Projected shrinkage algorithm for box-constrained $\ell_1$-minimization

Hui Zhang $^1$ · LiZhi Cheng $^1$

Abstract  Box-constrained $\ell_1$-minimization in some cases performs remarkably better than the classical $\ell_1$-minimization when appropriate box constraints are available. And also many practical $\ell_1$-minimization models indeed involve box constraints. In this paper, we propose an efficient iteration scheme, dubbed the projected shrinkage (ProShrink) algorithm, to solve a class of box-constrained $\ell_1$-minimization problems. A key component in our technique is that the proximal point operator of $\ell_1$-norm with box constraints can be equivalently simplified into a projected shrinkage operator which can be calculated directly. Theoretically, we prove that ProShrink enjoys convergence of both the primal and dual point sequences. On the numerical level, we demonstrate the benefit of adding box constraints via sparse recovery experiments.

Keywords  Proximal point operator · Projected shrinkage · Box constraints · $\ell_1$-minimization · Sparse recovery

1 Introduction

The past two decades have witnessed the wide application of $\ell_1$-minimization models in signal and image processing, compressive sensing, machine learning, statistic, and more. The success of $\ell_1$ minimization is mainly due to that the $\ell_1$-norm can be used to well reflect sparsity. Recently, it was observed that other auxiliary information of sparse solutions, such as partial support set [17] and nonnegative sparsity [3, 7], could
help to fit practical models. In this paper, instead of studying the theoretical benefit of adding auxiliary information, we are interested in designing efficient algorithms to solve the $\ell_1$-minimization problems with auxiliary box constraints:

$$\min_{x} \|x\|_1, \quad \text{s.t.} \quad Ax = b, \; x \in \mathcal{X}$$  \hspace{1cm} (1)

and

$$\min_{x} \|x\|_1 + \frac{1}{2\tau} \|x\|_2^2, \quad \text{s.t.} \quad Ax = b, \; x \in \mathcal{X}$$  \hspace{1cm} (2)

where $A \in \mathbb{R}^{n \times m}$, $b \in \mathbb{R}^m$ are given, $\tau$ is an augmented parameter, and $\mathcal{X}$ is some box-constrained set. The problems above are obtained respectively by imposing box constraints to the basis pursuit model [5]:

$$\min_{x} \|x\|_1, \quad \text{s.t.} \quad Ax = b$$  \hspace{1cm} (3)

and the augmented $\ell_1$ norm model [9]:

$$\min_{x} \|x\|_1 + \frac{1}{2\tau} \|x\|_2^2, \quad \text{s.t.} \quad Ax = b,$$  \hspace{1cm} (4)

both of which have been proved powerful for sparse recovery assignments. Adding box constraints to classical $\ell_1$-minimization problems on one hand extends the range of models (3) and (4) to include more practical models in application, and on the other hand can help to improve the ability of sparse recovery when correct box constraints are available. The similar benefit of adding box constraints to classical matrix completion has been observed in a recent paper [16] which was posted on arXiv at the time of writing this paper.

Due to the introduction of the strongly convex term $\frac{1}{\tau} \|x\|_2^2$, it has been explained in several papers [9,19,20] that models (2) and (4) have computational advantages over their correspondences (1) and (3). Besides, applying the proximal point algorithm [13] to models (1) or (3) generates a series of subproblems similar to (2) or (4). Therefore, the center assignment of solving problems (1)–(4) reduces to studying the following generalized problem

$$\min_{x} \|x\|_1 + \frac{1}{2\tau} \|x - u\|_2^2, \quad \text{s.t.} \quad Ax = b, \; x \in \mathcal{X}$$  \hspace{1cm} (5)

where $u$ is a given vector. With the help of the Lagrange dual analysis and by noticing the strong convexity of the objective function, we derive a projected shrinkage (ProShrink) algorithm for solving (5). By the Nesterov techniques [11], the proposed algorithm can be speeded up; we present an accelerated scheme as well. Theoretically, we prove the convergence of both the primal and dual point sequences of ProShink. A key component in our technique is that the proximal point operator [10] of $\ell_1$-norm

---

1 Let $\hat{x}$ be a sparse vector to be recovered and assume that its $i$-th entry $\hat{x}_i$ is nonzero. Then, we say that $[-|\hat{x}_i|, |\hat{x}_i|]$ is a correct box constraint, also called right size box.
Projected shrinkage algorithm for box-constrained...

with box constraints can be equivalently simplified into a projected shrinkage operator which can be calculated directly. Our technique can also be applied to simplifying the standard forward-backward splitting algorithm [1] for the boxed-constrained basis pursuit denoising problem:

$$\min_{x \in \mathcal{X}} \|x\|_1 + \frac{1}{2\lambda} \|Ax - b\|_2^2,$$

(6)

where $\lambda$ is a positive parameter.

The rest of the paper is organized as follows. In Sect. 2, we introduce some basic concepts of constrained convex optimization and derive important properties about the shrinkage operator. In Sect. 3, we propose the ProShrink algorithm, prove its convergence, and present detailed iteration schemes for solving models (1), (2), and (6). In Sect. 4, we do sparse recovery experiments to demonstrate the benefit of adding box constraints, and find out that box-constrained $\ell_1$-minimization models perform better but are usually harder to solve than the corresponding $\ell_1$-minimization models.

2 Notations and important properties

In this paper, we restrict our attention onto two classes of intervals. The first class is:

$$T_1 = \{I : I = [c, \infty), c > 0, \text{ or } I = (-\infty, c], c < 0, \text{ or } I = [c, d], 0 < c < d \text{ or } I = [d, c], d < c < 0\};$$

The second class is

$$T_2 = \{I : I = [c, d], c < 0 < d\}.$$

The box constraint $\mathcal{X}$ appeared in all models mentioned before is defined as $\mathcal{X} = I_1 \times I_2 \times \cdots \times I_n$, where $I_i \in T_1 \cup T_2$. Throughout this paper, we assume that $\mathcal{X} \cap \{x : Ax = b\} \neq \emptyset$.

2.1 Basic concepts and properties

First, we introduce the proximal point operator and its important properties.

**Definition 1** Let $f(x) : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a closed proper convex function. The proximal point operator [10] $\text{prox}_{f(\cdot)} : \mathbb{R}^n \to \mathbb{R}^n$ is defined by

$$\text{prox}_{f(\cdot)}(v) = \arg \min_x \left( f(x) + \frac{1}{2} \|x - v\|_2^2 \right).$$

(7)

Since the objective function is strongly convex and proper, $\text{prox}_{f(\cdot)}(v)$ is properly defined for every $v \in \mathbb{R}^n$. The following properties [10, 13] will be used in our analysis.
Lemma 1 Let $f(x) : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a closed proper convex function. Then, for all $x, y \in \mathbb{R}^n$ the proximal operator $\text{prox}_f$ satisfies the followings:

1. Firmly nonexpansive:

$$\|\text{prox}_f(x) - \text{prox}_f(y)\|_2^2 \leq \langle x - y, \text{prox}_f(x) - \text{prox}_f(y) \rangle$$

2. Lipschitz continuous:

$$\|\text{prox}_f(x) - \text{prox}_f(y)\|_2 \leq \|x - y\|_2.$$ 

If $f(x)$ is fully separable, meaning that $f(x) = \sum_{i=1}^n f_i(x_i)$, where $x = (x_1, x_2, \cdots, x_n)$ and $f_i(x_i) : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ are closed proper convex functions, then

$$\left( \text{prox}_f(x) \right)_i = \text{prox}_{f_i}(x_i). \quad (8)$$

Second, we need to introduce convex projected operator and projected subgradient to deal with box constraints.

Definition 2 (convex projected operator). $[x]^+_X := \arg \min_{y \in X} \|x - y\|$. 

The following property of projected operator shall be often encountered in our deduction.

Lemma 2 For any interval $I \in T_1 \cup T_2$, we have that $[\tau \cdot w]^+_I = \tau \cdot [w]^+_I/\tau$ holds for arbitrary $w \in \mathbb{R}$ and positive parameter $\tau$.

Proof We begin with the definition of projected operator and derive that

$$[\tau \cdot w]^+_I = \arg \min_{y \in I} \|y - \tau w\| = \arg \min_{y \in I} \|\tau^{-1} y - w\|$$

$$= \tau \cdot \arg \min_{z \in I/\tau} \|z - w\| = \tau \cdot [w]^+_I/\tau.$$ 

This completes the proof. 

Definition 3 (projected subgradient). Define $\partial^+_X f(x) := \{g : g = x - [x - h]^+_X, h \in \partial f(x)\}$. Without confusion, we also denote $x - [x - \partial f(x)]^+_X$ by $\partial^+_X f(x)$.

With projected subgradient, we can state a necessary and sufficient condition which guarantees a vector to be a minimizer to a class of constrained convex optimization problems.

Lemma 3 Let $f(x)$ be proper convex and $\mathcal{X}$ nonempty, closed, and convex. Then, we have

$$x^* \in \arg \min_{x \in \mathcal{X}} f(x) \iff 0 \in \partial^+_X f(x^*)$$

Proof The following two facts will be used in our deduction:

Fact 1 $z \in [x]^+_\mathcal{X} \iff \langle z - x, z - y \rangle \leq 0, \forall y \in \mathcal{X}$;

Fact 2 $\forall y \in \mathcal{X}, \exists h \in \partial f(x^*), \langle h, y - x^* \rangle \geq 0 \iff x^* \in \arg \min_{x \in \mathcal{X}} f(x)$. 

Springer
With these two facts, we derive that

\[
0 \in \partial_X^+ f(x^*) \iff x^* \in [x^* - \partial f(x^*)]_X^+ \quad (9a)
\]

\[
\iff \exists h \in \partial f(x^*), \text{ such that } x^* = [x^* - h]_X^+ \quad (9b)
\]

\[
\iff (x^* - (x^* - h), x^* - y) \geq 0, \forall y \in X \quad (9c)
\]

\[
\iff (h, y - x^*) \geq 0, \exists h \in \partial f(x^*), \forall y \in X \quad (9d)
\]

\[
\iff x^* \in \arg \min_{x \in X} f(x). \quad (9e)
\]

This completes the proof.

\[\Box\]

2.2 Projected shrinkage operator

In this part, we build an important formulation that links the proximal point operator of \(\ell_1\)-norm with box constraints and the projected shrinkage operation. In order to establish the formulation, we need two lemmas.

Lemma 4 Let \textit{shrink} be the shrinkage operator defined by \(\text{shrink}(s) = \text{sign}(s) \max(|s| - 1, 0)\) and let \(I \in T_1\). Then,

\[
[\text{shrink}(q)]_I^+ = [q - \text{sign}(c)]_I^+ \quad (10)
\]

holds for arbitrary \(q \in \mathbb{R}\), where \(c\) appears in the definition of \(I\).

Proof Recall that \(I = [c, \infty)\) with \(c > 0\), or \(I = (-\infty, c]\) with \(c < 0\), or \(I = [c, d]\) with \(0 < c < d\), or \(I = [d, c]\), with \(d < c < 0\). So it is easy to observe the following results:

(i) If \(c > 0\), then \([q + 1]_I^+ \equiv c\) for each \(q < -1\) and \([0]_I^+ \equiv c\);

(ii) If \(c < 0\), then \([q - 1]_I^+ \equiv c\) for each \(q > 1\) and \([0]_I^+ \equiv c\).

Thus, together with the definition of the shrinkage operator, for \(c > 0\) we derive that

\[
[\text{shrink}(q)]_I^+ = \begin{cases} 
[q - 1]_I^+, & q > 1 \\
[0]_I^+, & -1 \leq q \leq 1 \\
[q + 1]_I^+, & q < -1 
\end{cases} \quad (11)
\]

and for \(c < 0\) we have

\[
[\text{shrink}(q)]_I^+ = \begin{cases} 
[q - 1]_I^+, & q > 1 \\
c, & q \leq 1 = [q - 1]_I^+ 
\end{cases} \quad (12)
\]

On the other hand, \([q - \text{sign}(c)]_I^+ = [q - 1]_I^+\) when \(c > 0\) and \([q - \text{sign}(c)]_I^+ = [q + 1]_I^+\) when \(c < 0\). So the equality (10) holds. \[\Box\]
Lemma 5 Let $I \in T_1 \cup T_2$ and $I_\tau(t) = \tau \cdot |t| + \delta_I(t)$, where $\delta_I(\cdot)$ is the indicator function defined as $\delta(x) = 0$ if $x \in I$; otherwise $\delta(x) = \infty$. Then,

$$[\tau \cdot \text{shrink}(\tau^{-1}q)]_I^+ = \text{prox}_{I_\tau(\cdot)}(q)$$

(14)

holds for arbitrary $q \in \mathbb{R}$.

Proof By the definition of proximal point operator, we derive that

$$t^* := \text{prox}_{I_\tau(\cdot)}(q) = \arg \min_{t \in \mathbb{R}} \left[ I_\tau(t) + \frac{1}{2}(t - q)^2 \right] = \arg \min_{t \in I} \left[ \tau \cdot |t| + \frac{1}{2}(t - q)^2 \right].$$

(15a)

If $I \in T_1$, then $|t| = \text{sign}(c) \cdot t$ and hence

$$t^* = \arg \min_{t \in I} \left[ \tau \cdot |t| + \frac{1}{2}(t - q)^2 \right] = \arg \min_{t \in I} \left[ \tau \cdot \text{sign}(c) \cdot t + \frac{1}{2}(t - q)^2 \right].$$

Applying Lemma 3 yields to $t^* = [q - \tau \cdot \text{sign}(c)]_I^+$. Together with Lemmas 2 and 4, we derive that

$$t^* = \left[ \tau \cdot \text{shrink(} \tau^{-1}q \text{)} \right]_I^+ = \left[ \tau \cdot \text{shrink(} \tau^{-1}q \text{)} \right]_{I/\tau}.$$

So relationship (14) holds when $I \in T_1$.

If $I \in T_2$, then again invoking Lemma 3 yields to $t^* \in [q - \tau \cdot \partial |t|_{|t| = t^*}]_I^+$. Such $t^*$ must be the unique solution to problem (15a) because its objective functions is strongly convex. Thus, it suffices to show that $p(q) := [\tau \cdot \text{shrink}(\tau^{-1}q)]_I^+$ satisfies the following inclusion:

$$t \in [q - \tau \cdot \partial |t|]_I^+.$$
Case 2 \( p(q) < 0 \). Condition \( p(q) < 0 \) implies \( \tau^{-1} q < -1 \). Then, similarly to the argument in Case 1, we have that

\[
[q - \tau \cdot \|p(q)\|_1]_I^+ = [q - \tau]_I^+ = \tau \cdot \left( \tau^{-1} q + 1 \right)_{I/\tau}^+ = \tau \cdot \left[ \text{shrink} \left( \tau^{-1} q \right) \right]_{I/\tau}^+ = \tau \cdot p(q).
\]

Case 3 \( p(q) = 0 \). Condition \( p(q) = 0 \) implies \(-1 \leq \tau^{-1} \cdot q \leq 1\). Then, noting the fact that \([-1, 1] = \partial \|0\|_1\), we have that

\[
[q - \tau \cdot \|p(q)\|_1]_I^+ = [q - \tau \cdot \|0\|_1]_I^+ = \tau \cdot \left[ \tau^{-1} q - \partial \|0\|_1 \right]_{I/\tau}^+ \ni 0.
\]

This completes the proof.

Now, we are ready to build the most important formulation in this study.

**Corollary 1** Define the projected shrinkage operator \([\text{shrink}(v)]^+_X\) for a vector \( v \in \mathbb{R}^n \) via

\[
([\text{shrink}(v)]^+_X)_i = [\text{shrink}(v_i)]^+_I, \quad i = 1, 2, \ldots, n.
\]

And let \( \mathcal{X}_\tau(x) = \tau \cdot \|x\|_1 + \delta_{\mathcal{X}}(x) \). Then, it holds

\[
\left[ \tau \cdot \text{shrink} \left( \tau^{-1} v \right) \right]_{\mathcal{X}}^+ = \text{prox}_{\mathcal{X}_\tau(\cdot)}(v). \tag{16}
\]

**Proof** Noting that \( \mathcal{X}_\tau(x) = \sum_{i=1}^n \tau |x_i| + \delta_{I_i}(x_i) \) and the property (8), together with Lemma 5, the conclusion follows.

The significance of formulation (16) is twofold: the expression based on the proximal point operator of \( \ell_1 \)-norm will be used for convergence analysis; whilst the expression via the projected shrinkage operator is for computational consideration due to its simplicity. The main advantages of the projected shrinkage operator over the proximal point operator of \( \ell_1 \)-norm with box constraints lies in that the former can be calculated directly.

### 3 Projected shrinkage algorithm

In this section, we derive the Lagrange dual problem of (5) and then solve it with the help of formulation (16). Following the proof idea in paper [22], we prove the convergence of both the primal and dual point sequences of the proposed algorithm.
3.1 Lagrange dual analysis

The Lagrangian of the augmented convex model (5) is

\[ L(x, y) = \|x\|_1 + \frac{1}{2\tau} \|x - u\|_2^2 + \langle y, b - Ax \rangle. \] (17)

The Lagrange dual function is

\[ D(y) = \min_{x \in X} L(x, y). \] (18)

For any vector \( y \in \mathbb{R}^m \), the \( x \)-minimization problem above is a strongly convex program and hence has a unique solution \( x^*(y) \) that satisfies

\[ x^*(y) = \arg \min_{x \in X} \left[ \|x\|_1 + \frac{1}{2\tau} \|x - u\|_2^2 \right] \] (19a)

\[ = \arg \min_{x \in X} \left[ \|x\|_1 + \frac{1}{2\tau} \|x - u - \tau \cdot A^T y\|_2^2 \right] \] (19b)

\[ = \arg \min_{x \in X} \left[ \tau \cdot \|x\|_1 + \frac{1}{2} \|x - u - \tau \cdot A^T y\|_2^2 \right] = \text{prox}_{\chi\tau(\cdot)} \left( u + \tau \cdot A^T y \right). \] (19c)

where \( \chi\tau(x) = \tau \cdot \|x\|_1 + \delta_{\chi}(x) \) and

\[ L(x, y) = \|x\|_1 + \frac{1}{2\tau} \|x - u\|_2^2 + \langle y, b - Ax \rangle. \]

By formulation (16) in Corollary 1, we obtain

\[ x^*(y) = \text{prox}_{\chi\tau(\cdot)} \left( u + \tau \cdot A^T y \right) = \left[ \tau \cdot \text{shrink} \left( \tau^{-1}u + A^T y \right) \right]^+_{\chi}. \] (20)

Now, \( D(y) = L(x^*(y), y) = L([\tau \cdot \text{shrink}(\tau^{-1}u + A^T y)]^+_{\chi}, y) \). Thus, we can write down the Lagrange dual problem of (5) as follows:

\[ \max_y L \left( \left[ \tau \cdot \text{shrink} \left( \tau^{-1}u + A^T y \right) \right]^+_{\chi}, y \right). \] (21)

It is well known in convex analysis [14] that the dual objective function \( L([\tau \cdot \text{shrink}(\tau^{-1}u + A^T y)]^+_{\chi}, y) \) is gradient-Lipschitz-continuous due to the strong convexity of the primal objective function \( \|x\|_1 + \frac{1}{2\tau} \|x - u\|_2^2 \). And moreover, the gradient of dual objective function is given by

\[ \nabla D(y) = b - Ax^*(y) = b - A \left[ \tau \cdot \text{shrink} \left( \tau^{-1}u + A^T y \right) \right]^+_{\chi}. \]
Each solution to the dual problem (21) can generate the unique solution to the primal problem (5) via formulation (20). This fact is stated in the following lemma.

**Lemma 6** Let $x^*$ be the unique solution to problem (5) and $\mathcal{X} \cap \{x : Ax = b\} \neq \emptyset$. Then the dual solution set to problem (21) is

$$\mathcal{Y} = \left\{ y : x^* = \left[ \tau \cdot \text{shrink} \left( \tau^{-1}u + A^T y \right) \right]_\mathcal{X}^+ \right\},$$

which is nonempty and convex.

**Proof** By Lemma 1 and Corollary 1, for $\forall y, \tilde{y} \in \mathbb{R}^m$ we have that

$$\langle \nabla D(y) - \nabla D(\tilde{y}), y - \tilde{y} \rangle$$

$$= \left( \left[ \tau \cdot \text{shrink} \left( \tau^{-1}u + A^T y \right) \right]_\mathcal{X}^+ - \left[ \tau \cdot \text{shrink} \left( \tau^{-1}u + A^T \tilde{y} \right) \right]_\mathcal{X}^+ \right),$$

$$A^T y - A^T \tilde{y} \rangle$$

$$\geq \tau^{-1} \left\| \text{prox}_{\mathcal{X}(\cdot)}^{\tau} \left( u + \tau \cdot A^T y \right) - \text{prox}_{\mathcal{X}(\cdot)}^{\tau} \left( u + \tau \cdot A^T \tilde{y} \right) \right\|^2 \geq 0,$$

which implies that the dual objective function $D(y)$ is convex by Theorem 2.1.3 in [11]. Thus, the dual solution set is

$$\mathcal{Y}' = \{ y : \nabla D(y) = 0 \}$$

$$= \left\{ y : A \left[ \tau \cdot \text{shrink} \left( \tau^{-1}u + A^T y \right) \right]_\mathcal{X}^+ = b \right\},$$

which must be nonempty and convex by assumption $\mathcal{X} \cap \{x : Ax = b\} \neq \emptyset$ and the convexity of $D(y)$. Now, it suffices to show $\mathcal{Y} = \mathcal{Y}'$. On one hand, we have $\mathcal{Y} \subseteq \mathcal{Y}'$, since $Ax^* = b$. On the other hand, let $y' \in \mathcal{Y}'$, i.e., $y'$ is some dual solution. Then, $x' = [\tau \cdot \text{shrink}(\tau^{-1}u + A^T y')]_\mathcal{X}^+$ is a primal solution and it must equal $x^*$ by uniqueness. So $y' \in \mathcal{Y}$ and hence $\mathcal{Y}' \subseteq \mathcal{Y}$, which completes the proof. □

### 3.2 Algorithm schemes

Applying the gradient iteration to the dual objective $D(y)$ gives:

$$y^{k+1} = y^k + h \left( b - A \left[ \tau \cdot \text{shrink} \left( \tau^{-1}u + A^T y^k \right) \right]_\mathcal{X}^+ \right),$$

where $h > 0$ is the step size whose range shall be established later on. By setting $x^{k+1} = [\tau \cdot \text{shrink}(\tau^{-1}u + A^T y^k)]_\mathcal{X}^+$, we obtain the equivalent iteration in primal-dual form:

$$\begin{cases}
    x^{k+1} = [\tau \cdot \text{shrink}(\tau^{-1}u + A^T y^k)]_\mathcal{X}^+ \\
    y^{k+1} = y^k + h(b - Ax^{k+1}).
\end{cases}$$
Because the projected shrinkage operator is involved, we call (25) projected shrinkage algorithm. Recall that the linearized Bregman (LBreg) algorithm [4,18] has the following form:

\[
\begin{align*}
    x^{k+1} &= \tau \cdot \text{shrink}(A^Ty^k) \\
y^{k+1} &= y^k + h(b - Ax^{k+1})
\end{align*}
\] (26)

Therefore, the ProShrink algorithm can be viewed as a generalization of the LBreg algorithm.

Applying the Nesterov accelerated scheme [11], we obtain an accelerated ProShrink algorithm with the following form:

\[
\begin{align*}
x^{k+1} &= \left[\tau \cdot \text{shrink}(\tau^{-1}u + A^Ty^k)\right]^+ \\
y^{k+1} &= y^k + h(b - Ax^{k+1}) \\
\gamma_k &= (\sqrt{\theta_k} + 4 - \theta_k)/2 \\
\beta_{k+1} &= (1 - \theta_k)\gamma_k, \theta_{k+1} = \theta_k\gamma_k \\
y^{k+1} &= z^{k+1} + \beta_{k+1}(z^{k+1} - z^k)
\end{align*}
\] (27)

where \(\theta_0 = 1\). In addition, the adaptive restart technique developed in [12] may be applied to further accelerate the scheme (27); the acceleration for the LBreg algorithm was observed in paper [21].

Now, let us return to models (1) and (2). Model (2) can be solved by ProShrink (25) or its acceleration (27) with \(u = 0\). To solve model (1), we apply the proximal point algorithm [15] and obtain a series of subproblems as follows:

\[
z^{k+1} = \arg \min \left\{ \|x\|_1 + \frac{1}{2\lambda_k}\|x - z^k\|_2^2, Ax = b, x \in \mathcal{X} \right\}
\]

where \(\lambda_k\) are positive parameters. Each subproblem above can be well solved by ProShrink (25) as well. We write down the iteration scheme without deduction details:

\[
\begin{align*}
x^{k+1} &= \left[\lambda_k \cdot \text{shrink}\left(\frac{1}{\lambda_k}z^k + A^Ty^{k,i}\right)\right]^+ \\
y^{k+1} &= y^{k,i} + h(b - Ax^{k,i+1})
\end{align*}
\] (28)

The subproblem can also be solved by the accelerated ProShrink scheme (27).

**Proposition 1 ([8]).** Let \(f^* = \min\{\|x\|_1, Ax = b, x \in \mathcal{X}\}\). Then the following convergent result in terms of the objective residual holds:

\[
\|z^k\|_1 - f^* = O\left(\frac{1}{\sum_{i=0}^{k-1} \lambda_i}\right).
\]

Provided that \(\lambda_k \geq \lambda_0 > 0\), the iteration-complexity \(O(1/k)\) for the sequence \(\{\|z^k\|_1 - f^*\}_{k=0}^{\infty}\) is instantly implied.
At last, the standard forward–backward splitting algorithm \cite{2,6} for model (6) is

\[ x^{k+1} = \text{prox}_{\lambda \gamma_k} \left( x^k - \frac{\gamma_k}{\lambda} A^T (Ax^k - b) \right), \quad (29) \]

where \( \gamma_k \) are step sizes. The main difficulty of (29) is to compute the proximal point operator of \( \lambda \gamma_k \). Utilizing formulation (16), the iteration (29) can be simplified into

\[ x^{k+1} = \left[ \gamma_k \cdot \text{shrink} \left( \gamma_k^{-1} x^k - \frac{1}{\lambda} A^T (Ax^k - b) \right) \right]^+ \quad (30) \]

which now can be calculated directly.

### 3.3 Convergence analysis

In this part, we prove the convergence of primal sequence \( \{x^k\} \) and dual sequence \( \{y^k\} \) in the iteration (25).

**Theorem 1** Set step size \( h \in (0, \frac{2}{\tau \|A\|^2}) \) and \( y^0 = 0 \) in iteration (25). Let \( x^* \) be the unique minimizer to problem (5) and \( \mathcal{Y} \) be the solution set to problem (21). Then, \( \lim_{k \to +\infty} x^k = x^* \), and there exists a point \( \tilde{y} \in \mathcal{Y} \) such that \( \lim_{k \to +\infty} y^k = \tilde{y} \).

This theorem can be proved in the same manner as that in paper \cite{22}. For completeness, we provide a proof below.

**Proof** Let \( \hat{y} \in \mathcal{Y} \). By Lemma 6, we have \( x^* = [\tau \cdot \text{shrink}(\tau^{-1} u + A^T \hat{y})]^+ \). Together with \( x^{k+1} = [\tau \cdot \text{shrink}(\tau^{-1} u + A^T y^k)]^+ \), and Lemma 1 and Corollary 1, we derive

\[
\begin{align*}
\left\langle A^T y^k - A^T \hat{y}, x^{k+1} - x^* \right\rangle \\
= \left\langle A^T y^k - A^T \hat{y}, [\tau \cdot \text{shrink}(\tau^{-1} u + A^T y^k)]^+ \right\rangle \\
- \left\langle [\tau \cdot \text{shrink}(\tau^{-1} u + A^T \hat{y})]^+, x^{k+1} - x^* \right\rangle \\
= \left\langle A^T y^k - A^T \hat{y}, \text{prox}_{\lambda \mathcal{X}} \left( u + \tau \cdot A^T y^k \right) - \text{prox}_{\lambda \mathcal{X}} \left( u + \tau \cdot A^T \hat{y} \right) \right\rangle \\
\geq \tau^{-1} \cdot \left\| \text{prox}_{\lambda \mathcal{X}} \left( u + \tau \cdot A^T y^k \right) - \text{prox}_{\lambda \mathcal{X}} \left( u + \tau \cdot A^T \hat{y} \right) \right\|^2 \\
= \tau^{-1} \cdot \|x^{k+1} - x^*\|^2
\end{align*}
\]
Using this inequality, we have
\[
\|y^{k+1} - \hat{y}\|^2_2 = \|y^k - \hat{y} + h \left(b - Ax^{k+1}\right)\|^2_2 \\
= \|y^k - \hat{y} + h \left(Ax^* - Ax^{k+1}\right)\|^2_2 \\
= \|y^k - \hat{y}\|^2_2 - 2h \left(A^T y^k - A^T \hat{y}, x^{k+1} - x^* \right) + h^2 \left\|Ax^* - Ax^{k+1}\right\|^2_2 \\
\leq \|y^k - \hat{y}\|^2_2 - 2h \tau^{-1} \left\|x^{k+1} - x^*\right\|^2_2 + h^2 \left\|A\right\|^2 \left\|x^{k+1} - x^*\right\|^2_2 \\
= \|y^k - \hat{y}\|^2_2 - h \left(2\tau^{-1} - h \left\|A\right\|^2 \right) \left\|x^{k+1} - x^*\right\|^2_2.
\]

Therefore, under the assumption \(0 < h < \frac{2}{\tau \|A\|^2}\), we can make the following claims:

**Claim 1** \(\|y^{k+1} - \hat{y}\|_2\) is monotonically nonincreasing in \(k\) and thus converges to a limit.

**Claim 2** \(\|x^{k+1} - x^*\|_2\) converges to 0 as \(k\) tends to \(+\infty\), i.e., \(\lim_{k \to +\infty} x^{k+1} = x^*\).

From claim 1, it follows that \(\{y^k\}\) is bounded and thus has a converging subsequence \(y^{k_i}\). Let \(\bar{y} = \lim_{i \to \infty} y^{k_i}\). By the Lipschitz continuity of proximal point operator in Lemma 1 and Corollary 1, we have

\[
x^* = \lim_{i \to \infty} x^{k_i+1} = \lim_{i \to \infty} \left[\tau \cdot \text{shrink} \left(\tau^{-1} u + A^T y^{k_i+1}\right)\right]^+ \chi' \\
= \lim_{i \to \infty} \text{prox}_{\chi_{\tau}(\cdot)} \left(u + \tau \cdot A^T y^{k_i+1}\right) = \text{prox}_{\chi_{\tau}(\cdot)} \left(u + \tau \cdot A^T \bar{y}\right) \\
= \left[\tau \cdot \text{shrink} \left(\tau^{-1} u + A^T \bar{y}\right)\right]^+ \chi',
\]

so \(\bar{y} \in \mathcal{Y}\) by Lemma 6. Recall \(\hat{y} \in \mathcal{Y}\) is arbitrary. Hence, claim 1 holds for \(\hat{y} = \bar{y}\). If \(\{y^k\}\) had another limit point, then \(\|y^{k+1} - \bar{y}\|_2\) would fail to be monotonic. So, \(y^k\) converges to \(\bar{y} \in \mathcal{Y}\) (in norm).

4 Numerical experiment

In this section, we do sparse recovery experiments to show potential advantages of adding box constraints. It was shown in [9] when the augmented parameter \(\tau \geq 10\|x\|_{\infty}\), the augmented \(\ell_1\)-norm model (4) is equivalent to classical basis pursuit (3) if the sensing matrix \(A\) satisfies certain properties such as null-space property, or restricted isometry property. So we only test models (4) and (2) to observe possible advantages of adding box constraints. Model (4) was solved by the LBreg algorithm and model (2) by the ProShrink algorithm.

All the experiments were carried out on a personal computer with Intel(R) Core(TM)i5-3320 CPU @ 2.60GHz and 8G RAM, using MATLAB (version R2012b).
Projected shrinkage algorithm for box-constrained...

Fig. 1  Comparison of augmented $\ell_1$ norm models with or without box constraints for sparse recovery (corresponding to the ProShrink and the LBreg algorithms separately). a Comparison of exact recovery rate. b Comparison of average run time (s)

4.1 The benefit of adding box constraints

We used 100 random pairs $(A, x)$ with matrices $A$ of size $200 \times 400$ and vectors $x$ with 400 entries, out of which $s$ were nonzero entries set to $\pm 1$ uniformly randomly for $s = 1, 2, 3, \ldots, 80$, where $s$ denotes the sparsity level. Each entry of the sensing matrix $A$ was sampled independently from the standard Gaussian distribution. Thus, $b = Ax$ are given vectors. A relative error of $10^{-12}$ was considered as an exact recovery; the relative error is defined as $\frac{x - x_o}{x}$ where $x_o$ is finally generated by the LBreg or the ProShrink algorithms. The box-constrained set $X$ for the ProShrink algorithm was set as $X = [-1, 1]^{400}$.

We plot the exact recovery rate via sparsity levels in Fig. 1a from which we see that ProShrink performs remarkably better than LBreg as the sparse level increases. More precisely, when the sparse level is low, both LBreg and ProShrink can well recover sparse signals; but when the sparse level becomes higher, the exact recovery rate by LBreg is worse than that by ProShrink that indicates adding box constraints to the augmented $\ell_1$-norm model (4) indeed improves the recovery rate. The average CPU time via sparsity levels of the LBreg and the ProShrink algorithms was shown in Fig. 1b. It is fair to compare their CPU time when the exact sparse recovery occurs, so we only consider the interval of sparsity level less than 40. Figure 1b tells us that the ProShrink does not completely beat the LBreg in this particular case. More precisely, the ProShrink has a little advantage over the LBreg when the sparsity level is low, but when the sparsity level becomes relatively large the LBreg performs better.

4.2 The impact of box sizes

To see the impact of choosing larger size boxes, we repeated the experiments in Sect. 4.1. In the comparison, we added a new result which was obtained by running the ProShrink algorithm with the box constrained set $[-1.1, 1.1]^{400}$, abbreviated as 1.1X. Figure 2 shows that the ProShrink algorithm with larger boxes costs much more
Fig. 2 Exact sparse recovery comparisons via LBreg, ProShrink with box constrained set X, and ProShrink with box constrained set 1.1X. a Comparison of exact recovery rate. b Comparison of average run time (s)

CPU time than LBreg but does not help to improve the exact recovery rate. From this test, we suspect that in order to beat LBreg in terms of exact recovery rate, ProShrink must have some right size boxes as constraints.

In what follows, by constructing an example, we illustrate an interesting phenomenon; that is if more right size boxes can be utilized, then better results of sparse recovery might be obtained. We consider a sparse signal recovery problem with a signal $x^* \in \mathbb{R}^{400}$ which consists of 80 spikes with amplitudes 1, 2, 3, and 4, shown at the top plot of Fig. 3. Each entry of the sensing matrix $A \in \mathbb{R}^{210 \times 400}$ was sampled independently from the standard Gaussian distribution. Data vector $b = A x^*$ is given. Two types of box-constrained sets are tested: The first type is X1, taken as $[-4, 4]^{400}$, and the second is X2, taken as
Projected shrinkage algorithm for box-constrained...

\[ [-1, 1]^{100} \times [-2, 2]^{100} \times [-3, 3]^{100} \times [-4, 4]^{100}. \]

Obviously, X2 utilizes more right size boxes of \( x_\ast \) than X1. Figure 3 shows that both LBreg and ProShrink with X1 fail to exactly recover \( x_\ast \), while ProShrink with X2 generates the perfectly reconstructed result at the expense of CPU time.

5 Conclusion

In this paper, we proposed the projected shrinkage algorithm for boxed-constrained \( \ell_1 \)-minimization. The most important result in our study should be formulation (16) that establishes the relationship between projected shrinkage operator and the proximal point operator of \( \ell_1 \)-norm with box constraints. Numerically, we demonstrated that adding box constraints to classical \( \ell_1 \)-minimization may obtain better performance. In addition, by choosing different types of box-constrained sets, we might conclude that only right size boxes can help to improve the ability of sparse recovery, and that if more right size boxes are utilized, then better recovery results might be obtained. However, how to find the appropriate box size and giving theoretical explanation for these phenomena are open. We leave them for future work.

Acknowledgements We would like to thank Professor Wotao Yin (UCLA) for his comments and suggestion on numerical verification and Professor Jian-Feng Cai (Iowa U) for his insight of the projected shrinkage operator. We appreciate the constructive comments by anonymous reviewers, with which great improvements have been made in this manuscript. The work is supported by the National Science Foundation of China (Nos.11501569 and 61072118).

References

1. Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator theory in Hilbert Spaces. Springer, Berlin (2011)
2. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imaging Sci. 2(1), 183–202 (2009)
3. Bruckstein, A.M., Elad, M., Zibulevsky, M.: On the uniqueness of nonnegative sparse solutions to underdetermined systems of equations. IEEE Trans. Inf Theory 54(11), 4813–4820 (2008)
4. Cai, J.-F., Osher, S., Shen, Z.: Linearized Bregman iterations for compressed sensing. Math. Comput. 78(267), 1515–1536 (2009)
5. Chen, S., Donoho, D., Saunders, M.: Atomic decomposition by basis pursuit. SIAM Rev. 43(1), 129–159 (2001)
6. Combettes, P.L., Wajs, V.R.: Signal recovery by proximal forward–backward splitting. Multiscale Model. Simul. 4(4), 1168–1200 (2005)
7. Donoho, D.L., Tanner, J.: Sparse nonnegative solution of underdetermined linear equations by linear programming. Proc. Natl. Acad. Sci. 102(27), 9446–9451 (2005)
8. Güler, O.: On the convergence of the proximal point algorithm for convex minimization. SIAM J. Control Optim. 29(2), 403–419 (1991)
9. Lai, M.J., Yin, W.: Augmented \( \ell_1 \) and nuclear-norm models with a globally linearly convergent algorithm. SIAM. J. Imaging Sci. 6(2), 1059–1091 (2013)
10. Moreau, J.J.: Fonctions convexes duales et points proximaux dans un espace hilbertien. C. R. Acad. Sci. Paris Ser. A Math. 255, 2897–2899 (1962)
11. Nesterov, Y.: Introductory Lectures on Convex Optimization: A Basic Course. Kluwer Academic Publishers, Berlin (2004)
12. O’Donoghue, B., Candès, E.: Adaptive restart for accelerated gradient schemes. Found. Comput. Math. 15(3), 715–732 (2015)
13. Parikh, N., Boyd, S.: Proximal algorithm. Found. Trends Optim. 1(3), 123–231 (2014)
14. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton (1970)
15. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM J. Control Optim. 14(5), 877–898 (1976)
16. Takač, M., Mareček, J., Richtárik, P.: Inequality-Constrained Matrix Completion: Adding the Obvious Helps! arXiv:1408.2467v1 [math.OC] (2014)
17. Vaswani, N., Lu, W.: Modified-CS: modifying compressive sensing for problems with partially known support. IEEE Trans. Signal Process. 58(9), 4595–4607 (2010)
18. Yin, W., Osher, S., Goldfarb, D., Darbon, J.: Bregman iterative algorithms for $\ell_1$-minimization with applications to compressed sensing. SIAM J. Imaging Sci. 1(1), 143–168 (2008)
19. Zhang, H., Cheng, L., Zhu, W.: A lower bound guaranteeing exact matrix completion via singular value thresholding algorithm. Appl. Comput. Harmonic Anal. 31(3), 454–459 (2011)
20. Zhang, H., Cai, J.-F., Cheng, L., Zhu, J.: Strongly convex programming for exact matrix completion and robust principal component analysis. Inverse Probl. Imaging 6(2), 357–372 (2012)
21. Zhang, H., Yin, W.: Gradient methods for convex minimization: better rates under weaker conditions. In: UCLA CAM Report (13–17), 2013
22. Zhang, H., Cheng, L.Z., Yin, W.: Dual algorithm for a class of augmented convex signal recovery models. Commun. Math. Sci. 13(1), 103–112 (2015)