ℓ_p-ADMM Algorithm for Sparse Image Recovery Under Impulsive Noise

Dongbin Hao, Chunjie Zhang(✉), and Yingjun Hao

College of Information and Communication Engineering,
Harbin Engineering University, Harbin 150001, Heilongjiang, China
{dongbin_hao,zhangchunjie,443567277}@hrbeu.edu.cn

Abstract. The existing compressive sensing recovery algorithm has the problems of poor robustness, low peak signal-to-noise ratio (PSNR) and low applicability in images inpainting polluted by impulsive noise. In this paper, we proposed a robust algorithm for image recovery in the background of impulsive noise, called ℓ_p-ADMM algorithm. The proposed algorithm uses ℓ_1-norm substitute ℓ_2-norm residual term of cost function model to gain more image inpainting capability corrupted by impulsive noise and uses generalized non-convex penalty terms to ensure sparsity. The residual term of ℓ_1-norm is less sensitive to outliers in the observations than ℓ_1-norm. And using the non-convex penalty function can solve the offset problem of the ℓ_1-norm (not differential at zero point), so more accurate recovery can be obtained. The augmented Lagrange method is used to transform the constrained objective function model into an unconstrained model. Meanwhile, the alternating direction method can effectively improve the efficiently of ℓ_p-ADMM algorithm. Through numerical simulation results show that the proposed algorithm has better image inpainting performance in impulse noise environment by comparing with some state-of-the-art robust algorithms. Meanwhile, the proposed algorithm has flexible scalability for large-scale problem, which has better advantages for image progressing.

Keywords: Alternative direction method of multipliers · Augmented Lagrangian methods · Compressive sensing · Impulsive noise · Inpainting images

1 Introduction

Compressive Sensing (CS) illuminates that if signal is sparse or compressible, the measurement value of the target signal can be obtained by a non-adaptive linear mapping method far below the sampling frequency required by Shannon-Nyquist sampling theorem and recovering the signal from these measurement,

Supported by the National Natural Science Foundation of China under Grant 61671168, the Natural Science Foundation of Heilongjiang Province under Grant QC2016085 and the Fundamental Research Funds for the Central Universities (HEUCF160807).

S. Liu and G. Yang (Eds.): ADHIP 2018, LNICST 279, pp. 1–9, 2019.
https://doi.org/10.1007/978-3-030-19086-6_1

which has been a hot research topic in recent years [1,2]. Sparse representation of signals $\mathbf{x} \in \mathbb{R}^N$ the basic premise of CS theory application, but in real environment, many natural signals are not sparse in time domain. So it is necessary to transform the signal into other domains to make it sparse. The observation matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$ "senses" the signal $\mathbf{x}$ to obtain the observation signal value $\mathbf{y} = \mathbf{A}\mathbf{x} \in \mathbb{R}^M$, which is obtained by inner product of the row vector of the observation matrix and the signal. During the observation, it will be interfered by Gaussian white noise, and its measured value is

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n} \tag{1}$$

where $\mathbf{n} \in \mathbb{R}^M$ is additive gauss white noise. The recovery process of Compressed Sensing is the mapping process from low-dimensional data space to high-dimensional data space. The best cost function model for this process is

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \qquad \text{s.t} : \mathbf{y} = \mathbf{A}\mathbf{x} \tag{2}$$

$\|\mathbf{x}\|_0 = |\text{supp}(\mathbf{x})| = \#\{i : \text{x}(i) \neq 0\}$. $\text{supp}(\mathbf{x})$ represents the support range of $\mathbf{x}$, $|\text{supp}(\mathbf{x})|$ represents "cardinality", that is to say, the number of non-zero elements in the statistical vector $\mathbf{x}$, but Solving sparse solution of formula (2) will be NP-hard problem [3,4] with the increase of signal dimension. In order to reduce computational complexity, Candes and Donoho prove that the ℓ_0 norm model can be replaced by ℓ_1 norm under the condition that Restricted Isometry Property (RIP) criterion is met, and the obtained solution is very similar to that under ℓ_0 norm model. Many researchers have proposed the solution of formula (2), such as basis-pursuit denoising (BPDN) [5] or LASSO [6], which minimizes the ℓ_0-norm relaxes to the ℓ_1-norm.

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \qquad \text{s.t} : \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \boldsymbol{\epsilon}, \tag{3}$$

For formula (3), the constrained optimization problem of formula (3) can be converted into an unconstrained form by using Lagrange function

$$\widehat{\mathbf{x}} = \arg\min_{\mathbf{x} \in \mathbb{R}^{N \times 1}} \left\{ \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{x}\|_1 \right\} \tag{4}$$

where $\lambda > 0$ is a regularization parameter, which balances the twin cost of minimizing both error and sparsity. From a cost function model point of view, it plays a trade-off role.

Although the use of ℓ_1-regularization in the cost function model has good properties. The performance of the ℓ_1-regularization has two aspects drawbacks. First, ℓ_1 norm is non-differentiable at zero. Second, it would lead to biased estimation of large coefficients.

To address the above drawbacks, many improved methods have been proposed, such as the ℓ_q quasi-norm is used as the sparse term of the objective function, and its formula is modified to

$$\widehat{\mathbf{x}} = \arg\min_{\mathbf{x} \in \mathbb{R}^{N \times 1}} \left\{ \frac{1}{\lambda} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \|\mathbf{x}\|_q^q \right\}, \tag{5}$$

where $0 \le q < 1$, $\|\cdot\|_q^q$ is the ℓ_q quasi-norm defined as $\|\mathbf{x}\|_q^q = \sum_{i=0}^{N} |x_i|^q$.

At present, estimation methods based on CS sparse recovery mainly focus on robust denoising model under the Gaussian noise background. However, in practical applications, the measurement values are not only affected by Gauss noise, but also by non-gauss white noise. Impulse noise is discontinuous and the characteristics of short duration and large amplitude irregular pulses. Impulsive interfere may come from a sudden change in one bit of data during measurements process [7], and many image & video processing works [8,9]. It is well-known that ℓ_2-norm data-fitting is based on the least square method, so it is very sensitive to outliers in observed values. Moreover, the data-fitting efficiency using ℓ_2 norm is very low.

In recent years, various robust image processing methods have been proposed to suppress the interference of outliers in measurement. In [10], the Lorentzian-norm and Huber penalty function are used as residual terms of the objective function, and the objective function is optimized to recover sparse signals. In [11] the ℓ_1-norm is used as the residual term in the objective function and also as the sparse term, and is called ℓ_1-LA with the formula:

$$\widehat{\mathbf{x}} = \underset{\mathbf{x}\in\mathbb{R}^{N\times 1}}{\arg\min} \{\|\mathbf{Ax} - \mathbf{y}\|_1 + \lambda\|\mathbf{x}\|_1\}. \tag{6}$$

It has been shown in [11] that the ℓ_1-norm cost function has better suppression ability to impulse noise than ℓ_2-norm.

In this paper, using the ℓ_p quasi-norm $(0 \le p \le 1)$, as sparsity regular term of the objective function, the Eq. (6) can be rewritten as:

$$\widehat{\mathbf{x}} = \underset{\mathbf{x}\in\mathbb{R}^{N\times 1}}{\arg\min} \left\{\|\mathbf{Ax} - \mathbf{y}\|_1 + \lambda\|\mathbf{x}\|_p^p\right\}. \tag{7}$$

In order to reduce the operation time of solving the objective function model and improve the processing ability of high-dimensional data, the objective function of formula (7) is solved by efficient alternating direction methods, called ℓ_p-ADMM. For more details about ℓ_p-ADMM algorithm, seen 4.1

2 Symmetric α-Stable ($S\alpha S$) Distribution Model

α stable distribution does not have a unified and closed probability density function (PDF) expression, but its characteristic function (CF) can be expressed as [12]

$$\varphi(t) = \{\exp(jat - \gamma^\alpha |t|^\alpha)[1 + j\beta \mathrm{sign}(t)\,\omega(t,\alpha)]\}. \tag{8}$$

where sign(t) is sign function, $0 < \alpha \le 2$ is the characteristic exponent, a is the location parameter, $\gamma > 0$ is the scale parameter, and $\omega(t,\alpha)$ formulation is expressed as

$$\omega(t,\alpha) = \begin{cases} \tan(\alpha\pi/2), & \alpha \ne 1 \\ (2/\pi)\log|t|, & \alpha = 1. \end{cases} \tag{9}$$

In this paper, we just need to consider Symmetric α-Stable ($S\alpha S$) distribution model when $\beta = 0$ in (8). There α-Stable distribution has two special cases. When $\alpha = 2$ and $\beta = 0$ is Gauss distribution; $\alpha = 1$ and $\beta = 0$ is Cauchy distribution.

3 Proximity Operator for ℓ_p-Norm Function

Consider the proximity operator of a function $g(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^N$ with penalty η [13]

$$\operatorname{prox}_{g,\eta}(\mathbf{t}) = \arg\min_{\mathbf{x}} \left\{ a \|\mathbf{x}\|_p^p + \frac{\eta}{2} \|\mathbf{x} - \mathbf{t}\|_2^2 \right\} \tag{10}$$

where $0 \leq p \leq 1$ and $a > 0$.

Case 1: $p = 0$**.** The expression of proximity operator of formula (10) is:

$$\operatorname{prox}_{g,\eta}(\mathbf{t})_i = \begin{cases} 0, & |t_i| \leq \sqrt{2a/\eta} \\ t_i, & \text{others} \end{cases}, \quad i = 1, 2, \cdots, N \tag{11}$$

where t_i is the i-th element of the vector $\mathbf{t}$, and is well-known hard-thresholding operator.

Case 2: $0 < p < 1$**.** The proximity operator of formula (10) can be evaluated as [15]

$$\operatorname{prox}_{g,\eta}(\mathbf{t})_i = \begin{cases} 0, & |t_i| < \tau \\ \{0, \operatorname{sign}(\mathrm{t_i})\,\beta\}, & |t_i| = \tau, \quad i = 1, \cdots, N \\ \operatorname{sign}(\mathrm{t_i})\, z_{\mathrm{i}}, & |t_{\mathrm{i}}| > \tau \end{cases} \tag{12}$$

where $\beta = [2a(1-p)/\eta]^{1/(2-p)}$, $\tau = \beta + \left(ap\beta^{p-1}\right)/\eta$, z_i is the solution of $h_1(z) = paz^{p-1} + \eta z - \eta |t_i| = 0, \quad z \geq 0$ [14].

Case 3: $p = 1$**.** This is the well-known soft-thresholding operator, which the proximity operator can be written as:

$$\operatorname{prox}_{g,\eta}(\mathbf{t})_i = S_{a/\eta}(\mathbf{t})_i = \operatorname{sign}(t_i) \max\{|t_i| - \mathrm{a}/\eta, 0\} \tag{13}$$

4 Proposed ℓ_p-ADMM Algorithm

ADMM is parallel distributed algorithm, which is generally based on a convex optimization model with separable variables and is suitable for large-scale problems in cloud computing and image processing [16]. ADMM takes the form of a decomposition-coordination procedure, in which the solutions to small local subproblems are coordinated to find a global solution. ADMM mainly blend the benefits of dual decomposition and augmented Lagrangian methods for constrained optimization.

In the ADMM framework, the ℓ_1 loss term and the nonsmooth ℓ_p-regularization term are naturally separated. Using an auxiliary variable $\mathbf{v} \in \mathbb{R}^M$, the formulation (7) can be rewritten as

$$\min_{\mathbf{x},\mathbf{v}} \left\{ \frac{1}{\lambda} \|\mathbf{v}\|_1 + \|\mathbf{x}\|_p^p \right\} \qquad s.t.\, \mathbf{A}\mathbf{x} - \mathbf{y} = \mathbf{v}. \tag{14}$$

The augmented Lagrangian function of formula (14) can be written as

$$\mathcal{L}_\rho(\mathbf{v},\mathbf{x},\mathbf{w})=\frac{1}{\lambda}\|\mathbf{v}\|_1+\|\mathbf{x}\|_p^p-\langle\mathbf{w},\mathbf{A}\mathbf{x}-\mathbf{y}-\mathbf{v}\rangle+\frac{\rho}{2}\|\mathbf{A}\mathbf{x}-\mathbf{y}-\mathbf{v}\|_2^2 \tag{15}$$

where $\mathbf{w}$ is a the dual variable, $\rho>0$ is a penalty parameter. Then, ADMM is mainly consists of the following iterative steps:

$$\mathbf{v}^{k+1}:=\arg\min_{\mathbf{v}}\left(\frac{1}{\lambda}\|\mathbf{v}\|_1+\frac{\rho}{2}\left\|\mathbf{A}\mathbf{x}^k-\mathbf{y}-\mathbf{v}-\frac{\mathbf{w}^k}{\rho}\right\|_2^2\right) \tag{16}$$

$$\mathbf{x}^{k+1}:=\arg\min_{\mathbf{x}}\left(\|\mathbf{x}\|_p^p+\frac{\rho}{2}\left\|\mathbf{A}\mathbf{x}-\mathbf{y}-\mathbf{v}^{k+1}-\frac{\mathbf{w}^k}{\rho}\right\|_2^2\right) \tag{17}$$

$$\mathbf{w}^{k+1}:=\mathbf{w}^k-\rho\left(\mathbf{A}\mathbf{x}^{k+1}-\mathbf{y}-\mathbf{v}^{k+1}\right) \tag{18}$$

The $\mathbf{x}$-subproblem update step (17) actually resolved a penalized ℓ_1-ℓ_p problem. We use a basic method to speed up ADMM and approximate this subproblem. Let $\mathbf{u}^k=\mathbf{y}+\mathbf{v}^{k+1}+\mathbf{w}^k/\rho$, we can approximate the subproblem by linearizing the quadratic term of its cost function at point $\mathbf{x}^k$, which is expanded as follows:

$$\begin{aligned}\tfrac{1}{2}\left\|\mathbf{A}\mathbf{x}-\mathbf{u}^k\right\|_2^2\approx\tfrac{1}{2}&\left\|\mathbf{A}\mathbf{x}^k-\mathbf{u}^k\right\|_2^2\\&+\left\langle\mathbf{x}-\mathbf{x}^k,d\left(\mathbf{x}^k\right)\right\rangle+\tfrac{L_1}{2}\left\|\mathbf{x}-\mathbf{x}^k\right\|_2^2\end{aligned} \tag{19}$$

where $d\left(\mathbf{x}^k\right)=\mathbf{A}^T\left(\mathbf{A}\mathbf{x}^k-\mathbf{u}^k\right)$ is the gradient of the quadratic term, $L_1>0$ is a proximal parameter.

Based on the (19) approximation, the $\mathbf{x}$-subproblem (17) becomes easy to solve by proximity operator (10), which can be efficiently solved as

$$\mathbf{x}^{k+1}=\operatorname{prox}_{\|\mathbf{x}\|_p^p,\rho}\left(\mathbf{b}^k\right)=\begin{cases}\textit{solved as}\,(10)\,, & p=0\\ \textit{solved as}\,(11)\,, & 0<p<1\\ \textit{solved as}\,(12)\,, & p=1\end{cases} \tag{20}$$

with $\mathbf{b}^k=\mathbf{x}^k-(1/L_1)\,\mathbf{A}^T\left(\mathbf{A}\mathbf{x}^k-\mathbf{u}^k\right)$

Table 1. PSNR of the recovery image under the $S\alpha S$ noise environment.

Algorithm	L1LS-FISTA	LqLS-ADMM (q = 0.5)	YALL1	ℓ_p-ADMM (p = 0.2)	ℓ_p-ADMM (p = 0.5)	ℓ_p-ADMM (p = 0.7)
Shepp-Logan	13.04 (dB)	12.94 (dB)	29.44 (dB)	**41.29 (dB)**	41.02 (dB)	39.83 (dB)
MRI	15.51 (dB)	15.47 (dB)	25.27 (dB)	26.71 (dB)	27.23 (dB)	**27.39 (dB)**

The $\mathbf{v}$-update step (16) is a form of the proximity operator (13)

$$\mathbf{v}^{k+1}=S_{1/(\rho\lambda)}\left(\mathbf{A}\mathbf{x}^k-\mathbf{y}-\frac{\mathbf{w}^k}{\rho}\right) \tag{21}$$

For convex cases, the convergence property of the ADMM has been well solved. Recently, there are few explanations on the convergence of non-convex case [17].

5 Recovery of Images in the Impulsive Noise Environment

We evaluate recovery performance of the proposed algorithm in comparison with L1LS-FISTA [18], YALL1 [11] and LqLS-ADMM [19]. L1LS-FISTA solves the ℓ_1-Least Square formulation. YALL1 solves the ℓ_1-LA formulation (6). We mainly conduct reconstruction on the simulated images.

This experiment evaluates the performance of ℓ_p-ADMM algorithm on the image recovery under $S\alpha S$ impulsive noise environment. The test images are mainly "Shepp-Logan" and MRI images. The size of each image is 265×256, and this two-dimensional image is converted into one-dimensional image at the same time, which are set to $N = 65536$ and $M = round(0.4N)$. As shown in Fig. 1. Sensing matrix $\mathbf{A}$ is composed of discrete cosine transformation matrix as the measurement matrix and Haar wavelets as the basis functions. We only consider $S\alpha S$ the case of impulsive noise, whose parameters are set to $\alpha = 1$ and $\gamma = 10^{-4}$. PSNR is used to evaluate the recovery performance of the improved algorithm on images.

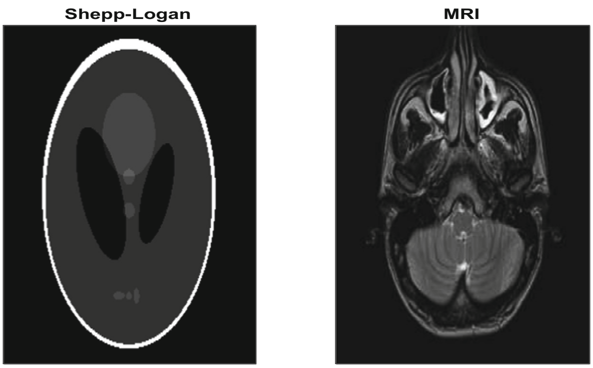
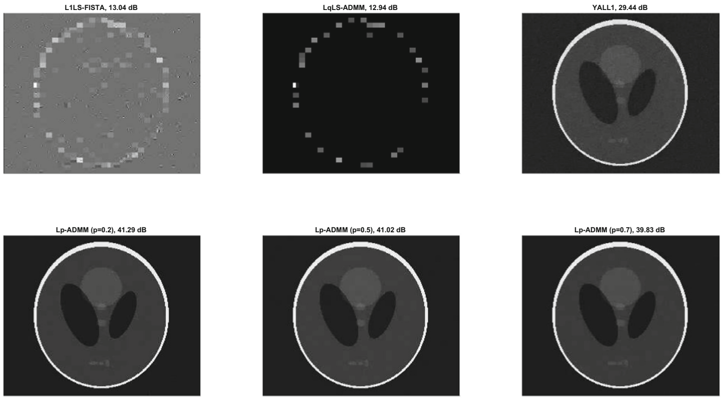
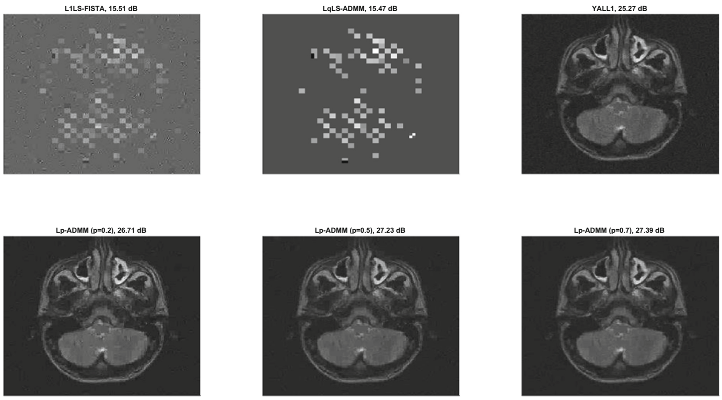


Fig. 1. Using two 256×256 images as test images.

The simulation results are shown in Table 1. It can be seen that the output PSNR of YALL1 algorithm for Logan and MRI images under $S\alpha S$ noise is higher than that of L1LS-FISTA and LqLs-ADMM algorithms. This is because the residual terms of L1LS-FISTA and LqLS-ADMM algorithms both adopt ℓ_2 norm, while ℓ_2 norm only has good suppression effect on Gaussian white noise and is very sensitive to noise. Therefore, L1LS-FISTA and LqLS-ADMM algorithms have poor recovery performance on images affected by impulse noise. However



(a)



(b)

Fig. 2. Recovery images performance of the compared algorithms in $S\alpha S$ noise; (a): Averaged PSNR of Shepp-Logan for different algorithm; (b): Averaged PSNR of MRI for different algorithm.

Lp-ADMM algorithm using ℓ_p quasi-norm in sparse terms is better than that of YALL1, because ℓ_p quasi-norm can solve the deficiency of ℓ_1 norm.

It can be seen from Fig. 2 that the improved algorithm proposed in this paper has better recovery performance than other comparison algorithms under the $S\alpha S$ noise environment. It can be seen that L1LS-FISTA and LqLS-ADMM based on ℓ_2 norm as residual term have failed, while the ℓ_1-loss based algorithm, YALL1 and ℓ_p-ADMM have work well. ℓ_p-ADMM algorithm performance advantage over other algorithms. Furthermore, the simulation results show that in recovering the MRI image, for ℓ_p-ADMM, $p = 0.7$ yield better performance than $p = 0.2$ and $p = 0.5$, which is different from the results of recovery "Shepp-Logan", where $p = 0.2$ PSNR significantly better performance than $p = 0.5$ and $p = 0.7$. This is because images in real-life are not as sparse as synthetic images, but compressible.

6 Conclusion

This paper presents a robust formula for images recovery in the $S\alpha S$ noise, which improves the ℓ_1-LA formula by replacing ℓ_1-regularization with generalized non-convex regularization (ℓ_p-norm, $0 < p < 1$). In order to effectively solve the non-convex and non-smooth minimization problem, a first-order algorithm based on ADMM and approximation operator is proposed. Simulation results on recovery images demonstrated that the proposed algorithm obtains considerable performance gain over the other algorithms such as the L1-FISTA,YALL1 and LqLS-ADMM in the $S\alpha S$ noise.

References

1. Candès, E.J., Romberg, J., Tao, T.: Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, pp. 489–509. IEEE Press (2006)
2. Donoho, D.L.: Compressed sensing. IEEE Trans. Inf. Theory **52**(4), 1289–1306 (2006)
3. Donoho, D.L.: For most large underdetermined systems of linear equations the minimal. Commun. Pure Appl. Math. **59**(6), 797–829 (2006)
4. Donoho, D.L., Elad, M., Temlyakov, V.N.: Stable recovery of sparse overcomplete representations in the presence of noise. IEEE Trans. Inf. Theory **52**(1), 6–18 (2006)
5. Candès, E.J., Romberg, J., Tao, T.: Stable signal recovery from incomplete and inaccurate measurements (2006)
6. Knight, K., Fu, W.: Asymptotics for lasso-type estimators. Ann. Stat. **28**, 1356 (2011)
7. Candès, E.J., Randall, P.A.: Highly robust error correction by convex programming. IEEE Trans. Inf. Theory **54**(7), 2829–2840 (2008)
8. Bar, L., Brook, A., Sochen, N., Kiryati, N.: Deblurring of color images corrupted by impulsive noise. IEEE Trans. Image Process. **16**(4), 1101–1111 (2007)
9. Civicioglu, P.: Using uncorrupted neighborhoods of the pixels for impulsive noise suppression with ANFIS. IEEE Trans. Image Process. **16**(3), 759–773 (2007)
10. Carrillo, R.E., Barner, K.E.: Lorentzian iterative hard thresholding: robust compressed sensing with prior information. IEEE Trans. Signal Process. **61**(19), 4822–4833 (2013)
11. Yang, J., Zhang, Y.: Alternating direction algorithms for l_1-problems in compressive sensing. SIAM J. Sci. Comput. **33**, 250–278 (2011). Society for Industrial and Applied Mathematics
12. Nolan, J.: Stable distributions: models for heavy-tailed data (2005). http://Academic2.american.edu/jpnolan

13. Combettes, P.L., Pesquet, J.C.: Proximal splitting methods in signal processing. In: Bauschke, H., Burachik, R., Combettes, P., Elser, V., Luke, D., Wolkowicz, H. (eds.) Fixed-Point Algorithms for Inverse Problems in Science and Engineering. SOIA, vol. 49, pp. 185–212. Springer, New York (2011). https://doi.org/10.1007/978-1-4419-9569-8_10
14. Wen, F., Liu, P., Liu, Y., et al.: Robust sparse recovery in impulsive noise via ℓ_p-ℓ_1 optimization. IEEE Trans. Signal Process. **65**(1), 105–18 (2017)
15. Marjanovic, G., Solo, V.: On l_q optimization and matrix completion. IEEE Trans. Signal Process. **60**(11), 5714–5724 (2012)
16. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn. **3**(1), 1–122 (2011)
17. Hong, M., Luo, Z., Razaviyayn, M.: Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems. SIAM J. Optim. **26**(1), 337–364 (2016)
18. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imaging Sci. **2**(1), 183–202 (2009)
19. Li, G., Pong, T.K.: Global convergence of splitting methods for nonconvex composite optimization. SIAM J. Optim. **25**(4), 2434–2460 (2014)