A CONVEX PENALTY FOR BLOCK-SPARSE SIGNALS WITH UNKNOWN STRUCTURES

Hiroki Kuroda, Daichi Kitahara, and Akira Hirabayashi

College of Information Science and Engineering, Ritsumeikan University
E-mail: {kuroda, d-kita, akirahrb}@media.ritsumei.ac.jp

ABSTRACT

We propose a novel convex penalty for block-sparse signals whose block partitions are unknown a priori. We first introduce a nonconvex penalty function, where the block partition is adjusted for the signal of interest by minimizing the mixed ℓ_2/ℓ_1 norm over all possible block partitions. Then, by exploiting a variational representation of the ℓ_2 norm, we derive the proposed penalty function as a suitable convex relaxation of the nonconvex penalty. For the resulting regularization model, we provide a proximal splitting-based algorithm which is guaranteed to converge to an optimal solution. Numerical experiments show the effectiveness of the proposed penalty.

***Index Terms*—** Block-sparsity, unknown partition, convex regularization, proximal splitting algorithm.

1. INTRODUCTION

Many natural signals exhibit block-sparsity, i.e., 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., [1–5], demonstrate the effectiveness of the mixed ℓ_2/ℓ_1 norm using the block partition. However, the information of the block partition is not available in many applications, e.g., recovery of acoustic [6, 7], image [8, 9], and radar signals [10–12]. For instance, the target signal of the phased array weather radar [11–14] is block-sparse in the Fourier domain due to the narrow bandwidth of power spectral density, but the block partition is unknown because it depends on the unknown mean and standard deviation of the Doppler frequency. In such situations, the recovery performance by the mixed ℓ_2/ℓ_1 norm often degrades due to a pre-fixed block partition, which is substituted for an ideal one.

Several attempts have been made to cope with the unknown block partitions. A commonly used modification is to use potentially overlapping blocks, e.g., [15, 16]. Among them, the latent group lasso (LGL) penalty [15] presents a clever approach where suitable blocks are selected from pre-defined potentially 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. However, since the problem size grows with the number of candidate blocks, the LGL penalty has a restriction on the size of candidate set, which restricts the recovery performance. Meanwhile, the greedy approach [17] has a similar restriction on the candidate set of blocks. Bayesian approaches [18, 19] have also been presented, but they have to solve challenging nonconvex optimization problems. Thus, it is important to realize a convex method which can flexibly cope with various block partitions.

In this paper, we propose a novel convex penalty function, named *latent optimally partitioned ℓ_2/ℓ_1 (LOP-ℓ_2/ℓ_1) penalty*, where the block partition is automatically adapted for the signal of interest. First, we introduce a nonconvex penalty function as the minimum of the mixed ℓ_2/ℓ_1 norm over all possible block partitions. Then, as a suitable convex relaxation of the nonconvex penalty function, we derive the LOP-ℓ_2/ℓ_1 penalty. More precisely, by utilizing a variational representation of the ℓ_2 norm, we represent the nonconvex penalty function as the minimization of a convex function over the ℓ_0 pseudo-norm constraint on latent variables. The LOP-ℓ_2/ℓ_1 penalty is derived by replacing the ℓ_0 pseudo-norm constraint with its convex envelope, i.e., the ℓ_1 norm. Moreover, with the aid of the computation of a proximity operator shown in [20, 21], we provide a proximal splitting algorithm which converges to an optimal solution of the regularization model using the LOP-ℓ_2/ℓ_1 penalty.

To show the effectiveness of the LOP-ℓ_2/ℓ_1 penalty, we conduct numerical experiments on synthetic examples and real-world data. The results show that the LOP-ℓ_2/ℓ_1 penalty achieves better estimation accuracy than existing penalties including the LGL penalty.

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 complex conjugate transpose respectively by $(\cdot)^\top$ and $(\cdot)^{\mathrm{H}}$. For $\boldsymbol{x} = (x_1, \ldots, x_N)^\top \in \mathbb{C}^N$ and an index set $\mathcal{I} \subset \{1, \ldots, N\}$, $\boldsymbol{x}_\mathcal{I} := (x_n)_{n\in\mathcal{I}}$ denotes the subvector of $\boldsymbol{x}$ indexed by $\mathcal{I}$. We define the support of $\boldsymbol{x} \in \mathbb{C}^N$ by $\mathrm{supp}(\boldsymbol{x}) := \{n \in \{1, \ldots, N\} \,|\, x_n \neq 0\}$. The cardinality of a set $\mathcal{S}$ is denoted by $|\mathcal{S}|$. The ℓ_2 norm, the ℓ_1 norm, and the ℓ_0 pseudo-norm of $\boldsymbol{x} \in \mathbb{C}^N$ are respectively denoted by $\|\boldsymbol{x}\|_2 := \sqrt{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}}$, $\|\boldsymbol{x}\|_1 := \sum_{n=1}^N |x_n|$, and $\|\boldsymbol{x}\|_0 := |\mathrm{supp}(\boldsymbol{x})|$.

2. PRELIMINARIES

2.1. Block-Sparsity with Unknown Block Partition

We suppose that $\boldsymbol{x}^\star \in \mathbb{C}^N$ to be estimated is block-sparse over an unknown block partition $\mathcal{B}_1^\star, \ldots, \mathcal{B}_{K^\star}^\star$. Namely, the subvector $\boldsymbol{x}_{\mathcal{B}_k^\star}^\star$ contains only (approximately) zero components for many $k \in \{1, \ldots, K^\star\}$. We use the term *block-sparse* in a strict sense, i.e., $\mathcal{B}_k^\star$ consists of consecutive indices as $\mathcal{B}_k^\star = \{n_k^\star, n_k^\star + 1, \ldots, m_k^\star\}$.

2.2. Existing Penalties for Block-Sparse Signals

To enhance the block-sparsity of $\boldsymbol{x} \in \mathbb{C}^N$ over the known non-overlapping blocks $\mathcal{B}_1, \ldots, \mathcal{B}_K \subset \{1, \ldots, N\}$, a commonly used penalty function is the mixed ℓ_2/ℓ_1 norm defined as the sum (i.e. the ℓ_1 norm) of the blockwise ℓ_2 norms:

$$\|\boldsymbol{x}\|_{2,1} := \sum_{k=1}^{K} \sqrt{|\mathcal{B}_k|}\, \|\boldsymbol{x}_{\mathcal{B}_k}\|_2, \tag{1}$$

where we use the weight $\sqrt{|\mathcal{B}_k|}$ by following the suggestions in, e.g., [1, 2]. The mixed ℓ_2/ℓ_1 norm promotes the block sparsity by pushing components in $\mathcal{B}_k$ toward zeros together. However, the performance of the mixed ℓ_2/ℓ_1 norm degrades when $\mathcal{B}_1, \ldots, \mathcal{B}_K$ do not match with the ground-truth $\mathcal{B}_1^\star, \ldots, \mathcal{B}_{K^\star}^\star$. To cope with such difficulty, several extensions, e.g., [15, 16] have been designed by using potentially overlapping blocks $\bar{\mathcal{B}}_1, \ldots, \bar{\mathcal{B}}_{\bar{K}} \subset \{1, \ldots, N\}$. Among them, the latent group lasso (LGL) penalty [15] is a more adequate extension for major applications, and defined as[1]

$$\left.\begin{aligned} &\mathrm{LGL}(\boldsymbol{x}) := \min_{(\boldsymbol{v}_1,\ldots,\boldsymbol{v}_{\bar{K}})\in\mathbb{C}^{N\times\bar{K}}} \sum_{k=1}^{\bar{K}} \sqrt{|\bar{\mathcal{B}}_k|}\,\|\boldsymbol{v}_k\|_2 \\ &\quad \text{s.t. } \sum_{k=1}^{\bar{K}} \boldsymbol{v}_k = \boldsymbol{x} \text{ and } \mathrm{supp}(\boldsymbol{v}_k) \subset \bar{\mathcal{B}}_k \ \ (k=1,\ldots,\bar{K}). \end{aligned}\right\} \quad (2)$$

Although the LGL penalty can be used via convex optimization, it is intractable to use all possible blocks as $\bar{\mathcal{B}}_1, \ldots, \bar{\mathcal{B}}_{\bar{K}}$ since $\bar{K}$ is too large in this case. Due to this computational issue, $\bar{\mathcal{B}}_1, \ldots, \bar{\mathcal{B}}_{\bar{K}}$ are typically restricted to blocks of a fixed size.

3. PROPOSED METHOD

3.1. Derivation of Proposed Penalty Function

We first introduce a nonconvex penalty function $\psi_{\mathrm{nc}}(\boldsymbol{x})$, where the block partition is optimized for the signal of interest. Then, as a suitable convex relaxation of $\psi_{\mathrm{nc}}(\boldsymbol{x})$, we derive the proposed latent optimally partitioned ℓ_2/ℓ_1 (LOP-ℓ_2/ℓ_1) penalty $\psi(\boldsymbol{x})$.

We define $\psi_{\mathrm{nc}}(\boldsymbol{x})$ by taking the minimum of the mixed ℓ_2/ℓ_1 norm over at-most K block partitions:

$$\psi_{\mathrm{nc}}(\boldsymbol{x}) := \min_{j\in\{1,\ldots,K\}} \min_{(\mathcal{B}_1,\ldots,\mathcal{B}_j)\in\mathcal{P}_j} \sum_{k=1}^{j} \sqrt{|\mathcal{B}_k|}\,\|\boldsymbol{x}_{\mathcal{B}_k}\|_2, \quad (3)$$

where $\mathcal{P}_j$ contains all j block partitions of $\{1, \ldots, N\}$, i.e.,

$$(\mathcal{B}_1, \ldots, \mathcal{B}_j) \in \mathcal{P}_j \Leftrightarrow \begin{cases} \bigcup_{k=1}^{j} \mathcal{B}_k = \{1, \ldots, N\}, \\ \mathcal{B}_k \cap \mathcal{B}_{k'} = \varnothing \quad (k \neq k'), \\ \exists n_k, m_k \in \{1, \ldots, N\} \text{ s.t. } \mathcal{B}_k = \{n_k, n_k+1, \ldots, m_k\}. \end{cases}$$

Note that K can be set to an upper bound of the ground-truth $K^\star$. Note also that, since the sizes of $\mathcal{B}_1, \ldots, \mathcal{B}_j$ can also be adjusted, the non-overlappping condition $\mathcal{B}_k \cap \mathcal{B}_{k'} = \varnothing$ is not restrictive. To derive the convex relaxation of $\psi_{\mathrm{nc}}(\boldsymbol{x})$, we exploit the following lemma which shows a variational representation of the ℓ_2 norm.

Lemma 1. *Let $\phi\colon \mathbb{C}\times\mathbb{R}_+ \to \mathbb{R}_+\cup\{\infty\}$ be a lower semicontinuous convex function defined as*

$$\phi(x, \rho) := \begin{cases} \dfrac{|x|^2}{2\rho} + \dfrac{\rho}{2}, & \textit{if } \rho > 0; \\ 0, & \textit{if } x = 0 \textit{ and } \rho = 0; \\ \infty, & \textit{otherwise}. \end{cases}$$

Then, the ℓ_2 norm can be represented in the variational form using ϕ as

$$\sqrt{|\mathcal{B}_k|}\,\|\boldsymbol{x}_{\mathcal{B}_k}\|_2 = \min_{\rho\in\mathbb{R}_+} \sum_{n\in\mathcal{B}_k} \phi(x_n, \rho). \quad (4)$$

[1] By setting $\bar{\mathcal{B}}_k$ to other than blocks, the LGL penalty is applicable to other structural sparsity, while this paper focuses on the block-sparsity.

This lemma implies a variational representation of $\psi_{\mathrm{nc}}(\boldsymbol{x})$ in (3) as

$$\psi_{\mathrm{nc}}(\boldsymbol{x}) = \min_{j\in\{1,\ldots,K\}} \min_{(\mathcal{B}_1,\ldots,\mathcal{B}_j)\in\mathcal{P}_j,\, \boldsymbol{\rho}\in\mathbb{R}_+^j} \sum_{k=1}^{j} \sum_{n\in\mathcal{B}_k} \phi(x_n, \rho_k).$$

By letting a latent vector $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_N)^\top \in \mathbb{R}_+^N$ as

$$\sigma_n = \rho_k \ \ (n \in \mathcal{B}_k) \quad \text{for } k = 1, \ldots, j,$$

we see that $\boldsymbol{\sigma}$ is characterized by the condition that

$$\|\boldsymbol{D}\boldsymbol{\sigma}\|_0 \leq j - 1,$$

where $\boldsymbol{D}$ is the first discrete difference operator defined by

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

Thus, we can represent $\psi_{\mathrm{nc}}(\boldsymbol{x})$ as

$$\begin{aligned} \psi_{\mathrm{nc}}(\boldsymbol{x}) &= \min_{j\in\{1,\ldots,K\}} \min_{\boldsymbol{\sigma}\in\mathbb{R}_+^N} \sum_{n=1}^{N} \phi(x_n, \sigma_n) \text{ s.t. } \|\boldsymbol{D}\boldsymbol{\sigma}\|_0 \leq j - 1 \\ &= \min_{\boldsymbol{\sigma}\in\mathbb{R}_+^N} \sum_{n=1}^{N} \phi(x_n, \sigma_n) \text{ s.t. } \|\boldsymbol{D}\boldsymbol{\sigma}\|_0 \leq K - 1. \end{aligned}$$

Based on the fact that the ℓ_1 norm is the convex envelope of the ℓ_0 pseudo-norm in the constraint, we finally derive the LOP-ℓ_2/ℓ_1 penalty as

$$\psi(\boldsymbol{x}) = \min_{\boldsymbol{\sigma}\in\mathbb{R}_+^N} \sum_{n=1}^{N} \phi(x_n, \sigma_n) \text{ s.t. } \|\boldsymbol{D}\boldsymbol{\sigma}\|_1 \leq \alpha, \quad (5)$$

where $\alpha \geq 0$ is a tuning parameter related to the number of blocks.

Remark 1 (Generalization for matrices)**.** For a matrix $\boldsymbol{X} \in \mathbb{C}^{N\times M}$, the LOP-$\ell_2/\ell_1$ penalty can be naturally extended by replacing the 1D discrete difference operator with a 2D one:

$$\left.\begin{aligned} \psi_{2\mathrm{d}}(\boldsymbol{X}) = \min_{\boldsymbol{\Sigma}\in\mathbb{R}_+^{N\times M}} \sum_{n=1}^{N} \sum_{m=1}^{M} \phi(X_{n,m}, \Sigma_{n,m}) \\ \text{s.t. } \|D_{2\mathrm{d}}(\boldsymbol{\Sigma})\|_1 \leq \alpha, \end{aligned}\right\} \quad (6)$$

where $D_{2\mathrm{d}}\colon \mathbb{R}^{N\times M} \to \mathbb{R}^{N(M-1)+(N-1)M}$ computes discrete differences in row and column directions. Note that we do not focus on rectangular blocks, and $\psi_{2\mathrm{d}}$ can manage blocks of various shapes thanks to the flexibility of the 2D discrete difference operator $D_{2\mathrm{d}}$.

3.2. Regularization with Proposed Penalty

We show an application of the LOP-ℓ_2/ℓ_1 penalty (5) for the regularization model, and provide its efficient solver. For simplicity, we consider the following regularized least-squares model:

$$\operatorname*{minimize}_{\boldsymbol{x}\in\mathbb{C}^N} \tfrac{1}{2}\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \lambda\psi(\boldsymbol{x}),$$

which can be tackled as

$$\operatorname*{minimize}_{(\boldsymbol{x},\boldsymbol{\sigma})\in\mathbb{C}^N\times\mathbb{R}_+^N} \tfrac{1}{2}\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \lambda \sum_{n=1}^{N} \phi(x_n, \sigma_n) \text{ s.t. } \|\boldsymbol{D}\boldsymbol{\sigma}\|_1 \leq \alpha, \quad (7)$$

Algorithm 1: Solver for the proposed model (7)

Input: $\gamma > 0, \mu \in (0, 1/\sqrt{2}], \tau \in (0, 1/\sqrt{5}]$, and $\boldsymbol{x}^{(0)}, \boldsymbol{\sigma}^{(0)}, \boldsymbol{\xi}^{(0)}, \boldsymbol{u}^{(0)}, \boldsymbol{\eta}^{(0)}, \boldsymbol{\zeta}^{(0)}$.

while *a stopping criterion is not satisfied* **do**

- $\tilde{\boldsymbol{x}}^{(j+1)} = \boldsymbol{x}^{(j)} - \mu(\mu(\boldsymbol{x}^{(j)} - \boldsymbol{\xi}^{(j)}) - \boldsymbol{\eta}^{(j)})$;
- $\tilde{\boldsymbol{\sigma}}^{(j+1)} = \boldsymbol{\sigma}^{(j)} - \tau \boldsymbol{D}^\top(\tau(\boldsymbol{D}\boldsymbol{\sigma}^{(j)} - \boldsymbol{u}^{(j)}) - \boldsymbol{\zeta}^{(j)})$;
- $\tilde{\boldsymbol{\xi}}^{(j+1)} = \boldsymbol{\xi}^{(j)} + \mu(\mu(\boldsymbol{x}^{(j)} - \boldsymbol{\xi}^{(j)}) - \boldsymbol{\eta}^{(j)})$;
- $\tilde{\boldsymbol{u}}^{(j+1)} = \boldsymbol{u}^{(j)} + \tau(\tau(\boldsymbol{D}\boldsymbol{\sigma}^{(j)} - \boldsymbol{u}^{(j)}) - \boldsymbol{\zeta}^{(j)})$;
- **for** $n = 1, \ldots, N$ **do**
  - $(x_n^{(j+1)}, \sigma_n^{(j+1)}) = \mathrm{prox}_{\gamma\lambda\phi}(\tilde{x}_n^{(j+1)}, \tilde{\sigma}_n^{(j+1)})$ by (8);
- $\boldsymbol{\xi}^{(j+1)} = \left(\gamma \boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} + \boldsymbol{I}\right)^{-1}\left(\gamma \boldsymbol{A}^{\mathrm{H}}\boldsymbol{y} + \tilde{\boldsymbol{\xi}}^{(j+1)}\right)$;
- $\boldsymbol{u}^{(j+1)} = P_{B_1^\alpha}(\tilde{\boldsymbol{u}}^{(j+1)})$ by (9);
- $\boldsymbol{\eta}^{(j+1)} = \boldsymbol{\eta}^{(j)} - \mu(\boldsymbol{x}^{(j+1)} - \boldsymbol{\xi}^{(j+1)})$;
- $\boldsymbol{\zeta}^{(j+1)} = \boldsymbol{\zeta}^{(j)} - \tau(\boldsymbol{D}\boldsymbol{\sigma}^{(j+1)} - \boldsymbol{u}^{(j+1)})$;
- $j \leftarrow j + 1$

where $\boldsymbol{y} \in \mathbb{C}^d$ is the known observation vector, $\boldsymbol{A} \in \mathbb{C}^{d \times N}$ is the known measurement matrix, and $\lambda > 0$ is the regularization parameter.

The proposed model (7) can be decomposed into lower semi-continuous convex functions whose proximity operators are easily computed. Thus, by applying proximal splitting algorithms [22–27], we can develop an efficient iterative solver for (7). As an instance, by applying the primal-dual type method [24–27], we obtain Algorithm 1, where $(\boldsymbol{x}^{(j)}, \boldsymbol{\sigma}^{(j)})_{j=1}^{\infty}$ converges to an optimal solution of (7). Thanks to [20, 21], we can compute the proximity operator of ϕ of index $\kappa > 0$ used in Algorithm 1 as

$$\mathrm{prox}_{\kappa\phi}(x, \sigma) = \begin{cases} (0, 0), & \text{if } 2\kappa\sigma + |x|^2 \leq \kappa^2; \\ (0, \sigma - \frac{\kappa}{2}), & \text{else if } x = 0; \\ \left(x - \kappa t \frac{x}{|x|}, \sigma + \kappa \frac{t^2 - 1}{2}\right), & \text{otherwise,} \end{cases} \tag{8}$$

where $t > 0$ is the unique positive root of

$$t^3 + \left(\tfrac{2}{\kappa}\sigma + 1\right)t - \tfrac{2}{\kappa}|x| = 0,$$

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

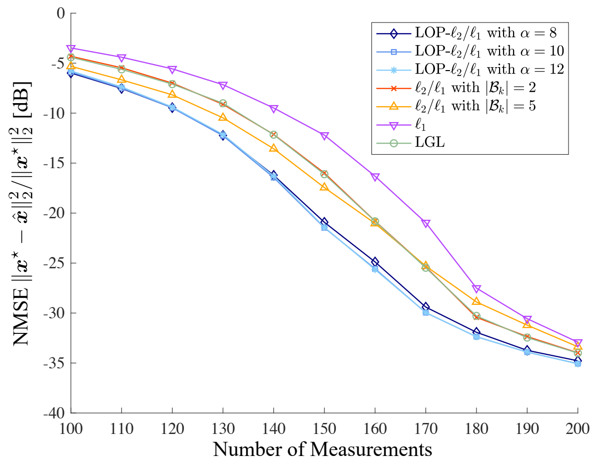
$$t = \begin{cases} \sqrt[3]{-\frac{q}{2} + \sqrt{-D}} + \sqrt[3]{-\frac{q}{2} - \sqrt{-D}}, & \text{if } D < 0; \\ 2\sqrt[3]{-\frac{q}{2}}, & \text{if } D = 0; \\ 2\sqrt[3]{\sqrt{\frac{q^2}{4} + D}} \cos\left(\frac{\arctan(-2\sqrt{D}/q)}{3}\right), & \text{if } D > 0, \end{cases}$$

where $\sqrt[3]{\cdot}$ denotes the real cubic root. In Algorithm 1, we also use the ℓ_1 ball projection, which can be computed as

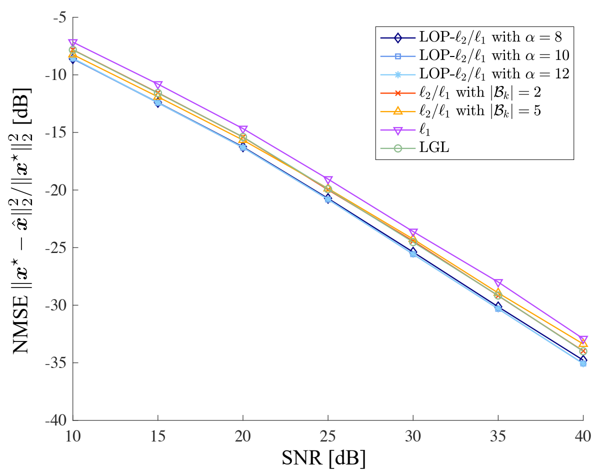
$$P_{B_1^\alpha}(\boldsymbol{u}) = \begin{cases} \boldsymbol{u}, & \text{if } \|\boldsymbol{u}\|_1 \leq \alpha; \\ (\mathrm{sign}(u_n)a_n)_{n=1}^{N-1}, & \text{otherwise,} \end{cases} \tag{9}$$

$$a_n := \max\{|u_n| - (\textstyle\sum_{i=1}^{I} r_i - \alpha)/I, 0\},$$
$$I := \max\{j \in \{1, \ldots, N-1\} \mid (\textstyle\sum_{i=1}^{j} r_i - \alpha)/j < r_j\},$$

where $r_1, \ldots, r_{N-1}$ are obtained by sorting $|u_1|, \ldots, |u_{N-1}|$ in descending order. Note that sophisticated ℓ_1 ball projection algorithms with $\mathcal{O}(N)$ expected complexity, e.g., [28], are also available. Thus, in Algorithm 1, computations regarding the LOP-ℓ_2/ℓ_1 penalty can be implemented with $\mathcal{O}(N)$ arithmetic operations.



(a) NMSE against the number of measurements where SNR is fixed to 40dB.



(b) NMSE against SNR where the number of measurements is fixed to 200.

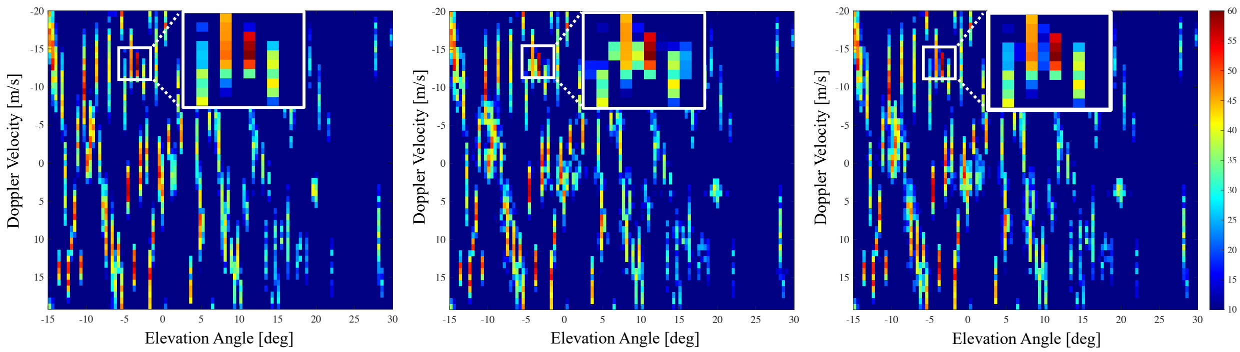
Fig. 1: Comparison of the penalties for synthetic examples, where the results are averaged over 100 independent trials.

4. NUMERICAL EXPERIMENTS

To demonstrate the effectiveness of the LOP-ℓ_2/ℓ_1 penalty (5), we conduct numerical experiments on synthetic and real-world data. We compare the LOP-ℓ_2/ℓ_1 penalty with the existing convex penalties: the ℓ_1 norm, the mixed ℓ_2/ℓ_1 norm (1), and the LGL penalty (2).

Synthetic Examples: We consider the estimation of a block-sparse signal $\boldsymbol{x}^\star \in \mathbb{R}^N$ from noisy compressive measurements. More precisely, we apply (7) for $\boldsymbol{y} := \boldsymbol{A}\boldsymbol{x}^\star + \boldsymbol{\varepsilon}$, where the entries of $\boldsymbol{A} \in \mathbb{R}^{d \times N}$ are drawn from i.i.d. Gaussian distribution $\mathcal{N}(0, 1)$, and $\boldsymbol{\varepsilon} \in \mathbb{R}^d$ is the white Gaussian noise. The block-sparse signal $\boldsymbol{x}^\star$ is randomly generated by the following scheme. We set $N = 250$, and $\boldsymbol{x}^\star$ has 80 nonzero components which are randomly divided into 4 blocks. The blocks are randomly located under the condition $|\mathrm{supp}(\boldsymbol{x}^\star)| = 80$. Amplitudes of nonzero components are drawn from i.i.d. $\mathcal{N}(0, 1)$.

We compare the penalties in terms of the normalized mean square error (NMSE) $\|\boldsymbol{x}^\star - \hat{\boldsymbol{x}}\|_2^2/\|\boldsymbol{x}^\star\|_2^2$, where $\hat{\boldsymbol{x}}$ is the solution of (7) for the LOP-ℓ_2/ℓ_1 penalty. The existing penalties are also combined with the square errors, where the regularization parameters are tuned independently to yield the best results. All regularization models are solved by proximal splitting-based iterative algorithms,



(a) Ground-truth (b) Estimate by LGL penalty (c) Estimate by LOP- ℓ_2/ℓ_1 penalty

Fig. 2: Power spectrums of ground-truth and estimates of the scatter signal shown in dB for a trial of phased array weather radar simulation.

where the iteration is terminated when the norm of the difference between successive iterates is below the threshold 10^{-4}.

In Figs. 1(a) and (b), we show the NMSE respectively against the number of measurements d and the SNR $\|\boldsymbol{A}\boldsymbol{x}^\star\|_2^2/\|\boldsymbol{\varepsilon}\|_2^2$, where the results are averaged over 100 independent trials. Note that the mixed ℓ_2/ℓ_1 norm are tested with block sizes 2 and 5, and the LGL penalty uses overlapping blocks of size 2 in (2). In Figs. 1(a) and (b), the LOP-ℓ_2/ℓ_1 penalty with several choices of α outperforms the existing penalties. In addition, the performance of the LOP-ℓ_2/ℓ_1 penalty is fairly robust against the tuning parameter α.

Application to Phased Array Weather Radar: A major goal of the phased array weather radar (PAWR) [13, 14] is to estimate the scatter signal $\boldsymbol{X}^\star \in \mathbb{C}^{N\times M}$ from degraded observations $\boldsymbol{Y} := \boldsymbol{S}\boldsymbol{X}^\star + \boldsymbol{\mathcal{E}} \in \mathbb{C}^{d\times M}$, where $\boldsymbol{S} \in \mathbb{C}^{d\times N}$ is a certain array manifold matrix, and $\boldsymbol{\mathcal{E}} \in \mathbb{C}^{d\times M}$ is the observation noise. Note that N is the number of elevation angles, M is the number of pulses, and d is the number of array elements. We conduct a numerical simulation in a setting similar to [12]. The scatter signal $\boldsymbol{X}^\star \in \mathbb{C}^{N\times M}$ is generated based on the reflection intensity measured by the PAWR at Osaka University, where the elevation angles are chosen uniformly from -15° to 30° degrees with $N = 110$. We set $M = 50$ and $d = 128$, and $\boldsymbol{\mathcal{E}}$ as the white Gaussian noise with the standard deviation $\sqrt{5}$. As shown in [12], the scatter signal $\boldsymbol{X}^\star$ exhibits block-sparsity in the Fourier domain for each elevation angle, but block partitions are unknown a priori. Thus, we apply the LOP-ℓ_2/ℓ_1 penalty as

$$\hat{\boldsymbol{X}} \in \arg\min_{\boldsymbol{X}\in\mathbb{C}^{N\times M}} \|\boldsymbol{Y} - \boldsymbol{S}\boldsymbol{X}\|_{\mathrm{fro}}^2 + \lambda\tilde{\psi}_{2\mathrm{d}}(\boldsymbol{F}\boldsymbol{X}^\top),$$

where $\|\cdot\|_{\mathrm{fro}}$ denotes the Frobenius norm, $\boldsymbol{F} \in \mathbb{C}^{M\times M}$ is the discrete Fourier transform matrix, and $\tilde{\psi}_{2\mathrm{d}}$ is slightly modified from (6) by omitting horizontal differences to exploit the column-wise block-sparsity of $\boldsymbol{V} = \boldsymbol{F}\boldsymbol{X}^\top$. Existing penalties are also combined with the square errors, and the regularization parameters are tuned independently so that the performances become best.

In Table 1, we show the NMSE $\|\boldsymbol{X}^\star - \hat{\boldsymbol{X}}\|_{\mathrm{fro}}^2/\|\boldsymbol{X}^\star\|_{\mathrm{fro}}^2$ averaged over 50 independent trails, where the LOP-ℓ_2/ℓ_1 penalty uses $\alpha = 130N$, the LGL penalty uses overlapping blocks of size 2, and ℓ_2/ℓ_1 (a) and (b) respectively use block sizes 2 and 4. The result shows that the LOP-ℓ_2/ℓ_1 penalty achieves the best performance. Moreover, as shown in Fig. 2 where we show the power spectrum of the estimate, i.e. squared magnitudes of $\boldsymbol{F}\hat{\boldsymbol{X}}^\top$, the LOP-$\ell_2/\ell_1$ penalty successfully reduces the artifacts caused in the second-best LGL penalty.

Application to Speech Denoising: To show the effectiveness of the proposed 2D extension (6), we conduct experiments on speech denosing using block-sparsity of the spectrogram. We generate noisy speech as $\boldsymbol{y} = \boldsymbol{s}^\star + \boldsymbol{\varepsilon} \in \mathbb{R}^d$, where $\boldsymbol{s}^\star$ is a 2-second clip of speech taken from [29] with 16kHz sampling rate (i.e. $d = 32000$), and $\boldsymbol{\varepsilon}$ is the white Gaussian noise with 10dB SNR. We use the 2D LOP-ℓ_2/ℓ_1 penalty to enforce the block-sparsity of the spectrogram as

$$\hat{\boldsymbol{X}} \in \arg\min_{\boldsymbol{X}\in\mathbb{C}^{N\times M}} \|\boldsymbol{y} - \mathcal{S}^{-1}(\boldsymbol{X})\|_2^2 + \lambda\psi_{2\mathrm{d}}(\boldsymbol{X}),$$

where $\mathcal{S}^{-1}$ denotes the inverse short-time Fourier transform with the hann window of 32ms with 75% overlap, and $N = 512$ and $M = 253$. The existing penalties are also used in similar ways, where the regularization parameters are tuned independently.

Table 1: Comparison of the penalties for simulation of phased array weather radar in terms of the NMSE in dB averaged over 50 independent trials.

LOP-ℓ_2/ℓ_1	ℓ_2/ℓ_1 (a)	ℓ_2/ℓ_1 (b)	ℓ_1	LGL
$\mathbf{-15.72}$	-14.78	-14.03	-13.39	-14.83

Table 2: Comparison of the penalties for speech denoising in terms of the NMSE in dB averaged over 20 independent trials.

	LOP-ℓ_2/ℓ_1	ℓ_2/ℓ_1 (a)	ℓ_2/ℓ_1 (b)	ℓ_1
Male	$\mathbf{-15.73}$	-15.56	-15.41	-15.28
Female	$\mathbf{-17.43}$	-17.04	-16.90	-16.85

In Table 2, we show the NMSE $\|\boldsymbol{s}^\star - \mathcal{S}^{-1}(\hat{\boldsymbol{X}})\|_2^2/\|\boldsymbol{s}^\star\|_2^2$ averaged over 20 independent trials respectively for male and female speech, where the LOP-ℓ_2/ℓ_1 penalty uses $\alpha = 0.007NM$, and ℓ_2/ℓ_1 (a) and (b) respectively use block sizes 2×4 and 2×2, which perform best among block sizes $\{1,2,4,8\}\times\{1,2,4,8\}$. It can be seen that the LOP-ℓ_2/ℓ_1 penalty improves the NMSE over the existing penalties.

5. CONCLUSION

We proposed the latent optimally partitioned ℓ_2/ℓ_1 (LOP-ℓ_2/ℓ_1) penalty (5) for block-sparse signals whose block partitions are not available a priori. We first introduce a nonconvex penalty function (3), where the block partition is optimized for the signal of interest based on the minimization of the mixed ℓ_2/ℓ_1 norm. Then, the LOP-ℓ_2/ℓ_1 penalty is derived as its suitable convex relaxation. We also provided a proximal splitting-based solver for the corresponding regularization model (7). Numerical experiments illustrate the effectiveness of the LOP-ℓ_2/ℓ_1 penalty.

Acknowledgment: We thank Prof. Tomoo Ushio, Dr. Hiroshi Kikuchi, and Dr. Eiichi Yoshikawa for providing the PAWR data. This work was supported in part by JSPS Grants-in-Aid 19K20361.

6. REFERENCES

[1] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. Roy. Statist. Soc. Ser. B*, vol. 68, no. 1, pp. 49–67, 2006.

[2] J. Huang, P. Breheny, and S. Ma, "A selective review of group selection in high-dimensional models," *Statist. Sci.*, vol. 27, no. 4, pp. 481–499, Nov. 2012.

[3] 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, June 2010.

[4] E. Elhamifar and R. Vidal, "Block-sparse recovery via convex optimization," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 4094–4107, Aug. 2012.

[5] 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, April 2011.

[6] G. Yu, S. Mallat, and E. Bacry, "Audio denoising by time-frequency block thresholding," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 1830–1839, May 2008.

[7] R. Gribonval and E. Bacry, "Harmonic decomposition of audio signals with matching pursuit," *IEEE Trans. Signal Process.*, vol. 51, no. 1, pp. 101–111, Jan. 2003.

[8] V. Cevher, A. Sankaranarayanan, M. F. Duarte, D. Reddy, R. G. Baraniuk, and R. Chellappa, "Compressive sensing for background subtraction," in *Proc. ECCV*, 2008, pp. 155–168.

[9] Z. Gao, L. Cheong, and Y. Wang, "Block-sparse RPCA for salient motion detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 10, pp. 1975–1987, Oct. 2014.

[10] L. Wang, L. Zhao, G. Bi, C. Wan, and L. Yang, "Enhanced ISAR imaging by exploiting the continuity of the target scene," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5736–5750, Sep. 2014.

[11] E. Yoshikawa, T. Ushio, Z. Kawasaki, S. Yoshida, T. Morimoto, F. Mizutani, and M. Wada, "MMSE beam forming on fast-scanning phased array weather radar," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 3077–3088, May 2013.

[12] D. Kitahara, M. Nakahara, A. Hirabayashi, E. Yoshikawa, H. Kikuchi, and T. Ushio, "Nonlinear beamforming via convex optimization for phased array weather radar," in *Proc. APSIPA ASC*, Nov. 2018, pp. 1831–1835.

[13] F. Mizutani, T. Ushio, E. Yoshikawa, S. Shimamura, H. Kikuchi, M. Wada, S. Satoh, and T. Iguchi, "Fast-scanning phased-array weather radar with angular imaging technique," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2664–2673, May 2018.

[14] B. Isom, R. Palmer, R. Kelley, J. Meier, D. Bodine, M. Yeary, B.-L. Cheong, Y. Zhang, T.-Y. Yu, and M. I. Biggerstaff, "The atmospheric imaging radar: Simultaneous volumetric observations using a phased array weather radar," *J. Atmos. Ocean. Technol.*, vol. 30, no. 4, pp. 655–675, April 2013.

[15] L. Jacob, G. Obozinski, and J.-P. Vert, "Group lasso with overlap and graph lasso," in *Proc. Annu. Int. Conf. Mach. Learn.*, 2009, pp. 433–440.

[16] R. Jenatton, J.-Y. Audibert, and F. Bach, "Structured variable selection with sparsity-inducing norms," *J. Mach. Learn. Res.*, vol. 12, pp. 2777–2824, 2011.

[17] J. Huang, T. Zhang, and D. Metaxas, "Learning with structured sparsity," *J. Mach. Learn. Res.*, vol. 12, no. 11, 2011.

[18] J. Fang, Y. Shen, H. Li, and P. Wang, "Pattern-coupled sparse bayesian learning for recovery of block-sparse signals," *IEEE Trans. Signal Process.*, vol. 63, no. 2, pp. 360–372, Jan. 2015.

[19] L. Wang, L. Zhao, S. Rahardja, and G. Bi, "Alternative to extended block sparse Bayesian learning and its relation to pattern-coupled sparse Bayesian learning," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2759–2771, May 2018.

[20] P. L. Combettes and C. L. Müller, "Perspective functions: Proximal calculus and applications in high-dimensional statistics," *J. Math. Anal. Appl.*, vol. 457, no. 2, pp. 1283–1306, 2018.

[21] P. L. Combettes and C. L. Müller, "Perspective maximum likelihood-type estimation via proximal decomposition," *Electron. J. Statist.*, vol. 14, no. 1, pp. 207–238, 2020.

[22] 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.

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

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

[25] I. Loris and C. Verhoeven, "On a generalization of the iterative soft-thresholding algorithm for the case of non-separable penalty," *Inverse Problems*, vol. 27, no. 12, 15 pages, Nov. 2011.

[26] M. Yamagishi and I. Yamada, "Nonexpansiveness of a linearized augmented Lagrangian operator for hierarchical convex optimization," *Inverse Problems*, vol. 33, no. 4, 35 pages, March 2017.

[27] L. Condat, D. Kitahara, A. Contreras, and A. Hirabayashi, "Proximal splitting algorithms: A tour of recent advances, with new twists," preprint arXiv:1912.00137, 2020.

[28] L. Condat, "Fast projection onto the simplex and the ℓ_1 ball," *Math. Programm.*, vol. 158, no. 1, pp. 575–585, 2016.

[29] J. Kominek and A. W. Black, "The CMU Arctic speech databases," in *ISCA Workshop on Speech Synthesis*, 2004.