

RESEARCH

Open Access



Optimality condition and iterative thresholding algorithm for l_p -regularization problems

Hongwei Jiao^{1*}, Yongqiang Chen² and Jingben Yin¹

*Correspondence:

jiaohongwei@126.com

¹ School of Mathematical Sciences, Henan Institute of Science and Technology, Xinxiang 453003, China
Full list of author information is available at the end of the article

Abstract

This paper investigates the l_p -regularization problems, which has a broad applications in compressive sensing, variable selection problems and sparse least squares fitting for high dimensional data. We derive the exact lower bounds for the absolute value of nonzero entries in each global optimal solution of the model, which clearly demonstrates the relation between the sparsity of the optimum solution and the choice of the regularization parameter and norm. We also establish the necessary condition for global optimum solutions of l_p -regularization problems, i.e., the global optimum solutions are fixed points of a vector thresholding operator. In addition, by selecting parameters carefully, a global minimizer which will have certain desired sparsity can be obtained. Finally, an iterative thresholding algorithm is designed for solving the l_p -regularization problems, and any accumulation point of the sequence generated by the designed algorithm is convergent to a fixed point of the vector thresholding operator.

Keywords: Optimality condition, l_p -regularization problems, Iterative thresholding algorithm, Global optimum solution, Fixed point

Mathematics Subject Classification: 90C26, 90C46, 90C90

Background

In this paper, we investigate the following l_p -regularization problems

$$\min_{s \in \mathbb{R}^n} f_\lambda(s) := \|As - b\|_2^2 + \lambda \|s\|_p^p \quad (1)$$

where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $\lambda \in (0, \infty)$, $\|s\|_p^p = \sum_{i=1}^n |s_i|^p$, $p \in (0, 1)$. The problem (1) has a broad applications in compressive sensing, variable selection problems and sparse least squares fitting for high dimensional data (see Chartrand and Staneva 2008; Fan and Li 2001; Foucart and Lai 2009; Frank and Freidman 1993; Ge et al. 2011; Huang et al. 2008; Knight and Wu 2000; Lai and Wang 2011; Natarajan 1995). The objective function of the problem (1) is consisted by a data fitting term $\|As - b\|_2^2$ and a regularization term $\lambda \|s\|_p^p$. In Chen et al. (2014) point out that the l_2 - l_p minimization problem (1) is a strongly NP-hard problem. Comparing with using the l_1 norm, using the l_p quasi-norm in the regularization term we can find sparser solution, which has been extensively discussed in Candès et al. (2008), Chartrand (2007a, b), Chartrand and Yin (2008), Chen et al. (2010),

Tian and Huang (2013), Tian and Jiao (2015), Xu et al. (2010, 2012), Shehu et al. (2013, 2015), Bredies et al. (2015), Fan et al. (2016). In Chen et al. (2010), Chen et al. derive the lower bounds for the absolute value of nonzero entries in each local optimum solution of the model. Xu et al. (2012) presented an analytical expression in a thresholding form for the resolvent of gradient of $\|s\|_{1/2}^{1/2}$ and developed an alternative feature theorem on optimum solutions of the $L_{1/2}$ regularization problem, and proposed an iterative half thresholding algorithm for fast solving the problem. But there is no result for the characteristics of the global optimum solution for the problem (1).

In this article, we pay more attention to derive the characteristics of the global optimum solution of problem (1), which is inspired by Xu et al. (2012). The remaining sections of the paper are organized as follows. In “[Technical preliminaries](#)” section, we portray some important technical results. “[Lower bound and optimality conditions](#)” section first develop the proximal operator associated with a non-convex l_p quasi-norm, which can be looked as an extension of the well-known proximal operator associated with convex functions. Next, an exact lower bound for the absolute value of nonzero entries in every global optimum solution of (1) is derived, which clearly demonstrates the relation between the sparsity of the optimum solution and the choice of the regularization parameter and norm. We also establish the necessary condition for global optimum solutions of the l_p -regularization problems, i.e., the global optimum solutions are fixed points of a vector thresholding operator. In “[Choosing the parameter \$\lambda\$ for sparsity](#)” section, we also propose a sufficient condition on the selection of λ to meet the sparsity requirement of global minimizers of the l_p -regularization problems. “[Iterative thresholding algorithm and its convergence](#)” section proposes an iterative thresholding algorithm for the l_p -regularization problems, and any accumulation point of the sequence produced by the designed algorithm is convergent to a fixed point of the vector thresholding operator. Finally, some conclusions are drawn in “[Numerical experiments](#)” section.

Technical preliminaries

By utilizing the objective function’s separability and the operator splitting technique, the l_p -regularization problems (1) can be converted into n homologous single variable minimization problems defined on $(-\infty, +\infty)$. Therefore, at first we investigate the homologous single variable minimization problem

$$\min_{s \in R} g_r(s) := s^2 - 2rs + \lambda|s|^p, \tag{2}$$

where $\lambda > 0$ and $p \in (0, 1)$ are all any real numbers, $s \in R$ is a variable and $r \in R$ is a parameter. Besides, we only need to consider the following two sub-problems

$$\min_{s \geq 0} g_r(s) = s^2 - 2rs + \lambda s^p, \tag{3}$$

$$\min_{s \leq 0} g_r(s) = s^2 - 2rs + \lambda(-s)^p. \tag{4}$$

In Chen et al. (2014), investigated the subproblem (3) and presented some results, which can be used to derive our conclusions. Let

$$\bar{r} := \frac{2-p}{1-p} [\lambda p(1-p)/2]^{1/(2-p)} > 0, \tag{5}$$

$$\bar{s} := [\lambda p(1-p)/2]^{1/(2-p)} > 0. \tag{6}$$

Lemma 1 (Lemma.2.2, Chen et al. 2014) *For any $s > 0$, denote $G(s, r) := [g_r(s)]' = 2s - 2r + \lambda ps^{p-1}$. For any known $r_0 > \bar{r}$, set s_0 ($s_0 > \bar{s}$) be the positive root of the equation $G(s, r_0) = 0$, where \bar{r} and \bar{s} are given in (5) and (6). Then, there is a unique implicit function $s = h_{\lambda,p}(r)$ define on $(\bar{r}, +\infty)$, which satisfies $s_0 = h_{\lambda,p}(r_0)$, $h_{\lambda,p}(r) > \bar{s}$ and $G(h_{\lambda,p}(r), r) \equiv 0$ for $\forall r \in (\bar{r}, +\infty)$. Furthermore, for the function $s = h_{\lambda,p}(r)$, the following conclusions hold:*

1. $s = h_{\lambda,p}(r)$ is a continuous function defined on $(\bar{r}, +\infty)$.
 2. $s = h_{\lambda,p}(r)$ is a differentiable function over $(\bar{r}, +\infty)$ and $h'_{\lambda,p}(r) = \frac{2}{2+\lambda p(p-1)h_{\lambda,p}^{p-2}(r)}$.
 3. $s = h_{\lambda,p}(r)$ is a strictly increasing function over $(\bar{r}, +\infty)$.
- Moreover, if $r > \bar{r}$, then $s = h_{\lambda,p}(r)$ is the sole local minimizer of $g_r(s)$ over $(0, +\infty)$.

Lemma 2 (Prop.2.4, Chen et al. 2014) *Set s^* be the global optimum solution for the problem (3), then we have*

$$s^* = h_{\lambda}(r) := \begin{cases} h_{\lambda,p}(r), & r > r^* \\ (\lambda(1-p))^{1/(2-p)} \text{ or } 0, & r = r^* \\ 0, & r < r^* \end{cases}$$

where $r^* := \frac{2-p}{2(1-p)} [\lambda(1-p)]^{1/(2-p)}$, $h_{\lambda,p}(r)$ is defined by Lemma 1.

Proposition 1 *Set s^* be the global optimum solution for the problem (2), then we have*

$$s^* = h_{\lambda}(r) := \begin{cases} h_{\lambda,p}(r), & r > r^* \\ L \text{ or } 0, & r = r^* \\ 0, & -r^* < r < r^* \\ -L \text{ or } 0, & r = -r^* \\ -h_{\lambda,p}(-r), & r < -r^* \end{cases} \tag{7}$$

where $r^* := \frac{2-p}{2(1-p)} [\lambda(1-p)]^{1/(2-p)}$, $h_{\lambda,p}(r)$ is defined in Lemma 1 and $L := (\lambda(1-p))^{1/(2-p)}$.

Proof If $s \geq 0$, then $g_r(s) = s^2 - 2rs + \lambda s^p$. Let s_1^* is a global optimum solution for the problem (3), then from Lemma 2, we have

$$s_1^* = h_{\lambda}(r) = \begin{cases} h_{\lambda,p}(r), & r > r^* \\ (\lambda(1-p))^{1/(2-p)} \text{ or } 0, & r = r^* \\ 0. & r < r^* \end{cases} \tag{8}$$

If $s \leq 0$, then $g_r(s) = s^2 - 2rs + \lambda(-s)^p = (-s)^2 + 2r(-s) + \lambda(-s)^p$. Let $y = -s$, we have $y \geq 0$ and $g_{(-r)}(y) = y^2 + 2ry + \lambda y^p$, we follow the first case. If y^* is a global optimum solution for the problem $g_{(-r)}(y)$ over $[0, +\infty)$, then from Lemma 2, we have

$$y^* = h_{\lambda}(-r) = \begin{cases} h_{\lambda,p}(-r), & -r > r^* \\ (\lambda(1-p))^{1/(2-p)} \text{ or } 0, & -r = r^* \\ 0. & -r < r^* \end{cases}$$

Therefore, if $s \leq 0$, s_2^* is a global optimum solution for the problem $\min_{s \in R_-} g_r(s) = s^2 - 2rs + \lambda(-s)^p$, then we have

$$s_2^* = -y^* = \begin{cases} -h_{\lambda,p}(-r), & r < -r^* \\ -(\lambda(1-p))^{1/(2-p)} \text{ or } 0, & r = -r^* \\ 0. & r > -r^* \end{cases} \tag{9}$$

Combining (8) and (9) together, we can get (7). Therefore, the proof is complete. □

Proposition 2 Assume that s^* is a global optimum solution for the problem (2). When $|r| = r^*$ given in Proposition 1, set $s^* = h_{\lambda}(r)$ be simultaneously zero or nonzero. Then the following conclusions hold:

1. The function $h_{\lambda}(r)$ is an odd function over $(-\infty, +\infty)$.
2. The function $h_{\lambda}(r)$ is continuous over $(r^*, +\infty)$, furthermore, $\lim_{r \downarrow r^*} h_{\lambda}(r) = L$.
3. The function $h_{\lambda}(r)$ is differentiable over $(r^*, +\infty)$.
4. The function $h_{\lambda}(r)$ is strictly increasing over $(r^*, +\infty)$.

Proof By Proposition 1 and Lemma 1, this proposition can be followed. □

When $p = 1/2$, in Xu et al. (2012), $h_{v,p}(r)$ of (7) has the following analytic corollary.

Corollary 1 (Theo. 1, Lemm. 1 and 2, Xu et al. 2012) When $p = 1/2$, the global optimum solution s^* of problem (2) has the following results:

$$s^* = h_{\lambda}(r) := \begin{cases} h_{\lambda,1/2}(r), & r > r^* \\ (\lambda/2)^{2/3} \text{ or } 0, & r = r^* \\ 0, & -r^* < r < r^* \\ -(\lambda/2)^{2/3} \text{ or } 0, & r = -r^* \\ -h_{\lambda,1/2}(-r), & r < -r^* \end{cases} \tag{10}$$

where $h_{\lambda,1/2}(r) = \frac{2}{3}r(1 + \cos(\frac{2\pi}{3} - \frac{2\varphi(r)}{3}))$, $\varphi(r) = \arccos(\frac{\lambda}{8}(\frac{|r|}{3})^{-3/2})$ and $r^* = \frac{\sqrt[3]{54}}{4}\lambda^{2/3}$.

Proof A brief proof is presented here for completeness. When $p = 1/2$, we have $r^* = \frac{\sqrt[3]{54}}{4}\lambda^{2/3}$. When $|r| > r^*$, $s^* = h_{\lambda}(r) \neq 0$, by Proposition 2, then $h_{\lambda}(r)$ is the root of the equation

$$s - r + \frac{\lambda \text{sign}(s)}{4\sqrt{|s|}} = 0,$$

which is followed by the first order optimum condition of (2). By Theorem 1 of Xu et al. (2012), we have $h_{\lambda,1/2}(r) = \frac{2}{3}r(1 + \cos(\frac{2\pi}{3} - \frac{2\varphi(r)}{3}))$, $\varphi(r) = \arccos(\frac{1}{8}(\frac{|r|}{3})^{-3/2})$. The proof is completed. \square

Lower bound and optimality conditions

In this section, by using function’s separability and the operator splitting technique, we propose the proximal operator associated with l_p quasi-norm. Next, we present the properties of the global optimum solutions of the l_p -regularization problems (1). For convenience, first of all, we define the following thresholding function and thresholding operators.

Definition 1 (*p thresholding function*) Assume that $r \in R$, for any $\lambda > 0$, the function $h_\lambda(r)$ defined in (7) is called as a p thresholding function.

Definition 2 (*Vector p thresholding operator*) Assume that $s \in R^n$, for any $\lambda > 0$, the vector p thresholding operator $H_\lambda(s)$ is defined as

$$H_\lambda(s) := (h_\lambda(s_1), h_\lambda(s_2), \dots, h_\lambda(s_n))^T.$$

In this section, one of the main results is a proximal operator associated with the non-convex l_p ($0 < p < 1$) quasi-norm, and which can be also looked as an extension of the well-known proximal operator associated with convex functions.

Theorem 1 For given a vector $y \in R^n$ and constants $\lambda > 0$, $0 < p < 1$. Assume that s^* be the global optimum solution of the following problem

$$\min_{s \in R^n} f(s) := \|s - y\|_2^2 + \lambda \|s\|_p^p, \tag{11}$$

then s^* can be expressed as

$$s^* = H_\lambda(y).$$

Furthermore, we can get the exact number of global optimum solutions for the problem.

Proof From

$$\begin{aligned} f(s) &= \|s - y\|_2^2 + \lambda \|s\|_p^p = \|s\|_2^2 - 2\langle s, y \rangle + \|y\|_2^2 + \lambda \|s\|_p^p \\ &= \sum_{i=1}^n (s_i^2 - 2y_i s_i + \lambda |s_i|^p) + \|y\|_2^2. \end{aligned}$$

Let $g_{y_i}(s_i) = s_i^2 - 2y_i s_i + \lambda |s_i|^p$, then

$$f(s) = \sum_{i=1}^n g_{y_i}(s_i) + \|y\|_2^2.$$

Therefore, to solve the problem (11) is equivalent to solving the following n problems, for each $i = 1, 2, \dots, n$,

$$\min_{s_i \in R} g_{y_i}(s_i). \tag{12}$$

By Proposition 1, for each $i = 1, 2, \dots, n$, we can follow

$$s_i^* = \arg \min_{s_i \in R} g_{y_i}(s_i) = h_\lambda(y_i),$$

and if $|y_i| = r^* := \frac{2-p}{2(1-p)}[\lambda(1-p)]^{1/(2-p)}$, the problem (12) has two solutions; else, unique solution. Hence we can know the exact number of global optimum solutions of (11). The proof is thus complete. \square

For any $\lambda, \mu > 0, 0 < p < 1$, and $z \in R^n$, let

$$f_\mu(s, z) := \mu(f_\lambda(s) - \|As - Az\|_2^2) + \|s - z\|_2^2, \tag{13}$$

For simplicity, let

$$B_\mu(z) := z + \mu A^T(b - Az). \tag{14}$$

Theorem 2 Assume that $s^* \in R^n$ be the global minimizer of $f_\mu(s, z)$ for any fixed $\lambda > 0, \mu > 0$ and $z \in R^n$, then we have

$$s^* = H_{\lambda\mu}(B_\mu(z)). \tag{15}$$

Proof Without loss of generality, $f_\mu(s, z)$ can be rewritten as

$$\begin{aligned} f_\mu(s, z) &= \mu(\|As - b\|_2^2 + \lambda\|s\|_p^p - \|As - Az\|_2^2) + \|s - z\|_2^2 \\ &= \lambda\mu\|s\|_p^p + \|s\|_2^2 - 2\langle s, z + \mu A^T(b - Az) \rangle + \|z\|_2^2 + \mu\|b\|_2^2 - \mu\|Az\|_2^2 \\ &= \|s - B_\mu(z)\|_2^2 + \lambda\mu\|s\|_p^p + \|z\|_2^2 + \mu\|b\|_2^2 - \mu\|Az\|_2^2 - \|B_\mu(z)\|_2^2. \end{aligned}$$

Therefore, to solve $\min_{s \in R^n} f_\mu(s, z)$ for any fixed v, μ and Y is equivalent to solving

$$\min_{s \in R^n} \{ \|s - B_\mu(z)\|_2^2 + \lambda\mu\|s\|_p^p \}.$$

By Theorem 1, thus the proof is complete. \square

Lemma 3 If $s^* \in R^n$ is a global minimizer of the problem (1) for any fixed $v > 0$ and for any fixed μ which satisfies $0 < \mu \leq \|A\|^{-2}$, then s^* is also a global minimizer of $f_\mu(s, s^*)$, that is,

$$f_\mu(s^*, s^*) \leq f_\mu(s, s^*) \quad \text{for all } s \in R^n.$$

Proof For any $s \in R^n$, Since $0 < \mu \leq \|A\|^{-2}$, we have

$$\|s - s^*\|_2^2 - \mu\|As - As^*\|_2^2 \geq \|s - s^*\|_2^2 - \mu\|A\|^2\|s - s^*\|_2^2 \geq 0.$$

Hence,

$$\begin{aligned}
 f_\mu(s, s^*) &= \mu(f_\lambda(s) - \|As - As^*\|_2^2) + \|s - s^*\|_2^2 \\
 &= \mu(\|As - b\|_2^2 + \lambda\|s\|_p^p) + (\|s - s^*\|_2^2 - \mu\|As - As^*\|_2^2) \\
 &\geq \mu(\|As - b\|_2^2 + \lambda\|s\|_p^p) \\
 &= \mu f_\lambda(s) \geq \mu f_\lambda(s^*) \\
 &= f_\mu(s^*, s^*)
 \end{aligned}$$

the proof is complete. □

Theorem 3 For any given $\lambda > 0$, $0 < \mu \leq \|A\|^{-2}$, if s^* be the global optimum solution of the problem (1), then s^* satisfies

$$s^* = H_{\lambda\mu}(B_\mu(s^*)). \tag{16}$$

Especially, we have

$$\begin{aligned}
 s_i^* &= h_{\lambda\mu}([B_\mu(s^*)]_i) \\
 &= \begin{cases} h_{\lambda\mu,p}([B_\mu(s^*)]_i), & \text{if } [B_\mu(s^*)]_i > r^* \\ L \text{ or } 0, & \text{if } [B_\mu(s^*)]_i = r^* \\ 0, & \text{if } -r^* < [B_\mu(s^*)]_i < r^* \\ -L \text{ or } 0, & \text{if } [B_\mu(s^*)]_i = -r^* \\ -h_{\lambda\mu,p}(-[B_\mu(s^*)]_i), & \text{if } [B_\mu(s^*)]_i < -r^* \end{cases} \tag{17}
 \end{aligned}$$

where $r^* := \frac{2-p}{2(1-p)}[\lambda\mu(1-p)]^{1/(2-p)}$ and $L := (\lambda\mu(1-p))^{1/(2-p)}$.

Furthermore, we have: if $s_i^* \in (-L, L)$, then $s_i^* = 0$.

Proof Since s^* is a global minimizer of $f_\mu(s, z)$ for given $z = s^*$, by Theorem 2 and Lemma 3, we can directly get (16) and (17). By proposition 2, we can follow that

$$\lim_{r \downarrow r^*} h_{\lambda\mu}(r) = [\lambda\mu(1-p)]^{\frac{1}{2-p}} =: L.$$

By Proposition 2, combining with the strict monotonicity of $h_{\lambda\mu}(\cdot)$ on $(\bar{r}, +\infty)$ and $(-\infty, -\bar{r})$, we can follow that $s_i^* > L$ as $[B_\mu(s^*)]_i > r^*$, $s_i^* < -L$ as $[B_\mu(s^*)]_i < -r^*$ and $|s_i^*| = L$ as $|[B_\mu(s^*)]_i| = r^*$. Therefore, the proof is completed. □

Remark 1 In Theorem 3, the necessary condition for global optimum solutions of the l_p -regularization problems is established, which is a thresholding expression associated with the global optimum solutions. Particularly, the global optimum solutions for the problem (1) are the fixed points of a vector-valued thresholding operator. In contrast, the conclusion does not hold in general, i.e., a point satisfying (16) is not the global optimum solution for the l_p -regularization problems (1) in general. This is related to the nature of the matrix A , for an instance, when $A \equiv I$ and $\mu = 1$, a fixed point of (16) is the global optimum solution for the l_p -regularization problems (1) (i.e., Theorem 1).

Remark 2 In Theorem 3, the exact lower bound for the absolute value of nonzero entries in every global optimum solution of the model is also provided, which can be used to identify zero entries precisely in any global optimum solution. These lower bounds clearly demonstrate the relationship between the sparsity of the global optimum solution and the choices of the regularization parameter and norm, therefore, our theorem can be used to select the desiring model parameters and norms.

Choosing the parameter λ for sparsity

In many applications such that sparse solution reconstruction and variable selection, one need to seek out least square estimators with no more than k nonzero entries. Chen et al. (2014) present a sufficient condition on λ for global minimizers of the l_p -regularization problems, which have desirable sparsity, and which are based on the lower bound theory in local optimum solutions. In this paper, we also present a sufficient condition on λ for global minimizers of the l_p -regularization problems, which also have desirable sparsity, but which are based on the lower bound theory in global optimum solutions.

Theorem 4 Set

$$\beta(k) = k^{(p-2)/2}[\mu(1-p)]^{-p/2}\|b\|^{2-p}, \quad 1 \leq k \leq n. \tag{18}$$

The following conclusions hold.

1. If $\lambda \geq \beta(k)$, then any global minimizer s^* of the l_p -regularization problems (1) satisfies $\|s^*\|_0 < k$ for $1 \leq k \leq n$.
2. If $\lambda \geq \beta(1)$, then $s^* = 0$ is the unique global minimizer of the l_p -regularization problems (1).

Proof Assume that $s^* \neq 0$ is a global minimizer of the l_p -regularization problems (1). Let $B = A_T \in R^{m \times |T|}$, where $T = \text{support}(s^*)$ and $|T| = \|s^*\|_0$ is the cardinality of the set T . Therefore, according to the first order necessary condition, s^* must satisfy

$$2B^T(Bs_T^* - b) + \lambda p(|s_T^*|^{p-2} \cdot (s_T^*)) = 0, \tag{19}$$

which shows $As^* - b = Bs_T^* - b \neq 0$. Hence, we have

$$f_\lambda(s^*) = \|As^* - b\|^2 + \lambda \|s^*\|_p^p > \lambda \sum_{i \in T} |s_i^*|^p. \tag{20}$$

By Theorem 3, we can follow that

$$|s_i^*| \geq (\lambda\mu(1-p))^{1/(2-p)}, \quad i \in T.$$

Therefore, we have

$$f_\lambda(s^*) > \lambda|T|(\lambda\mu(1-p))^{p/(2-p)}. \tag{21}$$

In the following, we will discuss different cases:

1. Assume that $\lambda \geq \beta(k)$, we shall prove it through apagoge. If $\|s^*\|_0 \geq k \geq 1$, then by (3.11) and the definition of $\beta(k)$ in (3.8), we have

$$\begin{aligned} f_\lambda(s^*) &> \lambda|T|(\lambda\mu(1-p))^{p/(2-p)} = k\lambda^{2/(2-p)}(\mu(1-p))^{p/(2-p)} \\ &\geq kk^{-1}\|b\|^2 \\ &= \|b\|^2 = f_\lambda(0). \end{aligned}$$

This is in contradiction with that s^* is a global minimizer of (1). Therefore, we have $\|s^*\|_0 < k$.

2. Assume that $\lambda \geq \beta(1)$, we shall prove it through apagoge. If $s^* \neq 0$, then there exists i_0 satisfying $s_{i_0}^* \neq 0$ and

$$f_\lambda(s^*) = \|As^* - b\|^2 + \lambda\|s^*\|_p^p > \lambda|s_{i_0}^*|^p \geq \lambda(\lambda\mu(1-p))^{p/(2-p)} \geq \|b\|^2 = f_p(0).$$

This is in contradiction with that s^* is a global minimizer of (1). Therefore, $s^* = 0$ must be the unique global minimizer of (1). □

Iterative thresholding algorithm and its convergence

By the thresholding representation formula (16), an iterative thresholding formula of the problem (1) can be presented in the following: initialized $s^0 \in R^n$,

$$s^{k+1} = H_{\lambda\mu}(s^k + \mu A^T(b - As^k)), \tag{22}$$

where

$$h_{\lambda\mu}(r) := \begin{cases} h_{\lambda\mu,p}(r), & r > r^* \\ 0, & -r^* \leq r \leq r^* \\ -h_{\lambda\mu,p}(-r), & r < -r^* \end{cases} \tag{23}$$

When $|r| = r^*$, the adjustment here is, we only select $h_{\lambda\mu}(r) = 0$.

Firstly, some important lemmas are given in the following.

Lemma 4 *Let $0 < \mu < \|A\|^{-2}$ and $\{s^k\}$ be the sequence produced by the algorithm (22), then we can follow that the sequences $\{f_\lambda(s^k)\}_k$ and $\{f_\mu(s^{k+1}, s^k)\}_k$ are non-increasing.*

Proof For $0 < \mu < \|A\|^{-2}$, we have

$$\|s^{k+1} - s^k\|_2^2 - \mu\|As^{k+1} - As^k\|_2^2 \geq 0.$$

Hence,

$$\begin{aligned} f_\lambda(s^{k+1}) &\leq \mu^{-1}(\mu f_\lambda(s^{k+1}) + \|s^{k+1} - s^k\|_2^2 - \mu\|As^{k+1} - As^k\|_2^2) \\ &= \mu^{-1}f_\mu(s^{k+1}, s^k) \\ &\leq \mu^{-1}f_\mu(s^k, s^k) \\ &= f_\lambda(s^k) \\ &\leq \mu^{-1}(\mu f_\lambda(s^k) + \|s^k - s^{k-1}\|_2^2 - \mu\|As^k - As^{k-1}\|_2^2) \\ &= \mu^{-1}f_\mu(s^k, s^{k-1}). \end{aligned}$$

The first equality can be followed from the definition of $f_\mu(s, z)$. The second inequality is because that the s^{k+1} is the minimizer of $f_\mu(s, s^k)$. \square

This lemma demonstrate that, from iteration to iteration, the objective function $f_\lambda(s)$ does not increase, moreover, using the proposed algorithm does not lead to worse results than not using the proposed algorithm. The algorithm (22) does not have a unique fixed point, therefore it is very important to analyze the fixed points in detail.

Lemma 5 Let $\Gamma_0 = \{i : s_i^* = 0\}$ and $\Gamma_1 = \{i : |s_i^*| > (\lambda\mu(1 - p))^{1/(2-p)}\}$. The point s^* is a fixed point for the algorithm (18) if and only if

$$|A_i^T(b - As^*)| \leq \frac{2-p}{2} \lambda^{1/(2-p)} [\mu(1-p)]^{(p-1)/(2-p)}, \quad \text{if } i \in \Gamma_0,$$

$$s_i^* = h_{\lambda,\mu,p}(s_i^* + \mu A_i^T(b - As^*)), \quad \text{if } i \in \Gamma_1.$$

Proof A fixed point of the algorithm (22) is any s^* satisfying $s^* = H_{\lambda,\mu}(s^* + \mu A^T(b - As^*))$, i.e., $s_i^* = h_{\lambda,\mu}(s_i^* + \mu A_i^T(b - As^*))$. If $i \in \Gamma_0$, the equality holds when and only when $|\mu A_i^T(b - As^*)| \leq \frac{2-p}{2(1-p)} [\lambda\mu(1-p)]^{1/(2-p)}$, i.e., $|A_i^T(b - As^*)| \leq \frac{2-p}{2} \lambda^{1/(2-p)} [\mu(1-p)]^{(p-1)/(2-p)}$. Similarly, $i \in \Gamma_1$ when and only when $s_i^* = h_{\lambda,\mu,p}(s_i^* + \mu A_i^T(b - As^*))$. \square

The following lemma demonstrate that the sequence $\{s^k\}$ produced by the algorithm (22) is asymptotically regular, i.e., $\lim_{k \rightarrow \infty} \|s^{k+1} - s^k\|_2 = 0$.

Lemma 6 If $f_\lambda(s^0) < \infty$, $0 < \mu < \|A\|^{-2}$ and assume that $\{s^k\}$ be the sequence produced by the algorithm (22), $\forall \epsilon > 0$, $\exists K$ satisfying $\forall k > K$, $\|s^{k+1} - s^k\|_2^2 \leq \epsilon$.

Proof We prove the convergence of $\sum_{k=0}^K \|s^{k+1} - s^k\|_2^2$, which implies the lemma. First of all, we prove that $\sum_{k=0}^K \|s^{k+1} - s^k\|_2^2$ is monotonically increasing. We can follow monotonicity from

$$\begin{aligned} \sum_{k=0}^K \|s^{k+1} - s^k\|_2^2 &= \sum_{k=0}^{K-1} \|s^{k+1} - s^k\|_2^2 + \|s^{K+1} - s^K\|_2^2 \\ &\geq \sum_{k=0}^{K-1} \|s^{k+1} - s^k\|_2^2. \end{aligned}$$

Then, we will show the boundness of $\sum_{k=0}^K \|s^{k+1} - s^k\|_2^2$. For $0 < \mu < \|A\|^{-2}$, we have $0 < \delta := 1 - \mu\|A\|^2 < 1$ and

$$\|s^{k+1} - s^k\|_2^2 \leq \delta^{-1} (\|s^{k+1} - s^k\|_2^2 - \mu \|As^{k+1} - As^k\|_2^2).$$

Therefore,

$$\begin{aligned} \sum_{k=0}^K \|s^{k+1} - s^k\|_2^2 &\leq \delta^{-1} \sum_{k=0}^K (\|s^{k+1} - s^k\|_2^2 - \mu \|As^{k+1} - As^k\|_2^2) \\ &\leq \delta^{-1} \sum_{k=0}^K \mu (f_\lambda(s^k) - f_\lambda(s^{k+1})) \\ &= \mu \delta^{-1} (f_\lambda(s^0) - f_\lambda(s^{K+1})) \\ &\leq \mu \delta^{-1} f_\lambda(s^0) < \infty. \end{aligned}$$

The second inequality can be followed from the proof of Lemma 4 and the last inequality can be taken from $f_\lambda(s^0) < \infty$. □

In the following, we present an very important property of the algorithm, i.e., any accumulation point of the sequence $\{s^k\}$ is a fixed point of the algorithm (22). Therefore, we have the following theorem and conclusion.

Theorem 5 *If $f_\lambda(s^0) < \infty$ and $0 < \mu < \|A\|^{-2}$, then we have the following conclusion: any accumulation point of the sequence $\{s^k\}$ produced by the algorithm (22) is a fixed point of (22).*

Proof In Lemma 6, we take $\epsilon < \lambda$. If $|s_i^k| > (\lambda\mu(1-p))^{1/(2-p)}$ and $s_i^{k+1} = 0$, then we have $\|s^{k+1} - s^k\|_2^2 \geq \lambda$, by Lemma 6 which is impossible for $k > K$ for some K . Therefore, for large K , the set of zero and non-zero coefficients will not change and $|s_i^k| > (\lambda\mu(1-p))^{1/(2-p)}$, $\forall i \in \Gamma_1, k > K$. Assume that $\{s^{k_j}\}$ be a convergent subsequence and s^* be its limit point, i.e.,

$$s^{k_j} \rightarrow s^*, \quad \text{as } k_j \rightarrow +\infty. \tag{24}$$

By the limitation (24) and Lemma 6, we have

$$\|s^{k_j+1} - s^*\|_2 \leq \|s^{k_j+1} - s^{k_j}\|_2 + \|s^{k_j} - s^*\|_2 \rightarrow 0, \quad \text{as } k_j \rightarrow +\infty,$$

which implies that the sequence $\{s^{k_j+1}\}$ is also convergent to s^* . Note that $s^{k_j+1} = H_{\lambda\mu}(B_\mu(s^{k_j}))$, i.e., $s_i^{k_j+1} = h_{\lambda\mu}(s_i^{k_j} + \mu A_i^T(b - As^{k_j}))$, for all $i = 1, 2, \dots, n$.

Let $\Gamma_0 = \{i : s_i^* = 0\}$ and $\Gamma_1 = \{i : s_i^* \neq 0\}$. For $s^{k_j}, k_j > K$ for some K , if $i \in \Gamma_0$, then by (23) and (7) we have

$$|A_i^T(b - As^{k_j})| \leq \frac{2-p}{2} \lambda^{1/(2-p)} [\mu(1-p)]^{(p-1)/(2-p)},$$

therefore, $|A_i^T(b - As^*)| \leq \frac{2-p}{2} \lambda^{1/(2-p)} [\mu(1-p)]^{(p-1)/(2-p)}$. Similarly, if $i \in \Gamma_1$, then by (23) and (7) we have

$$s_i^{k_j+1} = h_{\lambda\mu}(s_i^{k_j} + \mu A_i^T(b - As^{k_j})), \quad |s_i^{k_j} + \mu A_i^T(b - As^{k_j})| > r^*,$$

where $r^* := \frac{2-p}{2(1-p)} [\lambda\mu(1-p)]^{1/(2-p)}$. By Proposition 2, we can follow that the function $h_{\lambda\mu,p}(\cdot)$ is continuous over $(r^*, +\infty)$ and $(-\infty, r^*)$. Therefore, we follow that $s_i^* = h_{\lambda\mu,p}(s_i^* + \mu A_i^T(b - As^*))$. By Lemma 5, s^* is a fixed point of (22). □

Numerical experiments

Now we report numerical results to compare the performance of Iterative thresholding algorithm (ITA) ($p = 0.5$) for solving (1) (Signal reconstruction) with LASSO to find sparse solutions. The computational test was conducted on a Intel(R) Core(TM)2 Duo CPU E 8400 @3.00GHZ Dell desktop computer with 2.0GHz of memory with using Matlab R2010A.

Table 1 Comparison of ITA and LASSO algorithm

Problems			LASSO		ITA	
<i>n</i>	<i>T</i>	<i>m</i>	Time	Error	Time	Error
800	60	150	0.572	4.16e−4	0.375	1.15e−4
800	80	180	0.461	3.58e−4	0.252	1.06e−4
2000	160	300	0.853	5.75e−4	0.516	1.62e−4
2000	200	500	0.853	5.86e−4	0.553	1.73e−4

Consider a real-valued, finite-length signal $x \in R^n$. Suppose x is T -sparse, that is, only T of the signal coefficients are nonzero and the others are zero. We use the following Matlab code to generate the original signal, a matrix A and a vector b .

$$x_{or} = \text{zeros}(n, 1); q = \text{randperm}(n); x_{or}(q(1 : T)) = 2 * \text{randn}(T, 1);$$

$$A = \text{randn}(m, n); A = \text{orth}(A)'; b = A * x_{or};$$

The computational results for this experiment are displayed in Table 1.

From Table 1 we find that ITA has smaller prediction accuracy than LASSO in shorter time.

Conclusion

In this paper, an exact lower bound for the absolute value of nonzero entries in each global optimum solution of the problem (1) is established. And the necessary condition for global optimum solutions of the l_p -regularization problems is derived, i.e., the global optimum solutions are the fixed points of a vector thresholding operator. In addition, we have derived a sufficient condition on the selection of λ for the desired sparsity of global minimizers of the problem (1) with the given (A, b, p) . Finally, an iterative thresholding algorithm is designed for solving the l_p -regularization problems, and the convergence of algorithm is proved.

Authors' contributions

All authors are common first author, all authors contribute equally to the manuscript. All authors have a good contribution to derive the exact lower bounds, to establish the global optimum condition and to design the iterative thresholding algorithm, and to perform the numerical experiments of this research work. All authors read and approved the final manuscript.

Author details

¹ School of Mathematical Sciences, Henan Institute of Science and Technology, Xinxiang 453003, China. ² College of Mathematics and Information Science, Henan Normal University, Xinxiang 453007, China.

Acknowledgements

This paper is supported by the National Natural Science Foundation of China under Grant (11171094), the Natural Science Foundation of Henan Province (152300410097), the Key Scientific Research Project of Universities in Henan Province (14A110024), (16A110014) and (15A110022), the Major Scientific Research Projects of Henan Institute of Science and Technology (2015ZD07), the High-level Scientific Research Personnel Project for Henan Institute of Science and Technology (2015037), the Science and Technology Innovation Project for Henan Institute of Science and Technology.

Competing interests

The authors declare that they have no competing interests.

Received: 22 May 2016 Accepted: 11 October 2016

Published online: 26 October 2016

References

- Bredies K, Lorenz DA, Reiterer S (2015) Minimization of non-smooth, non-convex functionals by iterative thresholding. *J Optim Theory Appl* 165:78–112
- Candès EJ, Wakin MB, Boyd SP (2008) Enhancing sparsity by reweighted l_1 minimization. *J Fourier Anal Appl* 14:877–905
- Chartrand R (2007a) Nonconvex regularization for shape preservation. In: Proceedings of IEEE international conference on image processing
- Chartrand R (2007b) Exact reconstruction of sparse signals via nonconvex minimization. *IEEE Signal Process Lett* 14:707–710
- Chartrand R, Staneva V (2008) Restricted isometry properties and nonconvex compressive sensing. *Inverse Probl* 24:1–14
- Chartrand R, Yin W (2008) Iteratively reweighted algorithms for compressive sensing. In: Proceedings of international conference on acoustics, speech, signal processing (ICASSP)
- Chen X, Xu F, Ye Y (2010) Lower bound theory of nonzero entries in solutions of l_2 - l_p minimization. *SIAM J Sci Comput* 32:2832–2852
- Chen X, Ge D, Wang Z, Ye Y (2014) Complexity of unconstrained $l_2 - l_p$ minimization. *Math Program* 143:371–383
- Chen YQ, Xiu NH, Peng DT (2014) Global solutions of non-lipschitz $s_2 - s_p$ minimization over the semidefinite cones. *Optim Lett* 8(7):2053–2064
- Fan Q, Wu W, Zurada JM (2016) Convergence of batch gradient learning with smoothing regularization and adaptive momentum for neural networks. *SpringerPlus* 2016(5):1–17
- Fan J, Li R (2001) Variable selection via nonconcave penalized likelihood and its oracle properties. *J Am Stat Ass* 96:1348–1360
- Foucart S, Lai MJ (2009) Sparsest solutions of under-determined linear systems via l_q minimization for $0 < q \leq 1$. *Appl Comput Harmonic Anal* 26:395–407
- Frank IE, Freidman JH (1993) A statistical view of some chemometrics regression tools (with discussion). *Technometrics* 35:109–148
- Ge D, Jiang X, Ye Y (2011) A note on the complexity of l_p minimization. *Math Program* 129:285–299
- Huang J, Horowitz JL, Ma S (2008) Asymptotic properties of bridge estimators in sparse high-dimensional regression models. *Ann Stat* 36:587–613
- Knight K, Wu JF (2000) Asymptotics for lasso-type estimators. *Ann Stat* 28:1356–1378
- Lai M, Wang Y (2011) An unconstrained l_q minimization with $0 < q < 1$ for sparse solution of under-determined linear systems. *SIAM J Optim* 21:82–101
- Natarajan BK (1995) Sparse approximate solutions to linear systems. *SIAM J Comput* 24:227–234
- Shehu Y, Iyiola OS, Enyi CD (2013) Iterative approximation of solutions for constrained convex minimization problem. *Arab J Math* 2:393–402
- Shehu Y, Cai G, Iyiola OS (2015) Iterative approximation of solutions for proximal split feasibility problems. *Fixed Point Theory Appl* 2015(123):1–18
- Tian M, Huang LH (2013) Iterative methods for constrained convex minimization problem in hilbert spaces. *Fixed Point Theory Appl* 2013(105):1–18
- Tian M, Jiao S-W (2015) Regularized gradient-projection methods for the constrained convex minimization problem and the zero points of maximal monotone operator. *Fixed Point Theory Appl* 11:1–23
- Xu Z, Zhang H, Wang Y, Chang X (2010) $l_1/2$ regularizer. *Sci China Inf Sci* 53:1159–1169
- Xu Z, Chang X, Xu F, Zhang H (2012) $l_1/2$ regularization: a thresholding representation theory and a fast solver. *IEEE Trans Neural Netw Learn Syst* 23:1013–1027

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com