$L_{2/3}$ regularization: convergence of iterative thresholding algorithm

Yong Zhang, Wanzhou Ye*

Department of Mathematics, College of Science, Shanghai University, Shanghai, 200444, China

Abstract

The $L_{2/3}$ regularization is a nonconvex and nonsmooth optimization problem. Cao et al. (Cao et al. (2013) [15]) investigated that the $L_{2/3}$ regularization is more effective in imaging deconvolution. The convergence issue of the iterative thresholding algorithm of $L_{2/3}$ regularization problem (the $L_{2/3}$ algorithm) hasn't been addressed in (Cao et al. (2013) [15]). In this paper, we study the convergence of the $L_{2/3}$ algorithm. As the main result, we show that under certain conditions, the sequence $\{x^{(n)}\}$ generated by the $L_{2/3}$ algorithm converges to a local minimizer of $L_{2/3}$ regularization, and its asymptotical convergence rate is linear. We provide a set of experiments to verify our theoretical assertions and show the performance of the algorithm on sparse signal recovery. The established results provide a theoretical guarantee for a wide range of applications of the algorithm.

Keywords: $L_{1/2}$ regularization, $L_{2/3}$ regularization, Iterative thresholding algorithm, Thresholding formula, Convergence

1. Introduction

The L_q ($0 < q < 1$) regularization [1, 2, 3, 4] can be modeled as following

$$\min_{x \in \mathbb{R}^N} \{\|y - Ax\|_2^2 + \lambda \|x\|_q^q\}, \quad (1)$$

where $\|\cdot\|_2$ denotes the Euclidean norm, $A \in \mathbb{R}^{m \times N}$ (commonly $m < N$), $y \in \mathbb{R}^m$ and $\|x\|_q$ represents the L_q quasi-norm, defined by $\|x\|_q = (\sum_{i=1}^N |x_i|^q)^{1/q}$, and $\lambda > 0$ is a regularization parameter. This regularization formulation has been studied extensively in variable selection [5, 6, 7, 8], sparse signal recovery [9, 10, 11], low-rank matrix completion [12, 13] and imaging processing [14, 15].

The L_q ($0 < q < 1$) regularization is nonconvex and nonsmooth. In general, such problems are difficult to solve fast and efficiently. However, in particular, Xu et al [4] developed a thresholding representation theory for $L_{1/2}$ regularization and proposed an iterative half thresholding algorithm. They analyzed that among L_q regularization with $q \in (0, 1)$, only $L_{1/2}$ and $L_{2/3}$ regularizations permit an analytically thresholding representation, that is, the L_q ($q = 1/2, 2/3$) regularization has

closed-form thresholding formulas. In [15], Cao et al deduced the closed-form thresholding formula for the $L_{2/3}$ regularization. Based on the closed-form thresholding formulas for L_q ($q = 1/2, 2/3$) regularization, extensive experiments indicate that $L_{2/3}$ regularization is more effective than L_0 [16], $L_{1/2}$ [4] and L_1 [17, 18] regularization in image deconvolution.

Although the L_q regularization has wide applications in practice [19, 20], the convergence issue of the L_q regularization is very difficult to tackle. The convergence issue, however, is very crucial as it characterizes the circumstance in which an algorithm can or cannot be successfully applied. Fortunately, the convergence of the half algorithm for $L_{1/2}$ regularization has been partially analyzed in [4] and studied in [21]. Besides, in [22], relying on the restricted isometry property (RIP), Liu et al proved the subsequential convergence of the $L_{2/3}$ algorithm for $L_{2/3}$ regularization and further provided an error bound for limit point of any convergent subsequence. Based on the importance of the convergence issue of L_q regularization, we therefore investigate the convergence of the $L_{2/3}$ algorithm for $L_{2/3}$ regularization. We mainly focus on the following questions:

(i) under what conditions the $L_{2/3}$ algorithm converges to a local minimizer of $L_{2/3}$ regularization;

(ii) how fast the algorithm converges eventually. We provide answers to these questions in this paper.

*Corresponding author.

Email addresses: 13820161@shu.edu.cn (Yong Zhang), wzhy@shu.edu.cn (Wanzhou Ye)

Preprint submitted to Elsevier

October 10, 2015

As the main contribution of the present research, we present a direct and simple proof of the convergence of the $L_{2/3}$ algorithm for $L_{2/3}$ regularization. More specifically, we first investigate the detailed properties of the limit point set of a sequence $\{x^{(n)}\}$ generated by the $L_{2/3}$ algorithm, and then based on these results we study convergence of the $L_{2/3}$ algorithm. We show that under certain conditions, the $L_{2/3}$ algorithm converges to a local minimizer of $L_{2/3}$ regularization. Under the same conditions, we further verify that the asymptotical convergence rate of the $L_{2/3}$ algorithm is linear.

The rest of the paper is organized as follows. In Section 2, we will introduce the $L_{2/3}$ thresholding representation theory and the $L_{2/3}$ algorithm. In Section 3, we prove the convergence results for $L_{2/3}$ regularization. In Section 4, we present an asymptotically linear convergence rate of the $L_{2/3}$ algorithm. Numerical experiments are presented in Section 5 to verify our theoretical assertions for the convergence of $L_{2/3}$ algorithm. Finally, we summarize this paper in Section 6.

Before continuing, we present some notations that will be used in this paper. For any $a \in R$, $\text{sign}(a)$ be the sign function of a . For any vectors $x, z \in R^N$, $\text{supp}(x)$ denotes the support set of x , i.e., $\text{supp}(x) = \{i : x_i \neq 0, i = 1, 2, \dots, N\}$, and $\text{sign}(x) = \text{sign}(z)$ denotes that the two vectors have the same sign of corresponding components. For any index set S , we denote x_S as the subvector of x with components restricted to S , A_S as the submatrix of A with the columns restricted to S and $\sigma_{\min}(A_S^T A_S)$ as the minimal singular value of $A_S^T A_S$. Other notations are explained as they appear in the paper.

2. Iterative thresholding algorithm

In this section, we review some known results from [4, 15, 22] for thresholding representation theory of $L_{2/3}$ regularization and the corresponding $L_{2/3}$ algorithm, which will be used in the next sections.

2.1. Thresholding representation theory for $L_{2/3}$ regularization

First, we introduce the thresholding function of $L_{2/3}$ regularization, which is used to minimize a real-valued function.

Lemma 1. Let $t \in R$, $\lambda > 0$ be two given real numbers. Suppose that $x^* \in R$ is a solution of the problem

$$\min_{x \in R} f(x) = x^2 - 2tx + \lambda|x|^{2/3}. \quad (2)$$

Then x^* can be analytically expressed as

$$x^* = h_{\lambda,2/3}(t) = \begin{cases} f_{\lambda,2/3}(t), & |t| > \frac{\sqrt[4]{48}}{3} \lambda^{3/4} \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

where $f_{\lambda,2/3}(t)$ is defined by

$$f_{\lambda,2/3}(t) = \left(\frac{|\Phi_\lambda(t)| + \sqrt{\frac{2}{|\Phi_\lambda(t)|} - |\Phi_\lambda(t)|^2}}{2} \right)^3 \text{sign}(t), \quad (4)$$

with $\Phi_\lambda(t) = \frac{2}{\sqrt{3}} \lambda^{1/4} \left(\cosh\left(\frac{1}{3} \text{arccosh}\left(\frac{27}{16} \lambda^{-3/2} t^2\right)\right) \right)^{1/2}$.

Now, we consider the $L_{2/3}$ regularization problem:

$$\min_{x \in R^N} \left\{ C_\lambda(x) = \|y - Ax\|_2^2 + \lambda \|x\|_{2/3}^{2/3} \right\}, \quad (5)$$

where $\lambda > 0$ is a regularization parameter. For any $\lambda, \mu > 0$ and $z \in R^N$, let

$$C_{\lambda,\mu}(x, z) = \mu \left(C_\lambda(x) - \|Ax - Az\|_2^2 \right) + \|x - z\|_2^2, \quad (6)$$

$$B_\mu(z) = z - \mu A^T (Az - y).$$

where A^T is the transpose of A .

When $0 < \mu < \|A\|_2^{-2}$, we have

$$\|x - z\|_2^2 - \mu \|Ax - Az\|_2^2 \geq 0.$$

Therefore, we can get

$$\begin{aligned} C_{\lambda,\mu}(x, z) &= \mu C_\lambda(x) + (\|x - z\|_2^2 - \mu \|Ax - Az\|_2^2) \\ &\geq \mu C_\lambda(x). \end{aligned} \quad (7)$$

By Eq.(7), we would expect minimizing the $L_{2/3}$ regularization problem (5) by minimizing $C_{\lambda,\mu}(x, z)$.

Under the condition $0 < \mu < \|A\|_2^{-2}$, if $x^* \in R^N$ is a minimizer of $L_{2/3}$ regularization problem (5), then

$$\begin{aligned} C_{\lambda,\mu}(x, x^*) &= \mu C_\lambda(x) + (\|x - x^*\|_2^2 - \mu \|Ax - Ax^*\|_2^2) \\ &\geq \mu C_\lambda(x) \\ &\geq \mu C_\lambda(x^*) \\ &= C_{\lambda,\mu}(x^*, x^*). \end{aligned}$$

This implies that x^* is also a minimizer of $C_{\lambda,\mu}(x, x^*)$.

On the other hand, we note that $C_{\lambda,\mu}(x, z)$ can be re-expressed as

$$\begin{aligned} C_{\lambda,\mu}(x, z) &= \mu \left(C_\lambda(x) - \|Ax - Az\|_2^2 \right) + \|x - z\|_2^2 \\ &= \|x\|_2^2 - 2\langle x, z - \mu A^T (Az - y) \rangle + \lambda \mu \|x\|_{2/3}^{2/3} + \|z\|_2^2 \\ &\quad + \mu \|y\|_2^2 - \mu \|Az\|_2^2 \\ &= \|x\|_2^2 - 2\langle x, B_\mu(z) \rangle + \lambda \mu \|x\|_{2/3}^{2/3} + \|z\|_2^2 + \mu \|y\|_2^2 - \mu \|Az\|_2^2 \\ &= \|x - B_\mu(z)\|_2^2 + \lambda \mu \|x\|_{2/3}^{2/3} + \|z\|_2^2 + \mu \|y\|_2^2 \\ &\quad - \mu \|Az\|_2^2 - \|B_\mu(z)\|_2^2, \end{aligned}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product. This implies that for any fixed λ, μ and z , minimizing $C_{\lambda\mu}(x, z)$ is equivalent to solving

$$\min_{x \in R^N} \left\{ \|x - B_\mu(z)\|_2^2 + \lambda\mu \|x\|_{2/3}^{2/3} \right\},$$

i.e.,

$$\min_{x \in R^N} \left\{ \sum_{i=1}^N (|x_i - [B_\mu(z)]_i|^2 + \lambda\mu |x_i|^{2/3}) \right\}, \quad (8)$$

where $[B_\mu(z)]_i$ represents the i -th component of vector $B_\mu(z)$. Note that the summation of Eq.(8) is separable, hence, solving Eq.(8) is equivalent to solving the following N subproblem, for $i \in \{1, 2, \dots, N\}$,

$$\min_{x_i \in R} \left\{ (x_i - [B_\mu(z)]_i)^2 + \lambda\mu |x_i|^{2/3} \right\}. \quad (9)$$

By Lemma 1, the minimizer x_i^s of Eq.(9) is given by

$$x_i^s = \begin{cases} h_{\lambda\mu, 2/3}([B_\mu(z)]_i) \\ f_{\lambda\mu, 2/3}([B_\mu(z)]_i), & |[B_\mu(z)]_i| > \frac{\sqrt[4]{48}}{3}(\lambda\mu)^{3/4} \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, we get the minimizer $x^s \in R^N$ of $C_{\lambda\mu}(x, z)$ as follows

$$x^s = H_{\lambda\mu, 2/3}(B_\mu(z)),$$

where $H_{\lambda\mu, 2/3}$ is the thresholding operator defined as

$$H_{\lambda\mu, 2/3}(B_\mu(z)) = (h_{\lambda\mu, 2/3}([B_\mu(z)]_1), \dots, h_{\lambda\mu, 2/3}([B_\mu(z)]_N))^T. \quad (10)$$

It was analyzed in [4] that $L_{2/3}$ regularization permits a thresholding representation theory for its solution, and in detail its closed-form thresholding formula was deduced in [15, 22]. Through previous analysis, we thus get the following theorem.

Theorem 1. For any fixed $\mu \in (0, \|A\|_2^{-2})$, any minimizer $x^* \in R^N$ of $L_{2/3}$ regularization satisfies

$$x^* = H_{\lambda\mu, 2/3}(B_\mu(x^*)). \quad (11)$$

The following lemma summarizes the property of thresholding function of $L_{2/3}$ regularization.

Lemma 2. For any $u \in R$, $f_{\lambda\mu, 2/3}(u)$ has the following properties:

- (1) $f_{\lambda\mu, 2/3}(u)$ is an odd function.
- (2) For any $u > \frac{\sqrt[4]{48}}{3}(\lambda\mu)^{3/4}$, $f_{\lambda\mu, 2/3}(u)$ is strictly increasing and $f_{\lambda\mu, 2/3}(u) > (\frac{\lambda\mu}{3})^{3/4}$.

(3) For any $|u| > \frac{\sqrt[4]{48}}{3}(\lambda\mu)^{3/4}$, $f_{\lambda\mu, 2/3}(u)$ solves the equation

$$f_{\lambda\mu, 2/3}(u) + \frac{\lambda\mu}{3\sqrt[3]{f_{\lambda\mu, 2/3}(u)}} = u.$$

The Lemma 2 (1) and (2) can be easily derived from Lemma 1 and Lemma 2 (3) can be obtained according to the results in [4].

2.2. The $L_{2/3}$ algorithm

Since Xu et al [4] analyzed that $L_{2/3}$ regularization permits a thresholding representation in the form of Eq.(11) and Cao et al [15] deduced the closed-form thresholding formula for $L_{2/3}$ regularization, therefore, an iterative thresholding algorithm, called the $L_{2/3}$ algorithm, for $L_{2/3}$ regularization was naturally defined as

$$x^{(n+1)} = H_{\lambda\mu, 2/3}(x^{(n)} - \mu A^T(Ax^{(n)} - y)), \quad (12)$$

where μ is a step size parameter.

The convergence issue of the $L_{2/3}$ algorithm is very important in guaranteeing that the algorithm can be successfully applied. In this paper, we provide a direct and simple proof of convergence of the $L_{2/3}$ algorithm.

In the following, we first introduce the following theorem 2, which will be used later in this paper.

Theorem 2. Let $\{x^{(n)}\}$ be the sequence generated by the $L_{2/3}$ algorithm, and $\mu \in (0, \|A\|_2^{-2})$, then

- (1) $\{x^{(n)}\}$ is a minimization sequence, and $C_\lambda(x^{(n)})$ converges to $C_\lambda(x^*)$, where x^* is a limit point of minimization sequence $\{x^{(n)}\}$;
- (2) $\{x^{(n)}\}$ is asymptotically regular, i.e.,

$$\lim_{n \rightarrow \infty} \|x^{(n+1)} - x^{(n)}\|_2 = 0.$$

For the proof of Theorem 2, refer to the proof of Theorem 3 in [4].

Remark 1. From Theorem 2 (1), we obtain that the sequence $\{x^{(n)}\}$ is bounded. If we denote the set of all limit points of $\{x^{(n)}\}$ by $Y_{\{x^{(n)}\}}$, then from Theorem 2 (1), we know that $C_\lambda(x^*) = \text{constant}$ for $x^* \in Y_{\{x^{(n)}\}}$, i.e., all the limit points are on an equal-value plane of the real function $f = C_\lambda(x)$ of N variables.

Next, we study the detailed properties of $Y_{\{x^{(n)}\}}$, which is a basic and key result in this paper.

Lemma 3. Given that $\lambda > 0$ and $\mu \in (0, \|A\|_2^{-2})$. Assume $\{x^{(n)}\}$ to be the sequence generated by the $L_{2/3}$ algorithm, then there exists an n_0 such that whenever $n > n_0$,

- (1) $\text{supp}(x^{(n)}) = \text{supp}(x^{(n_0)})$, $\text{sign}(x^{(n)}) = \text{sign}(x^{(n_0)})$, and for each $i \in \{1, 2, \dots, N\}$,

$$x_i^{(n)} \in (-\infty, -(\frac{\lambda\mu}{3})^{3/4}) \cup \{0\} \cup ((\frac{\lambda\mu}{3})^{3/4}, +\infty);$$

(2) for any $x^* \in Y_{\{x^{(n)}\}}$, there holds $\text{supp}(x^*) = \text{supp}(x^{(n_0)})$, $\text{sign}(x^*) = \text{sign}(x^{(n_0)})$, and for each $i \in \{1, 2, \dots, N\}$,

$$x_i^* \in (-\infty, -(\frac{\lambda\mu}{3})^{3/4}) \cup \{0\} \cup [(\frac{\lambda\mu}{3})^{3/4}, +\infty);$$

(3) for each $x^* \in Y_{\{x^{(n)}\}}$ and $i \in I = \text{supp}(x^*)$, we have

$$x_i^* + \frac{\lambda\mu}{3\sqrt[3]{x_i^*}} = [B_\mu(x^*)]_i,$$

or equivalently

$$A_i^T(Ax^* - y) = -\frac{\lambda}{3\sqrt[3]{x_i^*}};$$

where $A_i \in R^m$ denotes the i -th column of A and A_i^T denotes the transpose of A_i .

(4) for any $x^* \in Y_{\{x^{(n)}\}}$ and $i \in I^c$, we have

$$|A_i^T(Ax^* - y)| \leq \frac{\sqrt[4]{48}}{3} \lambda^{3/4} \mu^{-1/4}.$$

where I^c represents the complementary set of I .

Proof. By Theorem 2 (2), there exists an n_0 such that whenever $n \geq n_0$, there holds $\|x^{(n+1)} - x^{(n)}\|_2 < (\frac{\lambda\mu}{3})^{3/4}$, and therefore we have

$$\max_{1 \leq i \leq N} |x_i^{(n+1)} - x_i^{(n)}| < (\frac{\lambda\mu}{3})^{3/4}. \quad (13)$$

By the definition of the $L_{2/3}$ algorithm, for all $n \geq 1$ and $i \in \{1, 2, \dots, N\}$ we have

$$x_i^{(n)} \in (-\infty, -(\frac{\lambda\mu}{3})^{3/4}) \cup \{0\} \cup [(\frac{\lambda\mu}{3})^{3/4}, +\infty). \quad (14)$$

(1) For each $i \in \{1, 2, \dots, N\}$ and $n > n_0$,

if $x_i^{(n_0)} = 0$, then by Eq.(13) and Eq.(14) it follows that $x_i^{(n_0+1)} = 0$, and by induction we have $x_i^{(n)} = 0$;

if $x_i^{(n_0)} > (\frac{\lambda\mu}{3})^{3/4}$, then by Eq.(13) and Eq.(14) it follows that $x_i^{(n_0+1)} > (\frac{\lambda\mu}{3})^{3/4}$, and by induction we have $x_i^{(n)} > (\frac{\lambda\mu}{3})^{3/4}$;

if $x_i^{(n_0)} < -(\frac{\lambda\mu}{3})^{3/4}$, then, by a similar argument as the above, we have that $x_i^{(n)} < -(\frac{\lambda\mu}{3})^{3/4}$.

These imply that (1) holds.

(2) For any $x^* \in Y_{\{x^{(n)}\}}$, there exists a subsequence $\{x^{(n_p)}\}$ such that $x^{(n_p)} \rightarrow x^*$ as $p \rightarrow \infty$. By (1), for $n_p \geq n_0$, we have $\text{supp}(x^{(n_p)}) = \text{supp}(x^{(n_0)})$ and $\text{sign}(x^{(n_p)}) = \text{sign}(x^{(n_0)})$. Therefore,

if $i \in \{1, 2, \dots, N\} - \text{supp}(x^{(n_0)})$, then we have that, for all $n_p \geq n_0$, $x_i^{(n_p)} = 0$, so $x_i^* = \lim_{p \rightarrow \infty} x_i^{(n_p)} = 0$;

if $i \in \text{supp}(x^{(n_0)})$ and $x_i^{(n_0)} > (\frac{\lambda\mu}{3})^{3/4}$, then, for all $n_p \geq n_0$, $x_i^{(n_p)} > (\frac{\lambda\mu}{3})^{3/4}$, so $x_i^* = \lim_{p \rightarrow \infty} x_i^{(n_p)} \geq (\frac{\lambda\mu}{3})^{3/4}$;

if $i \in \text{supp}(x^{(n_0)})$ and $x_i^{(n_0)} < -(\frac{\lambda\mu}{3})^{3/4}$, then, for all $n_p \geq n_0$, $x_i^{(n_p)} < -(\frac{\lambda\mu}{3})^{3/4}$, so $x_i^* = \lim_{p \rightarrow \infty} x_i^{(n_p)} \leq -(\frac{\lambda\mu}{3})^{3/4}$.

These imply that (2) holds.

(3) By (1) and (2), for any $n \geq n_0$ and $x^* \in Y_{\{x^{(n)}\}}$, we have $\text{supp}(x^{(n)}) = \text{supp}(x^*)$ and $\text{sign}(x^{(n)}) = \text{sign}(x^*)$. For any $x^* \in Y_{\{x^{(n)}\}}$ and $i \in I = \text{supp}(x^*)$, there exists a subsequence $\{x^{(n_p)}\}$ such that $x^{(n_p)} \rightarrow x^*$ as $p \rightarrow \infty$.

If $x_i^* \geq (\frac{\lambda\mu}{3})^{3/4}$, then $x_i^{(n_p)} > (\frac{\lambda\mu}{3})^{3/4}$ and $x_i^{(n_p+1)} > (\frac{\lambda\mu}{3})^{3/4}$ for all $n_p \geq n_0$. So by Lemma 2 (3), for all n_p , we have

$$x_i^{(n_p+1)} + \frac{\lambda\mu}{3\sqrt[3]{x_i^{(n_p+1)}}} = [B_\mu(x^{(n_p)})]_i.$$

By letting $p \rightarrow \infty$ on both sides of the above equation, we get

$$x_i^* + \frac{\lambda\mu}{3\sqrt[3]{x_i^*}} = [B_\mu(x^*)]_i. \quad (15)$$

If $x_i^* \leq -(\frac{\lambda\mu}{3})^{3/4}$, then by a similar argument as the above, we also obtain

$$x_i^* + \frac{\lambda\mu}{3\sqrt[3]{x_i^*}} = [B_\mu(x^*)]_i.$$

Thus, for any $i \in I = \text{supp}(x^*)$, x_i^* satisfies Eq.(15). Substituting $[B_\mu(x^*)]_i = x_i^* - \mu A_i^T(Ax^* - y)$ into Eq.(15), for any $i \in I = \text{supp}(x^*)$, we get

$$A_i^T(Ax^* - y) = -\frac{\lambda}{3\sqrt[3]{x_i^*}}.$$

(4) For any $x^* \in Y_{\{x^{(n)}\}}$, there exists a subsequence $\{x^{(n_p)}\}$ such that $x^{(n_p)} \rightarrow x^*$ as $p \rightarrow \infty$. By (1) and (2), for $n \geq n_0$, we have that $\text{supp}(x^{(n)}) = \text{supp}(x^{(n_0)}) = \text{supp}(x^*) = I$. Therefore, when $i \in I^c$, $x_i^* = 0$, and for all $n_p \geq n_0$, $x_i^{(n_p+1)} = 0$. By the definition of $x_i^{(n_p+1)}$, when $x_i^{(n_p+1)} = 0$ it implies that $|[B_\mu(x^{(n_p)})]_i| \leq \frac{\sqrt[4]{48}}{3} (\lambda\mu)^{3/4}$, i.e., $|x_i^{(n_p)} - \mu A_i^T(Ax^{(n_p)} - y)| \leq \frac{\sqrt[4]{48}}{3} (\lambda\mu)^{3/4}$, by taking limits

$$\lim_{p \rightarrow \infty} |x_i^{(n_p)} - \mu A_i^T(Ax^{(n_p)} - y)| = |x_i^* - \mu A_i^T(Ax^* - y)|,$$

hence we have

$$|A_i^T(Ax^* - y)| \leq \frac{\sqrt[4]{48}}{3} \lambda^{3/4} \mu^{-1/4}.$$

This completes the proof of Lemma 3.

From Lemma 3 it can be seen that, under the conditions $\lambda > 0$ and $\mu \in (0, \|A\|_2^{-2})$, any sequence $\{x^{(n)}\}$ generated by the $L_{2/3}$ algorithm satisfies that all its limit points share a common support set, denoted by $I = I(Y_{\{x^{(n)}\}})$ in the succeeding sections.

3. Convergence of the $L_{2/3}$ algorithm

As it is concerned with the properties of set $Y_{\{x^{(n)}\}}$ in Lemma 3, we may be more interested in investigating under what conditions the set $Y_{\{x^{(n)}\}}$ contains a unique element, i.e., the sequence $\{x^{(n)}\}$ is convergent, and whether its limit is a global minimizer of Eq.(5) or more possibly, a local minimizer of Eq.(5) due to the nonconvexity of the $L_{2/3}$ regularization. To provide answers to these questions, we present our main result in the following.

Theorem 3. Given that $\lambda > 0$ and $\mu \in (0, \|A\|_2^{-2})$. Assume that the sequence $\{x^{(n)}\}$ is generated by the $L_{2/3}$ algorithm, and let $I = I(Y_{\{x^{(n)}\}})$. If

$$\mu > \frac{1}{3\sigma_{\min}(A_I^T A_I)}, \quad (16)$$

then

- (1) $\{x^{(n)}\}$ is convergent;
- (2) its limit is a local minimizer of $L_{2/3}$ regularization.

Proof. (1) Since the sequence $\{x^{(n)}\}$ is bounded (see Remark 1), in order to prove that $\{x^{(n)}\}$ is convergent, we only prove that the set $Y_{\{x^{(n)}\}}$ contains a unique element, that is, for any $x^*, x^{*'} \in Y_{\{x^{(n)}\}}$, we have $x^* = x^{*'}$.

By Lemma 3 (2) and (3), for any $i \in I$, we have $\text{supp}(x^*) = \text{supp}(x^{*'})$, $\text{sign}(x^*) = \text{sign}(x^{*'})$. Hence,

$$x_i^{*'} + \frac{\lambda\mu}{3\sqrt[3]{x_i^{*'}}} = [B_\mu(x^*)]_i, \quad (17)$$

$$x_i^{*'} + \frac{\lambda\mu}{3\sqrt[3]{x_i^{*'}}} = [B_\mu(x^{*'})]_i. \quad (18)$$

Since $[B_\mu(z)]_i = z_i - \mu A_i^T (Az - y)$, we have

$$\frac{\lambda\mu}{3} \left(\frac{1}{\sqrt[3]{x_i^*}} - \frac{1}{\sqrt[3]{x_i^{*'}}} \right) = -\mu A_i^T A (x^* - x^{*'}). \quad (19)$$

By Lagrange mean value theorem, we can conclude that for any $i \in I$, there exists a $x_i^{\xi_i}$, which lies between x_i^* and $x_i^{*'}$, such that

$$\frac{1}{\sqrt[3]{x_i^*}} - \frac{1}{\sqrt[3]{x_i^{*'}}} = -\frac{1}{3(x_i^{\xi_i})^{4/3}} (x_i^* - x_i^{*'}).$$

Substituting the above equation into Eq.(19), we get

$$\mu A_i^T A (x^* - x^{*'}) = \frac{\lambda\mu}{9(x_i^{\xi_i})^{4/3}} (x_i^* - x_i^{*'}). \quad (20)$$

Denoting the diagonal matrix with $\Lambda(i, i) = \frac{\lambda\mu}{9(x_i^{\xi_i})^{4/3}}$ by Λ , we can write

$$\mu A_I^T A_I (x_I^* - x_I^{*'}) = \Lambda (x_I^* - x_I^{*'}). \quad (21)$$

Since $x_i^{\xi_i}$ is between x_i^* and $x_i^{*'}$, by Lemma 3 (2), for any $i \in I$, $|x_i^{\xi_i}| \geq (\frac{\lambda\mu}{3})^{3/4}$. So for any $i \in I$, $|\Lambda(i, i)| \leq \frac{1}{3}$. Furthermore, by Eq.(21) we have

$$\mu \|A_I^T A_I (x_I^* - x_I^{*'})\|_2 = \|\Lambda (x_I^* - x_I^{*'})\|_2. \quad (22)$$

From Eq.(22), on one hand, we have

$$\|\Lambda (x_I^* - x_I^{*'})\|_2 \leq \|\Lambda\|_2 \|x_I^* - x_I^{*'}\|_2 \leq \frac{1}{3} \|x_I^* - x_I^{*'}\|_2. \quad (23)$$

On the other hand, we have

$$\mu \|A_I^T A_I (x_I^* - x_I^{*'})\|_2 \geq \mu \sigma_{\min}(A_I^T A_I) \|x_I^* - x_I^{*'}\|_2. \quad (24)$$

Thus, we can assert that

$$\mu \sigma_{\min}(A_I^T A_I) \|x_I^* - x_I^{*'}\|_2 \leq \frac{1}{3} \|x_I^* - x_I^{*'}\|_2,$$

or equivalently,

$$(3\mu \sigma_{\min}(A_I^T A_I) - 1) \|x_I^* - x_I^{*'}\|_2 \leq 0.$$

By condition Eq.(16), we have $3\mu \sigma_{\min}(A_I^T A_I) - 1 > 0$, thus $\|x^* - x^{*'}\|_2 = \|x_I^* - x_I^{*'}\|_2 = 0$, that is, $x^* = x^{*'}$. Therefore, $\{x^{(n)}\}$ is convergent.

(2) We provide this proof in Appendix A.

Note that from the analysis of matrix A in [21], we know that when A satisfies a certain restricted isometry property (RIP), the condition Eq.(16) in Theorem 3 can therefore be satisfied.

Remark 2. In Theorem 3 (1), we proved convergence of the $L_{2/3}$ algorithm mainly based on properties of $Y_{\{x^{(n)}\}}$ [see Lemma 3]. In particular, our proof is more direct and simpler compared to [21].

4. Linear convergence rate

In Section 3, we have justified that the convergence of the $L_{2/3}$ algorithm to a local minimizer of $L_{2/3}$ regularization. In this section, we consider how fast the $L_{2/3}$

algorithm converges. We show that under the same conditions as Theorem 3, the asymptotical convergence rate of the $L_{2/3}$ algorithm is linear.

Theorem 4. Let $\{x^{(n)}\}$ be the sequence generated by the $L_{2/3}$ algorithm. If the sequence $\{x^{(n)}\}$ satisfies the conditions of Theorem 3 and let $\rho_0 = \frac{3}{2}(1 - \mu\sigma_{\min}(A_I^T A_I))$, then when n is sufficiently large,

- (1) $\|x^{(n+1)} - x^*\|_2 \leq \rho_0 \|x^{(n)} - x^*\|_2$;
- (2) $\|x^{(n+1)} - x^{(n)}\|_2 \leq \rho_0 \|x^{(n)} - x^{(n-1)}\|_2$.

Proof. By the condition $\frac{1}{3\sigma_{\min}(A_I^T A_I)} < \mu < \|A\|_2^{-2}$, we have $0 < \rho_0 < 1$.

(1). By using $x^{(n)} \rightarrow x^*$ as $n \rightarrow \infty$, Lemma 3 and Lemma 2 (3), there exists an n_0 such that when $n > n_0$,

$$\text{supp}(x^{(n)}) = \text{supp}(x^*) = I, \text{sign}(x^*) = \text{sign}(x^{(n)}), \quad (25)$$

and for any $i \in I$,

$$x_i^{(n+1)} + \frac{\lambda\mu}{3\sqrt[3]{x_i^{(n+1)}}} = [B_\mu(x^{(n)})]_i,$$

$$x_i^* + \frac{\lambda\mu}{3\sqrt[3]{x_i^*}} = [B_\mu(x^*)]_i,$$

where $B_\mu(z) = z - \mu A^T(Az - y)$. Therefore, we have

$$\begin{aligned} x_i^{(n+1)} - x_i^* + \frac{\lambda\mu}{3} \left(\frac{1}{\sqrt[3]{x_i^{(n+1)}}} - \frac{1}{\sqrt[3]{x_i^*}} \right) \\ = [B_\mu(x^{(n)})]_i - [B_\mu(x^*)]_i. \end{aligned}$$

Similarly by Lagrange mean value theorem, we can conclude that for any $i \in I$, there exists a $x_i^{\xi_i}$, which is between $x_i^{(n+1)}$ and x_i^* , such that

$$\frac{1}{\sqrt[3]{x_i^{(n+1)}}} - \frac{1}{\sqrt[3]{x_i^*}} = -\frac{1}{3(x_i^{\xi_i})^{4/3}}(x_i^{(n+1)} - x_i^*).$$

Thus, we get

$$[B_\mu(x^{(n)})]_i - [B_\mu(x^*)]_i = (1 - \frac{\lambda\mu}{9(x_i^{\xi_i})^{4/3}})(x_i^{(n+1)} - x_i^*).$$

Denoting the diagonal matrix with $\Lambda(i, i) = 1 - \frac{\lambda\mu}{9(x_i^{\xi_i})^{4/3}}$ by Λ , we can write

$$[B_\mu(x^{(n)})]_I - [B_\mu(x^*)]_I = \Lambda(x_I^{(n+1)} - x_I^*). \quad (26)$$

By the definition of $B_\mu(z)$, we have

$$\begin{aligned} [B_\mu(x^{(n)})]_i - [B_\mu(x^*)]_i \\ = (x_i^{(n)} - x_i^*) - \mu(A_i^T(Ax^{(n)} - y) - A_i^T(Ax^* - y)). \end{aligned}$$

Thus, we get

$$[B_\mu(x^{(n)})]_I - [B_\mu(x^*)]_I = (\mathbf{E} - \mu A_I^T A_I)(x_I^{(n)} - x_I^*), \quad (27)$$

where $\mathbf{E}$ is the identity matrix.

By Eq.(26) and Eq.(27), we conclude that

$$\Lambda(x_I^{(n+1)} - x_I^*) = (\mathbf{E} - \mu A_I^T A_I)(x_I^{(n)} - x_I^*).$$

Since $x_i^{\xi_i}$ is between $x_i^{(n+1)}$ and x_i^* , by Lemma 3 (1) and (2), for any $i \in I$, $|x_i^{\xi_i}| \geq (\frac{\lambda\mu}{3})^{3/4}$, so

$$|\Lambda(i, i)| \geq 1 - \frac{\lambda\mu}{9(\frac{\lambda\mu}{3})} = \frac{2}{3} > 0. \quad (28)$$

Thus, it follows that

$$x_I^{(n+1)} - x_I^* = \Lambda^{-1}(\mathbf{E} - \mu A_I^T A_I)(x_I^{(n)} - x_I^*). \quad (29)$$

Furthermore, we have

$$\|x_I^{(n+1)} - x_I^*\|_2 \leq \|\Lambda^{-1}\|_2 \|(\mathbf{E} - \mu A_I^T A_I)\|_2 \|x_I^{(n)} - x_I^*\|_2.$$

By Eq.(28), we have that $\|\Lambda^{-1}\|_2 \leq \frac{3}{2}$, hence

$$\begin{aligned} \|\Lambda^{-1}\|_2 \|(\mathbf{E} - \mu A_I^T A_I)\|_2 \|x_I^{(n)} - x_I^*\|_2 \\ \leq \frac{3(1 - \mu\sigma_{\min}(A_I^T A_I))}{2} \|x_I^{(n)} - x_I^*\|_2 \\ = \rho_0 \|x_I^{(n)} - x_I^*\|_2, \end{aligned}$$

Consequently, we obtain

$$\|x_I^{(n+1)} - x_I^*\|_2 \leq \rho_0 \|x_I^{(n)} - x_I^*\|_2. \quad (30)$$

Thus, the convergence rate of $\{x^{(n)}\}$ is asymptotically linear.

(2). By Lemma 2 (3), Lemma 3 (1) and Theorem 2 (2), there exists an n_1 such that when $n > n_1$, for any $i \in I$,

$$\text{supp}(x^{(n)}) = I, \text{sign}(x^{(n)}) = \text{sign}(x^{(n_1)}), \quad (31)$$

$$x_i^{(n)} \in (-\infty, -(\frac{\lambda\mu}{3})^{3/4}) \cup \{0\} \cup ((\frac{\lambda\mu}{3})^{3/4}, +\infty),$$

$$x_i^{(n+1)} + \frac{\lambda\mu}{3\sqrt[3]{x_i^{(n+1)}}} = [B_\mu(x^{(n)})]_i,$$

and

$$x_i^{(n)} + \frac{\lambda\mu}{3\sqrt[3]{x_i^{(n)}}} = [B_\mu(x^{(n-1)})]_i,$$

where $B_\mu(z) = z - \mu A^T(Az - y)$. Therefore, we have

$$\begin{aligned} x_i^{(n+1)} - x_i^{(n)} &+ \frac{\lambda\mu}{3} \left(\frac{1}{\sqrt[3]{x_i^{(n+1)}}} - \frac{1}{\sqrt[3]{x_i^{(n)}}} \right) \\ &= [B_\mu(x^{(n)})]_i - [B_\mu(x^{(n-1)})]_i. \end{aligned}$$

By Lagrange mean value theorem, we can conclude that for any $i \in I$, there exists a $x_i^{\xi_i}$, which between $x_i^{(n+1)}$ and $x_i^{(n)}$, such that

$$\frac{1}{\sqrt[3]{x_i^{(n+1)}}} - \frac{1}{\sqrt[3]{x_i^{(n)}}} = -\frac{1}{3(x_i^{\xi_i})^{4/3}}(x_i^{(n+1)} - x_i^{(n)}).$$

Thus, we get

$$[B_\mu(x^{(n)})]_i - [B_\mu(x^{(n-1)})]_i = (1 - \frac{\lambda\mu}{9(x_i^{\xi_i})^{4/3}})(x_i^{(n+1)} - x_i^{(n)}).$$

Denoting the diagonal matrix with $\Lambda(i, i) = 1 - \frac{\lambda\mu}{9(x_i^{\xi_i})^{4/3}}$ by Λ , we can write

$$[B_\mu(x^{(n)})]_I - [B_\mu(x^{(n-1)})]_I = \Lambda(x_I^{(n+1)} - x_I^{(n)}). \quad (32)$$

By the definition of $B_\mu(z)$, we have

$$\begin{aligned} [B_\mu(x^{(n)})]_i - [B_\mu(x^{(n-1)})]_i &= (x_i^{(n)} - x_i^{(n-1)}) \\ &- \mu(A_i^T(Ax^{(n)} - y) - A_i^T(Ax^{(n-1)} - y)). \end{aligned}$$

Thus, we get

$$[B_\mu(x^{(n)})]_I - [B_\mu(x^{(n-1)})]_I = (\mathbf{E} - \mu A_I^T A_I)(x_I^{(n)} - x_I^{(n-1)}), \quad (33)$$

where $\mathbf{E}$ is the identity matrix.

By Eq.(32) and Eq.(33), we deduce that

$$\Lambda(x_I^{(n+1)} - x_I^{(n)}) = (\mathbf{E} - \mu A_I^T A_I)(x_I^{(n)} - x_I^{(n-1)}). \quad (34)$$

Since $x_i^{\xi_i}$ is between $x_i^{(n)}$ and $x_i^{(n+1)}$, by Lemma 3 (1), we have $x_i^{\xi_i} > (\frac{\lambda\mu}{3})^{3/4}$. Thus, for all $i \in I$, it follows that

$$|\Lambda(i, i)| > 1 - \frac{\lambda\mu}{9(\frac{\lambda\mu}{3})} = \frac{2}{3} > 0. \quad (35)$$

Thus, by Eq.(34) we can obtain that

$$x_I^{(n+1)} - x_I^{(n)} = \Lambda^{-1}(\mathbf{E} - \mu A_I^T A_I)(x_I^{(n)} - x_I^{(n-1)}). \quad (36)$$

Furthermore, we have

$$\|x_I^{(n+1)} - x_I^{(n)}\|_2 \leq \|\Lambda^{-1}\|_2 \|(\mathbf{E} - \mu A_I^T A_I)\|_2 \|x_I^{(n)} - x_I^{(n-1)}\|_2.$$

By Eq.(35), we have $\|\Lambda^{-1}\|_2 < \frac{3}{2}$, hence

$$\begin{aligned} &\|\Lambda^{-1}\|_2 \|(\mathbf{E} - \mu A_I^T A_I)\|_2 \|x_I^{(n)} - x_I^{(n-1)}\|_2 \\ &\leq \frac{3(1 - \mu\sigma_{\min}(A_I^T A_I))}{2} \|x_I^{(n)} - x_I^{(n-1)}\|_2 \\ &= \rho_0 \|x_I^{(n)} - x_I^{(n-1)}\|_2, \end{aligned}$$

Consequently, we obtain

$$\|x_I^{(n+1)} - x_I^{(n)}\|_2 \leq \rho_0 \|x_I^{(n)} - x_I^{(n-1)}\|_2. \quad (37)$$

Thus, the asymptotically regular convergence rate of $\{x^{(n)}\}$ is linear. This completes the proof of Theorem 4.

By the Theorem 3 and 4, we can conclude that under the conditions $\lambda > 0$, $\frac{1}{3\sigma_{\min}(A_I^T A_I)} < \mu < \|A\|_2^{-2}$, the sequence $\{x^{(n)}\}$ generated by the $L_{2/3}$ algorithm is convergent, and its limit is a local minimizer of $L_{2/3}$ regularization. Thus, in the setting, the $L_{2/3}$ algorithm converges with an asymptotically linear convergence rate.

5. Experiments

In this section, we present numerical experiments to illustrate the linear convergence rate of $L_{2/3}$ algorithm.

We provide the numerical results that confirm the linear convergence rate of $L_{2/3}$ algorithm for $L_{2/3}$ regularization and its robust recovery of sparse signal. In our experiments, the original sparse signal x was generated randomly according to the standard Gaussian distribution with dimension $N = 500$ and sparsity $k = 15$. The location of nonzeros was uniformly and randomly generated. The measurement matrix A is of dimension $m \times N = 250 \times 500$ with its entries i.i.d. to Gaussian $N(0, 1/250)$. Such measurement matrix is known to satisfy (with high probability) the RIP with optimal bounds [26, 27] and thus the condition of Theorem 3 can be satisfied. We consider the signal recovery problem with observation y through $y = Ax$. We applied the $L_{2/3}$ algorithm to the problem with $\lambda = 0.001$ and $\mu = 0.99\|A\|_2^{-2}$. The mean square error (MSE) between the recovered signal and the original signal (i.e. $\text{MSE} = \|x^{(n)} - x\|_2^2 / N$) was recorded. The experiment results are shown in Figure 1 and 2.

In Figure 1, we show how the recovery MSEs of the iterations varies, from which we can see that, after approximately 1600 iterations, the $L_{2/3}$ algorithm converges to a stationary point with a linear decay rate. In Figure 2, we show the comparison between the recovered signal and the original signal, from which we can see that the original sparse signal has been recovered

Figure 1: The trend of MSEs for eventual linear convergence rate.

with very high accuracy ($\text{MSE} \approx 5.6259E - 09$). The numerical experiments clearly confirm the convergence properties of the $L_{2/3}$ algorithm and the expected linear rate of convergence of the algorithm.

6. Conclusion

In this paper, we present a detailed proof of convergence of the $L_{2/3}$ algorithm for $L_{2/3}$ regularization. The main contribution of this paper can be summarized as follows. First, we study the properties of the set of all limit points of $\{x^{(n)}\}$ generated by the $L_{2/3}$ algorithm. And then we show that under certain conditions, the $L_{2/3}$ algorithm converges to a local minimizer of $L_{2/3}$ regularization. We further consider the convergence rate of the $L_{2/3}$ algorithm, and we obtain that the asymptotical convergence rate of the $L_{2/3}$ algorithm is linear. Finally, numerical experiments were provided to verify the linear convergence rate of the theoretical analysis.

The established results that the convergence of the $L_{2/3}$ algorithm presented in this paper provide a theoretical guarantee for a wide range of applications of the $L_{2/3}$ algorithm. The $L_{2/3}$ regularization problem has wide applications beyond the image deconvolution, it has widely applied in other fields [23, 24, 25], such as compressive sensing, super-resolution and denoising. In our future work, we will extend the applications of the $L_{2/3}$ regularization to other related problems, such as non-negative matrix factorization [28] and low-rank matrix recovery.

Acknowledgments

We thank deeply the referees and the editor for their helpful suggestions and comments on the paper, which

Figure 2: Recovery sparse signal with recovery MSE is $5.6259E - 09$.

have improved the presentation. This paper is partially supported by the Shanghai Municipal Natural Science Foundation (15ZR1416300), NNSF of China (61071186, 11171205).

Appendix A. The proof of Theorem 3 (2)

Proof. Let $\beta^* = 1 - \left(\frac{1}{3\mu\sigma_{\min}(A_I^T A_I)}\right)^{3/4}$. Since $\frac{1}{3\mu\sigma_{\min}(A_I^T A_I)} < \mu < \|A\|_2^{-2}$, we have $0 < \beta^* < 1$. Denote

$$\beta_0 = \min \left\{ \left(\frac{\lambda\mu}{3}\right)^{3/4}\beta^*, \frac{1}{\left(\frac{2\sqrt[4]{48}}{3(\lambda\mu)^{1/4}} + \frac{2}{\lambda}\|A_{I^c}^T A_I\|_2\left(\frac{\lambda\mu}{3}\right)^{3/4}\right)^3} \right\}.$$

Let x^* be the limit of the sequence $\{x^{(n)}\}$. We will proceed to show that whenever $\|h\|_2 < \beta_0$, it holds

$$C_\lambda(x^* + h) - C_\lambda(x^*) \geq 0.$$

Actually, we have

$$\begin{aligned}
& C_\lambda(x^* + h) - C_\lambda(x^*) \\
&= 2h^T A^T (Ax^* - y) + h^T (A^T A)h \\
&+ \lambda \sum_{i=1}^N (|x_i^* + h_i|^{2/3} - |x_i^*|^{2/3}) \\
&= \lambda \sum_{i=1}^N \left(|x_i^* + h_i|^{2/3} - |x_i^*|^{2/3} + \frac{2}{\lambda} A_i^T (Ax^* - y) h_i \right) \\
&+ h^T (A^T A)h \\
&= \lambda \sum_{i \in I} \left(|x_i^* + h_i|^{2/3} - |x_i^*|^{2/3} + \frac{2}{\lambda} A_i^T (Ax^* - y) h_i \right) \\
&+ \lambda \sum_{i \in I^c} \left(|h_i|^{2/3} + \frac{2}{\lambda} A_i^T (Ax^* - y) h_i \right) + h^T (A^T A)h.
\end{aligned} \tag{A.1}$$

By Lemma 3 (3) and (4), for any $i \in I^c$, we have

$$|h_i|^{2/3} + \frac{2}{\lambda} A_i^T (Ax^* - y) h_i \geq |h_i|^{2/3} - \frac{2\sqrt[4]{48}}{3(\lambda\mu)^{1/4}} |h_i|, \tag{A.2}$$

and for $i \in I$,

$$\begin{aligned}
& |x_i^* + h_i|^{2/3} - |x_i^*|^{2/3} + \frac{2}{\lambda} A_i^T (Ax^* - y) h_i \\
&= |x_i^* + h_i|^{2/3} - |x_i^*|^{2/3} - \frac{2}{3\sqrt[3]{x_i^*}} h_i.
\end{aligned} \tag{A.3}$$

By Taylor expansion, for any $i \in I$, we obtain

$$|x_i^* + h_i|^{2/3} - |x_i^*|^{2/3} - \frac{2}{3\sqrt[3]{x_i^*}} h_i = \frac{-h_i^2}{9|x_i^* + \xi_i h_i|^{4/3}}, \tag{A.4}$$

for some constant $\xi_i \in (0, 1)$. Furthermore, since

$$\max_{i \in I} |h_i| \leq \|h\|_2 < \beta_0,$$

we have for any $i \in I$,

$$|x_i^* + \xi_i h_i| \geq |x_i^*| - \xi_i |h_i| \geq \left(\frac{\lambda\mu}{3}\right)^{3/4} - \beta_0 > 0,$$

which shows

$$\begin{aligned}
& |x_i^* + h_i|^{2/3} - |x_i^*|^{2/3} + \frac{2}{\lambda} A_i^T (Ax^* - y) h_i \\
&\geq -\frac{h_i^2}{9((\frac{\lambda\mu}{3})^{3/4} - \beta_0)^{4/3}}.
\end{aligned} \tag{A.5}$$

From Eq.(A.1)- Eq.(A.5), it thus follows that

$$\begin{aligned}
& C_\lambda(x^* + h) - C_\lambda(x^*) \\
&\geq \|Ah\|_2^2 + \lambda \sum_{i \in I^c} \left(|h_i|^{2/3} - \frac{2\sqrt[4]{48}}{3(\lambda\mu)^{1/4}} |h_i| \right) \\
&- \lambda \sum_{i \in I} \frac{h_i^2}{9((\frac{\lambda\mu}{3})^{3/4} - \beta_0)^{4/3}} \\
&\geq \left(\|A_I h_I\|_2^2 - \frac{\lambda}{9((\frac{\lambda\mu}{3})^{3/4} - \beta_0)^{4/3}} \|h_I\|_2^2 \right) + \|A_{I^c} h_{I^c}\|_2^2 \\
&+ 2(A_I h_I)^T A_{I^c} h_{I^c} + \lambda \sum_{i \in I^c} \left(|h_i|^{2/3} - \frac{2\sqrt[4]{48}}{3(\lambda\mu)^{1/4}} |h_i| \right) \\
&\geq \lambda \sum_{i \in I^c} \left(|h_i|^{2/3} - \frac{2\sqrt[4]{48}}{3(\lambda\mu)^{1/4}} |h_i| - \frac{2}{\lambda} |A_i^T (A_I h_I)| |h_i| \right) \\
&+ \left(\sigma_{\min}(A_I^T A_I) - \frac{\lambda}{9((\frac{\lambda\mu}{3})^{3/4} - \beta_0)^{4/3}} \right) \|h_I\|_2^2.
\end{aligned} \tag{A.6}$$

Since $\|h\|_2 < \beta_0$, for any $i \in I^c$, we find that

$$\begin{aligned}
& |h_i|^{1/3} \leq \frac{1}{\frac{2\sqrt[4]{48}}{3(\lambda\mu)^{1/4}} + \frac{2}{\lambda} \|A_{I^c}^T A_I\|_2 (\frac{\lambda\mu}{3})^{3/4}} \\
&\leq \frac{1}{\frac{2\sqrt[4]{48}}{3(\lambda\mu)^{1/4}} + \frac{2}{\lambda} \|A_{I^c}^T A_I\|_2 \|h_I\|_2} \\
&\leq \frac{1}{\frac{2\sqrt[4]{48}}{3(\lambda\mu)^{1/4}} + \frac{2}{\lambda} \|A_{I^c}^T A_I h_I\|_2} \\
&\leq \frac{1}{\frac{2\sqrt[4]{48}}{3(\lambda\mu)^{1/4}} + \frac{2}{\lambda} \max_{i \in I^c} |A_i^T A_I h_I|} \\
&\leq \frac{1}{\frac{2\sqrt[4]{48}}{3(\lambda\mu)^{1/4}} + \frac{2}{\lambda} |A_i^T A_I h_I|}.
\end{aligned} \tag{A.7}$$

Consequently,

$$\begin{aligned}
& |h_i|^{2/3} - \frac{2\sqrt[4]{48}}{3(\lambda\mu)^{1/4}} |h_i| - \frac{2}{\lambda} |A_i^T (A_I h_I)| |h_i| \\
&= \left(1 - \left(\frac{2\sqrt[4]{48}}{3(\lambda\mu)^{1/4}} + \frac{2}{\lambda} |A_i^T A_I h_I| \right) |h_i|^{1/3} \right) |h_i|^{2/3} \\
&\geq 0.
\end{aligned} \tag{A.8}$$

By Eq.(A.6)- Eq.(A.8), we then obtain

$$\begin{aligned}
& C_\lambda(x^* + h) - C_\lambda(x^*) \\
&\geq \left(\sigma_{\min}(A_I^T A_I) - \frac{\lambda}{9((\frac{\lambda\mu}{3})^{3/4} - \beta_0)^{4/3}} \right) \|h_I\|_2^2.
\end{aligned}$$

Since $\beta_0 \leq (\frac{\lambda\mu}{3})^{3/4}\beta^*$, this implies

$$\begin{aligned} \sigma_{\min}(A_I^T A_I) - \frac{\lambda}{9((\frac{\lambda\mu}{3})^{3/4} - \beta_0)^{4/3}} \\ \geq \sigma_{\min}(A_I^T A_I) - \frac{\lambda}{9((\frac{\lambda\mu}{3})^{3/4} - (\frac{\lambda\mu}{3})^{3/4}\beta^*)^{4/3}} \\ \geq 0, \end{aligned}$$

and consequently,

$$C_\lambda(x^* + h) - C_\lambda(x^*) \geq 0.$$

Thus, x^* is a local minimizer of $L_{2/3}$ regularization.

References

- [1] R. Chartrand, Exact reconstruction of sparse signals via non-convex minimization, *IEEE Signal Process. Lett.* 14(10) (2007) 707-710.
- [2] R. Chartrand, V. Staneva, Restricted isometry properties and nonconvex compressive sensing, *Inverse Prob.* 24(3) (2008) 1-14.
- [3] Z. B. Xu, H. Zhang, Y. Wang, X. Y. Chang, $L_{1/2}$ regularization, *Sci. China* 53(6) (2010) 1159-1169.
- [4] Z. B. Xu, X. Y. Chang, F. M. Xu, H. Zhang, $L_{1/2}$ regularization: a thresholding representation theory and a fast solver, *IEEE Trans. Neural Networks Learning Syst.* 23(7) (2012) 1013-1027.
- [5] R. Tibshirani, Regression shrinkage and selection via the lasso, *J. Roy. Statist. Soc., Series B(Methodolog.)* 58(1) (1996) 267-288.
- [6] F. P. Nie, H. Huang, X. Cai, C. Ding, Efficient and robust feature selection via joint $\ell_{2,1}$ -norms minimization, *Proc. Adv. Neural Inf. Process. Syst. (NIPS)* (2010) 1813-1821.
- [7] H. Tao, C. P. Hou, F. P. Nie, Y. Y. Jiao, D. Y. Yi, Effective discriminative feature selection with non-trivial solutions, *arXiv preprint* (2015) arXiv: 1504.05408.
- [8] T. Y. Zhou, D. C. Tao, Double shrinking sparse dimension reduction, *IEEE Trans. Image Process.* 22(1) (2013) 244-257.
- [9] D. L. Donoho, Compressed sensing, *IEEE Trans. Inf. Theory* 52(4) (2006) 1289-1306.
- [10] E. Candes, J. Romberg, T. Tao, Robust uncertainty principles : exact signal reconstruction from highly incomplete frequency information, *IEEE Trans. Inf. Theory* 52(2) (2006) 489-509.
- [11] E. J. Candès, M. B. Wakin, S. P. Boyd, Enhancing sparsity by reweighted l_1 minimization, *J. Fourier Anal. Appl.* 14(5) (2008) 877-905.
- [12] F. P. Nie, H. Wang, H. Huang, C. Ding, Joint Schatten p -norm and ℓ_p -norm robust matrix completion for missing value recovery, *Knowl. Inf. Syst.* 42(3) (2015) 525-544.
- [13] F. P. Nie, H. Huang, C. Ding, Low-Rank matrix recovery via efficient Schatten p -norm minimization, *Proc. Twenty-Sixth AAAI Conference on Artificial Intelligence.* (2012) 655-661.
- [14] D. Krishnan, R. Fergus, Fast image deconvolution using hyper-Laplacian priors, *Proc. Adv. Neural Inf. Process. Syst. (NIPS)* 2009.
- [15] W. F. Cao, J. Sun, Z. B. Xu, Fast image deconvolution using closed-form thresholding formulas of $L_q(q = 1/2, 2/3)$ regularization, *J. Vis. Commun. Image Represent.* 24(1) (2013) 1529-1542.
- [16] T. Blumensath, M. Yaghoobi, M. E. Davies, Iterative hard thresholding and L_0 regularisation, *IEEE Trans. Acoust. Speech Signal Process.* 3 (2007).
- [17] D. L. Donoho, Denoising by soft thresholding, *IEEE Trans. Inf. Theory* 41(3) (1995) 613-627.
- [18] I. Daubechies, M. Defrise, D. M. Christine, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, *Commun. Pure and Appl. Math.* 57(11) (2004) 1413-1457.
- [19] J. Gui, D. C. Tao, Z. N. Sun, Y. Luo, X. G. You, Y. Y. Tang, Group sparse multiview patch alignment framework with view consistency for image classification, *IEEE Trans. Image Process.* 23(7) (2014) 3126-3137.
- [20] X. M. Tian, D. C. Tao, Y. Rui, Sparse transfer learning for interactive video search reranking, *ACM Trans. Multim. Comput. Appl.* 8(3) (2012) 26:1-19.
- [21] J. S. Zeng, S. B. Lin, Y. Wang, Z. B. Xu, $L_{1/2}$ Regularization: Convergence of Iterative Half Thresholding Algorithm, *IEEE Trans. Signal Process.* 62(9) (2014) 2317-2329.
- [22] L. Liu, D. R. Chen, Convergence of $\ell_{2/3}$ regularization for sparse signal recovery, *Asia-Pacific J. Operational Research*, 32(3) (2015) 1550023.
- [23] S. Foucart, M. Lai, Sparsest solutions of underdetermined linear systems via l_q minimization for $0 < q \leq 1$, *Appl. Comput. Harmonic Anal.* 26(3) (2009) 395-407.
- [24] Q. Sun, Recovery of sparsest signals via l_q minimization, *Appl. Comput. Harmonic Anal.* 32 (2012) 329-341.
- [25] Y. B. Sun, Q. S. Liu, J. H. Tang, D. C. Tao, Learning discriminative dictionary for group sparse representation, *IEEE Trans. Image Process.* 23(9) (2014) 3816-3828.
- [26] M. Rudelson, R. Vershynin, On sparse reconstruction from Fourier and Gaussian measurements, *Comm. Pure Appl. Math.* 61 (2008) 1025-1045.
- [27] R. Baraniuk, M. Davenport, R. Devore, M. B. Wakin, A simple proof of the restricted isometry property for random matrices, *Constr. Approx.* 28(3) (2008) 253-263.
- [28] T. L. Liu, D. C. Tao, On the Performance of Manhattan Non-negative Matrix Factorization, (2015) <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7192641>.

Figure

Figure