Signal Processing 239 (2026) 110319

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

A new limited shrinkage thresholding iterative ADMM algorithm for compressed sensing signal reconstruction

Yuan-Min Li *, Hao Wang

School of Mathematics and Statistics, Xidian University, Xi'an 710071, PR China

ARTICLE INFO

Keywords:
Sparse recovery
Iterative hard thresholding
Limited shrinkage thresholding operator
Compressed sensing
ADMM

ABSTRACT

In recent years, solving optimization problems with ℓ_0-norm constraints has become a significant challenge due to their combinatorial nature. Traditional iterative hard thresholding (IHT) algorithms often fail to achieve satisfactory results for complex objective functions, primarily due to their high sensitivity to initialization and slow convergence properties. To address these limitations, we propose an iterative limited shrinkage thresholding ADMM algorithm for solving general non-convex problems with ℓ_0-norm constraints. Our approach integrates the limited shrinkage thresholding (LST) operator and the truncation technique, which significantly enhance the stability of the solutions compared to the traditional IHT methods. By introducing the concept of the LTP-stationary point, we give the optimality conditions of the problem. We prove that the proposed method converges to the LTP-stationary point. The convergence rate is also analyzed under the Kurdyka–Łojasiewicz (KŁ) framework. Extensive experiments on compressed sensing signal reconstruction and image deblurring demonstrate the superiority of our approach.

1. Introduction

The theory of compressed sensing (CS) was first proposed by Donoho, Candès, Romberg, and Tao (see [1,2]). It surpasses the limitations of the classical Nyquist Shannon sampling theory in signal processing [3]. Compressed sensing and its sparse optimization models have been extensively studied and applied in fields including image processing [4], data separation [5]. Generally speaking, CS can be formulated as the following ℓ_0-norm minimization problem

$$\min_x \|x\|_0 \ s.t.\ Ax = b, \tag{1}$$

where $b \in \mathbb{R}^m$ is the observation vector, $A \in \mathbb{R}^{m\times n}$ is the sensing matrix ($m \ll n$) and $x \in \mathbb{R}^n$ is the sparse signal, $\|x\|_0$ represents the number of non-zero elements in x, commonly referred to as the ℓ_0-norm. By the penalty function method, problem (1) can be formulated as the following regularization model

$$\min_x \|Ax - b\|_2^2 + \lambda\|x\|_0, \tag{2}$$

where $\lambda > 0$ is the regularization parameter. When sparse prior information is known, the above problem (1) can be reconsidered in the following

$$\min_x \|Ax - b\|_2^2 \ s.t.\ \|x\|_0 \le s, \tag{3}$$

where $s \in \mathbb{N}$ is the sparsity level. In the past decades, many algorithms have been developed to solve the sparse problems (2) or (3), such as the iterative hard-thresholding algorithm [6], the continuous exact ℓ_0 penalty (CEL0) [7], the forward–backward splitting (FBS) [8], and the mixed integer optimization method (MIO) [9], etc. Due to the non-convex and discontinuous properties of the ℓ_0-norm, minimizing the ℓ_0-norm is an NP-hard problem. To overcome this shortcoming, some relaxation methods have been proposed, such as ℓ_p-norm ($0 < p \le 1$). For $p = 1$, suppose that A satisfies the restricted isometry property (RIP) and $\delta_{2s} < \sqrt{2} - 1$, problem (2) is equivalent to minimizing the ℓ_1-norm [10]. ℓ_1-minimization and re-weighted ℓ_1-minimization are widely used to solve the problems (2) and (3) [11–13]. For $0 < p < 1$, Xu et al. pointed out that only when $p = 1/2$ or $2/3$, the proximal operator of ℓ_p-norm has closed-form thresholding formulas [14]. Moreover, some excellent relaxation functions include minimax concave plus (MCP) [15], smoothly clipped absolute deviation (SCAD) [16], log-sum penalty (LSP) [17,18], and other relaxation methods [19–23]. A comprehensive characterization of the above sparsity-promoting functions can be referred to [24]. The authors also proposed a method for constructing non convex sparsity promoting functions from convex sparsity promoting functions and their Moreau envelopes, which is very helpful for exploring and studying more sparsity promoting functions. In addition, there are many other interesting methods, such as homotopy algorithm [25] for solving (2), which utilizes the piece-wise

* Corresponding author.
E-mail addresses: ymli@xidian.edu.cn (Y.-M. Li), wanghao_23@stu.xidian.edu.cn (H. Wang).

https://doi.org/10.1016/j.sigpro.2025.110319
Received 19 March 2025; Received in revised form 28 June 2025; Accepted 23 September 2025
Available online 29 September 2025

constant property of regularization paths to construct skip search and introduces backtracking mechanism to avoid local optimal.

It is worth mentioning that Hu et al. proposed a limited shrinkage thresholding (LST) operator [26], which provides a unified framework for proximal mappings of some sparsity promoting functions, such as SCAD penalty, MCP penalty, etc. Through the combination of the LST operator with continuation and truncation techniques, their theoretical analysis demonstrates that when the sensing matrix A satisfies the RIP, the proposed algorithms achieve convergence to solutions in close truth sparse signals. Empirical evaluations reveal that the sparse solutions obtained by this method significantly outperform the stationary solutions generated by standard proximal gradient algorithms. These compelling results motivate us to further investigate into the fundamental characteristics of the LST operator, particularly its convergence rate analysis under non-convex optimization settings.

Regarding ℓ_0-norm constraint problem (3), several low-complexity greedy algorithms have been proposed. These include orthogonal matching pursuit (OMP) [27], iterative hard thresholding (IHT) [28], hard thresholding pursuit (HTP) [29,30], compressive sampling matching pursuit (CoSaMP) [31], fast recursive algorithm (RASU) [32] and subspace pursuit (SP) [33]. Notably, the IHT algorithm [34], characterized by its simple structure and well-established theoretical framework, has inspired numerous enhanced variants. These developments encompass normalized iterative hard thresholding (NIHT) [35], accelerated iterative hard-thresholding (AIHT) methods [36]. By combining the conjugate gradient method, Blanchard, Tanner and Wei developed a conjugate gradient iterative hard-thresholding (CGIHT) for the problem of the compressive sensing and matrix restoration, and then established the restoration bound theory [37]. Zhou, Xiu, and Qi proposed a Newton hard-thresholding pursuit (NHTP) for the sparse optimization problem of general continuous differentiable functions. They also obtained the global quadratic convergence results under the assumptions of restricted strongly convex and smooth functions [38]. Additionally, Zhao introduced a novel optimal k-thresholding technique [39], which successfully overcome the numerical oscillation phenomenon in HTP. From the above analysis, sparse constraint is a main obstacle to design a global convergent algorithm. This trigger us to develop an efficient algorithm for sparse optimization.

The ADMM, initially proposed in [40,41], has emerged as a powerful iterative framework for constrained optimization. Its versatility has driven widespread adoption in image processing [42] and compressive sensing [43]. However, convergence analysis remains nontrivial when handling non-convex objectives—a challenge addressed by Bolte et al. through their proximal alternating linearized minimization (PALM) method [44]. By exploiting the Kurdyka–Łojasiewicz (KŁ) property, the convergence of the PALM can be guaranteed for nonsmooth non-convex optimization, with subsequent works establishing its convergence rate under metric sub-regularity conditions [45,46]. For perturbed compressive sensing, Li et al. proposed a proximal alternating partial linear minimization (PAPLM), and combined with KŁ property, proved that any generated sequence will converge to the critical point of the objective function [47]. Recent advances focus on ADMM variants handling composite constraints [48], where a methodology combines augmented Lagrangian techniques for simple constraints with projection operators for intricate ones. This paradigm proves particularly effective for ℓ_0-norm constrained problems, as demonstrated in Teng et al. [49]. The authors investigated the optimality conditions for ℓ_0-constrained optimization, integrating PALM for continuous constraints and projection operator for combinatorial constraints. In addition, a new deep learning architecture (ADMM-CSNet) was proposed by combining traditional model-based compressive sensing methods with deep learning methods and introducing the ADMM to reconstruct images from sparse sampling measurements [50]. Thanks to the ADMM technique, we will develop a new limited shrinkage thresholding iterative ADMM algorithm for sparse optimization with the ℓ_0-norm constraint.

Consider the ℓ_0-norm constrained sparse optimization problem formulated as the following:

$$\min_x f(Ax+b)\ s.t.\ \|x\|_0 \le s, \tag{4}$$

where $f : \mathbb{R}^m \to \mathbb{R}$ is a differentiable function with a lower bound, $A \in \mathbb{R}^{m\times n}$ is a real matrix with $m < n$, and $b \in \mathbb{R}^m$. The objective is a composite function of f and an affine function $Ax+b$. Through variable substitution $y = Ax + b$, we equivalently reformulate (4) as

$$\min_y f(y)\ s.t.\ \|x\|_0 \le s,\ y = Ax + b. \tag{5}$$

By utilizing the LST operator and ADMM technique, we propose a limited shrinkage thresholding ADMM for sparse recovery problem (4). Besides global convergence, we further analyze the convergence rate of the algorithm using KŁ properties. Our contributions are specified as follows:

(1) We investigate the properties of the LST operator, propose the concept of LTP-stationary point and study the optimality conditions. We prove that an LTP-stationary point can be considered as an approximate P-stationary point.
(2) In conjunction with the ADMM, we propose a new limited shrinkage thresholding ADMM algorithm framework and prove that under certain conditions, this algorithm converges to an approximate sparse solution for problem (5). Additionally, combining the idea of inertial acceleration, we provide an accelerated form of the algorithm.
(3) We prove that our algorithm has sufficient descent property when penalty factor ρ is sufficiently large, and analyze the convergence rate of the algorithm based on KŁ property and RIP condition. It addresses the shortcomings in the study of convergence rates related to problem (5).
(4) We conduct numerical experiments in compressed sensing signal reconstruction and image deblurring by employing the proposed algorithms. In comparison with the classical IHT algorithm, recent algorithms ALPA and ILSTAT, our algorithm demonstrates the excellent performance.

The rest of this paper is organized as follows. In Section 2, we introduce some preliminaries. In Section 3, we propose optimality conditions for problem (5). In Section 4, an iterative limited shrinkage thresholding algorithm with ADMM is proposed to solve the problem (5). And we analyze the convergence rate of the algorithm based on KŁ properties. Numerical experiments and conclusions are presented in Sections 5 and 6, respectively.

2. Some preliminaries

In this section, we will introduce some basic definitions and lemmas. First, these are some notations. We use $\mathbb{R}^{m\times n}$ to represent the real matrix of m times n. And we represent n-dimensional Euclidean space with $\mathbb{R}^n$, $\mathbb{R}_+ = \{x : x \ge 0, x \in \mathbb{R}\}$, $\mathbb{R}_{++} = \{x : x > 0, x \in \mathbb{R}\}$. We use $h\circ f(x)$ to represent $h(f(x))$. Given a vector $x \in \mathbb{R}^n$, the l_2-norm of x is defined as $\|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$, and the ℓ_0-norm of x is defined as the number of non-zero elements in x, denoted by $\|x\|_0$. If $\|x\|_0 \le s$, then vector x is called an s-sparse vector, and we use $supp(x) = \{i : x_i \ne 0\}$ to represent the set of positional indicators for non-zero elements in x. For any $i \in \{1, 2, \dots, n\}$, x_i represents the ith component of $x \in \mathbb{R}^n$ and A_i represents the ith column vector of $A \in \mathbb{R}^{m\times n}$. We provide some definitions that will be used later.

Definition 1. For a proper, lower semi-continuous (l.s.c.) function $f : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ and a parameter value $\lambda > 0$, the proximal mapping $prox_{\lambda f}(x)$ is defined by

$$prox_{\lambda\phi}(x) = \arg\min_\omega \left\{ \phi(\omega) + \frac{1}{2\lambda}\|\omega - x\|_2^2 \right\}. \tag{6}$$

Table 1
Sparsity promoting functions and the corresponding LST operators.

Sparsity promoting functions	Proximal operator (LST operator)	$\kappa(\lambda)$ in LST
$\phi_{SCAD}(x)=\begin{cases}\lvert x\rvert, & \lvert x\rvert<\lambda\\ \frac{2a\lvert x\rvert-x^2-\lambda}{2(a-1)}, & \lambda\le\lvert x\rvert<a\lambda\\ \frac{(a+1)\lambda}{2}, & \lvert x\rvert>a\lambda\end{cases}$ [16]	$prox_{\phi_{SCAD}}(x)=\begin{cases}0, & \lvert x\rvert\le\lambda\\ \mathrm{sign}(x)(\lvert x\rvert-\lambda), & \lambda<\lvert x\rvert\le 2\lambda\\ \frac{(a-1)x-\mathrm{sign}(x)a\lambda}{a-2}, & 2\lambda<\lvert x\rvert\le a\lambda\\ x, & \lvert x\rvert>a\lambda\end{cases}$	λ
$\phi_{MCP}(x)=\begin{cases}\lvert x\rvert-\frac{x^2}{2\gamma\lambda}, & \lvert x\rvert<\gamma\lambda\\ \frac{1}{2}\gamma\lambda, & \lvert x\rvert\ge\gamma\lambda\end{cases}$ [15]	$prox_{\lambda\phi_{MCP}}(x)=\begin{cases}0, & \lvert x\rvert\le\lambda\\ \mathrm{sign}(x)\frac{\lvert x\rvert-\lambda}{1-1/\gamma}, & \lambda<\lvert x\rvert\le\gamma\lambda\\ x, & \lvert x\rvert>\gamma\lambda\end{cases}$	λ
$\phi_{\ell_p}(x)=\lvert x\rvert^p,\ 0\le p\le 1$ [51]	$prox_{\lambda\phi_{\ell_p}}(x)=\begin{cases}0, & \lvert x\rvert\le\alpha_p\lambda^{\frac{1}{2-p}}\\ t^*, & \lvert x\rvert>\alpha_p\lambda^{\frac{1}{2-p}}\end{cases}$	$\alpha_p\lambda^{\frac{1}{2-p}}$
$\phi_{log}(x)=log(1+\frac{\lvert x\rvert}{\epsilon})$ [17,18]	$prox_{\lambda\phi_{log}}(x)=\begin{cases}0, & \lvert x\rvert\le\frac{\lambda}{\epsilon}\\ \mathrm{sgn}(x)r(\lvert x\rvert), & \lvert x\rvert>\frac{\lambda}{\epsilon}\end{cases}$	$\frac{\lambda}{\epsilon}$

Most common penalty functions and the corresponding proximal operators, such as ϕ_{SCAD} and ϕ_{MCP}, are summarized in Table 1. The sparsity of the regularization penalty functions can be explained by their proximal operators. A further detailed description can be refereed to [24]. In [24], besides the functions mentioned in Table 1, a series of other sparsity promoting functions were formally defined and studied.

In Table 1, t^* is the unique (non-zero) solution to problem $\min_{\omega\in\mathbb{R}}\{\lambda|\omega|^p+\frac{1}{2}(\omega-x)^2\}$, α_p and $r(|x|)$ are

$$\alpha_p=\begin{cases}1, & p=1\\ (2-p)(2-2p)^{-\frac{1-p}{2-p}}, & p\in[0,1),\end{cases} \tag{7}$$

$$r(|x|)=\frac{1}{2}(|x|-\epsilon)+\sqrt{\frac{1}{4}(|x|+\epsilon)^2-\lambda}. \tag{8}$$

In [26], the authors unified the proximal mappings of some sparsity promoting functions into limited shrinkage thresholding (LST) operator. And we give the definition of the LST operator here.

Definition 2 (*See [26]*)**.** Let $\kappa:\mathbb{R}_{++}\to\mathbb{R}_+$ and $\lambda\in\mathbb{R}_+$.

(1) $\mathbb{T}_\lambda:\mathbb{R}\to\mathbb{R}$ is said to be a thresholding operator relative to $\kappa(\lambda)$ if the following thresholding property is satisfied

$$\mathbb{T}_\lambda(t)=0 \quad whenever \quad |t|\le\kappa(\lambda). \tag{9}$$

(2) $\mathbb{LT}_\lambda:\mathbb{R}\to\mathbb{R}$ is said to be a limited thresholding operator relative to $\kappa(\lambda)$ if (9) and the following limited shrinkage property are satisfied

$$\left|\mathbb{LT}_\lambda(t)-t\right|\le\kappa(\lambda) \quad for\ each\ t\in\mathbb{R}. \tag{10}$$

From [52], we know that the penalty functions mentioned in Table 1 are prox-bounded. Since the LST operators are generalizations of the proximal operators, we must require the LST operators to satisfy some conditions similar to proximal operators. In the following, we give an assumption which can ensure the compatibility between the LST operators and the proximal operators. This assumption is also a fundamental condition to analyze the convergence of the following proposing algorithms.

Assumption 1. If $\omega^k\in\mathbb{LT}_\lambda(x^k)$, $\lim_{k\to\infty}x^k=x^*$ and the sequence $\{\omega^k\}_{k\in\mathbb{N}}$ is bounded. Then for all the cluster points ω^* of $\{\omega^k\}$ lie in $\mathbb{LT}_\lambda(x^*)$.

Remark 1. According to [52], if an LST operator can be represented as a proximal operator, then it satisfies this assumption. But, if an LST operator is not defined by a proximal operator, whether this assumption is met requires further verification.

In sparse recovery, we usually assume that the recovery matrix satisfies the restricted isometry property (RIP), which is a regular condition measuring how close the submatrices are orthonormal restricted to sparse subspaces.

Definition 3 (*See [10]*)**.** Let $A\in\mathbb{R}^{m\times n}$ and $s\in\mathbb{N}$. The s-restricted isometry constant δ_s is defined to be the smallest quantity such that

$$(1-\delta_s)\|x\|_2^2\le\|Ax\|_2^2\le(1+\delta_s)\|x\|_2^2, \tag{11}$$

for each $x\in\mathbb{R}^n$ with $\|x\|_0\le s$. The matrix A is said to satisfy the s-RIP with δ_s if $0<\delta_s<1$.

In numerous studies, Bernoulli and Gaussian random matrices have been shown to exhibit high probability of satisfying the restricted isometry property (RIP) [2]. Matrices satisfying RIP, termed RIP matrices, are fundamental in compressed sensing analysis. Additionally, we assume that f is an L_f-smooth function, which is useful in the subsequent analysis.

Definition 4. A differentiable function $f(x)$ is called L_f-smooth, if for any $x_1,x_2\in\mathbb{R}^n$, there exists $L_f<+\infty$ such that

$$\|\nabla f(x_1)-\nabla f(x_2)\|_2\le L_f\|x_1-x_2\|_2. \tag{12}$$

Lemma 1. *Let $f:\mathbb{R}^n\to(-\infty,\infty]$ be an L_f-smooth function. Then, for any $x_1,x_2\in\mathbb{R}^n$, we have*

$$f(x_1)\le f(x_2)+\langle\nabla f(x_2),x_1-x_2\rangle+\frac{L_f}{2}\|x_1-x_2\|_2^2. \tag{13}$$

Let $\eta\in(0,\infty]$. We define $\boldsymbol{\Phi}_\eta$ as the class of concave continuous functions $\psi:[0,\eta)\to\mathbb{R}$ satisfying that $\psi(0)=0$ and ψ is continuously differentiable on $(0,\eta)$ with $\psi'(t)>0$ for all $t\in(0,\eta)$. We now state the Kurdyka-Łojasiewicz (KŁ) property, which is common used in analyzing the convergence of iterative algorithms.

Definition 5 (*Kurdyka–Łojasiewicz property*)**.** Let $\sigma:\mathbb{R}^d\to(-\infty,+\infty]$ be a proper and lower semi-continuous function.

(1) The function σ is said to have the Kurdyka–Łojasiewicz (KŁ) property at $u\in\mathrm{dom}\,\partial\sigma:=\{u\in\mathbb{R}^d:\partial\sigma(u)\neq\emptyset\}$ if there exist $\eta\in(0,+\infty]$, a neighborhood U of u, and a function $\psi\in\boldsymbol{\Phi}_\eta$, such that for all

$$u\in U\cap\{\sigma(u^*)<\sigma(u)<\sigma(u^*)+\eta\}, \tag{14}$$

the following inequality holds:

$$\psi'(\sigma(u)-\sigma(u^*))dist(0,\partial\sigma(u))\ge 1. \tag{15}$$

(2) If σ satisfies the KŁ property at each point of $\mathrm{dom}\,\partial\sigma$, then σ is called a KŁ function.

If σ is differentiable, then $\partial\sigma(x) = \nabla\sigma(x)$. Assuming $\psi(s) = cs^{1-\theta}$ for some $c > 0$ and $\theta \in (0, 1)$, then the KŁ inequality (15) becomes

$$\bar{c}\|\nabla\sigma(u)\|_2^{\frac{1}{\theta}} \geq \sigma(u) - \sigma(u^*), \tag{16}$$

for all u near u^* with $\sigma(u) \geq \sigma(u^*)$, where $\bar{c} = (c(1-\theta))^{\frac{1}{\theta}}$.

3. Optimality conditions

In this section, we introduce the definitions of some different types of stationary points, discuss the interconnections and then derive the optimality conditions for the sparse optimization problems. Unlike classical stationary points that only require zero gradients, the stationary points in sparse optimization needs support-set stability, which is a practical necessity for iterative algorithms to avoid oscillating between different sparse patterns.

For the original sparse problem (4), now we provide the definitions of L-stationary point and LT–stationary point and then analyze the interrelations.

Definition 6 (*See [34]*)**.** An s-sparse vector x^* is called an L-stationary point of (4) if it satisfies the relation

$$x^* \in \mathbb{H}_s(x^* - \frac{1}{\upsilon}A^T\nabla f(Ax^* + b)), \tag{17}$$

where $1/\upsilon$ is the step size and $\mathbb{H}_s$ is the hard thresholding operator, which is to set all but the largest s elements of a vector (in magnitude) to zero [53].

It should be pointed out that if x^* is the optimal solution of (4), then x^* satisfies (17). Conversely, it may not necessarily hold true. Regarding the L-stationary point, we have the following equivalent characterization.

Lemma 2 (*See [34]*)**.** *For any $\upsilon > 0$, x^* satisfies (17) if and only if $\|x^*\|_0 \leq s$ and*

$$\left|A_i^T\nabla f(Ax^* + b)\right| \begin{cases} \leq \upsilon\left|x^*_{(s)}\right|, & i \notin supp(x^*), \\ = 0, & i \in supp(x^*), \end{cases} \tag{18}$$

where $x^*_{(s)}$ is the sth largest absolute value component in x^*.

In [26], the authors point out that by combining the hard thresholding operator and the LST operator, it helps to ensure the iterative point converge to an approximate true solution with the required sparsity of (4). Based on this fact, we propose the definition of the following LT-stationary point.

Definition 7. An s-sparse vector x^* is called an LT-stationary point of (4) if it satisfies the relation

$$x^* \in \mathbb{H}_s \circ \mathbb{LT}_{\upsilon\lambda}(x^* - \frac{1}{\upsilon}A^T\nabla f(Ax^* + b)). \tag{19}$$

In the following, we provide a necessary condition for the LT-stationary point. Based on this result, we can derive that the LT-stationary point can be regarded as an approximation of the L-stationary point.

Lemma 3. *For any $\upsilon > 0$, if x^* satisfies (19), then*

$$\left|A_i^T\nabla f(Ax^* + b)\right| \leq \begin{cases} \upsilon\kappa(\upsilon\lambda) + \upsilon\left|x^*_{(s)}\right|, & i \notin supp(x^*), \\ \upsilon\kappa(\upsilon\lambda), & i \in supp(x^*). \end{cases} \tag{20}$$

When $\kappa(\upsilon\lambda) \to 0$, the LT-stationary point tends to L-stationary point.

Proof. The meaning of the operator $\mathbb{H}_s$ is to retain the s components with the highest absolute value and set the remaining components to zero. For convenience, note $u^* = \mathbb{LT}_{\upsilon\lambda}(x^* - \frac{1}{\upsilon}A^T\nabla f(Ax^* + b))$ and arrange the first s components according to their absolute values as follows:

$$u^*_{(1)} \geq u^*_{(2)} \geq \cdots \geq u^*_{(s)}. \tag{21}$$

Let $x^* = \mathbb{H}_s(u^*)$, $S_\lambda = supp(x^*)$. Then if $x_i^* \neq 0$, we have $x_i^* = u_i^*$.

If $|S_\lambda| = s$, for any $i \in S_\lambda$, we have

$$x_i^* = \mathbb{LT}_{\upsilon\lambda}(x_i^* - \frac{1}{\upsilon}A_i^T\nabla f(Ax^* + b)). \tag{22}$$

According to Definition 3 and (22), we have the following inequality:

$$|\frac{1}{\upsilon}A_i^T\nabla f(Ax^* + b)| = |x_i^* - (x_i^* - \frac{1}{\upsilon}A_i^T\nabla f(Ax^* + b))| \leq \kappa(\upsilon\lambda), \tag{23}$$

that is $|A_i^T\nabla f(Ax^* + b)| \leq \upsilon\kappa(\upsilon\lambda)$. For any $i \notin S_\lambda$, we have $x_i^* = 0$ and

$$|u_i^*| = |\mathbb{LT}_{\upsilon\lambda}(-\frac{1}{\upsilon}A_i^T\nabla f(Ax^* + b))| \leq x^*_{(s)}. \tag{24}$$

Then the following inequality holds:

$$\begin{aligned} |\frac{1}{\upsilon}A_i^T\nabla f(Ax^* + b)| &\leq |\mathbb{LT}_{\upsilon\lambda}(-\frac{1}{\upsilon}A_i^T\nabla f(Ax^* + b)) + \frac{1}{\upsilon}A_i^T\nabla f(Ax^* + b)| \\ &\quad + |\mathbb{LT}_{\upsilon\lambda}(-\frac{1}{\upsilon}A_i^T\nabla f(Ax^* + b))| \\ &\leq \kappa(\upsilon\lambda) + |u_i^*| \\ &\leq \kappa(\upsilon\lambda) + |x^*_{(s)}|, \end{aligned} \tag{25}$$

that is $|A_i^T\nabla f(Ax^* + b)| \leq \upsilon\kappa(\upsilon\lambda) + \upsilon|x^*_{(s)}|$.

If $|S_\lambda| < s$, we have $x^*_{(s)} = u^*_{(s)} = 0$ and $x^* = \mathbb{LT}_{\upsilon\lambda}(x^* - \frac{1}{\upsilon}A^T\nabla f(Ax^* + b))$. According to Definition 2, it is not difficult to obtain

$$\|A^T\nabla f(Ax^* + b)\|_\infty \leq \upsilon\kappa(\upsilon\lambda) + \upsilon|x^*_{(s)}|. \tag{26}$$

Combining (25) and (26), we get the result. □

To solve problem (5), we introduce an augmented Lagrangian function:

$$L_\rho(x, y, z) = f(y) - \langle z, Ax + b - y\rangle + \frac{\rho}{2}\|Ax + b - y\|_2^2, \tag{27}$$

where z is a Lagrange multiplier and ρ is a penalty factor. Then we establish the following model:

$$\min L_\rho(x, y, z) \quad s.t. \quad \|x\|_0 \leq s. \tag{28}$$

Regarding to this problem, we give the definitions of P-stationary point and LTP-stationary point.

Definition 8. A pairs of vector (x^*, y^*) is called a P-stationary point of (28) if there exists a vector z^* and constant $\upsilon > 0$ to satisfy the relation

$$\begin{cases} \nabla f(y^*) + z^* = 0 \\ x^* \in \mathbb{H}_s(x^* - \frac{1}{\upsilon}A^T\nabla f(Ax^* + b)). \\ Ax^* + b - y^* = 0 \end{cases} \tag{29}$$

Proposition 1 (*[54]*)**.** *Let the point x^* satisfy $\|x^*\|_0 \leq s$ and suppose that matrix A is row full rank. Then the following proposition holds true.*

(1) Suppose that x^* is a local minimizer of (4). Then there exists a vector z^* such that (x^*, y^*) is a P-stationary point of (28).

(2) Let f be a convex function. Suppose that (x^*, y^*) is a P-stationary point of (28). If $\|x^*\|_0 < s$, then x^* is the global minimizer of (4); otherwise, x^* is the local minimizer of (4);

Proposition 1 points out the relationship between the optimal solution of (4) and the P-stationary point of (28). Obviously, if (x^*, y^*) is the P-stationary point of (28), then x^* is also an L-stationary point of (4). This means that we only need to solve problem (28) to obtain the L-stationary point of problem (4).

Corresponding to the LT-stationary point, we define the following LTP-stationary point for the Lagrangian model with sparse constraint (28).

Definition 9. A pairs of vector (x^*, y^*) is called an LTP-stationary point of (28) if there exists a vector z^* and constant $\upsilon > 0$ satisfies the relation

$$\begin{cases} \nabla f(y^*) + z^* = 0 \\ x^* \in \mathbb{H}_s \circ \mathbb{LT}_{\upsilon\lambda}(x^* - \frac{1}{\upsilon} A^T \nabla f(Ax^* + b)). \\ Ax^* + b - y^* = 0 \end{cases} \tag{30}$$

Proposition 2. *Let the point x^* satisfy $\|x^*\|_0 \le s$. Then we have the following results.*

(1) If pairs of vector (x^*, y^*) is the LTP-stationary point of problem (28), then x^* is also the LT-stationary point of problem (4);
(2) Assume that $(\bar{x}, \bar{y})$ and (x^*, y^*) are the P-stationary point and LTP-stationary point of problem (28), respectively. If $\kappa(\upsilon\lambda) \to 0$, then the LTP-stationary point will approach the P-stationary point.

Proof. (1) The result of the proposition is obvious. Both the L-stationary point and the P-stationary point satisfy $x^* = \mathbb{H}_s(x^* - \frac{1}{\upsilon} A^T \nabla f(Ax^* + b))$.

(2) According to Lemmas 2 and 3, when $\kappa(\upsilon\lambda)$ tends to 0, x^* tending to $\bar{x}$. Also, because of $\|\bar{y} - y^*\| \le \|A\|_2 \|\bar{x} - x^*\|$, y^* tends to $\bar{y}$. This means that if $\kappa(\upsilon\lambda) \to 0$, then the LTP-stationary will approach the P-stationary. □

Combining Propositions 1 and 2, we just need to find an LTP-stationary point of (28) with a small $\kappa(\upsilon\lambda)$, which can be regarded as a solution to (4).

4. Limited shrinkage thresholding ADMM algorithm

4.1. Algorithm description

In this section, by using the LST operator and the ADMM technique, we propose an iterative limited shrinkage thresholding ADMM Algorithm (ILSTAT-ADMM), which is described in Algorithm 1. To further enhance the efficiency of this algorithm, we introduce an accelerated variant of ILSTAT-ADMM by incorporating the technique of inertial acceleration.

The standard iteration scheme of ADMM is outlined as follows:

$$x^{k+1} = \arg\min_x \left\{ L_\rho(x, y^k, z^k) : \|x\|_0 \le s \right\}, \tag{31}$$

$$y^{k+1} = \arg\min_y L_\rho(x^{k+1}, y, z^k), \tag{32}$$

$$z^{k+1} = z^k + \rho(y^{k+1} - Ax^{k+1} - b). \tag{33}$$

In fact, problem (31) is equivalent to

$$\begin{aligned} x^{k+1} &= \arg\min_x \left\{ L_\rho(x, y^k, z^k) : \|x\|_0 \le s \right\} \\ &= \arg\min_x \left\{ \frac{\rho}{2} \|Ax + b - y^k - \frac{z^k}{\rho}\|_2^2 : \|x\|_0 \le s \right\}. \end{aligned} \tag{37}$$

Let $b^k = y^k + \frac{z^k}{\rho} - b$, then

$$x^{k+1} = \arg\min_x \left\{ \frac{\rho}{2} \|Ax - b^k\|_2^2 : \|x\|_0 \le s \right\}, \tag{38}$$

which is a common sparse recovery problem. We use the iterative format of Algorithm 2 in [26] to solve this subproblem, i.e, for any $j \in \{1, 2, \dots, J\}$,

$$x^{k,j+1} \in \mathbb{H}_s \circ \mathbb{LT}_{\upsilon\lambda}(x^{k,j} - \frac{1}{\upsilon} A^T (Ax^{k,j} - b^k)). \tag{39}$$

According to Assumption 1, let $x^{k+1} = \lim_{J\to\infty} x^{k,J}$, then we have

$$x^{k+1} \in \mathbb{H}_s \circ \mathbb{LT}_{\upsilon\lambda}(x^{k+1} - \frac{1}{\upsilon} A^T (Ax^{k+1} - b^k)). \tag{40}$$

Algorithm 1 ILSTAT with ADMM

Input: A coercive, L_f-smooth with a lower bound function $f(Ax + b)$, a matrix A that satisfies $3s$-RIP, regularization parameter $\lambda > 0$, termination parameter $\varepsilon > 0$, sparse parameters $s \in \mathbb{N}$, random initial point $\omega^0 = (x^0, y^0)$, multiplier z^0 and number of iterations $k = 0$;
Output: $\omega^k = (x^k, y^k)$;
while $\|\omega^{k+1} - \omega^k\|_2 > \varepsilon$ **do**
Step 1: Let $b^k = y^k + \frac{z^k}{\rho} - b$, update x^k as follows

$$x^{k+1} \in \arg\min_x \left\{ L_\rho(x, y^k, z^k) : \|x\|_0 \le s \right\}. \tag{34}$$

Step 2: Update y^k as follows

$$y^{k+1} = \arg\min_y L_\rho(x^{k+1}, y, z^k). \tag{35}$$

Step 3: Update multiplier z^k

$$z^{k+1} = z^k + \rho(y^{k+1} - Ax^{k+1} - b). \tag{36}$$

Step 4: $k = k + 1$.
end while

.

Furthermore, according to the optimality condition and Eq. (36), we have

$$\nabla_y L_\rho(x^{k+1}, y^{k+1}, z^k) = \nabla f(y^{k+1}) + z^{k+1} = 0. \tag{41}$$

Remark 2. The first step of iteration can be seen as performing sparse recovery, and the second step is to minimize the objective function. To solve Eq. (35), we can employ the proximal linear minimization method.

To accelerate the convergence rate, we introduce the technology of inertial acceleration into the above proposed algorithm. Inertial acceleration technology can mitigate the oscillations caused by standard gradient descent, providing a more stable and globally optimal descent direction update vector through weighted averaging of historical gradient information. Drawing upon the methodology from [55], we integrate inertial acceleration into our approach to develop an accelerated algorithm (see Algorithm 2).

4.2. Convergence analysis

In this subsection, we aim to demonstrate two key properties of Algorithm 1. First, we will establish that when the parameter ρ is sufficiently large, Algorithm 1 satisfies the sufficient descent condition. Secondly, utilizing this property, we will demonstrate that the accumulation points of the sequence $\{x^k, y^k\}$ converge to an LTP-stationary point.

Lemma 4. *Assume that $f(Ax+b)$ is an L_f-smooth function and matrix A satisfies the 2s-RIP with δ_{2s}. Let $\rho > 0$ be the penalty factor in the augmented Lagrangian function and $\upsilon^{-1} > 0$ be the step length in the iterative process of x^k. Denote*

$$\eta_1 = \frac{\rho(1 - \delta_{2s})}{2} - \upsilon s, \ \eta_2 = \frac{\rho - 3L_f}{2} - \frac{L_f^2}{\rho}, \tag{47}$$

$$\mu = \min\left\{\eta_1, \eta_2\right\} - \sqrt{2s}\upsilon \frac{\kappa(\upsilon\lambda)}{\varepsilon} > 0. \tag{48}$$

Then we have

$$L_\rho(x^k, y^k, z^k) - L_\rho(x^{k+1}, y^{k+1}, z^{k+1}) \ge \mu \|\omega^{k+1} - \omega^k\|_2^2, \tag{49}$$

where $\omega^k = (x^k, y^k)$.

Algorithm 2 ILSTAT with iADMM

Input: A coercive, L_f-smooth with a lower bound function $f(Ax+b)$, a matrix A that satisfies $3s$-RIP, regularization parameter $\lambda > 0$, termination parameter $\varepsilon > 0$, sparse parameters $s \in \mathbb{N}$, random initial point $\omega^0 = (x^0, y^0)$, multiplier z^0 ,inertial parameters δ, ζ, and number of iterations $k = 0$.

Output: $\omega^k = (x^k, y^k)$;

while $\|\omega^{k+1} - \omega^k\| > \varepsilon$ **do**

Step 1: Let $b^k = y^k + \frac{z^k}{\rho} - b$, update x^k as follows

$$\hat{x}^k = x^k + \delta(x^k - x^{k-1}). \tag{42}$$

Using $\hat{x}^k$ as the initial point, and using iterative format (39)

$$x^{k+1} = \arg\min_x \left\{ \frac{\rho}{2}\|Ax - b^k\|_2^2 : \|x\|_0 \le s \right\}. \tag{43}$$

Step 2: Update y as follows

$$\hat{y}^k = y^k + \zeta(y^k - y^{k-1}), \tag{44}$$

$$y^{k+1} = \arg\min_y \left\{ h(y) + \frac{\rho}{2}\|Ax^k + b - y\|_2^2 \right\}, \tag{45}$$

where $h(y) = \langle y - \hat{y}^k, z^k + \nabla f(\hat{y}^k)\rangle + \frac{L_f}{2}\|y - \hat{y}^k\|_2^2$.

Step 3: Update multiplier

$$z^{k+1} = z^k + \rho(y^{k+1} - Ax^{k+1} - b). \tag{46}$$

Step 4: $k = k + 1$.

end while

Proof. Let $L^k = L_\rho(x^k, y^k, z^k)$, then

$$\begin{aligned} L^k - L^{k+1} &= L_\rho(x^k, y^k, z^k) - L_\rho(x^{k+1}, y^k, z^k) \\ &+ L_\rho(x^{k+1}, y^k, z^k) - L_\rho(x^{k+1}, y^{k+1}, z^k) \\ &+ L_\rho(x^{k+1}, y^{k+1}, z^k) - L_\rho(x^{k+1}, y^{k+1}, z^{k+1}). \end{aligned} \tag{50}$$

For convenience, we note

$$H_\rho(x, y, z) = -\langle z, Ax + b - y\rangle + \frac{\rho}{2}\|Ax + b - y\|_2^2. \tag{51}$$

Let $h^k = \nabla_x H_\rho(x^{k+1}, y^k, z^k)$, $S^k = supp(x^k)$ and $A^k = L_\rho(x^k, y^k, z^k) - L_\rho(x^{k+1}, y^k, z^k)$, then for any $j \in S^k \cup S^{k+1}$, we have

$$\begin{aligned} A^k &= \langle h^k, x^k - x^{k+1}\rangle + \frac{\rho}{2}\left\|A(x^k - x^{k+1})\right\|_2^2 \\ &\overset{(11)}{\ge} -|\langle h^k, x^k - x^{k+1}\rangle| + \frac{\rho(1-\delta_{2s})}{2}\|x^k - x^{k+1}\|_2^2 \\ &= -\sum_{j\in S^k\cup S^{k+1}} |h_j^k(x_j^k - x_j^{k+1})| + \frac{\rho(1-\delta_{2s})}{2}\|x^k - x^{k+1}\|_2^2 \\ &\ge -\upsilon\kappa(\upsilon\lambda)\sum_{j\in S^k\cup S^{k+1}} |x_j^k - x_j^{k+1}| + \frac{\rho(1-\delta_{2s})}{2}\|x^k - x^{k+1}\|_2^2 \\ &- \upsilon|x_{(s)}^{k+1}|\sum_{j\in S^k\cap S^{k+1}} |x_j^k - x_j^{k+1}| \\ &\ge -\sqrt{2s}\upsilon\kappa(\upsilon\lambda)\|x^{k+1} - x^k\|_2 + \eta_1\|x^k - x^{k+1}\|_2^2 \\ &\ge -\sqrt{2s}\upsilon\kappa(\upsilon\lambda)\|\omega^{k+1} - \omega^k\|_2 + \eta_1\|x^k - x^{k+1}\|_2^2 \\ &\ge -\sqrt{2s}\upsilon\frac{\kappa(\upsilon\lambda)}{\varepsilon}\|\omega^{k+1} - \omega^k\|_2^2 + \eta_1\|x^k - x^{k+1}\|_2^2. \end{aligned} \tag{52}$$

The second inequality is derived from Eq. (40) and Lemma 3. Meanwhile, the last inequality is established based on the following fact: if $\|\omega^{k+1} - \omega^k\|_2 \ge \varepsilon$, then it follows that $\|\omega^{k+1} - \omega^k\|_2 \le \frac{1}{\varepsilon}\|\omega^{k+1} - \omega^k\|_2^2$.

Since f is an L_f-smooth function, according to Lemma 1, we have

$$f(y^{k+1}) \le f(y^k) + \langle \nabla f(y^k), y^{k+1} - y^k\rangle + \frac{L_f}{2}\|y^{k+1} - y^k\|_2^2. \tag{53}$$

Let $B^k = L_\rho(x^{k+1}, y^k, z^k) - L_\rho(x^{k+1}, y^{k+1}, z^k)$, then

$$\begin{aligned} B^k &= f(y^k) - f(y^{k+1}) + H_\rho(x^{k+1}, y^k, z^k) - H_\rho(x^{k+1}, y^{k+1}, z^k) \\ &\overset{(53)}{\ge} \frac{\rho - L_f}{2}\|y^{k+1} - y^k\|_2^2 + \langle y^k - y^{k+1}, \nabla_y L_\rho(x^{k+1}, y^{k+1}, z^k)\rangle \\ &+ \langle y^k - y^{k+1}, \nabla f(y^k) - \nabla f(y^{k+1})\rangle \\ &\overset{(12)(41)}{\ge} \frac{\rho - 3L_f}{2}\|y^{k+1} - y^k\|_2^2. \end{aligned} \tag{54}$$

According to (12) and (41), we have

$$\|z^{k+1} - z^k\|_2 = \|\nabla f(y^{k+1}) - \nabla f(y^k)\|_2 \le L_f\|y^{k+1} - y^k\|_2. \tag{55}$$

Let $C^k = L_\rho(x^{k+1}, y^{k+1}, z^k) - L_\rho(x^{k+1}, y^{k+1}, z^{k+1})$, then

$$\begin{aligned} C^k &= \langle z^{k+1} - z^k, Ax^{k+1} + b - y^{k+1}\rangle \\ &\overset{(36)}{=} -\frac{1}{\rho}\|z^{k+1} - z^k\|_2^2 \\ &\ge -\frac{L_f^2}{\rho}\|y^{k+1} - y^k\|_2^2. \end{aligned} \tag{56}$$

The combination of inequalities (52), (54) and (56) leads to the conclusion. □

Theorem 1. *If f is a coercive function with a lower bound, then $\{\omega^k\} = \{(x^k, y^k)\}$ is a bounded sequence and satisfies*

$$\sum_{k=0}^{\infty}\|\omega^{k+1} - \omega^k\|_2^2 < \infty, \tag{57}$$

$$\lim_{k\to\infty}\|\omega^{k+1} - \omega^k\|_2 = 0. \tag{58}$$

Proof. First, since $f(y)$ has a lower bound, $L_\rho(x, y, z)$ also possesses a lower bound. By Lemma 4, we can conclude that the sequence $L_\rho(x^k, y^k, z^k)$ converges and satisfies $L_\rho(x^k, y^k, z^k) \ge \lim_{k\to\infty} L_\rho(x^k, y^k, z^k) = L^*$. Therefore, we can establish that

$$\sum_{k=0}^{\infty}\|\omega^{k+1} - \omega^k\|_2^2 \le \frac{1}{\mu}(L_\rho(x^0, y^0, z^0) - L^*) < \infty. \tag{59}$$

From this inequality, it follows that $\lim_{k\to\infty}\|\omega^{k+1} - \omega^k\|_2 = 0$. Furthermore, suppose that the sequence $\{\omega^k\}$ is unbounded. Then, there exists a subsequence $\{\omega^{k_n}\}$ such that $\lim_{n\to\infty}\|\omega^{k_n}\|_2 = +\infty$. Given that $f(y)$ is a coercive function, it can be shown that $L_\rho(x, y, z)$ is also coercive with respect to (x, y). However, according to Lemma 4, we have $L_\rho(x^0, y^0, z^0) \ge L_\rho(x^k, y^k, z^k)$, which is a contradiction. Therefore, $\{\omega^k\} = \{(x^k, y^k)\}$ is a bounded sequence. □

Theorem 2. *If f is a coercive and L_f-smooth function with a lower bound, then the accumulation points of sequence $\{\omega^k\}$ satisfy (30), that is, they correspond to the LTP-stationary points of the problem (28).*

Proof. Combining Assumption 1 and Theorem 1, it is not difficult for us to draw this conclusion. □

Remark 3. Theorems 1 and 2 establish that ILSTAT-ADMM converges to an LTP-stationary point, which serves as an approximate solution to (28). Furthermore, the accuracy of this approximate solution improves as λ decreases.

4.3. Convergence rate analysis

Assume that matrix A satisfies the $3s$-RIP. Let $\bar{x}^k$ satisfy $b^k = A\bar{x}^k$. From [26], we know that if step size $1/\upsilon$ satisfies

$$(1 - \delta_{3s}) \le \upsilon < 2(1 - \delta_{3s}), \tag{60}$$

then

$$\|x^{k,j} - \bar{x}^k\|_2 \le r^J\|x^k - \bar{x}^k\|_2 + \frac{2\sqrt{2}s}{1-r}\kappa(\upsilon\lambda), \tag{61}$$

where $r = 2(1 - \frac{1}{v} + \frac{1}{v}\delta_{3s})$. Let $x^{k+1} = x^{k,J}$, $b^k = Ax^{k+1} + e^k$. From (61), we can obtain

$$\|e^k\|_2 \le \gamma\|x^{k+1} - \bar{x}^k\|_2, \tag{62}$$

where $\gamma = \|A\|_2$, the largest singular value of matrix A.

Lemma 5. *Let $\{x^k, y^k, z^k\}$ be the sequence generated by Algorithm 1, $\nabla L_\rho(x^{k+1}, y^{k+1}, z^{k+1}) = d^{k+1} = (d_x^{k+1}, d_y^{k+1}, d_z^{k+1})$. We have*

$$\|d^{k+1}\|_2 \le \rho\gamma\|e^k\|_2 + M\|\omega^{k+1} - \omega^k\|_2, \tag{63}$$

where $M = L_f(1 + \frac{1}{\rho} + \gamma) + \rho\gamma$.

Proof. Firstly, we have

$$d_x^{k+1} = \nabla_x L_\rho(x^{k+1}, y^{k+1}, z^{k+1}) = -\rho A^T(Ax^{k+1} - b^{k+1}). \tag{64}$$

From (62), we can obtain

$$\|d_x^{k+1}\|_2 \le \rho\|e^k\| + \rho\gamma\|b^{k+1} - b^k\|_2, \tag{65}$$

where $b^k = y^k + \frac{z^k}{\rho} - b$. And we have

$$\|b^{k+1} - b^k\|_2 \le \gamma(\rho + L_f)\|\omega^{k+1} - \omega^k\|_2. \tag{66}$$

According to (41), (36), we have

$$\|d_y^{k+1}\|_2 = \|z^{k+1} - z^k\|_2 \le L_f\|\omega^{k+1} - \omega^k\|_2, \tag{67}$$

$$\|d_z^{k+1}\| = \frac{L_f}{\rho}\|\omega^{k+1} - \omega^k\|_2. \tag{68}$$

To sum up, it is not difficult for us to obtain

$$\|d^{k+1}\|_2 \le \rho\gamma\|e^k\|_2 + M\|\omega^{k+1} - \omega^k\|_2. \quad \square \tag{69}$$

Lemma 6 (*Uniformized KŁ property [44]*). *Let Ω be a compact set and let $\sigma : \mathbb{R}^d \to (-\infty, \infty]$ be a proper and l.s.c function. Assume that σ is constant on Ω and satisfies the KŁ property at each point of Ω. Then, there exist $\epsilon > 0$, $\eta > 0$, and $\varphi \in \Phi_\eta$ such that for all $u \in \Omega$ and all u in the following intersection:*

$$\{u \in \mathbb{R}^d : dist(u, \Omega) < \epsilon\} \cap [\sigma(u) < \sigma(u) < \sigma(u) + \eta], \tag{70}$$

we have

$$\varphi'(\sigma(u) - \sigma(u))dist(0, \partial\sigma(u)) \ge 1. \tag{71}$$

Based on the uniformized KŁ property and the above lemmas, we give the analysis on the convergence rate of Algorithm 1.

Theorem 3. *Let f be a KŁ function, $\{x^k, y^k, z^k\}$ be the sequence generated by Algorithm 1, $\{x^*, y^*, z^*\}$ be the accumulation point of $\{x^k, y^k, z^k\}$ and denote $\psi(s) = cs^{1-\theta}$ and $R_k = L_\rho(x^k, y^k, z^k) - L_\rho(x^*, y^*, z^*)$. Then there exists a K, and when $k > K$, it satisfies*

$$|R_k|^\theta \le (1 - \theta)c\|d^k\|_2. \tag{72}$$

Moreover,

(1) if $\theta = 0$, then R_k converges to zero in a finite number of iterations.
(2) if $\theta \in (0, \frac{1}{2}]$, then for all $k > K$, it holds

$$R_k \le \frac{\max\{R_i : 1 \le i \le K\}}{(1 + C_1 R_K^{2\theta-1})^{k-K}} + \frac{2\sqrt{2}C_2\gamma s\kappa(v\lambda)}{(1 - r)C_1 R_K^{2\theta-1}}. \tag{73}$$

(3) if $\theta \in (\frac{1}{2}, 1)$, when ρ is sufficiently large, there exists a $\hat{\delta} > 0$ such that for all $k \ge K$ holds

$$R_k \le \left(\frac{1}{\hat{\delta}(k - K) + R_K^{1-2\theta}}\right)^{\frac{1}{2\theta-1}}. \tag{74}$$

Proof. Let Ω be the set of all gathering points of $\{x^k, y^k, z^k\}$. Then for any $\epsilon > 0$, there exists a k_1, when $k > k_1$, we have

$$dist((x^k, y^k, z^k), \Omega) < \epsilon. \tag{75}$$

In addition, we have $\lim_{k\to\infty} L_\rho(x^k, y^k, z^k) = L_\rho(x^*, y^*, z^*)$. $L_\rho(x^k, y^k, z^k)$ is a non-increasing sequence. Therefore for any $\eta > 0$, there exits a k_0, when $k > k_0$, we have

$$L_\rho(x^*, y^*, z^*) < L_\rho(x^k, y^k, z^k) < L_\rho(x^*, y^*, z^*) + \eta. \tag{76}$$

Let $K = \max\{k_0, k_1\}$, $L_\rho^k = L_\rho(x^k, y^k, z^k)$ and $L_\rho^* = L_\rho(x^*, y^*, z^*)$. If $k > K$, we have

$$\psi'(L_\rho^k - L_\rho^*)\ dist(0, \nabla L_\rho(x^k, y^k, z^k)) \ge 1, \tag{77}$$

that is

$$|R_k|^\theta \le (1 - \theta)c\|d^k\|_2. \tag{78}$$

According to Lemmas 4 and 5, we have

$$C_1 R_k^{2\theta} \le R_{k-1} - R_k + C_2\|e^{k-1}\|_2, \tag{79}$$

where $C_1 = \frac{\mu}{2M^2(1-\theta)^2c^2}$, $C_2 = \frac{\mu\rho^2\gamma^2}{M^2}$.

(1) Let $\theta = 0$. If $R^k > 0$ for all $k > K$, then

$$R_{k-1} - R_k \ge C_1 - C_2\|e^{k-1}\|_2. \tag{80}$$

In fact, form (61) and (62), when J goes to infinity, $\|e^{k-1}\|_2 \le \frac{2\gamma\sqrt{2}s}{1-r}\kappa(v\lambda)$. If λ is small enough, $\|e^{k-1}\|_2$ is small enough as well, which means that we can always find a λ that makes $C_1 - C_2\|e^{k-1}\|_2 > 0$. However, the left side of (80) approaches zero, which leads to a contradiction. Additionally, since R_k is monotonically decreasing, therefore we have $R^k = 0$ for any $k > K$.

(2) If $\theta \in (0, \frac{1}{2}]$, then $2\theta - 1 < 0$, we have

$$C_1 R_K^{2\theta-1} R_k \le C_1 R_k^{2\theta} \le R_{k-1} - R_k + C_2\|e^{k-1}\|_2. \tag{81}$$

For convenience, note $t = \frac{1}{1 + C_1 R_K^{2\theta-1}} < 1$. For any $k > K$, we have

$$\begin{aligned} R_k &\le tR_{k-1} + C_2 t\|e^{k-1}\|_2 \le tR_{k-1} + \frac{2\sqrt{2}C_2\gamma s}{1 - r}\kappa(v\lambda)t \\ &\le t^2 R_{k-2} + \frac{2\sqrt{2}C_2\gamma s}{1 - r}\kappa(v\lambda)(t + t^2) \le \dots \\ &\le t^{k-K} R_K + \frac{2\sqrt{2}C_2\gamma s}{1 - r}\kappa(v\lambda)\sum_{i=1}^{\infty} t^i \\ &= t^{k-K} R_K + \frac{2\sqrt{2}C_2\gamma s}{1 - r}\kappa(v\lambda)\frac{t}{1 - t} \\ &= \frac{R_K}{(1 + C_1 R_K^{2\theta-1})^{k-K}} + \frac{2\sqrt{2}C_2\gamma s\kappa(v\lambda)}{(1 - r)C_1 R_K^{2\theta-1}}. \end{aligned} \tag{82}$$

Then we get

$$R_k \le \frac{\max\{R_i : 1 \le i \le K\}}{(1 + C_1 R_K^{2\theta-1})^{k-K}} + \frac{2\sqrt{2}C_2\gamma s\kappa(v\lambda)}{(1 - r)C_1 R_K^{2\theta-1}}. \tag{83}$$

(3) If $\theta \in (\frac{1}{2}, 1)$, we can obtain $C_1 \le (R_{k-1} - R_k)R_k^{-2\theta} + C_2\|e^{k-1}\|_2 R_k^{-2\theta}$. Denote $h(s) = s^{-2\theta}$ for $s \in \mathbb{R}_+$. Obviously, $h(R_{k-1}) \le h(R_k)$ for all $k > K$. We go on the proof in the following two cases. First, let

$r_0 \in (1,\infty)$ such that $h(R_k) \le r_0 h(R_{k-1}), \forall k > K$. We can obtain

$$\begin{aligned} C_1 &\le (R_{k-1}-R_k)R_k^{-2\theta} + C_2\|e^{k-1}\|_2 R_k^{-2\theta} \\ &\le r_0(R_{k-1}-R_k)h(R_{k-1}) + C_2\|e^{k-1}\|_2 R_k^{-2\theta} \\ &= r_0 h(R_{k-1})\int_{R_k}^{R_{k-1}} 1ds + C_2\|e^{k-1}\|_2 R_k^{-2\theta} \\ &\le r_0\int_{R_k}^{R_{k-1}} h(s)ds + C_2\|e^{k-1}\|_2 R_k^{-2\theta} \\ &= r_0\int_{R_k}^{R_{k-1}} s^{-2\theta}ds + C_2\|e^{k-1}\|_2 R_k^{-2\theta} \\ &= \frac{r_0}{1-2\theta}(R_{k-1}^{1-2\theta} - R_k^{1-2\theta}) + C_2\|e^{k-1}\|_2 R_k^{-2\theta}. \end{aligned} \tag{84}$$

Let $\delta_1 = \frac{C_1(2\theta-1)}{r_0}$, we have

$$\delta_1 \le R_k^{1-2\theta} - R_{k-1}^{1-2\theta} + \frac{C_2(2\theta-1)}{r_0}\|e^{k-1}\|_2 R_k^{-2\theta}. \tag{85}$$

Next, we consider the case $r_0 h(R_{k-1}) \le h(R_k)$. This yields $r_0 R_{k-1}^{-2\theta} \le R_k^{-2\theta}$, that is $r_0^{-\frac{1}{2\theta}} R_{k-1} \ge R_k$. Let $q = 1-2\theta < 0$, we have

$$(r_0^{-\frac{1}{2\theta}} - 1)R_{k-1}^q \le R_k^q - R_{k-1}^q. \tag{86}$$

By the fact $r_0^{-\frac{1}{2\theta}} - 1 > 0$ and $R_k \to 0_+$ as $k \to \infty$, there exists $\delta_2 > 0$ such that $(r_0^{-\frac{1}{2\theta}} - 1)R_{k-1}^q \ge \delta_2$. Therefore we obtain

$$0 < \delta_2 \le R_k^q - R_{k-1}^q, \forall k > K. \tag{87}$$

Choose $\delta' = \max\{\delta_1, \delta_2\}$, we can obtain

$$\delta' \le R_k^{1-2\theta} - R_{k-1}^{1-2\theta} + \frac{C_2(2\theta-1)}{r_0}\|e^{k-1}\|_2 R_k^{-2\theta}. \tag{88}$$

Summing this inequality from $K+1$ to some $k \ge K+1$, we get

$$\sum_{i=K+1}^{k}\delta' \le \sum_{i=K+1}^{k}(R_i^{1-2\theta} - R_{i-1}^{1-2\theta} + \frac{C_2(2\theta-1)}{r_0}\|e^{i-1}\|_2 R_i^{-2\theta}). \tag{89}$$

Thus,

$$\delta'(k-K) \le R_k^{1-2\theta} - R_K^{1-2\theta} + \frac{C_2(2\theta-1)}{r_0}\sum_{i=K+1}^{k}\|e^{i-1}\|_2 R_i^{-2\theta}. \tag{90}$$

According to Lemma 5 and the termination conditions of the algorithm, we have

$$\sum_{i=K+1}^{k}\|e^{i-1}\|_2 R_i^{-2\theta} \le (k-K)\frac{2\gamma\sqrt{2}s}{\mu\varepsilon^2(1-r)}\kappa(\upsilon\lambda). \tag{91}$$

Hence, for all $k > K$, we have

$$R_k \le \left(\frac{1}{\hat{\delta}(k-K) + R_K^{1-2\theta}}\right)^{\frac{1}{2\theta-1}}, \tag{92}$$

where $\hat{\delta} = \delta' + \frac{C_2(1-2\theta)}{r_0}\frac{2\gamma\sqrt{2}s}{\mu\varepsilon^2(1-r)}\kappa(\upsilon\lambda)$. Combining the value of δ' and (47), we can conclude that when ρ is sufficiently large, $\hat{\delta} > 0$. □

Remark 4.

(1) Theorem 3 indicates that the convergence speed of the algorithm will be affected by ρ. If $\theta \in \left(0, \frac{1}{2}\right)$, from (73), if C_1 increases, the convergence rate will decrease; On the contrary, the convergence rate will increase. Since $C_1 = \frac{\mu}{2M^2(1-\theta)^2c^2}$, where μ is defined by (47). By substituting μ into C_1, we obtain

$$C_1 = \frac{\min\left\{\frac{\rho(1-\delta_{2s})}{2} - \upsilon s, \frac{\rho-3L_f}{2} - \frac{L_f^2}{\rho}\right\} - \sqrt{2s}\upsilon\frac{\kappa(\upsilon\lambda)}{\varepsilon}}{2M^2(1-\theta)^2c^2}. \tag{93}$$

This indicates that C_1 increases with the increase of ρ. Therefore, an increase in ρ will result in a decrease in convergence rate. For $\theta \in \left(\frac{1}{2}, 1\right)$, we also have a similar situation with $\hat{\delta}$.

(2) When ρ is too small, the update method of the Lagrange multiplier z in the ADMM algorithm is $z^{k+1} = z^k + \rho(y^{k+1} - Ax^{k+1} - b)$. By making a simple change to this inequality, we have

$$\frac{z^{k+1}-z^k}{\rho} = y^{k+1} - Ax^{k+1} - b. \tag{94}$$

Insufficient update step size of z may require more iterations to obtain the optimal solution. In addition, due to the appearance of ρ as the denominator, a too small ρ will amplify the parts that do not satisfy the constraints, which may cause the algorithm to fail to converge.

5. Numerical experiments

In this section, we conduct numerical experiments to evaluate the performance of our proposed algorithms. In order to verify the efficiency of our proposed algorithms, we selected some different LST operators shown in Table 1. In addition, taking the SCAD function as an example, we compared our algorithm with some classical algorithms and the latest algorithms. All numerical experiments are performed on a personal desktop computer with the following specifications: Intel Core i5-13400F, 2.50 GHz, and 32.00 GB of RAM.

To avoid the randomness of a single experiment, we need to conduct multiple repeated experiments and take the average of the experiments as the result. So we generate a set of classical Gaussian matrices $\{A_k\}$. And each Gaussian matrix A_k satisfies $A_k A_k^T = I$. For a matrix A, we randomly generate an s-sparse solution $x^* \in \mathbb{R}^n$ by randomly selecting s nonzero entries. The observation vector b is then generated according to

$$b = Ax^* + \sigma\epsilon, \tag{95}$$

where $\sigma \in \mathbb{R}$ represents the noise level, and $\epsilon \sim \mathcal{N}(0,1)$ denotes Gaussian noise. We use the relative error to measure the degree of approximation. Relative error is defined as follows:

$$RE = \frac{\|x^k - x^*\|_2}{\|x^*\|_2}, \tag{96}$$

where x^k is the current solution after k iterations using Algorithm 1 or Algorithm 2 and x^* is the true solution of the original sparse problem of (4).

5.1. Compressed sensing signal reconstruction

Set $m = 128, n = 512$, sparsity $s = 16$, stepsize $\upsilon = 1$ and maximum number of iterations $itermax = 50$. The initial point $\omega^0 = (0, -b)$, and $\rho = 1$ in Algorithm 1.

Firstly, we tested the sensitivity of different LST operators to λ, as shown in Fig. 1. We note that a larger λ is not necessarily advantageous, as it may affect the sparsity of x^k through $\kappa(\upsilon\lambda)$. Excessively large values of λ can lead to significant information loss in the signal prior to performing a hard thresholding operation. Choosing different λ values for different sparsity promoting functions is advisable.

In addition, the experimental results indicate that the ℓ_p-norm ($p = 0, 1/2, 1$) is more susceptible to the interference of λ compared to the other three sparsity promoting functions. Though the experimental results, choosing different λ values for different sparsity promoting functions is recommended. For the ℓ_p-norm ($p = 0, 1/2, 1$), λ can be selected from $[10^{-4}, 0.002]$. For the log-sum function, λ can be selected from $[10^{-4}, 0.02]$. For the SCAD and MCP functions, λ can be selected from $[10^{-4}, 0.05]$. Therefore, the selection of λ should be appropriately adjusted based on the size of x^k.

Secondly, we conducted compressed sensing signal reconstruction experiments using different LST operators under $\lambda = 0.005$, as shown in Figs. 2–4. Fig. 2 shows the convergence result at sparsity $s = 16$, which indicates that different LST operators have almost no effect on the convergence speed of the algorithm. Different LST operators mainly

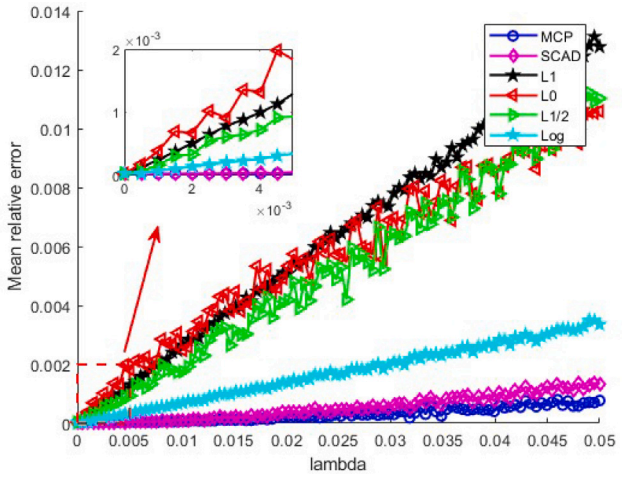


Fig. 1. Sensitivity to different λ of different LST operators.

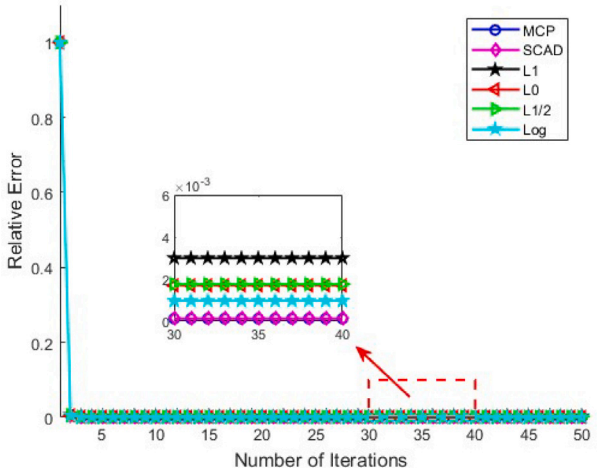


Fig. 2. Convergence behavior of different LST operators when $s = 16$, $\sigma = 0$, $\lambda = 0.005$.

affect the relative error of convergence. In addition, we independently repeat 100 experiments at different sparsity levels and use the success frequency as the success rate. The results are show in Fig. 3, which indicate that the MCP and SCAD functions can still maintain good reconstruction performance at different sparsity levels. The corresponding signal reconstruction results of a single experiment are shown in Fig. 4.

Next, we examined the influence of ρ on the performance of our algorithms, as show in Fig. 5. The results reveal that an excessively small ρ leads to degraded convergence performance, whereas an overly large ρ slows down the convergence rate, which is consistent with our previous theoretical analysis results.

After previous experiments, we selected the SCAD function as the sparsity promoting function and chose parameters $\lambda = 0.01$ and $\rho = 1$ for Algorithm 1. Additionally, we set the parameters $L = 0.9, \delta = \zeta = 1$ in the iADMM-ILSTAT algorithm and increase the number of iterations to 100. We compared the algorithm IHT [34], ILSTAT [26], and ALPA [49] with our algorithms. From Fig. 6, it demonstrates the advantage of introducing the LST operator in helping to converge to the true solution. In the case of noise intensity σ=0.5%, we conducted a single compression sensing experiment and the results are shown in Fig. 7.

Finally, under different sparsity levels, we conducted the experiments on the success rate of signal recovery. The success frequency

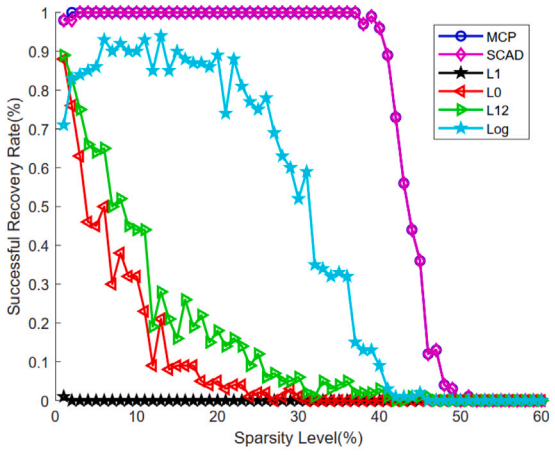


Fig. 3. Success rate of signal reconstruction under different sparsity levels when $\lambda = 0.005$.

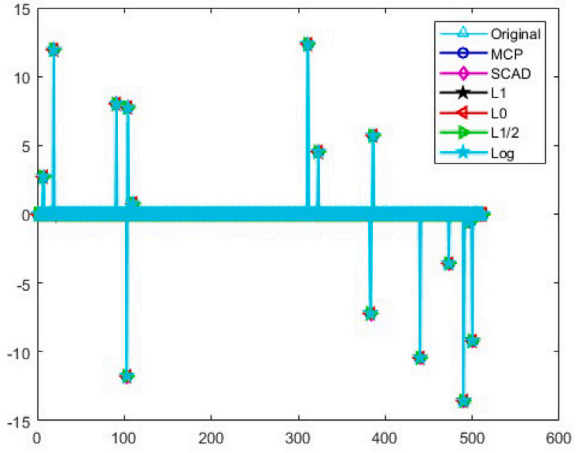


Fig. 4. Reconstructed signal using different LST operators when $\lambda = 0.005$, $s = 16$.

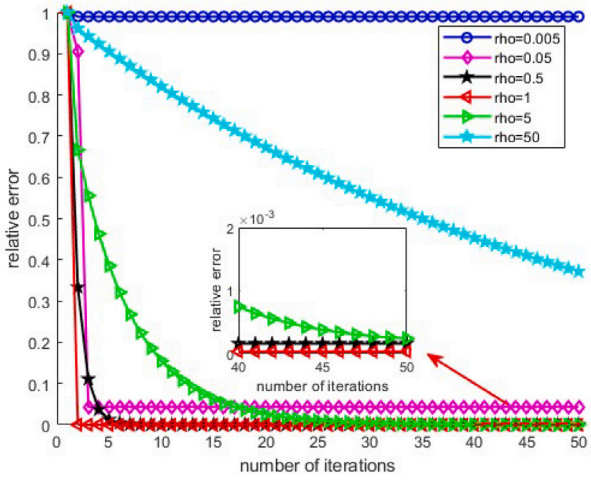


Fig. 5. Impact of ρ on the convergence rate when $\lambda = 0.002$.

is used as the success rate of signal recovery. When the relative error satisfies $RE \leq 10^{-3}$, it is considered that the signal recovery is successful. Repeat the experiment 100 times with $m = 128, n = 512$, and the numerical experimental results are shown in Fig. 8. It can be

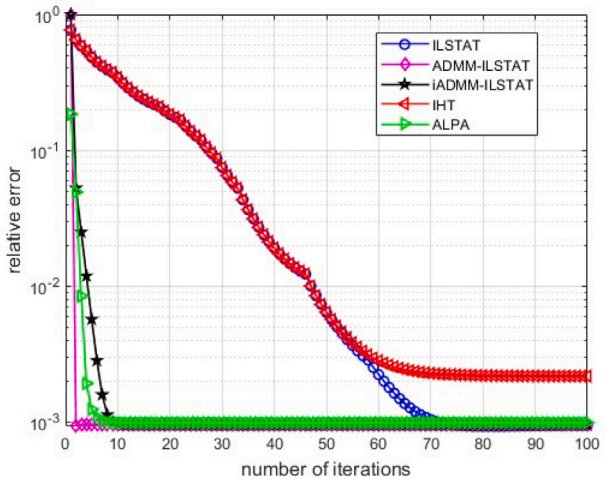


Fig. 6. Convergence behavior of different algorithms.

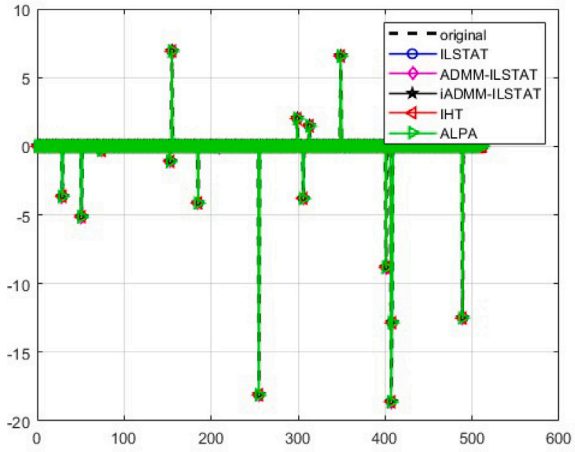


Fig. 7. Reconstructed signal using different algorithms when $\lambda = 0.002$, $s = 16$, $\sigma = 0.5\%$.

seen that the success rate of signal recovery has significantly improved compared to the ALPA and ILSTAT after combining the ADMM algorithm. Through Fig. 8(c) and Fig. 8(d), it can be observed that when λ is very small, ILSTAT and IHT almost overlap. Increasing λ appropriately can improve the recovery success rate of ILSTAT. In addition, from the perspective of whether it carries noise or not, ALPA has a fast iteration speed, but there are some unstable phenomena in the face of noise values compared with our algorithms.

5.2. Image deblurring

Let Φ be a sparse basis and B be a blurred image, H be a blurring kernel matrix, and $X \in \mathbb{R}^{m\times n}$ be an original image. For a two-dimensional image, we have the following sparse model:

$$B = HX + \sigma e \ s.t. \ \|\Phi X\|_0 \le s, \tag{97}$$

where e is a noise matrix composed of independent Gaussian noise.

Now, we conduct image recovery experiment. Chose discrete cosine transform (DCT) basis as the sparse basis. DCT is a transformation method that plays an important role in the fields of digital signal processing and image processing. It can transform signals from the time domain (or spatial domain) to a frequency domain expression centered around cosine basis functions, which is defined as follows:

$$\phi_{k,l}[m,n] = c_k \cdot c_l \cdot \cos\left(\frac{\pi}{N}\left(m+\frac{1}{2}\right)k\right) \cdot \cos\left(\frac{\pi}{N}\left(n+\frac{1}{2}\right)l\right), \tag{98}$$

Table 2
PSNR of baboon image.

Blurring kernel	ILSTAT	IST	ALPA	OUR
'gaussian' (20,5)	19.82	19.49	20.03	20.21
'disk' (8)	19.85	19.54	20.47	20.51
'motion' (20,45)	19.68	19.98	21.12	20.85
'average' (6)	20.89	20.33	21.13	22.92

Table 3
PSNR of woman image.

Blurring kernel	ILSTAT	IST	ALPA	OUR
'gaussian' (20,5)	21.27	21.90	22.24	22.56
'disk' (8)	21.95	21.38	22.53	22.33
'motion' (20,45)	21.65	22.17	23.40	23.65
'average' (6)	23.28	22.69	23.39	25.12

where N is the block size of the image, m and n are spatial domain coordinates and k and l are frequency domain coordinates. c_k and c_l are normalization coefficients defined as: $c_u = \begin{cases} \sqrt{\frac{1}{N}} & \text{when } u = 0 \\ \sqrt{\frac{2}{N}} & \text{when } u \ge 1 \end{cases}$, where $u \in \{k, l\}$.

Set the sparsity s to the top 15% of the coefficients of the image in the DCT domain, thresholding parameter $\lambda = 10^{-4}$, $\rho = 1$, $\sigma = 0.5\%$. The maximum number of iterations is set to 100, and we evaluate the quality of image restoration through the following indicators:

$$PSNR = 10\lg\frac{MAX^2}{MSE}, \tag{99}$$

where MAX represents the possible maximum value of image pixel values and MSE stands for the mean squared error which is defined as follows:

$$\text{MSE} = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}(X(i,j) - X^*(i,j))^2, \tag{100}$$

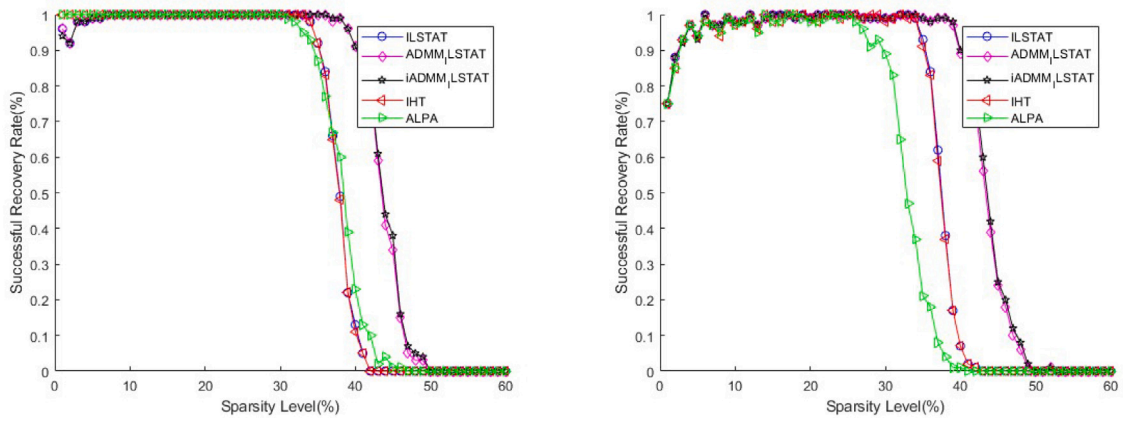
where X^* is the original image.

We conducted the simulation experiments using different blurring methods and compared the results with the iterative soft thresholding algorithm (IST), ALPA and ILSTAT. Here are several common blur kernels listed: Average (M): M represents the size of the blur kernel. Gaussian (M, σ): M represents the size of the blur kernel and σ represents the standard deviation of Gaussian distribution. Motion (L, θ): L represents the length of motion, θ represents the angle of motion. Disk (R): R represents the radius of the disk. For the image of 512×512 baboon image, we applied average blur with parameter (6), as shown in Fig. 9(a). For the image of 512×512 woman image, we applied Gaussian blur with parameters (20, 5), as shown in Fig. 9(b).

From Fig. 9, it can be observed that our algorithm achieves a good recovery effect compared to other algorithms. The corresponding PSNR values for these two images under different blurring methods are listed in Tables 2 and 3. We can observe that our algorithm consistently achieves good results in PSNR compared with other algorithms under different images and different blur methods. In some fuzzy methods, such as average blur, it provides better performance.

6. Conclusion

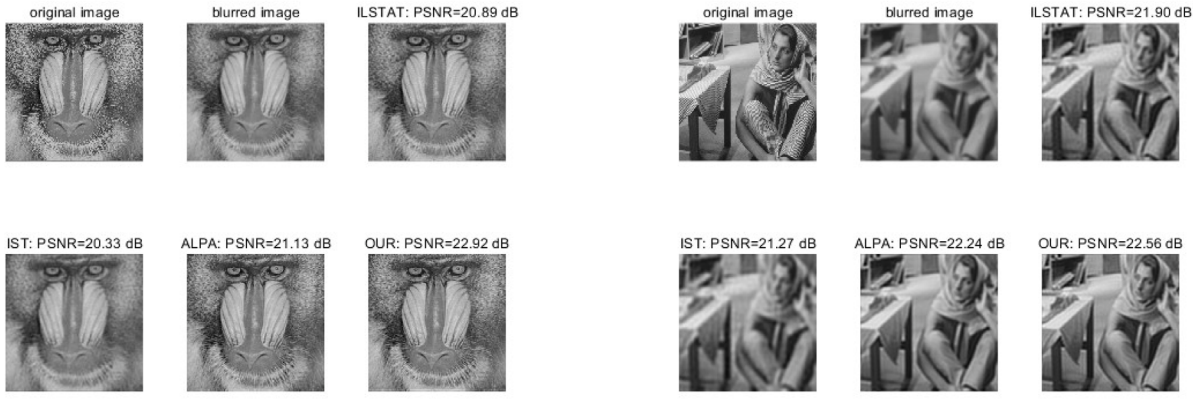
In this paper, we propose an iterative limited shrinkage thresholding ADMM algorithm for solving general non-convex problems with l_0-norm constraints. Then we explore the optimality conditions by utilizing the LST operator and defining the LTP-stationary point. Subsequently, we prove that the algorithm exhibits sufficient descent properties when the penalty parameter ρ is sufficiently large, and all limit points of the sequence converge to the LTP-stationary point. Furthermore, we analyze the convergence rate of the algorithm based on the



(a) $\lambda = 0.01, \sigma = 0$ (b) $\lambda = 0.01, \sigma = 0.5\%$

(c) $\lambda = 10^{-4}, \sigma = 0$ (d) $\lambda = 10^{-4}, \sigma = 0.5\%$

Fig. 8. Success rate of different algorithms in the recovery of a random sparse signal with different sparsity levels(%)



(a) Average blur (6) (b) Motion blur (20,10)

Fig. 9. Image deblurring performance.

RIP condition and the KŁ properties. Finally, we conduct experiments on compressive sensing signal reconstruction and image deblurring using the proposed algorithm. The experimental results demonstrate that the algorithm is more efficient in finding sparse solutions and improving the success rate of recovery.

CRediT authorship contribution statement

Yuan-Min Li: Writing – review & editing, Methodology, Conceptualization. **Hao Wang:** Writing – original draft, Methodology, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 62371364, and in part by Shaanxi Fundamental Science Research Project for Mathematics and Physics, China (Grant No. 23JSY038).

Thanks to the editor and the anonymous reviewers for their helpful and valuable comments and suggestions, which has greatly improved the clarity and quality of this paper.

Data availability

Data will be made available on request.

References

[1] D.L. Donoho, Compressed sensing, IEEE Trans. Inform. Theory 52 (4) (2006) 1289–1306.
[2] E. Candès, J. Romberg, T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, IEEE Trans. Inform. Theory 52 (2) (2006) 489–509.
[3] C.E. Shannon, Communication in the presence of noise, Proc. IRE 37 (1) (1949) 10–21.
[4] Y.-M. Li, et al., A novel image encryption algorithm based on compressive sensing and a two-dimensional linear canonical transform, Fractal Fract. 8 (92) (2024).
[5] Y. Eldar, G. Kutyniok, Compressed Sensing: Theory and Applications, Cambridge Univ. Press, Cambridge, U.K., 2012.
[6] T. Blumensath, Iterative thresholding for sparse approximations, J. Fourier Anal. Appl. 14 (5) (2008) 629–654.
[7] E. Soubies, B.F. Laure, G. Aubert, A continuous exact ℓ_0 penalty (CEL0) for least squares regularized problem, SIAM J. Imaging Sci. 8 (3) (2015) 1607–1639.
[8] H. Attouch, J. Bolte, B.F. Svaiter, Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods, Math. Program. 137 (1) (2013) 91–129.
[9] D. Bertsimas, A. King, R. Mazumder, Best subset selection via a modern optimization lens, Ann. Statist. 44 (2) (2016) 813–852.
[10] E. Candès, The restricted isometry property and its implications for compressed sensing, C. R. Math. 346 (9) (2008) 589–592.
[11] B.M. Wakin, P.S. Boyd, Enhancing sparsity by reweighted ℓ_1 minimization, J. Fourier Anal. Appl. 14 (2) (2008) 877–905.
[12] S.S. Chen, D.L. Donoho, M.A. Saunders, Atomic decomposition by basis pursuit, SIAM Rev. 43 (1) (2001) 129–159.
[13] Y.B. Zhao, D. Li, Reweighted ℓ_1 minimization for sparse solutions to underdetermined linear systems, SIAM J. Optim. 22 (3) (2012) 1065–1088.
[14] Z.B. Xu, X.Y. Chang, F.M. Xu, H. Zhang, $\ell_{1/2}$ regularization: A thresholding representation theory and a fast solver, IEEE Trans. Neural Netw. Learn. Syst. 23 (7) (2012) 1013–1027.
[15] C.H. Zhang, Nearly unbiased variable selection under minimax concave penalty, Ann. Statist. 38 (2) (2010) 894–942.
[16] J.Q. Fan, R.Z. Li, Variable selection via nonconcave penalized likelihood and its oracle properties, J. Amer. Statist. Assoc. 96 (456) (2001) 1348–1360.
[17] A. Prater-Bennette, L. Shen, E.E. Tripp, The proximity operator of the log-sum penalty, J. Sci. Comput. 93 (3) (2022) 67.
[18] V. Cerone, S.M. Fosson, D. Regruto, Fast sparse optimization via adaptive shrinkage, IFAC-PapersOnLine 56 (2) (2023) 10390–10395.
[19] J. Wang, Q. Ma, The variant of the iterative shrinkage-thresholding algorithm for minimization of the ℓ_1 over ℓ_∞ norms, Signal Process. 211 (2023) 109104.
[20] Z. He, Q. Shu, Y. Wang, J. Wen, A ReLU-based hard-thresholding algorithm for non-negative sparse signal recovery, Signal Process. 215 (2024) 109260.
[21] W. Bian, F. Wu, Accelerated smoothing hard thresholding algorithms for l_0 regularized nonsmooth convex regression problem, J. Sci. Comput. 96 (2) (2023).
[22] Y.-M. Li, T. Lei, Delayed neural network based on a new complementarity function for the NCP, Expert Syst. Appl. 251 (2024) 123980.
[23] K.-K. Wen, J.-X. He, P. Li, Sparse recovery using expanders via hard thresholding algorithm, Signal Process. 227 (2025) 109715.
[24] L. Shen, B.W. Suter, E.E. Tripp, Structured sparsity promoting functions, J. Optim. Theory Appl. 183 (2) (2019) 386–421.
[25] C. Soussen, J. Idier, J. Duan, D. Brie, Homotopy-based algorithms for ℓ_0-regularized least-squares, IEEE Trans. Signal Process. 63 (13) (2015) 3301–3316.
[26] Y.H. Hu, X.L. Hu, X.Q. Yang, On convergence of iterative thresholding algorithms to approximate sparse solution for composite nonconvex optimization, Math. Program. 211 (2025) 181–206.
[27] J.A. Tropp, C.A. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, IEEE Trans. Inform. Theory 53 (12) (2007) 4655–4666.
[28] S. Foucart, S. Subramanian, Iterative hard thresholding for low-rank recovery from rank-one projections, Linear Algebra Appl. 572 (2019) 117–134.
[29] S. Foucart, Hard thresholding pursuit: An algorithm for compressive sensing, SIAM J. Numer. Anal. 49 (6) (2011) 2543–2563.
[30] J.L. Bouchot, S. Foucart, P. Hitczenko, Hard thresholding pursuit algorithms: Number of iterations, Appl. Comput. Harmon. Anal. 41 (2) (2016) 412–435.
[31] D. Needell, J.A. Tropp, CoSaMP: Iterative signal recovery from incomplete and inaccurate samples, Appl. Comput. Harmon. Anal. 26 (3) (2009) 301–321.
[32] M. Xiang, Z.Y. Zhang, Fast recursive greedy methods for sparse signal recovery, IEEE Trans. Signal Process. 72 (2024) 4381–4394.
[33] W. Dai, O. Milenkovic, Subspace pursuit for compressive sensing signal reconstruction, IEEE Trans. Inform. Theory 55 (5) (2009) 2230–2249.
[34] A. Beck, Y.C. Eldar, Sparsity constrained nonlinear optimization: Optimality conditions and algorithms, SIAM J. Optim. 23 (3) (2013) 1480–1509.
[35] T. Blumensath, M.E. Davies, Normalized iterative hard thresholding: Guaranteed stability and performance, IEEE J. Sel. Top. Signal Process. 4 (2) (2010) 298–309.
[36] F. Wu, W. Bian, Accelerated iterative hard thresholding algorithm for l0 regularized regression problem, J. Global Optim. 76 (4) (2020) 819–840.
[37] J.D. Blanchard, J. Tanner, K. Wei, CGIHT: conjugate gradient iterative hard thresholding for compressed sensing and matrix completion, Inf. Inference: A J. IMA 4 (4) (2015) 289–327.
[38] S.L. Zhou, N.H. Xiu, H.D. Qi, Global and quadratic convergence of Newton hard-thresholding pursuit, J. Mach. Learn. Res. 22 (12) (2021) 1–45.
[39] Y.B. Zhao, Optimal k-thresholding algorithms for sparse optimization problems, SIAM J. Optim. 30 (1) (2020) 31–55.
[40] R. Glowinski, A. Marroco, Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de dirichlet non linéaires, Ann. l'Inst. H. Poincaré Anal. Numér. 280 (1975) 287–290.
[41] G. Daniel, M. Bertrand, A dual algorithm for the solution of nonlinear variational problems via finite element approximation, Comput. Math. Appl. 2 (1) (1976) 17–40.
[42] C. Wu, X. Chen, Q. Jin, J.-S. Chen, Applying smoothing technique and semi-proximal ADMM for image deblurring, CALCOLO 59 (4) (2022) 40.
[43] S.B. Shah, P. Pradhan, W. Pu, R. Randhi, M.R.D. Rodrigues, Y.C. Eldar, Optimization guarantees of unfolded ISTA and ADMM networks with smooth soft-thresholding, IEEE Trans. Signal Process. 72 (2024) 3272–3286.
[44] J. Bolte, S. Sabach, M. Teboulle, Proximal alternating linearized minimization for nonconvex and nonsmooth problems, Math. Program. 146 (1) (2014) 459–494.
[45] R. Bot, D. Nguyen, The proximal alternating direction method of multipliers in the nonconvex setting: Convergence analysis and rates, Math. Oper. Res. 45 (2) (2020) 682–712.
[46] M. Yashtini, Convergence and rate analysis of a proximal linearized admm for nonconvex nonsmooth optimization, J. Global Optim. 84 (4) (2022) 913–939.
[47] J. Li, W. Zhou, X. Li, Proximal alternating partially linearized minimization for perturbed compressive sensing, IEEE Trans. Signal Process. 71 (2023) 3373–3384.
[48] X. Jia, C. Kanzow, P. Mehlitz, G. Wachsmuth, An augmented lagrangian method for optimization problems with structured geometric constraints, Math. Program. 199 (1) (2023) 1365–1415.
[49] Y. Teng, L. Yang, X.L. Song, An augmented Lagrangian proximal alternating method for sparse discrete optimization problems, Numer. Algorithms 83 (3) (2020) 833–866.
[50] Y. Yang, J. Sun, H.B. Li, Z.B. Xu, ADMM-CSNet: A deep learning approach for image compressive sensing, IEEE Trans. Pattern Anal. Mach. Intell. 42 (3) (2020) 521–538.
[51] X.J. Chen, F.M. Xu, Y.Y. Ye, Lower bound theory of nonzero entries in solutions of $\ell_2 - \ell_p$ minimization, SIAM J. Sci. Comput. 32 (5) (2010) 2832–2852.
[52] R.T. Rockafellar, R. Wets, Variational Analysis, Springer, Berlin, Heidelberg, 1998.
[53] T. Blumensath, M.E. Davies, Iterative hard thresholding for compressed sensing, Appl. Comput. Harmon. Anal. 27 (3) (2009) 265–274.
[54] L. Pan, N.H. Xiu, J. Fan, Optimality conditions for sparse nonlinear programming, Sci. China Math. 60 (5) (2017) 759–776.
[55] L.T.K. Hien, D.N. Phan, N. Gillis, Inertial alternating direction method of multipliers for non-convex non-smooth optimization, Ann. Statist. 83 (1) (2022) 247–285.