# Iterative positive thresholding algorithm for nonnegative sparse optimization

Lufang Zhang[a], Yaohua Hu[b], Carisa Kwok Wai Yu[c] and Jinhua Wang[d]

[a]School of Mathematical Sciences, Zhejiang University, Hangzhou 310027, P. R. China
[b]College of Mathematics and Statistics, Shenzhen University, Shenzhen 518060, P. R. China;
[c]Department of Mathematics and Statistics, Hang Seng Management College, Hong Kong;
[d]Department of Mathematics, Zhejiang University of Technology, Hangzhou 310032, P. R. China

**ARTICLE HISTORY**
Compiled April 24, 2018

**Abstract**
The nonnegative $\ell_1$ regularization problem has been widely studied for finding nonnegative sparse solutions of linear inverse problems and gained successful applications in various application areas. In the present paper, we propose an iterative positive thresholding algorithm (IPTA) to solve the nonnegative $\ell_1$ regularization problem and investigate its convergence properties in finite- or infinite-dimensional Hilbert spaces. The significant advantage of the IPTA is that it is very simple and of low computation cost, and thus, it is practically attractive, especially for large-scale problems. The global convergence of the IPTA is achieved under some mild assumptions on algorithmic parameters. Furthermore, we introduce a notion of positive orthogonal sparsity pattern, and use it to establish the linear convergence rate of the IPTA to a global minimum. Finally, the numerical study on compressive sensing shows that the proposed IPTA is efficient in approaching the nonnegative sparse solutions of linear inverse problems and outperforms several existing algorithms in sparse optimization.

**KEYWORDS**
Nonnegative sparse optimization; Nonnegative $\ell_1$ regularization problems; Iterative positive thresholding algorithm; Global convergence; Linear convergence rate

## 1. Introduction

Nowadays, sparse optimization has become a very popular research topic in many disciplines of applied science and gained successful applications in a wide range of fields, which aims to find a sparse approximate solution of an underdetermined linear system from the underlying data. The $\ell_1$ regularization problem, also called Lasso [39] or Basis Pursuit [8], has been accepted as one of the most useful methodologies for sparse optimization, that is,

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}\|Ax-b\|^2 + \lambda\|x\|_1, \tag{1}$$

CONTACT Yaohua Hu. Email: mayhhu@szu.edu.cn

where $\|x\|_1 := \sum_{i=1}^n |x_i|$ is a sparsity promoting norm, and $\lambda$ is a regularization parameter providing a tradeoff between accuracy and sparsity. The $\ell_1$ regularization problem has been successfully and widely applied in various application areas, such as machine learning [1], systems biology [33], image science [15], compressive sensing [7,13]. It has also been investigated in infinite-dimensional Hilbert spaces [4,11] and applied in Fourier analysis [6] and Harmonic analysis [12].

Motivated by its successful applications, a great amount of attention has been attracted to the development of optimization algorithms, and many efficient algorithms have been proposed to solve the $\ell_1$ regularization problem; see [3,10,18,21,24,26,27, 30,43] and references therein. In particular, the iterative soft thresholding algorithm (ISTA) independently proposed by [11,17] is one of the most widely studied first-order iterative algorithms for solving problem (1). Tremendous efforts have been devoted to investigating the convergence properties of the ISTA; see [3,4,20,31,38] and references therein. In [11], the global (strong) convergence result of the ISTA was established under some assumptions on the algorithmic parameters. Moreover, the linear convergence of the ISTA has been well investigated under some additional assumptions. For example, Hale et al. [20] proved a linear convergence of the ISTA to a solution of problem (1) in Euclidean spaces under the assumption that $A$ satisfies a basis injectivity (BI) property or that a strict complementary condition (SCC) is satisfied at the solution. Extending to the infinite-dimensional Hilbert spaces, Bredies and Lorenz in [4] showed the linear convergence of the ISTA under the assumption of a finite basis injectivity (FBI) property or a strict sparsity pattern (SSP). Improving these results, Zhang et al. [45] introduced a notion of orthogonal sparsity pattern (OSP) that is weaker than either FBI or SSP, and established the linear convergence of the ISTA under the assumption of OSP in either finite- or infinite-dimensional spaces.

In recent years, a great amount of attention has been attracted to the structured sparse optimization, that is to enhance the sparse recovery capability by employing the special structures of practical applications; see [1,16,23,37,41] and references therein. One important structure is the nonnegativity of sparse variables. That is, a nonnegative $\ell_1$ regularization problem (i.e., (1) with an additional nonnegative constraint) is solved to approach a nonnegative sparse solution of the linear inverse problem. The nonnegative $\ell_1$ regularization problem was originally introduced by Donoho and Tanner in [14] and widely investigated in [5,9,16,19] and references therein. Numerous applications have been discovered in many fields, such as face recognition [22], compress sensing [25], statistics [44] and spectrometry analysis [16,35]. Several optimization algorithms have been proposed to solve the nonnegative $\ell_1$ regularization problem, such as nonnegative OMP [42] and nonnegative ADMM [16].

The present paper aims to continue the development of optimization algorithms for nonnegative sparse optimization. In order to include the studies in finite- and infinite-dimensional spaces in a unified framework, we adopt the following notations. Let $\mathcal{H}$ be a Hilbert space, and let $\ell^2$ be the Hilbert space consisting of all square-summable sequences. Let $N \in \mathbb{N} \cup \{+\infty\}$ be fixed, and write

$$\ell_N^2 := \begin{cases} \mathbb{R}^N, & \text{if } N \in \mathbb{N}, \\ \ell^2, & \text{otherwise}, \end{cases} \quad \text{and} \quad \mathcal{I}_N := \begin{cases} \{1, \dots, N\}, & \text{if } N \in \mathbb{N}, \\ \mathbb{N}, & \text{otherwise}, \end{cases} \tag{2}$$

and $\ell_N^+ := \{u \in \ell_N^2 : u \geq 0\}$. In the present paper, we consider the following nonnegative $\ell_1$ regularization problem

$$\min_{u\in\ell_N^+} \frac{1}{2}\|Ku-h\|^2 + \sum_{k=1}^{N} \omega_k u_k, \tag{3}$$

where $K:\ell_N^2 \to \mathcal{H}$ is a bounded linear operator and $\omega := (\omega_k)$ is a sequence of weights satisfying

$$\omega_k \geq \underline{\omega} > 0 \quad \text{for each } k \in \mathcal{I}_N. \tag{4}$$

Let $S$ denote the solution set of problem (3). Inspired by the ideas of the ISTA [11], we propose an iterative positive thresholding algorithm (IPTA) to solve the nonnegative $\ell_1$ regularization problem (3). The main difference is that the IPTA employs a positive thresholding operator to replace the soft thresholding operator in the ISTA. Consequently, the IPTA inherits the remarkable advantages of the ISTA that it is very simple and of low computation cost, and hence, the IPTA is practically attractive, especially for large-scale problems. A clear convergence analysis of the IPTA is provided to advance our understanding of its strength for solving the nonnegative $\ell_1$ regularization problem (3). In particular, the global convergence result of the IPTA is achieved under some mild assumptions on algorithmic parameters, which are same as the ones in [4, Theorem 1] for the ISTA. Furthermore, we introduce a notion of positive orthogonal sparsity pattern (POSP), and use it to establish the linear convergence of the IPTA to a global minimum of problem (3). As a byproduct, some sufficient conditions for ensuring the POSP are provided in terms of the FBI/SCC/OSP.

Furthermore, we conduct some numerical experiments in the simulation of compressive sensing to demonstrate the numerical performance of the proposed IPTA. The numerical results validate the linear convergence rate of the IPTA, and show that the IPTA is efficient in approaching the nonnegative sparse solutions of linear inverse problems and outperforms several existing algorithms in sparse optimization, including ISTA [11], NADMM [16], ADMM [43], NOMP [42] and OMP [40], on both accuracy and robustness.

The paper is organized as follows. In Section 2, we present the notations and preliminary results to be used in the present paper. In Section 3, we propose the IPTA to solve the nonnegative $\ell_1$ regularization problem (3) and investigate its convergence properties, including the global convergence and the linear convergence rate to a global minimum of problem (3). Finally, numerical results of the IPTA on compressive sensing are demonstrated in Section 4.

## 2. Notation and preliminary results

In the present paper, we consider a Hilbert space $\mathcal{H}$ with inner product $\langle\cdot\,,\cdot\rangle$ and the associated norm $\|\cdot\|$. As usual, $\mathbb{R}_+$ denotes the set of all non-negative real numbers, and, for $u \in \ell_N^2$, $u_+$ denotes the vector with entries equal to $\max\{u_i, 0\}$ and $\mathrm{supp}(u)$ denotes the support set of $u$, i.e.,

$$\mathrm{supp}(u) := \{k \in \mathcal{I}_N : u_k \neq 0\}.$$

Recall that $\ell^2_N$ and $\mathcal{I}_N$ are defined by (2). For $C \subseteq \ell^2_N$ and $I \subseteq \mathcal{I}_N$, the orthogonal complement of $C$ and the complementary of $I$ are denoted by $C^\perp$ and $I^c$, respectively. We denote a subspace $E_I$ by

$$E_I := \{u \in \ell^2_N : u_k = 0 \text{ for each } k \in I^c\}.$$

For $x \in \ell^2_N$, the classical metric projection of $x$ onto $C$ and the distance of $x$ from $C$, denoted by $P_C(x)$ and $\mathrm{d}_C(x)$, are respectively defined by

$$P_C(x) := \mathrm{argmin}\{\|x - y\| : \ y \in C\} \quad \text{and} \quad \mathrm{d}_C(x) := \inf\{\|x - y\| : \ y \in C\}.$$

We adopt the convention that $P_\emptyset(x) = 0$. It is trivial to see that, for each $u \in \ell^2_N$,

$$(P_{E_I}(u))_i = u_i \text{ when } i \in I, \text{ and } (P_{E_I}(u))_i = 0 \text{ otherwise.}$$

The following lemma recalls some basic properties of the projection operator, in which (a) is taken from [2, Theorem 3.14]; (b) and (c) are from [2, Corollary 3.22(iii)(vi)]; while (d) is known in [2, Proposition 3.19].

**Lemma 2.1.** *Let $C$ be a closed linear subspace of $\ell^2_N$ and $x \in \ell^2_N$. Then the following assertions hold:*

(a) *$z = P_C(x)$ if and only if $z \in C$ and $x - z \in C^\perp$;*
(b) *$P_C$ is a linear and continuous operator with $\|P_C\| \le 1$;*
(c) *$P_C^* = P_C$;*
(d) *$P_C$ is idempotent, i.e., $P_C^2 = P_C$.*

Let $K : \ell^2_N \to \mathcal{H}$ be a bounded linear operator. The conjugate of $K$ is denoted by $K^*$, and the kernel and image of $K$ are respectively defined by

$$\mathrm{ker} K := \{u \in \ell^2_N : Ku = 0\} \quad \text{and} \quad \mathrm{im} K := \{Ku : u \in \ell^2_N\}.$$

The restriction of $K$ on $C \subseteq \ell^2_N$ is denoted by $K|_C : C \to H$ and defined by

$$K|_C(u) := Ku \quad \text{ for each } u \in C.$$

In particular, for $I \subseteq \mathcal{I}_N$, we write $K|_I$ for $K|_{E_I}$ for simplicity. Some useful properties are recalled in the following lemmas, in which Lemma 2.2 is a direct consequence of [2, Fact 2.18] and Lemma 2.1(c), while Lemma 2.3 is taken from [45, Lemma 2.3].

**Lemma 2.2.** *Let $I \subseteq \mathcal{I}_N$ be a finite index set. Then $(\ker(KP_{E_I}))^\perp = \mathrm{im}(P_{E_I}K^*)$.*

**Lemma 2.3.** *Let $I_1, I_2 \subseteq \mathcal{I}_N$ be such that $I_1 \cap I_2 = \emptyset$ and $\langle KP_{E_{I_1}}(u), KP_{E_{I_2}}(u)\rangle = 0$ for any $u \in \ell^2_N$. Then $\langle KP_{E_{I_1}}(x), KP_{E_{I_2}}(y)\rangle = 0$ for any $x, y \in \ell^2_N$.*

We end this section by recalling the notions of subdifferential and proximal operator of convex functions, as well as some useful facts. Let $u \in \ell^2_N$, and let $f : \ell^2_N \to \overline{\mathbb{R}}\,(:= \mathbb{R} \cup \{+\infty\})$ be a proper, lower semi-continuous (lsc) and convex function. The subdifferential of $f$ at $u$ is defined by

$$\partial f(u) := \{\xi \in \ell^2_N : f(v) \ge f(u) + \langle \xi, v - u\rangle \text{ for each } v \in \ell^2_N\} \tag{5}$$

The proximal operator of $f$ is defined by

$$\text{Prox}_f(u) := \arg\min_{v\in\ell_N^2} \left\{ f(v) + \frac{1}{2}\|v-u\|^2 \right\} \quad \text{for each } u \in \ell_N^2. \tag{6}$$

It was shown in [36, Proposition 1] that $\text{Prox}_f$ is nonexpansive if $f$ is proper, lsc and convex, that is,

$$\|\text{Prox}_f(u) - \text{Prox}_f(v)\| \le \|u - v\| \quad \text{for each } u, v \in \ell_N^2. \tag{7}$$

## 3. Iterative positive thresholding algorithm

Inspired by the ideas of the ISTA [11], this section aims to propose an iterative positive thresholding algorithm (IPTA) to solve the nonnegative $\ell_1$ regularization problem (3) and to investigate its convergence properties, including the global convergence and the linear convergence rate to a solution of problem (3). The IPTA is formally stated as follows.

**Algorithm IPTA**. Select an initial point $u^0 \in \ell_N^+$ satisfying $\sum_{k=1}^N \omega_k u_k^0 < \infty$ and two constants $\underline{s}$ and $\bar{s}$ satisfying $0 < \underline{s} \le \bar{s} < \frac{2}{\|K\|^2}$. For each $n \in \mathbb{N}$, having $u^n$, we choose a stepsize $s_n \in [\underline{s}, \bar{s}]$ and determine $u^{n+1}$ by

$$u^{n+1} := \left(u^n - s_n K^*(Ku^n - h) - s_n\omega\right)_+ . \tag{8}$$

By the stepsize rule in Algorithm IPTA, one sees that

$$0 < \underline{s} \le s_n \le \bar{s} < \frac{2}{\|K\|^2} \quad \text{for any } n \in \mathbb{N}. \tag{9}$$

The relationships between the IPTA and some well-known numerical algorithms are provided in the following remark, so as to advance our understanding of the proposed algorithm.

**Remark 1.** (i) Algorithm IPTA can be regarded as an application of the well-known proximal gradient algorithm (PGA) [3,10,23] to solve problem (10), which is equivalent to the nonnegative $\ell_1$ regularization problem (3). Indeed, problem (3) can be reformulated as a composite optimization problem:

$$\min_{u\in\ell_N^2} \frac{1}{2}\|Ku - h\|^2 + \varphi(u), \tag{10}$$

where $\varphi : \ell_N^2 \to \overline{\mathbb{R}}$ is defined by

$$\varphi(u) := \sum_{k=1}^N \left(\omega_k u_k + \delta_{\mathbb{R}_+}(u_k)\right) \quad \text{for each } u \in \ell_N^2 \tag{11}$$

($\delta_X$ denotes an indicator function on $X$). The PGA for solving problem (10) (see

[3,10,23] for details) has the iterative formula:

$$u^{n+1} := \mathrm{Prox}_{s_n\varphi}(u^n - s_n K^*(Ku^n - h)). \tag{12}$$

By (6) and the optimality condition of convex optimization [2, Theorem 16.2], we obtain that

$$0 \in u^{n+1} - (u^n - s_n K^*(Ku^n - h)) + s_n \partial\varphi(u^{n+1}). \tag{13}$$

By definition (5), one checks that

$$\partial\delta_{\mathbb{R}_+}(t) = \begin{cases} \{0\}, & t > 0, \\ (-\infty, 0], & t = 0, \\ \emptyset, & t < 0, \end{cases} \tag{14}$$

and by [2, Corollary 16.38] that

$$(\partial\varphi(u))_k = \omega_k + \partial\delta_{\mathbb{R}_+}(u_k) = \begin{cases} \{\omega_k\}, & u_k > 0, \\ (-\infty, \omega_k], & u_k = 0, \\ \emptyset, & u_k < 0, \end{cases} \quad \text{for each } k \in \mathcal{I}_N. \tag{15}$$

This, together with (13), shows that

$$u_k^{n+1} \in (u^n - s_n K^*(Ku^n - h))_k - s_n\omega_k - s_n\partial\delta_{\mathbb{R}_+}(u_k^{n+1}) \quad \text{for each } k \in \mathcal{I}_N;$$

consequently, one checks by (14) that $u^{n+1}$ given by (12) is of form (8), as desired.

(ii) The ISTA [11] is a popular iterative algorithm for solving the $\ell_1$ regularization problem, i.e., (10) with $\varphi$ replaced by $\psi : \ell_N^2 \to \overline{\mathbb{R}}$ is defined by

$$\psi(u) := \sum_{k=1}^{N} \omega_k |u_k| \quad \text{for each } u \in \ell_N^2.$$

The ISTA has the following iterative formula:

$$u^{n+1} := \mathbf{S}_{s_n\omega}(u^n - s_n K^*(Ku^n - h)),$$

where $\mathbf{S}_\tau : \ell_N^2 \to \ell_N^2$ is a soft thresholding operator, defined by

$$\mathbf{S}_\tau(v) := \mathrm{sign}(v) \odot (|v| - \tau)_+ \quad \text{for each } v \in \ell_N^2,$$

in which $\mathrm{sign}(\cdot)$ and $\odot$ operate the entrywise signum and the entrywise product, respectively. Note by (8) that the iterative process of the IPTA can be rewritten as

$$u^{n+1} := \mathbf{P}_{s_n\omega}(u^n - s_n K^*(Ku^n - h)),$$

where $\mathbf{P}_\tau : \ell_N^2 \to \ell_N^2$ is a positive thresholding operator, defined by

$$\mathbf{P}_\tau(v) := (|v| - \tau)_+ \quad \text{for each } v \in \ell_N^2.$$

Obviously, the main difference between the IPTA and the ISTA is the thresholding operator, where the IPTA maintains only the positive dominant components, while the ISTA preserves the dominant components, either positive or negative.

### 3.1. *Global convergence of IPTA*

This subsection is devoted to investigating the global convergence of the IPTA to a solution of problem (3). For the remainder of this section, we use $f : \ell^2_N \to \overline{\mathbb{R}}$ to denote the objective function of problem (10), that is,

$$f(u) := \frac{1}{2}\|Ku - h\|^2 + \varphi(u) \quad \text{for each } u \in \ell^2_N, \tag{16}$$

where $\varphi$ is defined by (11). For $\bar{u} \in S$ (the solution set of problem (3)), we write for simplicity that

$$\bar{v} := -K^*(K\bar{u} - h). \tag{17}$$

The equivalence between problem (3) and problem (10) has been stated in Remark 1(i), and so, $S$ is also the solution set of problem (10). Then, associated to problem (10), one can directly verify by using the optimality condition that

$$\bar{u} \in S \quad \Leftrightarrow \quad \bar{v}_k \begin{cases} = \omega_k, & \text{if } \bar{u}_k > 0, \\ \in (-\infty, \omega_k], & \text{if } \bar{u}_k = 0, \end{cases} \text{ for any } k \in \mathcal{I}_N. \tag{18}$$

Fix $\bar{u} \in S$. An index set $J$ is defined by

$$J := \{k \in \mathcal{I}_N : \bar{v}_k = \omega_k\}. \tag{19}$$

**Remark 2.** (i) The index set $J$ defined in (19) is a finite set. Indeed, it is trivial when $\ell^2_N = \mathbb{R}^N$; otherwise, by (4) and (19), one has that

$$|J|\underline{\omega}^2 \leq \sum_{k\in J} \omega_k^2 \leq \sum_{k\in J} (\bar{v}_k)^2 \leq \|\bar{v}\|^2 < \infty.$$

(ii) There exists $\rho \in (0, 1)$ such that $\bar{v}_k \leq \rho\,\omega_k$ for each $k \in J^c$. Indeed, set $\bar{v}_k := 0$ and $\omega_k := \underline{\omega}$ for each $k > N$ in the case when $N < \infty$. Then $\bar{v} \in \ell^2$ for each $N \in \mathbb{N} \cup \{+\infty\}$. Consequently, one has that $\lim_{k\to\infty} \bar{v}_k = 0$, and hence, it follows from (4) that $\lim_{k\to\infty} \frac{|\bar{v}_k|}{\omega_k} \leq \lim_{k\to\infty} \frac{|\bar{v}_k|}{\underline{\omega}} = 0$. Fix $\eta \in (0, 1)$. Then there exists $n \in \mathbb{N}$ such that $\frac{|\bar{v}_k|}{\omega_k} \leq \eta$ for each $k \geq n$. Define $\rho := \max\left\{\eta, \max\left\{\frac{\bar{v}_k}{\omega_k} : k \in J^c, k \leq n\right\}\right\}$. Then, one can check that $\rho \in (0, 1)$ and $\bar{v}_k \leq \rho\,\omega_k$ for each $k \in J^c$, as desired.

For the convergence analysis of the IPTA, we introduce an auxiliary function $R : \ell^2_N \to \overline{\mathbb{R}}$, defined by

$$R(u) := \varphi(u) - \varphi(\bar{u}) - \langle \bar{v}, u - \bar{u}\rangle \quad \text{for each } u \in \ell^2_N. \tag{20}$$

A lower bound of the function $R(\cdot)$ on a level set is provided in the following lemma, which is useful in proving the global convergence of the IPTA.

**Lemma 3.1.** *Let $\bar{u} \in S$ and $\alpha \in \mathbb{R}$. Then there exist $\tau > 0$ and a subspace $U \subseteq \ell_N^2$ such that $U^\perp$ is finite-dimensional and*

$$R(u) \geq \tau \|P_U(u - \bar{u})\|^2 \quad \textit{whenever } \varphi(u) \leq \alpha. \tag{21}$$

***Proof.*** Let $u \in \ell_N^2$ be such that $\varphi(u) \leq \alpha$. By (11), it can be shown that $u \in \ell_N^+$. Then, one has by (11) and (20) that

$$R(u) = \sum_{k=1}^{N} \omega_k u_k - \sum_{k=1}^{N} \omega_k \bar{u}_k - \langle \bar{v}, u - \bar{u} \rangle = \sum_{k=1}^{N} (\omega_k - \bar{v}_k)(u_k - \bar{u}_k).$$

Noting by (19) that $\omega_k = \bar{v}_k$ for each $k \in J$ and $\bar{u}_k = 0$ otherwise (cf. (18)), it follows that

$$R(u) = \sum_{k \in J^c} u_k (\omega_k - \bar{v}_k). \tag{22}$$

Note by Remark 2(ii) that there exists $\rho \in (0,1)$ such that $\bar{v}_k \leq \rho \omega_k$ for each $k \in J^c$. Recalling that $u \in \ell_N^+$, we obtain by (22) and (4) that

$$R(u) \geq \sum_{k \in J^c} (1 - \rho) \omega_k u_k \geq (1 - \rho) \underline{\omega} \left( \sum_{k \in J^c} u_k^2 \right)^{\frac{1}{2}}.$$

Note that $\underline{\omega} \left( \sum_{k \in J^c} u_k^2 \right)^{\frac{1}{2}} \leq \underline{\omega} \|u\| \leq \varphi(u) \leq \alpha$ and $\bar{u}_k = 0$ for each $k \in J^c$, and let $\tau := \frac{1}{\alpha} \underline{w}^2 (1 - \rho)$ and $U := E_{J^c}$ ($U^\perp = E_J$ is finite-dimensional, since $J$ is a finite set by Remark 2(i)). Then, it follows that

$$R(u) \geq \frac{1}{\alpha} \underline{w}^2 (1 - \rho) \sum_{k \in J^c} (u_k - \bar{u}_k)^2 = \tau \|P_U(u - \bar{u})\|^2,$$

which verifies (21), and the proof is complete. □

Now we establish the global convergence of the IPTA under the mild assumptions on algorithmic parameters made in Algorithm IPTA, which are same as the ones in [4, Theorem 1] for the ISTA.

**Theorem 3.2.** *Let $\{u^n\}$ be a sequence generated by Algorithm IPTA. Then $\{u^n\}$ strongly converges to a solution of problem* (3)*.*

***Proof.*** Let $\bar{u} \in S$, and define a sequence of scalars $\{r_n\}$ by

$$r_n := f(u^n) - f(\bar{u}) \quad \text{for each } n \in \mathbb{N}.$$

Then, one can check by (16), (17) and (20) that, for each $n \in \mathbb{N}$,

$$r_n - R(u^n) = \frac{1}{2} \|Ku^n - h\|^2 - \frac{1}{2} \|K\bar{u} - h\|^2 - \langle K^*(K\bar{u} - h), u^n - \bar{u} \rangle \geq 0. \tag{23}$$

Note by (11) that $\varphi$ is proper lsc and convex and by (16) that $f$ is coercive, and recall from Remark 1(i) that the IPTA is indeed the PGA for solving problem (10). Then, we conclude by [4, Proposition 2] that there exists $\rho > 0$ such that

$$r_n \le \rho n^{-1}, \tag{24}$$

and hence,

$$\varphi(u^n) \le f(u^n) = f(\bar{u}) + r_n \le f(\bar{u}) + \rho, \tag{25}$$

for each $n \in \mathbb{N}$. Then, one has by Lemma 3.1 (with $f(\bar{u}) + \rho$ in place of $\alpha$) that there exists $\tau > 0$ and a subspace $U \subseteq \ell^2_N$ such that $U^\perp$ is finite-dimensional and

$$R(u^n) \ge \tau \|P_U(u^n - \bar{u})\|^2 \quad \text{for each } n \in \mathbb{N}.$$

This, together with (23) and (24), shows that $\|P_U(u^n - \bar{u})\|^2 \le \frac{\rho}{\tau} n^{-1}$ for each $n \in \mathbb{N}$, and hence,

$$\lim_{n\to\infty} P_U(u^n) = P_U(\bar{u}). \tag{26}$$

Moreover, note by (4) and (25) that $\|P_{U^\perp}(u^n)\| \le \|u^n\| \le \underline{\omega}^{-1}\varphi(u^n) \le \underline{\omega}^{-1}(f(\bar{u}) + \rho)$ for each $n \in \mathbb{N}$. This shows that $\{P_{U^\perp}(u^n)\}$ is bounded. Recalling that $U^\perp$ is finite-dimensional, there exist a subsequence $\{u^{n_i}\}$ and $\tilde{u} \in U^\perp$ such that $\{P_{U^\perp}(u^{n_i})\}$ strongly converges to $\tilde{u}$. This, together with (26), implies that

$$\lim_{i\to\infty} u^{n_i} = \lim_{i\to\infty} P_U(u^{n_i}) + \lim_{i\to\infty} P_{U^\perp}(u^{n_i}) = P_U(\bar{u}) + \tilde{u}. \tag{27}$$

Note by (24) that each cluster point of $\{u^n\}$ is a solution of problem (3), and thus, $\bar{z} := P_U(\bar{u}) + \tilde{u} \in S$. Then, we obtain by Algorithm IPTA (i.e., PGA (12)), (7) and (9) that

$$\begin{aligned} \|u^{n+1} - \bar{z}\| &= \|\mathrm{Prox}_{s_n\varphi}(u^n - s_n K^*(Ku^n - h)) - \mathrm{Prox}_{s_n\varphi}(\bar{z} - s_n K^*(K\bar{z} - h))\| \\ &\le \|(I - s_n K^* K)(u^n - \bar{z})\| \\ &\le \|u^n - \bar{z}\|, \end{aligned}$$

which, together with (27), ensures the strong convergence of $\{x^n\}$ to $\bar{z}\,(\in S)$. The proof is complete. □

### 3.2. *Linear convergence of IPTA.*

This subsection is devoted to the linear convergence issue of the IPTA to a solution of problem (3). To this end, we first introduce a notion of positive orthogonal sparsity pattern (POSP). The POSP is inspired by the notion of orthogonal sparsity pattern (OSP), which was introduced in [45] to establish the linear convergence of the ISTA. The only difference between these two notions is that the index set $J$ is replaced by $\{k \in \mathcal{I}_N : |\bar{v}_k| = \omega_k\}$ in the OSP.

**Definition 3.3.** Let $\bar{u} \in S$, and let $J$ be defined by (19). A bounded linear operator $K : \ell_N^2 \to \mathcal{H}$ is said to satisfy the POSP at $\bar{u}$, if there exists an index set $I \subseteq \mathcal{I}_N$ with

$$\{k \in J : \bar{u}_k = 0\} \subseteq I \subseteq J \tag{28}$$

such that $K|_I$ is injective and

$$\langle KP_{E_I}(u), KP_{E_{J\setminus I}}(u)\rangle = 0 \quad \text{for any } u \in \ell_N^2. \tag{29}$$

We provide some sufficient conditions for the POSP in terms of $J$-basis injectivity property ($J$-BI), strict complementarity condition (SCC) or OSP.

(S1) $J$-BI: $K$ has the $J$-basis injective property (at $\bar{u} \in S$), i.e., $K|_J$ is injective.

(S2) SCC: The strict complementarity condition is satisfied at $\bar{u} \in S$, i.e.,

$$0 \in K^*(K\bar{u} - h) + \mathrm{ri}(\partial\varphi(\bar{u})), \tag{30}$$

where ri$C$ denotes the relative interior of $C$.

(S3) OSP: $K$ satisfies the OSP at $\bar{u} \in S$, i.e., set $\hat{J} := \{k \in \mathcal{I}_N : |\bar{v}_k| = \omega_k\}$, there exists $\hat{I} \subseteq \mathcal{I}_N$ with $\{k \in \hat{J} : \bar{u}_k = 0\} \subseteq \hat{I} \subseteq \hat{J}$ such that

$$K|_{\hat{I}} \text{ is injective and } \langle KP_{E_{\hat{I}}}(u), KP_{E_{\hat{J}\setminus\hat{I}}}(u)\rangle = 0 \text{ for each } u \in \ell_N^2. \tag{31}$$

**Remark 3.** (i) The $J$-BI property is natural to be satisfied in the context of (non-negative) sparse optimization, since the involved linear operators are often injective. Prominent examples are the Radon transform [28], solution operators for electrical impedance tomography [29] and Haar wavelet basis in image processing [3].

(ii) The SCC imposes a regular condition on the optimality condition of problem (3) (also the associated problem (10)). By (15), one can check that the SCC (30) is equivalent to either of the following conditions:

(a) $\mathrm{supp}(\bar{u}) = J$;

(b) $\bar{v}_k < \omega_k$ for each $k \notin \mathrm{supp}(\bar{u})$.

It is worth mentioning that the SCC of the $\ell_1$ regularization problem (1) was used in [20,38] to establish the linear convergence of the ISTA.

**Lemma 3.4.** *Let $K : \ell_N^2 \to \mathcal{H}$ be a bounded linear operator and $\bar{u} \in S$, and let $J$ be defined by* (19)*. Then $K$ satisfies the POSP at $\bar{u}$ provided either of* (S1)-(S3)*.*

***Proof.*** Suppose that (S1) holds. Then one checks by Definition 3.3 that the POSP is satisfied with $I := J$.

Suppose that (S2) holds. Then, by Remark 3(ii), one sees that $\{k \in J : \bar{u}_k = 0\} = \emptyset$; consequently, the POSP is satisfied with $I := \emptyset$.

Suppose that (S3) holds. Then there exists $\hat{I} \subseteq \mathcal{I}_N$ with $\{k \in \hat{J} : \bar{u}_k = 0\} \subseteq \hat{I} \subseteq \hat{J}$ such that (31) holds. Let $I := J \cap \hat{I}$. Clearly, $I \subseteq \hat{I}$, and thus, $K|_I$ is injective by (31). Noting by (19) and the definition of $\hat{J}$ in (S3) that $J \subseteq \hat{J}$, we obtain by the definition of $I$ that (28) holds and $J \setminus I = J \setminus \hat{I} \subseteq \hat{J} \setminus \hat{I}$. Fix $v \in \ell_N^2$. Let $u := P_{E_I}(v) + P_{E_{J\setminus I}}(v)$. Then it follows that

$$\langle KP_{E_I}(v), KP_{E_{J\setminus I}}(v)\rangle = \langle KP_{E_{\hat{I}}}(u), KP_{E_{\hat{J}\setminus\hat{I}}}(u)\rangle = 0$$

(due to (31)), that is, (29) holds. Therefore, the POSP is satisfied. The proof is complete. □

Below, we provide an example in infinite-dimensional space to show that the POSP is strictly weaker than the $J$-BI, SCC or OSP.

**Example 3.5.** Consider the problem (3) with $K : \ell^2 \to \ell^2$ being defined by

$$Ku := (u_1 + u_3, u_2 + u_4, u_5, u_6, u_7 \cdots)^T \quad \text{for each } u := (u_k) \in \ell^2,$$

$h := (2, -1, 3, 4, 5, \frac{1}{6}, \frac{1}{7}, \frac{1}{8}, \cdots)^T$ and $\omega := (1, 1, 1, 1, 3, 1, 1, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \cdots)^T$. Then it can be reformulated as

$$\sum_{k=1}^{2} \min_{u \geq 0} \left( \frac{1}{2}(u_k + u_{k+2} - h_k)^2 + u_k + u_{k+2} \right) + \sum_{k=5}^{\infty} \min_{u \geq 0} \left( \frac{1}{2}(u_k - h_{k-2})^2 + \omega_k u_k \right). \tag{32}$$

Let $\bar{u} \in S$. Clearly, the first two minimizations of problem (32) are equivalent to that $\bar{u}_1 + \bar{u}_3 = 1$ and $\bar{u}_2 = \bar{u}_4 = 0$, while the others are equivalent to that $\bar{u}_5 = 0$, $\bar{u}_6 = 3$, $\bar{u}_7 = 4$ and $\bar{u}_8 = 0$ for each $k \geq 8$. Hence, the solution set of problem (32) is

$$S = \left\{ \bar{u} = (a, 0, 1-a, 0, 0, 3, 4, 0, 0, \cdots)^T : 0 \leq a \leq 1 \right\}.$$

Write $\bar{x} := (1, 0, 0, 0, 0, 3, 4, 0, 0, \cdots)^T$, $\bar{y} := (0, 0, 1, 0, 0, 3, 4, 0, 0, \cdots)^T$. Then we have that

(i) none of $J$-BI, SCC or OSP is satisfied at any $\bar{u} \in S$;
(ii) POSP is satisfied at each $\bar{u} \in S \setminus \{\bar{x}, \bar{y}\}$.

Indeed, for each $\bar{u} \in S$, one checks that

$$\text{supp}(\bar{u}) = \begin{cases} \{1, 6, 7\}, & \text{if } \bar{u} = \bar{x}, \\ \{3, 6, 7\}, & \text{if } \bar{u} = \bar{y}, \\ \{1, 3, 6, 7\}, & \text{if } \bar{u} \in S \setminus \{\bar{x}, \bar{y}\}, \end{cases} \qquad \bar{v} = (1, -1, 1, -1, 3, 1, 1, \frac{1}{6}, \frac{1}{7}, \cdots)^T, \tag{33}$$

which, together with (19), implies that

$$J = \{1, 3, 5, 6, 7\} \quad \text{and} \quad K|_J = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}. \tag{34}$$

From (33) and (34), we observe that $K|_J$ is not injective and $\text{supp}(\bar{u}) \neq J$; consequently, we conclude by (S1) and Remark 3(ii) that neither $J$-BI or SCC is satisfied at any $\bar{u} \in S$. We write $\hat{J}_0 := \{k \in \hat{J} : \bar{u}_k = 0\}$ for simplicity. Then one checks by (33) that

$$\hat{J} = \{1, 2, 3, 4, 5, 6, 7\} \quad \text{and} \quad \hat{J}_0 = \begin{cases} \{3, 2, 4, 5\}, & \text{if } \bar{u} = \bar{x}, \\ \{1, 2, 4, 5\}, & \text{if } \bar{u} = \bar{y}, \\ \{2, 4, 5\}, & \text{if } \bar{u} \in S \setminus \{\bar{x}, \bar{y}\}. \end{cases}$$

This, together with (34), says that $K|_{\hat{J}_0}$ is not injective, and thus, OSP is not satisfied at any $\bar{u} \in S$. Therefore, assertion (i) is verified.

Fix $\bar{u} \in S \setminus \{\bar{x}, \bar{y}\}$. Note by (33) and (34) that $\{k \in J : \bar{u}_k = 0\} = \{5\}$. Let $I := \{5\}$. Then one checks by (34) that (28) and (29) hold and $K|_I$ is injective. Hence, the POSP is satisfies at each $\bar{u} \in S \setminus \{\bar{x}, \bar{y}\}$, i.e., assertion (ii) is proved.

The main theorem of this subsection is as follows, in which the Q-linear convergence of the IPTA is guaranteed under the assumption of the POSP.

**Theorem 3.6.** *Let $\{u^n\}$ be a sequence generated by Algorithm IPTA. Then $\{u^n\}$ strongly converges to a solution $\bar{u}$ of problem* (3). *Suppose that $K$ satisfies the POSP at $\bar{u}$. Then, $\{u^n\}$ converges linearly to $\bar{u}$, that is, there exist $\lambda \in (0, 1)$ and $M \in \mathbb{N}$ such that*

$$\|u^{n+1} - \bar{u}\| \leq \lambda \|u^n - \bar{u}\| \quad \textit{for any } n > M.$$

Recall from Remak 1(i) that the IPTA can be regard as an application of the PGA to problem (10), the properties of the PGA (see [4,45]) can be used to prove Theorem 3.6. The line of proof for Theorem 3.6 is similar to that of [45, Theorem 1.2], but with some technical differences. In particular, originating from the difference between the optimality conditions of the investigated models and the one between IPTA and ISTA (see Remak 1(ii)), the construction of the index set $J$ in (19) is different from the one in [45] (i.e., $\hat{J}$ defined in (S3)); consequently, the index set $I$ satisfying (28) is also different from the one in [45]. The proof of Theorem 3.6 consists of the analysis of the iterative procedure restricted on the index sets $I$, $J \setminus I$, and $J^c$, and $\{k \in \mathcal{I}_N : \bar{v}_k < -\omega_k\}$, a subset of $J^c$, is no longer empty, as in [45]. Hence there are some technical differences from that of [45, Theorem 1.2]. To make the paper self-contained, we provide the complete proof of Theorem 3.6 as follows.

To prove Theorem 3.6, we always assume, for the remainder of this section, that

(A1) $\{u^n\}$ is generated by Algorithm IPTA;
(A2) $\bar{u} := \lim_{n\to\infty} u^n \in S$.

Recall that $\bar{v}$ is defined by (17). For simplicity, we write

$$v^n := -K^*(Ku^n - h) \text{ for each } n \in \mathbb{N}; \tag{35}$$

consequently, it follows from (A2) that

$$\bar{v} = \lim_{n\to\infty} v^n. \tag{36}$$

The following four lemmas are presented for the proof of Theorem 3.6.

**Lemma 3.7.** *Let $V$ be a finite-dimensional subspace of $\ell^2_N$ and $\{s_n\}$ be a sequence satisfying* (9). *Suppose that $K|_V$ is injective. Then there exists $\lambda \in (0, 1)$ such that*

$$\|P_V - s_n P_V K^* K P_V\| \leq \lambda \quad \textit{for each } n \in \mathbb{N}. \tag{37}$$

***Proof.*** We have by definition that

$$
\begin{aligned}
&\|P_V - s_n P_V K^* K P_V\|^2 \\
&= \sup_{\|u\|=1} \langle (P_V - s_n P_V K^* K P_V)(u), (P_V - s_n P_V K^* K P_V)(u) \rangle \\
&= \sup_{\|u\|=1} (\|P_V(u)\|^2 - 2s_n \langle P_V(u), P_V K^* K P_V(u) \rangle + s_n^2 \|P_V K^* K P_V(u)\|^2) \\
&\le \sup_{\|u\|=1} \|P_V(u)\|^2 - 2s_n \left(1 - \tfrac{s_n}{2}\|K\|^2\right) \|K P_V(u)\|^2,
\end{aligned} \tag{38}
$$

where the last inequality follows from Lemma 2.1(b) and (d). By assumptions that $V$ is finite-dimensional and that $K|_V$ is injective, there exists $\alpha \in (0, \|K\|^2)$ such that $\|K P_V(u)\| \ge \alpha \|P_V(u)\|$ for any $u \in \ell_N^2$. Also note by (9) that $s_n \left(1 - \frac{s_n}{2}\|K\|^2\right) \ge \underline{s} \left(1 - \frac{\bar{s}}{2}\|K\|^2\right)$. Combining the above two inequalities, one deduces by (38) that

$$
\|P_V - s_n P_V K^* K P_V\|^2 \le \left(1 - 2\underline{s}\left(1 - \frac{\bar{s}}{2}\|K\|^2\right)\alpha^2\right) \sup_{\|u\|=1} \|P_V(u)\|^2.
$$

Noting by Lemma 2.1(b) that $\sup_{\|u\|=1} \|P_V(u)\|^2 \le 1$, (37) is seen to hold with $\lambda := \sqrt{1 - 2\underline{s}\left(1 - \frac{\bar{s}}{2}\|K\|^2\right)\alpha^2} \in (0,1)$. The proof is complete. □

**Lemma 3.8.** *Let $I \subseteq J$. Then, there exists $M \in \mathbb{N}$ such that*

$$
P_{E_{I^c}}(u^n - \bar{u}) = P_{E_{J \setminus I}}(u^n - \bar{u}) \quad \textit{for any } n > M. \tag{39}
$$

***Proof.*** By (18) and (19), one has that

$$
J^c = \{k \in \mathcal{I}_N : \bar{v}_k < \omega_k\} \subseteq \{k \in \mathcal{I}_N : \bar{u}_k = 0\}. \tag{40}
$$

Note by Remark 2(ii) that there exists $\rho \in (0,1)$ such that

$$
\bar{v}_k \le \rho\, \omega_k \quad \text{for each } k \in J^c. \tag{41}
$$

By assumption (A2) and (36), there exists $M \in \mathbb{N}$ such that

$$
\|u^n - \bar{u}\| \le \frac{1-\rho}{2}\underline{s}\ \underline{\omega} \quad \text{and} \quad \|v^n - \bar{v}\| \le \frac{1-\rho}{2}\underline{\omega} \quad \text{for each } n \ge M. \tag{42}
$$

Fix $i \in J^c$ and $n \ge M$. Note by (40) that $\bar{u}_i = 0$ and by (4) that $\omega_i \ge \underline{\omega} > 0$. Then, it follows from (42) and (9) that

$$
u_i^n \le |u_i^n| = |u_i^n - \bar{u}_i| \le \frac{1-\rho}{2}\underline{s}\ \underline{\omega} \le \frac{1-\rho}{2} s_n \omega_i,
$$

and from (41)-(42) and (4) that

$$
v_i^n \le |v_i^n - \bar{v}_i| + \bar{v}_i \le \frac{1-\rho}{2}\underline{\omega} + \rho\omega_i \le \frac{1+\rho}{2}\omega_i.
$$

Combining the above two inequalities, we obtain that $u_i^n + s_n v_i^n \le s_n \omega_i$; hence, one has by (8) and (35) that $u_i^{n+1} = (u_i^n + s_n v_i^n - s_n \omega_i)_+ = 0$. Since $i \in J^c$ is arbitrary, we have that $P_{E_{J^c}}(u^{n+1}) = 0$. Note that $E_{J\setminus I} \perp E_{J^c}$ and $E_{I^c} = E_{J\setminus I} + E_{J^c}$ (since $I \subseteq J$). Then, one has that

$$P_{E_{I^c}}(u^{n+1}) = P_{E_{J^c}}(u^{n+1}) + P_{E_{J\setminus I}}(u^{n+1}) = P_{E_{J\setminus I}}(u^{n+1}),$$

and by (40) that

$$P_{E_{I^c}}(\bar{u}) = P_{E_{J^c}}(\bar{u}) + P_{E_{J\setminus I}}(\bar{u}) = P_{E_{J\setminus I}}(\bar{u}).$$

Hence, (39) is obtained, and the proof is complete. □

**Lemma 3.9.** *Let $I \subseteq J$ be such that $K_I$ is injective and (29) is satisfied. Then, there exist $\lambda \in (0, 1)$ and $M \in \mathbb{N}$ such that*

$$\|P_{E_I}(u^{n+1} - \bar{u})\| \leq \lambda \|P_{E_I}(u^n - \bar{u})\| \quad \textit{for any } n > M. \tag{43}$$

***Proof.*** By assumption, Lemma 3.8 is applicable to concluding that there exists $M \in \mathbb{N}$ such that (39) holds. Fix $n > M$. In view of Algorithm IPTA (i.e., PGA (12)) and by (7) and Lemma 2.1(b), one has that

$$\begin{aligned}
&\|P_{E_I}(u^{n+1} - \bar{u})\| \\
&= \|P_{E_I}(\mathrm{Prox}_{s_n\varphi}(u^n - s_n K^*(Ku^n - h))) - P_{E_I}(\mathrm{Prox}_{s_n\varphi}(\bar{u} - s_n K^*(K\bar{u} - h)))\| \\
&= \|\mathrm{Prox}_{s_n\varphi}(P_{E_I}(u^n - s_n K^*(Ku^n - h))) - \mathrm{Prox}_{s_n\varphi}(P_{E_I}(\bar{u} - s_n K^*(K\bar{u} - h)))\| \\
&\leq \|P_{E_I}(u^n - s_n K^*(Ku^n - h)) - P_{E_I}(\bar{u} - s_n K^*(K\bar{u} - h))\| \\
&= \|P_{E_I}(I - s_n K^*K)(u^n - \bar{u})\| \\
&= \|P_{E_I}(I - s_n K^*K)P_{E_I}(u^n - \bar{u}) + P_{E_I}(I - s_n K^*K)P_{E_{I^c}}(u^n - \bar{u})\| \\
&= \|(P_{E_I} - s_n P_{E_I} K^*KP_{E_I})P_{E_I}(u^n - \bar{u}) - s_n P_{E_I} K^*KP_{E_{I^c}}(u^n - \bar{u})\|
\end{aligned} \tag{44}$$

(due to the fact that $P_{E_I}P_{E_{I^c}} = 0$ and $P_{E_I}$ is linear idempotent). Note by (39) and Lemma 2.1(c)-(d) that

$$\begin{aligned}
\|P_{E_I}K^*KP_{E_{I^c}}(u^n - \bar{u})\|^2 &= \|P_{E_I}K^*KP_{E_{J\setminus I}}(u^n - \bar{u})\|^2 \\
&= \langle KP_{E_{J\setminus I}}(u^n - \bar{u}), KP_{E_I}K^*KP_{E_{J\setminus I}}(u^n - \bar{u})\rangle.
\end{aligned}$$

By assumption that (29) is satisfied, Lemma 2.3 is applicable (with $I$, $J \setminus I$ in place of $I_1$, $I_2$); hence, we obtain from above that $P_{E_I}K^*KP_{E_{I^c}}(u^n - \bar{u}) = 0$. Then, (44) is reduced to

$$\|P_{E_I}(u^{n+1} - \bar{u})\| \leq \|P_{E_I} - s_n P_{E_I}K^*KP_{E_I}\|\|P_{E_I}(u^n - \bar{u})\|.$$

This, together with Lemma 3.7, confirms (43), and the proof is complete. □

**Lemma 3.10.** *Let $I \subseteq \mathcal{I}_N$ be such that (28) and (29) are satisfied. Then, there exist $\lambda \in (0, 1)$ and $M \in \mathbb{N}$ such that*

$$\|P_{E_{I^c}}(u^{n+1} - \bar{u})\| \leq \lambda \|P_{E_{I^c}}(u^n - \bar{u})\| \quad \textit{for any } n > M. \tag{45}$$

***Proof.*** Define an index set

$$T := \{k \in \mathcal{I}_N : \bar{u}_k > 0\}. \tag{46}$$

It follows from (18) and (19) that $T \subseteq J$, which is a finite set (see Remark 2(i)). Let $\tau := \min\{\bar{u}_k : k \in T\} > 0$ and fix $i \in T$. Then, we have by (18) that

$$\bar{u}_i \geq \tau > 0 \quad \text{and} \quad \bar{v}_i = \omega_i. \tag{47}$$

For this $\tau > 0$, by assumption (A2) and (36), there exists $M \in \mathbb{N}$ such that

$$|u_i^n - \bar{u}_i| \leq \frac{\tau}{2} \quad \text{and} \quad |v_i^n - v_i| \leq \frac{\tau}{2\bar{s}} \quad \text{for any } n > M. \tag{48}$$

Fix $n > M$. Note that $u_i^n + s_n v_i^n \geq \bar{u}_i - |u_i^n - \bar{u}_i| - s_n |v_i^n - \bar{v}_i| + s_n \bar{v}_i$. This, together (9) and (47)-(48), yields that $u_i^n + s_n v_i^n - s_n \omega_i \geq 0$. Then, one has by (8) and (47) that

$$u_i^{n+1} = u_i^n + s_n v_i^n - s_n \bar{v}_i = ((I - s_n K^* K)(u^n - \bar{u}))_i + \bar{u}_i$$

(due to (17) and (35)). Noting by (18), (28) and (46) that $J \setminus I \subseteq T$, and recalling that $i \in T$ is arbitrary, we conclude that

$$P_{E_{J\setminus I}}(u^{n+1}) = P_{E_{J\setminus I}}(I - s_n K^* K)(u^n - \bar{u}) + P_{E_{J\setminus I}}(\bar{u}). \tag{49}$$

Let $U := \ker(K P_{E_{J\setminus I}})$. By Lemma 2.2, one has that

$$U^\perp = \operatorname{im}(P_{E_{J\setminus I}} K^*). \tag{50}$$

Employing $P_U$ on both sides of (49), we obtain by Lemma 2.1(b) that

$$P_U P_{E_{J\setminus I}}(u^{n+1}) = P_U P_{E_{J\setminus I}}(u^n) - s_n P_U P_{E_{J\setminus I}} K^* K(u^n - \bar{u}). \tag{51}$$

Noting by (50) that $P_{E_{J\setminus I}} K^* K(u^n - \bar{u}) \in U^\perp$, it is easy to see from Lemma 2.1(a) that $P_U P_{E_{J\setminus I}} K^* K(u^n - \bar{u}) = 0$. This, together with (51), implies that

$$P_U P_{E_{J\setminus I}}(u^{n+1}) = P_U P_{E_{J\setminus I}}(u^n).$$

Noting by assumption (A2) that $\lim_{n\to\infty} u^n = \bar{u}$ and that $n > M$ is arbitrary, we obtain by Lemma 2.1(b) that $P_U P_{E_{J\setminus I}}(u^n) = P_U P_{E_{J\setminus I}}(\bar{u})$, and thus,

$$P_{E_{J\setminus I}}(u^n - \bar{u}) = P_U P_{E_{J\setminus I}}(u^n - \bar{u}) + P_{U^\perp} P_{E_{J\setminus I}}(u^n - \bar{u}) = P_{U^\perp} P_{E_{J\setminus I}}(u^n - \bar{u}). \tag{52}$$

Employing $P_{U^\perp}$ on both sides of (49), we have by Lemma 2.1(b) that

$$\begin{aligned} &P_{U^\perp} P_{E_{J\setminus I}}(u^{n+1} - \bar{u}) \\ &= P_{U^\perp} P_{E_{J\setminus I}}(I - s_n K^* K) P_{E_{I^c}}(u^n - \bar{u}) + P_{U^\perp} P_{E_{J\setminus I}}(I - s_n K^* K) P_{E_I}(u^n - \bar{u}) \\ &= P_{U^\perp} P_{E_{J\setminus I}}(I - s_n K^* K) P_{E_{I^c}}(u^n - \bar{u}) - s_n P_{U^\perp} P_{E_{J\setminus I}} K^* K P_{E_I}(u^n - \bar{u}). \end{aligned} \tag{53}$$

By assumption, Lemma 3.8 is applicable to concluding (39). Then it follows that

$$\begin{aligned} P_{U^\perp} P_{E_{J\setminus I}}(I - s_n K^* K) P_{E_{I^c}}(u^n - \bar{u}) &= P_{U^\perp} P_{E_{J\setminus I}}(I - s_n K^* K) P_{E_{J\setminus I}}(u^n - \bar{u}) \\ &= (P_{U^\perp} - s_n P_{U^\perp} P_{E_{J\setminus I}} K^* K P_{U^\perp}) P_{E_{J\setminus I}}(u^n - \bar{u}) \end{aligned} \tag{54}$$

(due to (52)). By definition of $U^\perp$ (cf. (50)), one has that $P_{U^\perp}P_{E_{J\setminus I}}K^*v = P_{U^\perp}K^*v$ for any $v \in H$. This, together with (54), implies that

$$P_{U^\perp}P_{E_{J\setminus I}}(I - s_nK^*K)P_{E_{I^c}}(u^n - \bar{u}) = (P_{U^\perp} - s_nP_{U^\perp}K^*KP_{U^\perp})P_{E_{J\setminus I}}(u^n - \bar{u}). \quad (55)$$

On the other hand, by (50) and Lemma 2.1(c)-(d), we have that

$$\|P_{U^\perp}P_{E_{J\setminus I}}K^*KP_{E_I}(u^n - \bar{u})\|^2 = \langle KP_{E_I}(u^n - \bar{u}), KP_{E_{J\setminus I}}P_{U^\perp}P_{E_{J\setminus I}}K^*KP_{E_I}(u^n - \bar{u})\rangle.$$

Since (29) is assumed, one sees by Lemma 2.3 that $P_{U^\perp}P_{E_{J\setminus I}}K^*KP_{E_I}(u^n - \bar{u}) = 0$. Further note by (39) and (52) that $P_{E_{I^c}}(u^n - \bar{u}) = P_{E_{J\setminus I}}(u^n - \bar{u}) = P_{U^\perp}P_{E_{J\setminus I}}(u^n - \bar{u})$. The above two equalities, together with (53) and (55), yields

$$P_{E_{I^c}}(u^{n+1} - \bar{u}) = (P_{U^\perp} - s_nP_{U^\perp}K^*KP_{U^\perp})P_{E_{I^c}}(u^n - \bar{u}). \quad (56)$$

Note by (50) that $U^\perp$ is finite-dimensional and $K|_{U^\perp}$ is injective. Thus, by the Lemma 3.7, we conclude that there exists $\lambda \in (0,1)$ such that $\|P_{U^\perp} - s_nP_{U^\perp}K^*KP_{U^\perp}\| \le \lambda$ for each $n \in \mathbb{N}$. This, together with (56), derives (45), and the proof is complete. □

Now we are ready to provide the proof of the linear convergence of the IPTA as follows.

***Proof of Theorem 3.6.*** It has been shown in Theorem 3.2 that $\{u^n\}$ strongly converges to $\bar{u} \in S$, and hence, the assumptions (A1) and (A2) made in this subsection are satisfied. By the assumption of the POSP, the assumptions of Lemmas 3.9 and 3.10 are satisfied. Then the conclusion follows. □

## 4. Numerical experiments

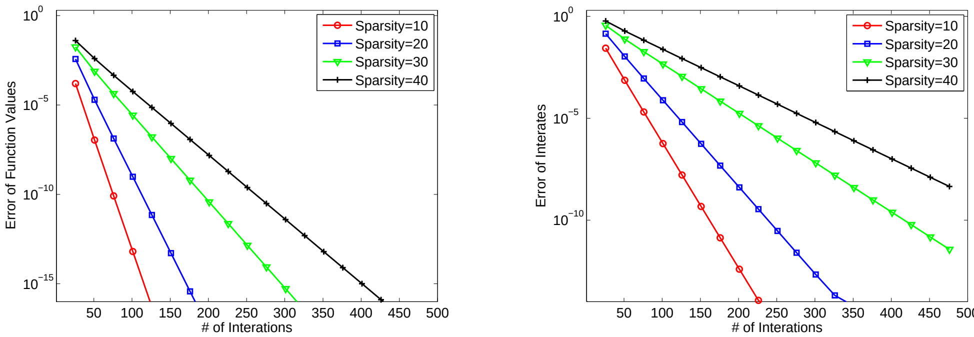
The purpose of this section is to demonstrate the numerical performance of the proposed IPTA for the nonnegative $\ell_1$ regularization problem, and to compare with several state-of-the-art algorithms in the simulation of compressive sensing. All numerical experiments are implemented in Matlab R2014a and executed on a personal desktop (Intel Core Duo E8500, 3.16 GHz, 4.00 GB of RAM).

In the numerical experiments, the simulation data are generated by the standard process of compressive sensing. The matrix $A$ is randomly generated via the following formats, which are popular linear transform matrices in compressive sensing:

- standard Gaussian: it is a standard transform matrix in compressive sensing [20,23,43], where we randomly generate an i.i.d. Gaussian ensemble $A \in \mathbb{R}^{m\times n}$ satisfying $A^\top A = I$.
- discrete cosine transform (DCT): DCT has been widely used in the digital signal and image processing [32,34], where $A$ is given by

$$A_{ij} := \frac{\min(j, \sqrt{2})}{\sqrt{n}} \cos\left(\frac{(2i-1)(j-1)\pi}{2n}\right) \quad \text{for any } 1 \le i, j \le n.$$

The nonnegative sparse solution $\bar{x} \in \mathbb{R}^n$ is generated via randomly picking $k$ of its components as active elements, whose entries are drawn from the standard uniform




(a) Linear convergence of function values.

(b) Linear convergence of iterates.

**Figure 1.** Linear convergence rate of IPTA.

**Table 1.** List of the algorithms compared in the numerical study.

| Abbreviations | Algorithms |
|---|---|
| IPTA | **I**terative **P**ositive **T**hresholding **A**lgorithm. |
| ISTA | **I**terative **S**oft **T**hresholding **A**lgorithm [11]. |
| NADMM | **N**onnegative **A**lternating **D**irection **M**ethod of **M**ultiplier [16]. |
| ADMM | **A**lternating **D**irection **M**ethod of **M**ultiplier [43]. |
| NOMP | **N**onnegative **O**rthogonal **M**atching **P**ursuit algorithm [42]. |
| OMP | **O**rthogonal **M**atching **P**ursuit algorithm [40]. |

distribution on $(0, 1)$, while the remaining components are all set to be zeros. Then we generate the observation data $b$ by the Matlab script

$$b = A * \bar{x} + sigma * randn(m, 1),$$

where $sigma$ is the standard deviation of additive Gaussian noise. The problem size is set as $n = 1024$ and $m = 256$ in all numerical experiments, and the stepsize and the initial point are respectively set as $v_k \equiv \frac{1}{2}$ and $x_0 = 0$ in the tests of IPTA.

We first verify the linear convergence rate of IPTA by conducting extensive simulations on standard Gaussian and DCT data, in which the noisy measurement is waived as $sigma = 0$. Figure 1 plots the error of function values or iterates along with the number of iterations at different sparsity levels in a random trial. Figure 1(a) illustrates the linear convergence to a minimal value, and Figure 1(b) demonstrates the linear convergence to a minimum, which are consistent with the theoretical result in Theorem 3.6. In this experiment, we also note that the short CPU times of IPTA, about 0.2 second per 500 iterations.

We then compare IPTA with several existing algorithms in the field of sparse optimization, including the well-known ISTA, NADMM, ADMM, NOMP and OMP, as listed in Table 1. Among these solvers, IPTA, NADMM and NOMP are nonnegative solvers for nonnegative sparse optimization problem, while ISTA, ADMM and OMP are solvers for sparse optimization problem. The noisy measurement is set as $sigma = 0.1\%$. All these solvers can successfully recover the signal when the solution is of low sparse level (i.e., $k$ is small). However, some of these solvers fails to obtain the nonnegative sparse solution along with the sparsity level increases. Figure 2 illustrates the signals estimated by these algorithms in a random trial on the standard Gaussian data at a sparsity level of $k = 70$, and Figure 3 displays the signals estimated by these algorithms in a random trial on the DCT data at a sparsity level of $k = 40$, in which the true signal is denoted by circles (red), and their estimates are denoted by asterisks (blue). It is illustrated from Figures 2 and 3 that the nonnegative solvers, including IPTA, NADMM and NOMP, achieve the nonnegative sparse solutions that approaches to the true signal, while the other solvers may not obtain the nonnegative sparse solutions. Among these three nonnegative solvers, it is further observed from Figure 2 that IPTA achieves a more exact prediction and accurate estimation than NADMM and NOMP for the standard Gaussian data, in which the former one does not obtain the accurate estimation of variables while the latter one is not able to predict the variables of small magnitude. Figure 3 shows that IPTA also outperforms NADMM and NOMP when dealing with the DCT data. This shows the advantage of IPTA in approaching the nonnegative sparse solutions of linear inverse problems.

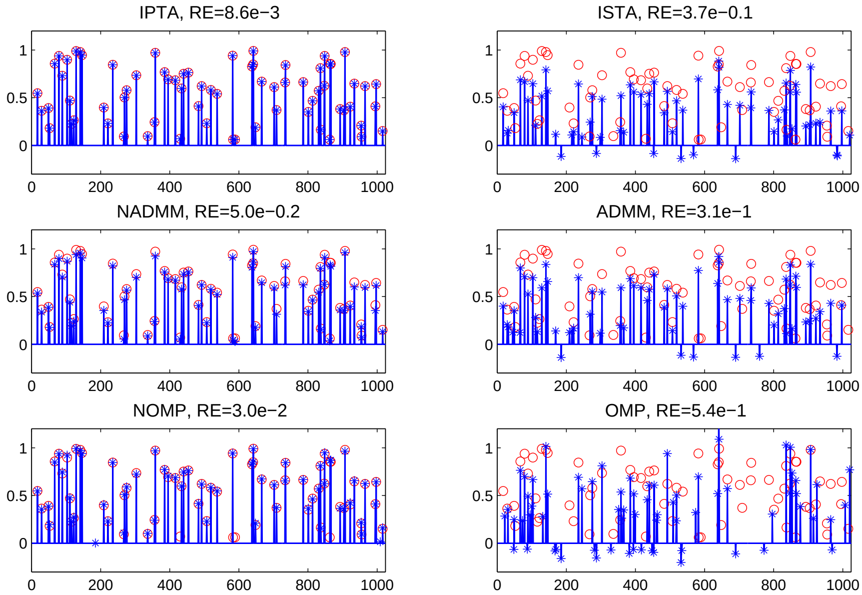



**Figure 2.** Simulation of IPTA and several existing algorithms for standard Gaussian.

Finally, we show the robustness analysis of these three nonnegative solvers (IPTA, NADMM and NOMP) by carrying out extensive simulations on standard Gaussian and DCT data. In particular, for each sparsity level, we randomly generate the data $A$, $\bar{x}$, $b$ (as above) 500 times, run the solver, and summarize these numerical results to illustrate the robustness of the solver. The noisy measurement is set as $sigma = 0.1\%$. The performance is measured by the successful recovery rate, where each recovery is defined as *success* when the relative error (RE) between the estimated solution and the true signal is smaller than 0.5%; otherwise, it is regarded as *failure*. Figure 4 demonstrates the overall performance of these nonnegative solvers by plotting the successful recovery rates along with different sparsity levels. It is indicated by Figure 4 that IPTA can achieve a higher successful recovery rate than NADMM and NOMP for both standard Gaussian and DCT data.

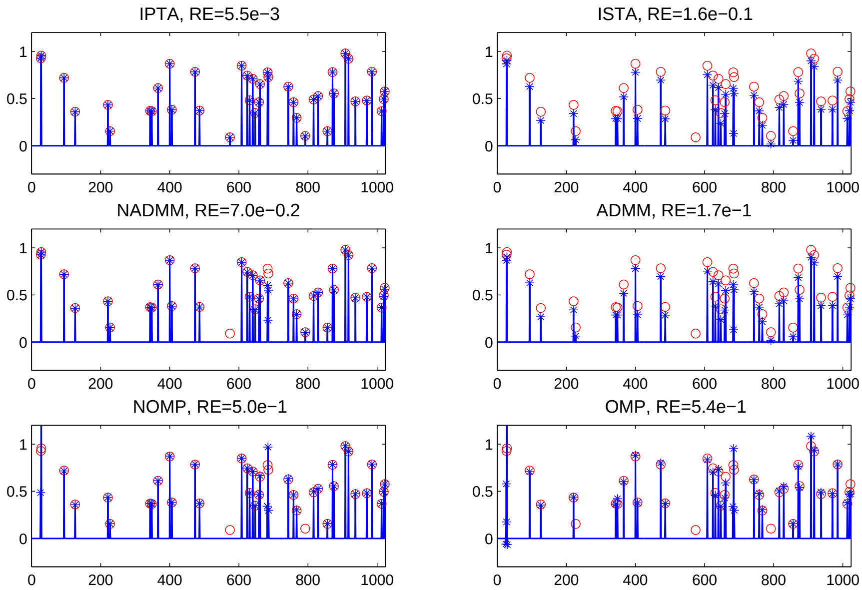
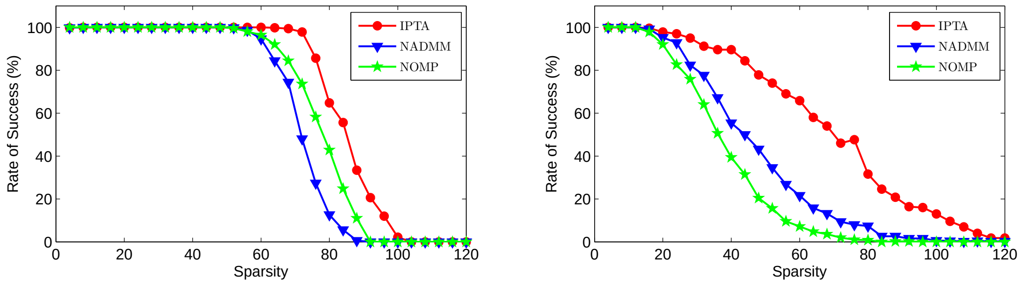



**Figure 3.** Simulation of IPTA and several existing algorithms for DCT.




(a) standard Gaussian (b) DCT

**Figure 4.** Probabilities of exact reconstruction by nonnegative solvers.

## Funding

Yaohua Hu was supported in part by the National Natural Science Foundation of China [grant number 11601343, 11601344]; Natural Science Foundation of Guangdong [grant number 2016A030310038, 2013B040403005]; Interdisciplinary Innovation Team of Shenzhen University; Carisa Kwok Wai Yu was supported in part by the Research Grants Council of the Hong Kong Special Administrative Region, China [grant number UGC/FDS14/P02/15]; Jinhua Wang was supported in part by the National Natural Science Foundation of China [grant number 11371325, 11771397]; Zhejiang Provincial Natural Science Foundation of China [grant number LY17A010021].

## Acknowledgements

The authors are grateful to two anonymous reviewers for their valuable suggestions and remarks which allow us to improve the original presentation of the paper. The authors are also grateful to Prof. Xiaoqi Yang (The Hong Kong Polytechnic University) and Prof. Chong Li (Zhejiang University) for their valuable comments and providing the research topics.

## References

[1] Bach F, Jenatton R, Mairal J, Obozinski G. Structured sparsity through convex optimization. Stat Sci. 2012;27:450–468.

[2] Bauschke HH, Combettes PL. Convex Analysis and Monotone Operator Theory in Hilbert Spaces. New York(NY): Springer;2011.

[3] Beck A, Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J Imaging Sci. 2009;2:183–202.

[4] Bredies K, Lorenz DA. Linear convergence of iterative soft-thresholding. J Fourier Anal Appl. 2008;14:813–837.

[5] Bruckstein AM, Elad E, Zibulevsky M. On the uniqueness of nonnegative sparse solutions to underdetermined systems of equations. IEEE Trans Inform Theory. 2008;14:4813–4820.

[6] Bazavan EG, Li F, Sminchisescu C. Fourier Kernel Learning. Proceedings of 12th European Conference on Computer Vision. Berlin: Springer; 2012;459–473.

[7] Candès E, Tao T. Decoding by linear programming. IEEE Trans Inform Theory. 2005;51:4203–4215.

[8] Chen SS, Donoho DL, Saunders MA. Atomic decomposition by basis pursuit. SIAM Rev. 2001;43:129–159.

[9] Chen X, Peng J, Zhang S. Sparse solutions to random standard quadratic optimization problems. Math Program. 2013;141:273–293.

[10] Combettes PL, Wajs VR. Signal recovery by proximal forward-backward splitting. Multiscale Model Sim. 2005;4:1168–1200.

[11] Daubechies I, Defrise M, Mol CD. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. Commun Pur Appl Math. 2004;57:1413–1457.

[12] Donoho DL. Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. Appl Comput Harmon Anal. 1995;2:101–126.

[13] Donoho DL. Compressed sensing. IEEE Trans Inform Theory. 2006;52:1289–1306.

[14] Donoho DL, Tanner J. Sparse nonnegative solution of underdetermined linear equations by linear programming. P Natl Acad Sci USA. 2005;102:9446–9451.

[15] Elad M. Sparse and Redundant Representations. New York(NY):Springer;2010.

[16] Esser E, Lou Y, Xin J. A method for finding structured sparse solutions to nonnegative least squares problems with applications. SIAM J Imaging Sci. 2013;6:2010–2046.

[17] Figueiredo MAT, Nowak RD. An EM algorithm for wavelet-based image restoration. IEEE Trans Image Process. 2003;12:906–916.

[18] Figueiredo MAT, Nowak RD, Wright SJ. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. IEEE J Sel Top Signa. 2007;1:586–597.

[19] Ge D, He R, He S. An improved algorithm for the $L_2$-$L_p$ minimization problem. Math Program. 2017;1–28.

[20] Hale ET, Yin W, Zhang Y. Fixed-point continuation for $\ell_1$-minimization: Methodology and convergence. SIAM J Optim. 2008;19:1107–1130.

[21] He B, Yuan X. On the $O(1/n)$ convergence rate of the Douglas-Rachford alternating direction method. SIAM J Numer Anal. 2012;50:700–709.

[22] He R, Zheng WS, Hu BG, Kong XW. Two-stage nonnegative sparse representation for

large-scale face recognition. IEEE Trans Neural Netw Learn Syst. 2013;24:35–46.

[23] Hu YH, Li C, Meng KW, Qin J, Yang XQ. Group sparse optimization via $\ell_{p,q}$ regularization. J Mach Learn Res. 2017;18:960–1011.

[24] Hu YH, Li C, Yang XQ. On convergence rates of linearized proximal algorithms for convex composite optimization with applications. SIAM J Optim. 2016;26:1207–1235.

[25] Khajehnejad MA, Dimakis AG, Xu W, Hassibi B. Sparse recovery of nonnegative signals with minimal expansion. IEEE Trans Signal Proces. 2011;59:196–208.

[26] Lu Z, Xiao L. On the complexity analysis of randomized block-coordinate descent methods. Math Program. 2015;152:615–642.

[27] Lu Z, Zhang Y, Lu J. $\ell_p$ Regularized low-rank approximation via iterative reweighted singular value minimization, Comput Optim Appl, 2017;68:619–642.

[28] Maass P. The interior Radon transform. SIAM J Appl Math. 1992;52:710–724.

[29] Nachman AI. Global uniqueness for a two-dimensional inverse boundary value problem. Ann of Math. 1996;143:71–96.

[30] Nesterov Y. Efficiency of coordinate descent methods on huge-scale optimization problems. SIAM J Optim. 2012;22:341–362.

[31] Nesterov Y. Gradient methods for minimizing composite functions. Math Program. 2013;14:125–161.

[32] Oppenheim A, Schafer R, Buck J. Discrete-Time Signal Processing. New Jersey (NJ): Prentice Hall;1999.

[33] Qin J, Hu YH, Xu F, Yalamanchili HK, Wang J. Inferring gene regulatory networks by integrating ChIP-seq/chip and transcriptome data via LASSO-type regularization methods. Methods. 2014;67:294–303.

[34] Rao K, Yip P. Discrete Cosine Transform: Algorithms, Advantages, Applications. Boston: Academic Press;1990.

[35] Rish I, Cecchi GA, Lozano A, Niculescu-Mizil A. Sparse Recovery for Protein Mass Spectrometry Data. Cambridge Massachusetts:MIT Press;2014.

[36] Rockafellar RT. Monotone operators and the proximal point algorithm. SIAM J Control Optim. 1976;14:877–898.

[37] Simon N, Friedman J, Hastie T, Tibshirani R. A sparse-group Lasso. J Comput Graph Stat. 2013;22:231–245.

[38] Tao S, Boley D, Zhang S. Local linear convergence of ISTA and FISTA on the LASSO problem. SIAM J Optim. 2016;26:313–336.

[39] Tibshirani R. Regression shrinkage and selection via the Lasso. J Roy Stat Soc B Met. 1996;58:267–288.

[40] Tropp JA. Greed is good: Algorithmic results for sparse approximation. IEEE Trans Inform Theory. 2004;50:2231–2242.

[41] Wang J, Hu YH, Li C, Yao JC. Linear convergence of CQ algorithms and applications in gene regulatory network inference. Inverse Problems. 2017;33:055017.

[42] Yaghoobi M, Wu D, Davies ME. Fast non-negative orthogonal matching pursuit. IEEE Signal Proc Let. 2015;22:1229–1233.

[43] Yang J, Zhang Y. Alternating direction algorithms for $\ell_1$-problems in compressive sensing. SIAM J Sci Comput. 2011;33:250–278.

[44] Yuan M, Lin Y. On the non-negative garrotte estimator. J Roy Stat Soc B. 2007;69:143–161.

[45] Zhang LF, Hu YH, Li C, Yao JC. A new linear convergence result for the iterative soft thresholding algorithm. Optimization. 2017;66:1177–1189.