第35回信号処理シンポジウム
2020年11月4日～6日（リモート開催）

Recovery of Block-Sparse Signals with Optimization of Block Partitions

Hiroki Kuroda† Daichi Kitahara† Akira Hirabayashi†
†College of Information Science and Engineering, Ritsumeikan University
Email: {kuroda, d-kita, akirahrb}@media.ritsumei.ac.jp

Abstract We propose a novel convex recovery model for block-sparse signals whose block partitions are unknown in advance. More precisely, we first introduce a nonconvex penalty function as the minimum of the mixed $\ell_{1,2}$ norm over all possible block partitions. Then, by utilizing a variational representation of the mixed $\ell_{1,2}$ norm, we derive the proposed penalty function as a convex relaxation of the nonconvex one. Numerical example shows the effectiveness of the recovery model regularized by the proposed penalty.

1 Introduction

Block-sparsity is a special kind of sparsity, which indicates sparse signals whose nonzero components are clustered in blocks. By assuming the knowledge on the block partition over which the signal is sparse, extensive researches, e.g., [1,2] demonstarate the effectiveness of the mixed $\ell_{1,2}$ norm using the block partition. However, the information of the block partition is not available in many applications such as recovery of acoustic, image, and radar signals. In such situations, a pre-fixed block partition used in the mixed $\ell_{1,2}$ norm often limits the recovery performance.

In this paper, we present a new convex penalty function where the block partition is automatically optimized for the signal to be estimated. We consider an optimal block partition as that yields the minimum of the mixed $\ell_{1,2}$ norm using the block partition. Namely, we introduce a nonconvex penalty function by minimizing the mixed $\ell_{1,2}$ over all possible block partitions. Subsequently, we derive the proposed convex penalty as a convex relaxation of the nonconvex penalty based on a variational representation of the ℓ_2 norm. We demonstrate the effectiveness of the proposed penalty for application to regularized least squares recovery of block-sparse signals.

2 Preliminaries

2.1 Notations and Problem Setting

$\mathbb{R}$ and $\mathbb{R}_+$ denote the sets of all real numbers and all nonnegative real numbers, respectively. For $\boldsymbol{x} = (x_1, \ldots, x_N)^\top \in \mathbb{R}^N$ 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 $\boldsymbol{x} \in \mathbb{R}^N$ by $\mathrm{supp}(\boldsymbol{x}) := \{n \in \{1, \ldots, N\} \mid x_n \neq 0\}$. We denote the cardinality of a set $\mathcal{S}$ by $|\mathcal{S}|$. The ℓ_2 norm, ℓ_1 norm, and the ℓ_0 pseudo-norm of $\boldsymbol{x} \in \mathbb{R}^N$ are respectively denoted by $\|\boldsymbol{x}\|_2 := \sqrt{\sum_{n=1}^N x_n^2}$, $\|\boldsymbol{x}\|_1 := \sum_{n=1}^N |x_n|$, and $\|\boldsymbol{x}\|_0 := |\mathrm{supp}(\boldsymbol{x})|$.

We consider the estimation of $\boldsymbol{x}^\star \in \mathbb{R}^N$, which is supposed to be block-sparse over an unknown block partition $\mathcal{B}_1^\star, \ldots, \mathcal{B}_{K^\star}^\star$. Namely, we suppose that $\boldsymbol{x}^\star_{\mathcal{B}_k^\star}$ contain (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 $\mathcal{B}_k^\star = \{n_k^\star, n_k^\star + 1, \ldots, m_k^\star\}$ for each $k = 1, \ldots, K$.

2.2 Existing Convex Penalties for Block-Sparse Signals

To enhance the block-sparsity of $\boldsymbol{x} \in \mathbb{R}^N$ 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 $\ell_{1,2}$ norm:

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

where we use the weight $\sqrt{|\mathcal{B}_k|}$ by following the suggestions in, e.g., [1,2]. A disadvantage of the mixed $\ell_{1,2}$ norm is that the performance 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 have been designed by using potentially overlapping blocks $\bar{\mathcal{B}}_1, \ldots, \bar{\mathcal{B}}_{\bar{K}} \subset \{1, \ldots, N\}$. Among them, a most reasonable extension is the so-called latent group lasso penalty [3], which is defined as

$$\min_{(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{\bar{K}}) \in \mathbb{R}^{N\times\bar{K}}} \sum_{k=1}^{\bar{K}} \sqrt{|\bar{\mathcal{B}}_k|}\, \|\boldsymbol{v}_k\|_2$$
$$\text{s.t.} \sum_{k=1}^{\bar{K}} \boldsymbol{v}_k = \boldsymbol{x} \text{ and } \mathrm{supp}(\boldsymbol{v}_k) \subset \bar{\mathcal{B}}_k \quad (k = 1, \ldots, \bar{K}).$$

However, using all possible blocks as $\bar{\mathcal{B}}_1, \ldots, \bar{\mathcal{B}}_{\bar{K}}$ is computationally intractable. Even if we restrict the blocks of size d, the latent group lasso penalty is computationally expensive for large d because the number of parameters to be optimized is $\sum_{k=1}^{\bar{K}} \left|\bar{\mathcal{B}}_k\right| = d(N - d + 1)$.

3 Proposed Penalty Function

We first introduce a nonconvex penalty function $\psi_{\mathrm{nc}}(\boldsymbol{x})$, from which the proposed convex penalty $\psi(\boldsymbol{x})$ is derived as a certain convex relaxation.

We define the nonconvex penalty function $\psi_{\rm nc}$ by taking the minimum of the mixed $\ell_{1,2}$ norm over at-most K block partitions:

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

where $\mathcal{P}_\ell$ contains all ℓ block partitions of $\{1,\ldots,N\}$, i.e.,

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

Note that K can be set to an upper bound of the ground-truth $K^\star$. To derive the proposed convex penalty function, we exploit the following variational representation of the ℓ_2 norm:

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

where $\phi\colon \mathbb{R}\times\mathbb{R}_+\to\mathbb{R}_+\cup\{\infty\}$ is defined as

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

This relation readily implies the variational representation of $\psi_{\rm nc}(\boldsymbol{x})$ as

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

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

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

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

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

where $\boldsymbol{D}\in\mathbb{R}^{(N-1)\times N}$ is the first discrete difference operator. Thus, we have

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

Finally, by replacing the ℓ_0 pseudo norm with the ℓ_1 norm in the constraint, we derive the proposed convex penalty function:

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

where $\alpha>0$ is a tuning parameter related to the number of blocks. Since ϕ is a lower semicontinuous convex function whose proximal operator is efficiently computed [4], we can develop efficient proximal-splitting based solvers, e.g., for the least squares model using the proposed penalty function shown in (1).

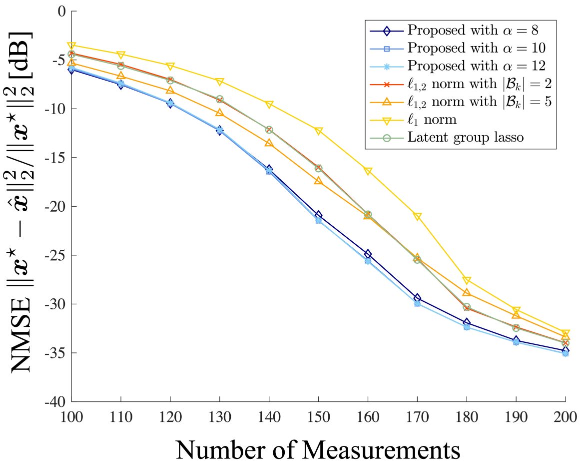

Fig. 1: Normalized mean square error between $\boldsymbol{x}^\star$ and $\hat{\boldsymbol{x}}$ averaged over 100 trials.

4 Numerical Example

We compare the proposed penalty function and existing penalty functions for the estimation of $\boldsymbol{x}^\star\in\mathbb{R}^N$ from noisy linear measurements $\boldsymbol{y} := \boldsymbol{A}\boldsymbol{x}^\star+\boldsymbol{\varepsilon}\in\mathbb{R}^M$, where $\boldsymbol{A}\in\mathbb{R}^{M\times N}$ is the measurement matrix, and $\boldsymbol{\varepsilon}\in\mathbb{R}^M$ is the noise term. Nonzero components of $\boldsymbol{x}^\star$ are randomly divided into 4 blocks, which are randomly located under the condition $|\mathrm{supp}(\boldsymbol{x}^\star)| = 80$, where we set $N=250$. Entries of $\boldsymbol{A}$ are drawn from i.i.d. Gaussian distribution $\mathcal{N}(0,1)$, and $\boldsymbol{\varepsilon}$ is set as the white Gaussian noise where the SNR $\|\boldsymbol{A}\boldsymbol{x}^\star\|^2/\|\boldsymbol{\varepsilon}\|^2$ is set to 40dB. We use the proposed penalty function in the regularized least squares model

$$\min_{\boldsymbol{x}\in\mathbb{R}^N} \|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\|_2^2 + \lambda\psi(\boldsymbol{x}) \tag{1}$$

where $\lambda>0$ is the regularization parameter tuned to give the best result. We obtain the estimate $\hat{\boldsymbol{x}}$ by terminating the iteration when the norm of the difference between successive iterates is below the threshold 10^{-4}. In Fig. 1, we show the normalized mean square error (NMSE) $\|\boldsymbol{x}^\star-\hat{\boldsymbol{x}}\|_2^2/\|\boldsymbol{x}^\star\|_2^2$ against the number of measurements, where the results are averaged over 100 independent trials. The result shows that the proposed penalty function outperforms the existing convex penalty functions including the latent group lasso penalty.

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] 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.

[4] 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.