Binary Vector Reconstruction via Discreteness-Aware Approximate Message Passing

Ryo Hayakawa* and Kazunori Hayashi†
* Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan
E-mail: rhayakawa@sys.i.kyoto-u.ac.jp
† Graduate School of Engineering, Osaka City University, Osaka 558-8585, Japan
E-mail: kazunori@eng.osaka-cu.ac.jp

***Abstract*—In this paper, we propose an algorithm to reconstruct a binary vector from its possibly underdetermined linear measurements. Taking advantage of the idea of the approximate message passing (AMP) algorithm for compressed sensing, we firstly formulate a probability distribution associated with the sum-of-absolute-value (SOAV) optimization. Then, by approximating the sum-product algorithm for the marginalization of the distribution, we obtain a low-complexity iterative algorithm, called discreteness-aware AMP (DAMP). We evaluate the performance of DAMP analytically via state evolution and derive a condition for the exact reconstruction with DAMP. Moreover, we also provide Bayes optimal DAMP for the binary vector reconstruction, which gives the minimum mean-square-error at each iteration in the large system limit. Simulation results show that DAMP can reconstruct the binary vector from underdetermined linear measurements and its performance can be well predicted by our theoretical results.**

I. Introduction

The reconstruction of a discrete-valued vector from its possibly underdetermined linear measurements is an important problem in signal processing. Since signals in digital communications are generally discrete-valued, there are many potential applications of the discrete-valued vector reconstruction in communications systems, such as multiuser detection [1]–[5], signal detection for overloaded multiple-input multiple-output (MIMO) systems [6], and faster-than-Nyquist signaling [7], [8]. The optimal maximum likelihood (ML) approach for the problem results in the combinatorial optimization problem and hence the computational complexity increases exponentially along the problem size. We thus need a low-complexity algorithm for the reconstruction, especially in the large-scale problem.

For the reconstruction of the discrete-valued vector, the regularization-based method and the transform-based method have been proposed [9]. These methods borrow the idea from compressed sensing [10], [11] in the formulation of the reconstruction leading to convex optimization problems. The optimization problems can be solved by interior point methods as the standard linear programming in the absence of observation noise. However, it may have a prohibitive computational complexity in large-scale systems. As the theoretical analysis, the required number of measurements has been derived for the binary vector reconstruction with the regularization-based method. For the transform-based method, a more general theorem has been proved for the uniformly distributed discrete-valued vector. For the vector with non-uniform distribution, however, no analytical result has been obtained.

Meanwhile, sum-of-absolute-value (SOAV) optimization has also been proposed for the discrete-valued vector reconstruction [12]. The SOAV optimization is similar to the regularization-based method and they are equivalent when the unknown vector is uniformly distributed. Since interior point methods for the SOAV optimization may have a huge computational complexity for large-scale problems, we have proposed a low-complexity algorithm for the SOAV optimization, referred to as discreteness-aware approximate message passing (DAMP) [13]. The DAMP algorithm is derived by employing the approach of the approximate message passing (AMP) algorithm [14], [15], which has been originally proposed for compressed sensing. The performance of DAMP has been analytically investigated via the state evolution framework [14], [16] and the required number of measurements for the perfect reconstruction has also been derived. The theoretical result for the reconstruction of the uniformly distributed vector coincides with the analysis of the regularization-based method and the transform-based method. In [13], however, the elements of the unknown discrete-valued vector is assumed to be symmetrically distributed.

In this paper, we propose the DAMP algorithm for the reconstruction of a binary vector whose distribution is possibly asymmetric. By this extension, for example, we can apply the DAMP algorithm to image processing for binary images [12] even when the number of black pixels is not equal to that of white pixels. In the derivation of the proposed algorithm, we firstly construct a probability distribution associated with the SOAV optimization, which is formulated differently from that in [13] to take account of the asymmetry. Next, we build the sum-product belief propagation algorithm [17], [18] over the corresponding factor graph. We then take the large system limit and approximate the message update rules to obtain a low-complexity iterative algorithm. The resultant

This work was supported in part by the Grants-in-Aid for Scientific Research no. 15K06064 and 15H02252 from the Ministry of Education, Culture, Sports, Science and Technology of Japan, and the Grant-in-Aid for JSPS Research Fellow no. 17J07055 from Japan Society for the Promotion of Science.

DAMP algorithm has basically the same form as that of the AMP algorithm for compressed sensing except for their soft thresholding functions. We evaluate the performance of DAMP analytically via the state evolution framework [14], [16] and derive a condition on a parameter and the required number of measurements for the exact reconstruction. Moreover, based on the state evolution, we also derive Bayes optimal DAMP, which gives the minimum mean-square-error (MSE) at each iteration in the large scale limit. Simulation results show that the proposed DAMP can reconstruct the binary vector from its underdetermined linear measurements and the performance can be well predicted by the theoretical result obtained with state evolution.

In the rest of the paper, we use the following notations. We represent the transpose by $(\cdot)^{\mathrm{T}}$, the vector whose elements are all 1 by $\mathbf{1}$, and the vector whose elements are all 0 by $\mathbf{0}$. For a vector $\boldsymbol{v} = [v_1 \cdots v_N]^{\mathrm{T}} \in \mathbb{R}^N$, we define the ℓ_1 and ℓ_2 norms of $\boldsymbol{v}$ as $\|\boldsymbol{v}\|_1 = \sum_{j=1}^N |v_j|$ and $\|\boldsymbol{v}\|_2 = \sqrt{\sum_{j=1}^N v_j^2}$, respectively. We represent the mean of the elements of $\boldsymbol{v}$ by $\langle \boldsymbol{v} \rangle = \frac{1}{N}\sum_{j=1}^N v_j$. For a function $h : \mathbb{R}^N \to \mathbb{R}$, the proximity operator [19] of h is defined as

$$\mathrm{prox}_h(\boldsymbol{v}) = \arg\min_{\boldsymbol{s}\in\mathbb{R}^N} \left\{ h(\boldsymbol{s}) + \frac{1}{2}\|\boldsymbol{s}-\boldsymbol{v}\|_2^2 \right\}. \tag{1}$$

We represent the sign function by $\mathrm{sgn}(\cdot)$.

II. Proposed Discreteness-aware AMP

In this section, we briefly review the SOAV optimization [12] and provide the proposed DAMP with the similar approach to that of the AMP algorithm for compressed sensing [14].

A. SOAV optimization

The SOAV optimization is a technique to reconstruct a binary vector such as $\boldsymbol{b} = [b_1 \cdots b_N]^{\mathrm{T}} \in \{r_1, \ldots, r_L\}^N \subset \mathbb{R}^N$ from its linear measurements

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{b}, \tag{2}$$

where $\boldsymbol{y} = [y_1 \cdots y_M]^{\mathrm{T}} \in \mathbb{R}^M$ and

$$\boldsymbol{A} = \begin{bmatrix} a_{1,1} & \cdots & a_{1,N} \\ \vdots & \ddots & \vdots \\ a_{M,1} & \cdots & a_{M,N} \end{bmatrix} \in \mathbb{R}^{M\times N}. \tag{3}$$

In this paper, we consider the binary case $\boldsymbol{b} \in \{r_1, r_2\}^N$ ($r_1 < r_2$) with the known distribution $\Pr(b_j = r_1) = p_1$ and $\Pr(b_j = r_2) = p_2$ ($= 1 - p_1$) ($j = 1, \ldots, N$). Taking advantage of the fact that $\boldsymbol{b} - r_1\mathbf{1}$ and $\boldsymbol{b} - r_2\mathbf{1}$ have approximately $p_1 N$ and $p_2 N$ zero elements, respectively, the SOAV optimization solves the following problem to obtain the estimate of $\boldsymbol{b}$.

$$\begin{aligned} \hat{\boldsymbol{b}} = \arg\min_{\boldsymbol{s}\in\mathbb{R}^N} \; & (q_1\|\boldsymbol{s} - r_1\mathbf{1}\|_1 + q_2\|\boldsymbol{s} - r_2\mathbf{1}\|_1) \\ \text{subject to} \; & \boldsymbol{y} = \boldsymbol{A}\boldsymbol{s}. \end{aligned} \tag{4}$$

In the original SOAV optimization [12], the coefficients $q_1, q_2 \geq 0$ are fixed as $q_1 = p_1$ and $q_2 = p_2$. The regularization-based method [9] solves (4) with $q_1 = q_2 = 1$. However, the validity of these selections has not been verified. We thus consider the coefficients as parameters to be optimized before solving (4), as described in Sec. III-B. Note that we can handle the case of the asymmetric distribution with $p_1 \neq p_2$ by using coefficients $q_1 \neq q_2$ in (4), while [13] considers only the case with symmetric distributions. While we consider only the noise-free case (2) in this paper, we can extend the DAMP algorithm for the noisy case as in the sparse vector reconstruction [15].

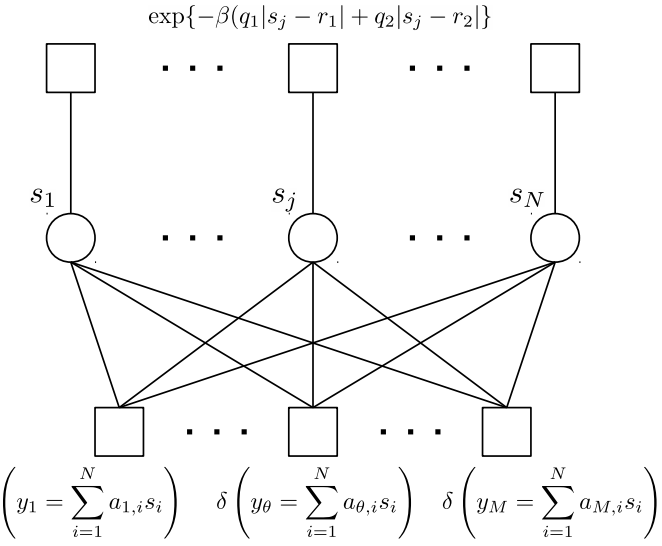

Fig. 1. Factor graph of (5): N circles denote the variable nodes $s_1, \ldots, s_N$. N squares in the upper side are the function nodes $\exp\{-\beta(q_1|s_j - r_1| + q_2|s_j - r_2|)\}$ $(j = 1, \ldots, N)$. M squares in the lower side are the function nodes $\delta\left(y_\theta = \sum_{i=1}^N a_{\theta,i}s_i\right)$ $(\theta = 1, \ldots, M)$.

B. Formulation of probability distribution

The derivation of DAMP begins with the sum-product algorithm for a probability distribution corresponding to the SOAV optimization (4). We first consider the following joint probability distribution

$$\begin{aligned} \mu(\boldsymbol{s}) \propto & \prod_{j=1}^N \exp\{-\beta(q_1|s_j - r_1| + q_2|s_j - r_2|)\} \\ & \cdot \prod_{\theta=1}^M \delta\left(y_\theta = \sum_{i=1}^N a_{\theta,i}s_i\right), \end{aligned} \tag{5}$$

where $\beta > 0$ and $\delta\left(y_\theta = \sum_{i=1}^N a_{\theta,i}s_i\right)$ represents the Dirac distribution on the hyperplane $y_\theta = \sum_{i=1}^N a_{\theta,i}s_i$. Note that, as $\beta \to \infty$, the mass of the distribution concentrates on the solution of (4). We obtain the marginal distribution of (5) via belief propagation with the sum-product algorithm [18]. The message passing procedure over the corresponding factor graph (See Fig. 1) can be written as

$$\nu_{j\to\theta}^{t+1}(s_j) \propto \exp\left\{-\beta\left(q_1|s_j-r_1|+q_2|s_j-r_2|\right)\right\} \cdot \prod_{\zeta\neq\theta}\hat{\nu}_{\zeta\to j}^{t}(s_j), \quad (6)$$

$$\hat{\nu}_{\theta\to j}^{t}(s_j) \propto \int \delta\left(y_\theta=\sum_{i=1}^{N}a_{\theta,i}s_i\right)\prod_{k\neq j}\nu_{k\to\theta}^{t}(s_k)\mathrm{d}\boldsymbol{s}_{\setminus j}, \quad (7)$$

where t is the iteration index and the integral $\int(\cdot)\mathrm{d}\boldsymbol{s}_{\setminus j}$ denotes the marginalization over all the variables except for s_j. Here, $\nu_{j\to\theta}^{t+1}(s_j)$ is the message from the variable node s_j to the function node $\delta\left(y_\theta=\sum_{i=1}^{N}a_{\theta,i}s_i\right)$, and $\hat{\nu}_{\theta\to j}^{t}(s_j)$ is the message from the function node to the variable node.

C. DAMP

Then, we consider the large system limit ($M,N\to\infty$ with fixed $M/N=\Delta$) and large β limit ($\beta\to\infty$) to approximate the message update rule (6), (7). In the derivation, we assume $a_{\theta,j}\in\{1/\sqrt{M},-1/\sqrt{M}\}$ to simplify the calculation as in [15]. This assumption is not very crucial because we can derive the same algorithm if $\boldsymbol{A}$ is composed of i.i.d. variables with zero mean and variance $1/M$. As in the derivation in [15], the message update rules (6), (7) can be reduced to

$$\boldsymbol{x}^{t+1}=\eta(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{z}^t+\boldsymbol{x}^t,\hat{\tau}^t), \quad (8)$$

$$\boldsymbol{z}^t=\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}^t+\frac{1}{\Delta}\boldsymbol{z}^{t-1}\langle\eta'(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{z}^{t-1}+\boldsymbol{x}^{t-1},\hat{\tau}^{t-1})\rangle, \quad (9)$$

$$\hat{\tau}^t=\frac{\hat{\tau}^{t-1}}{\Delta}\langle\eta'(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{z}^{t-1}+\boldsymbol{x}^{t-1},\hat{\tau}^{t-1})\rangle, \quad (10)$$

where $\boldsymbol{x}^t$ is the estimate of $\boldsymbol{b}$ at the tth iteration. The function η is given by

$$\eta(\boldsymbol{u},c)=\mathrm{prox}_{cJ}(\boldsymbol{u}), \quad (11)$$

where $J(\boldsymbol{s})=q_1\|\boldsymbol{s}-r_1\boldsymbol{1}\|_1+q_2\|\boldsymbol{s}-r_2\boldsymbol{1}\|_1$ is the objective function in (4). By the direct calculation described in [5], the jth element of $\mathrm{prox}_{cJ}(\boldsymbol{u})$ is written as

$$[\mathrm{prox}_{cJ}(\boldsymbol{u})]_j=\begin{cases}u_j-cQ_1 & (u_j<r_1+cQ_1)\\ r_1 & (r_1+cQ_1\le u_j<r_1+cQ_2)\\ u_j-cQ_2 & (r_1+cQ_2\le u_j<r_2+cQ_2)\\ r_2 & (r_2+cQ_2\le u_j<r_2+cQ_3)\\ u_j-cQ_3 & (r_2+cQ_3\le u_j)\end{cases}, \quad (12)$$

where $Q_1=-q_1-q_2, Q_2=q_1-q_2$ and $Q_3=q_1+q_2$. Since $[\mathrm{prox}_{cJ}(\boldsymbol{u})]_j$ is a function of only u_j, the function $\eta(\boldsymbol{u},c)$ is a element-wise function of $\boldsymbol{u}$. The jth element of $\eta'(\boldsymbol{u},c)$ is defined as $[\eta'(\boldsymbol{u},c)]_j=0$ if $[\eta(\boldsymbol{u},c)]_j\in\{r_1,r_2\}$, otherwise $[\eta'(\boldsymbol{u},c)]_j=1$.

It should be noted that the update equations of DAMP (8)–(10) are basically the same as those of the AMP algorithm for compressed sensing [15]. The only difference is the function $\eta(\boldsymbol{u},c)$, which is the soft thresholding function $[\eta(\boldsymbol{u},c)]_j=\mathrm{sgn}(u_j)\max\{|u_j|-c,0\}$ in the case of the sparse vector reconstruction. Hence, the function $\eta(\boldsymbol{u},c)$ given by (11) and (12) can be considered as the soft thresholding function for the binary vector reconstruction. Since (8)–(10) can be computed only with additions of vectors and multiplications of a matrix and a vector, the computational complexity of the algorithm is $O(MN)$ per iteration, which is lower than that of internal point methods $O(MN^2)$.

In the derivation of DAMP, we can also take a slightly different approach by considering the threshold level $\hat{\tau}^t$ as the parameter to be optimized. One of the choices is to replace $\hat{\tau}^t$ with $\tau\frac{\sigma_t}{\sqrt{\Delta}}$ as in the AMP algorithm for compressed sensing [14], where τ (≥ 0) is the parameter and $\sigma_t^2=\|\boldsymbol{x}^t-\boldsymbol{b}\|_2^2/N$ is the MSE of the estimate at the tth iteration. In this approach, (8) and (9) are rewritten as

$$\boldsymbol{x}^{t+1}=\eta\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{z}^t+\boldsymbol{x}^t,\tau\frac{\sigma_t}{\sqrt{\Delta}}\right), \quad (13)$$

$$\boldsymbol{z}^t=\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}^t+\frac{1}{\Delta}\boldsymbol{z}^{t-1}\left\langle\eta'\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{z}^{t-1}+\boldsymbol{x}^{t-1},\tau\frac{\sigma_{t-1}}{\sqrt{\Delta}}\right)\right\rangle. \quad (14)$$

Since the true solution $\boldsymbol{b}$ is unknown in practice, we use the alternative value for σ_t^2, e.g., $\hat{\sigma}_t^2=\|\boldsymbol{z}^t\|_2^2/N$ used in [20]. We consider this parameterized DAMP (13), (14) henceforth.

III. State Evolution for DAMP

In this section, we provide a theoretical analysis of DAMP with state evolution framework [14], [16]. By using state evolution, we give the required number of measurements for the perfect reconstruction and the parameter of the soft thresholding function minimizing the number of measurements in the large system limit. Moreover, we provide the Bayes optimal DAMP on the basis of state evolution.

A. State evolution

State evolution is a framework to analyze the performance of the AMP algorithm. In the large system limit, the MSE $\sigma_t^2=\|\boldsymbol{x}^t-\boldsymbol{b}\|_2^2/N$ of $\boldsymbol{x}^t$ can be predicted via state evolution. Similarly to the case of compressed sensing, the state evolution formula for the proposed DAMP (13), (14) is written as

$$\sigma_{t+1}^2=\Psi(\sigma_t^2) \quad (15)$$

under some assumptions, where

$$\Psi(\sigma^2)=\mathrm{E}\left[\left\{\eta\left(X+\frac{\sigma}{\sqrt{\Delta}}Z,\tau\frac{\sigma}{\sqrt{\Delta}}\right)-X\right\}^2\right]. \quad (16)$$

The random variable X has the same distribution as that of the unknown discrete variable, i.e., $\Pr(X=r_1)=p_1,\Pr(X=r_2)=p_2$, and Z is the standard Gaussian random variable independent of X. Since $\Psi(0)=0$, the sequence $\{\sigma_t^2\}_{t=0,1,\dots}$ with the recursion (15) converges to zero if $\Psi(\sigma^2)$ is concave and its derivative at $\sigma^2=0$ is smaller than one, i.e., $\left.\frac{\mathrm{d}\Psi}{\mathrm{d}(\sigma^2)}\right|_{\sigma^2\downarrow 0}<1$. In fact, the condition $\left.\frac{\mathrm{d}\Psi}{\mathrm{d}(\sigma^2)}\right|_{\sigma^2\downarrow 0}<1$ results in $\Psi(\sigma^2)<\sigma^2$ and hence we have $\sigma_{t+1}^2=\Psi(\sigma_t^2)<\sigma_t^2$. In this case, DAMP reconstructs the unknown vector $\boldsymbol{b}$ perfectly.

B. Condition for perfect reconstruction by DAMP

To obtain the condition for the perfect reconstruction, we evaluate $\left.\frac{\mathrm{d}\Psi}{\mathrm{d}(\sigma^2)}\right|_{\sigma\downarrow 0}$ analytically. By the direct calculation, $\left.\frac{\mathrm{d}\Psi}{\mathrm{d}(\sigma^2)}\right|_{\sigma\downarrow 0}$ can be obtained as

$$\begin{aligned}
&\left.\frac{\mathrm{d}\Psi}{\mathrm{d}(\sigma^2)}\right|_{\sigma\downarrow 0} \\
&= D(U_1, U_2, U_3) \qquad (17)\\
&:= \frac{p_1}{\Delta}\left\{U_1\phi(U_1) - U_2\phi(U_2)\right. \\
&\qquad \left. + \left(1+U_1^2\right)\Phi(U_1) + \left(1+U_2^2\right)(1-\Phi(U_2))\right\} \\
&\quad + \frac{p_2}{\Delta}\left\{U_2\phi(U_2) - U_3\phi(U_3)\right. \\
&\qquad \left. + \left(1+U_2^2\right)\Phi(U_2) + \left(1+U_3^2\right)(1-\Phi(U_3))\right\}, \qquad (18)
\end{aligned}$$

where $U_\ell = \tau Q_\ell$ $(\ell = 1, 2, 3)$. Here, $\phi(z) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right)$ and $\Phi(z) = \int_{-\infty}^{z}\phi(z')\mathrm{d}z'$ are the probability density function and the cumulative distribution function of the standard Gaussian distribution, respectively. Since we can choose any $q_1, q_2 \geq 0$ in (4) and $\tau \geq 0$, we can minimize (18) with respect to U_1, U_2, and U_3 as

$$D_{\min} = \min_{U_1,U_2,U_3} D(U_1, U_2, U_3) \text{ subject to } U_1 \leq U_2 \leq U_3. \qquad (19)$$

We can show that $U_1\phi(U_1) + (1+U_1^2)\Phi(U_1)$ in (18) is the monotone increasing function of U_1 and $-U_3\phi(U_3) + (1+U_3^2)(1-\Phi(U_3))$ is the monotone decreasing function of U_3. Hence, the optimal values of U_1 and U_3 are $U_1^{\mathrm{opt}} = -\infty$ and $U_3^{\mathrm{opt}} = \infty$, respectively. The minimization problem (19) can be rewritten as

$$\begin{aligned}
D_{\min} = \min_{U_2\in\mathbb{R}}\Bigg[&\frac{p_1}{\Delta}\left\{-U_2\phi(U_2) + \left(1+U_2^2\right)(1-\Phi(U_2))\right\} \\
&+\frac{p_2}{\Delta}\left\{U_2\phi(U_2) + \left(1+U_2^2\right)\Phi(U_2)\right\}\Bigg]. \qquad (20)
\end{aligned}$$

Since the objective function in (20) is the convex function of U_2, we can obtain the unique minimizer U_2^{opt} and the corresponding minimum value $D_{\min}$ of $\left.\frac{\mathrm{d}\Psi}{\mathrm{d}(\sigma^2)}\right|_{\sigma\downarrow 0}$. From (12), the soft thresholding function with the threshold $c = \tau\frac{\sigma}{\sqrt{\Delta}}$ and the optimal parameters of $U_\ell = \tau Q_\ell$ $(\ell = 1, 2, 3)$ is written as

$$\eta^{\mathrm{S}}(u) = \begin{cases} r_1 & \left(u < r_1 + \frac{\sigma}{\sqrt{\Delta}}U_2^{\mathrm{opt}}\right) \\ u - \frac{\sigma}{\sqrt{\Delta}}U_2^{\mathrm{opt}} & \left(r_1 + \frac{\sigma}{\sqrt{\Delta}}U_2^{\mathrm{opt}} \leq u < r_2 + \frac{\sigma}{\sqrt{\Delta}}U_2^{\mathrm{opt}}\right) \\ r_2 & \left(r_2 + \frac{\sigma}{\sqrt{\Delta}}U_2^{\mathrm{opt}} \leq u\right) \end{cases}. \qquad (21)$$

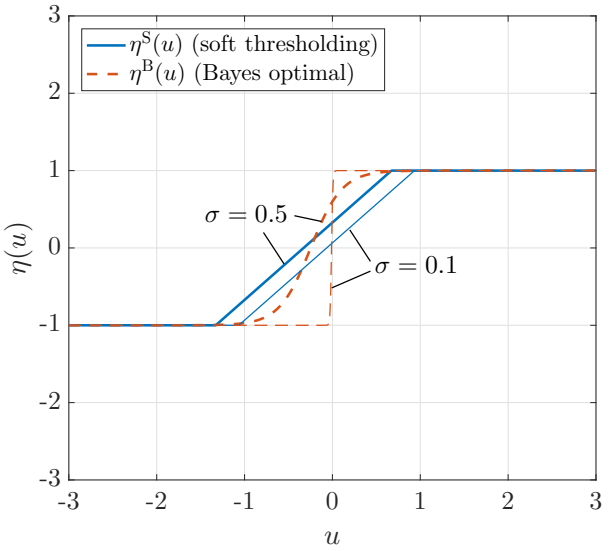

Fig. 2. Examples of $\eta^{\mathrm{S}}(u)$ and $\eta^{\mathrm{B}}(u)$ $(r_1 = -1, r_2 = 1, p_1 = 0.2, p_2 = 0.8, \Delta = 0.7,$ and $\sigma = 0.1, 0.5)$

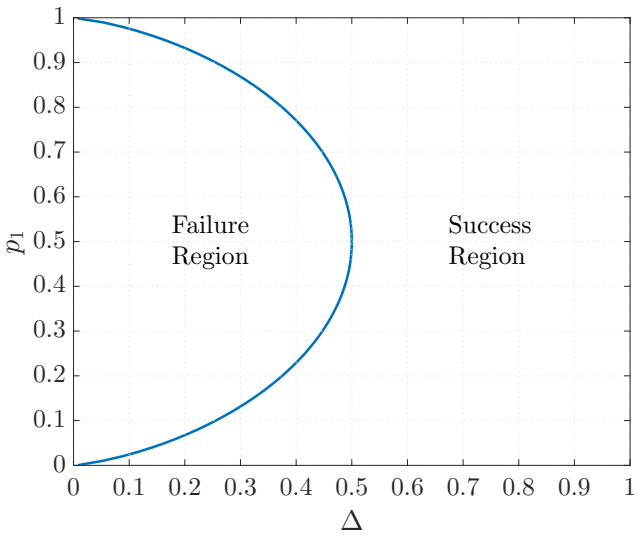

Fig. 3. Phase transition of DAMP with the soft thresholding function $\eta^{\mathrm{S}}(u)$ for binary vector

Fig. 2 shows an example of the soft thresholding function $\eta^{\mathrm{S}}(u)$ in the case with $r_1 = -1, r_2 = 1, p_1 = 0.2, p_2 = 0.8, \Delta = 0.7,$ and $\sigma = 0.1, 0.5$.

The DAMP algorithm with $\eta^{\mathrm{S}}(u)$ provides the perfect reconstruction in the large system limit if $D_{\min} < 1$. Fig. 3 shows the phase transition line of DAMP, where $D_{\min} = 1$. Note that the line is the boundary between the success and failure regions of DAMP in the large system limit. In the left region of the curve, the MSE of the estimate obtained by DAMP does not converge to zero. In the right region, DAMP can provide the perfect reconstruction of $\boldsymbol{b}$. For example, the figure shows that DAMP requires at least $N/2$ observations to accurately

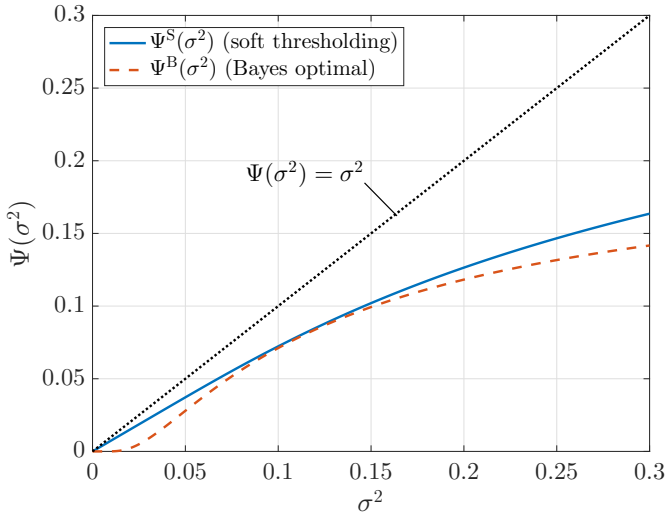

Fig. 4. Examples of $\Psi^{\mathrm{S}}(\sigma^2)$ and $\Psi^{\mathrm{B}}(\sigma^2)$ ($r_1=0, r_2=1, p_1=0.7, p_2=0.3$, and $\Delta=0.6$)

reconstruct an N-dimensional uniformly distributed binary vector. This result coincides with the theoretical analysis for the regularization-based method and the transform-based method [9]. Moreover, our analysis also provides the required number of measurements for the asymmetric distribution with $p_1 \neq 0.5$, which has not been obtained in [9]. It should be noted that $D_{\min}$ in (20) is independent of r_1 and r_2, and hence the phase transition line in Fig. 3 is identical for any r_1 and r_2.

We confirm that the function $\Psi(\sigma^2)$ in (15) is concave when we use the optimal soft thresholding function $\eta^{\mathrm{S}}(u)$. The second derivative of $\Psi^{\mathrm{S}}(\sigma^2)=\mathrm{E}\left[\left\{\eta^{\mathrm{S}}\left(X+\frac{\sigma}{\sqrt{\Delta}}Z\right)-X\right\}^2\right]$ can be obtained as

$$\frac{\mathrm{d}^2\Psi^{\mathrm{S}}}{\mathrm{d}(\sigma^2)^2} = \frac{\sqrt{\Delta}}{2\sigma^5}(r_1-r_2)^3\left\{p_1\phi\left(\frac{\sqrt{\Delta}}{\sigma}(-r_1+r_2)+U_2^{\mathrm{opt}}\right) + p_2\phi\left(\frac{\sqrt{\Delta}}{\sigma}(-r_2+r_1)+U_2^{\mathrm{opt}}\right)\right\}, \tag{22}$$

which is negative because $r_1 < r_2$, and hence $\Psi^{\mathrm{S}}(\sigma^2)$ is concave. Fig. 4 shows an example of $\Psi^{\mathrm{S}}(\sigma^2)$ in the case with $r_1=0, r_2=1, p_1=0.7, p_2=0.3$, and $\Delta=0.6$. We can see that $\Psi^{\mathrm{S}}(\sigma^2)$ is concave and $\Psi^{\mathrm{S}}(\sigma^2)\leq\sigma^2$ for all $\sigma^2\geq 0$.

C. Bayes optimal DAMP

In the DAMP algorithm (13), (14), we can use another function as η instead of the soft thresholding function (12), (21). The state evolution formula (15) can also be used for Lipschitz continuous functions other than the soft thresholding function. In the literature of compressed sensing, the AMP algorithm is called *Bayes optimal* if the function $\tilde{\eta}$ minimizing

$$\tilde{\Psi}(\sigma^2)=\mathrm{E}\left[\left\{\tilde{\eta}\left(X+\frac{\sigma}{\sqrt{\Delta}}Z\right)-X\right\}^2\right] \tag{23}$$

is used [20], [21]. Note that the minimizer $\eta^{\mathrm{B}}(u)$ of (23) can be written as the conditional expectation [16], i.e.,

$$\eta^{\mathrm{B}}(u)=\mathrm{E}\left[X\,\middle|\,X+\frac{\sigma}{\sqrt{\Delta}}Z=u\right]. \tag{24}$$

Although it is difficult in general to analytically calculate the optimal function $\eta^{\mathrm{B}}(u)$, we can obtain $\eta^{\mathrm{B}}(u)$ for the Bayes optimal DAMP because the distribution of X is discrete in our problem. We consider $\tilde{\Psi}(\sigma^2)$ as a functional of the function $\tilde{\eta}(u)$ and rewrite (23) as

$$\tilde{\Psi}(\sigma^2) = \frac{\sqrt{\Delta}}{\sigma}\left[p_1\int_{-\infty}^{\infty}\left\{\tilde{\eta}(u)-r_1\right\}^2\phi\left(\frac{\sqrt{\Delta}}{\sigma}(u-r_1)\right)\mathrm{d}u + p_2\int_{-\infty}^{\infty}\left\{\tilde{\eta}(u)-r_2\right\}^2\phi\left(\frac{\sqrt{\Delta}}{\sigma}(u-r_2)\right)\mathrm{d}u\right]. \tag{25}$$

The variation is given by

$$\frac{\mathrm{d}\tilde{\Psi}}{\mathrm{d}\tilde{\eta}} = \frac{2\sqrt{\Delta}}{\sigma}\left[p_1\left\{\tilde{\eta}(u)-r_1\right\}\phi\left(\frac{\sqrt{\Delta}}{\sigma}(u-r_1)\right) + p_2\left\{\tilde{\eta}(u)-r_2\right\}\phi\left(\frac{\sqrt{\Delta}}{\sigma}(u-r_2)\right)\right]. \tag{26}$$

By solving $\frac{\mathrm{d}\tilde{\Psi}}{\mathrm{d}\tilde{\eta}}=0$, the optimal function $\eta^{\mathrm{B}}(u)$ can be obtained as

$$\eta^{\mathrm{B}}(u)=\frac{p_1 r_1\phi\left(\frac{\sqrt{\Delta}}{\sigma}(u-r_1)\right)+p_2 r_2\phi\left(\frac{\sqrt{\Delta}}{\sigma}(u-r_2)\right)}{p_1\phi\left(\frac{\sqrt{\Delta}}{\sigma}(u-r_1)\right)+p_2\phi\left(\frac{\sqrt{\Delta}}{\sigma}(u-r_2)\right)}. \tag{27}$$

As a special case, when $r_1=-1, r_2=1$ and $p_1=p_2=0.5$, (27) can be reduced to

$$\eta^{\mathrm{B}}(u)=\tanh\left(\frac{\Delta}{\sigma^2}u\right), \tag{28}$$

which has been proposed for the code division multiple access (CDMA) multiuser detection [2], [22]. In Fig. 2, we show an example of $\eta^{\mathrm{B}}(u)$ as well as $\eta^{\mathrm{S}}(u)$ in the case with $r_1=-1, r_2=1, p_1=0.2, p_2=0.8, \Delta=0.7$, and $\sigma=0.1, 0.5$. We can see that the form of $\eta^{\mathrm{B}}(u)$ changes drastically against σ compared to $\eta^{\mathrm{S}}(u)$. We also show an example of $\Psi^{\mathrm{B}}(\sigma^2)=\mathrm{E}\left[\left\{\eta^{\mathrm{B}}\left(X+\frac{\sigma}{\sqrt{\Delta}}Z\right)-X\right\}^2\right]$ in Fig. 4 as well as $\Psi^{\mathrm{S}}(\sigma^2)$. The sequence of the MSE $\{\sigma_t^2\}_{t=0,1,\ldots}$ obtained by $\sigma_{t+1}^2=\Psi^{\mathrm{B}}(\sigma_t^2)$ converges to zero if $D_{\min}<1$, because $\sigma_{t+1}^2=\Psi^{\mathrm{B}}(\sigma_t^2)\leq\Psi^{\mathrm{S}}(\sigma_t^2)<\sigma_t^2$ in that case. However, since $\Psi^{\mathrm{B}}(\sigma^2)$ is not concave unlike $\Psi^{\mathrm{S}}(\sigma^2)$, it is difficult to obtain the necessary condition analytically for the perfect reconstruction by Bayes optimal DMAP.

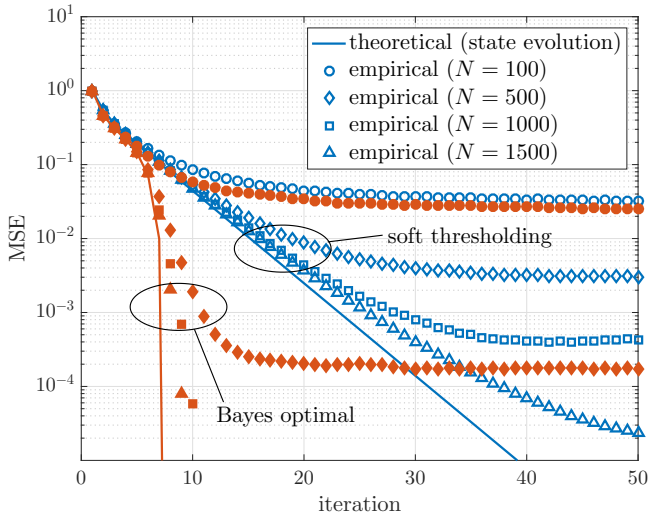

Fig. 5. State evolution and empirical performance ($r_1 = -1, r_2 = 1, p_1 = 0.2, p_2 = 0.8$, and $\Delta = 0.5$)

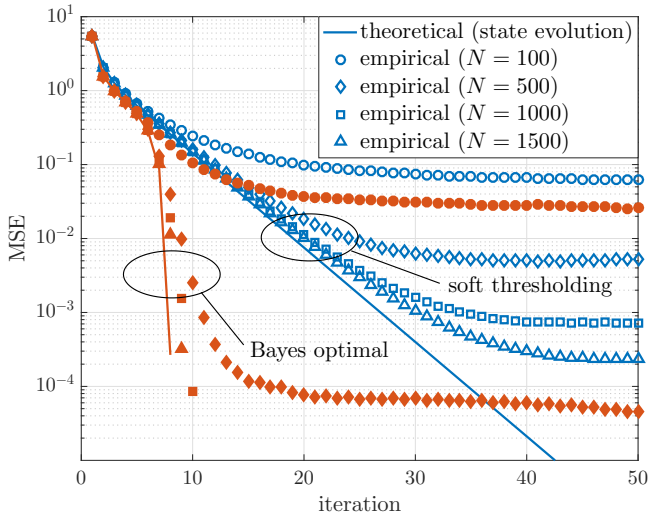

Fig. 6. State evolution and empirical performance ($r_1 = 1, r_2 = 4, p_1 = 0.7, p_2 = 0.3$, and $\Delta = 0.6$)

IV. Simulation Results

In this section, we evaluate the performance of the proposed DAMP via computer simulations. In the simulations, the sensing matrix $\boldsymbol{A} \in \mathbb{R}^{M \times N}$ is composed of i.i.d. Gaussian variables with zero mean and variance $1/M$. The initialization of the algorithm is given by $\boldsymbol{x}^{-1} = \boldsymbol{x}^0 = \boldsymbol{0}, \boldsymbol{z}^{-1} = \boldsymbol{0}$.

Figs. 5 and 6 show the prediction via state evolution and the empirical MSE $\sigma_t^2 = \|\boldsymbol{x}^t - \boldsymbol{b}\|_2^2/N$ with DAMP obtained by simulations. We set $r_1 = -1, r_2 = 1, p_1 = 0.2, p_2 = 0.8, \Delta = 0.5$ in Fig. 5 and $r_1 = 1, r_2 = 4, p_1 = 0.7, p_2 = 0.3, \Delta = 0.6$ in Fig. 6. We evaluate the performance for the different problem sizes of $N = 100, 500, 1000$, and 1500. In the figures, "soft thresholding" denotes the performance of DAMP with the soft thresholding function $\eta^{\mathrm{S}}(u)$ and "Bayes optimal" denotes that of Bayes optimal DAMP with $\eta^{\mathrm{B}}(u)$. We can see that Bayes optimal DAMP has much smaller MSE with less number of iterations than DAMP with the soft thresholding function. The figures also show that the prediction with state evolution is close to the empirical performance in the large-scale systems.

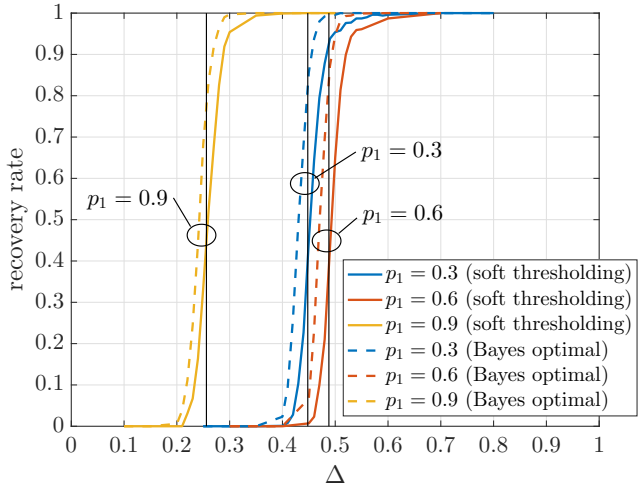

Fig. 7. Recovery rate ($\boldsymbol{b} \in \{0,1\}^N, N = 1000$)

In Fig. 7, we empirically evaluate the rate of the success recovery in the sense that $\sigma_t^2 < 10^{-3}$ after $t = 500$ iterations. The unknown vector is $\boldsymbol{b} \in \{0,1\}^N$ and $N = 1000$. The vertical lines for each p_1 correspond to the required Δ obtained from Fig. 3. In the large system limit, the left side of each vertical line is the failure region and the right side is the success region. For all p_1, the recovery rate of DAMP with soft thresholding rapidly increases around the corresponding vertical line. We can also see that the recovery rate of Bayes optimal DAMP is better than DAMP with soft thresholding. It should be noted that the recovery rate is not equal to one in the success region near the boundary. One reason is that we restrict the maximum number of iterations as $t = 500$. Another reason is that the problem size here is finite and not large enough.

V. Conclusion

In this paper, we have proposed the algorithm for the binary vector reconstruction, referred to as DAMP. We have analytically evaluated the performance of DAMP and have derived the condition for the perfect reconstruction with DAMP via state evolution. Moreover, we have also provided Bayes optimal DAMP, which gives much smaller MSE compared to DAMP with the soft thresholding function. Numerical results show that DAMP can reconstruct the binary vector from its underdetermined linear measurements and the empirical performance agrees well with our theoretical results. Future work includes the extension of DAMP for the noisy observation cases and its application to various communication systems.

REFERENCES

[1] S. Verdú, *Multiuser detection*, Cambridge University Press, 1998.

[2] Y. Kabashima, "A CDMA multiuser detection algorithm on the basis of belief propagation," *J. Phys. A*, vol. 36, pp. 11111–11121, Oct. 2003.

[3] H. Zhu and G. B. Giannakis, "Exploiting sparse user activity in multiuser detection," *IEEE Trans. Commun.*, vol. 59, no. 2, pp. 454–465, Feb. 2011.

[4] H. Sasahara, K. Hayashi, and M. Nagahara, "Multiuser detection by MAP estimation with sum-of-absolute-values relaxation," in *Proc. IEEE ICC 2016*, May 2016.

[5] H. Sasahara, K. Hayashi, and M. Nagahara, "Multiuser detection based on MAP estimation with sum-of-absolute-values relaxation," *IEEE Trans. Signal Process*, vol. 65, no. 21, pp. 5621–5634, Nov. 2017.

[6] R. Hayakawa, K. Hayashi, H. Sasahara and M. Nagahara, "Massive overloaded MIMO signal detection via convex optimization with proximal splitting," in *Proc. EUSIPCO 2016*, pp. 1383–1387, Aug.–Sept. 2016.

[7] J. E. Mazo, "Faster-than-Nyquist signaling," *Bell System Tech. J.*, vol. 54, no. 8, pp. 1451–1462, Oct. 1975.

[8] H. Sasahara, K. Hayashi, and M. Nagahara, "Faster-than-Nyquist signaling by sum-of-absolute-values optimization," in *Proc. SICE ISCS 2016*, pp. 7–8, Mar. 2016.

[9] A. A-El-Bey, D. Pastor, S. M. A. Sbaï, and Y. Fadlallah, "Sparsity-based recovery of finite alphabet solutions to underdetermined linear systems," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 2008–2018, Apr. 2015.

[10] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[11] K. Hayashi, M. Nagahara, and T. Tanaka, "A user's guide to compressed sensing for communications systems," *IEICE Trans. Commun.*, vol. E96-B, no. 3, pp. 685–712, Mar. 2013.

[12] M. Nagahara, "Discrete signal reconstruction by sum of absolute values," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp.1575–1579, Oct. 2015.

[13] R. Hayakawa and K. Hayashi, "Discreteness-aware AMP for reconstruction of symmetrically distributed discrete variables," in *Proc. IEEE SPAWC 2017*, July 2017.

[14] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci.*, vol. 106, no. 45, pp. 18914–18919, Nov. 2009.

[15] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing: I. motivation and construction," in *Proc. IEEE Inf. Theory Workshop*, pp. 1–5, Jan. 2010.

[16] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 764–785, Feb. 2011.

[17] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*, Morgan Kaufmann Publishers Inc., 1988.

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

[19] P. Combettes and J. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, ser. Springer Optimization and Its Applications. Springer New York, vol. 49, pp. 185–212, 2011.

[20] D. L. Donoho, A. Maleki, and A. Montanari, "The noise-sensitivity phase transition in compressed sensing," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6920–6941, Oct. 2011.

[21] K. Mimura, "On introducing damping to Bayes optimal approximate message passing for compressed sensing," in *Proc. APSIPA ASC 2015*, pp. 659–662, Dec. 2015.

[22] A. Montanari and D. Tse, "Analysis of belief propagation for non-linear problems: The example of CDMA (or: how to prove Tanaka's formula)," in *Proc. IEEE Inf. Theory Workshop*, Mar. 2006.