Optim Lett (2014) 8:2053–2064
DOI 10.1007/s11590-013-0701-y

ORIGINAL PAPER

Global solutions of non-Lipschitz S_2–S_p minimization over the positive semidefinite cone

Yongqiang Chen · NaihuaXiu · Dingtao Peng

Received: 3 June 2013 / Accepted: 17 October 2013 / Published online: 14 December 2013

Abstract The S_2–S_p minimization over the positive semidefinite cone is the semidefinite least squares problem with Schatten p-quasi ($0 < p < 1$) norm regularization term. It has wide applications in many areas including compressed sensing, control, statistics, signal and image processing, etc. In this paper, by developing the symmetric matrix p-thresholding operator representation theory, we establish the necessary condition for global optimal solutions of S_2–S_p minimization, and also provide the exact lower bound for the positive eigenvalues at global optimal solutions.

Keywords S_2–S_p minimization · Positive semidefinite cone · Thresholding operator · Global solution · Fixed point

1 Introduction

We consider the following non-Lipschitz S_2–S_p minimization problem

$$\begin{aligned}&\min_{X\in\mathbb{S}^n}\ f_\nu(X) := \|\mathcal{A}(X) - B\|_F^2 + \nu\|X\|_{S_p}^p\\ &\text{s.t.}\quad X \succcurlyeq 0,\end{aligned} \tag{1.1}$$

Y. Chen (✉) · N. Xiu · D. Peng
Department of Mathematics, School of Science, Beijing Jiaotong University,
Beijing 100044, People's Republic of China
e-mail: yqchenchn@gmail.com

N. Xiu
e-mail: nhxiu@bjtu.edu.cn

D. Peng
e-mail: dingtaopeng@126.com

Y. Chen
College of Mathematics and Information Science, Henan Normal University,
Xinxiang 453007, China

where $X \in \mathbb{S}^n$ is the matrix variable, $\mathcal{A}$ is a linear transformation from $\mathbb{S}^n$ to $\mathbb{R}^{m_1 \times m_2}$ which is given by

$$[\mathcal{A}(X)]_{ij} := \langle A_{ij}, X\rangle, \ i = 1, 2, \cdots, m_1, \ j = 1, 2, \cdots, m_2,$$

with $A_{ij} \in \mathbb{S}^n (i = 1, 2, \cdots, m_1, j = 1, 2, \cdots, m_2)$, $B \in \mathbb{R}^{m_1 \times m_2}$ and $\nu > 0$. $X \succeq 0$ says that X is positive semidefinite, i.e., $X \in \mathbb{S}^n_+$. The function $\|Z\|_{S_p}^p (0 < p < 1)$ is Schatten p-quasi norm and defined as $\|Z\|_{S_p}^p := \sum_{i=1}^m \sigma_i^p(Z)$, where $\sigma_i(Z)(i = 1, 2, \cdots, m)$ are the singular values of $Z \in \mathbb{R}^{m\times n}(m \leq n)$. Obviously, as $p \downarrow 0$, we have $\|Z\|_{S_p}^p \to \text{rank(Z)}$; as $p \uparrow 1$, $\|Z\|_{S_p}^p \to \|Z\|_* = \sum_{i=1}^m \sigma_i(Z)$. When X is diagonal, $m_1 = m$ and $m_2 = 1$, problem (1.1) reduces l_2-l_p minimization

$$\min_{x\in\mathbb{R}^n} \{\|Ax - b\|_2^2 + \nu\|x\|_p^p \mid x \geq 0\}. \tag{1.2}$$

The model (1.1) has wide applications in many areas including compressed sensing, control, statistics, signal and image processing, etc. To mention a few: the low-rank positive semidefinite matrix completion problem [4], the low-dimensional Euclidean distance matrix completion problem [10], the phase retrieval problem [2,3], the low-rank correlation matrix [15], and the sensor network localization [8].

Since problem (1.1) is nonconvex, nonsmooth and non-Lipschitz, most algorithms for solving the problem can only provide an approximate local optimal solution. Lu, Zhang and Wu [12] presented the lower bound for the nonzero singular values at local optimal solutions of problem (1.1) without constraint $X \succeq 0$. In the vector case, the problem (1.2) is of great practical interest since nonnegative data are naturally used in image processing, DNA microarrays, network monitoring, hidden Markov models, compressed sensing, and so on [9,11]. Problem (1.2) without constraint $x \geq 0$ is extensively studied in recent years, for example, Chen, Xu and Ye [7] gave the lower bound estimates of nonzero entries at local solutions, Chen, Ge, Wang and Ye [6] showed that the problem is strongly NP-hard, Chen [5] introduced the smoothing technique to tackle the nonconvex, non-Lipschiz regularization term $\|x\|_p^p$ and gave a SQP-type algorithm. However, there is no result for the global optimal solutions of problem (1.1).

In this paper, motivated by [14,16], we establish the necessary condition for global optimal solutions of S_2–S_p minimization, i.e., the global optimal solutions of problem (1.1) are fixed points of a symmetric matrix p-thresholding operator (see Theorem 3.8). We also provide the exact lower bound for the positive eigenvalues at global optimal solutions (see Theorem 3.8), which can be used to identify zero eigenvalues precisely at any global numerical solution. These lower bounds clearly show the relationship among the lowest rank of the global solution, the regularization parameter and p norm, so that our theorem can be used for selecting desired model parameters and norms.

This paper is organized as follows. In Sect. 2, we give some lemmas and propositions. In Sect. 3, we obtain the main results of this paper. Concluding remarks are made in Sect. 4.

2 Some Lemmas

By using the separability of the objective function and the technique of operator splitting, S_2–S_p minimization problem (1.1) can be converted into n corresponding single variable minimization problems on $[0, +\infty)$. Hence, in this section we research the corresponding single variable minimization problem

$$\min_{x \geq 0} g_t(x) := x^2 - 2tx + \nu x^p, \tag{2.1}$$

where $\nu > 0$ *and* $0 < p < 1$ are any given real numbers, $x \in \mathbb{R}$ is variable and $t \in \mathbb{R}$ is a parameter. Denote

$$\bar{t} := \frac{2-p}{1-p}[\nu p(1-p)/2]^{1/(2-p)} > 0, \tag{2.2}$$

$$\bar{x} := [\nu p(1-p)/2]^{1/(2-p)} > 0. \tag{2.3}$$

We first discuss the positive stationary point of $g_t(x)$. The relationship between the positive stationary point and parameter t is showed in the following lemma.

Lemma 2.1 *Let $g_t(x), \bar{t}, \ \bar{x}$ be defined by (2.1)–(2.3), there hold:*

(i) *when $t < \bar{t}$, $g_t(x)$ has no positive stationary point.*
(ii) *when $t = \bar{t}$, $g_t(x)$ has unique positive stationary point $\bar{x}$ and $g_{\bar{t}}(\bar{x}) > 0$.*
(iii) *when $t > \bar{t}$, $g_t(x)$ must have two different positive stationary points $\hat{x}_1 < \hat{x}_2$ and $\hat{x}_1 < \bar{x} < \hat{x}_2$.*

Proof From (2.1) and for any $x > 0$, we have

$$[g_t(x)]' = 2x - 2t + \nu p x^{p-1}, \tag{2.4}$$

$$[g_t(x)]'' = 2 + \nu p(p-1)x^{p-2}, \tag{2.5}$$

$$[g_t(x)]''' = \nu p(p-1)(p-2)x^{p-3}. \tag{2.6}$$

By (2.6), we have $[g_t(x)]''' > 0$ for any $x > 0$, so the function $[g_t(x)]'$ is convex on $(0, +\infty)$. The function $[g_t(x)]'$ is continuously differentiable on $(0, +\infty)$ by (2.4), and $\bar{x}$ defined by (2.3) is the unique positive root of the equation $[g_t(x)]'' = 0$, hence, $\bar{x}$ is the unique minimizer of $[g_t(x)]'$ on $(0, +\infty)$. From the definition of $\bar{t}$ and (2.4), we have $[g_{\bar{t}}(\bar{x})]' = 0$. By direct computation, we get

$$\begin{aligned} g_{\bar{t}}(\bar{x}) &= [\nu p(1-p)/2]^{2/(2-p)}[(p-3)/(1-p) + \nu[\nu p(1-p)/2]^{-1}] \\ &= (2/p - 1)[\nu p(1-p)/2]^{2/(2-p)} > 0, \end{aligned}$$

the last inequality is due to $0 < p < 1$. When $t < \bar{t}$, $[g_t(x)]' = 0$ has no root on $(0, +\infty)$ by (2.4). Hence, (i) and (ii) hold.

Note that $[g_t(x)]' \to +\infty$ as $x \to 0^+$ and $[g_t(x)]' \to +\infty$ as $x \to +\infty$. When $t > \bar{t}$, $[g_t(\bar{x})]' < 0$ by (2.4), together with the continuity of $[g_t(x)]'$, the equation $[g_t(x)]^{'} = 0$ must have two different positive roots $\hat{x}_1$, $\hat{x}_2$. Without loss of generality, assume $\hat{x}_1 < \hat{x}_2$, then $\hat{x}_1 < \bar{x} < \hat{x}_2$, i.e., (iii) is right. The whole proof is completed. □

The following lemma will show that there is the unique local minimizer of $g_t(x)$ on $(0, +\infty)$, which is the function in variable t on $(\bar{t}, +\infty)$.

Lemma 2.2 *Define* $G(x, t) := [g_t(x)]^{'} = 2x - 2t + \nu p x^{p-1}$. *For any given* $t_0 > \bar{t}$, *let* x_0 $(x_0 > \bar{x})$ *be the positive root of the equation* $G(x, t_0) = 0$, *where* $\bar{t}$, $\bar{x}$ *are defined by* (2.2) *and* (2.3). *Then, there exists a unique implicit function* $x = h_{\nu,p}(t)$ *on* $(\bar{t}, +\infty)$ *such that* $x_0 = h_{\nu,p}(t_0)$, $h_{\nu,p}(t) > \bar{x}$ *and* $G(h_{\nu,p}(t), t) \equiv 0$ *for any* $t \in (\bar{t}, +\infty)$. *Furthermore, for the function* $x = h_{\nu,p}(t)$, *there hold:*

(i) *The function* $x = h_{\nu,p}(t)$ *is continuous on* $(\bar{t}, +\infty)$.
(ii) *The function* $x = h_{\nu,p}(t)$ *is differentiable on* $(\bar{t}, +\infty)$ *and* $h^{'}_{\nu,p}(t) = \frac{2}{2+\nu p(p-1)h^{p-2}_{\nu,p}(t)}$.
(iii) *The function* $x = h_{\nu,p}(t)$ *is strictly increasing on* $(\bar{t}, +\infty)$.

Moreover, if $t > \bar{t}$, $x = h_{\nu,p}(t)$ *is the unique local minimizer of* $g_t(x)$ *on* $(0, +\infty)$.

Proof We firstly show that $G^{'}_x(x_0, t_0) \neq 0$. Since $G(x, t) = 2x - 2t + \nu p x^{p-1}$, we have

$$G^{'}_x(x, t) = 2 + \nu p(p - 1)x^{p-2},$$

and $\bar{x}$ defined by (2.3) is the unique positive root of the equation $G^{'}_x(x, t) = 0$. From $x_0 > \bar{x} > 0$, we have $G^{'}_x(x_0, t_0) \neq 0$. Obviously, the function G is continuous on the region $D := (\bar{t}, +\infty) \times (\bar{x}, +\infty) \subset \mathbb{R}^2$. On the other hand, the partial derivative function $G^{'}_x(x, t)$ is continuous on D. Note that $G(x_0, t_0) = 0$. By the implicit function theorem, we obtain that the existence and uniqueness of the implicit function $x = h_{\nu,p}(t)$ and that (i) is right. Additionally, $G^{'}_t(x, t)$ is continuous on D, hence the function $x = h_{\nu,p}(t)$ is differentiable on $(\bar{t}, +\infty)$ and

$$h^{'}_{\nu,p}(t) = -\frac{G^{'}_t(x, t)}{G^{'}_x(x, t)} = \frac{2}{2 + \nu p(p - 1)h^{p-2}_{\nu,p}(t)},$$

i.e., (ii) holds. The function $G^{'}_x(x, t)$ is strictly increasing in x on $(0, +\infty)$ by (2.6), together with $h_{\nu,p}(t) > \bar{x}$, then we have

$$G^{'}_x(h_{\nu,p}(t), t) = 2 + \nu p(p - 1)h^{p-2}_{\nu,p}(t) > G^{'}_x(\bar{x}, t) = 0,$$

then (iii) is correct. By (2.6), $[g_t(x)]^{''}$ is strictly increasing on $(0, +\infty)$. Note that $\bar{x}$ is the unique positive root of the equation $[g_t(x)]^{''} = 0$ by (2.3) and (2.5). From $h_{\nu,p}(t) > \bar{x}$ for any $t \in (\bar{t}, +\infty)$, we have $[g_t(x)]^{''}|_{x=h_{\nu,p}(t)} > 0$, then $h_{\nu,p}(t)$ is a local minimizer of $g_t(x)$ on $(0, +\infty)$. By the (iii) of Lemma 2.1, the uniqueness is right. Thus, the proof is completed. □

Lemma 2.3 *The function* $\hat{g}(t) := g_t(h_{\nu,p}(t))$ *is strictly decreasing on* $(\bar{t}, +\infty)$.

Proof From (2.1), we have

$$\hat{g}(t) = g_t(h_{\nu,p}(t)) = h_{\nu,p}^2(t) - 2t h_{\nu,p}(t) + \nu[h_{\nu,p}(t)]^p.$$

By Lemma 2.2, we know that $h_{\nu,p}(t)$ is differentiable on $(\bar{t}, +\infty)$ and that $h_{\nu,p}(t) > 0$. Additionally, the function $g_t(x)$ is differentiable on $(0, +\infty)$, hence we have

$$\begin{aligned} \hat{g}(t)' &= 2h_{\nu,p}(t)h'_{\nu,p}(t) - 2h_{\nu,p}(t) - 2t h'_{\nu,p}(t) + \nu p[h_{\nu,p}(t)]^{p-1}h'_{\nu,p}(t) \\ &= h'_{\nu,p}(t)[2h_{\nu,p}(t) - 2t + \nu p(h_{\nu,p}(t))^{p-1}] - 2h_{\nu,p}(t) \\ &= -2h_{\nu,p}(t) < 0. \end{aligned}$$

where the last equality is due to Lemma 2.2 which shows that $h_{\nu,p}(t)$ is a root of $[g_t(x)]' = 0$. The proof is thus completed. □

By applying Lemmas 2.1–2.3, we can obtain the global minimizer of problem (2.1) in the following proposition.

Proposition 2.4 *Let* x^* *be the global solution of the problem* (2.1). *Then*

$$x^* = h_\nu(t) := \begin{cases} h_{\nu,p}(t), & t > t^* \\ (\nu(1-p))^{1/(2-p)} \text{ or } 0, & t = t^* \\ 0, & t < t^* \end{cases} \tag{2.7}$$

where $t^* := \frac{2-p}{2(1-p)}[\nu(1-p)]^{1/(2-p)}$, $h_{\nu,p}(t)$ *is defined in Lemma* 2.2.

Proof If $t \le 0$, then $g_t(x)$ is monotonically increasing on $[0, +\infty)$ and $x^* = 0$. If $0 < t \le \bar{t}$, by Lemma 2.1 we have $x^* = 0$.

From now on, we consider only $t > \bar{t}$. By (i) of Lemma 2.2, we know that $x = h_{\nu,p}(t)$ is the unique local minimizer of $g_t(x)$ on $(0, +\infty)$. Therefore, the remainder thing is to compare the values between $g_t(h_{\nu,p}(t))$ and $g_t(0)$.

From the definition of t^*, it is easy to check $t^* > \bar{t}$. We shall firstly show that $\tilde{x} := [\nu(1-p)]^{1/(2-p)}$ is the local minimizer of $g_{t^*}(x)$ on $(0, +\infty)$. From $t^* = \frac{2-p}{2(1-p)}[\nu(1-p)]^{1/(2-p)}$ as well as (2.4) and (2.5), we have

$$\begin{aligned} [g_{t^*}(\tilde{x})]' &= [\nu(1-p)]^{1/(2-p)} + \nu(p-1)[\nu(1-p)]^{(p-1)/(2-p)} = 0, \\ [g_{t^*}(\tilde{x})]'' &= 2 + \nu p(p-1)[\nu(1-p)]^{-1} = 2 - p > 0, \end{aligned}$$

i.e., $\tilde{x}$ is the local minimizer of $g_{t^*}(x)$ on $[0, +\infty)$. By direct computation, we have

$$g_{t^*}(\tilde{x}) = [\nu(1-p)]^{2/(2-p)}[-\frac{1}{1-p} + \nu[\nu(1-p)]^{-1}] = 0 = g_{t^*}(0).$$

hence, when $t = t^*, \tilde{x}$ and 0 are both global minimizers of (2.1), that is, $x^* = (\nu(1-p))^{1/(2-p)}$ or 0.

When $t > t^*, h_{\nu,p}(t) > \tilde{x}$ by (iii) of Lemma 2.2, furthermore we have $g_t(h_{\nu,p}(t)) < g_{t^*}(\tilde{x}) = g_{t^*}(0)$ by Lemma 2.3, so $x^* = h_{\nu,p}(t)$. When $\bar{t} < t < t^*$, similarly we have $g_t(h_{\nu,p}(t)) > g_{t^*}(\tilde{x}) = g_{t^*}(0)$, so $x^* = 0$. The proof is thus complete. □

When $p = 1/2$, $h_{\nu,p}(t)$ of (2.7) has the analytic expression in the following corollary (Xu et al. [16]).

Corollary 2.5 *When $p = 1/2$, the global solution x^* of problem* (2.1) *has analytic expression*

$$x^* = h_\nu(t) := \begin{cases} h_{\nu,1/2}(t), & t > t^* \\ (\nu/2)^{2/3} \ \text{or} \ 0, & t = t^* \\ 0, & t < t^* \end{cases} \tag{2.8}$$

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

Proof A brief proof is given here for completeness. When $p = 1/2$, we have $t^* = \frac{\sqrt[3]{54}}{4}\nu^{2/3}$. When $t > t^* = \frac{\sqrt[3]{54}}{4}\nu^{2/3}$, $x^* = h_{\nu,1/2}(t)$ according to Proposition 2.4, and $h_{\nu,1/2}(t)$ is the root of the equation

$$x - t + \frac{\nu}{4\sqrt{x}} = 0,$$

which is due to the first order optimal condition of (2.1). By Theorem 1 in [16], we have

$$h_{\nu,1/2}(t) = \frac{2}{3}t\left(1+\cos\left(\frac{2\pi}{3} - \frac{2\varphi(t)}{3}\right)\right),\ \varphi(t) = \arccos\left(\frac{\nu}{8}\left(\frac{t}{3}\right)^{-3/2}\right).$$

The proof is complete. □

3 Main results

In this section, we firstly define p-thresholding function, vector p-thresholding operator and symmetric matrix p-thresholding operator. Then, by using the separability of the function and the technique of operator splitting, we develop the symmetric matrix p-thresholding operator representation theorem. And then, we portray the characteristics of the global optimal solutions of S_2–S_p minimization problem (1.1).

Definition 3.1 (p-thresholding function) Suppose $t \in \mathbb{R}$, for any $\nu > 0$, the function $h_\nu(t)$ defined by (2.7) is called a p-thresholding function.

Definition 3.2 (Vector p-thresholding operator) Suppose $x \in \mathbb{R}^n$, for any $\nu > 0$, the vector p-thresholding operator $H_\nu(x)$ is defined as

$$H_\nu(x) := (h_\nu(x_1), h_\nu(x_2), \cdots, h_\nu(x_n))^T.$$

For simplicity, let $\mathcal{O}^n$ denote the set of all n dimensional orthogonal matrices.

Definition 3.3 (Symmetric matrix p-thresholding operator) Suppose $Z \in \mathbb{S}^n$ and the eigenvalue decomposition (ED) of the Z is

$$Z = Q\mathbf{Diag}(\lambda)Q^T,$$

where $Q \in \mathcal{O}^n$ and the vector $\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_n)^T$ is the eigenvalues vector of Z sorted by nonincreasing order. For any $\nu > 0$, the symmetric matrix p-thresholding operator $\mathcal{H}_\nu(Z) : \mathbb{S}^n \to \mathbb{S}^n$ *is defined as*

$$\mathcal{H}_\nu(Z) := Q\mathbf{Diag}(H_\nu(\lambda))Q^T.$$

The following theorem plays an important role in the proof of our main results.

Theorem 3.4 *(*p*-thresholding operator representation) Given a matrix $Y \in \mathbb{S}^n$ and constants $\nu > 0,\ 0 < p < 1$, let X^* be the global solution of the following problem*

$$\min_{X \succeq 0} \{\|X - Y\|_F^2 + \nu\|X\|_{S_p}^p\}. \tag{3.1}$$

Then

$$X^* = \mathcal{H}_\nu(Y).$$

Proof Following the idea used in the proof of [14, Theorem 1], let the eigenvalue decomposition of any $X \in \mathbb{S}_+^n$ be given by $X = \bar{Q}\mathbf{Diag}(\bar{\lambda})\bar{Q}^T = \sum_{i=1}^n \bar{\lambda}_i \bar{q}_i \bar{q}_i^T$, where $\bar{Q} = [\bar{q}_1, \bar{q}_2, \cdots, \bar{q}_n],\ \bar{q}_i \in \mathbb{R}^n,\ \bar{\lambda}_1 \ge \bar{\lambda}_2 \ge \cdots \ge \bar{\lambda}_n \ge 0,\ i = 1, 2, \cdots, n$. Denote by $t_i = \bar{q}_i^T Y \bar{q}_i$ for each $i = 1, 2, \cdots, n$. Note that $\langle X, Y\rangle = \langle \sum_{i=1}^n \bar{\lambda}_i \bar{q}_i \bar{q}_i^T, Y\rangle = \sum_{i=1}^n \bar{\lambda}_i \langle \bar{q}_i \bar{q}_i^T, Y\rangle = \sum_{i=1}^n t_i \bar{\lambda}_i$. Then $\|X - Y\|_F^2 + \nu\|X\|_{S_p}^p = \sum_{i=1}^n ((\bar{\lambda}_i)^2 - 2t_i\bar{\lambda}_i + \nu\bar{\lambda}_i^p) + \|Y\|_F^2$. So, problem (3.1) is equivalent to

$$\min_{X \succeq 0} \{\|X\|_F^2 - 2\langle X, Y\rangle + \nu\|X\|_{S_p}^p\}. \tag{3.2}$$

Let $G(\bar{Q}) = \min\{\sum_{i=1}^n ((\bar{\lambda}_i)^2 - 2t_i\bar{\lambda}_i + \nu\bar{\lambda}_i^p) |\ \bar{\lambda}_i \ge 0, i = 1, 2, \cdots, n\}$, then problem (3.2) is equivalent to

$$\min_{\bar{Q}} \{G(\bar{Q}) \mid \bar{Q}^T\bar{Q} = I_n\}. \tag{3.3}$$

Let $g_{t_i}(\bar{\lambda}_i) := (\bar{\lambda}_i)^2 - 2t_i\bar{\lambda}_i + \nu\bar{\lambda}_i^p$, then

$$G(\bar{Q}) = \min\left\{\sum_{i=1}^{n} g_{t_i}(\bar{\lambda}_i) \mid \bar{\lambda}_i \geq 0, i = 1, 2, \cdots, n\right\}. \tag{3.4}$$

Fixing $\bar{Q}$, note that $\sum_{i=1}^{n} g_{t_i}(\bar{\lambda}_i)$ is separable as to $(\bar{\lambda}_1, \bar{\lambda}_2, \cdots, \bar{\lambda}_n)$. So, solving problem (3.4) is equivalent to solving the following n problems, for each $i = 1, 2, \cdots, n$,

$$\min_{\bar{\lambda}_i \geq 0} \; g_{t_i}(\bar{\lambda}_i).$$

By Proposition 2.4, for each $i = 1, 2, \cdots, n$, we can obtain $\lambda_i^* = \operatorname{argmin}_{\bar{\lambda}_i \geq 0} g_{t_i}(\bar{\lambda}_i) = h_\nu(t_i)$. Hence

$$g_{t_i}(\lambda_i^*) = h_\nu^2(t_i) - 2t_i h_\nu(t_i) + \nu \, h_\nu^p(t_i).$$

According to (2.7), when $t_i \leq t^*$, $g_{t_i}(\lambda_i^*) = 0$; When $t_i > t^*$, $\lambda_i^* > 0$, $g_{t_i}(\lambda_i^*)$ is strictly decreasing in the variable $t_i = \bar{q}_i^T Y \bar{q}_i$ by Lemma 2.3.

Let Y admit the eigenvalue decomposition $Y = QDiag(\lambda)Q^T$, $Q \in \mathcal{O}^n$ and $\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_n)^T$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. By (3.4), solving (3.3) is equivalent to solving the following n problems one by one from $i = 1$ to $i = n$.

$$\begin{aligned} \max_{\bar{q}_i} \quad & \sum_{j=1}^{n} \lambda_j (\bar{q}_i^T q_j)^2 \\ \text{s.t.} \quad & \|\bar{q}_i\|_2 = 1, \\ & \bar{q}_i \perp \{q_1^*, q_2^*, \cdots, q_{i-1}^*\}, \end{aligned}$$

where $Q = [q_1, q_2, \cdots, q_n]$, $q_j \in \mathbb{R}^n$, $j = 1, 2, \cdots, n$. The optimal solution is $q_i^* = q_i$, the optimal objective function value is $t_i^* = q_i^T Y q_i = \lambda_i$. Hence,

$$X^* = \sum_{j=1}^{n} \lambda_i^* q_i^* q_i^{*T} = \sum_{j=1}^{n} h_\nu(t_i^*) q_i q_i{}^T = \sum_{j=1}^{n} h_\nu(\lambda_i) q_i q_i{}^T = \mathcal{H}_\nu(Y).$$

The proof is thus complete. □

Remark 3.1 Naturally, one expects that using the S_p quasi-norm regularization term can find lower rank solution than using the S_1 norm. The adjustment here is, when $t = t^*$, we only choose $h_\nu(t) = 0$ in (2.7) i.e.,

$$h_\nu(t) := \begin{cases} h_{\nu,p}(t), & t > t^* \\ 0, & \text{otherwise}, \end{cases} \tag{3.5}$$

where t^* is defined in Proportion 2.4. Thus, the lowest-rank global solution of (3.1) is uniquely defined by $X^{**} = \mathcal{H}_\nu(Y)$, where $h_\nu(\cdot)$ is defined by (3.5), and rank $(X**)$ is the number of positive eigenvalues of Y which are greater than t^*.

From Theorem 3.4, Corollary 2.5 and (3.5), we obtain the following corollary, which is Theorem 1 of Rao et al. [14]. Their numerical experiments showed that the alternating direction method based on the analytic expression performs very well in solving the sparse and low-rank matrix decomposition problem.

Corollary 3.5 *When $p = 1/2$, the unique lowest-rank global solution of problem* (3.1)*, denoted by X^*, can be analytically expressed by*

$$X^* = \mathcal{H}_\nu(Y),$$

where $h_\nu(\cdot)$ is defined by (3.5) *and* (2.8).

Now, we consider the global solutions of problem (1.1). For any given $\nu, \mu > 0$ and $Y \in \mathbb{S}^n$, define the following auxiliary problem

$$\min_{X\in\mathbb{S}^n_+} f_{\nu,\mu}(X, Y) := \mu(f_\nu(X) - \|\mathcal{A}(X) - \mathcal{A}(Y)\|_F^2) + \|X - Y\|_F^2. \tag{3.6}$$

For simplicity, denote

$$A_\mu(Y) := Y + \mu\mathcal{A}^*(B - \mathcal{A}(Y)) \in \mathbb{S}^n.$$

Lemma 3.6 *Let X^* be the global solution of the problem* (3.6) *for any fixed $\nu > 0$, $\mu > 0$ and $Y \in \mathbb{S}^n$, then*

$$X^* = \mathcal{H}_{\nu\mu}(A_\mu(Y)).$$

Furthermore, if $h_{\nu\mu}(\cdot)$ is taken as (3.5)*, then X^* is the unique lowest-rank global solution of problem* (3.6)*, where $t^* := \frac{2-p}{2(1-p)}[\nu\mu(1-p)]^{1/(2-p)}$.*

Proof Note that $f_{\nu,\mu}(X, Y)$ can be reexpressed as

$$\begin{aligned}
f_{\nu,\mu}(X, Y) &= \mu(\nu\|X\|_{S_p}^p + \|\mathcal{A}(X) - B\|_F^2 - \|\mathcal{A}(X) - \mathcal{A}(Y)\|_F^2) + \|X - Y\|_F^2 \\
&= \nu\mu\|X\|_{S_p}^p + \|X\|_F^2 - 2\langle X, Y + \mu\mathcal{A}^*(B - \mathcal{A}(Y))\rangle \\
&\quad + \|Y\|_F^2 + \mu\|B\|_F^2 - \mu\|\mathcal{A}(Y)\|_F^2 \\
&= \nu\mu\|X\|_{S_p}^p + \|X - A_\mu(Y)\|_F^2 + \|Y\|_F^2 + \mu\|B\|_F^2 \\
&\quad - \mu\|\mathcal{A}(Y)\|_F^2 - \|A_\mu(Y)\|_F^2.
\end{aligned}$$

Hence, the problem (3.6) for any fixed ν, μ and Y is equivalent to

$$\min_{X\in\mathbb{S}^n_+} \{\|X - A_\mu(Y)\|_F^2 + \nu\mu\|X\|_{S_p}^p\}.$$

The proof is thus complete according to Theorem 3.4 and Remark 3.1. □

Lemma 3.7 *If X^* is a global minimizer of S_2–S_p minimization problem (1.1) for any fixed number $\nu > 0$ and μ is any fixed number satisfying $0 < \mu \leq \|\mathcal{A}\|^{-2}$, then X^* is also a global minimizer of the problem (3.6) for given $Y = X^*$, i.e.,*

$$f_{\nu,\mu}(X^*, X^*) \leq f_{\nu,\mu}(X, X^*) \text{ for all } X \in \mathbb{S}^n_+.$$

Proof For any $X \in \mathbb{S}^n_+$, since $0 < \mu \leq \|\mathcal{A}\|^{-2}$, we have

$$\|X - X^*\|_F^2 - \mu\|\mathcal{A}(X) - \mathcal{A}(X^*)\|_F^2 \geq \|X - X^*\|_F^2 - \mu\|\mathcal{A}\|^2\|X - X^*\|_F^2 \geq 0.$$

So,

$$\begin{aligned} f_{\nu,\mu}(X, X^*) &= \mu(f_\nu(X) - \|\mathcal{A}(X) - \mathcal{A}(X^*)\|_F^2) + \|X - X^*\|_F^2 \\ &= \mu(\nu\|X\|_{S_p}^p + \|\mathcal{A}(X) - B\|_F^2) + (\|X - X^*\|_F^2 - \mu\|\mathcal{A}(X) - \mathcal{A}(X^*)\|_F^2) \\ &\geq \mu(\nu\|X\|_{S_p}^p + \|\mathcal{A}(X) - B\|_F^2) \\ &= \mu f_\nu(X) \geq \mu f_\nu(X^*) = f_{\nu.\mu}(X^*, X^*), \end{aligned}$$

the last inequality is from that X^* is a global minimizer, which completes the proof. □

By applying Lemmas 2.2, 3.6 and 3.7, we derive the main results of this paper.

Theorem 3.8 *If X^* is a global solution of S_2–S_p minimization problem (1.1), then for any given $\mu \in (0, \|\mathcal{A}\|^{-2}]$, X^* satisfies*

$$X^* = \mathcal{H}_{\nu\mu}(A_\mu(X^*)). \tag{3.7}$$

Furthermore, let $L := (\nu\mu(1 - p))^{1/(2-p)}$, we have

$$either\ \ [\lambda(X^*)]_i > L\ \ or\ \ [\lambda(A_\mu(X^*))]_i \leq t^*, i = 1, 2, \cdots, n, \tag{3.8}$$

where $t^* := \frac{2-p}{2(1-p)}[\nu\mu(1 - p)]^{1/(2-p)}$.

Proof Since X^* is also a global minimizer of the problem (3.6) for given $Y = X^*$, we can directly obtain (3.7) from Lemmas 3.6 and 3.7. By the (3.7) and (2.7), we have

$$\begin{aligned} [\lambda(X^*)]_i &= h_{\nu\mu}([\lambda(A_\mu(X^*))]_i) \\ &= \begin{cases} h_{\nu\mu,p}([\lambda(A_\mu(X^*))]_i), & \text{if } [\lambda(A_\mu(X^*))]_i > t^* \\ (\nu\mu(1-p))^{1/(2-p)} \text{ or } 0, & \text{if } [\lambda(A_\mu(X^*))]_i = t^* \\ 0. & \text{if } [\lambda(A_\mu(X^*))]_i < t^* \end{cases} \end{aligned}$$

By the continuity of $h_{\nu\mu}(t)$ on $(t^*, +\infty)$ (by Lemma 2.2) and together with (2.7), we have

$$\lim_{t \downarrow t^*} h_{\nu\mu}(t) = L.$$

Together with the strict monotonicity of $h_{\nu\mu}(t)$ on $(t^*, +\infty)$ (by Lemma 2.2), we can obtain that $[\lambda(X^*)]_i > L$ as $[\lambda(A_\mu(X^*))]_i > t^*$. The proof is complete. □

Remark 3.2 If $h_{\nu\mu}(\cdot)$ is taken as (3.5), from Theorem 3.8 we have either $[\lambda(X^*)]_i > L$ or $[\lambda(X^*)]_i = 0$ for $i = 1, 2, \cdots, n$. In some sense, this result can be regarded as an extension of Theorem 2.1 in [7].

Remark 3.3 Theorem 3.8 establishes the necessary condition for global optimal solutions of S_2–S_p minimization, i.e., the global optimal solutions of problem (1.1) are fixed points of a symmetric matrix p-thresholding operator. The reverse does not hold generally.

At the same time, Theorem 3.8 also provides the exact lower bound for the positive eigenvalues at global optimal solutions, which can be used to identify zero eigenvalues precisely at any global numerical solution. Here we give a small example. Consider the following S_2–$S_{1/2}$ minimization problem

$$\begin{aligned} &\min_{X\in\mathbb{S}^5} f_\nu(X) = \|X - B\|_F^2 + 0.98\|X\|_{S_{1/2}}^{1/2} \\ &\text{s.t.} \quad X \succeq 0, \end{aligned} \tag{3.9}$$

where

$$B = \begin{pmatrix} 1 & 0 & 0 & 0.2 & 0 \\ 0 & 0 & 0.1 & 0 & 0 \\ 0 & 0.1 & 1 & 0 & 0 \\ 0.2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \in \mathbb{S}^5,$$

and $\mathcal{A}(X) = X$. From Theorem 3.4, we can get the global optimal solution of (3.9)

$$X^* = \mathcal{H}_{0.98}(B) = \begin{pmatrix} 0.7299 & 0 & 0 & 0.1406 & 0 \\ 0 & 0.007 & 0.0707 & 0 & 0 \\ 0 & 0.0707 & 0.7144 & 0 & 0 \\ 0.1406 & 0 & 0 & 0.0271 & 0 \\ 0 & 0 & 0 & 0 & 0.7090 \end{pmatrix}.$$

Furthermore, we have $\mathbf{rank}(X^*) = 3$, $\lambda(X^*) = (0.7569, 0.7215, 0.7090, 0, 0)^T$. For a given $\mu = 0.5$, applying Theorem 3.8 and $\|\mathcal{A}\|^{-2} = 1$, we have the lower bound $L = 0.3915 > 0$.

4 Concluding remarks

In this paper, we have developed the symmetric matrix p-thresholding operator representation theory, and then established the necessary condition for global optimal solutions of S_2–S_p minimization, i.e., the global optimal solutions of problem (1.1) are fixed points of a symmetric matrix p-thresholding operator. It is analogous to the fixed point representation of nuclear norm regularization solution associated with the so-called singular value shrinkage operator (see, for example, [1,13]). On the other hand, it is more complicated than the singular value shrinkage operator due to non-convex, nonsmooth and non-Lipschitz minimization problem. Furthermore, we have established the exact lower bound for the positive eigenvalues at global optimal solutions, which can be used to identify zero eigenvalues precisely in any global numerical solution.

In the future, we shall study under what conditions the fixed points of the symmetric matrix p-thresholding operator are global optimal solutions of problem (1.1). We shall also develop some numerical algorithm for solving problem (1.1) based on the formula (3.7).

Acknowledgments The work was supported in part by the National Basic Research Program of China (2010CB732501), the National Natural Science Foundation of China (71271021, 11171018), the Fundamental Research Funds for the Central Universities of China (2012YJS118).

References

1. Cai, J.F., Candès, E.J., Shen, Z.: A singular value thresholding algorithm for matrix completion. SIAM J. Optim. **20**, 1956–1982 (2010)
2. Candès, E.J., Eldar, Y., Strohmer, T., Voroninski, V.: Phase retrieval via matrix completion. SIAM J. Imaging Sci. **6**(1), 199–225 (2013)
3. Candès, E.J., Strohmer, T., Voroninski, V.: PhaseLift: exact and stable signal recovery from magnitude measurements via convex programming. Commun. Pure Appl. Math. **66**(8), 1241–1274 (2013)
4. Chen, C.H., He, B.S., Yuan, X.M.: Matrix completion via an alternating direction method. IMA J. Numer. Anal. **32**, 227–245 (2012)
5. Chen, X.: Smoothing methods for nonsmooth, nonconvex minimization. Math. Program. **134**, 71–99 (2012)
6. Chen, X., Ge, D., Wang, Z., Ye, Y.: Complexity of the unconstrained L_2-L_p minimization. Math. Program. (2012). doi:10.1007/s10107-012-0613-0
7. Chen, X., Xu, F., Ye, Y.: Lower bound theory of nonzero entries in solutions of ℓ_2-ℓ_p minimization. SIAM Sci. Comput. **32**(5), 2832–2852 (2010)
8. Ji, S., Sze, K.F., Zhou, Z., So, A.M.C., Ye, Y.Y.: Beyond convex relaxation: a polynomial-time non-convex optimization approach to network localization. In: The Proceedings of the 32nd IEEE International Conference on Computer Communications (INFOCOM 2013), pp. 2499–2507. Torino (2013)
9. Kong, L.C., Sun, J., Xiu, N.H.: S-semigoodness for low-rank semidefinite matrix recovery. http://www.optimization-online.org/DB_HTML/2013/01/3732.html
10. Krislock, N., Wolkowicz, H.: Euclidean distance matrices and applications. In: Anjos, M., Lasserre, J. (eds.) Handbook of Semidefinite, Cone and Polynomial Optimization, pp. 879–914. Springer, Heidelberg (2002)
11. Li, J.T., Jia, Y.M., Li, W.L.: Adaptive huberized support vector machine and its application to microarray classification. Neural Comput. Appl. **20**, 123–132 (2011)
12. Lu, Y., Zhang, L.W., Wu, J.: A smoothing majorization method for l_2-l_p matrix minimization. http://www.optimization-online.org/DB_HTML/2013/01/3735.html
13. Ma, S., Goldfarb, D., Chen, L.: Fixed point and bregman iterative methods for matrix rank minimization. Math. Program. **128**, 321–353 (2011)
14. Rao, G., Peng, Y., Xu, Z.B.: Robust sparse and low-rank matrix decomposition based on $S_{1/2}$ modeling. Science China Ser. F **43**(6), 733–748 (2013). In Chinese
15. Sun, D.F.: The low-rank correlation matrix problems: nonconvex regularization and successive convex relaxations. Tech. Rep., Department of Mathematics, National University of Singapore (2012)
16. Xu, Z.B., Chang, X., Xu, F.M., Zhang, H.: $L_{1/2}$ regularization: a thresholding representation theory and a fast solver. IEEE Trans. Neural Netw. Learning Sys. **23**, 1013–1027 (2012)