On Introducing Damping to Bayes Optimal Approximate Message Passing for Compressed Sensing

Kazushi Mimura
Department of Information Sciences, Hiroshima City University,
Hiroshima 731-3194, Japan
Email: mimura@hiroshima-cu.ac.jp

***Abstract*—Some damping methods are introduced to the Bayes optimal approximate message passing for compressed sensing. We show that these damping methods make the algorithm converge against a typical problem that algorithms without damping fail to converge by numerical experiments.**

I. Introduction

Let us consider a problem that an N-dimensional vector $\boldsymbol{x} \in \mathbb{R}^N$ is estimated from an M-dimensional ($M < N$) vector $\boldsymbol{y} \in \mathbb{R}^M$:

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

through a known $M \times N$ matrix $A \in \mathbb{R}^{M\times N}$. Here, $\boldsymbol{w} \in \mathbb{R}^M$ denotes a noise vector $\boldsymbol{w} \sim \mathcal{N}(\boldsymbol{0}, \sigma_w^2 I)$ and $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the multi-dimensional Normal distribution with a mean vector $\boldsymbol{\mu}$ and a covariance matrix $\boldsymbol{\Sigma}$. When the ratio $\delta = M/N (\leq 1)$, which is called the *compression rate*, this system of equations are undetermined. The original vector $\boldsymbol{x}$ may be able to be recovered if we have some knowledge of the original vector such as the sparsity. This problem [6], [18], [10] is termed the reconstruction problem of compressed sensing [11], [3], [4], [5].

The linear programming (LP) methods are widely applied and investigated [11], [3], [4], [5]. As one of promising methods to solve such undetermined systems, the approximate message passing (AMP) proposed by Donoho et al. is widely discussed so far [12]. The another message-passing-based algorithm is the generalized AMP (GAMP) developed by Rangan, which estimates an i.i.d. random vector observed through a noisy linear measurement channel [16].

The performance and convergence of AMP are guaranteed by a solid theory for measurement matrices having i.i.d entries of zero mean [12], [16], [1]. Unfortunately, one has observed that AMP often fail to converge for more general matrices. To treat a given general non-Gaussian sensing matrix, it is important to improve the convergence. Rangan et al. discussed the convergence of AMP by introducing the damped GAMP that a Damping factor is introduced [17]. They provided sufficient conditions for the convergence of a damped GAMP. Another approach for the convergence issue is the mean removal [2].

In this paper, we focus on some methods to introduce damping to *Bayes optimal approximate message passing (BoAMP)*, that is a special case of the spatially coupled AMP [13], i.e., not a coupled system but a single system. Note that BoAMP can be also regarded as a special case of GAMP. We consider two kinds of damping, namely an ordinary damping, that is used in the damped GAMP, and *MSE damping* introduced here.

This paper is organized as follows. The next section introduces some reconstruction algorithms including BoAMP. In Section III, we define damping. Section IV explains its experiments. The final section is devoted to a summary.

II. Preliminalies

We assume the following to simplify the problem. Each element of the *original vector* $\boldsymbol{x} = (x_j) \in \mathbb{R}^N$, is an i.i.d. random variable which obeys the distribution $p_X(x)$

A. AMP

Donoho et. al. have developed the following iterative algorithm achieving the performance of LP-based reconstruction. AMP iteratively calculates an estimate

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \|\boldsymbol{x}\|_p \text{ s.t. } \boldsymbol{y} = A\boldsymbol{x} \tag{2}$$

by applying ℓ_p minimization.

Definition 1 (AMP[12]): Starting from an initial guess $\boldsymbol{x}^{(0)} = \boldsymbol{0}$ and $\boldsymbol{z}^{(0)} = \boldsymbol{y}$, AMP iteratively proceeds by

$$\boldsymbol{x}^{(t+1)} = \eta_t(A^\top \boldsymbol{z}^{(t)} + \boldsymbol{x}^{(t)}), \tag{3}$$

$$\boldsymbol{z}^{(t)} = \boldsymbol{y} - A\boldsymbol{x}^{(t)} + \frac{1}{\delta}\boldsymbol{z}^{(t-1)}\langle \eta'_{t-1}(A^\top \boldsymbol{z}^{(t-1)} + \boldsymbol{x}^{(t-1)})\rangle. \tag{4}$$

$\boldsymbol{x}^{(t)} \in \mathbb{R}^N$ is an estimate of the original vector $\boldsymbol{x}$ at stage t. Here, $\{\eta_t\}$ is an appropriate sequence of denoiser functions (applied componentwise):

$$\eta_t(m) = \{m - v_t \operatorname{sgn}(m)\}\mathbb{I}(|m| > v_t) \tag{5}$$

and v_t is an (empirical) mean squared error, e.g., $v_t = \|\boldsymbol{z}^{(t)}\|_2^2/M$. $A^\top$ denotes the transpose of A and $\eta'_t(u) = \partial\eta_t(u)/\partial u$. For a vector $\boldsymbol{v} = (v_1, \cdots, v_N)$, $\langle \boldsymbol{v}\rangle \triangleq N^{-1}\sum_{n=1}^N v_n$. □

Here, $\mathbb{I}(A)$ denotes an indicator function that takes 1 if A is true and 0 otherwise.

B. Bayes Optimal AMP

For a given prior distribution p_X, an optimal estimator which minimizes the mean square error (MSE)

$$\hat{\boldsymbol{x}} = \int_{\mathbb{R}^N} \boldsymbol{x} p(\boldsymbol{x}|A,\boldsymbol{y})\mathrm{d}\boldsymbol{x} \tag{6}$$

by applying the Bayesian inference, $p(\boldsymbol{x}|A,\boldsymbol{y}) = p(\boldsymbol{y}|A,\boldsymbol{x})\, p_X(\boldsymbol{x})$ denotes a posterior distribution and $p(\boldsymbol{y}|A,\boldsymbol{x})$ is a likelihood that represents a noise distribution. This estimator is called as the minimum mean square error (MMSE) estimator. Applying the message passing algorithm, the following is obtained.

Definition 2 (Bayes optimal AMP[13]): Starting from an initial guess $\boldsymbol{x}^{(0)} = \mathbf{0}$ and $\boldsymbol{z}^{(0)} = \boldsymbol{y}$, BoAMP iteratively proceeds by

$$\begin{aligned} \boldsymbol{x}^{(t+1)} &= \eta_t(A^\top \boldsymbol{z}^{(t)} + \boldsymbol{x}^{(t)}), & (7)\\ \boldsymbol{z}^{(t)} &= \boldsymbol{y} - A\boldsymbol{x}^{(t)} \\ &\quad + \frac{1}{\delta}\boldsymbol{z}^{(t-1)}\langle \eta'_{t-1}(A^\top \boldsymbol{z}^{(t-1)} + \boldsymbol{x}^{(t-1)})\rangle. & (8)\end{aligned}$$

$\boldsymbol{x}^{(t)} \in \mathbb{R}^N$ is an estimate. Here, the denoiser function η_t is

$$\eta_t(m) = \mathbb{E}_{X,Z}\{X|X + v_t^{-1/2}Z = m\},\ Z \sim \mathcal{N}(0,1) \tag{9}$$

and v_t is again an (empirical) mean squared error, e.g., $v_t = \|\boldsymbol{z}^{(t)}\|_2^2/M$. □

Note that (7) and (8) are identical to (3) and (4), respectively. The difference between AMP and BoAMP is the definition of denoiser function η_t. The denoiser function depends on the prior p_X. For example, $\eta_t(m) = \epsilon m(1 + v_t^{-1})^{-1}\ \gamma(m; 1 + v_t^{-1})/\{\epsilon\gamma(m; 1 + v_t^{-1})\ + (1-\epsilon)\gamma(m; v_t^{-1})\}$ for the Gauss-Bernoulli type prior $p_X = (1-\epsilon)\delta(0) + \epsilon\mathcal{N}(0,1)$, where $\gamma(m; s) := (2\pi s)^{-1/2}\exp[-m^2/(2s)]$ and $\delta(m)$ denotes the Dirac's delta.

III. Damped BoAMP

First, we explain the concept of damping. For the following iterative equation with a variable

$$x^{(t+1)} = f(x^{(t)}), \tag{10}$$

the damping is introduced as

$$x^{(t+1)} = (1-\theta)x^{(t)} + \theta f(x^{(t)}). \tag{11}$$

We here call this the simple damping. Since damping is empirically introduced, there are several methods to introduce damping other than this. The stationary equation of (10) can be given by substituting x into $x^{(t)}$, which gives $x = f(x)$. That of (11) is also equal to $x = f(x)$.

Here we restrict ourselves to two kinds of damping. The first one is the simple damping as follows.

Definition 3 (S-DBoAMP): Let $\theta \in (0,1]$ be a damping constant. The iterative equations of BoAMP is modified to

$$\begin{aligned} \boldsymbol{x}^{(t+1)} &= (1-\theta)\boldsymbol{x}^{(t)} + \theta\eta_t(A^\top \boldsymbol{z}^{(t)} + \boldsymbol{x}^{(t)}), & (12)\\ \boldsymbol{z}^{(t)} &= \boldsymbol{y} - A\boldsymbol{x}^{(t)} \\ &\quad + \theta\frac{1}{\delta}\boldsymbol{z}^{(t-1)}\langle \eta'_{t-1}(A^\top \boldsymbol{z}^{(t-1)} + \boldsymbol{x}^{(t-1)})\rangle. & (13)\end{aligned}$$

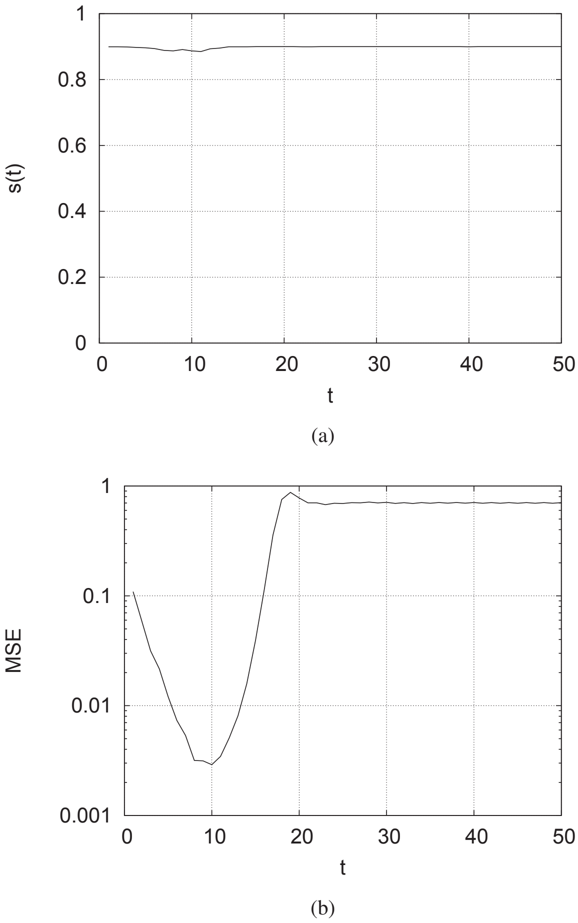


Fig. 1. Expamle of dynamics of BoAMP without damping. $\gamma = 2.5$. (a) $s(t)$. (b) MSE.

Other settings are same to BoAMP. We here call this algorithm the simple damping Bayes optimal AMP (S-DBoAMP). □

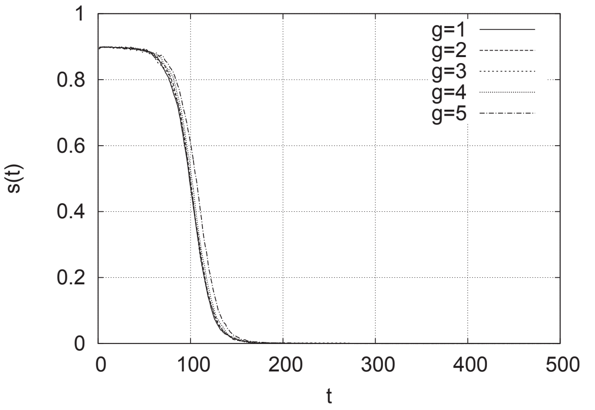
This algorithm has a close relationship to the damped GAMP [17]. When $\theta = 1$, the simple damping BoAMP reduces to BoAMP without damping. The other methods is to damp a estimate value before denoising as follows.

Definition 4 (MSE-DBoAMP): Let $c^{(t)} \in (0,1]$ be a damping function, which is monotonically increasing. The iterative equations of BoAMP is modified to

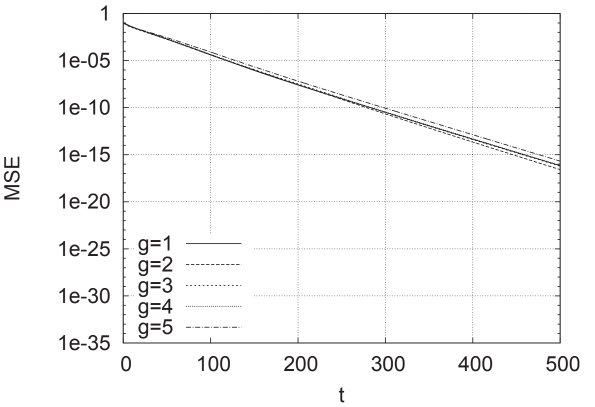
$$\begin{aligned} \boldsymbol{x}^{(t+1)} &= \eta_t(c^{(t)}A^\top \boldsymbol{z}^{(t)} + \boldsymbol{x}^{(t)}), & (14)\\ \boldsymbol{z}^{(t)} &= \boldsymbol{y} - A\boldsymbol{x}^{(t)} \\ &\quad + c^{(t)}\frac{1}{\delta}\boldsymbol{z}^{(t-1)}\langle \eta'_{t-1}(c^{(t)}A^\top \boldsymbol{z}^{(t-1)} + \boldsymbol{x}^{(t-1)})\rangle. & (15)\end{aligned}$$

Here, the denoiser function η_t (9) is defined by using a damped (empirical) MSE $v_t = c^{(t)}\|\boldsymbol{z}^{(t)}\|_2^2/M$. Other settings are same to BoAMP. This algorithm is referred to as the MSE damping Bayes optimal AMP (MSE-DBoAMP). □

If $c^{(t)}$ is small, this algorithm has a similar effect of the simple damped BoAMP. To accelerate convergence, an increasing

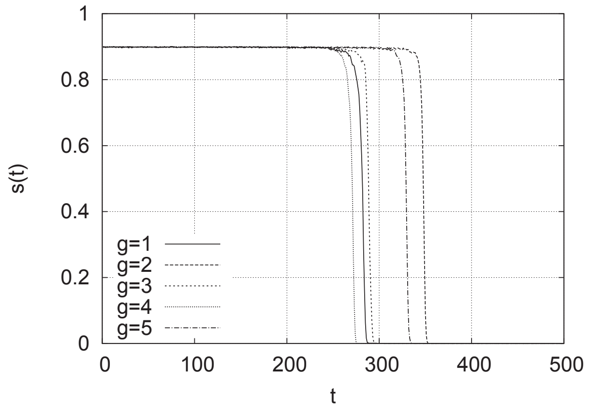


(a)

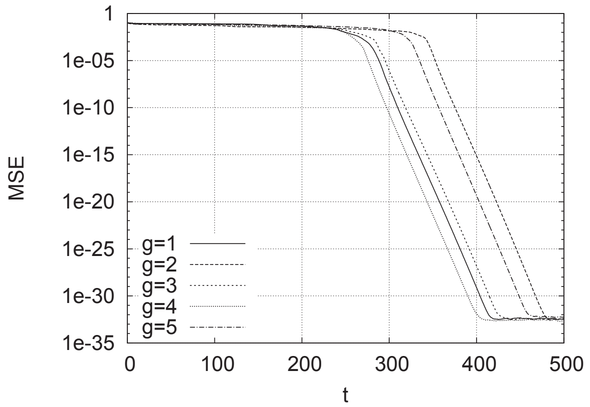


(b)

Fig. 2. Examples of dynamics of S-DBoAMP. $\gamma \in \{1, 2, 3, 4, 5\}$. (a) $s(t)$. (b) MSE.



(a)



(b)

Fig. 3. Examples of dynamics of MSE-DBoAMP. $\gamma \in \{1, 2, 3, 4, 5\}$. (a) $s(t)$. (b) MSE.

function is used as $c^{(t)}$. In both damping methods, the damping constant or function must be properly chosen to converge iterations.

IV. Experiments

For non-zero mean sensing matrices, AMP like algorithms fail to converge [2]. As a simple example for this situation, we, therefore, consider matrices with entries generated as follows

$$a_{i,j} = \frac{\gamma}{N} + \frac{1}{\sqrt{M}}\mathcal{N}(0, 1). \tag{16}$$

The mean and variance need to be scaled as $O(1/N)$ to have a $O(1)$ output for a $O(1)$ input. The dimension of the observation M is supposed to have the same order.

To validate the results obtained above, we performed numerical experiments in $N = 2,000$ systems. The dimension of the observation $\boldsymbol{y}$ is $M = 1,000$. The prior distribution is the Gauss-Bernoulli type prior $p_X = (1 - \epsilon)\delta(0) + \epsilon\mathcal{N}(0, 1)$ with the sparsity $\epsilon = 0.1$. The non-zero mean of the sensing matrix is $\gamma = 200$. In this problem setting, BoAMP without damping fails to converge.

To evaluate the quality of estimates, we define the following quantity $s(t) = |\operatorname{supp}(\boldsymbol{x}) - \operatorname{supp}(\boldsymbol{x}^{(t)})|/N$, which is a cardinality (ratio) of a difference set between support sets of the original vector and estimate. Therefore, $s(t) = 0$ means perfect support estimation. Figure 2 shows examples of dynamics of S-DBoAMP at $\gamma \in \{1, 2, 3, 4, 5\}$. We set $\theta = 0.3$, which is a fixed value. Figure 3 shows examples of dynamics of MSE-DBoAMP. We set $c^{(t)} = 0.4 \times 1.001^t \mathbb{I}(0.4 \times 1.001^t < 1) + \mathbb{I}(0.4 \times 1.001^t \geq 1)$. It can be confirmed that both algorithms successfully converge. It should be noted that since MSE-DBoAMP gets closet to BoAMP without damping, one might have to choose proper termination condition, e.g., the empirical MSE is less than 10^{-10}.

V. Summary

We introduced two kinds of damping to the Bayes optimal approximate message passing algorithm. Both can converge dynamics for nonzero mean sensing matrices. These can be also applied to systems with non-Gaussian sensing matrices and spatially coupled systems.

Since the damping factor or damping function must be properly chosen, the sufficient condition should be derived analytically. It might be better to discuss whether the dynamical property can improve or not, when several damping methods are used. These problems are under way.

ACKNOWLEDGMENT

This work was partially supported by a Grant-in-Aid for Scientific Research (B) No. 25289114 and (C) No. 22500136 from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan.

REFERENCES

[1] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *Proc. of the 2010 IEEE Int'l Sympo. on Info. Theory (ISIT)*, 1528 (2010). [the longer version will appear in *IEEE Transactions on Information Theory*]

[2] F. C. Caltagirone, L. Zdeborová, and F. Krzakala, "On convergence of Approximate Message Passing," *Proc. of the 2014 IEEE Int'l Sympo. on Info. Theory (ISIT)*, 1812 (2014).

[3] E. J. Candés and T. Tao, "Decoding by linear programming," *IEEE Trans. Info. Theory*, vol. 51, no. 12, 4203 (2005).

[4] E. J. Candés, J. Romberg and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Info. Theory*, vol. 52, no. , 489 (2006).

[5] E. J. Candés and T. Tao, "Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies?" *IEEE Trans. Info. Theory*, vol. 52, no. 12, 5406 (2006).

[6] J. F. Claerbout and F. Muir, "Robust modeling with erratic data," *Geophysics*, 38, 826 (1973).

[7] A. C. C. Coolen, *The Mathematical Theory of Minority Games*, Oxford Univ. Press (2005).

[8] E. T. Copson, *Asymptotic Expansions*, Cambridge Univ. Press (1965).

[9] I. Daubechies, M. Defrise and C. De-Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, no. 11, 1413 (2004).

[10] D. L. Donoho, "Uncertainty Principles and Signal Recovery," *SIAM J. Appl. Math.*, vol. 49, no. 3, 906 (1989).

[11] *IEEE Trans. Info. Theory*, vol. 52, no. 4, 1289 (2006).

[12] D. L. Donoho, A. Maleki and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. of the National Academy of Sciences (PNAS)*, vol. 106, no. 45, 18914 (2009).

[13] D. L. Donoho, A. Javanmard, and A. Montanari, "Information-Theoretically Optimal Compressed Sensing via Spatial Coupling and Approximate Message Passing,"' *IEEE Trans. Info. Theory*, vol. 59, no. 11, 7434 (2013).
D. L. Donoho, A. Maleki and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. of the National Academy of Sciences (PNAS)*, vol. 106, no. 45, 18914 (2009).

[14] H. Eissfeller and M. Opper, "New method for studying the dynamics of disordered spin systems without finite-size effects." *Phys. Rev. Lett.*, vol. 68, no. 13, 2094 (1992).

[15] K. Mimura K and M. Okada, "Generating functional analysis of CDMA detection dynamics," *J. Phys. A: Math. Gen.*, vol. 38, no. 46, 9917 (2005).

[16] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," arXiv:1010.5141v1 [cs.IT], (2010).

[17] S. Rangan, P. Schniter, and A. Fletcher, "On the convergence of Approximate Message Passing with Arbitrary Matrices," *Proc. of the 2014 IEEE Int'l Sympo. on Info. Theory (ISIT)*, 236 (2014).

[18] F. Santosa and W. W. Symes, "Linear Inversion of Band-Limited Reflection Seismograms," *SIAM J. Sci. and Stat. Comput.*, vol. 7, no. 4, 1307 (1986).

[19] T. Tanaka and M. Okada, "Approximate Belief Propagation, Density Evolution, and Statistical Neurodynamics for CDMA Multiuser Detection," *IEEE Trans. Info. Theory*, vol. 51, no. 2, 700 (2005).

[20] J. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proc. of IEEE Appl. & Comput. Math.*, vol. 98, no. 6, 948 (2010).

[21] M. Zibulevsky and M. Elad, "L1-L2 Optimization in Signal and Image Processing," *IEEE Signal Proc. Mag.*, vol. 27, no. 3, 76 (2010).