

Article

Exact Solution Analysis of Strongly Convex Programming for Principal Component Pursuit

Qingshan You and Qun Wan *

School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China; youlin_2001@163.com

* Correspondence: wanqun@uestc.edu.cn; Tel.: +86-28-61830358

Academic Editor: Willy Susilo

Received: 14 October 2016; Accepted: 25 January 2017; Published: 2 February 2017

Abstract: In this paper, we address strongly convex programming for principal component analysis, which recovers a target matrix that is a superposition of low-complexity structures from a small set of linear measurements. In this paper, we firstly provide sufficient conditions under which the strongly convex models lead to the exact low-rank matrix recovery. Secondly, we also give suggestions that will guide us how to choose suitable parameters in practical algorithms. Finally, the proposed result is extended to the principal component pursuit with reduced linear measurements and we provide numerical experiments.

Keywords: low-complexity structure; strongly convex programming; principal component pursuit; reduced linear measurements

1. Introduction

Recently, much attention has been focussed on the problem of recovering a target matrix with low-complexity structure from a small set of linear measurements. This problem has regained great concern since the publication of the pioneering works of E.J. Candés et al. [1–4]. It can be found in many different fields, such as medical imaging [5–7], seismology [8], information retrieval [9] and machine learning [10], especially the detection of moving objects [11]. In the case of detection of moving objects, the columns of matrix M are the video frames, and the low-rank matrix L_0 and the sparse matrix S_0 are the stationary background and the moving objects in the foreground respectively. According to [12], the main problem of detection of moving objects is how to recover the low-rank matrix L_0 and the sparse matrix S_0 from the given data matrix $M = L_0 + S_0$, where $L_0 \in \mathbb{R}^{n \times n}$ has low-rank, and S_0 is a sparse matrix. In the paper [12], E.J. Candés et al. have proved that that most low-rank matrices and the sparse components can be recovered, provided that the rank of the low-rank component is not too large, and that the sparse component is reasonably sparse; and more importantly they proved that this can be done by solving a simple convex optimization problem, i.e., provided that the rank of the matrix L and the cardinality of the sparse matrix S obey some suitable conditions, most matrices L_0 of rank r and the sparse component S_0 can be perfectly recovered by solving the simple optimization problem as follows:

$$\begin{aligned} & \text{minimize} && \|L\|_* + \lambda \|S\|_1 \\ & \text{subject to} && L + S = M \end{aligned} \quad (1)$$

wherein $\|L\|_*$ is the nuclear norm of matrix L , and $\|S\|_1$ is the sum of absolute values of all matrix entries.

Strongly convex optimizations have many advantages, e.g., unique optimal solution. Many scholars suggest solving their strongly convex approximations (see, e.g., [13–16]), instead of directly

solving the original convex optimizations. J.F. Cai et al. addressed the strongly convex optimization $\tau\|X\|_* + \frac{1}{2}\|X\|_F^2$ ($\|X\|_F$ denoting the Frobenius norm) instead of the original convex optimization $\|X\|_*$, and the authors introduced an important algorithm (singular value thresholding algorithm) to solve matrix completion based on this strongly convex optimization [15]; J. Wright et al. also addressed the strongly convex optimization $\|L\|_* + \lambda\|S\|_1 + \frac{1}{2\tau}\|L\|_F^2 + \frac{1}{2\tau}\|S\|_F^2$ instead of the original convex optimization $\|L\|_* + \lambda\|S\|_1$, and the authors proposed the iterative thresholding(IT) algorithm to solve robust principal component analysis [14]. J. Wright et al. only confirm performance of iterative thresholding by numerical experiments; however, the authors do not provide the sufficient conditions which can guarantee strongly convex optimization $\|L\|_* + \lambda\|S\|_1 + \frac{1}{2\tau}\|L\|_F^2 + \frac{1}{2\tau}\|S\|_F^2$ and original convex optimization $\|L\|_* + \lambda\|S\|_1$ have the same optimal solution. In this article, the authors have given many suitable sufficient conditions that would lead the strongly convex models to the exact low-rank and sparse matrix recovery. Some suggestions have been given in [16] on how to choose suitable parameters in practical algorithms. However, the results shown in [16] are limited to a special condition, i.e., $Q = \mathbb{R}^{n \times n}$. In this paper, we extend this result to the principal component pursuit with reduced linear measurements, that is Q^\perp is a p -dimensional random subspace instead of $Q = \mathbb{R}^{n \times n}$. It is easy to prove that the results in [16] are only a special case of those that we proposed.

1.1. Basic Problem Formulations

In this subsection, we will interpret an important strongly convex programming to be addressed in this paper, and list its existence and uniqueness theorems, which will be proved in the next sections. In [17], the authors have studied principal component pursuit with reduced linear measurements and given sufficient conditions under which L_0 and S_0 can be perfectly recovered.

In this paper, we address a strongly convex programming, and prove that it has the capability to guarantee exact low-rank matrix recovery. The proposed optimization is realized in the following way:

$$\begin{aligned} & \text{minimize} && \|L\|_* + \lambda\|S\|_1 + \frac{1}{2\tau}\|L\|_F^2 + \frac{1}{2\tau}\|S\|_F^2 \\ & \text{subject to} && \mathcal{P}_Q M = \mathcal{P}_Q(L + S) \end{aligned} \tag{2}$$

wherein, $\tau \geq 0$ refers to some positive penalty parameter and P_Q the orthogonal projection onto the linear subspace Q . We also assume that Q^\perp is a random subspace. The existence and uniqueness theorems when $\tau = \infty$ in (2) are provided in [17], and listed below. In the end, how to choose suitable parameters in the optimization model (2) is discussed.

Theorem 1 ([17]). *If we fix any $C_p > 0$, and let Q^\perp be a p -dimensional random subspace of $\mathbb{R}^{n \times n}$; L_0 obeys incoherence condition with parameter μ , and $\text{supp}(S_0) \sim \text{Ber}(\rho)$. With high probability, the solution of problem (2) with $\lambda = \frac{1}{\sqrt{n}}$ is exact, i.e., $\hat{L} = L_0$ and $\hat{S} = S_0$, provided that*

$$\text{Rank}(L_0) < C_r n \mu^{-1} (\log n)^{-2} \quad p < C_p n \quad \text{and} \quad \rho < \rho_0 \tag{3}$$

wherein, C_r, C_p and ρ are positive numerical constants and $\rho_0 < 1$.

1.2. Contents and Notations

We provide a brief summary of the notations which are used throughout the paper. $\|X\|$ denotes the operator norm of matrix X , $\|X\|_F$ denotes the Frobenius norm, $\|X\|_*$ the nuclear norm, and the dual norm of $\|X\|_{(i)}$ by $\|X\|_{(i)}^*$. The Euclidean inner product between two matrices X, Y is defined by the formula $\langle X, Y \rangle = \text{trace}(X^*Y)$. It's easy to note that $\|X\|_F^2 = \langle X, X \rangle$. The Cauchy-Schwarz inequality which will often be used in next sections gives $\langle X, Y \rangle \leq \|X\|_F \|Y\|_F$, and it is well known that $\langle X, Y \rangle \leq \|X\|_{(i)} \|Y\|_{(i)}^*$, (see e.g., [2,18]). Linear transformations which act on the space of matrices are denoted by $\mathcal{P}X$. It can be easily seen that the operator of \mathcal{P} is high dimension matrix in substance. The operator norm of this operator is signified by $\|\mathcal{P}\|$. It should also be noted

that $\|\mathcal{P}\| = \sup_{\{\|X\|_F=1\}} \|\mathcal{P}X\|_F$. We say an event E occurs with high probability if $\mathbb{P}[E] = Cm^{-\alpha}$. We denote the reduced singular value decomposition (SVD) of the low-rank matrix L_0 as $L_0 = U\Sigma V^*$, and define a linear subspace T as follows:

$$T = \{UX^* + YV^* : X \in \mathbb{R}^{n \times r}, Y \in \mathbb{R}^{m \times r}\}$$

We denote the support of the sparse matrix S_0 as Ω , by a slight abuse of notation, we also denote Ω as the subspace of matrices whose support is contained in the support of S_0 .

The rest of this paper is organized as follows. In Section 2, we firstly list many important lemmas, at the same time we prove a key lemma on which our main result is based. Secondly suggestions are given in Section 3, these conditions will guide us to choose suitable parameters in practical algorithms. Thirdly, the numerical result is given in Section 4. Finally, conclusions and results are discussed in Section 5.

2. Important Lemmas

In this section, we first list some useful lemmas which will be used throughout this paper and then prove a main lemma. Although the main lemma is similar to the corresponding one in [17], they have significant difference, in which the construction of W^Q is very different. That leads to our necessary additional work.

Lemma 1 ([17]). *Let $\Gamma = Q \cap T^\perp$, then we have $\Gamma^\perp = Q^\perp \oplus T$, and $\dim(Q^\perp \oplus T \oplus \Omega) = \dim(Q^\perp) + \dim(T) + \dim(\Omega)$. At the same time we assume that $\|\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp}\| < 1/2$ and $\lambda < 1$. Then (L_0, S_0) is the unique optimal solution to (2) if there is a pair $(W, F) \in \mathbb{R}^{n \times n} \times \mathbb{R}^{n \times n}$ satisfying the following conditions*

$$UV^* + W = \lambda(\text{sgn}(S_0) + F + \mathcal{P}_\Omega D) \in Q$$

In which $\mathcal{P}_T = 0$, $\|W\| < 1/2$, $\mathcal{P}_\Omega F = 0$, $\|F\|_\infty < 1/2$, and $\|\mathcal{P}_\Omega D\|_F \leq 1/4$.

Lemma 2 ([17]). *If $\Omega \sim \text{Ber}(\rho)$ for some small $\rho \in (0, 1)$ and the other conditions of Theorem 1 are true. Then, the matrix W^L obeys the below inequalities with high probability.*

- (a). $\|W^L\| < 1/4$
- (b). $\|\mathcal{P}_\Omega(UV^* + W^L)\|_F < \lambda/4$
- (c). $\|\mathcal{P}_{\Omega^\perp}(UV^* + W^L)\|_\infty < \lambda/4$

Lemma 3 ([17]). *In addition to the assumptions in the previous lemma, suppose that the signs of the non-zero entries of S_0 are i.i.d. random. Then the matrix W^S obeys the below inequalities with high probability.*

- (a). $\|W^S\| < 1/8$
- (b). $\|\mathcal{P}_{\Omega^\perp} W^S\|_\infty < \lambda/8$

The construction of W^L and W^S can be found in the paper [17]. However, the matrix W^Q constructed in the paper [17] does not satisfy the requirement of our problem, so we have to modify this construction in order to satisfy the problem (2). Firstly we will give explicit construction of W^Q , and then prove that the modification of W^Q satisfies the proper property.

Construction of W^Q with least modification. We define W^Q by the following least squares problem:

$$\begin{aligned} W^Q &= \arg \min_X \|X\|_F \\ \text{subject to} \quad &\mathcal{P}_{Q^\perp} X = -\mathcal{P}_{Q^\perp}(UV^* + \frac{1}{\tau} L_0) \\ &\mathcal{P}_\Pi X = 0 \end{aligned}$$

wherein $\Pi = T \oplus \Omega$. This construction of W^Q satisfies Lemma 5 in the paper [17], also has the below proper property.

Lemma 4. *If $\tau \geq \|M\|_F$, $\Omega \sim \text{Ber}(\rho)$ for some small $\rho \in (0, 1)$ and the assumptions of Theorem 1 are true. Then the matrix W^Q obeys the below inequalities with high probability.*

- (a). $\|W^Q\| < 1/8$
- (b). $\|\mathcal{P}_{\Omega^\perp} W^Q\|_\infty < \lambda/8$

Proof. A: Bounding the Frobenius norm of $UV^* + \frac{1}{\tau}L_0$. For convenience, let $\xi := \|UV^* + \frac{1}{\tau}L_0\|_F$. According to triangle inequality, we have

$$\|L_0\|_F = \|M - S_0\|_F \leq \|M\|_F + \|S_0\|_F = \|M\|_F + \|\mathcal{P}_\Omega S_0\|_F$$

In the last equality, we have used $S_0 \in \Omega$. Note that

$$\|\mathcal{P}_\Omega S_0\|_F = \|\mathcal{P}_\Omega(M - L_0)\|_F \leq \|\mathcal{P}_\Omega M\|_F + \|\mathcal{P}_\Omega L_0\|_F$$

According to the derivation in [16], the below inequality is true with high probability

$$\|\mathcal{P}_\Omega L_0\|_F \leq \frac{\sqrt{3}}{3} \|\mathcal{P}_{\Omega^\perp} M\|_F \leq \frac{\sqrt{3}}{3} \|M\|_F$$

Putting those all together, we can obtain

$$\|L_0\|_F \leq (\sqrt{3}/3 + 2) \|M\|_F$$

Combining with $\tau \geq \|M\|_F$, we can obtain

$$\xi \leq \|UV^*\|_F + \frac{\|L_0\|_F}{\tau} \leq r + \frac{(\sqrt{3}/3 + 2) \|M\|_F}{\tau} \leq r + \sqrt{3}/3 + 2$$

W^Q is the optimum solution of least squares problem, due to this we can use the convergent Neumann series expansion. It's easy to note that

$$W^Q = \mathcal{P}_{\Pi^\perp} \sum_{k>0} (\mathcal{P}_{Q^\perp} \mathcal{P}_\Pi \mathcal{P}_{Q^\perp})^k (\mathcal{P}_{Q^\perp} (-UV^* - \frac{1}{\tau}L_0))$$

According to triangle inequality, we have

$$\|W^Q\|_F \leq \left\| \sum_{k>0} (\mathcal{P}_{Q^\perp} \mathcal{P}_\Pi \mathcal{P}_{Q^\perp})^k \right\| \|\mathcal{P}_{Q^\perp} (-UV^* - \frac{1}{\tau}L_0)\|_F \tag{4}$$

B: Estimating the first inequality of Lemma 4. In order to bound $\|W^Q\|_F$, we first have to bound the norm of $\sum_{k>0} (\mathcal{P}_{Q^\perp} \mathcal{P}_\Pi \mathcal{P}_{Q^\perp})^k$ and bound the Frobenius norm of $\mathcal{P}_{Q^\perp} (-UV^* - \frac{1}{\tau}L_0)$. The norm of $\sum_{k>0} (\mathcal{P}_{Q^\perp} \mathcal{P}_\Pi \mathcal{P}_{Q^\perp})^k$ satisfies

$$\begin{aligned} \left\| \sum_{k>0} (\mathcal{P}_{Q^\perp} \mathcal{P}_\Pi \mathcal{P}_{Q^\perp})^k \right\| &\leq \sum_{k>0} \|(\mathcal{P}_{Q^\perp} \mathcal{P}_\Pi \mathcal{P}_{Q^\perp})^k\| \\ &\leq \sum_{k>0} \|\mathcal{P}_{Q^\perp} \mathcal{P}_\Pi\|^{2k} \end{aligned}$$

According to Lemma 11 in the paper [17], the following inequality is true with high probability for any $\epsilon > 0$,

$$\|\mathcal{P}_{Q^\perp} \mathcal{P}_\Pi\|^2 \leq \frac{64}{1 - \sqrt{\rho + \epsilon}} \left(\left(\sqrt{\frac{p}{n^2}} + \sqrt{\frac{5\rho}{4}} \right)^2 + \left(\sqrt{\frac{p}{n^2}} + \sqrt{\frac{2r}{n}} \right)^2 \right)$$

According to the paper [17], the following inequality is true with high probability:

$$\left\| \sum_{k>0} (\mathcal{P}_{Q^\perp} \mathcal{P}_\Pi \mathcal{P}_{Q^\perp})^k \right\| \leq \frac{4}{3}$$

Secondly, we will bound the Frobenius norm of $\mathcal{P}_{Q^\perp}(-UV^* - \frac{1}{\tau}L_0)$. \mathcal{P}_{Q^\perp} has the same distribution as $H(H^*H)^{-1}H^*$, in which $H \in \mathbb{R}^{n^2 \times p}$ is a random Gaussian matrix with i.i.d. entries satisfied $\mathcal{N}(0, 1/n^2)$. Therefore, we can obtain the below inequality

$$\begin{aligned} & \|\mathcal{P}_{Q^\perp}(UV^* + \frac{1}{\tau}L_0)\|_F \\ &= \|H(H^*H)^{-1}H^* \text{vec}(UV^* + \frac{1}{\tau}L_0)\|_F \\ &\leq \|H(H^*H)^{-1}\| \|H^* \text{vec}(UV^* + \frac{1}{\tau}L_0)\|_2 \end{aligned}$$

Together with Lemma 7 in the paper [17], we can obtain

$$\mathbb{P}[\|H(H^*H)^{-1}\| \geq 4] \leq e^{-\frac{n^2}{32}}$$

It is easy to note that any entries of $H^* \text{vec}(UV^* + \frac{1}{\tau}L_0)$ have the same distribution as $\langle G, UV^* + \frac{1}{\tau}L_0 \rangle$, in which $G_{ij} \sim \mathcal{N}(0, 1/n^2)$ are independent identically distributed. It is obvious to see that

$$\mathbb{E}\{\langle G, UV^* + \frac{1}{\tau}L_0 \rangle\} = \langle \mathbb{E}\{G\}, UV^* + \frac{1}{\tau}L_0 \rangle = 0$$

and

$$\begin{aligned} \text{Var}\{\langle G, UV^* + \frac{1}{\tau}L_0 \rangle\} &= \sum_{ij} (UV^* + \frac{1}{\tau}L_0)_{ij}^2 \text{Var}\{G_{ij}\} \\ &= \xi^2/n^4 \end{aligned}$$

Therefore $\langle G, UV^* + \frac{1}{\tau}L_0 \rangle$ is distributed with $\mathcal{N}(0, \xi/n^2)$, where $\xi := \|UV^* + \frac{1}{\tau}L_0\|_F$. For simplicity, we define $Z := H^* \text{vec}(UV^* + \frac{1}{\tau}L_0)$. Using the Jensen inequality, we can obtain

$$\mathbb{E}[\|Z\|_2] \leq (\mathbb{E}[\|Z\|_2^2])^{1/2} = \sqrt{\frac{p\xi}{n^2}}$$

According to the Proposition 2.18 in [18], we can obtain

$$\mathbb{P}\left[\|Z\|_2 \geq \mathbb{E}[\|Z\|_2] + t\sqrt{\frac{\xi}{n^2}}\right] \leq e^{-t^2/2}$$

Setting $t = \sqrt{6\log n}$, after a simple inference, we can obtain the below inequality with high probability.

$$\|W^Q\| \leq \|W^Q\|_F \leq \frac{16}{3} \left(\frac{\sqrt{p\xi}}{n} + \frac{\sqrt{6\xi \log n}}{n} \right)$$

For sufficiently large n , the first inequality of Lemma 4 is established. We will estimate the second inequality of Lemma 4 further.

C: Estimating the second inequality of Lemma 4, Note that

$$W^Q = \mathcal{P}_{\Pi^\perp} \mathcal{P}_{Q^\perp} \sum_{k>0} (\mathcal{P}_{Q^\perp} \mathcal{P}_{\Pi} \mathcal{P}_{Q^\perp})^k (\mathcal{P}_{Q^\perp} (-UV^* - \frac{1}{\tau} L_0))$$

After a simple inference, we can obtain the below inequality.

$$\|\mathcal{P}_{\Omega^\perp} W^Q\|_\infty \leq \frac{C\sqrt{\xi}}{n^2} (\sqrt{p} + \sqrt{6 \log n})^2$$

wherein C is some constant. Note that the second inequality of Lemma 4 is established for sufficiently large n . \square

3. Estimating Parameter τ

In this section, we shall provide sufficient conditions under which $(L_0; S_0)$ is the unique and exact solution of the strongly convex programming (2) with high probability, i.e., the solution of problem (2) is exact $\hat{L} = L_0$ and $\hat{S} = S_0$. Afterwards, an explicit lower bound of τ will be provided, which will further guide us to choose suitable parameters in practical algorithms.

Theorem 2. Let $\Gamma = Q \cap T^\perp$, then we have $\Gamma^\perp = Q^\perp \oplus T$, and $\dim(Q^\perp \oplus T \oplus \Omega) = \dim(Q^\perp) + \dim(T) + \dim(\Omega)$. Assume that $\|\mathcal{P}_{\Omega} \mathcal{P}_{\Gamma^\perp}\| < 1/2$ and $\lambda < 1$. If there is a pair $(W, F) \in \mathbb{R}^{n \times n} \times \mathbb{R}^{n \times n}$ and a matrix D satisfying

$$UV^* + W + \frac{1}{\tau} L_0 = \lambda(\text{sgn}(S_0) + F + \mathcal{P}_{\Omega} D) + \frac{1}{\tau} S_0 \in Q$$

with

$$\mathcal{P}_T W = 0, \|W\| \leq \beta, \mathcal{P}_{\Omega} F = 0, \|F\|_\infty \leq \beta, \|\mathcal{P}_{\Omega} D\|_F \leq \alpha \tag{5}$$

where α, β are positive parameters satisfying

$$\alpha + \beta \leq 1 \tag{6}$$

Then (L_0, S_0) is the unique solution of the strongly convex programming (2).

Proof. For any feasible perturbation (H_L, H_S) , it's easy to note that $\mathcal{P}_Q H_L = \mathcal{P}_Q H_S$. According to the definition of Γ , we have $\Gamma \subset Q$. Therefore $\mathcal{P}_\Gamma H_L = \mathcal{P}_\Gamma H_S$. For simplicity, we define $f(L, S) = \|L\|_* + \lambda \|S\|_1 + \frac{1}{2\tau} \|L\|_F^2 + \frac{1}{2\tau} \|S\|_F^2$, and we can obtain the below inequality

$$\begin{aligned} & f(L_0 + H_L, S_0 - H_S) \\ \geq & f(L_0, S_0) + \langle UV^* + W_0 + \frac{1}{\tau} L_0, H_L \rangle - \langle \lambda \text{sgn}(S_0) + \lambda F_0 + \frac{1}{\tau} S_0, H_S \rangle \\ \geq & f(L_0, S_0) + \langle W_0, H_L \rangle - \langle W, H_L \rangle + \langle UV^* + W + \frac{1}{\tau} L_0, \mathcal{P}_Q H_L \rangle \\ & - \langle \lambda F_0, H_S \rangle + \langle \lambda F, H_S \rangle - \langle \lambda \text{sgn}(S_0) + \lambda F + \frac{1}{\tau} S_0, \mathcal{P}_Q H_S \rangle \\ \geq & f(L_0, S_0) + \langle W_0, \mathcal{P}_{T^\perp} H_L \rangle - \langle W, \mathcal{P}_{T^\perp} H_L \rangle \\ & - \langle \lambda F_0, \mathcal{P}_{\Omega^\perp} H_S \rangle + \langle \lambda F, \mathcal{P}_{\Omega^\perp} H_S \rangle - \langle \lambda \mathcal{P}_{\Omega} D, \mathcal{P}_Q H_S \rangle \\ \geq & f(L_0, S_0) + (1 - \beta) \|\mathcal{P}_{T^\perp} H_L\|_* + (1 - \beta) \lambda \|\mathcal{P}_{\Omega^\perp} H_S\|_1 - \alpha \lambda \|\mathcal{P}_{\Omega} H_S\|_F \end{aligned}$$

In the second inequality above, we have used the facts

$$UV^* + W + \frac{1}{\tau}L_0 = \lambda(\text{sgn}(S_0) + F + \mathcal{P}_Q D) + \frac{1}{\tau}S_0 \in Q$$

In the third inequality above, we have used the property $\mathcal{P}_Q H_L = \mathcal{P}_Q H_S$.

We have provided the bound of $f(L_0 + H_L, S_0 - H_S)$, and then we will give the bound of $\|\mathcal{P}_\Omega H_S\|_F$. According to the definition of Γ , we can obtain

$$\begin{aligned} \|\mathcal{P}_\Omega H_S\|_F &\leq \|\mathcal{P}_\Omega \mathcal{P}_\Gamma H_S\|_F + \|\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp} H_S\|_F \\ &\leq \|\mathcal{P}_\Omega \mathcal{P}_\Gamma H_L\|_F + \frac{1}{2}\|H_S\|_F \\ &\leq \|\mathcal{P}_\Gamma H_L\|_F + \frac{1}{2}\|\mathcal{P}_\Omega H_S\|_F + \frac{1}{2}\|\mathcal{P}_{\Omega^\perp} H_S\|_F \\ &\leq \|\mathcal{P}_{T^\perp} H_L\|_F + \frac{1}{2}\|\mathcal{P}_\Omega H_S\|_F + \frac{1}{2}\|\mathcal{P}_{\Omega^\perp} H_S\|_F \end{aligned}$$

Therefore

$$\|\mathcal{P}_\Omega H_S\|_F \leq 2\|\mathcal{P}_{T^\perp} H_L\|_F + \|\mathcal{P}_{\Omega^\perp} H_S\|_F \leq 2\|\mathcal{P}_{T^\perp} H_L\|_* + \|\mathcal{P}_{\Omega^\perp} H_S\|_1$$

Putting those all together, we get

$$\begin{aligned} &f(L_0 + H_L, S_0 - H_S) \\ &\geq f(L_0, S_0) + (1 - \beta - 2\alpha\lambda)\|\mathcal{P}_{T^\perp} H_L\|_* + (1 - \beta - \alpha)\lambda\|\mathcal{P}_{\Omega^\perp} H_S\|_1 \end{aligned}$$

According to (6), the inequality above implies that (L_0, S_0) is a solution to (2), i.e., $(L_0; S_0)$ is the exact solution of the strongly convex programming (2) with high probability. The uniqueness follows from the strong convexity of the objective in (2).

In the practice, the choose of parameter τ is very difficult, therefore we will provide the criterion of the value of τ in the next section which will guide us to choose suitable parameters in practical algorithms. Theorem 3 and 4 provide the criterion of the value of τ , and the bound of τ in Theorem 4 is more explicit and useful in practice. \square

Theorem 3. Let $\tau_1 = \frac{\|\mathcal{P}_{\Omega^\perp} L_0\|_\infty}{(\beta - \frac{1}{2})\lambda}$, $\tau_2 = \frac{\|\mathcal{P}_\Omega(L_0 - S_0)\|_F}{(\alpha - \frac{1}{4})\lambda}$, and $\tau_3 = \frac{4(\|\mathcal{P}_{\Omega^\perp} L_0\|_\infty + \|\mathcal{P}_\Omega(L_0 - S_0)\|_F)}{\lambda}$. Assume

$$\tau \geq \max(\tau_1, \tau_2, \tau_3, \|M\|_F) \tag{7}$$

Then, under the other assumptions of Theorem 1, (L_0, S_0) is the unique solution to the strongly convex programming (2) with high probability.

Proof. In order to check the conditions in Theorem 2, we will prove the existence of a matrix W obeying

$$\begin{cases} \mathcal{P}_T W = 0 \\ \|W\| \leq \beta \\ \mathcal{P}_{Q^\perp} W = -\mathcal{P}_{Q^\perp}(UV^* + \frac{1}{\tau}L_0) \\ \|\mathcal{P}_{\Omega^\perp}(UV^* + W + \frac{1}{\tau}L_0 - \frac{1}{\tau}S_0)\|_\infty \leq \beta\lambda \\ \|\mathcal{P}_\Omega(UV^* + W - \lambda\text{sgn}(S_0) + \frac{1}{\tau}L_0 - \frac{1}{\tau}S_0)\|_F \leq \alpha\lambda \end{cases} \tag{8}$$

Note that $W = W^L + W^S + W^Q$. We will check above conditions hold true one by one. For simplicity, we define

$$\gamma := \|\mathcal{P}_{\Omega^\perp}(L_0 - S_0)\|_\infty, \quad \delta := \|\mathcal{P}_\Omega(L_0 - S_0)\|_F$$

Without loss of generality, let $\beta > 1/2$. With the help of the construction of W^L , W^S and W^Q , it is easy to check the first and second conditions are true. According to the modification of W^Q constructed in Lemma 4, and $\mathcal{P}_{Q^\perp}W^L = 0$ and $\mathcal{P}_{Q^\perp}W^S = 0$, we have $\mathcal{P}_{Q^\perp}W^Q = -\mathcal{P}_{Q^\perp}(UV^* + \frac{1}{\tau}L_0)$. It is easy to check that $\mathcal{P}_{Q^\perp}W = \mathcal{P}_{Q^\perp}W^L + \mathcal{P}_{Q^\perp}W^S + \mathcal{P}_{Q^\perp}W^Q = -\mathcal{P}_{Q^\perp}(UV^* + \frac{1}{\tau}L_0)$, which implies that the third condition holds true. Consequently, we will provide the last two conditions also hold true under some suitable assumptions. Pertaining to the fourth inequality, we have

$$\begin{aligned} & \|\mathcal{P}_{\Omega^\perp}(UV^* + W + \frac{1}{\tau}L_0 - \frac{1}{\tau}S_0)\|_\infty \\ \leq & \|\mathcal{P}_{\Omega^\perp}(UV^* + W^L)\|_\infty + \|\mathcal{P}_{\Omega^\perp}W^S\|_\infty \\ & + \|\mathcal{P}_{\Omega^\perp}W^Q\|_\infty + \frac{1}{\tau}\|\mathcal{P}_{\Omega^\perp}(L_0 - S_0)\|_\infty \\ \leq & \frac{\lambda}{4} + \frac{\lambda}{8} + \frac{\lambda}{8} + \frac{1}{\tau}\|\mathcal{P}_{\Omega^\perp}(L_0 - S_0)\|_\infty \\ \leq & \frac{\lambda}{2} + \frac{\gamma}{\tau} \end{aligned}$$

For the last inequality, noting that $\mathcal{P}_\Omega(W^S) = \lambda \text{sgn}(S_0)$ and $\mathcal{P}_\Omega(W^Q) = 0$, we can obtain

$$\begin{aligned} & \|\mathcal{P}_\Omega(UV^* + W - \lambda \text{sgn}(S_0) + \frac{1}{\tau}L_0 - \frac{1}{\tau}S_0)\|_F \\ = & \|\mathcal{P}_\Omega(UV^* + W^L + \frac{1}{\tau}L_0 - \frac{1}{\tau}S_0)\|_F \\ \leq & \|\mathcal{P}_\Omega(UV^* + W^L)\|_F + \frac{1}{\tau}\|\mathcal{P}_\Omega(L_0 - S_0)\|_F \\ \leq & \frac{\lambda}{4} + \frac{\delta}{\tau} \end{aligned}$$

In order to satisfy the condition (8), we choose a τ obeying

$$\frac{\lambda}{2} + \frac{\gamma}{\tau} \leq \beta\lambda, \text{ and } \frac{\lambda}{4} + \frac{\delta}{\tau} \leq \alpha\lambda \tag{9}$$

Therefore

$$\tau \geq \max\left(\frac{\gamma}{(\beta - \frac{1}{2})\lambda}, \frac{\delta}{(\alpha - \frac{1}{4})\lambda}\right) \tag{10}$$

Combining (9) with (6), we can obtain

$$\frac{\lambda}{2} + \frac{\gamma}{\tau} + \frac{\lambda}{4} + \frac{\delta}{\tau} \leq \beta\lambda + \alpha\lambda \leq \lambda$$

Therefore

$$\tau \geq \frac{4(\gamma + \delta)}{\lambda} \tag{11}$$

Together with (10) and (11), the Theorem 3 is established.

In order to simplify the Formula (7), we suppose $\alpha = 3/8$ and $\beta = 5/8$, which satisfy the conditions above. Therefore

$$\tau \geq \max\left(\frac{8\|\mathcal{P}_{\Omega^\perp}L_0\|_\infty}{\lambda}, \frac{8\|\mathcal{P}_\Omega(L_0 - S_0)\|_F}{\lambda}\right) \tag{12}$$

However, note that the exact lower bound is very hard to get, because we only have the information about the given data matrix M in practical problem. Noting that

$$\|\mathcal{P}_{\Omega^\perp} M\|_\infty \leq \|M\|_\infty$$

And according to the paper [16], we have

$$\|\mathcal{P}_\Omega(L_0 - S_0)\|_F \leq \frac{\sqrt{15}}{3} \|M\|_F$$

Therefore, we can choose

$$\tau \geq \max\left(\frac{8\|M\|_\infty}{\lambda}, \frac{8\sqrt{15}\|M\|_F}{3\lambda}\right)$$

It is obvious that $\|M\|_\infty \leq \|M\|_F$. Therefore, we can obtain Theorem 3.3 as follows. \square

Theorem 4. *Assuming*

$$\tau \geq \frac{8\sqrt{15}\|M\|_F}{3\lambda}$$

and the other assumptions of Theorem 1, (L_0, S_0) is the unique solution to the strongly convex programming (2) with high probability.

4. Numerical Results

In this section, we provide numerical experiments to certify the Theorem 4. Without loss of generality, we assume that $r = 2$, $M = L_0 + S_0$, and a rank- r matrix $L_0 = XY^T$ where X and Y are 15×2 and 30×2 matrices with entries independently sampled from a $\mathcal{N}(0; \delta^2)$ distribution, the sparse matrix $S_0 = \mathcal{P}_\Omega E$ with the support set of size $k_s = \rho_s mn$ uniformly at random. Assume that (L_1, S_1) and (L_2, S_2) are the optimal solutions of optimization problem (1) and strongly convex optimization problem (2) respectively. Numerical experiments are given under $\|M\|_F = 1$. Figure 1 shows probability of correct recovery ($\|L_1 - L_2\|_F^2 + \|S_1 - S_2\|_F^2 \leq 10^{-3}$) with different values of τ . It's noted that when $\frac{1}{\tau} < \frac{3\lambda}{8\sqrt{15}\|M\|_F} = 0.03$, the probability of correct recovery is nearly 100%; however when $\frac{1}{\tau} > 0.03$, the probability of correct recovery decreases fast. This phenomenon verifies the Theorem 4 by number experiments.

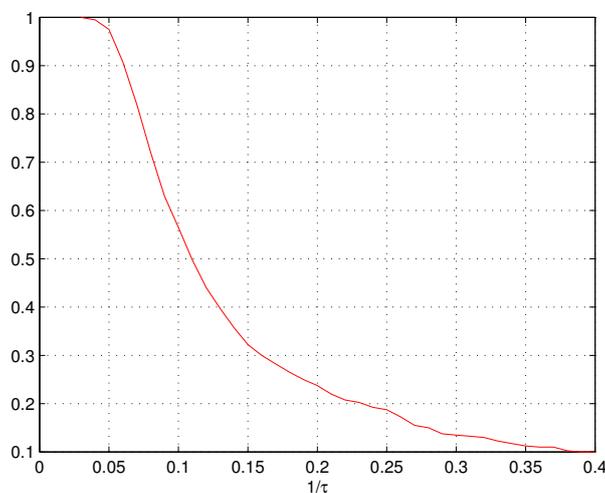


Figure 1. probability of correct recovery with different values of τ .

5. Results and Conclusions

In this paper, we have studied strongly convex programming for principal component pursuit with reduced linear measurements.

We firstly provide sufficient conditions under which the strongly convex models lead to the exact low rank and sparse components recovery, i.e.,

Assuming

$$\tau \geq \frac{8\sqrt{15}\|M\|_F}{3\lambda}$$

and the other assumptions of Theorem 1, (L_0, S_0) is the unique solution to the strongly convex programming (2) with high probability.

Secondly, we give the criterion of the choice of the value of τ , which gives very useful advice on how to set the suitable parameters in designing efficient algorithms. In particular, it is easy to note that the main results of paper [16] are only the special case of our results. In some sense, we extend the result of choosing suitable parameters to the general problem.

Acknowledgments: The author would like to thank the anonymous reviewers for their comments that helped to improve the quality of the paper. This research was supported by the National Natural Science Foundation of China (NSFC) under Grant U1533125, and the Scientific Research Program of the Education Department of Sichuan under Grant 16ZB0032.

Author Contributions: Qingshan You and Qun Wan conceived and designed the experiments; Qingshan You performed the experiments; Qingshan You and Qun Wan analyzed the data; Qingshan You contributed reagents/materials/analysis tools; Qingshan You wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Fazel, M. Matrix Rank Minimization with Applications. Ph.D. Thesis, Stanford University, Stanford, CA, USA, 2002.
2. Candès, E.J.; Recht, B. Exact matrix completion via convex optimization. *Found. Comput. Math.* **2009**, *9*, 717–772.
3. Candès, E.J.; Plan, Y. Matrix completion with noise. *Proc. IEEE* **2010**, *98*, 925–936.
4. Candès, E.J.; Tao, T. The power of convex relaxation: Near-optimal matrix completion. *IEEE Trans. Inf. Theory* **2010**, *56*, 2053–2080.
5. Ellenberg, J. Fill in the blanks: Using math to turn lo-res datasets into hi-res samples. *Wired* **2010**. Available online: https://www.wired.com/2010/02/ff_algorithm/all/1 (accessed on 26 January 2016).
6. Wright, J.; Yang, A.; Ganesh, A.; Ma, Y.; Sastry, S. Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal.* **2009**, *31*, 210–227.
7. Chambolle, A.; Lions, P.L. Image recovery via total variation minimization and related problems. *Numer. Math.* **1997**, *76*, 167–188.
8. Claerbout, J.F.; Muir, F. Robust modeling of erratic data. *Geophysics* **1973**, *38*, 826–844.
9. Papadimitriou, C.; Raghavan, P.; Tamaki, H.; Vempala, S. Latent semantic indexing: A probabilistic analysis. In Proceedings of the Seventeenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, Seattle, WA, USA, 1–4 June 1998; Volume 61.
10. Argyriou, A.; Evgeniou, T.; Pontil, M. Convex multi-task feature learning. *Mach. Learn.* **2008**, *73*, 243–272.
11. Bouwmans, T.; Sobral, A.; Javed, S.; Jung, S.; Zahzah, E. Decomposition into Low-rank plus Additive Matrices for Background/Foreground Separation: A Review for a Comparative Evaluation with a Large-Scale Dataset. *Comput. Vis. Pattern Recognit.* **2016**, doi:10.1016/j.cosrev.2016.11.001.
12. Candès, E.J.; Li, X.; Ma, Y.; Wright, J. Robust principal component analysis? *J. ACM* **2009**, *58*, 1–37.
13. Cai, J.F.; Osher, S.; Shen, Z. Linearized Bregman Iterations for Compressed Sensing. *Math. Comp.* **2009**, *78*, 1515–1536.
14. Wright, J.; Ganesh, A.; Rao, S.; Ma, Y. Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization. **2009**, arXiv:0905.0233.

15. Cai, J.-F.; Canès, E.J.; Shen, Z. A singular value thresholding algorithm for matrix completion. *SIAM J. Optim.* **2008**, *20*, 1956–1982.
16. Zhang, H.; Cai, J.-F.; Cheng, L.; Zhu, J. Strongly Convex Programming for Exact Matrix Completion and Robust Principal Component Analysis. *Inverse Probl. Imaging* **2012**, *6*, 357–372.
17. Ganesh, A.; Min, K.; Wright, J.; Ma, Y. Principal Component Pursuit with Reduced Linear Measurements. **2012**, arXiv:1202.6445v1.
18. Ledoux, M. *The Concentration of Measure Phenomenon*; American Mathematical Society: Providence, RI, USA, 2001.



© 2017 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).