A new linear convergence result for the iterative soft thresholding algorithm

Lufang Zhang* Yaohua Hu† Chong Li‡ Jen-Chih Yao§

Abstract

The iterative soft thresholding algorithm (ISTA) is one of the most popular optimization algorithms for solving the ℓ_1 regularized least square problem, and its linear convergence has been investigated under the assumption of finite basis injectivity property or strict sparsity pattern. In this paper, we consider the ℓ_1 regularized least square problem in finite- or infinite-dimensional Hilbert space, introduce a weaker notion of orthogonal sparsity pattern (OSP), and establish the Q-linear convergence of ISTA under the assumption of OSP. Examples are provided to illustrate the cases where the linear convergence of ISTA can be established only by our result, but cannot be ensured by any existing result in the literature.

Keyword: Iterative soft thresholding algorithm; Linear convergence analysis; Linear inverse problems; Sparsity pattern

Mathematics Subject Classification (2000): 65K05, 46N10, 65J22

1 Introduction

Let H be a Hilbert space, and let l^2 denote the Hilbert space consisting of all square-summable sequences. Let $N \in \mathbb{N} \cup \{+\infty\}$ be fixed, and write

$$l_N^2 := \begin{cases} \mathbb{R}^N, & \text{if } N \in \mathbb{N}, \\ l^2, & \text{otherwise}, \end{cases} \quad \text{and} \quad \mathcal{I}_N := \begin{cases} \{1, \ldots, N\}, & \text{if } N \in \mathbb{N}, \\ \mathbb{N}, & \text{otherwise}. \end{cases}$$

In this paper, we consider the following ℓ_1 regularized least square problem

$$\min_{u \in l^2} \frac{1}{2}\|Ku - h\|^2 + \sum_{k=1}^{N} \omega_k |u_k|, \tag{1.1}$$

*School of Mathematical Sciences, Zhejiang University, Hangzhou 310027, P. R. China (21435030@zju.edu.cn).

†College of Mathematics and Statistics, Shenzhen University, Shenzhen 518060, P. R. China (mayhhu@szu.edu.cn). This author was partially supported by the National Natural Science Foundation of China (11601343), Natural Science Foundation of Guangdong (2016A030310038), and Foundation for Distinguished Young Talents in Higher Education of Guangdong (2015KQNCX145).

‡School of Mathematical Sciences, Zhejiang University, Hangzhou 310027, P. R. China (cli@zju.edu.cn). This author was partially supported by the National Natural Science Foundation of China (11571308).

§Center for General Education, China Medical University, Taichung 40402, Taiwan (yaojc@maill.cmu.edu.tw). This author was partially supported by the grant MOST 105-2115-M-039-002-MY3.

where $K : l^2_N \to H$ is a bounded linear operator, and $\omega := (\omega_k)$ is a sequence of weights satisfying

$$\omega_k \geq \underline{\omega} > 0 \quad \text{for any } k \in \mathcal{I}_N. \tag{1.2}$$

In the last decade, problem (1.1) has been widely studied to approach a sparse approximate solution of the linear inverse problem and gained successful applications in a wide range of fields, such as compressive sensing [8,9,11], image science [4,6,13], systems biology [25,27,30] and machine learning [1, 19, 22] in finite-dimensional spaces; and Fourier analysis [2, 7] and Harmonic analysis [12, 16] in infinite-dimensional spaces.

Motivated by successful applications of the ℓ_1 regularization problem (1.1), many practical and efficient optimization algorithms have been proposed to solve problem (1.1); see [9, 15, 18–23, 31, 32] and references therein. In particular, the iterative soft thresholding algorithm (in short, ISTA) is one of the most widely studied first-order iterative algorithms for solving problem (1.1). The ISTA was originally proposed to solve the image deconvolution problem in Euclidean spaces, independently introduced by Figueiredo and Nowak [14] to approach a penalized maximum likelihood estimator under the name of EM algorithm, and by Starck et al. [28] to minimize a total-variation regularized least square problem; and it was first investigated in [10] for Hilbert spaces. The ISTA is formally described as follows.

Algorithm 1. Let an initial point $u^0 \in l^2_N$ be given. Having u^n, we choose a step size $s_n > 0$ and determine u^{n+1} by

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

where $\mathbf{S}_{s_n\omega} : l^2_N \to l^2_N$ is a soft thresholding operator, defined by

$$\mathbf{S}_{s_n\omega}(v) := (\operatorname{sign}(v_k) \cdot (|v_k| - s_n\omega_k)_+) \quad \text{for each } v := (v_k) \in l^2_N. \tag{1.3}$$

Under the assumption that the step sizes $\{s_n\}$ satisfy

$$0 < \underline{s} \leq s_n \leq \overline{s} < \frac{2}{\|K\|^2} \quad \text{for any } n \in \mathbb{N}, \tag{1.4}$$

the (strong) convergence result of the ISTA with the initial point $u^0 \in l^2_N$ satisfying

$$\sum_{k=1}^{N} \omega_k |u^0_k| < \infty \tag{1.5}$$

(noting that this condition holds automatically for any $u^0 \in l^2_N$ in the case when $N < \infty$) has been established in [5] for finite-dimensional spaces and in [10] for infinite-dimensional spaces, respectively.

In recent years, many articles have been devoted to the study of convergence rates of the ISTA, including convergence rates in terms of objective values and iterates; see [4,7,17,24,29] and references therein. In this paper, we concentrate on the linear convergence of the ISTA in terms of iterates. For $I \subseteq \mathcal{I}_N$, we define E_I and $K|_I : E_I \to H$ respectively by

$$E_I := \{u \in l^2_N : u_k = 0 \text{ for each } k \in I^c\} \quad \text{and} \quad K|_I(u) := Ku \text{ for each } u \in E_I. \tag{1.6}$$

K is said to satisfy the I-basis injective property (in short, I-BI) if $K|_I$ is injective.

Based on the convergence result and under the basic assumptions (1.4) and (1.5), the linear convergence of the ISTA has been established in [7, 17, 29] under some additional assumptions. In the case when $N < \infty$, Hale et al. [17] proved the linear convergence of the ISTA for the special case when $\omega_k \equiv \mu$ and under either of the following assumptions held at the limiting point u^* of the ISTA:

- J-BI: The operator K satisfies the J-BI (at x^*) with

$$J := \{k \in \mathcal{I}_N : |(K^*(Ku^* - h))_k| = \omega_k\}; \tag{1.7}$$

- SCC: The strict complementarity condition is satisfied at u^*, i.e., $\text{supp}(u^*) = J$.

In 2016, Tao et al. [29] developed a new approach, based on spectral analysis, to provide a linear convergence analysis of the ISTA under the assumptions that problem (1.1) has a unique solution and that the SCC is satisfied at this solution; this new approach can also be used to establish the linear convergence of the FISTA, which was proposed by Beck and Teboulle [4]. Considering the infinite-dimensional case (i.e., $l_N^2 = l^2$), Bredies and Lorenz [7] established the linear convergence of the ISTA under either of the following assumptions:

- FBI: The operator K has the finite basis injectivity property, i.e., $K|_I$ is injective for any finite subset $I \subseteq \mathbb{N}$;

- SSP: The strict sparsity pattern is satisfied at the limiting point u^* of the ISTA, i.e., it holds for any $k \in \mathbb{N}$ that

$$u_k^* = 0 \quad \Rightarrow \quad |(K^*(Ku^* - h))_k| < \omega_k.$$

Note that the FBI does not depend on the limiting point u^*, while the J-BI is crucial in the establishment of linear convergence of the ISTA; see [7, Remark 8] for details. In the special case when $N < \infty$, the FBI implies the J-BI, and that the SSP is equivalent to the SCC.

Inspired by the notions of FBI and SSP, we introduce a new notion of the orthogonal sparsity pattern (in short, OSP), which is weaker than either FBI or SSP; see Remark 1.2 for details. Let the solution set of problem (1.1) be denoted by S. For a closed linear subspace C of l_N^2, we use P_C to denote the metric projection onto C, that is,

$$P_C(u) := \arg\min_{v \in C} \|v - u\| \quad \text{for any } u \in l_N^2.$$

In particular, $(P_C(u))_i = u_i$ when $i \in I$ and $(P_C(u))_i = 0$ otherwise, where $C = E_I$ defined by (1.6).

Definition 1.1. Let $u^* \in S$, and let J be defined by (1.7). A bounded linear operator $K : l_N^2 \to H$ is said to have the OSP at u^*, if there exists an index set $I \subseteq \mathcal{I}_N$ with

$$\{k \in J : u_k^* = 0\} \subseteq I \subseteq J \tag{1.8}$$

such that K satisfies the I-BI (i.e., $K|_I$ is injective) and

$$\langle KP_{E_I}(u), KP_{E_{J \setminus I}}(u) \rangle = 0 \quad \text{for any } u \in l_N^2. \tag{1.9}$$

Remark 1.2. (i) The set J defined in (1.7) is a finite set. Indeed, it is trivial when $l_N^2 = \mathbb{R}^N$; otherwise, by (1.2) and (1.7), one has that

$$|J|\underline{\omega}^2 \le \sum_{k\in J} |\omega_k|^2 = \sum_{k\in J} |(K^*(Ku^* - h))_k|^2 \le \sum_{k\in \mathbb{N}} |(K^*(Ku^* - h))_k|^2 < \infty.$$

This shows that J is a finite set.

(ii) By (i) of this remark, the following implications/equivalence are true by definition.

$$\text{FBI} \quad \Rightarrow \quad J\text{-BI} \quad \Rightarrow \quad \text{OSP};$$

and

$$\text{SSP} \quad \Leftrightarrow \quad \text{SSC} \quad \Rightarrow \quad \text{OSP}.$$

The main result of this paper is presented in the following theorem, where the Q-linear convergence of the ISTA is ensured provided the OSP, a weaker assumption than the one assumed in [7,17].

Theorem 1.3. *Let $\{u^n\}$ be a sequence generated by Algorithm 1 satisfying* (1.4) *and* (1.5). *Then $\{u^n\}$ converges to a solution u^* of problem* (1.1). *Suppose that K possesses the OSP at u^*. Then $\{u^n\}$ linearly converges to u^*, that is, there exist $\lambda \in (0,1)$ and $M \in \mathbb{N}$ such that*

$$\| u^{n+1} - u^*\| \le \lambda \|u^n - u^*\| \quad \text{for any } n > M.$$

By Theorem 1.3 and Remark 1.2, we directly obtain the following corollary, which was proved in [17] for the case when $N \in \mathbb{N}$ and in [7] for the case when $N = \infty$.

Corollary 1.4. *Let $\{u^n\}$ be a sequence generated by Algorithm 1 satisfying* (1.4) *and* (1.5). *Then $\{u^n\}$ converges to a solution u^* of problem* (1.1). *Moreover, $\{u^n\}$ linearly converges to u^* provided either of the following assumptions:*

(a) *the J-BI is satisfied;*

(b) *the SSP is satisfied at u^*.*

The proof of Theorem 1.3 is presented in the next section. Examples are provided in section 3 to show the cases where our result in this paper is available but neither the one in [17] nor the one in [7].

2 Proof of Theorem 1.3

Let $I \subseteq \mathcal{I}_N$ and $C \subseteq l_N^2$. As usual, we use $C^\perp$ and I^c to denote the orthogonal complement of C and the complementary of I, respectively. As presented in the preceding section, let $K : l_N^2 \to H$ be a bounded linear operator. The kernel and image of K are respectively defined by

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

The restriction of K on C is denoted by $K|_C : C \to H$ and defined by

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

Note by (1.6) that $K|_I = K|_{E_I}$ for each index set $I \subseteq \mathcal{I}_N$.

To accomplish the proof of Theorem 1.3, we first present some basic properties of the projection operator in the following lemmas, in which Lemma 2.1(a) (resp. (b), (c), (d)) is taken from Theorem 3.14 (resp. Corollary 3.22(iii), (vi), Proposition 3.19) of [3], Lemma 2.2 is a direct consequence of [3, Fact 2.18] and Lemma 2.1(c).

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

(a) *$z = P_C(x)$ if and only if $z \in C$ and $x - z \in C^\perp$ for any $y \in C$;*

(b) *P_C is a linear and continuous operator with $\| P_C \| \leq 1$;*

(c) *$P_C^* = P_C$;*

(d) *P_C is idempotent, i.e., $P_C^2 = P_C$.*

Lemma 2.2. *Let $I \subseteq \mathcal{I}_N$. Then the following assertion holds:*

$$(\ker(KP_{E_I}))^\perp = \mathrm{im}(P_{E_I}K^*).$$

Lemma 2.3. *Let $x \in l_N^2$ and $y \in l_N^2$, and let $I_1 \subseteq \mathcal{I}_N$ and $I_2 \subseteq \mathcal{I}_N$ be such that $I_1 \cap I_2 = \emptyset$ and*

$$\langle KP_{E_{I_1}}(u), KP_{E_{I_2}}(u) \rangle = 0 \quad \textit{for any } u \in l_N^2. \tag{2.1}$$

Then the following assertion holds:

$$\langle KP_{E_{I_1}}(x), KP_{E_{I_2}}(y) \rangle = 0.$$

Proof. Let $u := P_{E_{I_1}}(x) + P_{E_{I_2}}(y) \in l_N^2$. Then it follows that

$$\langle KP_{E_{I_1}}(x), KP_{E_{I_2}}(y) \rangle = \langle KP_{E_{I_1}}(u), KP_{E_{I_2}}(u) \rangle = 0$$

(due to (2.1) and Lemma 2.1). The proof is complete. □

Associated to problem (1.1), one can directly check by using the optimality condition of convex optimization [26] that

$$u^* \in S \quad \Leftrightarrow \quad (-K^*(Ku^* - h))_k \begin{cases} = \omega_k, & \text{if } u_k^* > 0, \\ \in [-\omega_k, \omega_k], & \text{if } u_k^* = 0, \\ = -\omega_k, & \text{if } u_k^* < 0, \end{cases} \quad \text{for any } k \in \mathcal{I}_N. \tag{2.2}$$

For the remainder of this section, we always assume that

(A1) $\{u^n\}$ is generated by Algorithm 1 satisfying (1.4) and (1.5);

(A2) $u^* := \lim_{n\to\infty} u^n \in S$.

For simplicity, we further write

$$v^* := -K^*(Ku^* - h) \quad \text{and} \quad v^n := -K^*(Ku^n - h) \text{ for each } n \in \mathbb{N}. \tag{2.3}$$

Then we have that

$$v^* = \lim_{n\to\infty} v^n. \tag{2.4}$$

Next, we provide the following three lemmas for the iterative procedure of Algorithm 1. Recall that J is defined by (1.7).

Lemma 2.4. *Let $I \subseteq J$. Then there exists $M \in \mathbb{N}$ such that*

$$P_{E_{I^c}}(u^n - u^*) = P_{E_{J\setminus I}}(u^n - u^*) \quad \textit{for any } n > M. \tag{2.5}$$

Proof. By (1.7), (2.2) and (2.3), one has that

$$J^c = \{k \in \mathcal{I}_N : |v_k^*| < \omega_k\} \subseteq \{k \in \mathcal{I}_N : u_k^* = 0\}. \tag{2.6}$$

Set $v_k^* = 0$ and $\omega_k = \underline{\omega}$ for each $k > N$ in the case when $N < +\infty$. Then $v^* \in l^2$ for each $N \in \mathbb{N} \cup \{+\infty\}$. Hence it follows that $\lim_{k\to\infty} |v_k^*| = 0$, and so it follows from (1.2) that $\lim_{k\to\infty} \frac{|v_k^*|}{\omega_k} \leq \lim_{k\to\infty} \frac{|v_k^*|}{\underline{\omega}} = 0$. Fix $\tau_0 \in (0,1)$. Then there exists $M \in \mathbb{N}$ such that

$$\frac{|v_k^*|}{\omega_k} \leq \tau_0 \quad \text{for any } n \geq M.$$

Let $\tau := \max\left\{\tau_0, \max\left\{\frac{|v_k^*|}{\omega_k} : k \in J^c, k \leq M\right\}\right\}$. Hence we have that $\tau \in (0,1)$ and that

$$v_k^* \in [-\tau\omega_k, \tau\omega_k] \quad \text{for any } k \in J^c. \tag{2.7}$$

By assumption (A2) and (2.4), there exists $M \in \mathbb{N}$ such that

$$\|u^n - u^*\| \leq \frac{1-\tau}{2}\underline{s}\,\underline{\omega} \quad \text{and} \quad \|v^n - v^*\| \leq \frac{1-\tau}{2}\underline{\omega} \quad \text{for any } n \geq M. \tag{2.8}$$

Fix $i \in J^c$ and $n \geq M$. Note by (2.6) that $u_i^* = 0$ and by (1.2) that $\omega_i \geq \underline{\omega} > 0$. Then it follows from (2.8) and (1.4) that

$$|u_i^n| = |u_i^n - u_i^*| \leq \frac{1-\tau}{2}\underline{s}\,\underline{\omega} \leq \frac{1-\tau}{2}s_n\omega_i,$$

and from (2.7) and (2.8) that

$$|v_i^n| \leq |v_i^n - v_i^*| + |v_i^*| \leq \frac{1-\tau}{2}\underline{\omega} + \tau\omega_i \leq \frac{1-\tau}{2}\omega_i + \tau\omega_i = \frac{1+\tau}{2}\omega_i.$$

Combining the above two inequalities, we obtain that $|u_i^n + s_n v_i^n| \leq |u_i^n| + s_n|v_i^n| \leq s_n\omega_i$. Hence, in view of Algorithm 1 and by (2.3), one has that

$$u_i^{n+1} = \text{sign}(u_i^n + s_n v_i^n) \cdot (|u_i^n + s_n v_i^n| - s_n\omega_i)_+ = 0.$$

Since $i \in J^c$ is arbitrary, we have that

$$P_{E_{J^c}}(u^{n+1}) = 0. \tag{2.9}$$

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 it follows that

$$P_{E_{I^c}}(u^{n+1}) = P_{E_{J^c}}(u^{n+1}) + P_{E_{J\setminus I}}(u^{n+1}) = P_{E_{J\setminus I}}(u^{n+1}) \tag{2.10}$$

(due to (2.9)). By (2.6), we obtain that

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

This, together with (2.10) and Lemma 2.1(b), implies (2.5), and the proof is complete. □

Lemma 2.5. *Let $I \subseteq J$ be such that the I-BI and* (1.9) *are satisfied. Then there exist $\lambda \in (0,1)$ and $M \in \mathbb{N}$ such that*

$$\|P_{E_I}(u^{n+1} - u^*)\| \le \lambda \|P_{E_I}(u^n - u^*)\| \quad \text{for any } n > M. \tag{2.11}$$

Proof. By assumption, Lemma 2.4 is applicable to concluding that there exists $M \in \mathbb{N}$ such that (2.5) holds. One checks by definition (cf. (1.3)) that $\mathbf{S}_{s_n\omega}$ is nonexpansive, that is,

$$\|\mathbf{S}_{s_n\omega}(u) - \mathbf{S}_{s_n\omega}(v)\| \le \|u - v\| \quad \text{for any } u, v \in l_N^2. \tag{2.12}$$

Fix $n > M$. In view of Algorithm 1, by (2.12) and Lemma 2.1(b), one has that

$$\begin{aligned}
\|P_{E_I}(u^{n+1} - u^*)\| &= \|P_{E_I}(\mathbf{S}_{s_n\omega}(u^n - s_nK^*(Ku^n - h)) - \mathbf{S}_{s_n\omega}(u^* - s_nK^*(Ku^* - h)))\| \\
&= \|\mathbf{S}_{s_n\omega}P_{E_I}(u^n - s_nK^*(Ku^n - h)) - \mathbf{S}_{s_n\omega}P_{E_I}(u^* - s_nK^*(Ku^* - h))\| \\
&\le \|P_{E_I}(u^n - s_nK^*(Ku^n - h) - (u^* - s_nK^*(Ku^* - h)))\| \\
&= \|P_{E_I}(I - s_nK^*K)(u^n - u^*)\|.
\end{aligned}$$

Noting that $u = P_{E_{I^c}}(u) + P_{E_I}(u)$ for any $u \in l_N^2$, we obtain by Lemma 2.1 that $P_{E_I}P_{E_{I^c}} = 0$ and P_{E_I} is linear idempotent, and then it follows from above that

$$\begin{aligned}
&\|P_{E_I}(u^{n+1} - u^*)\| \\
&\le \|P_{E_I}(I - s_nK^*K)P_{E_I}(u^n - u^*) + P_{E_I}(I - s_nK^*K)P_{E_{I^c}}(u^n - u^*)\| \\
&= \|(P_{E_I} - s_nP_{E_I}K^*KP_{E_I})P_{E_I}(u^n - u^*) - s_nP_{E_I}K^*KP_{E_{I^c}}(u^n - u^*)\|.
\end{aligned} \tag{2.13}$$

Then we claim that

$$P_{E_I}K^*KP_{E_{I^c}}(u^n - u^*) = 0. \tag{2.14}$$

Indeed, by (2.5) and Lemma 2.1(c)-(d), we obtain that

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

By assumption that (1.9) is satisfied, Lemma 2.3 is applicable (with I, $J \setminus I$ in place of I_1, I_2); hence we proved (2.14).

Together with (2.14), (2.13) is reduced to

$$\begin{aligned}\|P_{E_I}(u^{n+1}-u^*)\| &\le \|(P_{E_I}-s_nP_{E_I}K^*KP_{E_I})P_{E_I}(u^n-u^*)\| \\ &\le \|P_{E_I}-s_nP_{E_I}K^*KP_{E_I}\|\|P_{E_I}(u^n-u^*)\|.\end{aligned} \tag{2.15}$$

We end this proof by estimating $\|P_{E_I}-s_nP_{E_I}K^*KP_{E_I}\|$. One has by definition that

$$\begin{aligned}&\|P_{E_I}-s_nP_{E_I}K^*KP_{E_I}\|^2 \\ &= \sup_{\|u\|=1}\langle (P_{E_I}-s_nP_{E_I}K^*KP_{E_I})(u), (P_{E_I}-s_nP_{E_I}K^*KP_{E_I})(u)\rangle \\ &= \sup_{\|u\|=1}\|P_{E_I}(u)\|^2 - 2s_n\langle P_{E_I}(u), P_{E_I}K^*KP_{E_I}(u)\rangle + s_n^2\|P_{E_I}K^*KP_{E_I}(u)\|^2.\end{aligned} \tag{2.16}$$

By Lemma 2.1, one has that

$$\langle P_{E_I}(u), P_{E_I}K^*KP_{E_I}(u)\rangle = \langle KP_{E_I}(u), KP_{E_I}(u)\rangle = \|KP_{E_I}(u)\|^2, \tag{2.17}$$

and that

$$\|P_{E_I}K^*KP_{E_I}(u)\|^2 \le \|KP_{E_I}\|^2\|KP_{E_I}(u)\|^2 \le \|K\|^2\|KP_{E_I}(u)\|^2. \tag{2.18}$$

Together with (2.17) and (2.18), (2.16) implies that

$$\|P_{E_I}-s_nP_{E_I}K^*KP_{E_I}\|^2 \le \sup_{\|u\|=1}\|P_{E_I}(u)\|^2 - 2s_n\left(1-\frac{s_n}{2}\|K\|^2\right)\|KP_{E_I}(u)\|^2. \tag{2.19}$$

Note by assumptions that E_I is finite-dimensional and that $K|_I$ is injective. There exists $\alpha \in (0, \|K\|^2)$ such that

$$\|KP_{E_I}(u)\| \ge \alpha\|P_{E_I}(u)\| \quad \text{for any } u \in l_N^2. \tag{2.20}$$

Also note by (1.4) that

$$s_n\left(1-\frac{s_n}{2}\|K\|^2\right) \ge \underline{s}\left(1-\frac{\overline{s}}{2}\|K\|^2\right) > 0. \tag{2.21}$$

Together with (2.20) and (2.21), (2.19) yields that

$$\|P_{E_I}-s_nP_{E_I}K^*KP_{E_I}\|^2 \le \sup_{\|u\|=1}\left(1-2\underline{s}\left(1-\frac{\overline{s}}{2}\|K\|^2\right)\alpha^2\right)\|P_{E_I}(u)\|^2. \tag{2.22}$$

Let $\lambda := \sqrt{1-2\underline{s}\left(1-\frac{\overline{s}}{2}\|K\|^2\right)\alpha^2} \in (0,1)$ (by (1.4)). Noting by Lemma 2.1(b) that $\|P_{E_I}(u)\| \le \|u\|$ we have by (2.22) that

$$\|P_{E_I}-s_nP_{E_I}K^*KP_{E_I}\| \le \lambda. \tag{2.23}$$

This, together with (2.15), implies (2.11), and the proof is complete. □

Lemma 2.6. *Let $I \subseteq \mathcal{I}_N$ be such that* (1.8) *and* (1.9) *are satisfied. Then there exist $\lambda \in (0,1)$ and $M \in \mathbb{N}$ such that*

$$\|P_{E_{I^c}}(u^{n+1}-u^*)\| \le \lambda\|P_{E_{I^c}}(u^n-u^*)\| \quad \textit{for any } n > M. \tag{2.24}$$

Proof. By assumption (A2) and (2.4), there exists $M \in \mathbb{N}$ such that

$$|u_i^n - u_i^*| \leq \frac{\tau}{2} \quad \text{and} \quad |v_i^n - v_i^*| \leq \frac{\tau}{2\overline{s}} \quad \text{for any } n > M. \tag{2.25}$$

Fix $n > M$. We first show that

$$P_{E_{J\setminus I}}(u^{n+1}) = P_{E_{J\setminus I}}(I - s_n K^*K)(u^n - u^*) + P_{E_{J\setminus I}}(u^*). \tag{2.26}$$

To this end, we define

$$T := \{k \in \mathcal{I}_N : u_k^* \neq 0\}. \tag{2.27}$$

It follows from (1.7) and (2.2) that $T \subseteq J$, which is a finite set (see Remark 1.2). Let $\tau := \min\{\underline{\omega}, \min\{|u_k^*| : k \in T\}\} > 0$ and fix $i \in T$. Then it follows that $|u_i^*| \geq \tau > 0$. Without loss of generality, we assume that

$$u_i^* \geq \tau > 0; \tag{2.28}$$

so we obtain by (2.2) that

$$v_i^* = \omega_i. \tag{2.29}$$

Note that

$$u_i^n + s_n v_i^n = u_i^* + u_i^n - u_i^* + s_n(v_i^n - v_i^*) + s_n v_i^* \geq u_i^* - |u_i^n - u_i^*| - s_n|v_i^n - v_i^*| + s_n v_i^*.$$

This, together (2.28), (2.25) and (2.29), yields that

$$u_i^n + s_n v_i^n \geq \tau - \frac{\tau}{2} - \frac{s_n \tau}{2\overline{s}} + s_n \omega_i \geq s_n \omega_i > 0$$

(due to (1.4)). This says that $\operatorname{sign}(u_i^n + s_n v_i^n) > 0$ and that $|u_i^n + s_n v_i^n| - s_n\omega_i = u_i^n + s_n v_i^n - s_n \omega_i \geq 0$. Therefore, in view of Algorithm 1, one has by (2.29) that

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

(due to (2.3)). Noting by (1.8) and (2.27) that $J \setminus I \subseteq T$ and recalling that $i \in T$ is arbitrary, we obtained (2.26).

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

$$U^\perp = \operatorname{im}(P_{E_{J\setminus I}} K^*). \tag{2.30}$$

Next, we show that

$$\|P_{I^c}(u^{n+1} - u^*)\| \leq \|P_{U^\perp} - s_n P_{U^\perp} K^*K P_{U^\perp}\| \|P_{E_{I^c}}(u^n - u^*)\|. \tag{2.31}$$

To show this, employing P_U on both sides of (2.26), we obtain by Lemma 2.1(b) that

$$\begin{aligned} P_U P_{E_{J\setminus I}}(u^{n+1}) &= P_U P_{E_{J\setminus I}}(u^n - u^*) - s_n P_U P_{E_{J\setminus I}} K^*K(u^n - u^*) + P_U P_{E_{J\setminus I}}(u^*) \\ &= P_U P_{E_{J\setminus I}}(u^n) - s_n P_U P_{E_{J\setminus I}} K^*K(u^n - u^*). \end{aligned} \tag{2.32}$$

Noting by (2.30) that $P_{E_{J\setminus I}}K^*K(u^n-u^*)\in U^\perp$, it is easy to see from Lemma 2.1(a) that $P_UP_{E_{J\setminus I}}K^*K(u^n-u^*)=0$. This, together with (2.32), implies that

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

Noting by assumption (A2) that $\lim_{n\to\infty}u^n=u^*$ and that $n>M$ is arbitrary, we obtain by Lemma 2.1(b) that

$$P_UP_{E_{J\setminus I}}(u^n)=P_UP_{E_{J\setminus I}}(u^*). \tag{2.33}$$

Then it follows from (2.33) and Lemma 2.1(b) that

$$P_{E_{J\setminus I}}(u^n-u^*)=P_UP_{E_{J\setminus I}}(u^n-u^*)+P_{U^\perp}P_{E_{J\setminus I}}(u^n-u^*)=P_{U^\perp}P_{E_{J\setminus I}}(u^n-u^*). \tag{2.34}$$

Employing $P_{U^\perp}$ on both side of (2.26), we have by Lemma 2.1(b) that

$$\begin{aligned}&P_{U^\perp}P_{E_{J\setminus I}}(u^{n+1}-u^*)\\&=P_{U^\perp}P_{E_{J\setminus I}}(I-s_nK^*K)(u^n-u^*)\\&=P_{U^\perp}P_{E_{J\setminus I}}(I-s_nK^*K)P_{E_{I^c}}(u^n-u^*)+P_{U^\perp}P_{E_{J\setminus I}}(I-s_nK^*K)P_{E_I}(u^n-u^*)\\&=P_{U^\perp}P_{E_{J\setminus I}}(I-s_nK^*K)P_{E_{I^c}}(u^n-u^*)-s_nP_{U^\perp}P_{E_{J\setminus I}}K^*KP_{E_I}(u^n-u^*).\end{aligned} \tag{2.35}$$

By assumption, Lemma 2.4 is applicable to ensuring (2.5). Then it follows from Lemma 2.1(b) and (d) that

$$\begin{aligned}&P_{U^\perp}P_{E_{J\setminus I}}(I-s_nK^*K)P_{E_{I^c}}(u^n-u^*)\\&=P_{U^\perp}P_{E_{J\setminus I}}(I-s_nK^*K)P_{E_{J\setminus I}}(u^n-u^*)\\&=(P_{U^\perp}-s_nP_{U^\perp}P_{E_{J\setminus I}}K^*KP_{U^\perp})P_{E_{J\setminus I}}(u^n-u^*)\end{aligned} \tag{2.36}$$

(due to (2.34)). By definition of $U^\perp$ (cf. (2.30)), 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 (2.36), implies that

$$P_{U^\perp}P_{E_{J\setminus I}}(I-s_nK^*K)P_{E_{I^c}}(u^n-u^*)=(P_{U^\perp}-s_nP_{U^\perp}K^*KP_{U^\perp})P_{E_{J\setminus I}}(u^n-u^*). \tag{2.37}$$

On the other hand, by (2.30) and Lemma 2.1(c)-(d), we have that

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

Note by (1.9) that Lemma 2.3 is applicable to concluding that

$$P_{U^\perp}P_{E_{J\setminus I}}K^*KP_{E_I}(u^n-u^*)=0. \tag{2.38}$$

Together with (2.37) and (2.38), (2.35) is reduced to

$$P_{U^\perp}P_{E_{J\setminus I}}(u^{n+1}-u^*)=(P_{U^\perp}-s_nP_{U^\perp}K^*KP_{U^\perp})P_{E_{J\setminus I}}(u^n-u^*). \tag{2.39}$$

Note by (2.5) and (2.34) that $P_{I^c}(u^n-u^*)=P_{E_{J\setminus I}}(u^n-u^*)=P_{U^\perp}P_{E_{J\setminus I}}(u^n-u^*)$. This, together with (2.39), yields (2.31).

Recall that J is finite (see Remark 1.2), and so $U^{\perp}$ is finite-dimensional (by (2.30)). We also claim that $K|_{U^{\perp}}$ is injective. Indeed, let $u \in U^{\perp}$ be such that

$$Ku = 0. \tag{2.40}$$

By (2.30), there exists $v \in H$ such that $u = P_{E_{J\setminus I}} K^* v$; hence (2.40) says that $KP_{E_{J\setminus I}} K^* v = 0$. Then it follows from Lemma 2.1(c)-(d) that

$$\|u\|^2 = \|P_{E_{J\setminus I}} K^* v\|^2 = \langle P_{E_{J\setminus I}} K^* v, P_{E_{J\setminus I}} K^* v\rangle = \langle KP_{E_{J\setminus I}} K^* v, v\rangle = 0.$$

Thus we proved that $K|_{U^{\perp}}$ is injective, as desired.

Thus, by the arguments as we did for (2.23) (and by (2.30)), we obtain that there exists $\lambda \in (0, 1)$ such that

$$\|P_{U^{\perp}} - s_n P_{U^{\perp}} K^* K P_{U^{\perp}}\| \leq \lambda.$$

This, together with (2.31), yields (2.24), and the proof is complete. □

Now we are ready to provide the proof of Theorem 1.3 as follows.

Proof of Theorem 1.3. As mentioned in the preceding section, $\{u^n\}$ strongly converges to $u^* \in S$. By assumptions of Theorem 1.3, the blanket assumptions in this section and the assumptions of Lemmas 2.5 and 2.6 are satisfied. Then the conclusion follows. □

3 Examples

In this section, we provide two examples to show the cases where our result in this paper is available but neither the one in [17] nor the one in [7]. The first example is illustrated in Euclidean space, and the second one is demonstrated in infinite-dimensional space.

Example 3.1. Consider problem (1.1) with

$$K = \begin{pmatrix} -1 & 1 & -1 & 2 \\ 1 & 1 & 1 & 2 \end{pmatrix}, \quad h = (-3, 5)^T \quad \text{and} \quad \omega = (4, 2, 4, 8)^T.$$

By (2.2), we have that $u^* \in S$ if and only if

$$\begin{aligned} 0 &\in 2u_1^* + 2u_3^* - 8 + 4\,\mathrm{Sign}(u_1^*), \\ 0 &\in 2u_2^* + 4u_4^* - 2 + 2\,\mathrm{Sign}(u_2^*), \\ 0 &\in 2u_1^* + 2u_3^* - 8 + 4\,\mathrm{Sign}(u_3^*), \\ 0 &\in 4u_2^* + 8u_4^* - 4 + 8\,\mathrm{Sign}(u_4^*), \end{aligned} \tag{3.1}$$

where

$$\mathrm{Sign}(t) := \begin{cases} \{1\}, & t > 0, \\ [-1, 1], & t = 0, \\ \{-1\}, & t < 0, \end{cases} \quad \text{for any } t \in \mathbb{R}.$$

Clearly, the second and fourth inclusions of (3.1) are equivalent to that $u_4^* = 0$ and $u_2^* = 0$; while the first and third inclusions of (3.1) are equivalent to that $u_1^* + u_3^* = 2$. Therefore, we have that

$$S = \left\{(a, 0, 2 - a, 0)^T : 0 \leq a \leq 2\right\}. \tag{3.2}$$

Write $x^* := (2, 0, 0, 0)^T$ and $y^* := (0, 0, 2, 0)^T$. Then we have that

(i) neither J-BI nor SSP is satisfied at any $u^* \in S$;

(ii) OSP is satisfied at each $u^* \in S \setminus \{x^*, y^*\}$.

Indeed, by (3.2), it is clear that

$$\mathrm{supp}(u^*) = \begin{cases} \{1\}, & \text{if } u^* = x^*, \\ \{3\}, & \text{if } u^* = y^*, \\ \{1,3\}, & \text{if } u^* \in S \setminus \{x^*, y^*\}. \end{cases} \tag{3.3}$$

For each $u^* \in S$, one checks that

$$|K^*(Ku^* - h)| = (4, 2, 4, 4)^T, \quad J = \{1, 2, 3\}, \quad K|_J = \begin{pmatrix} -1 & 1 & -1 \\ 1 & -1 & 1 \end{pmatrix}; \tag{3.4}$$

hence we observe from (3.3) and (3.4) that $K|_J$ is not injective and that $\mathrm{supp}(u^*) \neq J$. Thus assertion (i) is verified. Fix $u^* \in S \setminus \{x^*, y^*\}$. Note by (3.3) that $\mathrm{supp}(u^*) = \{1, 3\}$ and note by (3.4) that $J = \{1, 2, 3\}$. Let $I = \{2\}$. Then one checks by (3.4) that (1.8), (1.9) and the I-BI are satisfied; consequently, assertion (ii) is proved.

Let $u^0 = 0$ and $s_n \equiv s \in (0, \frac{1}{5})$ ($\|K\| = \sqrt{10}$), and let $\{u^n\}$ be a sequence generated by Algorithm 1. Below, we show that $\{u^n\}$ linearly converges to a solution of problem (1.1). In view of Algorithm 1, one has that

$$u^n = (1 - (1 - 4s)^n, 0, 1 - (1 - 4s)^n, 0)^T \quad \text{for any } n \in \mathbb{N}.$$

Since $s \in (0, \frac{1}{5})$, it follows that

$$\lim_{n \to \infty} u^n = (1, 0, 1, 0)^T \in S$$

(due to (3.2)). Noting by assertion (ii) that the OSP is satisfied at $(1, 0, 1, 0)^T$, we obtain by Theorem 1.3 the linear convergence of $\{u^n\}$ to $(1, 0, 1, 0)^T$. However, according to assertion (i), neither the result of linear convergence in [17] nor the one in [7] is available when solving this example.

Example 3.2. Consider problem (1.1) with $K : l^2 \to l^2$ being defined by

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

$h := (2, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \frac{1}{6}, \cdots)^T$ and $\omega := (1, 1, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \cdots)^T$. Then problem (1.1) is equivalent to

$$\min \left(\frac{1}{2}(u_1 + u_2 - 2)^2 + |u_1| + |u_2| \right) + \sum_{k=3}^{\infty} \min \left(\frac{1}{2}(u_k - \frac{1}{k})^2 + \frac{1}{5}|u_k| \right). \tag{3.5}$$

Let $u^* \in S$. Clearly, the first minimization of problem (3.5) is equivalent to that $u_1^* + u_2^* = 2$; while the others are equivalent to that $u_3^* = \frac{2}{15}$, $u_4^* = \frac{1}{20}$ and $u_k^* = 0$ for each $k \geq 5$. Therefore, we have that

$$S = \left\{ (a, 1 - a, \frac{2}{15}, \frac{1}{20}, 0, 0, \cdots)^T : 0 \leq a \leq 1 \right\}. \tag{3.6}$$

Write $x^* := (1, 0, \frac{2}{15}, \frac{1}{20}, 0, 0, \cdots)^T$ and $y^* := (0, 1, \frac{2}{15}, \frac{1}{20}, 0, 0, \cdots)^T$. Then we have assertions (i) and (ii) in Example 3.1. Indeed, by (3.6), it is clear that

$$\mathrm{supp}(u^*) = \begin{cases} \{1,3,4\}, & \text{if } u^* = x^*, \\ \{2,3,4\}, & \text{if } u^* = y^*, \\ \{1,2,3,4\}, & \text{if } u^* \in S \setminus \{x^*, y^*\}. \end{cases} \tag{3.7}$$

For each $u^* \in S$, one checks that

$$|K^*(Ku^* - h)| = (1, 1, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{6}, \frac{1}{7}, \cdots)^T,$$

and so

$$J = \{1,2,3,4,5\} \quad \text{and} \quad K|_J = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}; \tag{3.8}$$

hence we observe from (3.7) and (3.8) that $K|_J$ is not injective and that $\mathrm{supp}(u^*) \neq J$. That is, assertion (i) is verified. Fix $u^* \in S \setminus \{x^*, y^*\}$. Note by (3.7) that $\mathrm{supp}(u^*) = \{1,2,3,4\}$ and note by (3.8) that $J = \{1,2,3,4,5\}$. Let $I = \{5\}$. Then one checks by (3.8) that (1.8), (1.9) and the I-BI are satisfied; hence assertion (ii) is proved.

Let $u^0 = 0$ and $s_n \equiv s \in (0,1)$ ($\|K\| = \sqrt{2}$), and let $\{u^n\}$ be a sequence generated by Algorithm 1. Below, we show that $\{u^n\}$ linearly converges to a solution of problem (1.1). In view of Algorithm 1, one has that

$$u^n = \begin{pmatrix} \frac{1}{2} - \frac{1}{2}(1-2s)^n \\ \frac{1}{2} - \frac{1}{2}(1-2s)^n \\ \frac{2}{15} - \frac{2}{15}(1-s)^n \\ \frac{1}{20} - \frac{1}{20}(1-s)^n \\ 0 \\ 0 \\ \vdots \end{pmatrix} \quad \text{for any } n \in \mathbb{N}.$$

Since $s \in (0,1)$, it follows that

$$\lim_{n\to\infty} u^n = (\frac{1}{2}, \frac{1}{2}, \frac{2}{15}, \frac{1}{20}, 0, 0, \cdots)^T \in S$$

(due to (3.6)). Noting by assertion (ii) that the OSP is satisfied at $(\frac{1}{2}, \frac{1}{2}, \frac{2}{15}, \frac{1}{20}, 0, 0, \cdots)^T$, Theorem 1.3 is applicable to concluding that $\{u^n\}$ linearly converges to $(\frac{1}{2}, \frac{1}{2}, \frac{2}{15}, \frac{1}{20}, 0, 0, \cdots)^T$. However, according to assertion (i), the result of linear convergence in [7] is not available when solving this example.

References

[1] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, *Structured sparsity through convex optimization*, Stat. Sci., 27 (2012), pp. 450–468.

[2] E. G. Bazavan, F. Li, and C. Sminchisescu, *Fourier kernel learning*, in: A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, C. Schmid (Eds.), Proc. 12th Europ. Conf. Comput. Vision, 2012, pp. 459–473.

[3] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, New York, 2011.

[4] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., 2 (2009), pp. 183–202.

[5] J. Bect, L. Blanc-Fèraud, G. Aubert, and A. Chambolle, *A l^1 unified variational framework for image restoration*, in: T. Pajdla and J. Matas (Eds.), Proc. Eighth Europ. Conf. Comput. Vision, 2004, pp. 1–13.

[6] K. Bredies and D. A. Lorenz, *Iterated hard shrinkage for minimization problems with sparsity constraints*, SIAM J. Sci. Comput., 30 (2008), pp. 657–683.

[7] K. Bredies and D. A. Lorenz, *Linear convergence of iterative soft-thresholding*, J. Fourier Anal. Appl., 14 (2008), pp. 813–837.

[8] E. Candès and T. Tao, *Decoding by linear programming*, IEEE Trans. Inform. Theory, 51 (2005), pp. 4203–4215.

[9] P. L. Combettes and V. R. Wajs, *Signal recovery by proximal forward-backward splitting*, Multiscale Model. Sim., 4 (2005), pp. 1168–1200.

[10] I. Daubechies, M. Defrise, and C. D. Mol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Commun. Pur. Appl. Math., 57 (2004), pp. 1413–1457.

[11] D. L. Donoho, *Compressed sensing*, IEEE Trans. Inform. Theory, 52 (2006), pp. 1289–1306.

[12] D. L. Donoho, *Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition*, Appl. Comput. Harmon. Anal. 2 (1995), pp. 101–126.

[13] M. Elad, *Sparse and Redundant Representations*, Springer, New York, 2010.

[14] M. A. T. Figueiredo and R. D. Nowak, *An EM algorithm for wavelet-based image restoration*, IEEE Trans. Image Process., 12 (2003), pp. 906–916.

[15] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, *Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems*, IEEE J. Sel. Top. Signa., 1 (2007), pp. 586–597.

[16] R. Gribonval and M. Nielsen, *Highly sparse representations from dictionaries are unique and independent of the sparseness measure*, Appl. Comput. Harmon. Anal., 22 (2007), pp. 335–355.

[17] E. T. Hale, W. Yin, and Y. Zhang, *Fixed-point continuation for ℓ_1-minimization: Methodology and convergence*, SIAM J. Optim., 19 (2008), pp. 1107–1130.

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

[19] Y. Hu, C. Li, K. Meng, J. Qin, and X. Yang, *Group sparse optimization via $\ell_{p,q}$ regularization*, J. Mach. Learn. Res., 18 (2017), pp. 1–52.

[20] Y. Hu, C. Li, and X. Q. Yang, *On convergence rates of linearized proximal algorithms for convex composite optimization with applications*, SIAM J. Optim., 26 (2016), pp. 1207–1235.

[21] Z. Lu and L. Xiao, *On the complexity analysis of randomized block-coordinate descent methods*, Math. Program., 152 (2015), pp. 615–642.

[22] J. Mairal, *Incremental majorization-minimization optimization with application to large-scale machine learning*, SIAM J. Optim., 25 (2015), pp. 829–855.

[23] Y. Nesterov, *Efficiency of coordinate descent methods on huge-scale optimization problems*, SIAM J. Optim., 22 (2012), pp. 341–362.

[24] Y. Nesterov, *Gradient methods for minimizing composite functions*, Math. Program., 140 (2013), pp. 125–161.

[25] J. Qin, Y. H. Hu, F. Xu, H. K. Yalamanchili, and J. Wang, *Inferring gene regulatory networks by integrating ChIP-seq/chip and transcriptome data via LASSO-type regularization methods*, Methods, 67 (2014), pp. 294–303.

[26] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, 1970.

[27] N. Simon, J. Friedman, T. Hastie, and R. Tibshirani, *A sparse-group Lasso*, J. Comput. Graph. Stat., 22 (2013), pp. 231–245.

[28] J.-L. Starck, M. K. Nguyen, and F. Murtagh, *Wavelets and curvelets for image deconvolution: A combined approach*, Signal Process., 83 (2003), pp. 2279–2283.

[29] S. Tao, D. Boley, and S. Zhang, *Local linear convergence of ISTA and FISTA on the LASSO problem*, SIAM J. Optim., 26 (2016), pp. 313–336.

[30] J. Wang, Y. Hu, C. Li, and J.-C. Yao, *Linear convergence of CQ algorithms and applications in gene regulatory network inference*, Inverse Probl., to appear, 2017.

[31] L. Xiao and T. Zhang, *A proximal-gradient homotopy method for the sparse least-squares problem*, SIAM J. Optim., 23 (2013), pp. 1062–1091.

[32] J. Yang and Y. Zhang, *Alternating direction algorithms for ℓ_1-problems in compressive sensing*, SIAM J. Sci. Comput., 33 (2011), pp. 250–278.