

Group sparse recovery in impulsive noise via ADM*

Jianwen Huang^a, Feng Zhang^a, Jianjun Wang^{a,b}, Wendong Wang^a

^aSchool of Mathematics and Statistics, Southwest University, Chongqing, 400715, China

^bResearch Center for Artificial Intelligence & Education Big Data, Southwest University, Chongqing, 400715, China

Abstract. In this paper, we consider the recovery of group sparse signals corrupted by impulsive noise. In some recent literature, researchers have utilized stable data fitting models, like l_1 -norm, Huber penalty function and Lorentzian-norm, to substitute the l_2 -norm data fidelity model to obtain more robust performance. In this paper, a stable model is developed, which exploits the generalized l_p -norm as the measure for the error for sparse reconstruction. In order to address this model, we propose an efficient alternative direction method, which includes the proximity operator of l_p -norm functions to the framework of Lagrangian methods. Besides, to guarantee the convergence of the algorithm in the case of $0 \leq p < 1$ (nonconvex case), we took advantage of a smoothing strategy. For both $0 \leq p < 1$ (nonconvex case) and $1 \leq p \leq 2$ (convex case), we have derived the conditions of the convergence for the proposed algorithm. Moreover, under the block restricted isometry property with constant $\delta_{\tau k_0} < \tau/(4 - \tau)$ for $0 < \tau < 4/3$ and $\delta_{\tau k_0} < \sqrt{(\tau - 1)/\tau}$ for $\tau \geq 4/3$, a sharp sufficient condition for group sparse recovery in the presence of impulsive noise and its associated error upper bound estimation are established. Numerical results based on the synthetic block sparse signals and the real-world FECG signals demonstrate the effectiveness and robustness of new algorithm in highly impulsive noise.

Keywords. Alternative direction method; Lagrangian methods; compressed sensing; group sparse; impulsive noise; sparse recovery.

AMS Classification(2010): 94A12, 94A15, 94A08, 68P30, 41A64, 15A52, 42C15

1 Introduction

In the last decade, the problem to find sparse solutions has attracted much interest and been extensively studied, especially in fields of applied mathematics, statistics, machine learning, signal processing [1, 2, 3], etc. Assume that $x \in \mathbb{R}^N$ is the unknown signal that we want to reconstruct. Meanwhile, we suppose that x is sparse in terms of the orthonormal basis, that is, the number of non-zero entries in x is far smaller than its dimensionality, which is denoted by $\|x\|_0$. Define $\Phi \in \mathbb{R}^{n \times N}$ be a measurement matrix with $n \ll N$ and $b \in \mathbb{R}^n$ be a vector of measurements, fulfilling the following relationship

$$b = \Phi x.$$

*Corresponding author. E-mail address: wjj@swu.edu.cn (J. Wang).

Then, one could recover the true signal x from the linear equations of underdetermined system $b = \Phi x$ by means of certain recovery algorithm. But, there exist infinite solutions, since the number of the unknown variable is far larger than that of the linear equations, i.e., it is ill-conditioned.

In the compressive sensing (CS), provided that x is (approximately) sparse, one could naturally consider looking for the sparsest solution among all those, i.e.,

$$\min_{x \in \mathbb{R}^N} \|x\|_0, \text{ subject to } b = \Phi x + z, \quad (1.1)$$

where $\|x\|_0$ counts the number of the non-zero components in x , and z is a vector of measurement noise. Unfortunately, l_0 -problem (1.1) is combinatorial and computationally complicated. In CS, an alternative approach is to substitute the l_0 -problem with the l_1 -problem, which is the so-called basis pursuit (BP) problem [4]:

$$\min_{x \in \mathbb{R}^N} \|x\|_1, \text{ subject to } b = \Phi x + z. \quad (1.2)$$

Candès [5] proved that the solutions to problem (1.2) are equivalent to those of (1.1) with overwhelming probability under some favorable conditions. The constrained minimization problem generally can be transformed into the regularized least squares problem (also called basis pursuit denoising or LASSO):

$$\min_{x \in \mathbb{R}^N} \|x\|_1 + \frac{1}{\nu} \|\Phi x - b\|_2^2, \quad (1.3)$$

where ν is a positive regularization parameter, which plays an important role in trading off both terms. It has been showed that l_1 minimization is an effective method to reconstruct sparse signals, since it is convex and easily tractable. Thus, its applications are extremely broad in the area of sparse recovery.

As in (1.3) and its variants, with respect to maximum likelihood context, the data fitting model in l_2 -norm is best for Gaussian noise. However, in a variety of practical scenarios, the measuring residual error may be of various types or combinations. Impulsive noise is a representative type. It has been extensively investigated in modern statistics and can be utilized to simulate large residual errors in the observations. A lot of image processing and nonlinear signal processing literature have proposed impulsive disturbance, which is caused by missing data in the sampling process, communication issues [6, 7], malfunctioning pixels [8], and buffer overflow [9]. Under these settings, it has been showed that the fluctuation of least squares estimation is remarkably large, since it is easily affected by outliers in the measuring process. Consequently, it is inefficient to use the data fitting model in l_2 -norm.

In recent years, researchers have proposed various stable models for CS to repress the outliers in the observations. In [10], the researchers have employed the l_1 -norm as the data fitting model to establish a stable formulation:

$$\min_{x \in \mathbb{R}^N} \|x\|_1 + \frac{1}{\nu} \|\Phi x - b\|_1. \quad (1.4)$$

The researchers in [10] have proved that when observation data is corrupted by impulsive noise, the formulation consisting of a data fitting model measured in the l_1 -norm performs better than that of the l_2 -norm. Later, the researchers in [11] have introduced more efficient alternating direction algorithm for above formulation. In 2017, the researchers in [12] substituted the l_1 -norm data

fidelity term of the problem (1.4) with the generalized l_p -norm ($p \in [0, 2)$) to give the stable formulation as follows:

$$\min_{x \in \mathbb{R}^N} \|x\|_1 + \frac{1}{\nu} \|\Phi x - b\|_p^p, \quad (1.5)$$

where $\|u\|_p = (\sum_{i=1}^n |u_i|^p)^{1/p}$, $u \in \mathbb{R}^n$. Observe that when $p = 2$ and $p = 1$, formulation (1.5) respectively degenerates to the formulations (1.3) and (1.4).

However, there are some real signals have additional structure information. In practical applications, it has been showed that a broad class of signals possess some “group sparsity” structure. This means the signal has a natural grouping of its coefficients, and the coefficients with a group are probably either all zeros or all non-zeros. Such group sparse signals have a large number of applications such as expression quantitative trait locus mapping [13], graphical statistics [14], Modeling disease progression [15], Video-to-Shot Tag Propagation [16], click through rate prediction in display advertising [17], classification problems [18], etc. Let $x \in \mathbb{R}^N$ be the signal that we wish to capture. Suppose that $\{x_{g_i} \in \mathbb{R}^{N_i} : i = 1, 2, \dots, k\}$ is the grouping of x with $g_i \subseteq \{1, 2, \dots, N\}$ standing for an index set associated with the i -th group, where x_{g_i} indicates the subvector of x that is indexed by g_i . In general, g_i can be any index set, and based on prior information, we could preset them. The recovery of group sparse signals recently has triggered many study activities. In particular, the block sparse optimization is one of the group sparse optimization which attracts many researchers’ interest; for more details, we refer readers to see [19, 20, 21, 22, 23, 24, 25, 26].

In this paper, to be more general, we propose the following model to stably reconstruct the group sparse signals contaminated by impulsive noise:

$$\min_{x \in \mathbb{R}^N} \|x\|_{2,\mathcal{I}} + \frac{1}{\nu} \|\Phi x - b\|_p^p, \quad (1.6)$$

where the definition of the parameters ν, p are the same as above mention, the $l_{2,\mathcal{I}}$ -norm is determined by $\|x\|_{2,\mathcal{I}} = \sum_{i=1}^k \|x_{g_i}\|_2$.

In order to study the theoretical analysis of new model (1.6), the restricted isometry property (RIP) [27] is extended to the block restricted isometry property (block-RIP) [40], which we will give in Section 3.

In recent year, researchers have done some work concerning the reconstruction of block sparse signals, which includes various central results about block-RIP condition or others. For the Gaussian noise case, see [40, 19, 20]. Furthermore, various sufficient conditions and other results on recovery of block sparse signals were gained in contributions [21, 22, 23, 24, 25, 26]. However, all these researches are discussed only in Gaussian noise case, that is, the observation measurement b is disturbed by Gaussian noise. From the viewpoint of application, the exploration on recovering block sparse signals in the setting of impulsive noise is more practical.

Main contributions of this paper are abbreviated here. Firstly, to the best of our knowledge, we first consider the issue of reconstructing block sparse signals disturbed by impulsive noise. Secondly, we provide a sharp sufficient condition and associated error upper bound estimation for recovery of block sparse signals in the presence of impulsive noise. Thirdly, we propose an efficient algorithm to solve the new model. Fourthly, for both the convex and nonconvex situations, we prove the convergence of the new algorithm. Finally, a series of numerical experiments are carried out to show better performance of the new algorithm.

The rest of the paper is organized as follows. In Section 2, we give some preliminaries, which is used in later section. Section 3 provides some analysis on the proposed formulation and a

new algorithm is proposed in Section 4 to solve that formulation. In Section 5, we present the convergence condition of the proposed algorithm. In Section 6, we offer simulation results. Finally, the conclusion is addressed in Section 7.

2 Some preliminaries

It is known that for $x \in \mathbb{R}^N$, the proximity operator of some function $f(x)$ with the regularization parameter μ is defined as follows:

$$\text{prox}_{f,\mu}(s) = \arg \min_x \left\{ f(x) + \frac{\mu}{2} \|x - s\|_2^2 \right\}. \quad (2.1)$$

When $f(x) = c\|x\|_p^p$ with $c \in (0, +\infty)$ and $p \in [0, 2)$, solving the problem (2.1) degenerates to resolving N single variable minimization problems. Accordingly, it is computationally no difficult to calculate. According to the range of p , the computation of $\text{prox}_{f,\mu}(s)$ is divided into following four cases:

a. $p = 0$. The proximity operator returns to the known hard thresholding operator

$$\text{prox}_{f,\mu}(s)_i = \begin{cases} 0, & |s_i| \leq \sqrt{2c/\mu}, \\ s_i, & \text{otherwise,} \end{cases} \quad (2.2)$$

where s_i denotes the i -th component of the vector s .

b. $p = 1$. The researchers [28] presented an explicit form of the proximity operator as follows:

$$\text{prox}_{f,\mu}(s)_i = \max \left\{ |s_i| - \frac{c}{\mu}, 0 \right\} \text{sign}(s_i) = \text{Shrink} \left(s_i, \frac{c}{\mu} \right), \quad (2.3)$$

where $\text{sign}(\cdot)$ is the sign function, and “Shrink (\cdot, \cdot) ” indicates the well-known one dimensional shrinkage or soft thresholding [10] [11].

c. $p \in (0, 1)$. Due to [29], the proximity operator obeys

$$\text{prox}_{f,\mu}(s)_i = \begin{cases} 0, & |s_i| \leq \sigma, \\ \{0, \text{sign}(s_i)\phi\}, & |s_i| = \sigma, \\ \text{sign}(s_i)z_i, & |s_i| > \sigma, \end{cases} \quad (2.4)$$

where $\sigma = cp\phi^{p-1}/\mu + \phi$, $\phi^{2-p} = 2c(1-p)/\mu$, $z_i \in (\phi, |s_i|)$ fulfils the equation $g_1(z) = cpz^{p-1} + \mu z - \mu|s_i| = 0$.

d. $p \in (1, 2)$. It is not hard to check that $f(x)$ has the smoothness and convexity. The researchers [30] gained the explicit expression of the proximity operator

$$\text{prox}_{f,\mu}(s)_i = \text{sign}(s_i)z_i, \quad (2.5)$$

where $z_i \geq 0$, and z_i satisfies the following equation

$$g_2(z) = cpz^{p-1} + \mu z - \mu|s_i| = 0. \quad (2.6)$$

One can easily verify that in the case of $s_i \neq 0$, $g_2(z)$ meets the inequalities $g_2(0) < 0 < g_2(|s_i|)$. Besides, the function $g_2(z)$ has the concavity, and it ascends as z increases. Therefore, the solution

of (2.6) such that $0 < z_i < |s_i|$ holds, where $s_i \neq 0$. And we could employ a Newton's method to calculate it. Set $\zeta = \mu|s_i|/(pc + \mu)$. One could pick a positive lower bound of the solution as the initial value as follows:

$$z_i^0 = \begin{cases} \zeta^{1/(p-1)}, & \zeta < 1, \\ \zeta, & \text{otherwise.} \end{cases} \quad (2.7)$$

However, in practical operation, in the case of $\zeta < 1$ and $p \rightarrow 1$, the computational value of $\zeta^{1/(p-1)}$ is probably extreme small. In order to resolve this problem, when $\zeta < 1$, the corresponding initial value is provided by

$$z_i^0 = \begin{cases} \chi, & g_2(\chi) \leq 0, \\ 0, & g_2(\chi) > 0, \end{cases}$$

where χ is a small positive constant which is preset like $\chi = 10^{-10}$.

In the following, we provide the definitions of three different impulsive noise and their corresponding p -moments.

(i) $S\tilde{\alpha}S$ noise

Researchers [31] [12] showed that symmetric $\tilde{\alpha}$ -stable (abbreviated $S\tilde{\alpha}S$) distribution can be used to model impulsive noise. Although one cannot analytically present the probability density function (pdf) for a general stable distribution, the general characteristic function can be. Any pdf is given by the Fourier transform of its characteristic function $\varphi(w)$ by:

$$\tilde{f}(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \varphi(w) e^{iwx} dw.$$

If the characteristic function is given by

$$\varphi(w) = \exp(iaw - \gamma^{\tilde{\alpha}}|w|^{\tilde{\alpha}}),$$

then a variable random X is said to follow the $S\tilde{\alpha}S$ distribution [33, 34]. Here, a is the location parameter, γ is the scale parameter, and $\tilde{\alpha}$ is the characteristic exponent measuring the thickness of the distributional tail with $\tilde{\alpha} \in (0, 2]$. If the value of $\tilde{\alpha}$ is smaller, then the tail of the $S\tilde{\alpha}S$ distribution is thicker and consequently the noise is more impulsive. Assume that $a = 0$, for independently identically distributed (iid) $S\tilde{\alpha}S$ noise, it follows from [12] that the p -th moment of noise vector z

$$E\{\|z\|_p^p\} = \begin{cases} nC(p, \tilde{\alpha})\gamma^p, & 0 < p < \tilde{\alpha}, \\ +\infty, & p \geq \tilde{\alpha}, \end{cases} \quad (2.8)$$

where $C(p, \tilde{\alpha}) = 2^{p+1}\Gamma(\frac{p+1}{2})\Gamma(-\frac{p}{\tilde{\alpha}})/(\tilde{\alpha}\sqrt{\pi}\Gamma(-\frac{p}{2}))$ and $\Gamma(\theta) = \int_0^{+\infty} x^{\theta-1}e^{-x} dx$ denotes the gamma function ($\Gamma(1/2) = \sqrt{\pi}$).

(ii) GGD/GED noise.

Since the pdf of $S\tilde{\alpha}S$ distribution doesn't have the explicit representation, it is hard to apply in some contexts. As one of its alternative, one can also employ the generalized Gaussian distribution (for short GGD) or general error distribution (for short GED) to model impulsive noise. The pdf of $X \sim GGD$ with zero mean is determined by

$$\tilde{f}(x) = \frac{\tilde{v}}{2\tilde{\sigma}\Gamma(\frac{1}{\tilde{v}})} \exp\left(-\left[\frac{|x|}{\tilde{\sigma}}\right]^{\tilde{v}}\right) \quad (2.9)$$

where $\tilde{\sigma} > 0$ is a scale parameter, and $\tilde{v} > 0$ is a shape parameter controlling the distribution shape. In some literature, the properties of it have been well studied, see [35, 36, 37, 38]. This is a parametric family of symmetric distributions. This family includes the Gaussian distribution when $\tilde{v} = 2$ and it incorporates the Laplace distribution when $\tilde{v} = 1$. This family allows for tails that are either heavier than Gaussian (when $0 < \tilde{v} < 2$) or lighter than Gaussian (when $\tilde{v} > 2$). Therefore, it models impulsive noise is appropriate in the case of $0 < \tilde{v} < 2$. For iid GGD noise, it is no difficult to verify that the p -th order moment is [12]

$$E\{\|z\|_p^p\} = n\tilde{\sigma}^p \Gamma\left(\frac{p+1}{\tilde{v}}\right) / \Gamma\left(\frac{1}{\tilde{v}}\right). \quad (2.10)$$

(iii) Gaussian mixture noise.

A two-term Gaussian mixture model is defined by

$$(1 - \lambda)N(0, \tilde{\sigma}^2) + \lambda N(0, \kappa\tilde{\sigma}^2), \quad (2.11)$$

where λ represents the portion of outliers in the noise with $\lambda \in [0, 1)$ and κ is the power of outliers with $\kappa > 1$. It is easy to see that its pdf is

$$\tilde{f}(x) = \frac{1}{\tilde{\sigma}\sqrt{2\pi(1-\lambda+\kappa\lambda)}} \exp\left(-\frac{x^2}{2(1-\lambda+\kappa\lambda)\tilde{\sigma}^2}\right), \quad x \in \mathbb{R}. \quad (2.12)$$

In this model, the first term stands for the background noise as well as another term denotes the impulsive property of the noise. For related recent research, see [39]. Observe that the total variance of the noise is $(1 - \lambda + \kappa\lambda)\tilde{\sigma}^2$. It follows from some elementary calculation that the p -th order moment of such noise is [12]

$$E\{\|z\|_p^p\} = \frac{n2^{\frac{p}{2}}(1 - \lambda + \kappa\lambda)^{\frac{p}{2}}\tilde{\sigma}^p \Gamma\left(\frac{p+1}{2}\right)}{\sqrt{\pi}}. \quad (2.13)$$

3 Analysis on the proposed formulation

In this place, we show in the case that the original signal is contaminated by impulsive noise, the proposed formulation can guarantee the stable recovery of the true signal. For simplification, suppose that the special grouping of x is as follows:

$$x = [\underbrace{x_1, \dots, x_{d_1}}_{x_{g_1}}, \underbrace{x_{d_1+1}, \dots, x_{d_1+d_2}}_{x_{g_2}}, \dots, \underbrace{x_{N-d_k+1}, \dots, x_N}_{x_{g_k}}]^T, \quad (3.1)$$

where $N \gg k \geq 2$ is an integer, and x_{g_i} stands for the i th group of x associated with the group size d_i and $N = d_1 + d_2 + \dots + d_k$.

Eldar and Mishali [40] introduced the definition of block restricted isometry property (block-RIP) of a measurement matrix to describe the condition under which the desired signal could be robustly reconstructed with small or zero error. Let $\Phi \in \mathbb{R}^{n \times N}$ be a measurement matrix. If there exists a constant δ_{k_0} such that

$$(1 - \delta_{k_0})\|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1 + \delta_{k_0})\|x\|_2^2 \quad (3.2)$$

holds, for every $x \in \mathbb{R}^N$ that is block k_0 -sparse over $\mathcal{J} = \{d_1, d_2, \dots, d_k\}$, then Φ is said to have the block-RIP over \mathcal{J} with constant δ_{k_0} , where a block k_0 -sparse signal x is a signal of the form (3.1) in which x_{g_i} is nonzero for at most k_0 indices g_i . Suppose that \tilde{x} is the solution to the following problem

$$\min_{x \in \mathbb{R}^N} \|x\|_{2,\mathcal{I}}, \text{ subject to } \|\Phi x - b\|_2 \leq \eta, \quad (3.3)$$

where $\|x\|_{2,\mathcal{I}}$ is a special case of $\|x\|_{w,2,\mathcal{I}}$, i.e., when $w_i = 1$, $i = 1, 2, \dots, k$. Eldar and Mishali [40] also proved that if Φ satisfies (3.2) with $\delta_{2k_0} < \sqrt{2} - 1$ and $\|z\|_2 \leq \eta$, then

$$\|x - \tilde{x}\|_2 \leq \frac{4\sqrt{\delta_{2k_0}}}{1 - (1 + \sqrt{2})\delta_{2k_0}}\eta. \quad (3.4)$$

The above result shows that based on the method (3.3), the unknown signal could be robustly reconstructed with an error depending on the noise level if the variance of the noise is finite. However, in the case of impulsive noise that its variance is infinite, the method (3.3) isn't stable. In this setting, the formulation (1.6) is better.

Theorem 3.1. *Let \tilde{x} be the solution to the problem (1.6). Assume that $x \in \mathbb{R}^N$ is some block k_0 -sparse signal that we want to recover and $\|z\|_p \leq \eta$.*

(a) *For $\tau \geq 4/3$, if the sensing matrix Φ satisfies the block-RIP with constant $\delta_{\tau k_0} < \sqrt{(\tau - 1)/\tau}$, the solution satisfies*

$$\|x - \tilde{x}\|_2 \leq \frac{2\sqrt{2\tau(\tau - 1)(1 + \delta_{\tau k_0})}}{\tau(\sqrt{\frac{\tau-1}{\tau}} - \delta_{\tau k_0})}\eta. \quad (3.5)$$

(b) *For $0 < \tau < 4/3$, if the sensing matrix Φ satisfies the block-RIP with constant $\delta_{\tau k_0} < \tau/(4 - \tau)$, the solution satisfies*

$$\|x - \tilde{x}\|_2 \leq \frac{2\sqrt{2(1 + \delta_{\tau k_0})} \max\{\sqrt{\tau}, \tau\}}{(4 - \tau)(\frac{\tau}{4 - \tau} - \delta_{\tau k_0})}\eta. \quad (3.6)$$

Remark 3.1. *The inequalities (3.5) and (3.6) present an upper bound estimation of the reconstructed error on the formulation (1.6). Compared with the assumption of the noise of the problem (3.3) that $\|z\|_2 \leq \eta$, in Theorem 3.1, it extends the assumption of the noise to $\|z\|_p \leq \eta$, $p \in [0, 2]$. The assumption of the noise on robust reconstruction is relaxed from that associating variance is finite to that associating p th-order moment is finite. This shows that when the noise is impulsive having infinite variance, the desired signal x could be robustly reconstructed by the proposed model. Furthermore, for both $0 < \tau < 4/3$ and $\tau \geq 4/3$ cases, the bound of block-RIP constant $\delta_{\tau k_0}$ are sharp, that is, it is impossible to improve the bound of constant $\delta_{\tau k_0}$; for more details, see [41, 42].*

Remark 3.2. *We show that different choices of τ can lead to different conditions. Observe that when $\tau = 2$, we get the condition $\delta_{2k_0} < \sqrt{2}/2 \approx 0.707$. Therefore, we obtain a weaker condition than $\delta_{2k_0} < \sqrt{2} - 1 \approx 0.414$ provided in [12]. In Figure 6.3, the error bound constant is plotted versus δ_{2k_0} . From the observation of Figure 6.3, the error bound constant in (3.5) is smaller than that of (9) in Theorem 1 in [12].*

Proof. Since the proofs of two cases are similar, we only give the proof of the case of $\tau \geq 4/3$. It is similar to that of Theorem 1 in [12]. Suppose that $\tilde{x} = x + h$ is a solution to the problem (1.6),

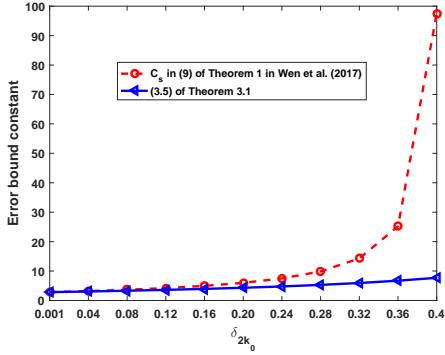


Fig. 3.1: Error bound constant versus δ_{2k_0}

where x is the original signal. Chen and Li [41] showed that if the matrix Φ fulfils the block-RIP with $\delta_{\tau k_0} < \sqrt{(\tau - 1)/\tau}$ for $\tau \geq 4/3$, then the following equation

$$\|h\|_2 \leq \frac{\sqrt{2\tau(\tau-1)(1+\delta_{\tau k_0})}}{\tau(\sqrt{\frac{\tau-1}{\tau}} - \delta_{\tau k_0})} \|\Phi h\|_2 \quad (3.7)$$

holds. By the condition of noise of the theorem, we get

$$\|\Phi x - b\|_p \leq \eta \text{ and } \|\Phi \tilde{x} - b\|_p \leq \eta. \quad (3.8)$$

Note that when $p \in (0, 2]$, for any $u \in \mathbb{R}^n$,

$$\|u\|_2 \leq \|u\|_p. \quad (3.9)$$

Combining with (3.8) and (3.9), we get

$$\begin{aligned} \|\Phi h\|_2 &\leq \|\Phi x - b\|_2 + \|\Phi \tilde{x} - b\|_2 \\ &\leq \|\Phi x - b\|_p + \|\Phi \tilde{x} - b\|_p \\ &\leq 2\eta. \end{aligned} \quad (3.10)$$

Substituting (3.10) into (3.7), it leads to (3.5). Accordingly, we complete the proof of the result. \square

4 Algorithm

In this section, based on the alternating direction method (ADM), we propose an efficient algorithm to address the formulation (1.6). Specifically, the construction of the algorithm incorporates two main steps: (a) to restate the formulation (1.6) into one satisfying separable objective function via introducing new variables and adding constraints; (b) to take advantage of an alternating direction method to the associated formulation.

A. Block- L_p -ADM without smoothing

With a new variable $u \in \mathbb{R}^n$, the problem (1.6) is equivalent to

$$\min_{x,u} \|x\|_{2,\mathcal{I}} + \frac{1}{\nu} \|u\|_p^p, \text{ s.t. } u = \Phi x - b. \quad (4.1)$$

Then the associated augmented Lagrangian function is

$$L_p(x, u, y) = \|x\|_{2,\mathcal{I}} + \frac{1}{\nu} \|u\|_p^p - \langle y, \Phi x - b - u \rangle + \frac{\alpha}{2} \|\Phi x - b - u\|_2^2, \quad (4.2)$$

where $y \in \mathbb{R}^n$ denotes the Lagrangian multiplier, and α represents a positive regularization parameter. Now, applying ADM to (4.2) and some elementary calculation produces the iterations as follows:

$$u^{t+1} = \arg \min_{u \in \mathbb{R}^n} \left\{ \frac{1}{\nu} \|u\|_p^p + \frac{\alpha}{2} \|\Phi x^t - b - u - \frac{y^t}{\alpha}\|_2^2 \right\}, \quad (4.3)$$

$$x^{t+1} = \arg \min_{x \in \mathbb{R}^N} \left\{ \|x\|_{2,\mathcal{I}} + \frac{\alpha}{2} \|\Phi x - b - u^{t+1} - \frac{y^t}{\alpha}\|_2^2 \right\}, \quad (4.4)$$

$$y^{t+1} = y^t - \gamma \alpha (\Phi x^{t+1} - b - u^{t+1}), \quad (4.5)$$

where γ is a positive constant. In the rest of this paper, set $\gamma = 1$.

Firstly, observe that the minimizer u^{t+1} of (4.3) is a form of the proximity operator (2.1), thus we can compute it as

$$u^{t+1} = \text{prox}_{\frac{1}{\nu} \|u\|_p^p, \alpha}(\xi^t) = \begin{cases} \text{solved as (2.2)}, & p = 0, \\ \text{solved as (2.4)}, & p \in (0, 1), \\ \text{Shrink } (\xi^t, \frac{1}{\alpha\nu}), & p = 1, \\ \text{solved as (2.5)}, & p \in (1, 2), \\ \frac{\nu\alpha\xi^t}{\nu\alpha+2}, & p = 2, \end{cases} \quad (4.6)$$

where $\xi^t = \Phi x^t - b - y^t/\alpha$, and Shrink(\cdot, \cdot) is component-wise.

Secondly, we consider the minimization problem (4.4). However, its analytical solution of (4.4) is not clear. Set $v^t = b + u^{t+1} + \frac{y^t}{\alpha}$. Let

$$h_1(x^t) = \Phi^\top (\Phi x^t - v^t)$$

denote the gradient of $\frac{1}{2} \|\Phi x - v^t\|_2^2$ at x^t . Instead of directly solving (4.4), it can be approximated by

$$\begin{aligned} x^{t+1} &\approx \arg \min_{x \in \mathbb{R}^N} \left\{ \|x\|_{2,\mathcal{I}} + \alpha \left((h_1(x^t))^\top (x - x^t) + \frac{\rho_1}{2} \|x - x^t\|_2^2 \right) \right\} \\ &= \arg \min_{x \in \mathbb{R}^N} \left\{ \|x\|_{2,\mathcal{I}} + \frac{\alpha\rho_1}{2} \|x - x^t + \frac{h_1(x^t)}{\rho_1}\|_2^2 \right\}, \end{aligned} \quad (4.7)$$

where ρ_1 is a positive proximal parameter. By simple computation, (4.7) is equivalent to

$$\arg \min_{x \in \mathbb{R}^N} \sum_{i=1}^k \left\{ \|x_{g_i}\|_2 + \frac{\alpha\rho_1}{2} \|x_{g_i} - r_i\|_2^2 \right\}, \quad (4.8)$$

where $r_i = (x^t)_{g_i} - (h_1(x^t))_{g_i}/\rho_1$. Applying the one-dimensional shrinkage formulate, we can obtain a closed form solution as follows:

$$x_{g_i} = \max \left\{ \|r_i\|_2 - \frac{1}{\alpha\rho_1}, 0 \right\} \frac{r_i}{\|r_i\|_2}, \text{ for } i = 1, 2, \dots, k. \quad (4.9)$$

The convention $0 \cdot 0/0 = 0$ is followed.

B. Block- L_p -ADM utilizing smoothed l_1 -regularization for nonconvex case

When $1 \leq p < 2$ (convex case), under the condition of a reasonable choice of the parameter ρ_1 , the convergence of the above Block- L_p -ADM algorithm can be ensured. But, when $0 \leq p < 1$ (nonconvex case), the convergence of this algorithm is not ensured. In order to resolve this problem, adding a smoothed parameter to the weighted $l_{2,\mathcal{I}}$ regularization term, we derive a smoothed version of the problem (1.6) as follows:

$$\min_{x \in \mathbb{R}^N} \|x\|_{2,\mathcal{I}}^\epsilon + \frac{1}{\nu} \|\Phi x - b\|_p^p, \quad (4.10)$$

where

$$\|x\|_{2,\mathcal{I}}^\epsilon = \sum_{i=1}^k (\epsilon^2 + \|x_{g_i}\|_2^2)^{\frac{1}{2}},$$

and ϵ is a positive constant. Similar to (4.1), the above problem can be converted into

$$\min_{x,u} \|x\|_{2,\mathcal{I}}^\epsilon + \frac{1}{\nu} \|u\|_p^p, \text{ s.t. } u = \Phi x - b. \quad (4.11)$$

The augmented lagrangian function is of the form

$$L_{p,\epsilon}(x, u, y) = \|x\|_{2,\mathcal{I}}^\epsilon + \frac{1}{\nu} \|u\|_p^p - \langle y, \Phi x - b - u \rangle + \frac{\alpha}{2} \|\Phi x - b - u\|_2^2. \quad (4.12)$$

For fixed $u = u^{t+1}$ and $y = y^t$, the minimizer x^{t+1} of (4.12) with respect to x is provided by

$$x^{t+1} = \arg \min_{x \in \mathbb{R}^N} \left\{ \|x\|_{2,\mathcal{I}}^\epsilon + \frac{\alpha}{2} \|\Phi x - v^t\|_2^2 \right\}. \quad (4.13)$$

Note that the objective function of (4.13) has the smoothness, thus one could make use of general iterative approaches to address this minimization problem. Nevertheless, in order to get efficiency of the method, the technique of ADM is employed again to resolve (4.13). More precisely, for a fixed x^t , the regularization term $\|x\|_{2,\mathcal{I}}^\epsilon$ is linearized as

$$\|x\|_{2,\mathcal{I}}^\epsilon \approx \|x^t\|_{2,\mathcal{I}}^\epsilon + (h_2(x^t))^\top (x - x^t) + \frac{\rho_2}{2} \|x - x^t\|_2^2, \quad (4.14)$$

where $h_2(x^t) = \nabla \|x^t\|_{2,\mathcal{I}}^\epsilon$ is the gradient of $\|x\|_{2,\mathcal{I}}^\epsilon$ at x^t , and ρ_2 is a positive proximal parameter. Plugging (4.14) into (4.13), we gain

$$\begin{aligned} x^{t+1} &\approx \arg \min_{x \in \mathbb{R}^N} \left\{ (h_2(x^t))^\top (x - x^t) + \frac{\rho_2}{2} \|x - x^t\|_2^2 + \frac{\alpha}{2} \|\Phi x - v^t\|_2^2 \right\} \\ &= \arg \min_{x \in \mathbb{R}^N} \left\{ \frac{1}{2} x^\top \left(\rho_2 I_N + \alpha \Phi^\top \Phi \right) x + \left(h_2(x^t) - \rho_2 x^t - \alpha \Phi^\top v^t \right)^\top x \right\}. \end{aligned} \quad (4.15)$$

Observe that it is a convex quadratic problem, consequently it degenerates to solving the linear system as follows:

$$\left(\rho_2 I_N + \alpha \Phi^\top \Phi \right) x = \rho_2 x^t - h_2(x^t) + \alpha \Phi^\top v^t. \quad (4.16)$$

Based on all above derivations, we state the Block- L_p -ADM algorithm for problem (1.6) as follows:

Algorithm 1 : Block- L_p -ADM

- 1: Initialize $x^0 \in \mathbb{R}^N$ and $y^0 \in \mathbb{R}^n$. Constants k, ν, α, ρ_1 , and ρ_2 . Set $t = 0$.
 - 2: **while** stopping criterion is not satisfied **do**
 - 3: $u^{t+1} \leftarrow$ applying (4.6);
 - 4: $x^{t+1} \leftarrow \begin{cases} \text{applying (4.9), } & 1 \leq p < 2, \\ \text{solving (4.16), } & 0 \leq p < 1; \end{cases}$
 - 5: $y^{t+1} \leftarrow y^t - \alpha (\Phi x^{t+1} - b - u^{t+1})$;
 - 6: $t = t + 1$;
-

5 Convergence analysis

First, for $1 \leq p < 2$, in the case that the minimizer x^{t+1} of the problem (4.4) is solved by (4.9), the condition of the convergence for Block- L_p -ADM will be established.

Theorem 5.1. *For any $\alpha > 0, 1 \leq p < 2$, under the assumption of $\rho_1 > \lambda_{\max}(\Phi^\top \Phi)$, the sequence (x^t, u^t, y^t) generated by Algorithm 1 from any initiated value (x^0, y^0) converges to a solution to (4.1).*

Proof. By using optimization theory, we have

$$\begin{aligned} 0 &\in \nabla_x L_p(x, u, y) = \partial \|x\|_{2,\mathcal{I}} - \Phi^\top y + \alpha \Phi^\top (\Phi x - b - u), \\ 0 &= \nabla_u L_p(x, u, y) = \frac{1}{\nu} \nabla \|u\|_p^p + y - \alpha (\Phi x - b - u). \end{aligned} \quad (5.1)$$

Define (\hat{x}, \hat{u}) be the solution of (4.1) with $\Phi \hat{x} - \hat{u} = b$. Then, (5.1) shows that there is $\hat{y} \in \mathbb{R}^n$ satisfying the following equations:

$$\Phi^\top \hat{y} \in \partial \|\hat{x}\|_{2,\mathcal{I}}, \quad \nabla \|\hat{u}\|_p^p + \nu \hat{y} = 0, \quad \text{and} \quad \Phi \hat{x} - \hat{u} = b. \quad (5.2)$$

Set $\tilde{x} = x^{t+1}, \tilde{u} = u^{t+1}$ and $\tilde{y} = y^t - \alpha (\Phi \tilde{x} - b - \tilde{u})$. For fixed $x = x^t$ and $y = y^t$, the minimizer u^{t+1} of (4.2) with respect to u such that

$$\frac{1}{\nu} \nabla \|u^{t+1}\|_p^p + y^t - \alpha (\Phi x^t - b - u^{t+1}) = 0 \quad (5.3)$$

holds. By the definitions of \tilde{x}, \tilde{u} and \tilde{y} , (5.3) is transformed into

$$\nabla \|\tilde{u}\|_p^p = \nu [-\tilde{y} - \alpha \Phi (\tilde{x} - x^t)]. \quad (5.4)$$

Combining with the convexity of $\|u\|_p^p$, (5.4) and the fact that $\nabla \|\hat{u}\|_p^p + \nu \hat{y} = 0$, we derive

$$(\tilde{u} - \hat{u})^\top (\hat{y} - \tilde{y} - \alpha \Phi (\tilde{x} - x^t)) \geq 0. \quad (5.5)$$

The inequality (5.5) is just the equation (A.3) of Theorem 2.1 in [10]. The reminder of the proof is from the proof of Theorem 2.1 in [10].

□

Next, in the case that the minimizer x^{t+1} is solved by (4.16), the condition of the convergence for Algorithm 1 will be established.

Theorem 5.2. Assume that $\epsilon > 0$ and $\Phi\Phi^\top - \vartheta I_n \succeq 0$ ($\vartheta > 0$). The sequence (x^t, u^t, y^t) generated by Algorithm 1 from any initiated value (x^0, y^0) tends to the critical point of (4.11) for any $0 \leq p < 1$, provided that $\rho_2 > \frac{C}{2}$ and

$$\alpha > \frac{2\rho_2^2 + 2(\rho_2 + C)}{\vartheta(\rho_2 - C)},$$

where $C = \sqrt{2(1 + d_{\max})}/\epsilon$, and $d_{\max} = \max\{d_1, d_2, \dots, d_k\}$ is the maximum of the group size $\{d_1, d_2, \dots, d_k\}$.

In order to prove the main result, some auxiliary lemmas are presented.

Lemma 5.1. $\nabla\|x\|_{2,\mathcal{I}}^\epsilon$ is C -Lipschitz continuous, i.e., for any $x, y \in \mathbb{R}^N$, the equation

$$\|\nabla\|x\|_{2,\mathcal{I}}^\epsilon - \nabla\|y\|_{2,\mathcal{I}}^\epsilon\|_2 \leq C\|x - y\|_2$$

holds, where $C = \sqrt{2(1 + d_{\max})}/\epsilon$, and $d_{\max} = \max\{d_1, \dots, d_k\}$.

Proof.

It is easy to check that the gradient of $\|x\|_{2,\mathcal{I}}^\epsilon$ is

$$\nabla\|x\|_{2,\mathcal{I}}^\epsilon = \left[\frac{x_1}{(\epsilon^2 + \|x_{g_1}\|_2^2)^{\frac{1}{2}}}, \frac{x_2}{(\epsilon^2 + \|x_{g_1}\|_2^2)^{\frac{1}{2}}}, \dots, \frac{x_N}{(\epsilon^2 + \|x_{g_k}\|_2^2)^{\frac{1}{2}}} \right]^\top. \quad (5.6)$$

Next, we compute the constant C that satisfies for any $x, y \in \mathbb{R}^N$,

$$\|\nabla\|x\|_{2,\mathcal{I}}^\epsilon - \nabla\|y\|_{2,\mathcal{I}}^\epsilon\|_2 \leq C\|x - y\|_2. \quad (5.7)$$

For convenience of discussion, (5.7) is rewritten as

$$\|\nabla\|x\|_{2,\mathcal{I}}^\epsilon - \nabla\|y\|_{2,\mathcal{I}}^\epsilon\|_2^2 \leq C^2\|x - y\|_2^2. \quad (5.8)$$

It follows from (5.6) that

$$\begin{aligned} & \|\nabla\|x\|_{2,\mathcal{I}}^\epsilon - \nabla\|y\|_{2,\mathcal{I}}^\epsilon\|_2^2 \\ &= \left(\frac{x_1}{(\epsilon^2 + \|x_{g_1}\|_2^2)^{\frac{1}{2}}} - \frac{y_1}{(\epsilon^2 + \|y_{g_1}\|_2^2)^{\frac{1}{2}}} \right)^2 + \left(\frac{x_2}{(\epsilon^2 + \|x_{g_1}\|_2^2)^{\frac{1}{2}}} - \frac{y_2}{(\epsilon^2 + \|y_{g_1}\|_2^2)^{\frac{1}{2}}} \right)^2 + \dots. \end{aligned} \quad (5.9)$$

Set

$$l_i = \left(\frac{x_i}{(\epsilon^2 + \|x_{g_j}\|_2^2)^{\frac{1}{2}}} - \frac{y_i}{(\epsilon^2 + \|y_{g_j}\|_2^2)^{\frac{1}{2}}} \right)^2, \quad i = 1, 2, \dots, N, j = 1, 2, \dots, k,$$

where x_i is the component of x_{g_j} . Observe that

$$l_i = \left\{ \frac{(x_i - y_i)(\epsilon^2 + \|y_{g_j}\|_2^2)^{\frac{1}{2}} + y_i[(\epsilon^2 + \|y_{g_j}\|_2^2)^{\frac{1}{2}} - (\epsilon^2 + \|x_{g_j}\|_2^2)^{\frac{1}{2}}]}{(\epsilon^2 + \|x_{g_j}\|_2^2)^{\frac{1}{2}}(\epsilon^2 + \|y_{g_j}\|_2^2)^{\frac{1}{2}}} \right\}^2. \quad (5.10)$$

Set

$$\Delta_i = \frac{(x_i - y_i)(\epsilon^2 + \|y_{g_j}\|_2^2)^{\frac{1}{2}} + y_i[(\epsilon^2 + \|y_{g_j}\|_2^2)^{\frac{1}{2}} - (\epsilon^2 + \|x_{g_j}\|_2^2)^{\frac{1}{2}}]}{(\epsilon^2 + \|x_{g_j}\|_2^2)^{\frac{1}{2}}(\epsilon^2 + \|y_{g_j}\|_2^2)^{\frac{1}{2}}}.$$

Applying the triangular inequality to above equality, we have

$$\begin{aligned} |\Delta_i| &\leq \frac{|x_i - y_i|}{(\epsilon^2 + \|x_{g_j}\|_2^2)^{\frac{1}{2}}} + \frac{|y_i| \left| (\epsilon^2 + \|y_{g_j}\|_2^2)^{\frac{1}{2}} - (\epsilon^2 + \|x_{g_j}\|_2^2)^{\frac{1}{2}} \right|}{(\epsilon^2 + \|x_{g_j}\|_2^2)^{\frac{1}{2}} (\epsilon^2 + \|y_{g_j}\|_2^2)^{\frac{1}{2}}} \\ &\leq \frac{|x_i - y_i|}{(\epsilon^2 + \|x_{g_j}\|_2^2)^{\frac{1}{2}}} + \frac{|y_i| \left| \|y_{g_j}\|_2^2 - \|x_{g_j}\|_2^2 \right|}{(\epsilon^2 + \|x_{g_j}\|_2^2)^{\frac{1}{2}} (\epsilon^2 + \|y_{g_j}\|_2^2)^{\frac{1}{2}} \left[(\epsilon^2 + \|y_{g_j}\|_2^2)^{\frac{1}{2}} + (\epsilon^2 + \|x_{g_j}\|_2^2)^{\frac{1}{2}} \right]} \end{aligned} \quad (5.11)$$

Note that

$$\begin{aligned} \|y_{g_j}\|_2^2 - \|x_{g_j}\|_2^2 &= (\|y_{g_j}\|_2 + \|x_{g_j}\|_2)(\|y_{g_j}\|_2 - \|x_{g_j}\|_2) \\ &\leq (\|y_{g_j}\|_2 + \|x_{g_j}\|_2) \|x_{g_j} - y_{g_j}\|_2, \end{aligned} \quad (5.12)$$

where for the above equality, we used the inverse triangular inequality. Plugging (5.12) into (5.11), we get

$$\begin{aligned} |\Delta_i| &\leq \frac{|x_i - y_i|}{(\epsilon^2 + \|x_{g_j}\|_2^2)^{\frac{1}{2}}} + \frac{|y_i| (\|y_{g_j}\|_2 + \|x_{g_j}\|_2) \|x_{g_j} - y_{g_j}\|_2}{(\epsilon^2 + \|x_{g_j}\|_2^2)^{\frac{1}{2}} (\epsilon^2 + \|y_{g_j}\|_2^2)^{\frac{1}{2}} \left[(\epsilon^2 + \|y_{g_j}\|_2^2)^{\frac{1}{2}} + (\epsilon^2 + \|x_{g_j}\|_2^2)^{\frac{1}{2}} \right]} \\ &\leq \frac{|x_i - y_i|}{(\epsilon^2 + \|x_{g_j}\|_2^2)^{\frac{1}{2}}} + \frac{\|x_{g_j} - y_{g_j}\|_2}{(\epsilon^2 + \|x_{g_j}\|_2^2)^{\frac{1}{2}}} \\ &\leq \frac{1}{\epsilon} (|x_i - y_i| + \|x_{g_j} - y_{g_j}\|_2). \end{aligned} \quad (5.13)$$

By (5.13), we get

$$\begin{aligned} l_i &\leq \frac{1}{\epsilon^2} (|x_i - y_i| + \|x_{g_j} - y_{g_j}\|_2)^2 \\ &\leq \frac{2}{\epsilon^2} (|x_i - y_i|^2 + \|x_{g_j} - y_{g_j}\|_2^2), \end{aligned} \quad (5.14)$$

where the second inequality follows from the fact that $\|u\|_1 \leq \sqrt{N} \|u\|_2$, for any $u \in \mathbb{R}^N$. Substituting (5.14) into (5.9), we get

$$\begin{aligned} &\|\nabla \|x\|_{2,\mathcal{I}}^\epsilon - \nabla \|y\|_{2,\mathcal{I}}^\epsilon\|_2^2 \\ &\leq \frac{2}{\epsilon^2} \sum_{i=1}^{d_1} |x_i - y_i|^2 + \frac{2}{\epsilon^2} \sum_{i=d_1+1}^{d_2} |x_i - y_i|^2 + \cdots + \frac{2}{\epsilon^2} \sum_{i=N-d_k+1}^N |x_i - y_i|^2 + \frac{2}{\epsilon^2} \sum_{j=1}^k d_j \|x_{g_j} - y_{g_j}\|_2^2 \\ &\leq \frac{2}{\epsilon^2} \left(\sum_{i=1}^N |x_i - y_i|^2 + d_{\max} \sum_{j=1}^k \|x_{g_j} - y_{g_j}\|_2^2 \right) \\ &\leq \frac{2}{\epsilon^2} (1 + d_{\max}) \|x - y\|_2^2 \\ &=: \mathcal{C}^2 \|x - y\|_2^2, \end{aligned} \quad (5.15)$$

where $d_{\max} = \max\{d_1, d_2, \dots, d_k\}$ indicates the maximum of the group size $\{d_1, d_2, \dots, d_k\}$. The equation (5.15) shows that the gradient of $\|x\|_{2,\mathcal{I}}^\epsilon$ is \mathcal{C} -Lipshitz continuous.

□

Lemma 5.2. Set $\tilde{L}_{p,\epsilon}(x, u, y, \hat{x}) = L_{p,\epsilon}(x, u, y) + a_1\|x - \hat{x}\|_2^2$ ($a_1 > 0$). Assume that $\epsilon > 0$, $\rho_2 > \frac{\mathcal{C}}{2}$ and for $\vartheta > 0$, $\Phi\Phi^\top \succeq \vartheta I_n$. If constant α obeys the following inequality

$$\alpha > \frac{2\rho_2^2 + 2(\rho_2 + \mathcal{C})^2}{\vartheta(\rho_2 - \frac{\mathcal{C}}{2})},$$

then

$$\tilde{L}_{p,\epsilon}(x^{t+1}, u^{t+1}, y^{t+1}, x^t) \leq \tilde{L}_{p,\epsilon}(x^t, u^t, y^t, x^{t-1}) - a_2\|x^{t+1} - x^t\|_2^2,$$

where

$$a_1 = \frac{2(\rho_2 + \mathcal{C})^2}{\alpha\vartheta}, \quad a_2 = \rho_2 - \frac{2\rho_2^2 + 2(\rho_2 + \mathcal{C})^2}{\vartheta\alpha} - \frac{\mathcal{C}}{2} > 0.$$

Proof.

By (4.15), solving minimizer x^{t+1} of (4.13) is equivalent to solving the minimizer of the function as follows:

$$(\nabla\|x^t\|_{2,\mathcal{I}}^\epsilon)^\top(x - x^t) + \frac{\rho_2}{2}\|x - x^t\|_2^2 + \frac{\alpha}{2}\|\Phi x - b - u^{t+1} - \frac{y^t}{\alpha}\|_2^2 =: S_{x^t}(x). \quad (5.16)$$

It follows from the above equation, we have

$$S_{x^t}(x^t) = \frac{\alpha}{2}\|\Phi x^t - b - u^{t+1} - \frac{y^t}{\alpha}\|_2^2. \quad (5.17)$$

According to the minimality of x^{t+1} , we get

$$\nabla S_{x^t}(x^{t+1}) = 0. \quad (5.18)$$

By ρ_2 -strongly convexity of $S_{x^t}(x)$, we get

$$S_{x^t}(x^t) \geq S_{x^t}(x^{t+1}) + (\nabla S_{x^t}(x^{t+1}))^\top(x^t - x^{t+1}) + \frac{\rho_2}{2}\|x^t - x^{t+1}\|_2^2, \quad (5.19)$$

for any $x^t \in \mathbb{R}^N$. Combining with (5.16)-(5.19), we get

$$\begin{aligned} & (\nabla\|x^t\|_{2,\mathcal{I}}^\epsilon)^\top(x^{t+1} - x^t) + \frac{\alpha}{2}\|\Phi x^{t+1} - b - u^{t+1} - \frac{y^t}{\alpha}\|_2^2 \\ & \leq \frac{\alpha}{2}\|\Phi x^t - b - u^{t+1} - \frac{y^t}{\alpha}\|_2^2 - \rho_2\|x^t - x^{t+1}\|_2^2. \end{aligned} \quad (5.20)$$

By Lemma 5.1 and the fact that $\|x\|_{2,\mathcal{I}}^\epsilon$ is convex, we get

$$\|x^{t+1}\|_{2,\mathcal{I}}^\epsilon \leq \|x^t\|_{2,\mathcal{I}}^\epsilon + (\nabla\|x^t\|_{2,\mathcal{I}}^\epsilon)^\top(x^{t+1} - x^t) + \frac{\mathcal{C}}{2}\|x^{t+1} - x^t\|_2^2, \quad (5.21)$$

for any $x^t, x^{t+1} \in \mathbb{R}^N$. A combination of (5.20) and (5.21), we have

$$\begin{aligned} & \|x^{t+1}\|_{2,\mathcal{I}}^\epsilon + \frac{\alpha}{2}\|\Phi x^{t+1} - b - u^{t+1} - \frac{y^t}{\alpha}\|_2^2 \\ & \leq \|x^t\|_{2,\mathcal{I}}^\epsilon + \frac{\alpha}{2}\|\Phi x^t - b - u^{t+1} - \frac{y^t}{\alpha}\|_2^2 + \left(\frac{\mathcal{C}}{2} - \rho_2\right)\|x^{t+1} - x^t\|_2^2. \end{aligned} \quad (5.22)$$

Since $\Phi\Phi^\top \succeq \vartheta I_n$ for $\vartheta > 0$, we get

$$\vartheta \|y^{t+1} - y^t\|_2^2 \leq \|\Phi^\top(y^{t+1} - y^t)\|_2^2. \quad (5.23)$$

By (5.18), we have

$$\nabla\|x^t\|_{2,\mathcal{I}}^\epsilon + \rho_2(x^{t+1} - x^t) + \alpha\Phi^\top(\Phi x^{t+1} - b - u^{t+1} - \frac{y^t}{\alpha}) = 0. \quad (5.24)$$

Substituting (4.5) into the above equality, we get

$$\Phi^\top y^{t+1} = \nabla\|x^t\|_{2,\mathcal{I}}^\epsilon + \rho_2(x^{t+1} - x^t). \quad (5.25)$$

By (5.25) and Lemma 5.1, we get

$$\begin{aligned} & \|\Phi^\top(y^{t+1} - y^t)\|_2^2 \\ & \leq \left(\|\nabla\|x^t\|_{2,\mathcal{I}}^\epsilon - \nabla\|x^{t-1}\|_{2,\mathcal{I}}^\epsilon \|_2 + \rho_2\|x^{t+1} - x^t\|_2 + \rho_2\|x^t - x^{t-1}\|_2 \right)^2 \\ & \leq \left(\mathcal{C}\|x^t - x^{t-1}\|_2 + \rho_2\|x^{t+1} - x^t\|_2 + \rho_2\|x^t - x^{t-1}\|_2 \right)^2 \\ & \leq 2(\mathcal{C} + \rho_2)^2\|x^t - x^{t-1}\|_2^2 + 2\rho_2^2\|x^{t+1} - x^t\|_2^2. \end{aligned} \quad (5.26)$$

Combining with (5.23) and (5.26), we have

$$\|y^{t+1} - y^t\|_2^2 \leq \frac{2(\mathcal{C} + \rho_2)^2}{\vartheta}\|x^t - x^{t-1}\|_2^2 + \frac{2\rho_2^2}{\vartheta}\|x^{t+1} - x^t\|_2^2. \quad (5.27)$$

By the definition of the minimizer u^{t+1} , we get

$$L_{p,\epsilon}(x^t, u^{t+1}, y^t) - L_{p,\epsilon}(x^t, u^t, y^t) \leq 0. \quad (5.28)$$

By (5.22), we get

$$L_{p,\epsilon}(x^{t+1}, u^{t+1}, y^t) - L_{p,\epsilon}(x^t, u^{t+1}, y^t) \leq \left(\frac{\mathcal{C}}{2} - \rho_2 \right) \|x^{t+1} - x^t\|_2^2. \quad (5.29)$$

By the definition of $L_{p,\epsilon}(x, u, y)$ and (4.5), we get

$$L_{p,\epsilon}(x^{t+1}, u^{t+1}, y^{t+1}) - L_{p,\epsilon}(x^{t+1}, u^{t+1}, y^t) \leq \frac{1}{\alpha}\|y^{t+1} - y^t\|_2^2. \quad (5.30)$$

Combining with (5.28)-(5.30) and (5.27), it follows that

$$\begin{aligned} & L_{p,\epsilon}(x^{t+1}, u^{t+1}, y^{t+1}) - L_{p,\epsilon}(x^t, u^t, y^t) \\ & \leq \left(\frac{2\rho_2^2}{\alpha\vartheta} + \frac{\mathcal{C}}{2} - \rho_2 \right) \|x^{t+1} - x^t\|_2^2 + \frac{2(\mathcal{C} + \rho_2)^2}{\vartheta\alpha}\|x^t - x^{t-1}\|_2^2. \end{aligned} \quad (5.31)$$

By (5.31) and some elementary manipulation, the desired result will be obtained. In order to ensure constant $a_2 > 0$, the regularization parameter α needs to satisfy the following equation

$$\rho_2 - \frac{\mathcal{C}}{2} > \frac{2\rho_2^2 + 2(\rho_2 + \mathcal{C})^2}{\vartheta\alpha}. \quad (5.32)$$

□

Lemma 5.3. Set $\tilde{z}^t = (u^t, x^t, y^t)$. Under the conditions of Lemma 5.2, we gain

$$\lim_{t \rightarrow \infty} \|\tilde{z}^{t+1} - \tilde{z}^t\|_2^2 = 0,$$

and any cluster point of \tilde{z}^t is one critical point of $L_{p,\epsilon}$.

Proof.

Due to (5.25) and the fact that $\|\nabla \|x\|_{2,\mathcal{I}}^\epsilon\|_2^2 \leq N$, for any $x \in \mathbb{R}^N$, we get

$$\begin{aligned} & \|\Phi^\top y^t\|_2^2 \\ & \leq (\|\nabla \|x^{t-1}\|_{2,\mathcal{I}}^\epsilon\|_2 + \rho_2 \|x^t - x^{t-1}\|_2)^2 \\ & \leq 2\|\nabla \|x^{t-1}\|_{2,\mathcal{I}}^\epsilon\|_2^2 + 2\rho_2^2 \|x^t - x^{t-1}\|_2^2 \\ & \leq 2N + 2\rho_2^2 \|x^t - x^{t-1}\|_2^2. \end{aligned} \quad (5.33)$$

Similar to (5.23), we get

$$\|\Phi^\top y^t\|_2^2 \geq \vartheta \|y^t\|_2^2. \quad (5.34)$$

By (5.33) and (5.34), we get

$$\|y^t\|_2^2 \leq \frac{2N}{\vartheta} + \frac{2\rho_2^2}{\vartheta} \|x^t - x^{t-1}\|_2^2. \quad (5.35)$$

Set $z^t = (u^t, x^t, y^t, x^{t-1})$. It follows from Lemma 5.2 and (5.35) that

$$\begin{aligned} & \tilde{L}_{p,\epsilon}(z^1) \geq \tilde{L}_{p,\epsilon}(z^t) \\ & = \|x^t\|_{2,\mathcal{I}}^\epsilon + \frac{1}{\nu} \|u^t\|_p^p + \frac{\alpha}{2} \|\Phi x^t - b - u^t - \frac{y^t}{\alpha}\|_2^2 - \frac{1}{2\alpha} \|y^t\|_2^2 + a_1 \|x^t - x^{t-1}\|_2^2 \\ & \geq \|x^t\|_{2,\mathcal{I}}^\epsilon + \frac{1}{\nu} \|u^t\|_p^p + \frac{\alpha}{2} \|\Phi x^t - b - u^t - \frac{y^t}{\alpha}\|_2^2 - \frac{N}{\alpha\vartheta} - \left(\frac{\rho_2^2}{\alpha\vartheta} - \frac{2(\rho_2 + C)^2}{\alpha\vartheta} \right) \|x^t - x^{t-1}\|_2^2. \end{aligned} \quad (5.36)$$

Furthermore, (5.36) shows that $\tilde{L}_{p,\epsilon}(z^t)$ is bounded. Notice that $\|x^t\|_{2,\mathcal{I}}^\epsilon$ and $\|u^t\|_p^p$ are coercive and using (5.35), x^t , u^t and y^t are bounded. Hence, z^t is bounded.

According to the boundedness of z^t , there is a convergent subsequence z^{t_j} tending to some cluster point z^* . By Lemma 5.2, we get

$$\begin{aligned} & a_2 \sum_{t=1}^m \|x^{t+1} - x^t\|_2^2 \\ & \leq \sum_{t=1}^m (\tilde{L}_{p,\epsilon}(z^t) - \tilde{L}_{p,\epsilon}(z^{t+1})) \\ & \leq \tilde{L}_{p,\epsilon}(z^1) - \tilde{L}_{p,\epsilon}(z^{m+1}) \\ & \leq \tilde{L}_{p,\epsilon}(z^1) - \tilde{L}_{p,\epsilon}(z^*) < \infty, \end{aligned} \quad (5.37)$$

where the third inequality follows from the facts that $\tilde{L}_{p,\epsilon}(z^t)$ is convergent and for any $t > 0$, $\tilde{L}_{p,\epsilon}(z^t) \geq \tilde{L}_{p,\epsilon}(z^*)$. When $m \rightarrow \infty$, we get

$$\sum_{t=1}^{\infty} \|x^{t+1} - x^t\|_2^2 < \infty. \quad (5.38)$$

By (5.27) and (5.38), we get

$$\sum_{t=1}^{\infty} \|y^{t+1} - y^t\|_2^2 < \infty. \quad (5.39)$$

By (4.5), we get

$$\|u^{t+1} - u^t\|_2 \leq \frac{1}{\alpha} \|y^{t+1} - y^t\|_2 + \frac{1}{\alpha} \|y^t - y^{t-1}\|_2 + \|\Phi\|_2 \|x^{t+1} - x^t\|_2. \quad (5.40)$$

By (5.38) and (5.39), we get

$$\sum_{t=1}^{\infty} \|u^{t+1} - u^t\|_2 < \infty. \quad (5.41)$$

Therefore, we derive

$$\sum_{t=1}^{\infty} \|\tilde{z}^{t+1} - \tilde{z}^t\|_2^2 < \infty \text{ and } \|\tilde{z}^{t+1} - \tilde{z}^t\|_2^2 \rightarrow 0, \text{ as } t \rightarrow \infty.$$

Now, we prove that any cluster point of sequence $\{\tilde{z}^t\}$ is a critical point of $L_{p,\epsilon}$. By the optimality theory and (4.5), we get

$$\begin{aligned} 0 &\in \partial \|u^{t+1}\|_p^p + \alpha\nu\Phi(x^{t+1} - x^t) + \nu y^{t+1}, \\ 0 &= \nabla \|x^t\|_{2,\mathcal{I}}^\epsilon + \rho_2(x^{t+1} - x^t) - \Phi^\top y^{t+1}, \\ \frac{y^{t+1}}{\alpha} &= \frac{y^t}{\alpha} - (\Phi x^{t+1} - b - u^{t+1}). \end{aligned} \quad (5.42)$$

Due to $\lim_{t \rightarrow \infty} \|\tilde{z}^{t+1} - \tilde{z}^t\|_2^2 = 0$, for a convergent subsequence \tilde{z}^{t_i} , both \tilde{z}^{t_i} and \tilde{z}^{t_i+1} converge to the point $z^* := (u^*, x^*, y^*)$. Taking the limit in (5.42) along the subsequence \tilde{z}^{t_i} results in

$$0 \in \partial \|u^*\|_p^p + \nu y^*, \quad \Phi^\top y^* = \nabla \|x^*\|_{2,\mathcal{I}}^\epsilon, \quad \text{and } \Phi x^* - u^* = b.$$

Consequently, z^* is one critical point of $L_{p,\epsilon}$.

□

Proof of Theorem 5.2:

The proof of Theorem 5.2 includes two steps as follows:

(i) There is a positive constant a_3 that satisfies

$$\text{dist}(\partial \tilde{L}_{p,\epsilon}(z^{t+1}), 0) \leq a_3 (\|x^{t+1} - x^t\|_2 + \|x^t - x^{t-1}\|_2 + \|x^{t-1} - x^{t-2}\|_2);$$

By Lemma 5.3, it follows that $\lim_{t \rightarrow \infty} \text{dist}(\partial \tilde{L}_{p,\epsilon}(z^{t+1}), 0) = 0$.

(ii) Set $\tilde{z}^t = (u^t, x^t, y^t)$. The sequence $\{\tilde{z}^t\}$ satisfies the following equation

$$\sum_{t=0}^{\infty} \|\tilde{z}^{t+1} - \tilde{z}^t\|_2 < \infty,$$

that is, its length is finite; It leads to the sequence $\{\tilde{z}^t\}$ is a Cauchy sequence; Accordingly, it is convergent.

Based on the above lemmas, the proofs of (i) and (ii) follow from the proof of Theorem III.3 and Theorem III.4 in [32]. Combining with (i), (ii) and Lemma 5.3, the desired result follows.

□

6 Numerical simulations

In this section, we carry out several numerical experiments to show the robustness of new method. Two kinds of signals are used as the test signals, which incorporate the synthetic block sparse signals and the real-world FECG signals (which can be regarded as approximately block sparse signals).

6.1 Experiments on synthetic signals

In our experiments, without loss of generality, we discuss the block sparse signal with even block size, i.e. $d_1 = d_2 = \dots = d_k = d$ and set the signal length $N = 256$. For each trial, we firstly randomly produce block sparse signal x with coefficients following a Gaussian distribution of mean 0 and variance 1, and randomly produce a 100×256 measurement matrix Φ from Gaussian ensemble. Employing x and Φ , we generate the measurements b by means of $b = \Phi x + z$, where z is (impulsive) bit errors like noise / Laplace noise / generalized Gaussian noise. In each experiment, the average results over independent 100 trials are reported.

To look for the better regularization parameter ν that derives the minimal recovery error, we conduct a set of trails. In the set of trails, we produce the signals with 10 nonzero blocks by choosing 128 blocks uniformly at random, i.e. $d = 2$. In Figure 6.2, the average normalized reconstruction error (RelError, $\text{RelError} = \|x^* - x\|_2 / \|x\|_2$) is plotted versus the regularization parameter ν for different p values, $p = 0.5, 0.8, 1$ in (impulsive) bit errors like noise and Laplace noise, respectively and the figure indicates that the parameter $\nu = 1 \times 10^4$ is an appropriate choice. Figure 6.3(a) the signal-to-noise ratio (SNR, $\text{SNR} = 20 \log_{10}(\|x\|_2 / \|x^* - x\|_2)$) is plotted versus the values p in four different impulsive noises and the values of p range from 0.2 to 2. Figure 6.3(a) shows that when $p = 0.8$, SNR is highest, so we choose $p = 0.8$ to conduct several simulation experiments for testing recovery performance of Block- L_p -ADM. Figure 6.3 (b) shows simulation results concerning the performance of the non-block algorithm and the block algorithm in several different impulsive noises, where $p = 0.8$. Two curves of SNR are described via L_p -ADM [12] and Block- L_p -ADM. Figure 6.3(b) demonstrates the signal structure is very significant in the signal recovery. In Figure 6.4, the number of non-zero components of k_0 ranges from 12 to 72. Figure 6.4 reveals that the performance of Block- L_p -ADM is better than that of L_p -ADM.

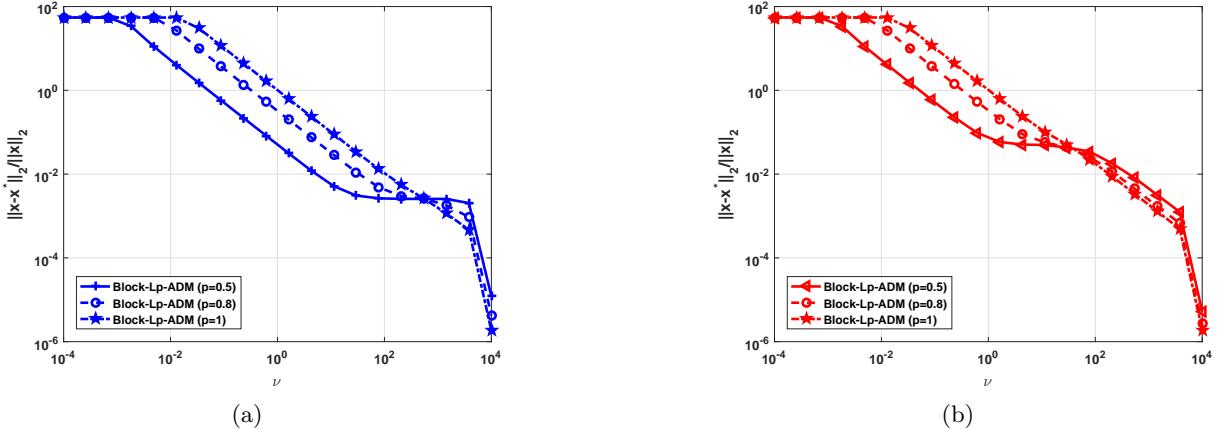


Fig. 6.2: Recovery performance of Block- L_p -ADM versus ν for the block size $d = 2$, (a) (impulsive) bit errors like noise case, (b) Laplace noise case

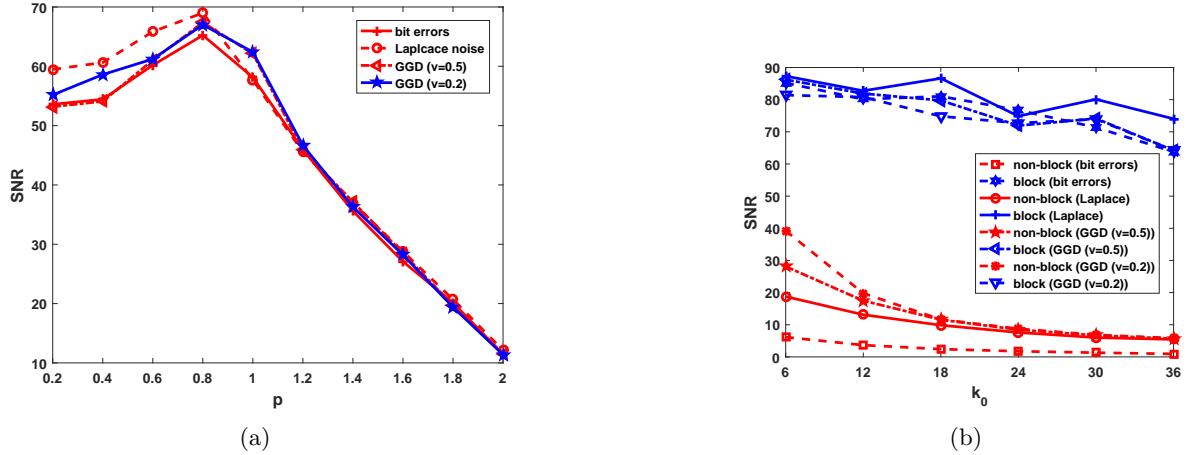


Fig. 6.3: (a) Recovery performance of Block- L_p -ADM versus the value of p for the block size $d = 2$, (b) Recovery performance of L_p -ADM and Block- L_p -ADM with $p = 0.8$

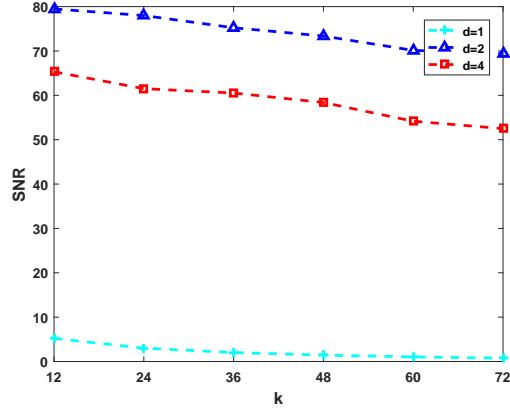


Fig. 6.4: SNR versus the number of non-zero components k with $p = 0.8$

Finally, we compared the performance of our Block- L_p -ADM algorithm for $p = 0.5, 0.8, 1$ with the other representative algorithms including Group Lasso algorithm (Group-Lasso) [43], Huber-fast iterative shrinkage/thresholding algorithm (Huber-FISTA) [44], L_q -regularized algorithm (L_q -min) [45], BP-SEP [46] and orthogonal greedy algorithm (OGA) [47]. We utilize the relative error (RelErr) to measure the algorithm capability. Figure 6.5 presents the relative error versus the sparsity k_0 . Observe that the performance of Block- L_p -ADM algorithm is much better than that of the other algorithms.

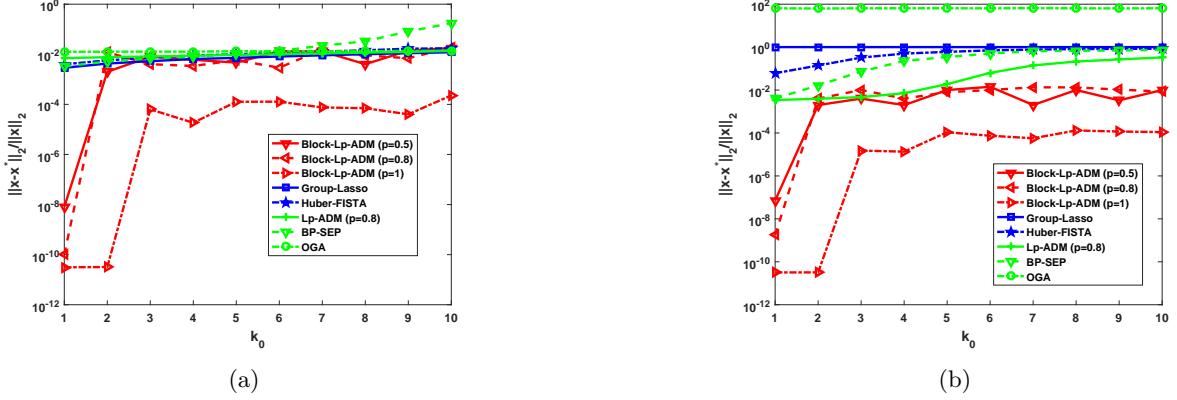


Fig. 6.5: Comparison of execution efficiency with respect to RelErr (a) Laplace noise and (b) (impulsive) bit errors like noise

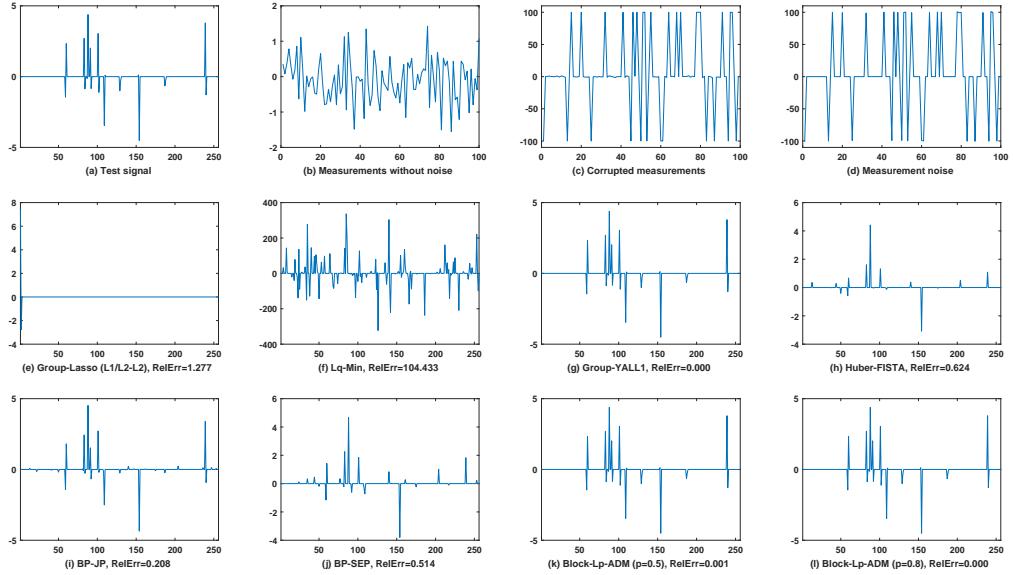


Fig. 6.6: Recovery performance of the compared algorithms in the presence of bit errors like corruption. (a) Test signal. (b) Measurements without noise. (c) Corrupted measurements. (d) Measurement noise. (e) Group-Lasso, RelErr= 1.429. (f) L_q -min, RelErr= 101.332. (g) Group-YALL1, RelErr= 0.001. (h) Huber-FISTA, RelErr= 0.739. (i) BP-JP, RelErr= 0.193. (j) BP-SEP, RelErr= 0.601. (k) block- L_p -ADM ($p = 0.5$), RelErr= 0.001. (l) block- L_p -ADM ($p = 0.8$), RelErr= 0.169.

6.2 Experiments on FECG signals

In order to further validate the recovery performance of our Block- L_p -ADM algorithm in some practical applications, we employ our proposed method, together with the methods we exploited to reconstruct the FECG signals [48]. Actually, compressed sensing and application communities have studied this sort of signals, see [49] and therein literature. Figure 6.7 displays a segment of such FECG signals. In this segment, we can regard the sections from 20 to 60 and from 200 to 250 time points as two major non-zero blocks, and other sections can be regarded as cascades of

Table 6.1: RelError results obtained by different methods

| Algorithm | a | b | c | d | e | f | g | h |
|-------------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| Block- L_p -ADM | 0.0013 | 0.0007 | 0.0010 | 0.0015 | 0.0008 | 0.0001 | 0.0001 | 0.0001 |
| L_p -ADM | 0.0660 | 0.0036 | 0.0595 | 0.1504 | 0.0042 | 0.0005 | 0.0059 | 0.0004 |
| Group-Lasso | 0.8927 | 0.8557 | 0.9242 | 0.9969 | 0.8140 | 0.7955 | 0.7504 | 0.7414 |

zero blocks. Approximately, we can regard this segment as a block 2-sparse signal. In general, since we beforehand don't know the position of non-zero coefficients in FECG, it is hard for us to utilize the block-structured ways diametrically. Therefore, analogy to [49], suppose that this segment comprises same 10 blocks with block size $d = 25$. In addition, similar with [49], the same matrix is used as the sensing matrix. In order to facilitate the optimal performance of these methods, their regularization parameters are selected from $\{10^{-4}, 10^{-3}, \dots, 10^8\}$, and return their best reconstructed signals which are determined by RelError as the final results. Figure 6.8 demonstrates the results which are derived by Block- L_p -ADM method with $p = 0.8$, L_p -ADM method with $p = 0.8$, group-lasso method and OGA method. It is easy to observe that both our Block- L_p -ADM method with $p = 0.8$ and L_p -ADM method with $p = 0.8$ perform much better than other methods, and the recovered segments are very approximate to the original segment. But, from the opinion of recovered relative error, our Block- L_p -ADM method with $p = 0.8$ expresses somewhat better than L_p -ADM method with $p = 0.8$. Then, to testify the performance of our Block- L_p -ADM method, the identical sensing matrix is utilized to compress all FECG signals which are given in Figure 6.9. Note that the dimension of each FECG signal is 2500, so we first equally partition each of them into 10 segments and then reconstruct those segments sequentially. Table 6.1 shows the derived results. It is easy to see that the recovering efficiency of our method is best, followed by L_p -ADM method and Group-Lasso sequentially. These results again verify the ability of our method.

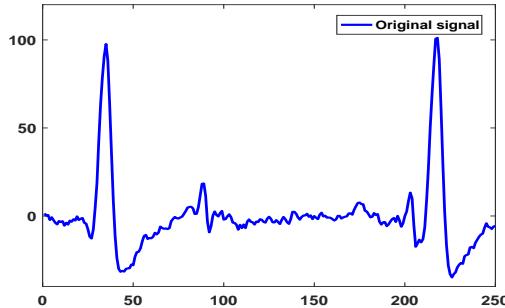


Fig. 6.7: Segment from FECG signals

7 Conclusions

This paper investigates the problem of the block sparse reconstruction which is corrupted by impulsive noise. We have put forward a robust model for block signal recovery, which exploited the generalized l_p -norm ($0 \leq p < 2$) to measure the residual error. For the model, we have proposed an efficient algorithm to solve it, comprising the approximate operator for l_p -norm functions into the frame of augmented Lagrangian methods. Based on block-RIP, we have provided a sharp sufficient condition and the error upper bound estimation of recovering block-sparse signals in the

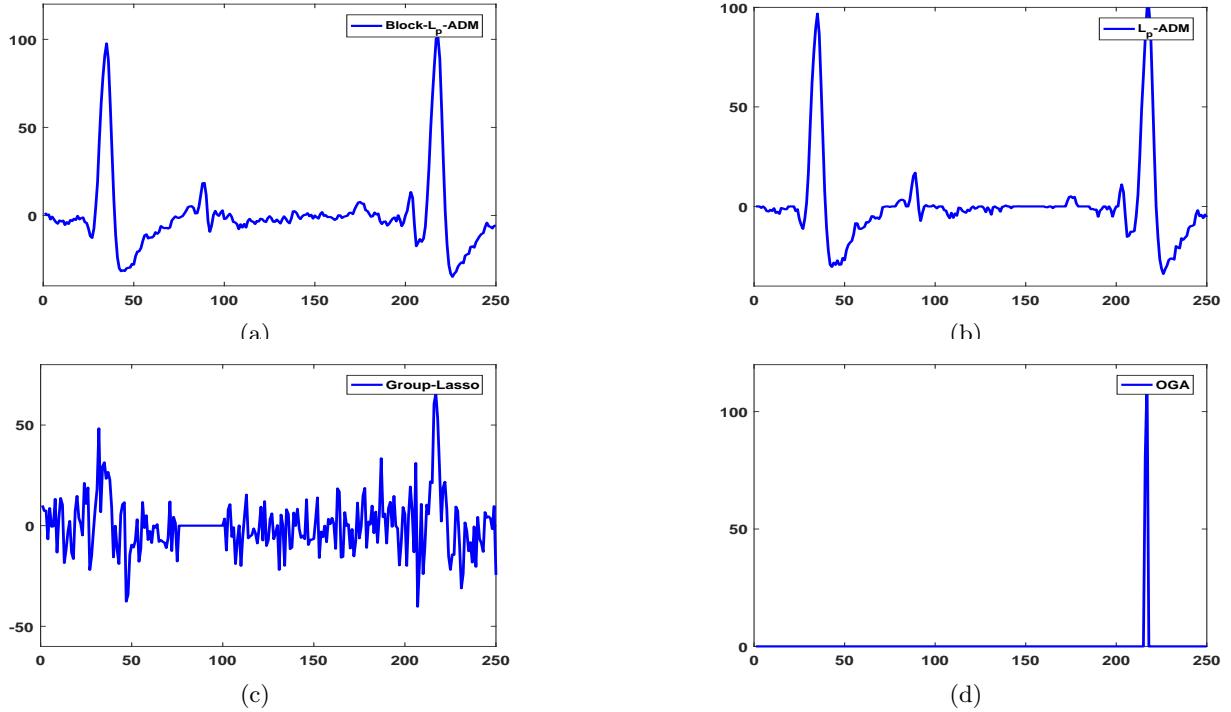


Fig. 6.8: Reconstructed results by different methods

(a) Block- L_p -ADM method with RelError=0.0005, (b) L_p -ADM method with RelError=0.0031, (c) Group-Lasso method with RelError=0.7791, (d) OGA method with RelError= 0.9257

presence of impulsive noise. Furthermore, the convergence condition of new algorithm for both the nonconvex ($0 \leq p < 1$) and convex ($1 \leq p < 2$) cases has been analyzed. Simulation experiments that are based on the synthetic block-sparse signals and the real-world FECG signals manifested that when observation measurement is disturbed by impulsive noise, the better performance of the Block- L_p -ADM algorithm is expressed by comparing with other well-known algorithms.

Acknowledgments

This work was supported by Natural Science Foundation of China (No. 61673015, 61273020, 11661084), Fundamental Research Funds for the Central Universities (No. XDK2015A007, SWU1809002), Science Computing and Intelligent Information Processing of GuangXi higher education key laboratory (No. GXSCIIIP201702), Science and Technology Plan Project of Guizhou Province(No. LH[2015]7053, LH[2015]7055), Science and technology Foundation of Guizhou province (Qian ke he Ji Chu [2016]1161), Guizhou province natural science foundation in China (Qian Jiao He KY [2016]255).

References

- [1] Figueiredo M A T, Nowak R D, Wright S J. Gradient Projection for Sparse Reconstruction: Application to Compressed Sensing and Other Inverse Problems. IEEE Journal of Selected Topics in Signal Processing, 2008, 1(4):586-597.
- [2] Bruckstein A M, Donoho D L, Elad M. From sparse solutions of systems of equations to sparse modeling of signals and images. SIAM Rev 2009; 51(1): 34-81.

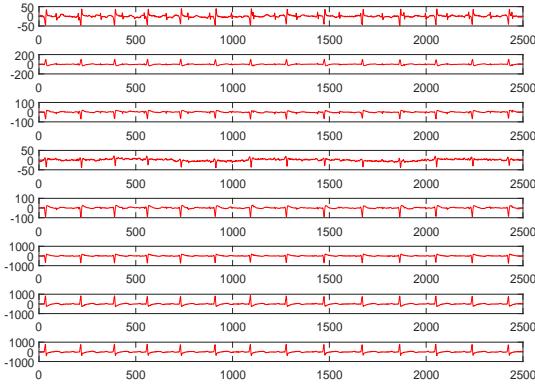


Fig. 6.9: All FECG signals

- [3] Zhu H, Xiao Y, Wu S Y. Large sparse signal recovery by conjugate gradient algorithm based on smoothing technique. *Computers & Mathematics with Applications*, 2013, 66(1):24-32.
- [4] Candès, E, Tao, T: Near-optimal signal recovery from random projections: universal encoding strategies. *IEEE Trans. Inf. Theory* 52(12), 5406-5425 (2006)
- [5] Candès, E: The restricted isometry property and its implications for compressed sensing. *C. R. Math. Acad. Sci. Paris, Sér. I* 346, 589-592 (2008)
- [6] E. J. Candès and P. A. Randall, Highly robust error correction by convex programming, *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 2829-2840, Jul. 2008.
- [7] B. Popilka, S. Setzer, and G. Steidl, Signal recovery from incomplete measurements in the presence of outliers, *Inverse Problems Imag.*, vol. 1, no. 4, pp. 661-672, Nov. 2007.
- [8] R. Chan, C. W. Ho, and M. Nikolova, Salt-and-pepper noise removal by median-type noise detectors and de tail-preserving regularization, *IEEE Trans. Image Process.*, vol. 14, no. 10, pp. 1479-1485, Oct. 2005.
- [9] T. Hashimoto, Bounds on a probability for the heavy tailed distribution and the probability of deficient decoding in sequential decoding, *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 990-1002, Mar. 2005.
- [10] J. F. Yang and Y. Zhang, Alternating direction algorithms for l_1 -problems in compressive sensing, *SIAM J. Sci. Comput.*, vol. 33, pp. 250-278, 2011.
- [11] Y. Xiao, H. Zhu, and S. Y. Wu, Primal and dual alternating direction algorithms for l_1-l_1 -norm minimization problems in compressive sensing, *Comput. Optim. Appl.*, vol. 54, no. 2, pp. 441-459, 2013.
- [12] Wen, F., Liu, P., Liu, Y., Qiu, R. C., Yu, W. (2017). Robust sparse recovery in impulsive noise via l_p - l_1 optimization. *IEEE Transactions on Signal Processing*, 65(1), 105-118.
- [13] Kim S, Xing E P. Tree-guided group lasso for multi-response regression with structured sparsity, with an application to eQTL mapping. *Annals of Applied Statistics*, 2012, 6(3):1095-1117.
- [14] Noah Simon, Jerome Friedman, Trevor Hastie, et al. A Sparse-Group Lasso. *Journal of Computational & Graphical Statistics*, 2013, 22(2):231-245.
- [15] Zhou J, Liu J, Narayan V A, et al. Modeling disease progression via fused sparse group lasso[C]// ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2012:1095-1103.

- [16] Zhu X, Huang Z, Cui J, et al. Video-to-Shot Tag Propagation by Graph Sparse Group Lasso. *IEEE Transactions on Multimedia*, 2013, 15(3):633-646.
- [17] Yan L, Li W J, Xue G R, et al. Coupled group lasso for web-scale CTR prediction in display advertising[C]// International Conference on Machine Learning. 2014:802-810.
- [18] Rao N, Nowak R, Cox C, et al. Classification With the Sparse Group Lasso. *IEEE Transactions on Signal Processing*, 2015, 64(2):448-463.
- [19] Deng, W., Yin, W. Zhang, Y. (2011) Group Sparse Optimization by Alternating Direction Method. Rice CAAM Report TR11-06.
- [20] J.H. Lin, S. Li, Block sparse recovery via mixed l_2/l_1 minimization, *Acta Math. Sin. Engl. Ser.* 29(7) (2013) 1401-1412.
- [21] Y. Wang, J.J. Wang, Z.B. Xu, On recovery of block-sparse signals via mixed l_2/l_p ($0 < p \leq 1$) norm minimization, *EURASIP J. Adv. Sig. Pr.* 2013:76 (2013).
- [22] Y. Wang, J.J. Wang, Z.B. Xu, Restricted p -isometry properties of nonconvex block-sparse compressed sensing, *Signal Process.* 104 (2014) 188-196.
- [23] Wang, W.D., Wang, J.J., Zhang, Z.L.: ‘Block-sparse signal recovery via l_2/l_{1-2} minimisation method’, *IET Signal Processing*, 2017, doi: 10.1049/iet-spr.2016.0381
- [24] Y. Gao, M.D. Ma, A new bound on the block restricted isometry constant in compressed sensing, *J. Inequal. Appl.* 2017:174 (2017).
- [25] Liu C Y, Wang J J, Wang W D, Wang Z. Nonconvex block-sparse compressed sensing with redundant dictionaries. *IET Signal Processing*, 2017, 11(2): 171-180.
- [26] Gao Y, Peng J, Yue S. Stability and robustness of the l_2/l_p -minimization for block sparse recovery. *Signal Processing*, 2017, 137: 287-297.
- [27] Candes E J, Tao T. Decoding by linear programming. *IEEE Transactions on Information Theory*, 2005, 51(12):4203-4215.
- [28] A. Beck and M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183-202, 2009.
- [29] G. Marjanovic and V. Solo, On l_q optimization and matrix completion, *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 5714-5724, Nov. 2012.
- [30] P. L. Combettes and J. C. Pesquet, Proximal splitting methods in signal processing, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*. New York, NY, USA: Springer, 2011, pp. 185-212.
- [31] F. Wen, P. Liu. Y. Liu, R. C. Qiu, and W. Yu, Robust sparse recovery for compressive sensing in impulsive noise using L_p -norm model fitting, in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., 2016, pp. 4643- 4647.
- [32] F. Wang, Z. Xu, and H. K. Xu, Convergence of Bregman alternating direction method with multipliers for nonconvex composite problems, arXiv preprint, arXiv:1410.8625, Dec. 2014.
- [33] Nolan, John P. Stable Distributions - Models for Heavy Tailed Data. Retrieved 2009-02-21.
- [34] Penson, K. A., Górska, K. (2010-11-17). Exact and Explicit Probability Densities for One-Sided Lévy Stable Distributions. *Physical Review Letters*. 105 (21): 210604.
- [35] Peng Zuoxiang, Tong Bin, Saralees Nadarajah. Tail Behavior of the General Error Distribution. *Communications in Statistics*, 2009, 38(11):1884-1892.

- [36] Peng, Z., Nadarajah, S., Lin, F. (2010). Convergence rate of extremes for the general error distribution. *Journal of Applied Probability*, 47(3), 668-679.
- [37] Chen, S., Wang, C., Zhang, G. (2012). Rates of convergence of extreme for general error distribution under power normalization. *Statistics & Probability Letters*, 82(2), 385-395.
- [38] Jia, P., Liao, X., Peng, Z. (2015). Asymptotic expansions of the moments of extremes from general error distribution. *Journal of Mathematical Analysis & Applications*, 422(2), 1131-1145.
- [39] Cao X, Zhao Q, Meng D, et al. Robust Low-rank Matrix Factorization under General Mixture Noise Distributions. *IEEE Transactions on Image Processing*, 2016, 25(10):4677-4690.
- [40] Y Eldar, M Mishali, Robust recovery of signals from a structured union of subspaces. *IEEE Trans. Inf. Theory*. 55(11), 5302-5316 (2009)
- [41] Chen W, Li Y. The high order block RIP condition for signal recovery. arXiv preprint, arXiv:1610.06294, Oct. 2016.
- [42] Huang, Jianwen, Wang, Jianjun, Wang, Wendong, et al. A sharp sufficient condition of block signal recovery via l_2/l_1 -minimization. arXiv:1710.11125v1, Oct, 2017.
- [43] Boyd S, Parikh N, Chu E, et al. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Foundations & Trends in Machine Learning*, 2011, 3(1):1-122.
- [44] D. S. Pham and S. Venkatesh, Efficient algorithms for robust recovery of images from compressed data, *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 4724-4737, Dec. 2013.
- [45] S. Foucart and M. J. Lai, Sparsest solutions of underdetermined linear systems via l_q minimization for $0 < q \leq 1$, *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 395-407, May 2009.
- [46] C. Studer and R. G. Baraniuk, Stable restoration and separation of approximately sparse signals, *Appl. Comput. Harmon. Anal.*, vol. 37, no. 1, pp. 12-35, 2014.
- [47] Petukhov A. Fast implementation of orthogonal greedy algorithm for tight wavelet frames. *Signal Processing*, 2006, 86(3):471-479.
- [48] Moor, B.D.: DaISy: database for the identification of systems. Available at <http://www.esat.kuleuven.ac.be/sista/daisy>, November 2011
- [49] Zhang, Z., Jung, T.P., Makeig, S., et al.: Compressed sensing for energyefficient wireless telemonitoring of noninvasive fetal ECG via block sparse Bayesian learning, *IEEE Trans. Biomed. Eng.*, 2013, 60, (2), pp. 300-309