Discreteness-Aware AMP for Reconstruction of Symmetrically Distributed Discrete Variables

Ryo Hayakawa
Graduate School of Informatics, Kyoto University,
Kyoto 606-8501, Japan
E-mail: rhayakawa@sys.i.kyoto-u.ac.jp

Kazunori Hayashi
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 a message passing-based algorithm to reconstruct a discrete-valued vector whose elements have a symmetric probability distribution. The proposed algorithm, referred to as discreteness-aware approximate message passing (DAMP), borrows the idea of the approximate message passing (AMP) algorithm for compressed sensing. We analytically evaluate the performance of DAMP via state evolution framework to derive a required number of linear measurements for the exact reconstruction with DAMP. The analysis also provides the optimal parameter minimizing the required number of measurements. Simulation results show that DAMP can reconstruct the discrete-valued vector and its performance agrees well with our theoretical results via the state evolution.**

***Index Terms*—approximate message passing, state evolution, sum-of-absolute-value optimization, discrete-valued vector reconstruction.**

I. Introduction

Many problems in signal processing can be regarded as the reconstruction of a *discrete-valued* vector from its possibly underdetermined linear measurements. For example, since we generally use discrete-valued signals in digital communications systems, there are many potential applications of the discrete-valued vector reconstruction, including multiuser detection in machine-to-machine communications [1], signal detection for massive overloaded multiple-input multiple-output (MIMO) systems [2], and faster-than-Nyquist signaling [3]. The optimal maximum likelihood (ML) method for the problem becomes a combinatorial optimization problem, whose computational complexity increases exponentially along the problem size. Since the complexity is prohibitive especially in the large-scale problem, a low-complexity algorithm is required for the reconstruction. Although there are several methods to reconstruct the discrete-valued *sparse* vector [4], [5], they might not be suitable for the non-sparse discrete-valued vector reconstruction. In fact, the performance for discrete-valued dense vectors has not been evaluated in [4] and [5].

For the discrete-valued vector reconstruction, the regularization-based method and the transform-based method have been proposed [6]. These two methods borrow the idea of compressed sensing [7], [8] to formulate the reconstruction problem as convex optimization, which can be solved with interior point methods [9]. The required number of measurements is analytically derived for the binary vector reconstruction using the regularization-based method. For multi-valued vectors other than the binary vector, however, no theoretical results have been shown for the method. Although a more general theorem has been theoretically provided for the transform-based method, it is assumed in the proof that the unknown vector is uniformly distributed. On the other hand, in [10], sum-of-absolute-value (SOAV) optimization similar to the regularization-based method has been proposed for the discrete-valued vector reconstruction. Since the SOAV optimization for uniform distributions is equivalent to the regularization-based method, its required number of measurements for the perfect reconstruction has been provided for the case of a uniformly distributed binary vector. For general discrete distributions, however, the required number of measurements for the SOAV optimization has not been obtained.

In this paper, using the idea of the SOAV optimization, we propose an algorithm for the reconstruction of discrete-valued vectors whose elements have a symmetric probability distribution. The proposed algorithm, referred to as discreteness-aware approximate message passing (DAMP), is derived with the similar approach to that of the approximate message passing (AMP) algorithm [11], [12], which has been originally proposed for compressed sensing. The DAMP algorithm requires a lower computational complexity than interior point methods. We also analytically evaluate the performance of DAMP via the state evolution framework [11], [13]. In the analysis, we derive a required number of measurements for the exact reconstruction and obtain the parameter of the DAMP algorithm minimizing the number of measurements. Moreover, we can analytically show that, if the unknown vector is discrete-valued, DAMP can correctly reconstruct the vector with a smaller number of measurements than the conventional AMP algorithm. Furthermore, unlike the conventional transform-based method, the analysis for DAMP can be applied to non-uniform distributions. Simulation results show that the proposed DAMP algorithm can reconstruct the discrete-valued vector from its underdetermined linear measurements and the performance can be well predicted with the theoretical analysis.

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.

978-1-5090-3009-5/17/$31.00 ©2017 IEEE

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 denote 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 [14] of h is defined as

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

The jth element of $\mathrm{sgn}(\boldsymbol{v})$ is the sign of v_j.

II. Proposed DAMP Algorithm

In this section, we shortly review the SOAV optimization [10] and describe the proposed DAMP algorithm based on the optimization problem.

A. SOAV optimization

The SOAV optimization is a technique to reconstruct a discrete-valued vector $\boldsymbol{b} = [b_1 \cdots b_N]^{\mathrm{T}} \in \mathbb{R}^N$ from its linear measurements $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{b} \in \mathbb{R}^M$, where $\boldsymbol{A} \in \mathbb{R}^{M\times N}$. In this paper, we assume $\boldsymbol{b} \in \{r_0, \pm r_1, \ldots, \pm r_L\}^N$ ($r_L > r_{L-1} > \cdots > r_1 > r_0 = 0$) and symmetric distribution $\Pr(b_j = r_\ell) = \Pr(b_j = -r_\ell) = p_\ell$, $\Pr(b_j = 0) = p_0$ ($j = 1,\ldots,N$ and $\ell = 1,\ldots,L$), where $b_1,\ldots,b_N$ are independent, $p_0 \geq 0$, $p_1,\ldots,p_L > 0$ and $p_0 + 2\sum_{\ell=1}^L p_\ell = 1$. The SOAV optimization to estimate $\boldsymbol{b}$ can be written as

$$\hat{\boldsymbol{b}} = \underset{\boldsymbol{s}\in\mathbb{R}^N}{\arg\min} \left\{ q_0\|\boldsymbol{s}\|_1 + \sum_{\ell=1}^L q_\ell \left(\|\boldsymbol{s} - r_\ell\mathbf{1}\|_1 + \|\boldsymbol{s} + r_\ell\mathbf{1}\|_1\right) \right\} \quad \text{subject to} \quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{s}, \tag{2}$$

which uses the idea of ℓ_1 optimization [7] and the fact that $\boldsymbol{b} - r_\ell\mathbf{1}$ has approximately $p_\ell N$ zero elements. The coefficients $q_0,\ldots,q_L \geq 0$ are fixed as $q_\ell = p_\ell$ ($\ell = 0,\ldots,L$) in the original SOAV optimization [10]. The regularization-based method [6] is also a special case of the problem (2) because it is obtained by setting $q_0 = \cdots = q_L = 1$. However, it is not clear whether these selections of the coefficients are appropriate or not. We thus consider them as parameters to be optimized before solving (2), as described in Sec. III. Although we consider only the noise-free case in this paper, we can straightforwardly extend DAMP for the noisy case as in the case of compressed sensing [12].

B. DAMP algorithm

The proposed DAMP algorithm for (2) can be derived in a similar approach to that of the AMP algorithm for compressed sensing [11]. In what follows, we assume that the elements of the sensing matrix $\boldsymbol{A}$ are i.i.d. with zero mean and variance $1/M$. In the large system limit ($M, N \to \infty$ with $M/N = \Delta$), a parameterized DAMP algorithm is written as

$$\boldsymbol{x}^{t+1} = \eta(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{z}^t + \boldsymbol{x}^t, \lambda\sigma_t), \tag{3}$$

$$\boldsymbol{z}^t = \boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}^t + \frac{1}{\Delta}\boldsymbol{z}^{t-1}\langle \eta'(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{z}^{t-1} + \boldsymbol{x}^{t-1}, \lambda\sigma_{t-1})\rangle, \tag{4}$$

where $\boldsymbol{x}^t$ is the estimate of $\boldsymbol{b}$ in the tth iteration, $\lambda \geq 0$ is the parameter, and $\sigma_t^2 = \|\boldsymbol{x}^t - \boldsymbol{b}\|_2^2/N$ is the mean squared error (MSE) of $\boldsymbol{x}^t$. Since the true solution $\boldsymbol{b}$ is unknown in practice, we use the alternative value for σ_t^2 calculated from the current estimate alone, e.g., $\hat{\sigma}_t^2 = \|\boldsymbol{x}^t - \mathcal{R}(\boldsymbol{x}^t)\|_2^2/N$, where $\mathcal{R}$ maps each element of $\boldsymbol{x}^t$ into the closest element in $\{r_0, \pm r_1, \ldots, \pm r_L\}$. The function η is the soft thresholding function given by $\eta(\boldsymbol{u}, \lambda\sigma) = \mathrm{prox}_{\lambda\sigma J}(\boldsymbol{u}) := \underset{\boldsymbol{s}\in\mathbb{R}^N}{\arg\min}\{\lambda\sigma J(\boldsymbol{s}) + \frac{1}{2}\|\boldsymbol{s}-\boldsymbol{u}\|_2^2\}$, where $J(\boldsymbol{s}) = q_0\|\boldsymbol{s}\|_1 + \sum_{\ell=1}^L q_\ell\left(\|\boldsymbol{s} - r_\ell\mathbf{1}\|_1 + \|\boldsymbol{s} + r_\ell\mathbf{1}\|_1\right)$ is the objective function in (2). From the proposition in [1], the jth element of $\mathrm{prox}_{\lambda\sigma J}(\boldsymbol{u})$ can be written as

$$[\mathrm{prox}_{\lambda\sigma J}(\boldsymbol{u})]_j = \begin{cases} u_j + Q_L\lambda\sigma & (u_j < -r_L - Q_L\lambda\sigma) \\ \vdots & \vdots \\ -r_k & (-r_k - Q_k\lambda\sigma \leq u_j < -r_k - Q_{k-1}\lambda\sigma) \\ u_j + Q_{k-1}\lambda\sigma & (-r_k - Q_{k-1}\lambda\sigma \leq u_j < -r_{k-1} - Q_{k-1}\lambda\sigma) \\ \vdots & \vdots \\ 0 & (-Q_0\lambda\sigma \leq u_j < Q_0\lambda\sigma) \\ \vdots & \vdots \\ u_j - Q_{k-1}\lambda\sigma & (r_{k-1} + Q_{k-1}\lambda\sigma \leq u_j < r_k + Q_{k-1}\lambda\sigma) \\ r_k & (r_k + Q_{k-1}\lambda\sigma \leq u_j < r_k + Q_k\lambda\sigma) \\ \vdots & \vdots \\ u_j - Q_L\lambda\sigma & (r_L + Q_L\lambda\sigma \leq u_j) \end{cases}, \tag{5}$$

where $Q_0 = q_0$ and $Q_k = q_0 + 2\sum_{\ell=1}^k q_\ell$ ($k = 1,\ldots,L$). Note that $[\mathrm{prox}_{\lambda\sigma J}(\boldsymbol{u})]_j$ is a function of only u_j and hence the soft thresholding function η is a element-wise function. Figure 1 shows an example of the soft thresholding function in the case with $L = 1$ and $r_1 = 1$, where $[\eta(\boldsymbol{u}, \lambda\sigma)]_j$ denotes the jth element of $\eta(\boldsymbol{u}, \lambda\sigma)$. The jth element of $\eta'(\boldsymbol{u}, \lambda\sigma)$ is defined as $[\eta'(\boldsymbol{u}, \lambda\sigma)]_j = 0$ if $[\eta(\boldsymbol{u}, \lambda\sigma)]_j \in \{r_0, \pm r_1 \ldots, \pm r_L\}$, otherwise $[\eta'(\boldsymbol{u}, \lambda\sigma)]_j = 1$.

The update equations of DAMP (3), (4) are basically the same as those of the AMP algorithm for compressed sensing [11]. The only difference is the soft thresholding function η, which is $[\eta(\boldsymbol{u}, \lambda\sigma)]_j = \mathrm{sgn}(u_j)\max\{|u_j| - \lambda\sigma, 0\}$ for the reconstruction of the sparse vector. Since the equations (3) and (4) can be calculated only with the addition of the vectors and the multiplication of the matrix and the vector, the computational complexity in each iteration is $O(MN)$. It is lower than the complexity of the interior point method used in the conventional regularization-based method and transform-based method, i.e., $O(MN^2)$ and $O((2L+1)^2(M+N)N^2)$ per iteration [6], respectively.

III. State Evolution for DAMP

State evolution [11], [13] is a framework to analyze the performance of the AMP algorithm. Similarly to the case of compressed sensing, the state evolution formula for the

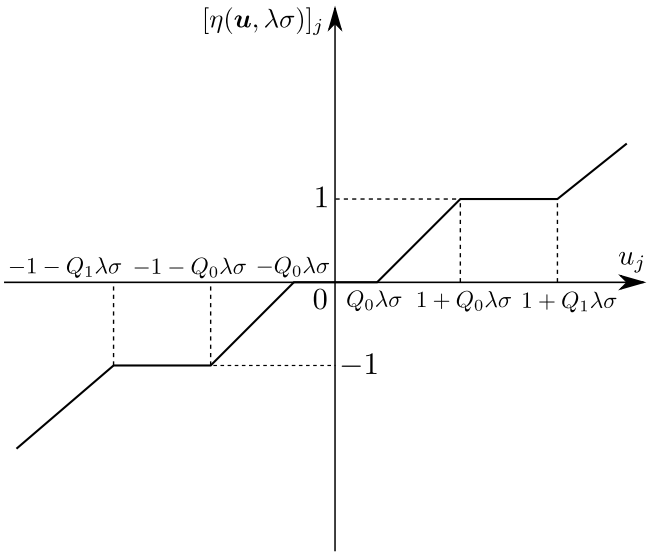

Fig. 1. Example of soft thresholding function $\eta(\boldsymbol{u},\lambda\sigma)$ $(L=1, r_1=1)$

proposed DAMP (3), (4) is given by

$$\sigma_{t+1}^2 = \Psi(\sigma_t^2), \tag{6}$$

where

$$\Psi(\sigma^2) = \mathrm{E}\left[\left\{\eta\left(X+\frac{\sigma}{\sqrt{\Delta}}Z,\lambda\sigma\right)-X\right\}^2\right]. \tag{7}$$

The random variable X has the same distribution as that of the unknown discrete variable, i.e., $\Pr(X=r_\ell)=\Pr(X=-r_\ell)=p_\ell$, $\Pr(X=0)=p_0$ $(\ell=1,\ldots,L)$. Z is the standard Gaussian random variable independent of X.

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 the state evolution (6). Since $\Psi(0)=0$, the sequence $\{\sigma_t^2\}_{t=0,1,\ldots}$ with the recursion (6) converges to zero if $\Psi(\sigma^2)$ is concave and the derivative of $\Psi(\sigma^2)$ at $\sigma^2=0$ is smaller than one, i.e., $\left.\frac{\mathrm{d}\Psi}{\mathrm{d}(\sigma^2)}\right|_{\sigma\downarrow 0}<1$. In fact, these conditions result in $\Psi(\sigma^2)<\sigma^2$ and hence we have $\sigma_{t+1}^2=\Psi(\sigma_t^2)<\sigma_t^2$.

Since we consider the coefficients $q_0,\ldots,q_L\geq 0$ in (2) as the parameters, all of the parameters in DAMP are λ and $Q_0,\ldots,Q_L$ in the soft thresholding function (5). Hence, assuming $Q_L=q_0+2\sum_{\ell=1}^L q_\ell=1$ without loss of generality, we minimize $\left.\frac{\mathrm{d}\Psi}{\mathrm{d}(\sigma^2)}\right|_{\sigma\downarrow 0}$ as

$$D_{\min}=\min_{\lambda,Q_0,\ldots,Q_L}\left.\frac{\mathrm{d}\Psi}{\mathrm{d}(\sigma^2)}\right|_{\sigma\downarrow 0} \quad \text{subject to } \lambda\geq 0,\ 0\leq Q_0\leq\cdots\leq Q_L=1 \tag{8}$$

and use the parameters corresponding to the minimum value $D_{\min}$. If $D_{\min}<1$ holds, the MSE $\{\sigma_t^2\}_{t=0,1,\ldots}$ obtained by DAMP with such parameters converges to zero and DAMP reconstructs the unknown vector $\boldsymbol{b}$ perfectly. To calculate $\left.\frac{\mathrm{d}\Psi}{\mathrm{d}(\sigma^2)}\right|_{\sigma\downarrow 0}$, we firstly obtain

$$\Psi(\sigma^2)=p_0\Psi_0(\sigma^2)+2\sum_{\ell=1}^L p_\ell\Psi_\ell(\sigma^2) \tag{9}$$

from (7), where

$$\Psi_\ell(\sigma^2)=\mathrm{E}\left[\left\{\eta\left(r_\ell+\frac{\sigma}{\sqrt{\Delta}}Z,\lambda\sigma\right)-r_\ell\right\}^2\right] \tag{10}$$

$$=\int_{-\infty}^{\infty}\left\{\eta\left(r_\ell+\frac{\sigma}{\sqrt{\Delta}}z,\lambda\sigma\right)-r_\ell\right\}^2\phi(z)\,\mathrm{d}z, \tag{11}$$

and $\phi(z)=\left(1/\sqrt{2\pi}\right)\exp\left(-z^2/2\right)$ is the probability distribution function of the standard Gaussian distribution. From (9), we have

$$\left.\frac{\mathrm{d}\Psi}{\mathrm{d}(\sigma^2)}\right|_{\sigma\downarrow 0}=p_0\left.\frac{\mathrm{d}\Psi_0}{\mathrm{d}(\sigma^2)}\right|_{\sigma\downarrow 0}+2\sum_{\ell=1}^L p_\ell\left.\frac{\mathrm{d}\Psi_\ell}{\mathrm{d}(\sigma^2)}\right|_{\sigma\downarrow 0}. \tag{12}$$

With the direct calculation of (11), we can obtain

$$\left.\frac{\mathrm{d}\Psi_\ell}{\mathrm{d}(\sigma^2)}\right|_{\sigma\downarrow 0}=D_\ell(\boldsymbol{U}) \tag{13}$$

$$:=\frac{1}{\Delta}\Big\{U_{\ell-1}\phi\left(U_{\ell-1}\right)-U_\ell\phi\left(U_\ell\right)+\left(1+U_{\ell-1}^2\right)\Phi\left(U_{\ell-1}\right)+\left(1+U_\ell^2\right)\left(1-\Phi\left(U_\ell\right)\right)\Big\}, \tag{14}$$

where $\boldsymbol{U}=[U_0\ \cdots\ U_L]^{\mathrm{T}}$, $U_\ell=Q_\ell\lambda\sqrt{\Delta}$, $Q_{-1}=-Q_0$, and $\Phi(z)=\int_{-\infty}^z\phi(z')\mathrm{d}z'$ is the cumulative distribution function of the standard Gaussian distribution. Let us define $D(\boldsymbol{U}):=\left.\frac{\mathrm{d}\Psi}{\mathrm{d}(\sigma^2)}\right|_{\sigma\downarrow 0}$ and rewrite (12) as

$$D(\boldsymbol{U})=p_0D_0(\boldsymbol{U})+2\sum_{\ell=1}^L p_\ell D_\ell(\boldsymbol{U}). \tag{15}$$

Finally, the problem (8) can be rewritten as

$$D_{\min}=\min_{\boldsymbol{U}}D(\boldsymbol{U}) \text{ subject to } 0\leq U_0\leq\cdots\leq U_L, \tag{16}$$

which can be solved via interior point methods [9] because the Hessian of $D(\boldsymbol{U})$ is positive semi-definite and hence $D(\boldsymbol{U})$ is a convex function of $\boldsymbol{U}$. By solving (16), we can obtain the soft thresholding function η minimizing $\left.\frac{\mathrm{d}\Psi}{\mathrm{d}(\sigma^2)}\right|_{\sigma\downarrow 0}$. Note that the above discussion assumes the large system limit and hence the condition $D_{\min}<1$ is not necessarily sufficient for the finite-scale problem.

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 $1/M$. As a discrete-valued vector $\boldsymbol{b}$, we consider two cases of $b_j\in\{0,\pm 1\}$ and $b_j\in\{0,\pm 1,\pm 3\}$. For simplicity, we assume that each nonzero

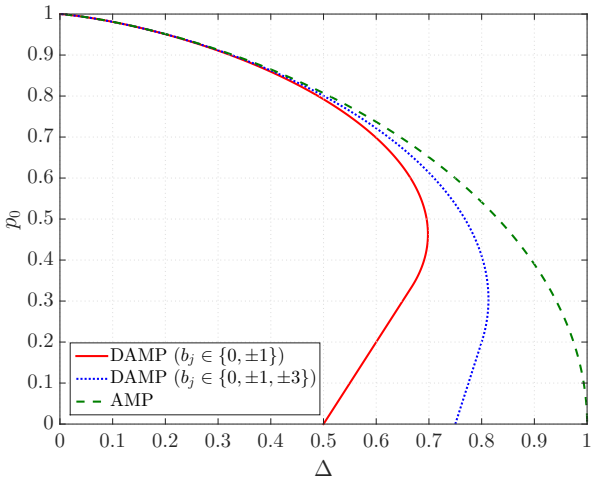

Fig. 2. Phase transition of DAMP and AMP

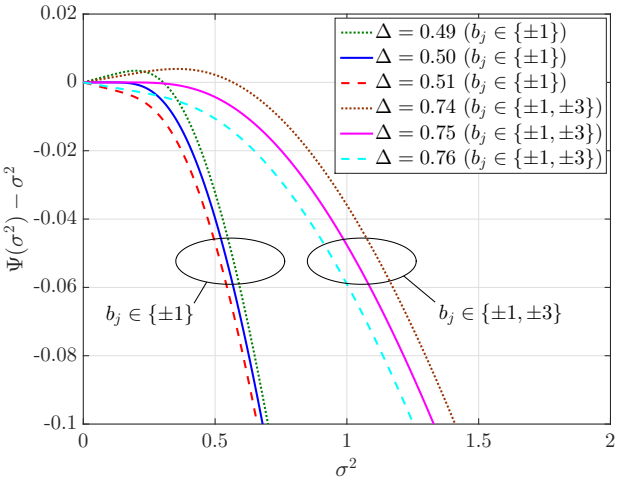

Fig. 3. $\Psi(\sigma^2) - \sigma^2$

element is drawn uniformly from $\{\pm 1\}$ and $\{\pm 1, \pm 3\}$, respectively. The initialization of DAMP in the simulation is given by $\boldsymbol{x}^{-1} = \boldsymbol{x}^0 = \boldsymbol{0}$ and $\boldsymbol{z}^{-1} = \boldsymbol{0}$.

Figure 2 shows the phase transition line of DAMP, where $D_{\min}$ obtained by (16) satisfies $D_{\min} = 1$. For comparison, we also show the phase transition of the AMP algorithm in [11]. The lines are the borders between the success and failure regions of the algorithms in the large system limit. In the left side of the line, we have $D_{\min} > 1$ and hence the MSE of the estimate cannot be zero. In the right side, $D_{\min} < 1$ holds and we can estimate $\boldsymbol{b}$ correctly. From the figure, we can observe that DAMP and AMP needs almost the same number of measurements to precisely reconstruct a discrete-valued sparse vector with large p_0. For a dense vector with small p_0, however, the required number of measurements for DAMP is much less than that for AMP because AMP uses only the sparsity to reconstruct $\boldsymbol{b}$ while DAMP can use both the sparsity and the discreteness. In particular, from the phase transition of DAMP for $p_0 = 0$, we can see that DAMP needs at least $N/2$ measurements to reconstruct N dimensional binary vector $\boldsymbol{b} \in \{\pm 1\}^N$ and $3N/4$ measurements to reconstruct $\boldsymbol{b} \in \{\pm 1, \pm 3\}^N$. For example, in the case with $b_j \in \{\pm 1\}$, the minimum value of $D(\boldsymbol{U})$ is actually $D_{\min} = 1/(2\Delta)$ with $\boldsymbol{U} = [0 \ \infty]^T$ and hence $\Delta > 1/2$ is required. The corresponding soft thresholding function is $[\eta(\boldsymbol{u}, \lambda\sigma)]_j = \min(\max(u_j, -1), 1)$, which is also used in [11] for $\boldsymbol{b} \in [-1, 1]^N$ with a few elements in the interior $(-1, 1)$.

In Fig. 3, we plot $\Psi(\sigma^2) - \sigma^2$ as a function σ^2 for the two cases with $b_j \in \{\pm 1\}$ and $b_j \in \{\pm 1, \pm 3\}$. From the figure, we can see that $\Psi(\sigma^2)$ is concave. The intersections of each curve and the horizontal line $\Psi(\sigma^2) - \sigma^2 = 0$ indicate the fixed points of $\Psi(\sigma^2)$. Figure 3 shows that $\Psi(\sigma^2)$ has no positive fixed point when Δ is larger than the threshold obtained from Fig. 2, i.e., 0.5 for $b_j \in \{\pm 1\}$ and 0.75 for $b_j \in \{\pm 1, \pm 3\}$. Since $\Psi(\sigma^2) < \sigma^2$ holds for any $\sigma^2 > 0$ in this case, the sequence $\{\sigma_t^2\}_{t=0,1,\ldots}$ obtained with the state evolution (6) converges to 0 and hence DAMP provides perfect reconstruction of $\boldsymbol{b}$. In other words, in the large system limit, DAMP can correctly estimate $\boldsymbol{b} \in \{\pm 1\}^N$ if $M > N/2$ and estimate $\boldsymbol{b} \in \{\pm 1, \pm 3\}^N$ if $M > 3N/4$.

The obtained result for DAMP with $p_0 = 0$ agrees with the theorem for the transform-based method in [6], which claims that it can reconstruct a discrete-valued vector in $\{c_1, \ldots, c_T\}^N$ if $M > (T-1)N/T$. In the proof of the theorem, however, it is assumed that the true solution is drawn uniformly from $\{c_1, \ldots, c_T\}^N$. On the other hand, the state evolution for DAMP can also be applied even to non-uniform discrete distributions. Similarly to the theorem for the transform-based method, it is proven that the regularization-based method [6] can reconstruct an N dimensional binary vector if $M > N/2$ in the large system limit, which also agrees with our result for DAMP. For a non-binary vector such as $\boldsymbol{b} \in \{\pm 1, \pm 3\}^N$, however, no theoretical analysis for the method is provided. In fact, it has been shown in [6] by computer simulations that the required number of measurements is larger than $3N/4$ for the reconstruction of $\boldsymbol{b} \in \{\pm 1, \pm 3\}^N$. It is because the coefficients $q_0, \ldots, q_L$ in the optimization problem (2) are all 1 in the method, while the parameters are optimized as (16) by using the state evolution in DAMP.

In Figs. 4 and 5, we compare the prediction with the state evolution (6) and the empirical behavior of the MSE $\sigma_t^2 = \|\boldsymbol{x}^t - \boldsymbol{b}\|_2^2/N$ with DAMP obtained by simulations. In Fig. 4, we fix $\Delta = 0.7$ and reconstruct $\boldsymbol{b} \in \{0, \pm 1\}^N$ with $p_0 = 0.2$ for several problem sizes of $N = 100, 500, 1000, 2000$, and 3000. In Fig. 5, we set $\Delta = 0.9$ and reconstruct $\boldsymbol{b} \in \{0, \pm 1, \pm 3\}^N$ with $p_0 = 0.2$. The figures show that the empirical performance approaches to the prediction with the state evolution as the problem size increases.

In Fig. 6, we empirically reconstruct $\boldsymbol{b} \in \{0, \pm 1\}^N$ with $N = 1000$, $p_0 = 0, 0.4$ to evaluate the rate of the success recovery, which is defined as $\sigma_t^2 < 10^{-3}$ after $t = 500$

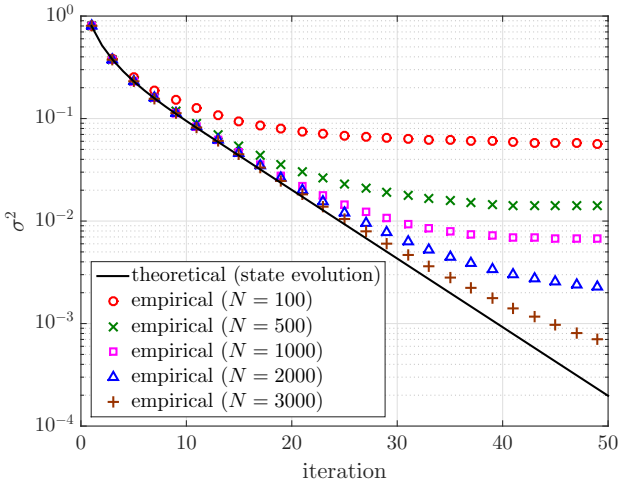

Fig. 4. Empirical performance of DAMP and theoretical prediction via state evolution ($p_0 = 0.2,\ \Delta = 0.7,\ \boldsymbol{b} \in \{0, \pm 1\}^N$)

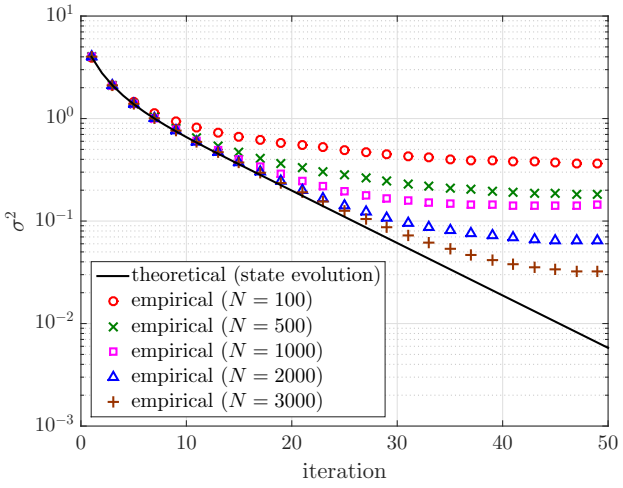

Fig. 5. Empirical performance of DAMP and theoretical prediction via state evolution ($p_0 = 0.2,\ \Delta = 0.9,\ \boldsymbol{b} \in \{0, \pm 1, \pm 3\}^N$)

iterations. The vertical lines correspond to the border between the success and failure region obtained from Fig. 2. In the large system limit, the left side of each vertical line is the failure region and the right side is the success region. We can see that the recovery rate rapidly increases around the border, while the recovery rate is not equal to one in the success region close to the border. One reason is that we restrict the maximum number of iterations as $t = 500$ and another reason is that the problem size here is finite and not large enough.

V. Conclusion

In this paper, we have proposed the DAMP algorithm for the discrete-valued vector reconstruction. Using the state evolution, we have analytically evaluated the performance of DAMP to derive the condition for the perfect reconstruction and to obtain the optimal values of the parameters. Simulation

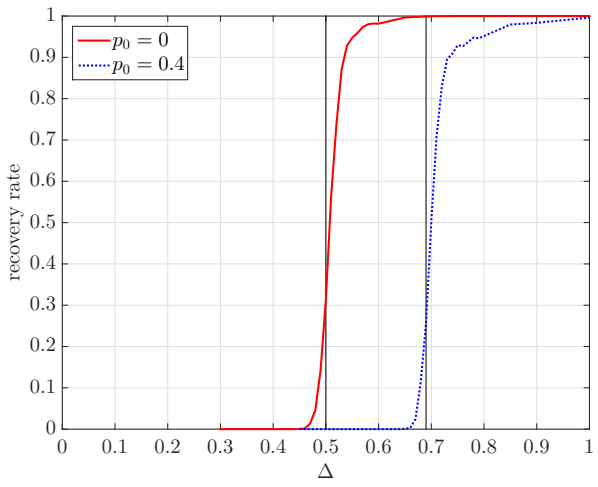

Fig. 6. Recovery rate of DAMP ($N = 1000,\ \boldsymbol{b} \in \{0, \pm 1\}^N$)

results show that DAMP can reconstruct some discrete-valued vectors from their underdetermined linear measurements and the empirical performance agrees well with our theoretical results. Future work includes the extension of DAMP for discrete-valued complex vector and its application.

References

[1] H. Sasahara, K. Hayashi, and M. Nagahara, "Multiuser detection based on MAP estimation with sum-of-absolute-values relaxation." (submitted)

[2] 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.

[3] H. Sasahara, K. Hayashi, and M. Nagahara, "Symbol detection for faster-than-Nyquist signaling by sum-of-absolute-values optimization," *IEEE Signal Process. Lett.*, vol. 23, no. 12, pp. 1853–1857, Dec. 2016.

[4] A. Müller, D. Sejdinovic, and R. Piechocki, "Approximate message passing under finite alphabet constraints," in *Proc. ICASSP 2012*, pp. 3177–3180, Mar. 2012.

[5] S. Sparrer and R. F. H. Fischer, "Enhanced iterative hard thresholding for the estimation of discrete-valued sparse signals," in *Proc. EUSIPCO 2016*, pp. 71–75, Aug.–Sept. 2016.

[6] 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.

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

[8] 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.

[9] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

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

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

[12] 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.

[13] 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.

[14] 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.