\boldsymbol{\mu}_{\mathsf{z}} = \boldsymbol{v} - \bar{v}\mathbf{1}_L$

$\sigma_{\mathsf{x}}^{2(l)} = (\sigma_{\mathsf{w}}^2 + L^{-1}\left\|\boldsymbol{\mu}_{\mathsf{z}}\right\|_2^2)(1 + L^{-1})$

All other variables are initialized to zero.

repeat

$\boldsymbol{\mu}_{\mathsf{x}}^{(l)} = \boldsymbol{A}^T \boldsymbol{\mu}_{\mathsf{z}} + \boldsymbol{\mu}_{\mathsf{x}}$

$\boldsymbol{\mu}_{\mathsf{x}} = \boldsymbol{F}(\boldsymbol{\mu}_{\mathsf{x}}^{(l)}, \sigma_{\mathsf{x}}^{2(l)})$

$\boldsymbol{\sigma}_{\mathsf{x}}^2 = \boldsymbol{G}(\boldsymbol{\mu}_{\mathsf{x}}^{(l)}, \sigma_{\mathsf{x}}^{2(l)})$

$\sigma_{vy\to\bar{\mathsf{x}}}^2 = \frac{1}{(N\mu_B)^2}\left(\frac{1}{L}\sum_i \sigma_{\mathsf{x},i}^2 + \sigma_{\mathsf{w}}^2\right)$

$\sigma_{\bar{\mathsf{x}}\to vy}^2 = \left(\frac{N^2}{\sum_i \sigma_{\mathsf{x},i}^2} + \frac{L-1}{\sigma_{vy\to\bar{\mathsf{x}}}^2}\right)^{-1}$

$\sigma_{vy\to y}^2 = \sigma_{\mathsf{w}}^2 + (N\mu_B)^2 \sigma_{\bar{\mathsf{x}}\to vy}^2$

$\sigma_{\mathsf{x}}^{2(l)} = \sigma_{vy\to y}^2 + \frac{1}{L}\sum_i \sigma_{\mathsf{x},i}^2$

$\mu_{f,\bar{\mathsf{x}}\to\bar{\mathsf{x}}} = \frac{1}{N}\sum_i \mu_{\mathsf{x},i}$

$\mu_{vy,a\to\bar{\mathsf{x}}} = \frac{1}{N\mu_B}\left(v_a - \sum_i A_{a,i}\mu_{\mathsf{x},i} + \frac{\mu_{\mathsf{z},a}}{L}\sum_i \frac{\partial F}{\partial \mu_{\mathsf{x},i}^{(l)}}\right)$

$\mu_{\bar{\mathsf{x}}\to vy} = \sigma_{\bar{\mathsf{x}}\to vy}^2 \left(\frac{\mu_{f,\bar{\mathsf{x}}\to\bar{\mathsf{x}}}}{\sigma_{f,\bar{\mathsf{x}}\to\bar{\mathsf{x}}}^2} + \frac{L-1}{L\sigma_{vy\to\bar{\mathsf{x}}}^2}\sum_b \mu_{vy,b\to\bar{\mathsf{x}}}\right)$

$\boldsymbol{\mu}_{vy\to\boldsymbol{y}} = \boldsymbol{v} - N\mu_B \mu_{\bar{\mathsf{x}}\to vy}\mathbf{1}_L$

$\boldsymbol{\mu}_{\mathsf{z}} = \boldsymbol{\mu}_{vy\to\boldsymbol{y}} - \boldsymbol{A}\boldsymbol{\mu}_{\mathsf{x}} + \boldsymbol{\mu}_{\mathsf{z}}\frac{1}{L}\sum_i \frac{\partial F}{\partial \mu_{\mathsf{x},i}^{(l)}}$

until $t > t_{\max}$ **or** $t > 1$ **and** $\left\|\boldsymbol{\mu}_{\mathsf{x}} - \boldsymbol{\mu}_{\mathsf{x}}^{[t-1]}\right\|_2^2 < \epsilon \left\|\boldsymbol{\mu}_{\mathsf{x}}\right\|_2^2$

$\hat{\boldsymbol{x}} = \boldsymbol{\mu}_{\mathsf{x}}$

5. NUMERICAL RESULTS

Performance is evaluated using the signal-to-distortion ratio:

$$\mathrm{SDR}_{\mathrm{dB}} = 10\log_{10}\left(\|\boldsymbol{x}\|_2^2 \,/\, \|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2^2\right). \tag{47}$$

A comparison of AMP, MR-AMP and MEAN-AMP is shown in Fig. 2. The performance of MEAN-AMP is almost identical to AMP for $\mu_B = 0$ and does not depend on μ_B, as expected. Meanwhile, MR-AMP is limited by the estimation error ε of $\bar{y}$. The dotted curves show state evolution estimates for AMP and MR-AMP while the dashed curve shows a state evolution estimate for MEAN-AMP at $\mu_B = 1$.

Results from state evolution can be seen in Fig. 3. Even though σ_κ^2 is small, failing to improve the estimate of $\bar{x}$ causes it to exhibit an "error floor", which limits the performance of MR-AMP. In fact, σ_κ^2 is so small that numerically, $\sigma_{\mathsf{x}}^{2(l)} = \sigma_\tau^2$, which explains why state evolution of MEAN-AMP is almost identical to AMP.

The estimated variance of $\mu_{\bar{\mathsf{x}}\to vy}$ agrees with the estimation error, as shown in Fig. 4.

6. CONCLUSION

By quantifying the error incurred by mean-removal, we were able to use standard AMP for recovery and accurately predict the limits of its performance. We also present an extended algorithm which iteratively improves upon the estimate obtained by mean-removal. This algorithm achieves results comparable to AMP in the zero-mean matrix regime. A derivation based on belief propagation is provided. Our state evolution formalism accurately predicts performance for both algorithms.

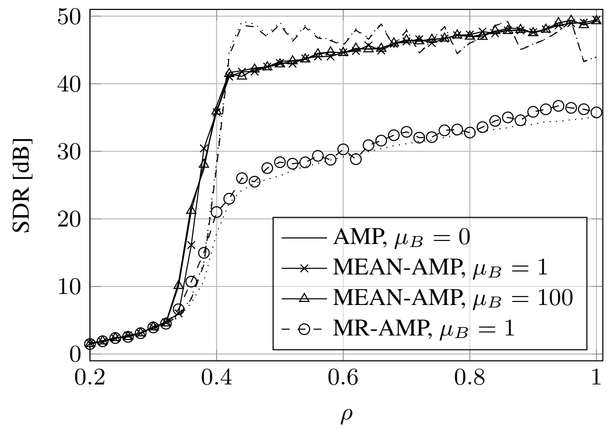

Fig. 2: Performance in terms of SDR vs. subsampling ratio $\rho = L/N$ with $N = 10^3, \sigma_{\mathsf{w}}^2 = 0$ and sparsity $\gamma = 0.2$.

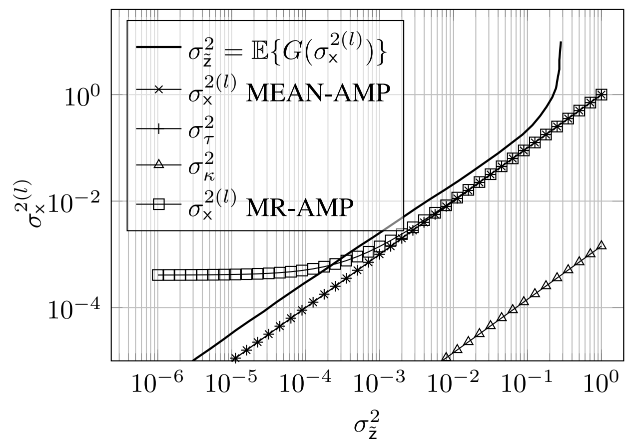

Fig. 3: "Exit-plot": the evolution of $\sigma_{\bar{\mathsf{z}}}^2$ and $\sigma_{\mathsf{x}}^{2(l)}$, estimated with state evolution for $\mu_B = 100, L = 700, N = 1000, \sigma_{\mathsf{w}}^2 = 0$ and $\gamma = 0.2$. The effective estimation noise is thus $\sigma_\varepsilon^2 \approx 4.1 \cdot 10^{-4}$.

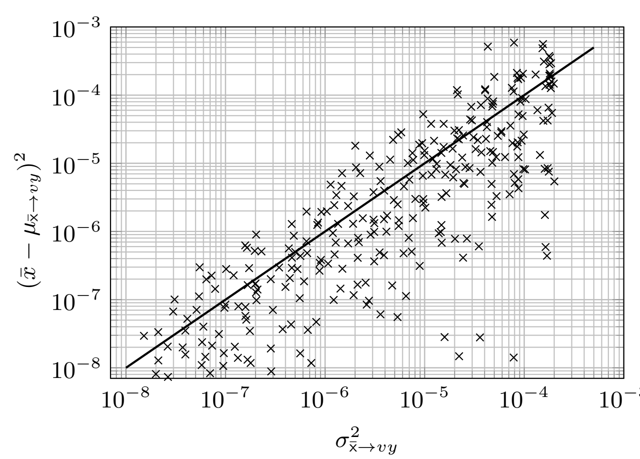

Fig. 4: Estimated variance of $\mu_{\bar{\mathsf{x}}\to vy}$ versus squared estimation error from 30 runs of MEAN-AMP with $\mu_B = 1$, $L = 700$, $N = 1000$, $\sigma_{\mathsf{w}}^2 = 0$ and $\gamma = 0.2$. The diagonal line indicates the region of equality of the two values.

7. REFERENCES

[1] I. Daubechies, M. Defrise, and C. De Mol, “An iterative thresholding algorithm for linear inverse problems with a sparsity constraint,” *Communications on Pure and Applied Mathematics*, vol. 57, no. 11, pp. 1413–1457, Nov. 2004.

[2] David L. Donoho, Arian Maleki, and Andrea Montanari, “Message-passing algorithms for compressed sensing,” *Proceedings of the National Academy of Sciences*, vol. 106, no. 45, pp. 18914–18919, Nov 2009.

[3] David L. Donoho, Arian Maleki, and Andrea 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.

[4] D.L. Donoho, A. Maleki, and A. Montanari, “Message passing algorithms for compressed sensing: II. analysis and validation,” in *Proceedings IEEE Information Theory Workshop (ITW)*, 2010, pp. 1–5.

[5] Sundeep Rangan, “Generalized approximate message passing for estimation with random linear mixing,” in *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*. IEEE, Jul 2011, pp. 2168–2172.

[6] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, “Single-pixel imaging via compressive sampling,” *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 83–91, Mar 2008.

[7] F. Caltagirone, L. Zdeborová, and F. Krzakala, “On convergence of approximate message passing,” in *2014 IEEE International Symposium on Information Theory*, June 2014, pp. 1812–1816.

[8] J. Vila, P. Schniter, S. Rangan, F. Krzakala, and L. Zdeborová, “Adaptive damping and mean removal for the generalized approximate message passing algorithm,” in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2015, pp. 2021–2025.

[9] S. Rangan, P. Schniter, and A. Fletcher, “On the convergence of approximate message passing with arbitrary matrices,” in *2014 IEEE International Symposium on Information Theory*, Jun 2014, pp. 236–240.

[10] C.M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

[11] 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.