Asymptotic Analysis of ADMM for Compressed Sensing

Ryo Hayakawa, Member, IEEE

Abstract—In this paper, we analyze the asymptotic behavior of alternating direction method of multipliers (ADMM) for compressed sensing, where we reconstruct an unknown structured signal from its underdetermined linear measurements. The analytical tool used in this paper is recently developed convex Gaussian min-max theorem (CGMT), which can be applied to various convex optimization problems to obtain its asymptotic error performance. In our analysis of ADMM, we analyze the convex subproblem in the update of ADMM and characterize the asymptotic distribution of the tentative estimate obtained at each iteration. The result shows that the update equations in ADMM can be decoupled into a scalar-valued stochastic process in the asymptotic regime with the large system limit. From the asymptotic result, we can predict the evolution of the error (e.g., mean-square-error (MSE) and symbol error rate (SER)) in ADMM for large-scale compressed sensing problems. Simulation results show that the empirical performance of ADMM and its theoretical prediction are close to each other in sparse vector reconstruction and binary vector reconstruction.

Index Terms—Compressed sensing, alternating direction method of multipliers, convex Gaussian min-max theorem, asymptotic performance

I. INTRODUCTION

Compressed sensing [1]–[4] becomes a key technology in the field of signal processing such as image processing [5], [6] and wireless communication [7], [8]. A basic problem in compressed sensing is to reconstruct an unknown sparse vector from its underdetermined linear measurements, where the number of measurements is less than that of unknown variables. The compressed sensing techniques take advantage of the sparsity as the prior knowledge to reconstruct the vector. The idea of compressed sensing can also be applied for other structured signals by appropriately utilizing the structures, e.g., group sparsity [9], low-rankness [10], [11], and discreteness [12], [13].

For compressed sensing, various algorithms have been proposed in the literature. Greedy algorithms such as matching pursuit (MP) [14] and orthogonal matching pursuit (OMP) [15], [16] iteratively update the support of the estimate of the unknown sparse vector. Several improved greedy algorithms have also been proposed to achieve better reconstruction performance [17]–[22]. Another approach for compressed sensing is based on message passing algorithms using Bayesian framework. Approximated belief propagation (BP) [23] and approximate message passing (AMP) [24] can reconstruct the unknown vector with low computational complexity. Moreover, the asymptotic performance can be predicted by the state evolution framework [24], [25]. The AMP algorithm can also be used when the unknown vector has some structure other than the sparsity [26], [27]. However, the AMP algorithm requires an assumption on the measurement matrix, and hence other message passing-based algorithms have also been proposed [28]–[31].

Convex optimization-based approaches have also been well studied for compressed sensing. The most popular convex optimization problem for compressed sensing is the \( \ell_1 \) optimization, where we utilize the \( \ell_1 \) norm as the regularizer to promote the sparsity of the estimate. Although the objective function is not differentiable, the iterative soft thresholding algorithm (ISTA) [32], [33] and the fast iterative soft thresholding algorithm (FISTA) [35] can solve the \( \ell_1 \) optimization problem with feasible computational complexity. Another promising algorithm is alternating direction method of multipliers (ADMM) [36]–[39], which can be applied to wider class of optimization problems than ISTA and FISTA. Moreover, ADMM can provide a sufficiently accurate solution with relatively small number of iterations in practice [39]. However, since the convergence speed largely depends on the parameter, it is important to determine the appropriate parameter in practical applications. For the parameter selection of ADMM, several approaches have been proposed [40]–[44]. However, they are rather heuristic or inapplicable to compressed sensing problems.

There are several theoretical analyses for convex optimization-based compressed sensing, e.g., [45]–[47]. In particular, recently developed convex Gaussian min-max theorem (CGMT) [48], [49] can be utilized to obtain the asymptotic error of various optimization problems in a precise manner. For example, the asymptotic mean-square-error (MSE) has been analyzed for various regularized estimators [39], [50]. The asymptotic symbol error rate (SER) has also been derived for convex optimization-based discrete-value vector reconstruction [51], [52]. The CGMT-based analysis has been extended for the optimization problem in the complex-valued domain [53], whereas above analyses consider optimization problems in the real-valued domain. These analyses focus on the performance of the optimizer, and do not deeply discuss the optimization algorithm to obtain the optimizer.

In this paper, we analyze the asymptotic behavior of ADMM for convex optimization-based compressed sensing. The main idea is that, when we use the squared loss function as the data fidelity term, the subproblem in the iterations of ADMM can...
be analyzed by the CGMT framework. We thus analyze the asymptotic property of the tentative estimate of the unknown vector at each iteration in ADMM. We show that the asymptotic distribution of the tentative estimate can be characterized by a scalar-valued stochastic process, which depends on the measurement ratio, the parameter in the optimization problem, the parameter in ADMM, the distribution of the unknown vector, and the noise variance. As a corollary, we can predict the evolution of the error such as MSE and SER in ADMM for large-scale compressed sensing problems. We can also utilize the asymptotic result to reveal the effect of the parameter in ADMM and tune it to achieve the fast convergence. As examples, we consider sparse vector reconstruction and binary vector reconstruction and then analyze the asymptotic result via computer simulations. Simulation results show that the asymptotic evolution of MSE converges to the MSE of the optimizer, which can be obtained with the previous CGMT-based analysis in the literature [49]. We also observe that the empirical performance of ADMM and its theoretical prediction are close to each other in both sparse vector reconstruction and binary vector reconstruction.

The rest of the paper is organized as follows. In Section II, we describe the ADMM-based compressed sensing and CGMT as the preliminary. Then we provide the main analytical results for ADMM in Section III. We then consider two examples of the reconstruction problem and show several simulation results. Finally, Section V presents some conclusions.

In this paper, we use the following notations. We denote the transpose by (·)T and the identity matrix by I. For a vector $z = [z_1 \cdots z_N]^T \in \mathbb{R}^N$, the $\ell_1$ norm and the $\ell_2$ norm are given by $||z||_1 = \sum_{n=1}^N |z_n|$ and $||z||_2 = \sqrt{\sum_{n=1}^N z_n^2}$, respectively. We denote the number of nonzero elements of $z$ by $||z||_0$, $\text{sign}(\cdot)$ denotes the sign function. For a lower semicontinuous convex function $\zeta: \mathbb{R}^N \rightarrow \mathbb{R} \cup \{+\infty\}$, we define the proximity operator as $\text{prox}_\zeta(z) = \arg \min_{u \in \mathbb{R}^N} \{\zeta(u) + \frac{1}{2} \|u - z\|_2^2\}$. The Gaussian distribution with mean $\mu$ and variance $\sigma^2$ is denoted as $\mathcal{N}(\mu, \sigma^2)$. When a sequence of random variables $\{\Theta_n\}$ ($n = 1, 2, \ldots$) converges in probability to $\Theta$, we denote $\Theta_n \xrightarrow{p} \Theta$ as $n \rightarrow \infty$ or $\text{plim}_{n \rightarrow \infty} \Theta_n = \Theta$.

II. Preliminaries

A. ADMM-Based Compressed Sensing

In this paper, we consider the reconstruction of an $N$ dimensional vector $x = [x_1 \cdots x_N]^T \in \mathbb{R}^N$ from its linear measurements given by

$$y = Ax + v \in \mathbb{R}^M.$$ 

(1)

Here, $A \in \mathbb{R}^{M \times N}$ is a known measurement matrix and $v \in \mathbb{R}^M$ is an additive Gaussian noise vector. We denote the measurement ratio by $\Delta = M/N$. In the scenario of compressed sensing, we focus on the underdetermined case with $\Delta < 1$ and utilize the structure of $x$ as the prior knowledge for the reconstruction. Note that we can use not only the sparsity but also other structures such as boundedness and discreteness [13], [54].

The convex optimization-based method is a promising approach for compressed sensing because we can flexibly design the objective function to utilize the structure of the unknown vector $x$. In this paper, we consider the following convex optimization problem

$$\min_{s \in \mathbb{R}^N} \left\{ \frac{1}{2} \|y - As\|_2^2 + \lambda f(s) \right\},$$

(2)

where $f(\cdot): \mathbb{R}^N \rightarrow \mathbb{R} \cup \{+\infty\}$ is a convex regularizer to utilize the prior knowledge of the unknown vector $x$. For example, $\ell_1$ regularization $f(s) = \|s\|_1$ is a popular convex regularizer for the reconstruction of the sparse vector. The regularization parameter $\lambda (>0)$ controls the balance between the data fidelity term $\frac{1}{2} \|y - As\|_2^2$ and the regularization term $\lambda f(s)$.

ADMM has been used in wide range of applications because it can be applied to various optimization problems. Moreover, we can obtain a sufficiently accurate solution with relatively small number of iterations in practice [39]. To derive ADMM for the optimization problem (2), we firstly rewrite (2) as

$$\min_{s, z \in \mathbb{R}^N} \left\{ \frac{1}{2} \|y - As\|_2^2 + \lambda f(z) \right\}$$

subject to $s = z$.

(3)

The update equations of ADMM for (3) are given by

$$s^{(k+1)} = \arg \min_{s \in \mathbb{R}^N} \left\{ \frac{1}{2} \|y - As\|_2^2 + \frac{\rho}{2} \|s - z^{(k)} + w^{(k)}\|_2^2 \right\}$$

(4)

$$= (A^T A + \rho I)^{-1} (A^T y + \rho (z^{(k)} - w^{(k)})),$$

(5)

$$z^{(k+1)} = \arg \min_{z \in \mathbb{R}^N} \left\{ \lambda f(z) + \frac{\rho}{2} \|s^{(k+1)} - z + w^{(k)}\|_2^2 \right\}$$

(6)

$$= \text{prox}_{\lambda f}(s^{(k+1)} + w^{(k)}),$$

(7)

$$w^{(k+1)} = w^{(k)} + s^{(k+1)} - z^{(k+1)},$$

(8)

where $k (=0, 1, 2, \ldots)$ is the iteration index in the algorithm and $\rho (>0)$ is the parameter. In this paper, we refer to $s^{(k+1)}$ as the tentative estimate of the unknown vector $x$ in ADMM. We use $z^{(0)} = w^{(0)} = 0$ as the initial value in this paper.

B. CGMT

CGMT associates the primary optimization (PO) problem with the auxiliary optimization (AO) problem given by

$$\text{(PO): } \Phi(G) = \min_{w \in S_w} \left\{ w^T G w + \xi(w, u) \right\},$$

(9)

$$\text{(AO): } \phi(g, h) = \min_{w \in S_w} \left\{ \|w\|_2 g^T w - \|w\|_2 h^T w + \xi(w, u) \right\},$$

(10)

respectively, where $G \in \mathbb{R}^{M \times N}$, $g \in \mathbb{R}^M$, $h \in \mathbb{R}^N$, $S_w \subset \mathbb{R}^N$, $S_u \subset \mathbb{R}^M$, and $\xi(\cdot, \cdot) : \mathbb{R}^N \times \mathbb{R}^M \rightarrow \mathbb{R}$. $S_w$ and $S_u$ are assumed to be closed compact sets. $\xi(\cdot, \cdot)$ is a continuous convex-concave function on $S_w \times S_u$. Also, the elements of $G$. 


\(g,\) and \(h\) are i.i.d. standard Gaussian random variables. From the following theorem, we can relate the optimizer \(\hat{w}_\Phi(G)\) of (PO) with the optimal value of (AO) in the large system limit of \(M, N \to \infty\) with a fixed ratio \(\Delta = M/N\). For simplicity, we denote the large system limit by \(N \to \infty\) in this paper.

**Theorem II.1** (CGMT [51]). Let \(S\) be an open set in \(S_w\) and \(S' = S_w \setminus S\). Also, let \(\phi_{S'}(g, h)\) be the optimal value of (AO) with the constraint \(w \in S'\). If there are constants \(\eta > 0\) and \(\phi\) satisfying (i) \(\phi_{S'}(g, h) \leq \phi + \eta\) and (ii) \(\phi_{S'}(g, h) \geq \phi + 2\eta\) with probability approaching 1, then we have \(\lim_{N \to \infty} \Pr(\hat{w}_\Phi(G) \in S) = 1\).

### III. Main Result

In this section, we provide the main analytical result for the behavior of ADMM for the problem (2). In the analysis, we use the following assumptions.

**Assumption III.1.** The unknown vector \(x\) is composed of independent and identically distributed (i.i.d.) random variables with a known distribution \(p_X\) which has some mean and variance. The measurement matrix \(A \in \mathbb{R}^{M \times N}\) is composed of i.i.d. Gaussian random variables with zero mean and variance \(1/N\). Moreover, the additive noise vector \(v \in \mathbb{R}^M\) is also Gaussian with zero mean and the covariance matrix \(\sigma^2 I\).

**Assumption III.2.** The regularizer \(f(\cdot) : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}\) is a lower semicontinuous convex function. Moreover, \(f(\cdot)\) is separable and expressed with the corresponding function \(\tilde{f}(\cdot) : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}\) as \(f(s) = \sum_{n=1}^N \tilde{f}(s_n)\), where \(s = [s_1 \cdots s_N]^T \in \mathbb{R}^N\). With the slight abuse of notation, we use the same \(f(\cdot)\) for the corresponding function \(\tilde{f}(\cdot)\).

In Assumption III.1 we assume that the elements of the measurement matrix \(A\) are Gaussian variables because CGMT requires the Gaussian assumption in the proof [49]. However, the universality [53, 57] of random matrices suggests that the result of the analysis can be applied when the measurement matrix is drawn from some other distributions. In fact, our theoretical result is valid for the random matrix from Bernoulli distribution with \(\{1/\sqrt{N}, -1/\sqrt{N}\}\) in computer simulations (See Example IV.1).

In Assumption III.2, we assume the separability of the regularizer \(f(\cdot)\). Under this assumption, the proximity operator \(\text{prox}_{\tilde{f}}(\cdot) : \mathbb{R}^N \to \mathbb{R}^N\) becomes an element-wise function, i.e., the \(n\)-th element of the output depends only on the corresponding \(n\)-th element of the input.

Under Assumptions III.1 and III.2, we present the following theorem.

**Theorem III.1.** We assume that \(x, A, v,\) and \(f(\cdot)\) satisfy the Assumptions III.1 and III.2. We consider the following stochastic process

\[
S_{k+1} = \hat{S}_{k+1}(\alpha_k^*, \beta_k^*)
\]

\[
Z_{k+1} = \text{prox}_{\tilde{f}}(S_{k+1} + W_k)
\]

\[
W_{k+1} = W_k + S_{k+1} - Z_{k+1},
\]

with the index \(k\), where \(\hat{S}_{k+1}(\alpha, \beta)\) is defined as

\[
\hat{S}_{k+1}(\alpha, \beta) = \frac{1}{\beta \sqrt{\Delta}} \left( \beta \sqrt{\Delta} \left( X + \frac{\alpha}{\sqrt{\Delta}} H \right) + \rho(Z_k - W_k) \right)
\]

with the random variables \(X \sim p_X\) and \(H \sim N(0, 1)\) \((Z_0 = W_0 = 0)\). We here assume the optimization problem

\[
\min_{\alpha > 0, \beta > 0} \left\{ \frac{\alpha \beta \sqrt{\Delta}}{2} + \frac{\beta^2 \sigma^2}{2} - \frac{1}{2} \beta^2 + E \left[ J^{(k+1)}(\alpha, \beta) \right] \right\}
\]

has a unique optimizer \((\alpha_k^*, \beta_k^*)\), where

\[
J^{(k+1)}(\alpha, \beta) = \frac{\beta \sqrt{\Delta}}{2} \left( \hat{S}_{k+1}(\alpha, \beta) - X \right)^2 - \beta H \left( \hat{S}_{k+1}(\alpha, \beta) - X \right) + \frac{\rho}{2} \left( \hat{S}_{k+1}(\alpha, \beta) - Z_k + W_k \right)^2.
\]

The expectation is taken over all random variables \(X, H, Z_k,\) and \(W_k\).

Let \(\mu_{s(k+1)}\) be the empirical distribution of \(s(k+1)\) corresponding to the cumulative distribution function (CDF) given by \(P_{\mu_{s(k+1)}}(s) = \frac{1}{N} \sum_{n=1}^N \left[ \mathbb{I}(s_n^{(k+1)} < s) \right]\), where we define \(\mathbb{I}(s_n^{(k+1)} < s) = 1\) if \(s_n^{(k+1)} < s\) and otherwise \(\mathbb{I}(s_n^{(k+1)} < s) = 0\). Moreover, we denote the distribution of the random variable \(S_{k+1}\) in (11) as \(\mu_{S_{k+1}}\). Then, the empirical distribution \(\mu_{s(k+1)}\) converges weakly in probability to \(\mu_{S_{k+1}}\) as \(N \to \infty\), i.e., \(\int g d\mu_{s(k+1)} \to \int g d\mu_{S_{k+1}}\) holds for any continuous compactly supported function \(g(\cdot) : \mathbb{R} \to \mathbb{R}\).

**Proof:** See Appendix A.

**Theorem III.1** means that the distribution of the elements of \(s^k\) is characterized by the random variable \(S_k\) in the asymptotic regime with \(M, N \to \infty\) \((M/N = \Delta)\). The update of \(S_k\) in (11) can be regarded as the ‘decoupled’ version of the update of \(s^k\) in (4). Figure II shows the comparison between the update of \(\hat{S}_{k+1}(\alpha, \beta)\) and its empirical version obtained from Theorem III.1. In the update of \(s(k)\), the measurement vector \(y\) is obtained through the linear transformation of \(x\) and additive Gaussian noise channel. On the other hand, in the decoupled system, the random variable \(X\) goes through only the additive Gaussian noise channel. We can also see that the update of \(s(k)\) and \(S_k\) have the similar form because they can be rewritten as

\[
s^{(k+1)} = (A^T A + \rho I)^{-1} \left( A^T Ax + A^T v + \rho \left( z^{(k)} - w^{(k)} \right) \right),
\]

\[
S_{k+1} = \frac{1}{\beta \sqrt{\Delta}} \left( \beta \sqrt{\Delta} X + \beta \sigma^2 H + \rho(Z_k - W_k) \right),
\]
reconstruction algorithm is MSE given by

\[ s^{(k+1)} = (A^T A + \rho I)^{-1} (A^T y + \rho (z^{(k)} - w^{(k)})) \]

\[ = (A^T A + \rho I)^{-1} (A^T Ax + A^T v + \rho (z^{(k)} - w^{(k)})) \]

(a) update of \( s^{(k)} \) in ADMM

respectively. The update of \( S_k \) in \( (11) \) and \( (14) \) shows that \( S_{k+1} \) is the weighted sum of \( X + \frac{\alpha_k^p}{\alpha_k^p} H \) and \( Z_k - W_k \) with the weights \( \beta_k^* \sqrt{\Delta} \) and \( \rho \), respectively. Since \( \rho \) is the parameter of ADMM, we can control the weight in the update of \( S_k \) by tuning \( \rho \).

One of the most important performance measures for the reconstruction algorithm is MSE given by \( \frac{1}{N} \| s^{(k)} - x \|_2^2 \). As in the CGMT-based analysis \[49\], the optimal value of \( \alpha \) is related to the asymptotic MSE. Specifically, from Theorem \[III.1\] the asymptotic MSE of the tentative estimate \( s^{(k)} \) in ADMM can be obtained as follows (See also \[52, Remark IV.1\]).

**Corollary III.1.** Under the assumptions in Theorem \[III.1\] the asymptotic MSE of \( s^{(k+1)} \) is given by

\[ \lim_{N \to \infty} \frac{1}{N} \| s^{(k+1)} - x \|_2^2 = (\alpha_k^p)^2 - \sigma^2. \] (19)

From the theoretical result in Theorem \[III.1\] (or Corollary \[III.1\]), we can tune the parameter \( \rho \) in ADMM to achieve the fast convergence. The conventional parameter tuning \[40\]–\[44\] focus on the difference between the tentative estimate and the optimizer of the optimization problem. On the other hand, the parameter tuning based on Theorem \[III.1\] can take account of the error from the true unknown vector in the asymptotic regime. Since the effect of \( \rho \) on \( \alpha_k^p \) and \( \beta_k^* \) is complicated, the explicit expression of the optimal \( \rho \) is difficult to obtain. By numerical simulations, however, we can predict the performance of ADMM and select the parameter \( \rho \) achieving the fast convergence. For instance, see Example \[IV.1\] in Section \[IV\].

**IV. Examples**

In this section, we consider two examples of the reconstruction problem and compare the empirical performance of ADMM and its prediction obtained by Theorem \[III.1\].

---

**Example IV.1 (Sparse Vector Reconstruction).** The \( \ell_1 \) optimization

\[ \min_{s \in \mathbb{R}^N} \left\{ \frac{1}{2} \| y - As \|_2^2 + \lambda \| s \|_1 \right\} \] (20)

with the \( \ell_1 \) norm is the most popular convex optimization problem for sparse vector reconstruction. The \( \ell_1 \) regularization promotes the sparsity of the estimate of the unknown vector in the reconstruction. We here assume that the distribution of the unknown vector \( x \) is given by the Bernoulli-Gaussian distribution as

\[ p_X(x) = p_0 \delta_0(x) + (1 - p_0) p_H(x), \] (21)

where \( p_0 \in (0, 1) \), \( \delta_0(\cdot) \) denotes the Dirac delta function, and \( p_H(\cdot) \) is the probability density function of the standard Gaussian distribution. When \( p_0 \) is large, the unknown vector becomes sparse. The proximity operator of the \( \ell_1 \) norm is given by

\[ \left[ \text{prox}_{\| s \|_1}(r) \right]_n = \text{sign}(r_n) \max(|r_n| - \gamma, 0), \] (22)

where \( r = [r_1 \cdots r_N]^T \in \mathbb{R}^N \), \( \gamma > 0 \), and \([\cdot]_n\) denotes the \( n \)-th element of the vector. By using (22), we can perform ADMM in (4)–(8) for the \( \ell_1 \) optimization (20). Theorem \[III.1\] enables us to predict the asymptotic behavior of ADMM for the \( \ell_1 \) optimization.

We firstly compare the empirical performance of the sparse vector reconstruction and its prediction obtained from Theorem \[III.1\]. Figure [2] shows that the MSE performance of the sparse vector reconstruction, where \( \Delta = 0.9 \), \( p_0 = 0.8 \), and \( \sigma^2 = 0.001 \). The measurement matrix \( A \) and the noise vector \( v \) satisfy Assumption \[III.1\]. The parameter \( \rho \) of ADMM is set as \( \rho = 0.1 \). In the figure, ‘empirical’ means the empirical performance obtained by ADMM in (4)–(8) when \( N = 50, 100, 500, \) and \( 1000 \). The empirical performance is obtained by averaging the results for 100 independent realizations of \( x, A, \) and \( v \). On the other hand, ‘prediction’
shows the theoretical performance prediction obtained by Theorem 3.1 (or Corollary 3.1) in the large system limit. Since the exact computation of the distribution of $(S_k, Z_k, W_k)$ is difficult in practice, we make 100,000 realizations of the random variables $(S_k, Z_k, W_k)$ and obtain the approximation of $(\alpha_k^*, \beta_k^*)$. For the optimization of $\alpha$ and $\beta$, we can use searching techniques such as ternary search and golden-section search [58]. In the simulations, we use the ternary search with the error tolerance $10^{-6}$. We also show the asymptotic MSE of the optimizer obtained by applying CGMT to the $\ell_1$ optimization problem as in [49]. The parameter $\lambda$ in [49] is determined by minimizing the asymptotic MSE. From Fig. 2 we can see that the empirical performance and its prediction are close to each other. Moreover, they converge to the asymptotic MSE of the optimizer in the original $\ell_1$ optimization problem. Precisely, there is a slight difference between the empirical performance and its prediction. One of possible reasons is that the empirical performance is evaluated for finite $N$, whereas the large system limit $N \to \infty$ is assumed in the asymptotic analysis. Another reason is that we create the many realizations of $(S_k, Z_k, W_k)$ for the theoretical prediction instead of computing their exact distributions.

Next, we evaluate the MSE performance for different matrix structures. Figure 3 shows the MSE performance when $N = 500$, $M = 250$, $p_0 = 0.9$, $\sigma_v^2 = 0.001$, and $\rho = 0.1$. In the figure, ‘Gaussian’ means the performance when the measurement matrix $A$ is composed of i.i.d. Gaussian elements and satisfies Assumption 3.1. On the other hand, ‘Bernoulli’ shows the performance when each element of measurement matrix is drawn uniformly from $\{1/\sqrt{N}, -1/\sqrt{N}\}$. The empirical performance is obtained by averaging the results for 500 independent realizations of $x$, $A$, and $v$. From Fig. 3 we observe that the empirical performance for the both cases is close to the theoretical prediction obtained by Theorem 3.1 (or Corollary 3.1).

We then evaluate the effects of the parameter $\rho$ in ADMM. Figure 4 shows the asymptotic MSE performance for $\rho = 0.05, 0.2$, and 0.5. In the figure, we set $\Delta = 0.8$, $p_0 = 0.9$, and $\sigma_v^2 = 0.005$. We can see that the value of the parameter $\rho$ significantly affects the convergence speed of ADMM. By using the theoretical prediction obtained from Theorem 3.1 we can adjust $\rho$ to achieve the fast convergence.

**Example IV.2** (Binary Vector Reconstruction). We consider the reconstruction of a binary vector $x \in \{1, -1\}^N$ with

$$ p_X(x) = \frac{1}{2} (\delta_0(x-1) + \delta_0(x+1)). $$

A reasonable approach to reconstruct $x \in \{1, -1\}^N$ is the box relaxation method [54], [59] given by

$$ \minimize_{s \in [-1,1]^N} \frac{1}{2} \|y - As\|^2_2, $$

which is a convex relaxation of the maximum likelihood approach

$$ \minimize_{s \in \{1,-1\}^N} \frac{1}{2} \|y - As\|^2_2. $$

The asymptotic performance of the final estimate obtained by the box relaxation method has been analyzed with CGMT in [51]. The optimization problem \( \text{(24)} \) is equivalent to \( \text{(2)} \) with \( f(s) = \sum_{n=1}^{N} \psi(s_n) \), where
\[
\psi(s) = \begin{cases} 
0 & (s \in [-1, 1]) \\
\infty & (s \notin [-1, 1])
\end{cases}.
\]
Since the proximity operator of \( \psi(\cdot) \) is given by the projection to \([-1, 1] \), i.e.,
\[
\text{prox}_{\psi}(r) = \min(\max(r, -1), 1),
\]
we can perform ADMM in \( \text{(4)}–\text{(8)} \) by using \( \text{(27)} \). From Theorem III.1, we can predict the asymptotic performance of ADMM for the box relaxation method.

We evaluate the SER performance defined as \( \|\text{sign}(s^{(k)}) - x\|_0/N \), which is important performance measure in binary vector reconstruction. From Theorem III.1, we can predict the asymptotic SER performance by \( \text{E} \{\text{sign}(S_k) - X\} \). Although the sign function is not continuous, we can approximate the function to use the result of Theorem III.1 (cf. [51, Lemma A.4]). Figure 5 shows the SER performance of ADMM for \( \Delta = 0.7, 0.8, \) and \( 0.9 \), where \( N = 500, \sigma_x^2 = 0.04, \) and \( \rho = 0.1 \). The empirical performance is obtained by averaging 300 results for independent realizations of \( x, A, \) and \( \nu \). The theoretical prediction is computed by making 100,000 realizations of the random variables \( (S_k, Z_k, W_k) \). From Fig. 5, we observe that the empirical CDF agrees well with the theoretical prediction at each iteration. We can also see that the distributions concentrate near 1 and −1 as the iteration proceeds.

V. Conclusion

In this paper, we have analyzed the asymptotic behavior of ADMM for compressed sensing. By using recently developed CGMT framework, we have shown that the asymptotic distribution of the tentative estimate in ADMM is characterized by the stochastic process \( \{S_k\}_{k=1,2,...} \). The main theorem enables us to predict the error evolution of ADMM in the large system limit. We can also tune the parameter in ADMM from the asymptotic result. Simulation results show that the empirical performance obtained by ADMM and its theoretical prediction are close to each other in terms of MSE and SER in sparse vector reconstruction and binary vector reconstruction, respectively.

We here show some possible research directions based on the analysis in this paper. Although we consider the fixed parameter \( \rho \) in ADMM, it is possible to use the different parameter \( \rho_k \) at each iteration and predict the asymptotic performance in the same manner. The theoretical result in this case would provide the faster convergence of the algorithm. Moreover, both ADMM and CGMT can be applied to the convex optimization problem in the complex-valued domain [53], [60], [61]. It would be also an interesting topic to analyze the performance of ADMM for compressed sensing problems in the complex-valued domain, which often appear in communication systems.

Appendix A

Proof of Theorem III.1

In Appendices A, B, we provide the proof of the main theorem in Theorem III.1. Figure 7 shows the overview of the proof in the appendices.

The equations \( \text{(11)}–\text{(13)} \) in Theorem III.1 correspond to the updates \( \text{(4)}–\text{(8)} \) in ADMM, respectively. Since the updates of \( z^{(k)} \) and \( w^{(k)} \) are element-wise from Assumption III.2, we can see that these updates can be characterized by \( \text{(12)} \) and \( \text{(13)} \), respectively. Hence, it is sufficient to show that the behavior of \( s^{(k+1)} \) in \( \text{(4)} \) is characterized with the random variable \( S_{k+1} \) in \( \text{(14)} \). By applying the standard approach with CGMT to the optimization problem \( \text{(3)} \), we can obtain the following lemma, which implies that \( S_{k+1} \) has the probabilistic property of \( S^{(k+1)} \) in the asymptotic regime.

Lemma A.1. Let
\[
\mathcal{L} = \{\psi(\cdot, \cdot) : \mathbb{R} \times \mathbb{R} \to \mathbb{R} | \psi( \cdot, x) \text{ is Lipschitz continuous for any } x \in \mathbb{R}\}.
\]
For any function \( \psi(\cdot, \cdot) \in \mathcal{L} \), we have
\[
\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \psi \left( s_n^{(k+1)} - x_n, x_n \right) = E \left[ \psi \left( S_{k+1} - X, X \right) \right].
\]
To prove Theorem III.1, we show
\[ \lim_{N \to \infty} \text{Pr} \left( \left| \int gd\mu_{x(k+1)} - \int gd\mu_{S_{k+1}} \right| > \varepsilon \right) = 0 \]  \hspace{1cm} (30)
for any continuous compactly supported function \( g(\cdot) : \mathbb{R} \to \mathbb{R} \) and any \( \varepsilon > 0 \). Since the function \( g(\cdot) \) has a compact support, there exists a polynomial \( \nu(\cdot) : \mathbb{R} \to \mathbb{R} \) such that
\[ |g(x) - \nu(x)| < \frac{\varepsilon}{3} \]  \hspace{1cm} (31)
for any \( x \) in the support from the Stone-Weierstrass theorem [62]. We thus have
\[ \left| \int gd\mu_{x(k+1)} - \int gd\mu_{S_{k+1}} \right| < \frac{2}{3} \varepsilon. \]  \hspace{1cm} (32)
Since the polynomial \( \nu(\cdot) \) is Lipschitz on the compact support, we define \( \psi(e, x) = \nu(e + x) \) in Lemma A.1 and obtain
\[ \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \nu\left( s^{(k+1)}_n \right) = E\left[ \nu\left( S_{k+1} \right) \right]. \]  \hspace{1cm} (34)
Since we have (30) from (33) and (34), we can obtain \( \int gd\mu_{x(k+1)} \stackrel{P}{\to} \int gd\mu_{S_{k+1}} \) as \( N \to \infty \), which is the result of Theorem III.1.

Appendix A (Proof of Theorem III.1)
Obtain the asymptotic distribution of \( s^{(k+1)} \)

Appendix B (Proof of Lemma A.1)
Analyze (4) via CGMT framework

Appendix C (Proof of Lemma B.1)
Analyze the solution of (AO)

Construct \( \delta_{k+1} \)

Apply CGMT (Theorem I.1)

Show the convergence of \( \mu_{s(k+1)} \)

---

Fig. 6. Comparison between empirical CDF and its prediction in binary vector reconstruction (\( N = 500, M = 400, \sigma^2 = 0.001, \rho = 0.1 \)).

Fig. 7. Overview of Appendices A–C.

Proof: See Appendix B
To prove Theorem III.1 we show
\[
\lim_{N \to \infty} \text{Pr} \left( \left| \int gd\mu_{x(k+1)} - \int gd\mu_{S_{k+1}} \right| < \varepsilon \right) = 1 \quad (30)
\]
for any continuous compactly supported function \( g(\cdot) : \mathbb{R} \to \mathbb{R} \) and any \( \varepsilon > 0 \). Since the function \( g(\cdot) \) has a compact support, there exists a polynomial \( \nu(\cdot) : \mathbb{R} \to \mathbb{R} \) such that
\[
|g(x) - \nu(x)| < \frac{\varepsilon}{3} \quad (31)
\]
for any \( x \) in the support from the Stone-Weierstrass theorem [62]. We thus have
\[
\left| \int gd\mu_{x(k+1)} - \int gd\mu_{S_{k+1}} \right| < \frac{2}{3} \varepsilon. \quad (32)
\]
Since the polynomial \( \nu(\cdot) \) is Lipschitz on the compact support, we define \( \psi(e, x) = \nu(e + x) \) in Lemma A.1 and obtain
\[
\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \nu\left( s^{(k+1)}_n \right) = E\left[ \nu\left( S_{k+1} \right) \right]. \quad (34)
\]
Since we have (30) from (33) and (34), we can obtain \( \int gd\mu_{x(k+1)} \stackrel{P}{\to} \int gd\mu_{S_{k+1}} \) as \( N \to \infty \), which is the result of Theorem III.1.

\section*{Appendix B
Proof of Lemma A.1

We investigate the asymptotic behavior of the update equation (4). Since the analysis for the optimization problem (4) is based on the standard approach with CGMT [49], we omit some details and show only the outline of the proof. For details of the CGMT-based analysis, see [49–52] and references therein.

\subsection*{A. (PO) Problem}
We firstly define the error vector \( e = s - x \) to rewrite the optimization problem (4) as
\[
\min_{e \in \mathbb{R}^n} \frac{1}{N} \left\{ \frac{1}{2} \| A e - v \|_2^2 + \frac{\rho}{2} \| e + z^{(k)} + w^{(k)} \|_2^2 \right\}, \quad (35)
\]
where the objective function is normalized by \( N \). By using
\[
\frac{1}{2} \| A e - v \|_2^2 = \max_{u \in \mathbb{R}^M} \left\{ \sqrt{N} u^T(A e - v) - \frac{N}{2} \| u \|_2^2 \right\}, \quad (36)
\]
we can obtain the equivalent (PO) problem given by

\[
\min_{e \in \mathbb{R}^N} \max_{u \in \mathbb{R}^M} \left\{ \frac{1}{N} u^T (\sqrt{N} A) e - \frac{1}{\sqrt{N}} v^T u - \frac{1}{2} \|u\|_2^2 + \frac{1}{2} \frac{\rho}{N} \|e + x - z^{(k)} + w^{(k)}\|_2^2 \right\}. \tag{37}
\]

B. (AO) Problem

The corresponding (AO) problem is given by

\[
\min_{e \in S_e} \max_{u \in S_u} \left\{ \frac{1}{N} (\|e\|_2 g^T u - \|u\|_2 h^T e) - \frac{1}{\sqrt{N}} v^T u \right\}
- \frac{1}{2} \|u\|_2 \|e + x - z^{(k)} + w^{(k)}\|_2^2 \right\}. \tag{38}
\]

Although the constraint set of the problem (37) is unbounded, we can introduce a bounded constraint with sufficiently large constraint sets \(S_e\) and \(S_u\) to apply CGMT. (For details, see [49] Appendix A). Since both \(g\) and \(v\) are Gaussian, the vector \(\frac{\sqrt{N}}{N} g - v\) is also Gaussian with zero mean and the covariance matrix \(\frac{\|e\|_2^2}{N} + \sigma^2 I\). Hence, we can rewrite

\[
\left(\frac{\|e\|_2}{\sqrt{N}} g - v\right)^T u \sqrt{\frac{\|e\|_2^2}{N} + \sigma^2} g^T u \text{ with the slight abuse of notation, where } g \text{ has i.i.d. standard Gaussian elements. We apply this technique to (38), set } \|u\|_2 = \beta, \text{ and use the identity}
\]

\[
\chi = \min_{\alpha > 0} \left( \frac{\alpha}{2} + \frac{\chi^2}{2\alpha} \right) \tag{39}
\]

for \(\chi (> 0)\) to rewrite (38) as

\[
\min_{\alpha > 0} \max_{\beta > 0} \left\{ \frac{\alpha \beta \|g\|_2}{2 \sqrt{N}} + \frac{\beta \|g\|_2}{2 \alpha \sqrt{N}} - \frac{1}{2} \beta^2 + \frac{1}{N} \sum_{n=1}^N J_n^{(k+1)}(e_n, \alpha, \beta) \right\}, \tag{40}
\]

where

\[
J_n^{(k+1)}(e_n, \alpha, \beta) = \frac{\beta \|g\|_2}{2 \sqrt{N}} e_n - \beta h_n e_n + \rho \left( e_n + x_n - z_n^{(k)} + w_n^{(k)} \right)^2. \tag{41}
\]

The minimum value of \(J_n^{(k+1)}(e_n, \alpha, \beta)\) is achieved when

\[
\hat{e}_n^{(k+1)}(\alpha, \beta) = \frac{1}{\beta \|g\|_2} \left( 3h_n - \rho \left( x_n - z_n^{(k)} + w_n^{(k)} \right) \right). \tag{42}
\]

We then define \(\hat{s}_n^{(k+1)}(\alpha, \beta) = \hat{e}_n^{(k+1)}(\alpha, \beta) + x_n\), which is given by

\[
\hat{s}_n^{(k+1)}(\alpha, \beta) = \frac{1}{\beta \|g\|_2} \left( \frac{\beta \|g\|_2}{\alpha \sqrt{N}} \left( x_n + \sqrt{N} \frac{\|g\|_2}{\alpha} h_n \right) \right) + \rho \left( z_n^{(k)} - w_n^{(k)} \right). \tag{43}
\]

The optimization problem (40) can be rewritten as

\[
\min_{\alpha > 0} \max_{\beta > 0} \left\{ \frac{\alpha \beta \|g\|_2}{2 \sqrt{N}} + \frac{\beta \|g\|_2}{2 \alpha \sqrt{N}} - \frac{1}{2} \beta^2 + \frac{1}{N} \sum_{n=1}^N J_n^{(k+1)}(\alpha_n, \beta_n) \right\}, \tag{44}
\]

As \(N \to \infty\), the objective function of (44) converges pointwise to

\[
\frac{\alpha \beta \sqrt{\Delta}}{2} + \frac{\beta \|g\|_2}{2 \alpha \sqrt{N}} - \frac{1}{2} \beta^2 + E \left( J_n^{(k+1)}(\alpha, \beta) \right), \tag{45}
\]

where

\[
J_n^{(k+1)}(\alpha, \beta) = \frac{\beta \sqrt{\Delta}}{2 \alpha} \left( \hat{S}_{k+1}(\alpha, \beta) - X \right)^2 - \beta H \left( \hat{S}_{k+1}(\alpha, \beta) - X \right) + \frac{\rho}{2} \left( \hat{S}_{k+1}(\alpha, \beta) - Z_k + W_k \right)^2. \tag{46}
\]

and \(\hat{S}_{k+1}(\alpha, \beta)\) is defined in (14). Note that the function (45) is the objective function of (15) in Theorem III.1.

C. Applying CGMT

To apply CGMT for the above (PO) and (AO), we consider the conditions (i) and (ii) in Theorem II.1. We denote the optimal value of the objective function in (44) and the corresponding solution as \(\phi_{k,N}^*\) and \((\alpha_{k,N}^*, \beta_{k,N}^*)\), respectively. The optimal value of \(e\) in (AO) is given by \(e_n^{(k+1)}(\alpha_{k,N}^*, \beta_{k,N}^*) = \left[ e_1^{(k+1)}(\alpha_{k,N}^*, \beta_{k,N}^*) \ldots e_N^{(k+1)}(\alpha_{k,N}^*, \beta_{k,N}^*) \right]^T \) from (40)-(42). Moreover, let \(\varphi_k^*\) be the optimal value of the objective function in (15) (= (45)) and recall that \((\alpha_k^*, \beta_k^*)\) is the corresponding optimal value of \((\alpha, \beta)\). By a similar discussion to [51] Lemma IV.1], we have \(\phi_{k,N}^* \xrightarrow{P} \varphi_k^*\) and \((\alpha_{k,N}^*, \beta_{k,N}^*) \xrightarrow{P} (\alpha_k^*, \beta_k^*)\) as \(N \to \infty\). Thus, the optimal value of (AO) satisfies the condition (i) in Theorem II.1 for \(\varphi = \varphi^*\) and any \(\eta (> 0)\).

Next, we investigate the condition (ii) in Theorem III.1. We use the following lemma to construct the set \(S\) in CGMT.

Lemma B.1. For any function \(\psi(\cdot, \cdot) \in \mathcal{L}\) (\(\mathcal{L}\) is defined in (28)), we have

\[
\operatorname{plim}_{N \to \infty} \frac{1}{N} \sum_{n=1}^N \psi \left( \hat{s}_n^{(k+1)}(\alpha_n^*, \beta_n^*) - x_n, x_n \right) = E \psi \left( \hat{S}_{k+1}(\alpha_k^*, \beta_k^*) - X, X \right). \tag{47}
\]

Proof: See Appendix C

From Lemma B.1 we define

\[
S_{k+1} = \left\{ z \in \mathbb{R}^N \left| \frac{1}{N} \sum_{n=1}^N \psi(z_n, x_n) - E \psi \left( \hat{S}_{k+1}(\alpha_k^*, \beta_k^*) - X, X \right) < \varepsilon \right. \right\}. \tag{48}
\]
and obtain $s_n^{(k+1)}(\alpha_k^*, \beta_k^*) \in S_{k+1}$ with probability approaching 1 for any $\varepsilon (> 0)$. By using the strong convexity of $J_n^{(k+1)}(\epsilon_n, \alpha, \beta)$ over $\epsilon_n$, we can see that there exists a constant $\eta$ satisfying the condition (ii) in CGMT with $\epsilon_{k+1}$. Hence, from CGMT, Lemma B.1 holds not only for the optimizer of (AO) in (38) but also for that of (PO) in (37), i.e., we have

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \psi \left( s_n^{(k+1)} - x_n, x_n \right) = E \left[ \psi \left( \tilde{S}_{k+1} - X, X \right) \right].$$

(49)

**APPENDIX C**

**PROOF OF LEMMA B.1**

Define

$$s_n^{(k+1)}(\alpha, \beta) = \frac{1}{\beta \sqrt{\Delta}} \frac{\alpha \sqrt{\Delta}}{\alpha + \rho} \left( x_n + \frac{\alpha \sqrt{\Delta}}{\alpha + \rho} h_n \right) + \rho \left( \tilde{z}_n^{(k)} - w_n^{(k)} \right),$$

where we replace $\frac{1}{\sqrt{N}} g$ in (48) with its asymptotic value $\sqrt{\Delta}$.

From the law of large numbers, we have

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \psi \left( s_n^{(k+1)}(\alpha_k^*, \beta_k^*) - x_n, x_n \right) = E \left[ \psi \left( \tilde{S}_{k+1}(\alpha_k^*, \beta_k^*) - X, X \right) \right].$$

(51)

Thus, it is sufficient to show

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \psi \left( s_n^{(k+1)}(\alpha_k^*, \beta_k^*) - x_n, x_n \right) - \frac{1}{N} \sum_{n=1}^{N} \psi \left( s_n^{(k+1)}(\alpha_k^*, \beta_k^*) - x_n, x_n \right) = 0.$$  

(52)

Since $\psi(\cdot, x_n)$ is Lipschitz, there is a constant $C_\psi (> 0)$ such that

$$\left| \frac{1}{N} \sum_{n=1}^{N} \psi \left( s_n^{(k+1)}(\alpha_k^*, \beta_k^*) - x_n, x_n \right) - \frac{1}{N} \sum_{n=1}^{N} \psi \left( s_n^{(k+1)}(\alpha_k^*, \beta_k^*) - x_n, x_n \right) \right| \leq \frac{C_\psi}{N} \sum_{n=1}^{N} \left| s_n^{(k+1)}(\alpha_k^*, \beta_k^*) - s_n^{(k+1)}(\alpha_k^*, \beta_k^*) \right| \leq 0 \quad (N \to \infty),$$

(53)

(54)

which completes the proof.

**REFERENCES**

[1] E. J. Candès and T. Tao, “Decoding by linear programming,” *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[2] E. J. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

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

[4] E. J. Candès and B. Recht, “Exact matrix completion via convex optimization,” *Found Comput Math.*, vol. 9, no. 6, pp. 717–772, Dec. 2009.

[5] E. J. Candès and T. Tao, “The power of convex relaxation: Near-optimal matrix completion,” *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2053–2080, May 2010.

[6] E. J. Candès and T. Tao, “The power of convex relaxation: Near-optimal matrix completion,” *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2053–2080, May 2010.

[7] A. Aissa-El-Bey, D. Pastor, S. M. A. Sba, 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.

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

[9] S. Mallat and Z. Zhang, “Matching pursuits with time-frequency dictionaries,” *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[10] E. J. Candès and T. Tao, “The power of convex relaxation: Near-optimal matrix completion,” *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2053–2080, May 2010.

[11] E. J. Candès and T. Tao, “The power of convex relaxation: Near-optimal matrix completion,” *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2053–2080, May 2010.

[12] A. Aissa-El-Bey, D. Pastor, S. M. A. Sba, 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.

[13] J. A. Tropp and A. C. Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit,” *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[14] D. Needell and J. A. Tropp, “Cosamp: Iterative signal recovery from incomplete and inaccurate samples,” *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, May 2009.

[15] E. Liu and V. N. Temlyakov, “The orthogonal super greedy algorithm and applications in compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 58, no. 4, pp. 2040–2047, Apr. 2012.

[16] D. L. Donoho, Y. Tsai, I. Drori, and J. L. Starck, “Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit,” *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 1094–1121, Feb. 2012.

[17] J. Wang, S. Kwon, and B. Shim, “Generalized orthogonal matching pursuit,” *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6202–6216, Dec. 2012.

[18] S. S. Kwon, J. Wang, and B. Shim, “Multipath matching pursuit,” *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2986–3001, May 2014.

[19] Y. Kabashima, “A CDMA multiuser detection algorithm on the basis of belief propagation,” *J. Phys. A: Math. Gen.*, vol. 36, no. 43, pp. 11111–11121, Oct. 2003.

[20] D. L. Donoho, A. Maleki, and A. Montanari, “Message-passing algorithms for compressed sensing,” *PNAS*, vol. 106, no. 45, pp. 18914–18919, Nov. 2009.

[21] 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.
C. Thrampoulidis, E. Abbasi, and B. Hassibi, “Precise error analysis of regularized $M$-estimators in high dimensions,” *IEEE Trans. Inf. Theory*, vol. 64, no. 8, pp. 5592–5628, Aug. 2018.

I. B. Atitallah, C. Thrampoulidis, A. Kammoun, T. Y. Al-Naffouri, M. Alouini, and B. Hassibi, “The BOX-LASSO with application to GSSK modulation in massive MIMO systems,” in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Jun. 2017, pp. 1082–1086.

C. Thrampoulidis, W. Xu, and B. Hassibi, “Symbol error rate performance of box-relaxation decoders in massive MIMO,” *IEEE Trans. Signal Process.*, vol. 66, no. 13, pp. 3377–3392, Jul. 2018.

R. Hayakawa and K. Hayashi, “Asymptotic performance of discrete-valued vector reconstruction via box-constrained optimization with sum of $f_1$ regularizers,” *IEEE Trans. Signal Process.*, vol. 68, pp. 4320–4335, 2020.

E. Abbasi, F. Salehi, and B. Hassibi, “Performance analysis of convex data detection in MIMO,” in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2019, pp. 4554–4558.

P. H. Tan, L. K. Rasmussen, and T. J. Lim, “Constrained maximum-likelihood detection in CDMA,” *IEEE Trans. Commun.*, vol. 49, no. 1, pp. 142–153, Jan. 2001.

M. Bayati, M. Lelarge, and A. Montanari, “Universality in polytope phase transitions and message passing algorithms,” *Ann. Appl. Probab.*, vol. 25, no. 2, pp. 753–822, Apr. 2015.

A. Panahi and B. Hassibi, “A universal analysis of large-scale regularized least squares solutions,” in *Proc. Advances in Neural Information Processing Systems*, 2017, pp. 3381–3390.

S. Oymak and J. A. Tropp, “Universality laws for randomized dimension reduction, with applications,” *Inf. Inference*, vol. 7, no. 3, pp. 337–446, Sep. 2018.

D. G. Luenberger and Y. Ye, “Basic Descent Methods,” in *Linear and Nonlinear Programming*, ser. International Series in Operations Research & Management Science. New York, NY: Springer US, 2008, pp. 215–262.

A. Yener, R. D. Yates, and S. Ulukus, “CDMA multiuser detection: A nonlinear programming approach,” *IEEE Trans. Commun.*, vol. 50, no. 6, pp. 1016–1024, Jun. 2002.

L. Li, X. Wang, and G. Wang, “Alternating direction method of multipliers for separable convex optimization of real functions in complex variables,” *Math. Probl. Eng.*, 2015.

R. Hayakawa and K. Hayashi, “Reconstruction of complex discrete-valued vector via convex optimization with sparse regularizers,” *IEEE Access*, vol. 6, pp. 66,499–66,514, 2018.

D. Pérez and Y. Quintana, “A survey on the Weierstrass approximation theorem,” *Divulg. Matemáticas*, vol. 16, no. 1, pp. 231–247, 2008.

Ryo Hayakawa received the bachelor’s degree in engineering, the master’s degree in informatics, and Ph.D. degree in informatics from Kyoto University, Kyoto, Japan, in 2015, 2017, and 2020, respectively. He is currently an Assistant Professor at Graduate School of Engineering Science, Osaka University. He was a Research Fellow (DC1) of the Japan Society for the Promotion of Science (JSPS) from 2017 to 2020. He received the 33rd Telecom System Technology Student Award, APSIPA ASC 2019 Best Special Session Paper Nomination Award, and the 16th IEEE Kansai Section Student Paper Award. His research interests include signal processing and mathematical optimization. He is a member of IEICE.