Discrete-Valued Vector Reconstruction by Optimization with Sum of Sparse Regularizers

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

Kazunori Hayashi
Graduate School of Engineering, Osaka City University
E-mail: kazunori@eng.osaka-cu.ac.jp

***Abstract*—In this paper, we propose a possibly nonconvex optimization problem to reconstruct a discrete-valued vector from its underdetermined linear measurements. The proposed sum of sparse regularizers (SSR) optimization uses the sum of sparse regularizers as a regularizer for the discrete-valued vector. We also propose two proximal splitting algorithms for the SSR optimization problem on the basis of alternating direction method of multipliers (ADMM) and primal-dual splitting (PDS). The ADMM based algorithm can achieve faster convergence, whereas the PDS based algorithm does not require the computation of any inverse matrix. Moreover, we extend the ADMM based approach for the reconstruction of complex discrete-valued vectors. Note that the proposed approach can use any sparse regularizer as long as its proximity operator can be efficiently computed. Simulation results show that the proposed algorithms with nonconvex regularizers can achieve good reconstruction performance.**

***Index Terms*—Discrete-valued vector reconstruction, nonconvex optimization, alternating direction method of multipliers, primal-dual splitting.**

I. Introduction

Discrete-valued vector reconstruction from its linear measurements is a common problem in signal processing for communications systems, e.g., multiple-input multiple-output (MIMO) signal detection [1]–[3] and multiuser detection in machine-to-machine communications [4]. Especially in underdetermined cases, where the number of measurements is less than that of unknown variables, the performance of linear reconstruction methods is severely degraded. Although the maximum likelihood approach can achieve excellent performance, it requires huge computational complexity in large-scale problems. Possible candidates of the low-complexity method are some message passing approaches [5]–[7], which can utilize the discreteness of the unknown vector effectively. However, since they require some assumptions on the measurement matrix, the performance may degrade for general measurement matrices. Other low-complexity approaches have been proposed on the basis of convex optimization [8]–[10]. These optimization problems also take advantage of the discrete nature of the unknown vector as a prior knowledge. However, since all these methods consider convex optimization problems obtained by convex relaxation techniques, the discreteness has not been taken full advantage of.

In this paper, to obtain better reconstruction performance without any explicit assumption on the measurement matrix, we propose a possibly nonconvex optimization problem named sum of sparse regularizers (SSR) optimization. By using the discreteness of the unknown vector and the idea of compressed sensing [11], [12], we utilize the sum of some sparse regularizers as a regularizer for the discrete-valued vector in the proposed SSR optimization. The SSR optimization can be considered as a generalization of the sum of absolute values (SOAV) optimization [10], [13]–[17], and is equivalent to the SOAV optimization when we use the convex ℓ_1 norm as the sparse regularizer. Other than the ℓ_1 norm, we can also use nonconvex regularizers such as the ℓ_p norm ($0 < p < 1$) [18]–[22], the ℓ_0 norm, and the $\ell_1 - \ell_2$ difference [23], [24]. For the SSR optimization, we propose an algorithm on the basis of alternating direction method of multipliers (ADMM) [25]–[28], which is known to achieve fast convergence in general, regardless of the convexity of the cost function. However, the ADMM based algorithm involves the computation of an inverse matrix, which may require prohibitive computational complexity in very large-scale problems. We thus also propose a primal-dual splitting (PDS) [29], [30] based algorithm, which can avoid the computation of the inverse matrix. Moreover, we extend the proposed approach to the reconstruction of discrete-valued vectors in the complex domain, which commonly emerges in the filed of communications. Simulation results show that the proposed algorithms with nonconvex regularizers can achieve better performance than that with the convex ℓ_1 regularizer, which corresponds to the conventional SOAV optimization.

In the rest of the paper, we use the following notations. $\mathbb{R}$ is the set of all real numbers and $\mathbb{C}$ is the set of all complex numbers. We represent the transpose by $(\cdot)^{\mathsf{T}}$, the Hermitian transpose by $(\cdot)^{\mathsf{H}}$, the imaginary unit by j, the $N \times N$ identity matrix by $\boldsymbol{I}_N$, the vector whose elements are all 1 by $\mathbf{1}$, and the vector whose elements are all 0 by $\mathbf{0}$. $\mathrm{Re}\{\cdot\}$ and $\mathrm{Im}\{\cdot\}$ denote the real part and the imaginary part, respectively. For a lower semicontinuous function $\phi : \mathbb{K}^N \to \mathbb{R} \cup \{\infty\}$ ($\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$), the proximity operator of $\phi(\cdot)$ is defined as $\mathrm{prox}_{\phi}(\boldsymbol{u}) = \arg\min_{\boldsymbol{s} \in \mathbb{K}^N} \left\{ \phi(\boldsymbol{s}) + \frac{1}{2} \left\| \boldsymbol{s} - \boldsymbol{u} \right\|_2^2 \right\}$.

This work was supported in part by the Grants-in-Aid for Scientific Research no. 18K04148 and 18H03765 from MEXT, the Grant-in-Aid for JSPS Research Fellow no. 17J07055 from JSPS, and the R&D contract (FY2017-2020) “Wired-and-Wireless Converged Radio Access Network for Massive IoT Traffic” for radio resource enhancement by the Ministry of Internal Affairs and Communications, Japan.

978-9-0827-9703-9/19/$31.00 ©2019 IEEE

II. Problem Statement

We consider the reconstruction of a discrete-valued vector $\boldsymbol{x} = [x_1 \; \cdots \; x_N]^\mathsf{T} \in \mathcal{R}^N \subset \mathbb{R}^N$ from its underdetermined linear measurement given by $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{v} \in \mathbb{R}^M$, where $M < N$. Here, $\mathcal{R} = \{r_1, \dots, r_L\}$ is the set of possible values that the elements of the unknown vector $\boldsymbol{x}$ take, where $L \ll N$. The distribution of x_n is assumed to be known and given by $\Pr(x_n = r_\ell) = p_\ell$ $(\ell = 1, \dots, L)$, where $\sum_{\ell=1}^L p_\ell = 1$. $\boldsymbol{A} \in \mathbb{R}^{M \times N}$ is a measurement matrix and $\boldsymbol{v} \in \mathbb{R}^M$ is an additive noise vector. In MIMO signal detection [1]–[3], for example, N, M, and L correspond to the number of transmit antennas, the number of receive antennas, and the constellation size, respectively.

III. Proposed SSR Optimization Problem

For the reconstruction of $\boldsymbol{x}$ from $\boldsymbol{y}$ and $\boldsymbol{A}$, we propose the SSR optimization problem

$$\underset{\boldsymbol{s} \in \mathbb{R}^N}{\text{minimize}} \left\{ \sum_{\ell=1}^{L} q_\ell h_\ell (\boldsymbol{s} - r_\ell \mathbf{1}) + \frac{\lambda}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{s}\|_2^2 \right\}, \tag{1}$$

where $\lambda \; (> 0)$ is the regularization parameter and $\sum_{\ell=1}^L q_\ell = 1$ $(q_\ell \geq 0)$. The function $h_\ell(\cdot)$ is a sparse regularizer and we assume that its proximity operator can be computed efficiently. The employment of sparse regularizers in the SSR optimization is based on the fact that the vector $\boldsymbol{x} - r_\ell \mathbf{1}$ has some zero elements, which has been utilized in the SOAV optimization [10]. We can thus consider $\sum_{\ell=1}^L q_\ell h_\ell (\boldsymbol{s} - r_\ell \mathbf{1})$ in the objective function as a regularizer for discrete-valued vectors in $\mathcal{R}^N$.

We show some examples of the sparse regularizer $h_\ell(\cdot)$ and the corresponding proximity operators, which are required for the proposed algorithms in Section IV. Note that we consider both convex and nonconvex regularizers in this paper, and we can use any sparse regularizer as far as its proximity operator can be computed.

Example 1 (ℓ_1 Norm). For the ℓ_1 norm based regularizer $h^{(1)}(\boldsymbol{u}) = \|\boldsymbol{u}\|_1 = \sum_{n=1}^N |u_n|$ $(\boldsymbol{u} = [u_1 \; \cdots \; u_N]^\mathsf{T} \in \mathbb{R}^N)$, the proximity operator $\mathrm{prox}_{\gamma h^{(1)}}(\cdot)$ is given by $\left[\mathrm{prox}_{\gamma h^{(1)}}(\boldsymbol{u})\right]_n = \mathrm{sign}(u_n) \max(|u_n| - \gamma, 0)$, where $[\cdot]_n$ denotes the nth element of the vector and $\mathrm{sign}(\cdot)$ is the sign function. The SSR optimization with the ℓ_1 regularizer is equivalent to the SOAV optimization [10].

Example 2 (ℓ_0 Norm). The nonconvex regularizer $h^{(0)}(\boldsymbol{u}) = \|\boldsymbol{u}\|_0$ based on the ℓ_0 norm, i.e., the number of nonzero elements of $\boldsymbol{u}$, has the proximity operator given by $\left[\mathrm{prox}_{\gamma h^{(0)}}(\boldsymbol{u})\right]_n = 0$ when $|u_n| < \sqrt{2\gamma}$, $\left[\mathrm{prox}_{\gamma h^{(0)}}(\boldsymbol{u})\right]_n = \{0, u_n\}$ when $|u_n| = \sqrt{2\gamma}$, and $\left[\mathrm{prox}_{\gamma h^{(0)}}(\boldsymbol{u})\right]_n = u_n$ when $|u_n| > \sqrt{2\gamma}$ $(n = 1, \dots, N)$.

Example 3 (ℓ_p Norm $(0 < p < 1)$). We also consider the nonconvex regularizer $h^{(p)}(\boldsymbol{u}) = \|\boldsymbol{u}\|_p^p = \sum_{n=1}^N |u_n|^p$ with the ℓ_p norm $(0 < p < 1)$. In Fig. 1, we compare the regularizer $(h^{(p)}(s+1) + h^{(p)}(s-1))/2$ in the binary case with $\mathcal{R} = \{-1, 1\}$ for different values of p. From the figure, we can see that the sums of nonconvex regularizers with

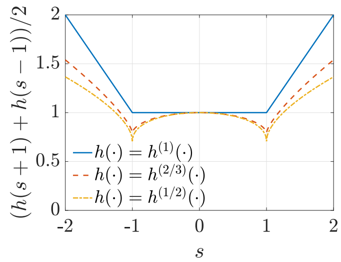

Fig. 1: $(h(s+1) + h(s-1))/2$

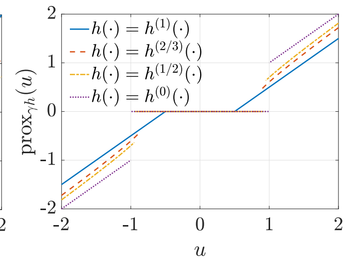

Fig. 2: $\mathrm{prox}_{\gamma h}(u)$ $(\gamma = 0.5)$

$h^{(1/2)}(\cdot)$ and $h^{(2/3)}(\cdot)$ can promote the discrete nature more effectively compared to the convex one with $h^{(1)}(\cdot)$, because the sums of nonconvex regularizers do not have their minimum values for $s \in (-1, 1)$ but only for $s = \pm 1$. The proximity operator of the ℓ_p norm based regularizers has been discussed in [19]–[21]. For arbitrary $p \in (0, 1)$, we can numerically compute the proximity operator, while the proximity operator for $p = 1/2, 2/3$ can be written explicitly. Figure 2 shows the proximity operators of $\gamma h^{(1)}(\cdot)$, $\gamma h^{(2/3)}(\cdot)$, $\gamma h^{(1/2)}(\cdot)$, and $\gamma h^{(0)}(\cdot)$ $(\gamma = 0.5)$. As we can see from the figure, the proximity operators of the nonconvex regularizers are not continuous.

Example 4 ($\ell_1 - \ell_2$ Difference). The nonconvex regularizer $h^{(1-2)}(\boldsymbol{u}) = \|\boldsymbol{u}\|_1 - \|\boldsymbol{u}\|_2$ based on the $\ell_1 - \ell_2$ difference has been proposed for compressed sensing [23], [24]. The proximity operator of $h^{(1-2)}(\cdot)$ can be computed with Lemma 1 in [24] or Proposition 7.1 in [31].

IV. Proximal Splitting Algorithms for SSR Optimization

A. ADMM Based Algorithm

We can rewrite the optimization problem of the SSR optimization (1) with new variables $\boldsymbol{z}_1, \dots, \boldsymbol{z}_L \in \mathbb{R}^N$ as

$$\begin{aligned} &\underset{\boldsymbol{s}, \boldsymbol{z}_1, \dots, \boldsymbol{z}_L \in \mathbb{R}^N}{\text{minimize}} \left\{ \sum_{\ell=1}^{L} q_\ell h_\ell (\boldsymbol{z}_\ell - r_\ell \mathbf{1}) + \frac{\lambda}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{s}\|_2^2 \right\} \\ &\text{subject to} \quad \boldsymbol{s} = \boldsymbol{z}_\ell \; (\ell = 1, \dots, L), \end{aligned} \tag{2}$$

which is further rewritten as

$$\underset{\boldsymbol{s} \in \mathbb{R}^N, \boldsymbol{z} \in \mathbb{R}^{LN}}{\text{minimize}} \{f(\boldsymbol{s}) + g(\boldsymbol{z})\} \quad \text{subject to} \quad \boldsymbol{\Phi}\boldsymbol{s} = \boldsymbol{z}. \tag{3}$$

Here, $\boldsymbol{z} = \left[\boldsymbol{z}_1^\mathsf{T} \; \cdots \; \boldsymbol{z}_L^\mathsf{T}\right]^\mathsf{T} \in \mathbb{R}^{LN}$, $\boldsymbol{\Phi} = [\boldsymbol{I}_N \; \cdots \; \boldsymbol{I}_N]^\mathsf{T}$, $f(\boldsymbol{s}) = \frac{\lambda}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{s}\|_2^2$, and $g(\boldsymbol{z}) = \sum_{\ell=1}^L q_\ell h_\ell (\boldsymbol{z}_\ell - r_\ell \mathbf{1})$.

We derive the proposed algorithm based on ADMM. The update equations of ADMM for (3) are given by

$$\boldsymbol{s}^{k+1} = \arg\min_{\boldsymbol{s} \in \mathbb{C}^N} \left\{ f(\boldsymbol{s}) + \frac{\rho}{2} \left\| \boldsymbol{\Phi}\boldsymbol{s} - \boldsymbol{z}^k + \boldsymbol{w}^k \right\|_2^2 \right\}, \tag{4}$$

$$\boldsymbol{z}^{k+1} = \arg\min_{\boldsymbol{z} \in \mathbb{C}^{LN}} \left\{ g(\boldsymbol{z}) + \frac{\rho}{2} \left\| \boldsymbol{\Phi}\boldsymbol{s}^{k+1} - \boldsymbol{z} + \boldsymbol{w}^k \right\|_2^2 \right\}, \tag{5}$$

$$\boldsymbol{w}^{k+1} = \boldsymbol{w}^k + \boldsymbol{\Phi}\boldsymbol{s}^{k+1} - \boldsymbol{z}^{k+1}, \tag{6}$$

where k is the iteration index, $\rho \; (> 0)$ is a parameter, and $\boldsymbol{w}^k \in \mathbb{R}^{LN}$. From $\frac{\partial}{\partial \boldsymbol{s}^\mathsf{T}} \left\{ f(\boldsymbol{s}) + \frac{\rho}{2} \left\| \boldsymbol{\Phi}\boldsymbol{s} - \boldsymbol{z}^k + \boldsymbol{w}^k \right\|_2^2 \right\} =$

Algorithm 1 ADMM-SSR: ADMM Based Algorithm for (1)

Input: $\boldsymbol{y} \in \mathbb{R}^M, \boldsymbol{A} \in \mathbb{R}^{M \times N}$
Output: $\hat{\boldsymbol{x}} \in \mathcal{R}^N$
Fix $\rho > 0$, $\boldsymbol{z}^0 \in \mathbb{R}^{NL}$, and $\boldsymbol{w}^0 \in \mathbb{R}^{NL}$
for $k = 0$ to $K-1$ **do**
 $\boldsymbol{s}^{k+1} = \left(\rho L \boldsymbol{I}_N + \lambda \boldsymbol{A}^\top \boldsymbol{A}\right)^{-1} \cdot \left(\rho \sum_{\ell=1}^{L} \left(\boldsymbol{z}_\ell^k - \boldsymbol{w}_\ell^k\right) + \lambda \boldsymbol{A}^\top \boldsymbol{y}\right)$
 $\boldsymbol{z}_\ell^{k+1} = r_\ell \boldsymbol{1} + \mathrm{prox}_{\frac{q_\ell}{\rho} h_\ell}\left(\boldsymbol{s}^{k+1} + \boldsymbol{w}_\ell^k - r_\ell \boldsymbol{1}\right)$ $(\ell = 1, \ldots, L)$
 $\boldsymbol{w}_\ell^{k+1} = \boldsymbol{w}_\ell^k + \boldsymbol{s}^{k+1} - \boldsymbol{z}_\ell^{k+1}$ $(\ell = 1, \ldots, L)$
end for
$\hat{\boldsymbol{x}} = \mathcal{Q}(\boldsymbol{s}^K)$

Algorithm 2 PDS-SSR: PDS Based Algorithm for (1)

Input: $\boldsymbol{y} \in \mathbb{R}^M, \boldsymbol{A} \in \mathbb{R}^{M \times N}$
Output: $\hat{\boldsymbol{x}} \in \mathcal{R}^N$
Fix $\rho_1 > 0$, $\rho_2 > 0$, $\boldsymbol{s}^0 \in \mathbb{R}^N$, and $\boldsymbol{w}^0 \in \mathbb{R}^{NL}$
for $k = 0$ to $K-1$ **do**
 $\boldsymbol{s}^{k+1} = \boldsymbol{s}^k - \rho_1 \left(\lambda \boldsymbol{A}^\top \left(\boldsymbol{A}\boldsymbol{s}^k - \boldsymbol{y}\right) + \sum_{\ell=1}^{L} \boldsymbol{w}_\ell^k\right)$
 $\boldsymbol{z}_\ell^{k+1} = \boldsymbol{w}_\ell^k + \rho_2 \left(2\boldsymbol{s}^{k+1} - \boldsymbol{s}^k\right)$ $(\ell = 1, \ldots, L)$
 $\boldsymbol{w}_\ell^{k+1} = \boldsymbol{z}_\ell^{k+1} - \rho_2 \left(r_\ell \boldsymbol{1} + \mathrm{prox}_{\frac{q_\ell}{\rho_2} h_\ell}\left(\frac{\boldsymbol{z}_\ell^{k+1}}{\rho_2} - r_\ell \boldsymbol{1}\right)\right)$ $(\ell = 1, \ldots, L)$
end for
$\hat{\boldsymbol{x}} = \mathcal{Q}(\boldsymbol{s}^K)$

$\boldsymbol{0}$, the update of $\boldsymbol{s}^k$ in (4) can be written as $\boldsymbol{s}^{k+1} = \left(\rho L \boldsymbol{I}_N + \lambda \boldsymbol{A}^\top \boldsymbol{A}\right)^{-1} \left(\rho \sum_{\ell=1}^{L} \left(\boldsymbol{z}_\ell^k - \boldsymbol{w}_\ell^k\right) + \lambda \boldsymbol{A}^\top \boldsymbol{y}\right)$, where $\boldsymbol{z}_\ell^k \in \mathbb{R}^N$ and $\boldsymbol{w}_\ell^k \in \mathbb{R}^N$ $(\ell = 1, \ldots, L)$ are subvectors of $\boldsymbol{z}^k = \left[\boldsymbol{z}_1^{k\top} \cdots \boldsymbol{z}_L^{k\top}\right]^\top$ and $\boldsymbol{w}^k = \left[\boldsymbol{w}_1^{k\top} \cdots \boldsymbol{w}_L^{k\top}\right]^\top$, respectively. The update of $\boldsymbol{z}^k$ in (5) can be written as

$$\boldsymbol{z}^{k+1} = \mathrm{prox}_{\frac{1}{\rho} g}\left(\boldsymbol{\Phi}\boldsymbol{s}^{k+1} + \boldsymbol{w}^k\right) \tag{7}$$

$$= \begin{bmatrix} r_1 \boldsymbol{1} + \mathrm{prox}_{\frac{q_1}{\rho} h_1}\left(\boldsymbol{s}^{k+1} + \boldsymbol{w}_1^k - r_1 \boldsymbol{1}\right) \\ \vdots \\ r_L \boldsymbol{1} + \mathrm{prox}_{\frac{q_L}{\rho} h_L}\left(\boldsymbol{s}^{k+1} + \boldsymbol{w}_L^k - r_L \boldsymbol{1}\right) \end{bmatrix}, \tag{8}$$

because the function $g(\cdot)$ is separable as $g(\boldsymbol{z}) = \sum_{\ell=1}^{L} q_\ell h_\ell (\boldsymbol{z}_\ell - r_\ell \boldsymbol{1})$. We also use the property of proximity operator for translation [27] in (8).

We summarize the ADMM based algorithm for the SSR optimization (1) as ADMM-SSR in Algorithm 1, where $\mathcal{Q}(\cdot)$ denotes the element-wise quantization operator which maps the input to its nearest value in $\mathcal{R}$. One of the advantages of ADMM-SSR is that we do not require the proximity operator of the whole regularizer $\sum_{\ell=1}^{L} q_\ell h_\ell (\boldsymbol{s} - r_\ell \boldsymbol{1})$ and we can implement ADMM-SSR as long as the proximity operator of $h_\ell(\cdot)$ can be calculated as in Examples 1–4. The computational complexity is dominated by the inverse matrix $\left(\rho L \boldsymbol{I}_N + \lambda \boldsymbol{A}^\top \boldsymbol{A}\right)^{-1}$, which usually requires $\mathcal{O}(N^3)$ complexity [32, Ch. 11].

B. PDS Based Algorithm

As we have mentioned in the previous subsection, ADMM-SSR requires the computation of the inverse matrix, which may require prohibitive computational complexity for very large-scale problems. To overcome this problem, we also propose an algorithm based on primal-dual splitting [30], which can avoid the computation of the inverse matrix.

We first rewrite the SSR optimization problem (1) as

$$\underset{\boldsymbol{s} \in \mathbb{R}^N}{\mathrm{minimize}} \ \left\{f(\boldsymbol{s}) + g(\boldsymbol{\Phi}\boldsymbol{s})\right\}, \tag{9}$$

where $f(\cdot)$ and $g(\cdot)$ are defined below (3). PDS is applicable to the problem of the form (9) and is given by

$$\boldsymbol{s}^{k+1} = \boldsymbol{s}^k - \rho_1 \left(\nabla f\left(\boldsymbol{s}^k\right) + \boldsymbol{\Phi}^\top \boldsymbol{w}^k\right), \tag{10}$$

$$\boldsymbol{z}^{k+1} = \boldsymbol{w}^k + \rho_2 \boldsymbol{\Phi}\left(2\boldsymbol{s}^{k+1} - \boldsymbol{s}^k\right), \tag{11}$$

$$\boldsymbol{w}^{k+1} = \mathrm{prox}_{\rho_2 g^*}\left(\boldsymbol{z}^{k+1}\right), \tag{12}$$

where ρ_1, ρ_2 (> 0) are the parameters, $\nabla f(\cdot)$ denotes the gradient of the function $f(\cdot)$, and $g^*(\cdot)$ represents the convex conjugate of $g(\cdot)$. The update of $\boldsymbol{s}^k$ in (10) can be written as $\boldsymbol{s}^{k+1} = \boldsymbol{s}^k - \rho_1 \left(\lambda \boldsymbol{A}^\top \left(\boldsymbol{A}\boldsymbol{s}^k - \boldsymbol{y}\right) + \sum_{\ell=1}^{L} \boldsymbol{w}_\ell^k\right)$ because $\nabla f(\boldsymbol{s}) = \lambda \boldsymbol{A}^\top (\boldsymbol{A}\boldsymbol{s} - \boldsymbol{y})$. The proximity operator $\mathrm{prox}_{\rho_2 g^*}(\cdot)$ in (12) is expressed as $\mathrm{prox}_{\rho_2 g^*}(\boldsymbol{u}) = \boldsymbol{u} - \rho_2 \mathrm{prox}_{g/\rho_2}(\boldsymbol{u}/\rho_2)$. Hence, from (8) and (12), we can update $\boldsymbol{w}_\ell^{k+1}$ as $\boldsymbol{w}_\ell^{k+1} = \boldsymbol{z}_\ell^{k+1} - \rho_2 \left(r_\ell \boldsymbol{1} + \mathrm{prox}_{\frac{q_\ell}{\rho_2} h_\ell}\left(\frac{\boldsymbol{z}_\ell^{k+1}}{\rho_2} - r_\ell \boldsymbol{1}\right)\right)$.

We summarize the PDS based algorithm named PDS-SSR in Algorithm 2. As is the case with ADMM-SSR, PDS-SSR also requires only the proximity operator of $h_\ell(\cdot)$. Since PDS-SSR computes only the addition of vectors and the multiplication of a matrix and a vector, it requires $\mathcal{O}(MN)$ complexity [32, Ch. 6], which is lower than that of ADMM-SSR.

C. Convergence of Proposed Algorithms

The convergence of the proposed algorithms depends on the convexity of the sparse regularizer $h_\ell(\cdot)$. When $h_1(\cdot), \ldots, h_L(\cdot)$ are all convex, the objective function of the SSR optimization is also convex. In this case, the sequence $\left\{\boldsymbol{s}^k\right\}$ obtained by ADMM-SSR converges to the optimizer of the problem from the general result for ADMM [26]. From Theorem 3.1 in [30], the sequence $\left\{\boldsymbol{s}^k\right\}$ obtained by PDS-SSR also converges if the parameters ρ_1 and ρ_2 satisfy $1/\rho_1 - \rho_2 L \geq \lambda \left\|\boldsymbol{A}^\top \boldsymbol{A}\right\|_2 / 2$. When $h_\ell(\cdot)$ is nonconvex, however, the convergence to the global optimizer is not guaranteed in general. Although some convergence property have been proved under several assumptions [33]–[37], their results cannot be directly used for the proposed algorithms.

V. Extension to Complex-Valued Case

In this section, we extend the proposed method to the reconstruction of the complex-valued vector $\tilde{\boldsymbol{x}} \in \mathcal{C}^N \subset \mathbb{C}^N$, where $\mathcal{C} = \{c_1, \ldots, c_L\}$ denotes the set of possible complex values.

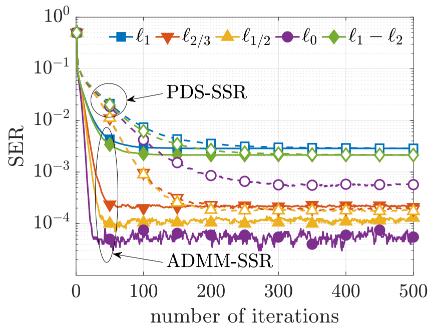

Fig. 3: SER versus number of iterations for binary vectors ($(N, M) = (200, 160)$, SNR = 15 dB)

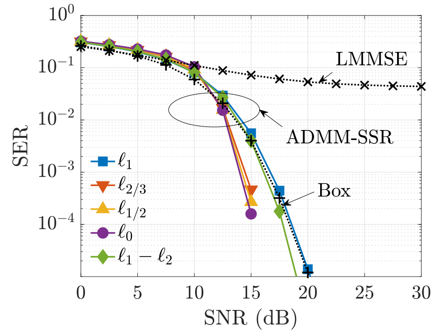

Fig. 4: SER versus SNR for binary vectors ($(N, M) = (200, 150)$)

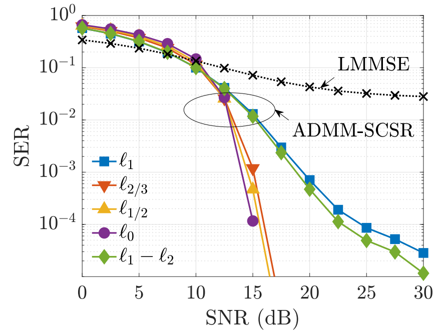

Fig. 5: SER versus SNR for complex-valued vectors ($(N, M) = (200, 160)$)

The measurement vector $\tilde{\boldsymbol{y}} \in \mathbb{C}^M$ is given by $\tilde{\boldsymbol{y}} = \tilde{\boldsymbol{A}}\tilde{\boldsymbol{x}} + \tilde{\boldsymbol{v}}$, where $\tilde{\boldsymbol{A}} \in \mathbb{C}^{M\times N}$ and $\tilde{\boldsymbol{v}} \in \mathbb{C}^M$ are a measurement matrix and an additive noise vector, respectively.

For the reconstruction of the complex discrete-valued vector, we extend the SSR optimization (1) to the problem

$$\underset{\tilde{\boldsymbol{s}}\in\mathbb{C}^N}{\text{minimize}} \left\{ \sum_{\ell=1}^{L} q_\ell \tilde{h}_\ell \left(\tilde{\boldsymbol{s}} - c_\ell \boldsymbol{1}\right) + \lambda \left\| \tilde{\boldsymbol{y}} - \tilde{\boldsymbol{A}}\tilde{\boldsymbol{s}} \right\|_2^2 \right\}, \quad (13)$$

which is referred to as the sum of complex sparse regularizers (SCSR) optimization hereafter. The function $\tilde{h}_\ell(\cdot)$ is a sparse regularizer for the complex-valued sparse vector. The SCSR optimization with the ℓ_1 regularizer has been proposed in [38], whereas we consider nonconvex regularizers as well in this paper. As discussed in [38], the optimization in the complex domain is more suitable than that in the real domain when the real part and the imaginary part of the unknown vector are not independent.

For the SCSR optimization (13), we newly consider two kinds of sparse regularizers as the candidates of $\tilde{h}_\ell(\cdot)$. For example, as the regularizers based on the ℓ_p norm, we define $\tilde{h}_\star^{(p)}(\tilde{\boldsymbol{u}}) = \sum_{n=1}^N |\tilde{u}_n|^p$ and $\tilde{h}_{\star\star}^{(p)}(\tilde{\boldsymbol{u}}) = \sum_{n=1}^N (|\mathrm{Re}\{\tilde{u}_n\}|^p + |\mathrm{Im}\{\tilde{u}_n\}|^p)$, where $\tilde{\boldsymbol{u}} = [\tilde{u}_1 \ \cdots \ \tilde{u}_N]^\mathsf{T} \in \mathbb{C}^N$. The first regularizer $\tilde{h}_\star^{(p)}(\cdot)$ is based on the modulus for complex numbers, whereas the second one $\tilde{h}_{\star\star}^{(p)}(\cdot)$ treats the real part and the imaginary part independently. We also define $h_\star^{(1)}(\cdot)$, $h_\star^{(0)}(\cdot)$, $h_\star^{(1-2)}(\cdot)$, $h_{\star\star}^{(1)}(\cdot)$, $h_{\star\star}^{(0)}(\cdot)$, and $h_{\star\star}^{(1-2)}(\cdot)$ in the same manner. The proximity operator of $\gamma\tilde{h}_\star(\cdot)$ ($\tilde{h}_\star(\cdot) = \tilde{h}_\star^{(1)}(\cdot), \tilde{h}_\star^{(0)}(\cdot), \tilde{h}_\star^{(p)}(\cdot), \tilde{h}_\star^{(1-2)}(\cdot)$) in the complex domain can be written with that of the corresponding regularizer $\gamma h(\cdot)$ ($h(\cdot) = h^{(1)}(\cdot), h^{(0)}(\cdot), h^{(p)}(\cdot), h^{(1-2)}(\cdot)$, respectively) in the real domain. Note that $\tilde{h}_\star(\cdot)$ satisfy $\tilde{h}_\star(\tilde{\boldsymbol{u}}) = h(|\tilde{\boldsymbol{u}}|)$, where we define $|\tilde{\boldsymbol{u}}| = [|\tilde{u}_1| \ \cdots \ |\tilde{u}_N|]^\mathsf{T}$. By using this property, the proximity operator of $\gamma\tilde{h}_\star(\cdot)$ can be derived as $\left[\mathrm{prox}_{\gamma\tilde{h}_\star}(\tilde{\boldsymbol{u}})\right]_n = \left[\mathrm{prox}_{\gamma h}(|\tilde{\boldsymbol{u}}|)\right]_n \dfrac{\tilde{u}_n}{|\tilde{u}_n|}$ with a simple manipulation. The proximity operator of $\gamma\tilde{h}_{\star\star}(\cdot)$ can also be written with the corresponding proximity operator $\mathrm{prox}_{\gamma h}(\cdot)$. Since we have $\tilde{h}_{\star\star}(\tilde{\boldsymbol{u}}) = h(\boldsymbol{u}_\mathrm{R}) + h(\boldsymbol{u}_\mathrm{I})$ from the definition, the proximity operator can be written as $\left[\mathrm{prox}_{\gamma\tilde{h}_{\star\star}}(\tilde{\boldsymbol{u}})\right]_n = \left[\mathrm{prox}_{\gamma h}(\boldsymbol{u}_\mathrm{R})\right]_n + j \cdot \left[\mathrm{prox}_{\gamma h}(\boldsymbol{u}_\mathrm{I})\right]_n$ by using a similar approach to [38], where $\boldsymbol{u}_\mathrm{R} = \mathrm{Re}\{\tilde{\boldsymbol{u}}\}$ and $\boldsymbol{u}_\mathrm{I} = \mathrm{Im}\{\tilde{\boldsymbol{u}}\}$.

Since ADMM with complex-valued variables have been discussed in the literature [38], [39], we propose the ADMM based algorithm for the SCSR optimization (13) by using the approach in [38]. The resultant algorithm is obtained by replacing $\mathbb{R}$, $(\cdot)^\mathsf{T}$, r_ℓ, and $\mathrm{prox}_{\frac{q_\ell}{\rho}h_\ell}(\cdot)$ in Algorithm 1 with $\mathbb{C}$, $(\cdot)^\mathsf{H}$, c_ℓ, and $\mathrm{prox}_{\frac{q_\ell}{2\rho}\tilde{h}_\ell}(\cdot)$, respectively.

VI. Simulation Results

In this section, we evaluate the performance of the proposed algorithms. In the simulation, the measurement matrix is composed of independent and identically distributed (i.i.d.) Gaussian variables with zero mean and unit variance. We also assume the zero mean additive white Gaussian noise. The initialization is given by $\boldsymbol{z}^0 = \boldsymbol{w}^0 = \boldsymbol{0}$ in ADMM-SSR and $\boldsymbol{s}^0 = \boldsymbol{0}, \boldsymbol{w}^0 = \boldsymbol{0}$ in PDS-SSR. Other parameters such as λ and ρ are chosen to achieve good performance in the simulation.

Figure 3 shows the symbol error rate (SER) versus number of iterations for the binary vector with $(r_1, r_2) = (-1, 1)$ and $(p_1, p_2) = (1/2, 1/2)$. The result is obtained by averaging the SER over $2{,}000$ independent realizations of the measurement matrix. The problem size is $(N, M) = (200, 160)$ and the signal-to-noise ratio (SNR) is 15 dB. The parameters are set as $q_1 = q_2 = 1/2$, $\lambda = 0.05$, $\rho = 3$, $\rho_1 = 2/\left(\lambda \left\|\boldsymbol{A}^\mathsf{T}\boldsymbol{A}\right\|_2 + 4\right)$, and $\rho_2 = 1/2$. In Fig. 3, we denote the sparse regularizers based on the ℓ_1 norm, the $\ell_{2/3}$ norm, the $\ell_{1/2}$ norm, the ℓ_0 norm, and the $\ell_1 - \ell_2$ difference by 'ℓ_1,' '$\ell_{2/3}$,' '$\ell_{1/2}$,' 'ℓ_0,' and '$\ell_1 - \ell_2$,' respectively. We can see that ADMM-SSR and PDS-SSR converge to the same SER when we use the convex ℓ_1 regularizer. The proposed algorithms with nonconvex regularizers, especially with the ℓ_p and ℓ_0 norms, can achieve much better SER performance.

In Fig. 4, we show the SER of ADMM-SSR versus SNR for the binary vector reconstruction with $(N, M) = (200, 150)$. For comparison, we also show the performance of the linear minimum mean-square-error method and the box relaxation method [8] as 'LMMSE' and 'Box,' respectively. The parameters in ADMM-SSR are set as $\lambda = 0.05$, $\rho = 3$, and $K = 300$. The nonconvex regularizers can outperform the convex ℓ_1 regularizer and the box relaxation method.

Figure 5 shows the SER versus SNR for the reconstruction of complex discrete-valued vectors with $(N, M) = (200, 160)$. The distribution of the unknown complex variable is given by $(c_1, c_2, c_3, c_4, c_5) = (0, 1 + j, -1 + j, -1 - j, 1 - j)$ and $(p_1, p_2, p_3, p_4, p_5) = (0.6, 0.1, 0.1, 0.1, 0.1)$, where $p_\ell = \Pr(x_n = c_\ell)$ $(\ell = 1, \ldots, 5)$. We use the sparse regularizers given by $\tilde{h}_1(\cdot) = \tilde{h}_\star(\cdot)$ and $\tilde{h}_\ell(\cdot) = \tilde{h}_{\star\star}(\cdot)$ $(\ell = 2, \ldots, 5)$ because the real part becomes zero only when the imaginary part is zero in this case. The parameters of the proposed algorithm are set as $q_\ell = p_\ell$, $\lambda = 0.05$, $\rho = 3$, and $K = 300$. From the figure, we can see that the proposed approach with nonconvex regularizer can achieve good performance even for the reconstruction of the complex discrete-valued vector.

VII. Conclusion

In this paper, we have proposed possibly nonconvex optimization problems for the discrete-valued vector reconstruction in both real- and complex-valued cases. The proposed method utilizes the sum of sparse regularizers as the regularizer for the discrete-valued vector. Simulation results show that the proposed algorithms with nonconvex regularizers can achieve better performance than that with the convex ℓ_1 regularizer. Future work includes the PDS based algorithm for the SCSR optimization and the convergence analysis for the proposed algorithms with nonconvex regularizers.

References

[1] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

[2] A. Chockalingam and B. S. Rajan, *Large MIMO systems*. Cambridge University Press, 2014.

[3] S. Yang and L. Hanzo, "Fifty years of MIMO detection: The road to large-scale MIMOs," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 1941–1988, Fourthquarter 2015.

[4] S. Verdú, *Multiuser Detection*. Cambridge University Press, 1998.

[5] C. Jeon, R. Ghods, A. Maleki, and C. Studer, "Optimality of large MIMO detection via approximate message passing," in *Proc. IEEE ISIT*, Jun. 2015.

[6] R. Hayakawa and K. Hayashi, "Discreteness-aware approximate message passing for discrete-valued vector reconstruction," *IEEE Trans. Signal Process.*, vol. 66, no. 24, pp. 6443–6457, Dec. 2018.

[7] J. Céspedes, P. M. Olmos, M. Sánchez-Fernández, and F. Perez-Cruz, "Expectation propagation detection for high-order high-dimensional MIMO systems," *IEEE Trans. Commun.*, vol. 62, no. 8, pp. 2840–2849, Aug. 2014.

[8] 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.

[9] A. Aïssa-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] 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, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[12] 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.

[13] 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.

[14] 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.

[15] R. Hayakawa and K. Hayashi, "Convex optimization-based signal detection for massive overloaded MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 11, pp. 7080–7091, Nov. 2017.

[16] R. Hayakawa and K. Hayashi, "Error recovery for massive MIMO signal detection via reconstruction of discrete-valued sparse vector," *IEICE Trans. Fundamentals*, vol. E100-A, no. 12, pp. 2671–2679, Dec. 2017.

[17] R. Hayakawa and K. Hayashi, "Discreteness-aware decoding for overloaded non-orthogonal STBCs via convex optimization," *IEEE Commun. Lett.*, vol. 22, no. 10, pp. 2080–2083, Oct. 2018.

[18] Z. Xu, H. Zhang, Y. Wang, X. Chang, and Y. Liang, "$L_{1/2}$ regularization," *Sci. China Inf. Sci.*, vol. 53, no. 6, pp. 1159–1169, Jun 2010.

[19] Z. Xu, X. Chang, F. Xu, and H. Zhang, "$L_{1/2}$ regularization: A thresholding representation theory and a fast solver," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 7, pp. 1013–1027, Jul. 2012.

[20] Y. Zhang and W. Ye, "$L_{2/3}$ regularization: Convergence of iterative thresholding algorithm," *J. Vis. Commun. Image Represent.*, vol. 33, pp. 350–357, 2015.

[21] F. Chen, L. Shen, and B. W. Suter, "Computing the proximity operator of the ℓ_p norm with $0 < p < 1$," *IET Signal Process.*, vol. 10, no. 5, pp. 557–565, Jun. 2016.

[22] Z. Li, S. Ding, Y. Li, Z. Yang, S. Xie, and W. Chen, "Manifold optimization-based analysis dictionary learning with an $\ell_{1/2}$-norm regularizer," *Neural Networks*, vol. 98, pp. 212–222, 2018.

[23] P. Yin, Y. Lou, Q. He, and J. Xin, "Minimization of ℓ_{1-2} for compressed sensing," *SIAM J. Sci. Comput.*, vol. 37, no. 1, pp. A536–A563, Feb. 2015.

[24] Y. Lou and M. Yan, "Fast L1–L2 minimization via a proximal operator," *J. Sci. Comput.*, vol. 74, no. 2, pp. 767–785, Feb. 2018.

[25] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximation," *Comput. Math. Appl.*, vol. 2, no. 1, pp. 17–40, 1976.

[26] J. Eckstein and D. P. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Program.*, vol. 55, no. 1, pp. 293–318, Apr. 1992.

[27] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-point algorithms for inverse problems in science and engineering*. Springer, 2011.

[28] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[29] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imaging Vision*, vol. 40, no. 1, pp. 120–145, May 2011.

[30] L. Condat, "A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms," *J. Optim. Theory Appl.*, vol. 158, no. 2, pp. 460–479, Aug. 2013.

[31] T. Liu and T. K. Pong, "Further properties of the forward–backward envelope with applications to difference-of-convex programming," *Comput. Optim. Appl.*, vol. 67, no. 3, pp. 489–520, Jul. 2017.

[32] S. Boyd and L. Vandenberghe, *Introduction to Applied Linear Algebra: Vectors, Matrices, and Least Squares*. Cambridge University Press, 2018.

[33] H. Attouch, J. Bolte, and B. F. Svaiter, "Convergence of descent methods for semi-algebraic and tame problems: Proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods," *Math. Program. Ser. A*, vol. 137, no. 1, pp. 91–129, Feb. 2013.

[34] G. Li and T. Pong, "Global convergence of splitting methods for nonconvex composite optimization," *SIAM J. Optim.*, vol. 25, no. 4, pp. 2434–2460, 2015.

[35] M. Hong, Z. Luo, and M. Razaviyayn, "Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems," *SIAM J. Optim.*, vol. 26, no. 1, pp. 337–364, 2016.

[36] D. Hajinezhad, M. Hong, T. Zhao, and Z. Wang, "NESTT: A nonconvex primal-dual splitting method for distributed and stochastic optimization," in *Proc. NIPS*, 2016.

[37] Y. Wang, W. Yin, and J. Zeng, "Global convergence of ADMM in nonconvex nonsmooth optimization," *J. Sci. Comput.*, Jun. 2018.

[38] R. Hayakawa and K. Hayashi, "Reconstruction of complex discrete-valued vector via convex optimization with sparse regularizers," *IEEE Access*, vol. 6, pp. 66 499–66 512, Dec. 2018.

[39] L. Li, X. Wang, and G. Wang, "Alternating direction method of multipliers for separable convex optimization of real functions in complex variables," *Mathematical Problems in Engineering*, vol. 2015, 2015.