

# Hyperspectral Unmixing with Row-Sparsity Enhancement: A Difference-of-Convex Approach

Kazuki Naganuma\* and Shunsuke Ono\*

\* Institute of Science Tokyo, Tokyo

E-mail: naganuma.k.491a@m.isct.ac.jp; ono.s.5af2@m.isct.ac.jp

**Abstract**—This paper proposes a new hyperspectral unmixing method that enhances the row sparsity of an abundance matrix via a difference-of-convex (DC) approach. One promising unmixing technique is to promote the row sparsity by designing and solving optimization problems. However, existing methods are unable to both fully enhance the row sparsity and guarantee the convergence of their optimization algorithms, resulting in performance degradation. To address this limitation, we introduce a DC approach to unmixing. First, we design a DC function that promotes the row sparsity while preserving the energies of some row vectors of an abundance matrix. Then, incorporating the DC function, we formulate unmixing as a constrained DC optimization problem. In addition, we develop an algorithm to solve the optimization problem based on the proximal linearized DC algorithm and the preconditioned primal-dual splitting algorithm and guarantee its convergence to a critical point of the problem. Experiments on synthetic and real HS images show that our method is superior to existing methods.

## I. INTRODUCTION

Hyperspectral (HS) images are cube data consisting of rich spectral information and have various applications, such as ecology, mineralogy, and agriculture [1]–[5]. Due to the trade-off between spatial and spectral resolution, HS sensors do not have sufficient spatial resolutions, resulting in containing multiple components (called endmembers) in a pixel [6], which is referred to as a mixel. The process of decomposing the mixel into endmembers and their abundances is called unmixing. Unmixing has been actively studied in the remote sensing field because it is essential for HS image analysis [7]–[9].

One major category of unmixing tasks is non-blind unmixing, which estimates abundances using an endmember library consisting of all possible pre-collected spectral signatures that may appear in a target HS image. This indicates that the HS image often contains a smaller number of spectral signatures of the endmember library, i.e., many of the corresponding abundances are zeros. Therefore, the matrix created by arranging the abundances corresponding to each endmember in the row direction is exactly row sparse (we call the matrix *abundance matrix*).

To capture the exact row sparsity, non-blind unmixing methods often solve optimization problems that incorporate a function promoting the row sparsity<sup>1</sup>. Many of these methods adopt a mixed  $\ell_{1,2}$ -type norm [10]–[15]. Since a mixed  $\ell_{1,2}$ -type norm is convex, these methods formulate unmixing as a convex optimization problem and thus can obtain stable results. As a more advanced method, the authors in [16] employ a

mixed  $\ell_{0,2}$  pseudo norm, a nonconvex function that more appropriately models row sparsity.

However, the existing methods have the following limitations. The convex approach is weak in enhancing the row sparsity because minimizing the mixed  $\ell_{1,2}$ -type norm of an abundance matrix almost equally reduces all the  $\ell_2$  norms of the row vectors of the matrix. In the nonconvex approach, the obtained abundance matrix is not fully row sparse because the sequence generated by the optimization algorithm does not converge to a solution of the optimization problem involving a mixed  $\ell_{0,2}$  pseudo norm. Therefore, to improve the unmixing performance, the following question arises: *Could we fully enhance the row sparsity while guaranteeing the convergence of the optimization algorithm?*

Based on the above discussion, we introduce a difference-of-convex (DC) approach to unmixing. The DC handles the difference of convex functions and convex functions, and thus provides the improvements of many signal processing performance [17]–[21], including the enhancement of (group) sparsity. In addition, a DC algorithm generates the sequence that converges to a critical point of a DC optimization problem. Our contributions are listed as follows.

- We design a new DC function that promotes row sparsity while preserving the energies of some row vectors of a matrix.
- We formulate unmixing as a constrained DC optimization problem that incorporates the DC function.
- Based on the proximal linearized DC algorithm (PLDC) [22] and the preconditioned primal-dual splitting algorithm (PPDS) [23], [24], we develop an algorithm that generates the sequence converging to a critical point of the DC optimization problem.

Finally, we illustrate the effectiveness of our method through unmixing experiments using both synthetic and real datasets.

## II. PRELIMINARIES

### A. Notations

Vectors and matrices are respectively denoted by lowercase and capitalized boldface letters (e.g.,  $\mathbf{x}$  and  $\mathbf{X}$ ), and the element at the  $i$ -th row and  $j$ -th column of matrix  $\mathbf{X}$  is

<sup>1</sup>A deep neural network-based method has also been proposed [25]. However, as experimentally shown in Section IV, row sparsity-based methods are more suitable for non-blind unmixing.

denoted by  $X_{i,j}$  or  $[X]_{i,j}$ . For matrices  $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{m \times n}$ , the inner product  $\langle \mathbf{X}, \mathbf{Y} \rangle$ , the  $\ell_1$ -norm  $\|\mathbf{X}\|_1$ , the Frobenius norm  $\|\mathbf{X}\|_F$ , the mixed  $\ell_{1,2}$ -norm grouped by row  $\|\mathbf{X}\|_{1,2}$  are defined by  $\langle \mathbf{X}, \mathbf{Y} \rangle = \sum_{i,j} X_{i,j} Y_{i,j}$ ,  $\|\mathbf{X}\|_1 = \sum_{i,j} |X_{i,j}|$ ,  $\|\mathbf{X}\|_F = \sqrt{\langle \mathbf{X}, \mathbf{X} \rangle}$ , and  $\|\mathbf{X}\|_{1,2} = \sum_i \sqrt{\sum_j X_{i,j}^2}$ , respectively. Let  $\mathbf{G} : \mathbb{R}^{m_1 \times n_1} \rightarrow \mathbb{R}^{m_2 \times n_2}$  be a linear operator. A linear operator  $\mathbf{G}^* : \mathbb{R}^{m_2 \times n_2} \rightarrow \mathbb{R}^{m_1 \times n_1}$  is called the adjoint operator of  $\mathbf{G}$  if it satisfies  $\langle \mathbf{G}(\mathbf{X}), \mathbf{Y} \rangle = \langle \mathbf{X}, \mathbf{G}^*(\mathbf{Y}) \rangle$  for any  $\mathbf{X} \in \mathbb{R}^{m_1 \times n_1}$  and  $\mathbf{Y} \in \mathbb{R}^{m_2 \times n_2}$ .

### B. Weighted Sorted Mixed $\ell_{1,2}$ Norm [26]

For a matrix  $\mathbf{A} = [\mathbf{a}_1 \ \dots \ \mathbf{a}_m]^\top \in \mathbb{R}^{m \times n}$ , the weighted sorted mixed  $\ell_{1,2}$  norm of  $\mathbf{A}$  with the weight  $\mathbf{w} \in \mathbb{R}^m$  is defined by

$$\phi_{\mathbf{w}}(\mathbf{A}) = \sum_{j=1}^m w_j \|\mathbf{a}_{\pi_{\mathbf{A}}(j)}\|_2, \quad (1)$$

where  $\pi_{\mathbf{A}}(j)$  ( $\forall j = 1, \dots, m$ ) is the index corresponding to the  $j$ -th vector of the vectors obtained by sorting  $\mathbf{a}_1, \dots, \mathbf{a}_m$  in descending order in an  $\ell_2$  norm, i.e.,  $\|\mathbf{a}_{\pi_{\mathbf{A}}(1)}\|_2 \geq \dots \geq \|\mathbf{a}_{\pi_{\mathbf{A}}(m)}\|_2$ . Note that if the weight vector  $\mathbf{w}$  satisfy  $w_1 \geq \dots \geq w_m$ ,  $\phi_{\mathbf{w}}$  is convex.

### C. Proximal Linearized Difference-of-Convex Algorithm [22]

Let  $g, h : \mathbb{R}^{m \times n} \rightarrow (-\infty, \infty]$  be proper lower-semicontinuous convex functions and  $g - h$  is bounded below. We consider *difference-of-convex* minimization (DC) problems of the form:

$$\min_{\mathbf{X} \in \mathbb{R}^{n_1 \times n_2}} g(\mathbf{X}) - h(\mathbf{X}). \quad (2)$$

First, we introduce the *proximity operator* of  $g$  with a parameter  $\gamma > 0$  as follows:

$$\text{prox}_{\gamma g}(\mathbf{X}) = \underset{\mathbf{Y}}{\text{argmin}} g(\mathbf{Y}) + \frac{1}{2\gamma} \|\mathbf{X} - \mathbf{Y}\|_F^2. \quad (3)$$

We also define the *subdifferential* of  $h$  by

$$\partial h(\mathbf{X}) = \{\mathbf{U} \in \mathbb{R}^{m \times n} \mid \forall \mathbf{Y}, \langle \mathbf{Y} - \mathbf{X}, \mathbf{U} \rangle + h(\mathbf{X}) \leq h(\mathbf{Y})\}. \quad (4)$$

Then, PLDC solves Prob. (2) by the following procedures: for a given  $\mu > 0$ , iterate

$$\left\{ \begin{array}{l} \text{Find } \mathbf{Z}^{(t+1)} \in \partial h(\mathbf{X}^{(t)}); \\ \mathbf{X}^{(t+1)} \leftarrow \text{prox}_{\mu g}(\mathbf{X}^{(t)} + \mu \mathbf{Z}^{(t+1)}); \\ t = t + 1; \end{array} \right. \quad (5)$$

We summarize the theoretical results for the convergence of PLDC as follows:

**Theorem II.1** (Convergence of the sequence generated by PLDC [22]). *Let  $\{\mathbf{X}^t\}_{t \in \mathbb{N}}$  be generated by Algorithm (5). Then, it is satisfied that  $\lim_{t \rightarrow +\infty} \|\mathbf{X}^{(t+1)} - \mathbf{X}^{(t)}\|_F = 0$ . Furthermore, suppose that  $\{\mathbf{X}^{(t)}\}_{t \in \mathbb{N}}$  is bounded, and let  $\mathbf{X}^*$  and  $\mathbf{Z}^*$  be cluster-point of  $\mathbf{X}^{(t)}$  and  $\mathbf{Z}^{(t)}$ , respectively. Then, every cluster-point of  $\{\mathbf{X}^{(t)}\}_{t \in \mathbb{N}}$  is a critical point of the objective function in Prob. (2), i.e.,  $\mathbf{Z}^* \in \partial g(\mathbf{X}^*) \cap \partial h(\mathbf{X}^*)$ .*

---

### Algorithm 1 An algorithm for solving Prob. (13)

---

**Input:**  $\mathbf{V}, \mathbf{E}, \lambda, \mu, \varepsilon$ , and  $\eta$

**Output:**  $\mathbf{A}^{(k)}, \mathbf{S}^{(k)}$

- 1: Initialize  $\mathbf{A}^{(0)}, \mathbf{S}^{(0)}, \mathbf{Y}_1^{(0)}, \mathbf{Y}_2^{(0)}, \mathbf{Y}_3^{(0)}, \mathbf{Y}_4^{(0)}$ , and  $\mathbf{Y}_5^{(0)}$ ;
- 2: Set  $\gamma_1 = \frac{1}{10 + \sigma_1(\mathbf{E})^2}$ ,  $\gamma_2 = \frac{1}{2}$ , and  $\gamma_3 = \frac{1}{2}$ ;
- 3: **while** until  $\frac{\|\mathbf{A}^{(k+1)} + \mathbf{S}^{(k+1)} - (\mathbf{A}^{(k)} + \mathbf{S}^{(k)})\|_F}{\|\mathbf{A}^{(k)} + \mathbf{S}^{(k)}\|_F} \leq 10^{-5}$  **do**
- 4:  $\tilde{\mathbf{A}} \leftarrow \mathbf{A}^{(k)} - \gamma_1(\mathbf{Y}_1^{(k)} + \mathbf{D}^*(\mathbf{Y}_2^{(k)}) + \mathbf{E}^\top \mathbf{Y}_3^{(k)} + \mathbf{Y}_4^{(k)})$ ;
- 5:  $\mathbf{A}^{(k+1)} \leftarrow \text{prox}_{\gamma_1 \ell_{[0,1]^{m \times n}}}(\tilde{\mathbf{A}})$ ;
- 6:  $\mathbf{S}^{(k+1)} \leftarrow \text{prox}_{\gamma_2 \ell_{\mathcal{B}_{1,\eta}}}(\mathbf{S}^{(k)} - \gamma_2(\mathbf{Y}_3^{(k)} + \mathbf{Y}_5^{(k)}))$ ;
- 7:  $\tilde{\mathbf{Y}}_1 \leftarrow \mathbf{Y}_1^{(k)} + \gamma_3(2\mathbf{A}^{(k+1)} - \mathbf{A}^{(k)})$ ;
- 8:  $\mathbf{Y}_1^{(k+1)} \leftarrow \tilde{\mathbf{Y}}_1 - \gamma_3 \text{prox}_{\frac{1}{\gamma_3} \|\cdot\|_{1,2}}(\frac{\tilde{\mathbf{Y}}_1}{\gamma_3})$ ;
- 9:  $\tilde{\mathbf{Y}}_2 \leftarrow \mathbf{Y}_2^{(k)} + \gamma_3 \mathbf{D}(2\mathbf{A}^{(k+1)} - \mathbf{A}^{(k)})$ ;
- 10:  $\mathbf{Y}_2^{(k+1)} \leftarrow \tilde{\mathbf{Y}}_2 - \gamma_3 \text{prox}_{\frac{\lambda}{\gamma_3} \|\cdot\|_1}(\frac{\tilde{\mathbf{Y}}_2}{\gamma_3})$ ;
- 11:  $\tilde{\mathbf{Y}}_3 \leftarrow \mathbf{Y}_3^{(k)} + \gamma_3(\mathbf{E}(2\mathbf{A}^{(k+1)} - \mathbf{A}^{(k)}) + (2\mathbf{S}^{(k+1)} - \mathbf{S}^{(k)}))$ ;
- 12:  $\mathbf{Y}_3^{(k+1)} \leftarrow \tilde{\mathbf{Y}}_3 - \gamma_3 \text{prox}_{\frac{1}{\gamma_3} \ell_{\mathcal{B}_{F,\varepsilon}}}(\frac{\tilde{\mathbf{Y}}_3}{\gamma_3})$ ;
- 13:  $\tilde{\mathbf{Y}}_4 \leftarrow \mathbf{Y}_4^{(k)} + \gamma_3(2\mathbf{A}^{(k+1)} - \mathbf{A}^{(k)})$ ;
- 14:  $\mathbf{Y}_4^{(k+1)} \leftarrow \tilde{\mathbf{Y}}_4 - \gamma_3 \text{prox}_{\frac{1}{2\mu_1 \gamma_3} \|\cdot - \mathbf{A}'\|_2}(\frac{\tilde{\mathbf{Y}}_4}{\gamma_3})$ ;
- 15:  $\tilde{\mathbf{Y}}_5 \leftarrow \mathbf{Y}_5^{(k)} + 2(\mathbf{S}^{(k+1)} - \mathbf{S}^{(k)})$ ;
- 16:  $\mathbf{Y}_5^{(k+1)} \leftarrow \tilde{\mathbf{Y}}_5 - \gamma_3 \text{prox}_{\frac{1}{2\mu_1 \gamma_3} \|\cdot - \mathbf{S}'\|_2}(\frac{\tilde{\mathbf{Y}}_5}{\gamma_3})$ ;
- 17:  $k \leftarrow k + 1$ ;

---

## III. PROPOSED METHOD

### A. Row-Sparsity Enhancement Weighted Sorted Function

Let a weight  $\mathbf{w} \in \mathbb{R}^m$  be for  $k \in \{1, \dots, m\}$

$$w_j = \begin{cases} 1, & \text{if } j \leq k; \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

For a matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$ , we define a new weighted sorted function with the weight  $\mathbf{w}$  by

$$\Phi_{\mathbf{w}}(\mathbf{A}) = \|\mathbf{A}\|_{1,2} - \phi_{\mathbf{w}}(\mathbf{A}) = \sum_{j=k+1}^m w_j \|\mathbf{a}_{\pi_{\mathbf{A}}(j)}\|_2. \quad (7)$$

By minimizing Eq. (7), we can promote the row sparsity of  $\mathbf{A}$  while preserving the energies of the top  $k$  vectors  $\mathbf{a}_{\pi_{\mathbf{A}}(1)}, \dots, \mathbf{a}_{\pi_{\mathbf{A}}(k)}$ . Since the weight  $\mathbf{w}$  in (6) satisfies  $w_1 \geq \dots \geq w_m$ ,  $\phi_{\mathbf{w}}$  in (7) is convex. Therefore,  $\Phi_{\mathbf{w}}$  is a DC function. In addition,  $\Phi_{\mathbf{w}}$  is bounded below by 0, i.e.,  $\Phi_{\mathbf{w}}(\mathbf{A}) \geq 0$  for any  $\mathbf{A} \in \mathbb{R}^{m \times n}$ .

### B. Problem Formulation

Let  $\mathbf{E} \in \mathbb{R}^{l \times m}$ ,  $\bar{\mathbf{A}} \in \mathbb{R}^{m \times n}$ ,  $\bar{\mathbf{S}} \in \mathbb{R}^{l \times n}$ , and  $\mathbf{N} \in \mathbb{R}^{l \times n}$  be an endmember library, a true abundance matrix, sparse noise, and Gaussian noise, respectively. Then, the observed HS image  $\mathbf{V}$  is modeled according to the linear mixing model as

$$\mathbf{V} = \mathbf{E}\bar{\mathbf{A}} + \bar{\mathbf{S}} + \mathbf{N}. \quad (8)$$

TABLE I  
SRES AND RMSES IN THE EXPERIMENTS.

Images	Noise $\sigma / p_S$	Metrics	Methods							
			CLSUnSAL [10]	JSTV [11]	LGSU [13]	MdLRR [14]	RSSUn-TV [16]	UnDIP [25]	Ours ( $k = 0.1m$ )	Ours ( $k = 0.2m$ )
Synth	0.1 / 0	SRE	3.22	5.19	5.78	6.37	3.14	-0.11	<b>11.91</b>	<u>11.90</u>
		RMSE	0.0779	0.0621	0.0580	0.0543	0.0787	0.1143	<b>0.0286</b>	<u>0.0287</u>
	0.1 / 0.05	SRE	1.60	5.10	2.41	2.40	0.73	0.02	<b>11.58</b>	<u>11.40</u>
		RMSE	0.0939	0.0628	0.0855	0.0857	0.1039	0.1126	<b>0.0298</b>	<u>0.0304</u>
Real	0.1 / 0	SRE	8.27	3.67	6.91	7.02	6.63	0.24	<b>9.47</b>	<b>9.47</b>
		RMSE	0.0469	0.0797	0.0549	0.0542	0.0566	0.1183	<b>0.0409</b>	<b>0.0409</b>
	0.1 / 0.05	SRE	6.55	3.56	3.70	2.88	3.66	0.44	<b>8.14</b>	<b>8.14</b>
		RMSE	0.0572	0.0807	0.0794	0.0873	0.0798	0.1156	<b>0.0476</b>	<b>0.0476</b>

Based on Eq. (8), we formulate an unmixing problem as the following constrained DC minimization problem:

$$\begin{aligned} \min_{\mathbf{A}, \mathbf{S}} \Phi_{\mathbf{w}}(\mathbf{A}) + \lambda \|\mathbf{D}(\mathbf{A})\|_1 \\ \text{s.t. } \mathbf{A} \in [0, 1]^{m \times n}, \mathbf{E}\mathbf{A} + \mathbf{S} \in \mathcal{B}_{F, \varepsilon}^{\mathbf{V}}, \mathbf{S} \in \mathcal{B}_{1, \eta}, \end{aligned} \quad (9)$$

where  $\lambda > 0$  is a balancing parameter and  $\mathbf{D}$  is a spatial difference operator. The first term promotes the row sparsity of  $\mathbf{A}$ . The second term captures the spatial piecewise smoothness of  $\mathbf{A}$ . The first constraint guarantees that abundances are within  $[0, 1]$ . Note that we do not explicitly adopt the abundance sum-to-one constraint because it tends to be a strong assumption for LMM-based unmixing in real-world situations due to spectral variabilities [7]. The second constraint serves as data-fidelity to the observed HS image  $\mathbf{V}$  with the Frobenius norm ball  $\mathcal{B}_{F, \varepsilon}^{\mathbf{V}} := \{\mathbf{X} \mid \|\mathbf{V} - \mathbf{X}\|_F \leq \varepsilon\}$  with the center  $\mathbf{V}$  and radius  $\varepsilon$ . The third constraint evaluates the sparsity of  $\mathbf{S}$  with the  $\ell_1$ -norm ball  $\mathcal{B}_{1, \eta} := \{\mathbf{X} \mid \|\mathbf{X}\|_1 \leq \eta\}$  with radius  $\eta$ . Using such constraints instead of data-fidelity and sparse terms makes it easy to adjust hyperparameters since the parameters  $\varepsilon$  and  $\eta$  can be determined based only on noise intensity<sup>2</sup>.

### C. Optimization Algorithm

Using the indicator functions<sup>3</sup> of  $[0, 1]^{m \times n}$ ,  $\mathcal{B}_{F, \varepsilon}^{\mathbf{V}}$ , and  $\mathcal{B}_{1, \eta}$ , we reformulate Prob. (9) as the following equivalent problem:

$$\begin{aligned} \min_{\mathbf{A}, \mathbf{S}} \|\mathbf{A}\|_{1,2} + \lambda \|\mathbf{D}(\mathbf{A})\|_1 + \iota_{[0,1]^{m \times n}}(\mathbf{A}) \\ + \iota_{\mathcal{B}_{F, \varepsilon}^{\mathbf{V}}}(\mathbf{E}\mathbf{A} + \mathbf{S}) + \iota_{\mathcal{B}_{1, \eta}}(\mathbf{S}) - \phi_{\mathbf{w}}(\mathbf{A}). \end{aligned} \quad (10)$$

Since  $\lambda \|\mathbf{D}(\mathbf{A})\|_1$ ,  $\iota_{[0,1]^{m \times n}}(\mathbf{A})$ ,  $\iota_{\mathcal{B}_{F, \varepsilon}^{\mathbf{V}}}$ , and  $\iota_{\mathcal{B}_{1, \eta}}$  are convex and  $\|\mathbf{A}\|_{1,2} - \phi_{\mathbf{w}}(\mathbf{A})$  is bounded below, the objective function in Prob. (10) is bounded below. Therefore, Prob. (10), i.e., Prob. (9) can be solved by PLDC in (5).

In what follows, we derive specific computations of our PLDC-based algorithm for solving (9).

<sup>2</sup>Indeed, this kind of constrained formulation has played an important role in facilitating parameter setup of signal recovery problems [27]–[31].

<sup>3</sup>For a given nonempty closed convex set  $C \subset \mathbb{R}^{m \times n}$ , the indicator function of  $C$  is defined by  $\iota_C(\mathbf{X}) := 0$ , if  $\mathbf{X} \in C$ ;  $\infty$ , otherwise.

1) *Step 1 of our PLDC-based algorithm for solving Prob. (9):* First, we need to calculate one of the subgradients of  $\phi_{\mathbf{w}}$  at  $\mathbf{A}^{(t)}$ , i.e.,  $\mathbf{Z}_{\mathbf{A}}^{(t)} \in \partial \phi_{\mathbf{w}}(\mathbf{A}^{(t)})$ . The following proposition helps us to calculate  $\mathbf{Z}_{\mathbf{A}}^{(t)}$ .

**Proposition III.1.** *The subdifferential of weighted sorted mixed  $\ell_{1,2}$  norm  $\phi_{\mathbf{w}}$  is given by*

$$\partial \phi_{\mathbf{w}}(\mathbf{A}) = \{[\mathbf{u}_1 \cdots \mathbf{u}_m]^\top \mid \forall j, \mathbf{u}_{\pi_{\mathbf{A}}(j)} \in \partial w_j \|\cdot\|_2(\mathbf{a}_{\pi_{\mathbf{A}}(j)})\}, \quad (11)$$

where

$$\partial w_j \|\cdot\|_2(\mathbf{a}_{\pi_{\mathbf{A}}(j)}) = \begin{cases} \frac{w_j}{\|\mathbf{a}_{\pi_{\mathbf{A}}(j)}\|_2} \mathbf{a}_{\pi_{\mathbf{A}}(j)}, & \text{if } \mathbf{a}_{\pi_{\mathbf{A}}(j)} \neq \mathbf{0}; \\ \{w_j \mathbf{z} \mid \|\mathbf{z}\|_2 \leq 1\}, & \text{if } \mathbf{a}_{\pi_{\mathbf{A}}(j)} = \mathbf{0}. \end{cases} \quad (12)$$

The following gives a rough sketch of the proof of Proposition III.1:  $\partial \phi_{\mathbf{w}}(\mathbf{A}^{(t)})$  satisfies Eq. (4) from the rearrangement inequality [32, Theorem 6.1]. When  $\mathbf{a}_{\pi_{\mathbf{A}}(j)} = \mathbf{0}$ , we recommend selecting  $\mathbf{z} = \mathbf{0}$ .

2) *Step 2 of our PLDC-based algorithm for solving Prob. (9):* Let  $\mathbf{A}' = \mathbf{A}^{(t)} + \mu \mathbf{Z}_{\mathbf{A}}^{(t)}$  and  $\mathbf{S}' = \mathbf{S}^{(t)}$ . Then,  $\mathbf{A}^{(t+1)}$  and  $\mathbf{S}^{(t+1)}$  are given by solving the following optimization problem:

$$\begin{aligned} \min_{\mathbf{A}, \mathbf{S}} \frac{1}{2\mu} \|\mathbf{A} - \mathbf{A}'\|_2^2 + \frac{1}{2\mu} \|\mathbf{S} - \mathbf{S}'\|_2^2 + \|\mathbf{A}\|_{1,2} + \lambda \|\mathbf{D}(\mathbf{A})\|_1 \\ + \iota_{[0,1]^{m \times n}}(\mathbf{A}) + \iota_{\mathcal{B}_{F, \varepsilon}^{\mathbf{V}}}(\mathbf{E}\mathbf{A} + \mathbf{S}) + \iota_{\mathcal{B}_{1, \eta}}(\mathbf{S}). \end{aligned} \quad (13)$$

Prob. (13) can be solved by the preconditioned primal-dual splitting method [23] and variable-wise diagonal preconditioning method [24]. The algorithm for solving Prob. (13) is summarized in Algorithm 1. The proximity operators in Algorithm 1 are calculated as shown in [33], [34].

**Remark III.1** (Convergence of our PLDC-based algorithm). The proximity operators of  $\iota_{[0,1]^{m \times n}}$  and  $\iota_{\mathcal{B}_{1, \eta}}$  are the projections onto  $[0, 1]^{m \times n}$  and  $\mathcal{B}_{1, \eta}$ . This indicates that for any  $t \in \mathbb{N}$ ,  $\mathbf{A}^{(t)}$  and  $\mathbf{S}^{(t)}$  satisfy  $\mathbf{A}^{(t)} \in [0, 1]^{m \times n}$  and  $\|\mathbf{S}^{(t)}\|_1 \leq \eta$ , i.e.,  $\{\mathbf{A}^{(t)}, \mathbf{S}^{(t)}\}_{t \in \mathbb{N}}$  is bounded. Therefore, from Theorem II.1,  $\{\mathbf{A}^{(t)}, \mathbf{S}^{(t)}\}_{t \in \mathbb{N}}$  is guaranteed to converge to a critical point of Prob. (9).

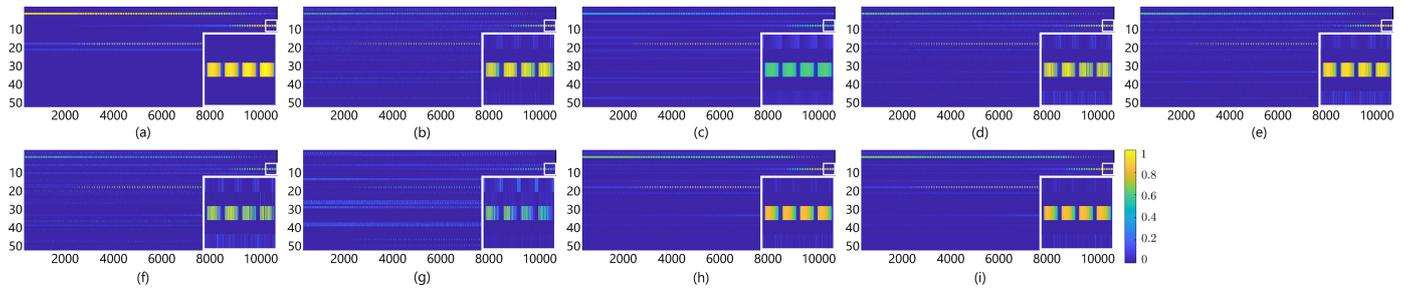


Fig. 1. Unmixing results of abundance map matrices for the experiments using the synthetic HS image with  $\sigma = 0.1$  of Gaussian noise. Vertical and horizontal axes are the indices of endmembers and pixels, respectively. Expanded area shows the 8th through 12th endmembers. (a): Original abundance maps. (b): CLSUnSAL [10], (c): JSTV [11], (d): LGSU [13], (e): MdLRR [14], (f): RSSUn-TV [16], (g): UnDIP [25], (h): Ours ( $k = 0.1m$ ), and (i): Ours ( $k = 0.2m$ ).

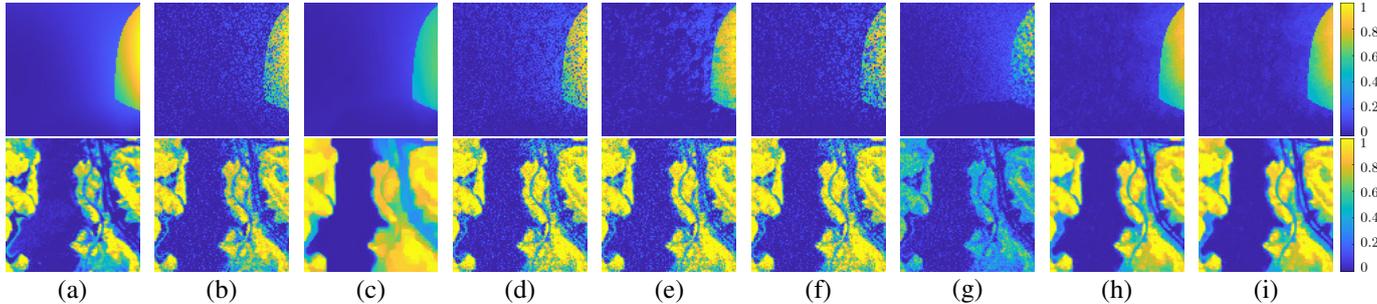


Fig. 2. Unmixing results of abundance maps. Top row shows the 22nd abundance maps of the experiments using the synthetic HS image with  $\sigma = 0.1$  of Gaussian noise. Bottom row shows the 35th abundance maps of the experiments using the real HS image with  $\sigma = 0.1$  of Gaussian noise and  $p_S = 0.05$  of sparse noise. (a): Original abundance maps. (b): CLSUnSAL [10], (c): JSTV [11], (d): LGSU [13], (e): MdLRR [14], (f): RSSUn-TV [16], (g): UnDIP [25], (h): Ours ( $k = 0.1m$ ), and (i): Ours ( $k = 0.2m$ ).

## IV. EXPERIMENTS

### A. Experimental Setup

To illustrate the effectiveness of our method, we compare it with six methods, including state-of-the-art ones: four convex-based methods CLSUnSAL [10], JSTV [11], LGSU [13], and MdLRR [14], one nonconvex-based method RSSUn-TV [16], and one deep neural network-based method UnDIP [25]. The parameters of the existing methods were set to the values recommended in their references.

Experiments were performed using a synthetic image and a real image. For a synthetic image, we generate a ground-truth image of size  $100 \times 100 \times 224$  using the hyperspectral data retrieval and analysis toolbox<sup>4</sup> and four spectral signatures from the U.S. Geological Survey Spectral Library<sup>5</sup>. For a real image, we use *Jasper Ridge* as a ground-truth image of size  $100 \times 100 \times 187$  after removing several noisy bands and cropping the image. For both image experiments, by adding 46 spectral signatures from the USGS Spectral Library, we used  $\mathbf{E}$  containing a total of  $m = 50$  spectral signatures. The pixel values were normalized into  $[0, 1]$ . Contaminating these ground-truth videos by white Gaussian noise with 0.1 of the standard deviation  $\sigma$  and Salt & Pepper noise with 0 and 0.05 of the rate  $p_S$ , we generated observed HS images.

For the PLDC-based algorithm, we set the stopping criterion as  $\|\mathbf{A}^{(t+1)} - \mathbf{A}^{(t)}\|_F / \|\mathbf{A}^{(t)}\|_F \leq 10^{-5}$ . We initialized  $\mathbf{A}^{(0)}$  and  $\mathbf{S}^{(0)}$  by zero matrices. The parameters  $\lambda$ ,  $\mu$ ,  $\varepsilon$ , and

$\eta$  were set to 0.1, 10,  $0.95\sigma\sqrt{(1-p_S)nl}$ , and  $0.45p_Snl$ , respectively. For the threshold  $k$  in Eq. (6), it is preferable to give the number of endmembers involved in a target HS image. However, in real-world situations, the number of endmembers is often unknown. Therefore, in this experiment,  $k$  was set in the following two ways based on the number of spectral signatures contained in the endmember libraries:  $k = 0.1m = 5$  and  $k = 0.2m = 10$ . Note that the former is closer to the actual value of 4. For the quantitative evaluation of abundance maps, we used the signal reconstruction error (SRE):  $\text{SRE}[\text{dB}] := 10 \log_{10}(\|\hat{\mathbf{A}}\|_F^2 / \|\hat{\mathbf{A}} - \mathbf{A}\|_F^2)$ , the root-mean-square error (RMSE):  $\text{RMSE} := \|\hat{\mathbf{A}} - \mathbf{A}\|_F / \sqrt{mn}$ .

### B. Results and Discussion

Table I shows SREs and RMSEs of unmixing results. The best and second-best results are highlighted in bold and underlined. Our method was superior to all the existing methods. Compared to when  $k = 5$ , the performance did not drop as much when  $k = 10$ .

Fig. 1 shows the abundance matrices for the experiments using a synthetic HS image with  $\sigma = 0.1$  of Gaussian noise. The abundance matrices of CLSUnSAL, RSSUnTV, LGSU, UnDIP, and MdLRR had slightly large values even when the corresponding endmembers did not actually exist in the ground truth HS images. The abundance matrix of JSTV had smaller values than the original abundance maps corresponding to the endmembers that exist in the ground truth HS images. These results indicate that all the existing methods did not accurately enhance the row sparsity. In contrast, our method enhanced the row sparsity of the abundance matrix.

<sup>4</sup>[https://www.ehu.es/ccwintco/index.php?title=Hyperspectral\\_Imagery\\_Synthesis\\_tools\\_for\\_MATLAB](https://www.ehu.es/ccwintco/index.php?title=Hyperspectral_Imagery_Synthesis_tools_for_MATLAB), accessed on Feb. 5, 2023

<sup>5</sup><https://www.usgs.gov/programs/usgs-library>, accessed on Aug. 7, 2023

Fig. 2 depicts the abundance maps. All the existing methods did not estimate the abundance maps while removing noise. Specifically, the abundance maps of CLSUnSAL, RSSUnTV, LGSU, and MdLRR have residual noise. The abundance maps of JSTV were over-smooth. UnDIP does not employ a function that explicitly promotes the row sparsity; therefore, the intensities of the abundance maps are lower than the original abundance maps. In contrast, our method exactly estimates the abundance maps even when  $k = 10$ .

## V. CONCLUSION

We proposed a new unmixing method via a DC approach. First, we designed a DC function that enhances the row sparsity of abundance maps and formulated unmixing as a constrained DC optimization problem that incorporates the DC function. The algorithm solving the optimization problem was developed based on PLDC and P-PDS. The experiments showed that our method can enhance the row sparsity of abundance maps and thus obtained superior results than existing methods.

## ACKNOWLEDGMENT

This work was supported in part by JST ACT-X Grant Number JPMJAX23CJ, JST PRESTO under Grant JPMJPR21C4, and JST AdCORP under Grant JPMJKB2307, in part by JSPS KAKENHI under Grant 22H03610, 22H00512, 23H01415, and 23K17461, and in part by Grant-in-Aid for JSPS Fellows under Grant 23KJ0912.

## REFERENCES

- [1] C-I Chang, *Hyperspectral imaging: techniques for spectral detection and classification*, vol. 1, Springer Science & Business Media, 2003.
- [2] W.-K. Ma, J. M. Bioucas-Dias, T.-H. Chan, N. Gillis, P. Gader, A. J. Plaza, A. Ambikapathi, and C.-Y. Chi, "A signal processing perspective on hyperspectral unmixing: Insights from remote sensing," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 67–81, 2014.
- [3] P. S. Thenkabail and J. G. Lyon, *Hyperspectral remote sensing of vegetation*, CRC press, 2016.
- [4] P. Ghamisi, N. Yokoya, J. Li, W. Liao, S. Liu, J. Plaza, B. Rasti, and A. Plaza, "Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 37–78, 2017.
- [5] B. Lu, P. D. Dao, J. Liu, Y. He, and J. Shang, "Recent advances of hyperspectral imaging technology and applications in agriculture," *Remote Sens.*, vol. 12, no. 16, pp. 2659, 2020.
- [6] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, 2002.
- [7] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, 2011.
- [8] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Top. Appl. Earth. Obs. Remote Sens.*, vol. 5, no. 2, pp. 354–379, 2012.
- [9] W.K. Ma, J. M. Bioucas-Dias, T.H. Chan, N. Gillis, P. Gader, A. Plaza, A. Ambikapathi, and C.Y. Chi, "A signal processing perspective on hyperspectral unmixing: Insights from remote sensing," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 67–81, Jan. 2014.
- [10] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Collaborative sparse regression for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 341–354, Jan. 2014.
- [11] H. K. Aggarwal and A. Majumdar, "Hyperspectral unmixing in the presence of mixed noise using joint-sparsity and total variation," *IEEE J. Sel. Top. Appl. Earth. Obs. Remote Sens.*, vol. 9, no. 9, pp. 4257–4266, 2016.
- [12] H Li, R Feng, L Wang, Y Zhong, and L Zhang, "Superpixel-based reweighted low-rank and total variation sparse unmixing for hyperspectral remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 629–647, 2020.
- [13] X Shen, H Liu, X Zhang, K Qin, and X Zhou, "Superpixel-guided local sparsity prior for hyperspectral sparse regression unmixing," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022, Art no. 6015105.
- [14] L. Wu, J. Huang, and M.-S. Guo, "Multidimensional low-rank representation for sparse hyperspectral unmixing," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, Mar 2023, Art no. 5502805.
- [15] K. Naganuma and S. Ono, "Toward robust hyperspectral unmixing: Mixed noise modeling and image-domain regularization," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 8117–8138, 2024.
- [16] J.J. Wang, T.Z. Huang, J. Huang, H.X. Dou, L.J. Deng, and X.L. Zhao, "Row-sparsity spectral unmixing via total variation," *IEEE J. Sel. Top. Appl. Earth. Obs. Remote Sens.*, vol. 12, no. 12, pp. 5009–5022, 2019.
- [17] E. Esser, Y. Lou, and J. Xin, "A method for finding structured sparse solutions to nonnegative least squares problems with applications," *SIAM J. Imag. Sci.*, vol. 6, no. 4, pp. 2010–2046, 2013.
- [18] Y. Lou, P. Yin, Q. He, and J. Xin, "Computing sparse representation in a highly coherent dictionary based on difference of  $L_1$  and  $L_2$ ," *J. Sci. Comput.*, vol. 64, pp. 178–196, 2015.
- [19] W. Li, W. Bian, and K.-C. Toh, "Difference-of-convex algorithms for a class of sparse group  $\ell_0$  regularized optimization problems," *SIAM J. Optim.*, vol. 32, no. 3, pp. 1614–1641, 2022.
- [20] Shunsuke Ono, Kazuki Naganuma, and Keitaro Yamashita, "Graph signal sampling under smoothness priors: A difference-of-convex approach," *arXiv preprint arXiv:2306.14634*, 2023.
- [21] Koyo Sato, Kazuki Naganuma, and Shunsuke Ono, "Enhancing hyperspectral anomaly detection by difference-of-convex sparse anomaly modeling," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2024, pp. 9921–9925.
- [22] J. C. O Souza, P. R. Oliveira, and A. Soubeyran, "Global convergence of a proximal linearized algorithm for difference of convex functions," *Optim. Lett.*, vol. 10, no. 7, pp. 1529–1539, 2016.
- [23] T Pock and A Chambolle, "Diagonal preconditioning for first order primal-dual algorithms in convex optimization," in *2011 International Conference on Computer Vision*. IEEE, 2011, pp. 1762–1769.
- [24] Kazuki Naganuma and Shunsuke Ono, "Variable-wise diagonal preconditioning for primal-dual splitting: Design and applications," *IEEE Transactions on Signal Processing*, vol. 71, pp. 3281–3295, 2023.
- [25] B. Rasti, B. Koirala, P. Scheunders, and P. Ghamisi, "UnDIP: Hyperspectral unmixing using deep image prior," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, 2022, Art no. 5504615.
- [26] D. Brzyski, A. Gossmann, W. Su, and M. Bogdan, "Group slope – adaptive selection of groups of predictors," *J. Amer. Stat. Assoc.*, vol. 114, no. 525, pp. 419–433, 2019.
- [27] M. Afonso, J.M. Bioucas-Dias, and M. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 681–695, 2011.
- [28] G. Chierchia, N. Pustelnik, J.C. Pesquet, and B. Pesquet-Popescu, "Epigraphical projection and proximal tools for solving constrained convex optimization problems," *Signal, Image Video Process.*, vol. 9, no. 8, pp. 1737–1749, 2015.
- [29] S. Ono and I. Yamada, "Signal recovery with certain involved convex data-fidelity constraints," *IEEE Trans. Signal Process.*, vol. 63, no. 22, pp. 6149–6163, Nov. 2015.
- [30] S. Ono, "Primal-dual plug-and-play image restoration," *IEEE Signal Process. Lett.*, vol. 24, no. 8, pp. 1108–1112, Aug. 2017.
- [31] S. Ono, "Efficient constrained signal reconstruction by randomized epigraphical projection," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*. IEEE, 2019, pp. 4993–4997.
- [32] E. F. Beckenbach and R. Bellman, *Inequalities*, vol. 30, Springer Science & Business Media, 2012.
- [33] G. Chierchia, E. Chouzenoux, P. L. Combettes, and J.-C. Pesquet, "The proximity operator repository," *User's guide <http://proximityoperator.net/download/guide.pdf>* (accessed October 3rd, 2021), 2020.
- [34] L. Condat, "Fast projection onto the simplex and the  $l_1$  ball," *Math. Program.*, vol. 158, no. 1, pp. 575–585, Jul. 2016.