Block-Sparse Recovery with Optimal Block Partition

Hiroki Kuroda, Member, IEEE, and Daichi Kitahara, Member, IEEE

Abstract—This paper presents a convex recovery method for block-sparse signals whose block partitions are unknown a priori. We first introduce a nonconvex penalty function, where the block partition is adapted for the signal of interest by minimizing the mixed $\ell_2/\ell_1$ norm over all possible block partitions. Then, by exploiting a variational representation of the $\ell_2$ norm, we derive the proposed penalty function as a suitable convex relaxation of the nonconvex one. For a block-sparse recovery model designed with the proposed penalty, we develop an iterative algorithm which is guaranteed to converge to a globally optimal solution. Numerical experiments demonstrate the effectiveness of the proposed method.

Index Terms—Block-sparsity, unknown partition, penalty function, convex optimization, proximal splitting algorithm.

I. INTRODUCTION

BLOCK-SPARSITY is a special type of sparsity where nonzero components are clustered in blocks. For recovery of a block-sparse signal whose block partition is known a priori, extensive researches, e.g., [1]–[10], show the effectiveness of the mixed $\ell_2/\ell_1$ norm using the known block partition. However, in fact, the information on the block partition is not available in many applications, e.g., acoustic signal recovery [11]–[13], image restoration [14]–[16], change detection [17]–[19], and radar signal processing [20]–[23]. For instance, the target signal of the phased array weather radar [21]–[25] is block-sparse in the Fourier domain due to the narrow bandwidth of the power spectrum, but the block partition is unknown because it depends on the unknown mean and standard deviation of the Doppler frequency. In such situations, the estimation accuracy of the mixed $\ell_2/\ell_1$ norm often degrades due to a pre-fixed block partition used instead of the ideal one.

Motivated by these observations, we consider the estimation problem of $x^* \in \mathbb{C}^N$ which is block-sparse across unknown non-overlapping blocks $B_1^*, \ldots, B_K^*$. (see Fig. 1 for an illustration). More precisely, we suppose that the subvector $x_{B_k^*}^* := \{x_n^*\}_{n \in B_k^*}$ contains only (nearly) zero components for many $k$, i.e.,

$$x_{B_k^*}^* \approx 0 \text{ for many } k \in \{1, \ldots, K^*\}.$$ 

We use the term block-sparse in a strict sense, i.e., $B_k^*$ consists of only consecutive indices for each $k = 1, \ldots, K^*$:

$$B_k^* = \{n \in \{1, \ldots, N\} \mid n_k^* \leq n \leq m_k^*\},$$

with some $n_k^*, m_k^* \in \{1, \ldots, N\}$. We also suppose that $B_1^*, \ldots, B_K^*$, form a partition of $\{1, \ldots, N\}$: $\bigcup_{k=1}^{K^*} B_k^* = \{1, \ldots, N\}$, $B_k^* \neq \emptyset$ ($k = 1, \ldots, K^*$), and $B_k^* \cap B_{k'}^* = \emptyset$ $(k \neq k')$. Note that the non-overlapping condition $B_k^* \cap B_{k'}^* = \emptyset$ is not restrictive since the sizes of $B_1^*, \ldots, B_K^*$, can be different.

Several attempts have been made to deal with the unknown blocks $B_1^*, \ldots, B_K^*$. A commonly used modification is to use overlapping blocks in the mixed $\ell_2/\ell_1$ norm, e.g., [26]–[33]. Among them, the latent group lasso (LGL) method [32], [33] presents a sophisticated approach to select relevant blocks from pre-defined overlapping blocks. The LGL penalty is defined as the minimum of a convex function, and thus the corresponding regularization model can be solved as a convex optimization problem (see Appendix A for detail). However, since the problem size grows with the number of candidate blocks, the LGL method has a restriction on the number of candidate blocks, which leads to the degradation of the estimation accuracy. Meanwhile, the greedy methods [34], [35] have similar restrictions on the number of candidate blocks due to the computational complexity. Bayesian approaches [36], [37] have also been presented, but they have to solve challenging nonconvex optimization problems. Thus, it is highly desired to realize a convex method which can cope with the unknown blocks $B_1^*, \ldots, B_K^*$, whose sizes are different.

In this paper, we propose a convex recovery method for $x^*$ block-sparse across the unknown blocks $B_1^*, \ldots, B_K^*$. Our main contribution is to introduce a novel convex penalty function, named latent optimally partitioned (LOP)-$\ell_2/\ell_1$ penalty, where the block partition is automatically adapted for the signal of interest. More precisely, we first design a nonconvex penalty function as the minimum of the mixed $\ell_2/\ell_1$ norm over all possible block partitions. Then, to derive a tight convex relaxation of the nonconvex penalty function, using a variational representation of the $\ell_2$ norm (see Lemma 1), we represent the nonconvex penalty function by the minimization of a convex function under the $\ell_0$ pseudo-norm constraint on latent variables. Finally, the LOP-$\ell_2/\ell_1$ penalty is derived by replacing the $\ell_0$ pseudo-norm constraint with its convex envelope, i.e., the $\ell_1$ norm constraint. To compute an optimal solution of a block-sparse recovery model designed with the LOP-$\ell_2/\ell_1$ penalty, we reformulate it into a convex optimization problem involving the latent variables of the LOP-$\ell_2/\ell_1$ penalty. By applying proximal splitting techniques, e.g., [38]–[46], to the reformulated problem with the aid of the computation of the proximity operator shown in [47], [48], we obtain an iterative algorithm with guaranteed convergence to the optimal solution. Numerical experiments on both synthetic examples and real-
world data show that the proposed method achieves superior estimation accuracy to existing convex methods including the LGL method.

The rest of this paper is organized as follows. In Section II, we introduce the LOP-\(\ell_2/\ell_1\) penalty, and also show its extension to higher dimensions. In Section III, we design the block-sparse recovery model with the LOP-\(\ell_2/\ell_1\) penalty, and develop the iterative algorithm which converges to the optimal solution of the model. Section IV provides numerical examples, followed by conclusion in Section V.

A preliminary short version of this paper is to be presented at a conference [49].

Notations: \(\mathbb{R}\), \(\mathbb{R}_+\), and \(\mathbb{C}\) respectively denote the sets of all real numbers, all nonnegative real numbers, and all complex numbers. For matrices or vectors, we denote the simple transpose and the Hermitian transpose respectively by \((\cdot)^T\) and \((\cdot)^H\). For \(x = (x_1, \ldots, x_N)^T \in \mathbb{C}^N\) and an index set \(I \subset \{1, \ldots, N\}\), \(x_I := (x_n)_{n \in I}\) denotes the subvector of \(x\) indexed by \(I\). We define the support of \(x \in \mathbb{C}^N\) by \(\text{supp}(x) := \{n \in \{1, \ldots, N\} \mid x_n \neq 0\}\). For a set \(S\), \(|S|\) denotes the cardinality of \(S\). The \(\ell_2\) norm, the \(\ell_1\) norm, and the \(\ell_0\) pseudo-norm of \(x \in \mathbb{C}^N\) are respectively defined by \(\|x\|_2 := \sqrt{\sum_{n=1}^{N} |x_n|^2}\), \(\|x\|_1 := \sum_{n=1}^{N} |x_n|\), and \(\|x\|_0 := |\text{supp}(x)|\). We define the operator norm of \(L \in \mathbb{C}^{M \times N}\) by \(\|L\|_{\text{op}} := \max_{x \neq 0} \|Lx\|_2/\|x\|_2\).

Let \((\mathcal{U}, \|\cdot\|)\) be a finite dimensional Hilbert space. A function \(f: \mathcal{U} \to \mathbb{R} \cup \{\infty\}\) is called proper if its effective domain \(\text{dom}(f) := \{u \in \mathcal{U} \mid f(u) < \infty\}\) is nonempty. A function \(f: \mathcal{U} \to \mathbb{R} \cup \{\infty\}\) is lower semicontinuous if its lower level set \(\{u \in \mathcal{U} \mid f(u) \leq a\}\) is closed for every \(a \in \mathbb{R}\). A function \(f: \mathcal{U} \to \mathbb{R} \cup \{\infty\}\) is convex if \(f(\beta u + (1- \beta) v) \leq \beta f(u) + (1- \beta) f(v)\) for every \(u, v \in \mathcal{U}\) and \(\beta \in (0,1)\). The set of all proper lower semicontinuous convex functions from \(\mathcal{U}\) to \(\mathbb{R} \cup \{\infty\}\) is denoted by \(\Gamma_0(\mathcal{U})\). A function \(f: \mathcal{U} \to \mathbb{R} \cup \{\infty\}\) is called coercive if \(\lim_{\|u\| \to \infty} f(u) = \infty\). The proximity operator of \(f \in \Gamma_0(\mathcal{U})\) is defined by \(\text{prox}_f(u) := \arg \min_{v \in \mathcal{U}} \left[ f(v) + \frac{1}{2} \|v - u\|^2 \right]\).

II. DESIGN OF PROPOSED PENALTY FUNCTION

To evaluate the block-sparsity of \(x \in \mathbb{C}^N\) across the fixed non-overlapping blocks \(B_1, \ldots, B_j \subset \{1, \ldots, N\}\), the mixed \(\ell_2/\ell_1\) norm is widely used as a convex penalty function. The mixed \(\ell_2/\ell_1\) norm is defined as the sum (i.e. the \(\ell_1\) norm) of the block-wise \(\ell_2\) norms:

\[
\|x\|_{2/1,1} := \sum_{k=1}^{j} \sqrt{|B_k|} \|x_{B_k}\|_2, \quad (1)
\]

where we employ the weight \(\sqrt{|B_k|}\) by following the suggestions in, e.g., [5]–[8]. The mixed \(\ell_2/\ell_1\) norm promotes the block-sparsity by pushing components in \(B_k\) toward zeros together. However, since the mixed \(\ell_2/\ell_1\) norm excessively penalizes blocks composed of both zero and nonzero components, its performance degrades when \(B_1, \ldots, B_j\) do not match with the ground-truth \(B_1^\ast, \ldots, B_K^\ast\).

To avoid the block mismatch, the block partition is adapted for \(x\) in the proposed penalty function named latent optimally partitioned (LOP)-\(\ell_2/\ell_1\) penalty. As shown in Fig. 2, we first introduce a nonconvex penalty function \(\psi_K(x)\) as the minimum of the mixed \(\ell_2/\ell_1\) norm over the partition of at most \(K\) blocks. Then, as a suitable convex relaxation of \(\psi_K(x)\), we derive the LOP-\(\ell_2/\ell_1\) penalty \(\Psi_\alpha(x)\). Concretely, we define \(\psi_K(x)\) by

\[
\psi_K(x) := \min_{j \in \{1, \ldots, K\}} \left[ \min_{(B_k)_{k=1}^{j} \in \mathcal{P}_j} \|x|_{2,1}^{(B_k)} \right],
\]

\[
\text{where } \mathcal{P}_j \text{ consists of all } j \text{ block partitions of } \{1, \ldots, N\}, \text{i.e.,}
\]

\[
\bigcup_{k=1}^{j} B_k = \{1, \ldots, N\}, \quad B_k \neq \emptyset \quad (k = 1, \ldots, j),
\]

\[
\bigcap_{k=1}^{j} B_k = \emptyset \quad (k \neq k'),
\]

\[
B_k = \{n \in \{1, \ldots, N\} \mid n_k \leq n \leq m_k\}
\]

\[
\text{for some } n_k, m_k \in \{1, \ldots, N\} \quad (k = 1, \ldots, j).
\]

Note that \(K\) can be set to an upper bound of the ground-truth \(K^\ast\). Note also that, since the sizes of \(B_1, \ldots, B_j\) are adapted,
the non-overlapping condition $B_k \cap B_{k'} = \emptyset$ is not restrictive.

To derive the convex relaxation of $\psi_K(x)$, we exploit the following lemma which shows a variational representation of the $\ell_2$ norm.

**Lemma 1.** Define $\phi : \mathbb{C} \times \mathbb{R} \to \mathbb{R}_+ \cup \{\infty\}$ by

$$\phi(x, \tau) := \begin{cases} \frac{|x|^2}{2\tau} + \frac{\tau}{2}, & \text{if } \tau > 0; \\ 0, & \text{if } x = 0 \text{ and } \tau = 0; \\ \infty, & \text{otherwise}, \end{cases}$$

which is a coercive lower semicontinuous convex function. Then, the block-wise $\ell_2$ norm is variationally represented as

$$\sqrt{|B|} \|x_B\|_2 = \min_{\tau \in \mathbb{R}} \sum_{n \in B} \phi(x_n, \tau),$$

for any $B \subset \{1, \ldots, N\}$.

**Proof.** See Appendix C.

**Remark 1.** Lemma 1 is a slight modification of the existing result

$$\sqrt{|B|} \|x_B\|_2 = \inf_{\tau > 0} \sum_{n \in B} \left[ \frac{|x_n|^2}{2\tau} + \frac{\tau}{2} \right]$$

used in, e.g., [33] for the analysis of the LGL penalty and [50] for the design of a graph-sparse penalty. Since the domain of $\sum_{n \in B} \left[ \frac{|x_n|^2}{2\tau} + \frac{\tau}{2} \right]$ is not closed, this representation is inappropriate for the use in concrete optimization algorithms, as the applicability of [50] is limited [51].

Lemma 1 resolves this issue by using the lower semicontinuous convex function $\phi$.

By applying Lemma 1 for each $\sqrt{|B_k|} \|x_{B_k}\|_2$ in (2), we can rewrite $\psi_K(x)$ as

$$\psi_K(x) = \min_{j \in \{1, \ldots, K\}} \left[ \min_{(B_k)_{k=1}^j \in \mathcal{P}, \tau_j \in \mathbb{R}} \sum_{k=1}^j \sum_{n \in B_k} \phi(x_n, \tau_k) \right].$$

Introducing a latent vector $\sigma = (\sigma_1, \ldots, \sigma_N)^T \in \mathbb{R}^N$ as

$$\sigma_n = \tau_k \text{ for } k = 1, \ldots, j,$$

as illustrated in Fig. 3, we see that $\sigma$ is characterized by the condition that

$$\|D\sigma\|_0 \leq j - 1,$$

where $D$ is the first difference operator defined by

$$D := \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -1 \end{bmatrix} \in \mathbb{R}^{(N-1) \times N},$$

because $B_k$ consists of only consecutive indices for each $k = 1, \ldots, j$. Moreover, from $\bigcup_{k=1}^j B_k = \{1, \ldots, N\}$ and $B_k \cap B_{k'} = \emptyset$ ($k \neq k'$) in (3), we have

$$\sum_{k=1}^j \sum_{n \in B_k} \phi(x_n, \tau_k) = \sum_{n=1}^N \phi(x_n, \sigma_n).$$

Thus, we can represent $\psi_K(x)$ as

$$\psi_K(x) = \min_{j \in \{1, \ldots, K\}} \left[ \min_{\sigma \in \mathbb{R}^N, \|D\sigma\|_0 \leq j - 1} \sum_{n=1}^N \phi(x_n, \sigma_n) \right].$$

Finally, based on the fact that the $\ell_1$ norm is the tightest convex relaxation of $\ell_0$ pseudo-norm in the constraint, we derive the LOP-$\ell_2/\ell_1$ penalty as

$$\Psi_\alpha(x) := \min_{\sigma \in \mathbb{R}^N, \|D\sigma\|_0 \leq \alpha} \sum_{n=1}^N \phi(x_n, \sigma_n),$$

where $\alpha \in \mathbb{R}_+$ is a tuning parameter related to the number of blocks.

**Theorem 1.** For any $\alpha \in \mathbb{R}_+$, $\Psi_\alpha$ takes finite values, i.e., $\text{dom}(\Psi_\alpha) = \mathbb{C}^N$. Moreover, $\Psi_\alpha$ is coercive, continuous, nonnegative, and convex. Note that this implies $\Psi_\alpha \in C_0(\mathbb{C}^N)$.

**Proof.** See Appendix D.

As its special instances, the LOP-$\ell_2/\ell_1$ penalty reproduces the mixed $\ell_2/\ell_1$ norm with the coarsest partition and the finest partition. This is more precisely shown in the next theorem.

**Theorem 2.** The LOP-$\ell_2/\ell_1$ penalty reproduces the mixed $\ell_2/\ell_1$ norm with the largest block $\{1, \ldots, N\}$ and the smallest blocks $\{k\}_{k=1}^N$ respectively for $\alpha = 0$ and $\alpha \to \infty$, i.e.,

$$\Psi_0(x) = \|x\|_2^{\{1, \ldots, N\}} = \sqrt{N} \|x\|_2, \quad \lim_{\alpha \to \infty} \Psi_\alpha(x) = \|x\|_2^{\{1\}} = \|x\|_1.$$

**Proof.** See Appendix E.

This theorem also suggests that the tuning parameter $\alpha$ indeed controls the number of blocks in the LOP-$\ell_2/\ell_1$ penalty, though blocks do not explicitly appear in (6). Note that, from our experiments shown in Section IV, the performance of the LOP-$\ell_2/\ell_1$ penalty seems fairly robust against the choice of $\alpha$. 
Remark 2 (Generalization to higher dimensions). For a matrix \( X \in \mathbb{C}^{N \times M} \), the LOP-\( \ell_2/\ell_1 \) penalty can be naturally extended by replacing the 1D difference operator with a 2D one:

\[
\psi^{(2d)}_\alpha(X) := \min_{\Sigma \in \mathbb{R}^{N \times M}} \sum_{n=1}^{N} \sum_{m=1}^{M} \phi(X_{n,m}, \Sigma_{n,m}),
\]

where \( D_{2d} : \mathbb{R}^{N \times M} \to \mathbb{R}^{(N-1)M + N(M-1)} \) computes differences in vertical and horizontal directions by

\[
D_{2d}(\Sigma) = \begin{bmatrix} D_v(\Sigma) \\ D_h(\Sigma) \end{bmatrix},
\]

\[
D_v(\Sigma) = (\Sigma_{n+1,m} - \Sigma_{n,m})_{(n,m) \in \{1, \ldots, N-1\} \times \{1, \ldots, M\}},
\]

\[
D_h(\Sigma) = (\Sigma_{n,m+1} - \Sigma_{n,m})_{(n,m) \in \{1, \ldots, N\} \times \{1, \ldots, M-1\}}.
\]

Note that we do not only focus on rectangular blocks, and \( \psi^{(2d)}_\alpha \) can manage blocks of various shapes because most components of \( D_{2d}(\Sigma) \) are zeros when \( \Sigma \) is constant on each block, irrespective of block shapes. The LOP-\( \ell_2/\ell_1 \) penalty can be naturally extended to signals in higher dimensions, e.g., \( p \)-th order tensors, by modifying the difference operator in similar ways.

III. PROPOSED BLOCK-SPARSE RECOVERY MODEL AND ITS OPTIMIZATION ALGORITHM

We present a block-sparse recovery model using the LOP-\( \ell_2/\ell_1 \) penalty (6), and develop an iterative algorithm which converges to an optimal solution of the model. Specifically, we consider the following regularization model:

\[
\minimize_{x \in \mathbb{C}^N} f(Lx) + \lambda \psi_\alpha(x),
\]

where \( f(Lx) \) is some convex data-fidelity function with \( f : \mathbb{C}^L \to \mathbb{R}^+ \) and \( L \in \mathbb{C}^{L \times N} \), and \( \lambda > 0 \) is the regularization parameter. We suppose that \( f \in \Gamma_0(\mathbb{C}^L) \), and its proximity operator can be computed efficiently. Such examples include the square error \( f(u) = \frac{1}{2} \| y - Au \|^2 \) and \( L = I \), and the absolute error \( f(u) = \| y - u \|_1 \) and \( L = A \), where \( y \) is the known observation vector, \( A \) is the known measurement matrix, and \( I \) denotes the identity matrix. The existence of the solution of (10) is guaranteed by the coercivity of \( \psi_\alpha \in \Gamma_0(\mathbb{C}^N) \), as shown in the next theorem.

Theorem 3. For any \( \lambda > 0 \), the proposed model (10) has an optimal solution.

Proof. It follows from Theorem 1 by [52, Corollary 11.16].

We show that an optimal solution of the proposed model (10) can be obtained by utilizing proximal splitting techniques [38]–[46] as follows. By plugging the definition of \( \psi_\alpha(x) \) in (6) into (10), the optimization problem (10) is translated into

\[
\minimize_{(x, \sigma) \in \mathbb{C}^N \times \mathbb{R}^N} f(Lx) + \lambda \varphi(x, \sigma),
\]

subject to \( \| D\sigma \|_1 \leq \alpha \),

where we let

\[
\varphi(x, \sigma) := \sum_{n=1}^{N} \phi(x_n, \sigma_n).
\]

To apply the primal-dual algorithm [38]–[43] shown in Appendix B, by introducing auxiliary variables \( u = Lx \) and \( \eta = D\sigma \), we further rewrite the optimization problem (11) as

\[
\begin{align*}
\minimize_{(x, \sigma, u, \eta) \in \mathbb{C}^N \times \mathbb{R}^N \times \mathbb{C}^L \times \mathbb{R}^{L^2}} & \quad f(u) + \lambda \varphi(x, \sigma) + \mu \beta^G_x(\eta) \\
\text{subject to} & \quad \begin{bmatrix} \mu_1 L & O & -\mu_1 I & O \\ O & \mu_2 D & O & -\mu_2 I \end{bmatrix} \begin{bmatrix} x \\ \sigma \\ u \\ \eta \end{bmatrix} = 0
\end{align*}
\]

(12)

where \( \beta^G_x(\eta) \) is the indicator function of the \( \ell_1 \) ball and defined by

\[
\beta^G_x(\eta) := \begin{cases} 0, & \text{if } \| \eta \|_1 \leq \alpha; \\
\infty, & \text{otherwise},
\end{cases}
\]

\( I \) and \( O \) respectively denote the identity and zero matrices of appropriate sizes, and \( \mu_1 \) and \( \mu_2 \) satisfying

\[
0 < \mu_1 \leq \frac{1}{\sqrt{\|L\|_{op}^2 + 1}},
\]

\[
0 < \mu_2 \leq \frac{1}{\sqrt{\|D\|_{op}^2 + 1}}
\]

(13)

are introduced so that \( \| H \|_{op} \leq 1 \).

Applying the primal-dual algorithm (26) in Appendix B to the problem (12), we obtain the proposed algorithm shown in Algorithm 1, where the updates are written in terms of \((x, \sigma, u, \eta)\). The proximity operator of \( \gamma G \) in (26) reduces to the proximity operators of \( \gamma \lambda \varphi \), \( \gamma f \), and \( \gamma \beta^G_x \), since \( \gamma G \) is the separable sum of these functions. Based on [48, Example 2.4], we can compute the proximity operator of \( \gamma \lambda \varphi \) as

\[
\text{prox}_{\gamma \lambda \varphi}(x, \sigma) = \left( \text{prox}_{\gamma \varphi}(x_n, \sigma_n) \right)_{n=1}^{N},
\]

(14)

with

\[
\text{prox}_{\gamma \varphi}(x, \sigma) = \begin{cases} (0, 0), & \text{if } 2\gamma \lambda \sigma + |x|^2 \leq \gamma^2 \lambda^2; \\
(0, \sigma - \frac{x}{\lambda^2}), & \text{if } x = 0 \text{ and } 2\sigma > \gamma \lambda; \\
\left( -x, \sigma - \frac{x^2}{\lambda^2}, \sigma + \gamma \lambda \frac{x^2 - 1}{2} \right), & \text{otherwise},
\end{cases}
\]

(15)

where \( s > 0 \) is the unique positive root of

\[
s^3 + \left( \frac{2\sigma}{s} + 1 \right)s - \frac{2\sigma}{s^3} = 0,
\]

and can be explicitly given via Cardano’s formula as follows. Let \( p = \frac{2\sigma}{s} + 1, q = -\frac{2\sigma}{s^3}, \) and \( D = -\frac{q^2}{4} - \frac{p^3}{27} \). Then,

\[
s = \begin{cases} \sqrt{-\frac{q}{2} + \sqrt{-D}} + \sqrt{-\frac{q}{2} - \sqrt{-D}}, & \text{if } D < 0; \\
2 \sqrt{-\frac{q}{2}}, & \text{if } D = 0; \\
2 \sqrt{-\frac{q}{4} + D \cos \left( \arctan(-2\sqrt{\frac{q}{D}}) \right)}, & \text{if } D > 0,
\end{cases}
\]

for \( s > 0 \),
Choose $\gamma > 0$, $\mu_1$, and $\mu_2$ satisfying (13), and initial values $x^{(0)} \in \mathbb{C}^N$, $\sigma^{(0)} \in \mathbb{R}^N$, $u^{(0)} \in \mathbb{C}^J$, $\eta^{(0)} \in \mathbb{R}^{N-1}$, $r_1^{(0)} \in \mathbb{C}^J$, and $r_2^{(0)} \in \mathbb{R}^{N-1}$. Then, $(x^{(i)})_{i=1}^{\infty}$ generated by Algorithm 1 converges to an optimal solution of (10).

Proof. See Appendix F.

**Remark 3** (Computational complexity of proposed method).
We summarize the computational cost of Algorithm 1 per an iteration. We can compute $\tilde{x}^{(i+1)}$, $\tilde{u}^{(i+1)}$, and $r_1^{(i+1)}$ with $\mathcal{O}(|\text{supp}(L)|)$ operations, and $\tilde{v}^{(i+1)}$, $\tilde{q}^{(i+1)}$, and $r_2^{(i+1)}$ with $\mathcal{O}(N)$ operations since $D$ has only $2(N-1)$ nonzero entries. The proximity operator of $\gamma \lambda_C$ can be computed by (14) and (15) with $\mathcal{O}(N)$ operations. By setting $L = A$, prox$_{f}$ can be computed with $\mathcal{O}(N)$ operations for both the absolute error and the square error (see footnote 1). Note that, for the square error, setting $L = I$ would be beneficial when, e.g., $A^H A$ is diagonal. The computation of the $\ell_1$ ball projection $P_{B_1}$ by (18) and (19) requires $\mathcal{O}(N \log N)$ operations for sorting. We can also use sophisticated $\ell_1$ ball projection algorithms having $\mathcal{O}(N)$ expected complexity, e.g., [53]. In particular, this discussion implies that computations regarding the LOP-$\ell_2/\ell_1$ penalty can be implemented with $\mathcal{O}(N)$ operations. Thus, per an iteration, the proposed method has the same computational complexity with the mixed $\ell_2/\ell_1$ regularization method using the non-overlapping blocks when the primal-dual algorithm (26) in Appendix B is applied.

**IV. Numerical Experiments**
To show the effectiveness of the proposed block-sparse recovery model (10) using the LOP-$\ell_2/\ell_1$ penalty (6), we conduct numerical experiments on both synthetic and real-world data. We compare the LOP-$\ell_2/\ell_1$ model with regularization models using existing convex penalties: the $\ell_1$ norm, the mixed $\ell_2/\ell_1$ norm (1), and the LGL penalty (24) in Appendix A.

**A. Synthetic Examples**
We consider the estimation of a block-sparse signal $x^* \in \mathbb{R}^N$ from noisy compressive measurements. More precisely, we define the measurements by $y := Ax^* + \varepsilon$, where the entries of $A \in \mathbb{R}^{d \times N}$ ($d < N$) are drawn from i.i.d. Gaussian distribution $\mathcal{N}(0, 1)$, and $\varepsilon \in \mathbb{R}^d$ is the white Gaussian noise. The block-sparse signal $x^*$ is randomly generated by the following scheme. We set $N = 250$, and $x^*$ has 80 nonzero components, which are randomly divided into 4 blocks. The blocks are randomly located under the condition $|\text{supp}(x^*)| = 80$. Amplitudes of nonzero components are drawn from i.i.d. $\mathcal{N}(0, 1)$.

In the proposed regularization model (10), we use the square error by setting $f(u) = \frac{1}{2}\|y - Au\|_2^2$ and $L = I$. The existing penalties are also combined with the square error in similar ways. Note that the regularization parameter is tuned independently for each model to obtain the best accuracy. The proposed model (10) is solved by Algorithm 1, where we set $\gamma = 10^{-3}$, $\mu_1 = 1/\sqrt{2}$, and $\mu_2 = 1/\sqrt{5}$ which satisfy the condition (13) for the convergence to an optimal solution. The existing models are also solved by applying the primal-dual algorithm shown in Appendix B. We terminate the iteration when the norm of the differences between the variables of successive iterates is below the threshold $10^{-4}$.

We compare the proposed model (10) and the existing models in terms of the normalized mean square error (NMSE) $\frac{\|x^* - \hat{x}\|_2^2}{\|x^*\|_2^2}$.
where $\hat{x}$ is the solution of each model. In Figs. 4(a) and (b), we show the NMSE respectively against the number $d$ of measurements and the SNR $E[\|Ax^*\|^2_2]/E[\|e\|^2_2]$, where the results are averaged over 100 independent trials. The mixed $\ell_2/\ell_1$ model performs better than the existing models. It can also be seen that the performance of the LOP-$\ell_2/\ell_1$ model is fairly robust against the choice of the tuning parameter $\alpha$ which indirectly controls the number of blocks.

### B. Application to Phased Array Weather Radar

A major goal of the phased array weather radar (PAWR) [24], [25] is to estimate the backscattered signal $X^* \in \mathbb{C}^{M \times N}$ from noisy observations $Y := AX^* + E \in \mathbb{C}^{d \times N}$ obtained by transmitting fan beams, where $A \in \mathbb{C}^{d \times M}$ is a known array manifold matrix, $E \in \mathbb{C}^{d \times N}$ is the white Gaussian noise, and $M$, $N$, and $d$ respectively are the numbers of elevation angles, pulses, and array elements. This estimation problem is called beamforming in the PAWR literature [21]–[23]. As shown in [22], [23], the backscattered signal $X^*$ exhibits block-sparsity in the Fourier domain for each elevation angle. Since the block partition depends on the unknown mean and standard deviation of the Doppler frequency at each elevation angle, it is also unknown and different for each elevation angle. Thus, the proposed approach is suitable for the PAWR beamforming. Specifically, we design a LOP-$\ell_2/\ell_1$ model for the PAWR beamforming as

$$\minimize_{X \in \mathbb{C}^{M \times N}} \frac{1}{2}\|Y - AX\|_F^2 + \lambda\Psi_\alpha^{(2d)}(FX^T), \quad (20)$$

where $\| \cdot \|_F$ denotes the Frobenius norm, and $F \in \mathbb{C}^{N \times N}$ is the normalized discrete Fourier transform matrix. To exploit the column-wise block-sparsity of $FX^T$, we use $\Psi_\alpha^{(2d)}$ slightly modified from (9) by replacing $D_{2d}$ with the vertical difference operator $D_v$. To apply Algorithm 1, we rewrite (20) in the form of (11) as follows. Since $F$ is a unitary matrix, we have

$$\|Y - AX\|_F^2 = \|Y^T - X^TA^T\|_F^2 = \|FY^T - FX^TA^T\|_F^2.$$ 

Thus, by letting $Z = FX^T$ and substituting the definition of $\tilde{\psi}_\alpha^{(2d)}$, the optimization problem (20) can be translated into

$$\minimize_{(Z, \Sigma) \in \mathbb{C}^{N \times M} \times \mathbb{R}^{N \times M}} \frac{1}{2}\|FY^T - Z\|_F^2 + \lambda\tilde{\varphi}(Z, \Sigma)$$

subject to $\|D_v(\Sigma)\|_1 \leq \alpha$, \quad (21)

where

$$\tilde{\varphi}(Z, \Sigma) := \sum_{n=1}^N \sum_{m=1}^M \phi(Z_{nm}, \Sigma_{nm}).$$

Further, by defining $\bar{A} \in \mathbb{C}^{dN \times NM}$ and $\bar{D}_v \in \mathbb{R}^{(N-1)M \times NM}$ satisfying

$$\begin{align*}
(\forall Z \in \mathbb{C}^{N \times M}) \quad & \text{vec}(ZA^T) = \bar{A}\text{vec}(Z), \\
(\forall \Sigma \in \mathbb{R}^{N \times M}) \quad & D_v(\Sigma) = \bar{D}_v\text{vec}(\Sigma),
\end{align*}$$

and $\bar{y} := \text{vec}(FY^T) \in \mathbb{C}^{dN}$, we can rewrite (21) as

$$\begin{align*}
\minimize_{(Z, \Sigma) \in \mathbb{C}^{N \times M} \times \mathbb{R}^{N \times M}} \frac{1}{2}\|\bar{y} - \bar{A}Z\|_2^2 + \lambda\tilde{\varphi}(\bar{z}, \bar{\Sigma})
\quad \text{subject to } \|\bar{D}_v(\Sigma)\|_1 \leq \alpha,
\end{align*}$$

\quad (22)

Since (22) is in the form of (11), we can solve (22) by Algorithm 1 with setting $f(\bar{z}) = \|\bar{y} - \bar{A}Z\|_2^2$, $L = \bar{I}$, and $D = \bar{D}_v$. From $\bar{z}$, the solution of (22), we obtain the solution of (20) by $X = (F^H\text{vec}^{-1}(\bar{z}))^T$, where $\text{vec}^{-1}$ is the inverse of the vectorization operator. Existing penalties are also combined with the square error and solved in similar ways.

2We define the vectorization operator by $\text{vec}(V) := (v_1^T, \ldots, v_q^T)^T \in \mathbb{C}^{pq}$ for $V := (v_1, \ldots, v_q) \in \mathbb{C}^{p \times q}$.
where the regularization parameters are tuned independently to obtain the best results.

We conduct a numerical simulation in a setting similar to [21]–[23]. The backscattered signal $X^* \in \mathbb{C}^{M \times N}$ is generated based on the real reflection intensity measured by the PAWR at Osaka University,\(^3\) where the elevation angles are chosen uniformly from $-15^\circ$ to $30^\circ$ degrees with $M = 110$, and the mean and the standard deviation of the Doppler frequency are respectively drawn from uniform distributions of $[-1.25, 1.25]$ and $[0.0314, 0.1257]$ in kHz for each elevation angle. We set $N = 50$ and $d = 128$, and $\varepsilon$ as the white Gaussian noise of the standard deviation $\sqrt{5}$. Table I shows the NMSE averaged over 50 independent trials, where the LGL model uses overlapping blocks of size 2, and $\ell_2/\ell_1$ (a) and (b) respectively refer to the mixed $\ell_2/\ell_1$ models using block sizes 2 and 4. The result shows that the LOP-$\ell_2/\ell_1$ model achieves better estimation accuracy than the existing models, and is quite robust against the choice of the tuning parameter $\alpha$. In Fig. 5, we show examples of the power spectrums of the ground-truth and the estimates, i.e., squared magnitudes of entries of $FX^*\mathbf{T}$ and $FX^\mathbf{T}$. It can be seen that the LOP-$\ell_2/\ell_1$ model reduces the artifacts caused in the LGL model, e.g., in the zoomed areas of Fig. 5.

\begin{table}
\centering
\caption{Comparison of the regularization models for simulation of phased array weather radar in terms of NMSE in dB averaged over 50 independent trials.}
\begin{tabular}{|c|c|c|c|c|c|c|}
\hline
 & $\text{LOP-}\ell_2/\ell_1$ ($\alpha = 120M$) & $\text{LOP-}\ell_2/\ell_1$ ($\alpha = 130M$) & $\text{LOP-}\ell_2/\ell_1$ ($\alpha = 140M$) & $\ell_2/\ell_1$ (a) & $\ell_2/\ell_1$ (b) & $\ell_1$ & LGL \\
\hline
NMSE & $-15.62$ & $-15.72$ & $-15.71$ & $-14.78$ & $-14.03$ & $-13.39$ & $-14.83$ \\
\hline
\end{tabular}
\end{table}

\subsection{Application of Proposed 2D Penalty to Speech Denoising}

To show the effectiveness of the 2D extension (9) of the LOP-$\ell_2/\ell_1$ penalty, we conduct experiments on the speech denoising exploiting block-sparsity of the spectrogram. We generate noisy speech as $y := s^* + \varepsilon \in \mathbb{R}^d$, where $s^*$ is a 2-second clip of speech taken from [54] with 16kHz sampling rate (i.e. $d = 32000$), and $\varepsilon$ is the white Gaussian noise. The spectrogram $X^* := \mathcal{F}(s^*)$ of the clean speech, where $\mathcal{F}$ denotes the short-time Fourier transform (STFT), is expected to be block-sparse because entries of $X^*$ in adjacent frequency bins and frames tend to be zeros or nonzeros together. The block partition depends on the characteristics of the speech to be recovered, and thus is unknown a priori. Hence, the proposed approach is suitable for exploiting the block-sparsity of the spectrogram. Specifically, we apply the 2D LOP-$\ell_2/\ell_1$ penalty in (9) to the speech denoising as

\[
\hat{X} \in \arg \min_{X \in \mathbb{C}^{N \times M}} \frac{1}{2} \|y - \mathcal{F}^{-1}(X)\|_F^2 + \lambda \Psi_{\text{S2d}}(X),
\]

where $N$ and $M$ are respectively the numbers of frequency bins and frames, and $\mathcal{F}^{-1}$ denotes the inverse STFT. We implement the STFT as a semi-orthogonal transformation by using the hann window of 32ms (i.e. 512 samples) with 75\% overlap and appropriate zero padding. Note that this setting implies $N = 512$ and $M = 253$. Since $D_{2d}$ and $\mathcal{F}^{-1}$ are linear operators, we can solve (23) by Algorithm I through reformulation with the vectorization of $\hat{X}$, similarly to the reformulation of (21) into (22), where we should choose $\mu_2 \leq 1/3$ according to the operator norm of $D_{2d}$. Note that $\mathcal{F}^{-1}$ can be computed as $\mathcal{F}^*$, the adjoint of $\mathcal{F}$, thanks to the semi-orthogonality of $\mathcal{F}$. The existing penalties are used in manners similar to (23), where the regularization parameters are tuned independently to obtain the best NMSE.

Experiments are conducted for male and female speech with input SNRs of 5dB and 10dB. In Table II, we show the NMSE

\[
\frac{\|s^* - \mathcal{F}^{-1}(\hat{X})\|_F^2}{\|s^*\|_2^2},
\]

averaged over 20 independent trials, where $\ell_2/\ell_1$ (a) and (b) respectively denote the mixed $\ell_2/\ell_1$ models using block sizes $2 \times 4$ and $2 \times 2$, which yield best results among
block sizes $\{1, 2, 4, 8\} \times \{1, 2, 4, 8\}$. We here do not include the LGL model since the use of 2D overlapping blocks is computationally expensive. From Table II, we see that the 2D LOP-$\ell_2/\ell_1$ model achieves better estimation accuracy than the existing models for every experimental condition. In addition, the performance of the 2D LOP-$\ell_2/\ell_1$ model is fairly robust against the choice of $\alpha$. Fig. 6 illustrates the magnitude spectrograms of the clean speech and the estimates, i.e., magnitudes of entries of $X^*$ and $\hat{X}$, for the denoising of male speech with 10dB input SNR. Rectangular artifacts are observed in Fig. 6(b) which shows the magnitude spectrogram obtained by the mixed $\ell_2/\ell_1$ model with the block size $2 \times 4$. These artifacts are considered to be caused by the rectangular blocks used in the mixed $\ell_2/\ell_1$ norm. Note that, since the blocks are used explicitly, it is difficult to employ blocks of various shapes in the mixed $\ell_2/\ell_1$ norm due to the computational complexity. On the other hand, the 2D LOP-$\ell_2/\ell_1$ penalty can handle blocks of various shapes because the blocks are implicitly controlled with the 2D difference operator $D_{2d}$ (see Remark 2). Indeed, from Fig. 6(c), we see that the 2D LOP-$\ell_2/\ell_1$ model considerably reduces such artifacts, and more accurately recovers detailed shapes of the spectrogram.

V. CONCLUSION

We presented a convex recovery method for block-sparse signals whose block partitions are not available a priori. We first introduce a nonconvex penalty function $\psi_K$ in (2), where the block partition is optimized for the signal of interest by minimizing the mixed $\ell_2/\ell_1$ norm. Then, as the tight convex relaxation of $\psi_K$, we derive the proposed latent optimally partitioned (LOP)-$\ell_2/\ell_1$ penalty $\Psi_\alpha$ in (6). For the block-sparse recovery model (10) using the LOP-$\ell_2/\ell_1$ penalty, by applying proximal splitting techniques, we develop the iterative solver as Algorithm 1 which is guaranteed to converge to an optimal solution. Numerical experiments on both synthetic and real-world data illustrate the effectiveness of the proposed method.

APPENDIX A

BRIEF REVIEW OF CONVEX BLOCK-SPARSE PENALTIES USING OVERLAPPING BLOCKS

A simple modification of the mixed $\ell_2/\ell_1$ norm is to use, in (1), overlapping blocks $\mathcal{B}_1, \ldots, \mathcal{B}_K \subset \{1, \ldots, N\}$ such that

$$\bigcup_{k=1}^K \mathcal{B}_k = \{1, \ldots, N\},$$

$$\mathcal{B}_k \neq \emptyset \quad (k = 1, \ldots, K),$$

$$\mathcal{B}_k = \{n \in \{1, \ldots, N\} \mid \bar{n}_k \leq n \leq \bar{m}_k\}$$

for some $\bar{n}_k, \bar{m}_k \in \{1, \ldots, N\}$ ($k = 1, \ldots, K$),

where the non-overlapping condition $\mathcal{B}_k \cap \mathcal{B}_{k'} = \emptyset$ ($k \neq k'$) is omitted, e.g., as in [26]–[31]. However, this modification cannot remove the blocks containing both zero and nonzero components. To alleviate this drawback, the latent group lasso (LGL) method [32], [33] is presented to select relevant blocks from the pre-defined overlapping blocks $\mathcal{B}_1, \ldots, \mathcal{B}_K$. Specifically, the LGL penalty is defined by

$$\text{LGL}(x) := \min_{(w_1, \ldots, w_K) \in \mathbb{C}^{N \times K}} \sum_{k=1}^K \sqrt{|\mathcal{B}_k|} \|w_k\|_2.$$  \hspace{1cm} (24)

*By setting $\mathcal{B}_k$ to other than blocks, the LGL method is applicable to other structural sparsity, while this paper focuses on the block-sparsity.
The LGL penalty can also be written as
\[
\text{LGL}(x) = \min_{\tilde{w} \in \mathbb{C}^{\tilde{N}}} \sum_{k=1}^{\tilde{K}} \sqrt{\|\mathcal{B}_k\|} \|\tilde{w}_{\mathcal{I}_k}\|_2,
\]
where \( \tilde{N} = \sum_{k=1}^{\tilde{K}} |\mathcal{B}_k| \).

\[
\mathcal{I}_k := \left\{ n \in \{1, \ldots, \tilde{N}\} \mid 1 + \sum_{m=1}^{k-1} |\mathcal{B}_m| \leq n \leq \sum_{m=1}^{k} |\mathcal{B}_m| \right\},
\]

\[
G := \begin{bmatrix} G_1 & \cdots & G_{\tilde{K}} \end{bmatrix} \in \mathbb{R}^{N \times \tilde{N}},
\]

\[
G_k := \begin{bmatrix} O_{n_k-1,|\mathcal{B}_k|} & I_{|\mathcal{B}_k|} \\ I_{|\mathcal{B}_k|} & O_{N-n_k,|\mathcal{B}_k|} \end{bmatrix},
\]

with \( I_p \in \mathbb{R}^{p \times p} \) and \( O_{p,q} \in \mathbb{R}^{p \times q} \) respectively denoting the identity matrix of order \( p \) and the zero matrix of size \( p \times q \). Thus, when the LGL penalty is combined with some convex data-fidelity function, the resulting regularization model can be solved as a convex optimization problem of \( \tilde{w} \in \mathbb{C}^{\tilde{N}} \). However, it is intractable to use all possible blocks as \( \mathcal{B}_1, \ldots, \mathcal{B}_{\tilde{K}} \) because the problem size becomes too large, i.e., \( \tilde{N} = N(N+1)(N+2)/6 \), in this case. Due to this issue, \( \mathcal{B}_1, \ldots, \mathcal{B}_{\tilde{K}} \) are typically restricted to blocks of a fixed size. On the other hand, the proposed approach is free from such issues because the block partition is implicitly optimized in the proposed penalty (6) with the latent vector \( \sigma \) of size \( N \).

**APPENDIX B**

**PRIMAL-DUAL ALGORITHM FOR NONSMOOTH CONVEX OPTIMIZATION PROBLEM**

The Chambolle-Pock primal-dual algorithm [38], [39] and the Loris-Verhoeven algorithm [40], [41] are widely used for solving nonsmooth convex optimization problems. As explained in [46], these algorithms are essentially the same for the following problem

\[
\text{minimize} \ G(v) \text{ subject to } Hv = 0,
\]

and the iteration can be written as

\[
\begin{align*}
\nu^{(i+1)} &= \nu^{(i)} + \mathcal{H}^*(r^{(i)} - Hv^{(i)}) \\
\nu^{(i+1)} &= \text{prox}_{\gamma G}(\tilde{v}^{(i+1)}) \\
\nu^{(i+1)} &= \tilde{v}^{(i+1)} - Hv^{(i+1)},
\end{align*}
\]

where we suppose that \( G \in \Gamma_0(V) \), \( H : V \to W \) is a linear operator, \( H^* : W \to V \) is the adjoint of \( H \), \( V \) and \( W \) are finite-dimensional Hilbert spaces, and \( \gamma > 0 \). This algorithm is also known as the linearized augmented Lagrangian algorithm [42], [43]. The convergence of (26) to an optimal solution of (25) is shown in the following lemma from, e.g., [46, Theorem 4.3] and [42, Corollary 1].

**Lemma 2.** Suppose that an optimal solution of (25) exists, the so-called qualification condition\(^3\) \( 0 \in \text{ri}(H(\text{dom}(G))) \) holds, \( \|H\|_{op} \leq 1 \), and \( \gamma > 0 \). Let \( v^{(0)} \in V \) and \( r^{(0)} \in W \). Then, \((v^{(i)})_{i=1}^{\infty} \) generated by (26) converges to an optimal solution of (25).

Note that, when \( \|H\|_{op} > 1 \), by replacing \( H \) in (25) with \( H/C \) where \( C \geq \|H\|_{op} \), the condition \( \|H\|_{op} \leq 1 \) can be satisfied without changing the solution set of (25). Thus, the iteration (26) can compute an optimal solution of (25) for general \( H \).

**APPENDIX C**

**PROOF OF LEMMA 1**

First, we prove the properties of \( \phi \) defined by (4). Notice that \( \phi \) can be written as

\[
\phi(x, \tau) = \frac{h(x, \tau)}{2} + \frac{\tau}{2},
\]

where

\[
h(x, \tau) := \begin{cases} |x|^2 & \text{if } \tau > 0; \\
0 & \text{if } x = 0 \text{ and } \tau = 0; \\
\infty & \text{otherwise.}
\end{cases}
\]

Since \( h \) is the perspective function of \( |x|^2 \), it is lower semicontinuous convex [55, Proposition 2.3 and Example 2.6], and thus \( \phi \) is lower semicontinuous convex. The coercivity of \( \phi \) is proven by showing \( \phi(x, \tau) \to \infty \) \( (|x| + |\tau|) \to \infty \) as follows.

a) (Case of \( |\tau| \to \infty \)) Clear from \( \phi(x, \tau) \geq \frac{\tau^2}{2} \).

b) (Case of \( |\tau| \) is bounded) In this case, we have \( |x| \to \infty \) and \( \exists R > 0 \) such that \( |\tau| \leq R \). Moreover, since \( \phi(x, \tau) = \infty \) when \( \tau \leq 0 \) and \( x \neq 0 \), we have

\[
(\forall x \neq 0) \ \ \phi(x, \tau) \geq \frac{|x|^2}{R},
\]

which implies \( \phi(x, \tau) \to \infty \) due to \( |x| \to \infty \).

Next, we prove the equation (5) as follows.

a) (Case of \( x_B \neq 0 \)) For \( \tau > 0 \), we have

\[
\sum_{n \in B} \phi(x_n, \tau) = \sum_{n \in B} \left( \frac{|x_n|^2}{2\tau} + \frac{\tau}{2} \right) = \frac{\|x_B\|^2_2}{2\tau} + \frac{|B|^2}{2\tau}.
\]

Since \( x_B \neq 0 \) in this case, \( \frac{\|x_B\|^2_2}{2\tau} + \frac{|B|^2}{2\tau} \) is a strictly convex function of \( \tau > 0 \), and the minimum is attained when \( \tau = \frac{\|x_B\|^2_2}{\|B\|^2} \). Thus, we have

\[
\min_{\tau > 0} \sum_{n \in B} \phi(x_n, \tau) = \sqrt{|B|\|x_B\|^2}.
\]

For \( \tau = 0 \), since there exists \( n \in B \) such that \( x_n \neq 0 \) in this case, we have \( \sum_{n \in B} \phi(x_n, 0) = \infty \). For \( \tau < 0 \), the definition of \( \phi \) in (4) readily implies \( \sum_{n \in B} \phi(x_n, \tau) = \infty \).

Summarizing, we have the equation (5).

b) (Case of \( x_B = 0 \)) For \( \tau > 0 \), since \( x_n = 0 \) for every \( n \in B \) in this case, we have

\[
\sum_{n \in B} \phi(x_n, \tau) = \sum_{n \in B} \left( \frac{|x_n|^2}{2\tau} + \frac{\tau}{2} \right) = \frac{\tau}{2} > 0.
\]

For \( \tau = 0 \), we have

\[
\sum_{n \in B} \phi(x_n, 0) = 0.
\]
For $\tau < 0$, we have $\sum_{n \in B} \phi(x_n, \tau) = \infty$. Thus, the minimum is attained when $\tau = 0$, and

$$\min_{\tau \in \mathbb{R}} \sum_{n \in \mathbb{R}} \phi(x_n, \tau) = 0,$$

which implies (5) since $x_B = 0$ in this case.

APPENDIX D
PROOF OF THEOREM 1

Since $\Psi_\alpha$ is defined by the partial minimization of the coercive lower semicontinuous convex function $\sum_{n=1}^{N} \phi(x_n, \sigma_n)$ under the closed convex constraint $\|D\sigma\|_1 \leq \alpha$, and

$$\left\{ \sigma \in \mathbb{R}^{N} \mid \sum_{n=1}^{N} \phi(x_n, \sigma_n) < \infty \text{ and } \|D\sigma\|_1 \leq \alpha \right\} \neq \emptyset$$

for any $\alpha \in \mathbb{R}_+$, $\Psi_\alpha$ takes finite values for every $x \in \mathbb{C}^N$, i.e., $\text{dom}(\Psi_\alpha) = \mathbb{C}^N$, by [52, Proposition 11.15]. Note that (27) holds for $\alpha \in \mathbb{R}_+$, since $D\sigma = 0$ for any $\sigma$ satisfying $\sigma_1 = \sigma_2 = \cdots = \sigma_N$. The convexity of $\Psi_\alpha$ follows from the convexity of $\sum_{n=1}^{N} \phi(x_n, \sigma_n)$ and $\|D\sigma\|_1 \leq \alpha$ by [52, Proposition 8.35]. Since $\Psi_\alpha$ is convex and $\text{dom}(\Psi_\alpha) = \mathbb{C}^N$, $\Psi_\alpha$ is continuous by [52, Corollary 8.40]. From $\sum_{n=1}^{N} \phi(x_n, \sigma_n) \geq 0$, $\Psi_\alpha$ is nonnegative. Lastly, the coercivity of $\Psi_\alpha$ follows from

$$\Psi_\alpha(x) = \min_{\sigma \in \mathbb{R}^{N}} \sum_{n=1}^{N} \phi(x_n, \sigma_n)$$

$$\geq \min_{\sigma \in \mathbb{R}^{N}} \sum_{n=1}^{N} \phi(x_n, \sigma_n)$$

$$= \sum_{n=1}^{N} \min_{\sigma_n \in \mathbb{R}} \phi(x_n, \sigma_n) = \|x\|_1,$$

where the last equality is obtained by applying Lemma 1 for each of $\min_{\sigma_n \in \mathbb{R}} \phi(x_n, \sigma_n)$ ($n = 1, \ldots, N$).

APPENDIX E
PROOF OF THEOREM 2

The equation (7) is proven as

$$\Psi_0(x) = \min_{\sigma \in \mathbb{R}^{N}} \sum_{n=1}^{N} \phi(x_n, \sigma_n)$$

$$= \min_{\sigma \in \mathbb{R}} \sum_{n=1}^{N} \phi(x_n, \sigma) = \sqrt{N} \|x\|_2,$$

where the second equality holds due to

$$\|D\sigma\|_1 \leq 0 \Leftrightarrow D\sigma = 0 \Leftrightarrow \sigma_1 = \sigma_2 = \cdots = \sigma_N,$$

and the last equality follows from Lemma 1. Next, (8) is proven as

$$\lim_{\alpha \to \infty} \Psi_\alpha(x) = \min_{\sigma \in \mathbb{R}^{N}} \sum_{n=1}^{N} \phi(x_n, \sigma_n) = \|x\|_1,$$

where the last equality holds by Lemma 1.

APPENDIX F
PROOF OF THEOREM 4

First, the existence of the solution of (12) is guaranteed from Theorem 3 and the equivalence between (10), (11), and (12). The qualification condition for (12) is confirmed as follows.

The effective domain of $G$ is given by

$$\text{dom}(G) = \text{dom}(\phi) \times \text{dom}(f) \times \text{dom}(\mu_{B_1^c})$$

$$= ((\mathbb{C}^N \times \mathbb{R}_+^{N}) \cup \{0\}) \times \mathbb{C}^J \times B_1^0,$$

where $\mathbb{R}_+$ denotes the set of all positive real numbers, and $B_1^0 = \{ \eta \in \mathbb{R}^{N-1} \mid \|\eta\|_1 \leq \alpha \}$. Thus, we have

$$H(\text{dom}(G)) = \left\{ \begin{array}{ll}
\mu_1 Lx - \mu_1 u \\
\mu_2 D\sigma - \mu_2 u
\end{array} \right\} \in \text{dom}(G)$$

$$= (L(\mathbb{C}^N - \mathbb{C}^J) \times (D(\mathbb{R}^N) - \mu_2 B_1^0),$$

for the relation $D(\mathbb{R}^N) = D(\mathbb{R}^N)$. This relation holds because, for any $\sigma \in \mathbb{R}^N$, $D(\sigma + c) = D\sigma$ and $\sigma + c \in \mathbb{R}^N$ can be satisfied with some $c \in \mathbb{R}_+$, where $1 := (1, 1, \ldots, 1)^T \in \mathbb{R}_+^N$. Since the expression (28) implies that $H(\text{dom}(G))$ is a nonempty convex set satisfying the symmetry $H(\text{dom}(G)) = -H(\text{dom}(G))$, the qualification condition 0 lies in $H(\text{dom}(G))$ holds by [52, Example 6.10].

The condition $\|H\|_\infty \leq 1$ is checked as follows. Since

$$HH^H = \begin{bmatrix}
\mu_1^2 (LL^H + I) \\
O
\end{bmatrix}$$

is block-diagonal, with $\rho(\cdot)$ denoting the largest eigenvalue of the argument, we have

$$\|H\|_\infty = \sqrt{\rho(HH^H)}$$

$$= \max \left\{ \sqrt{\rho(\mu_1^2 (LL^H + I))}, \sqrt{\rho(\mu_2^2 (DD^T + I))} \right\}$$

$$= \max \left\{ \mu_1 \sqrt{\rho(LL^H) + 1}, \mu_2 \sqrt{\rho(DD^T) + 1} \right\}$$

$$= \max \left\{ \mu_1 \sqrt{\|L\|_\infty^2 + 1}, \mu_2 \sqrt{\|D\|_\infty^2 + 1} \right\} \leq 1,$$

where the last inequality holds since $\mu_1$ and $\mu_2$ are chosen to satisfy (13). Thus, by Lemma 2, $(x^{(i)}, \sigma^{(i)}, u^{(i)}, \eta^{(i)})_{i=1}^{\infty}$ converges to an optimal solution of (12). The convergence of $(x^{(i)})_{i=1}^{\infty}$ to the solution of (10) follows from the relation between (10), (11), and (12).

ACKNOWLEDGMENT

We would like to acknowledge helpful discussions with Dr. Akira Hirabayashi. We would also like to thank Prof. Tomoo Ushio, Dr. Hiroshi Kikuchi, and Dr. Eiichi Yoshikawa for providing us with the PAWR data.

REFERENCES

[1] M. Stojnic, F. Parvaresh, and B. Hassibi, “On the reconstruction of block-sparse signals with an optimal number of measurements,” IEEE Trans. Signal Process., vol. 57, no. 8, pp. 3075–3085, Aug. 2009.
[2] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, “Block-sparse signals: Uncertainty relations and efficient recovery,” IEEE Trans. Signal Process., vol. 58, no. 6, pp. 3042–3054, Jun. 2010.
[3] X. Lv, G. Bi, and C. Wan, “The group lasso for stable recovery of block-sparse signal representations,” IEEE Trans. Signal Process., vol. 59, no. 4, pp. 1371–1382, Apr. 2011.
Hiroki Kuroda received the B.E. degree in computer science, and the M.E. and the Ph.D. degrees in information and communications engineering from the Tokyo Institute of Technology, Tokyo, Japan, in 2013, 2015 and 2019, respectively. From April 2019 to March 2020, he was a postdoctoral researcher with the National Institute of Advanced Industrial Science and Technology, Tsukuba, Japan. Since April 2020, he has been an Assistant Professor with the College of Information Science and Engineering, Ritsumeikan University, Shiga, Japan. His current research interests are in signal processing and its applications, sparse modeling, and optimization theory. He received the Seiichi Tejima Doctoral Dissertation Award from the Tokyo Institute of Technology in 2020.

Daichi Kitahara received the B.E. degree in computer science in 2012, and the M.E. and the Ph.D. degrees in communications and computer engineering in 2014 and 2017 from the Tokyo Institute of Technology, Tokyo, Japan, respectively. From April 2014 to March 2017, he was a recipient of the Research Fellowship of the Japan Society for the Promotion of Science. Since April 2017, he has been an Assistant Professor with the College of Information Science and Engineering, Ritsumeikan University, Shiga, Japan. His current research interests include signal processing and its applications, inverse problem, optimization theory, and multivariate spline theory.