LOCAL AND GLOBAL OPTIMALITY OF LP MINIMIZATION FOR SPARSE RECOVERY

Laming Chen and Yuantao Gu[*]

State Key Laboratory on Microwave and Digital Communications
Tsinghua National Laboratory for Information Science and Technology
Department of Electronic Engineering, Tsinghua University, Beijing 100084, CHINA

ABSTRACT

In solving the problem of sparse recovery, non-convex techniques have been paid much more attention than ever before, among which the most widely used one is ℓ_p minimization with $p \in (0, 1)$. It has been shown that the global optimality of ℓ_p minimization is guaranteed under weaker conditions than convex ℓ_1 minimization, but little interest is shown in the local optimality, which is also significant since practical non-convex approaches can only get local optimums. In this work, we derive a tight condition in guaranteeing the local optimality of ℓ_p minimization. For practical purposes, we study the performance of an approximated version of ℓ_p minimization, and show that its global optimality is equivalent to that of ℓ_p minimization when the penalty approaches the ℓ_p "norm". Simulations are implemented to show the recovery performance of the approximated optimization in sparse recovery.

***Index Terms*—** Sparse recovery, ℓ_p minimization, non-convex optimization, local optimality, global optimality.

1. INTRODUCTION

In many important applications such as source localization [1], image denoising [2], and face recognition [3], the key issues can be viewed as finding sparse solutions to underdetermined systems of linear equations. Mathematically, suppose $\mathbf{x}^* \in \mathbb{R}^N$ is an unknown sparse signal that we wish to recover, and it is observed through a sensing matrix $\mathbf{A} \in \mathbb{R}^{M\times N}$ with $M < N$,

$$\mathbf{y} = \mathbf{A}\mathbf{x}^*. \tag{1}$$

To determine the sparse signal from the sensing matrix $\mathbf{A}$ and the measurement vector $\mathbf{y}$, a natural idea is to adopt ℓ_0 minimization

$$\operatorname*{argmin}_{\mathbf{x}} \|\mathbf{x}\|_0 \ \text{ subject to } \ \mathbf{A}\mathbf{x} = \mathbf{y} \tag{2}$$

where $\|\mathbf{x}\|_0$ simply counts the nonzero entries of $\mathbf{x}$. Since ℓ_0 minimization is NP-hard [4], (2) is computationally intractable.

To tackle this problem, some works relax the non-convex discontinuous ℓ_0 "norm" to convex ℓ_1 norm [5,6] and show that, under some relaxed conditions, the optimum solutions to ℓ_0 minimization and ℓ_1 minimization

$$\operatorname*{argmin}_{\mathbf{x}} \|\mathbf{x}\|_1 \ \text{ subject to } \ \mathbf{A}\mathbf{x} = \mathbf{y} \tag{3}$$

are the same [7–9]. ℓ_1 minimization can be reformulated as a linear program and therefore can be solved efficiently [10].

This work was supported by National 973 Program of China (Grant No. 2013CB329203) and National Natural Science Foundation of China (NSFC 61371137). The corresponding author of this paper is Yuantao Gu (gyt@tsinghua.edu.cn).

Some other works try to find a tradeoff between the computational complexity and the recovery performance, and introduce ℓ_p minimization

$$\operatorname*{argmin}_{\mathbf{x}} \|\mathbf{x}\|_p^p \ \text{ subject to } \ \mathbf{A}\mathbf{x} = \mathbf{y} \tag{4}$$

where $\|\mathbf{x}\|_p = (\sum_{i=1}^N |x_i|^p)^{1/p}$ with $0 < p < 1$ [11–13]. It has been empirically shown that ℓ_p minimization tends to outperform convex ℓ_1 minimization in various aspects such as less number of measurements needed, larger sparsity level of signals allowed, and better denoising performance [12–16]. But since ℓ_p minimization is non-convex, its local optimality and global optimality should both be carefully studied.

The global optimality of ℓ_p minimization has been thoroughly studied before. Based on restricted isometry constant, sufficient conditions to guarantee the global optimality of ℓ_p minimization are obtained in a bunch of works [12, 13, 15]. Some other works derive tight conditions in indicting the performance with null space property [8, 11, 12, 17]. In [11], the global optimality of a class of non-convex optimizations besides ℓ_p minimization is derived, but their comparison with ℓ_p minimization is not discussed. When dealing with practical algorithms, [15] considers an approximated version of ℓ_p minimization, and shows that under the same sufficient condition as that guaranteeing the global optimality of ℓ_p minimization, the approximated optimization also returns the sparse signal. But since this work only derives the sufficient condition, it is still unclear whether the approximation would result in performance degeneration.

The main contribution of this paper is twofold. First, we discuss the local optimality of ℓ_p minimization, i.e., condition under which the sparse signal is the local optimum, which has rarely been considered in previous works. This is also of great importance since solving ℓ_p minimization is NP-hard and all feasible algorithms can only find local optimums [18]. Second, to avoid the infinite derivative of ℓ_p "norm" around the origin, we consider an approximated version of ℓ_p minimization, and study its performance equivalence with ℓ_p minimization in the aspect of global optimality.

2. PRELIMINARY

The sparse signal recovery problems introduced in Section 1 can be summarized as

$$\operatorname*{argmin}_{\mathbf{x}} J(\mathbf{x}) \ \text{ subject to } \ \mathbf{A}\mathbf{x} = \mathbf{y} \tag{5}$$

where $J(\mathbf{x}) = \sum_{i=1}^N F(|x_i|)$ and $F(\cdot)$ belongs to a class of sparseness measures [11], i.e., the following Definition 1.

Definition 1. (Definition 1 from [11]) The sparseness measure $F : [0, +\infty) \to [0, +\infty)$ satisfies

1) $F(0) = 0$ and $F(\cdot)$ is not identically zero;
2) $F(\cdot)$ is non-decreasing;
3) $F(x)/x$ is non-increasing in $x \in (0, +\infty)$.

Remark 1. It's easy to check that ℓ_0, ℓ_1, and ℓ_p minimizations are special cases of problem (5).

Some quantities are introduced in literatures to characterize the performance of sparse recovery problems and algorithms, such as spark [7], mutual coherence [19], restricted isometry constant [20], and null space constant [11, 17]. In this paper we adopt spark and null space constant introduced as follows. Let $\mathcal{N}(\mathbf{A})$ denote the null space of $\mathbf{A}$ and $\mathbf{z}_S$ be the vector generated by keeping the entries of $\mathbf{z}$ indexed by S and setting the rest entries to zeros.

Definition 2. (Definition 1 from [7]) The spark of a matrix $\mathbf{A}$, denoted as $\mathrm{Spark}(\mathbf{A})$, is the smallest number of columns from $\mathbf{A}$ that are linearly dependent.

Definition 3. (Summarized from [11, 17, 21]) For $J(\cdot)$ formed by $F(\cdot)$ satisfying Definition 1, define null space constant $\gamma(J, \mathbf{A}, K)$ as the smallest quantity such that

$$J(\mathbf{z}_S) \le \gamma(J, \mathbf{A}, K) J(\mathbf{z}_{S^c}) \tag{6}$$

holds for any set $S \subseteq \{1, 2, \ldots, N\}$ with $\#S \le K$ and for any vector $\mathbf{z} \in \mathcal{N}(\mathbf{A})$.

Based on null space constant, the global optimality of problem (5) can be characterized as follows.

Proposition 1. (Theorem 2, 3, and 5 from [11]) For $J(\cdot)$ formed by $F(\cdot)$ satisfying Definition 1,

1) Assume $\mathbf{x}^*$ is K-sparse and $\mathbf{y} = \mathbf{A}\mathbf{x}^*$ with $\mathbf{A}$ satisfying $\gamma(J, \mathbf{A}, K) < 1$. Then $\mathbf{x}^*$ is the global optimum to (5).
2) Assume that the sensing matrix $\mathbf{A}$ satisfies $\gamma(J, \mathbf{A}, K) > 1$. Then there exists a K-sparse signal $\mathbf{x}^*$ and $\mathbf{y} = \mathbf{A}\mathbf{x}^*$ such that $\mathbf{x}^*$ is not the global optimum to (5).
3) $\gamma(\ell_0, \mathbf{A}, K) \le \gamma(J, \mathbf{A}, K) \le \gamma(\ell_1, \mathbf{A}, K)$.

Remark 2. According to Proposition 1.1)-2), the null space constant is a tight quantity in indicating the global optimality of problem (5), especially ℓ_0, ℓ_1, and ℓ_p minimizations.

Remark 3. Proposition 1.3) reveals that among all problems with the form (5), the global optimality of ℓ_0 minimization is guaranteed under the weakest condition while that of ℓ_1 minimization is opposite.

3. MAIN CONTRIBUTION

In this section, we introduce the main theoretical contributions of this paper including the local optimality of ℓ_p minimization and the global optimality of an approximated version of ℓ_p minimization. To begin with, we consider a slightly more generalized class of penalties than ℓ_p "norm".

Definition 4. Let $\mathcal{J}$ be the set of penalties $J(\cdot)$ formed by $F(\cdot)$ satisfying Definition 1 and that $F(x)/x$ tends to positive infinity as x approaches zero.

Remark 4. It's easy to check that $\|\cdot\|_0 \in \mathcal{J}$ and $\|\cdot\|_p^p \in \mathcal{J}$ with $0 < p < 1$, but $\|\cdot\|_1 \notin \mathcal{J}$.

When discussing the performance of ℓ_p minimization $(0 < p < 1)$, most if not all works mainly consider the global optimality and neglect the local optimality. The latter is also quite important since any feasible non-convex algorithm for ℓ_p minimization can only find its local optimum [18]. The following two theorems fill this research gap by deriving a necessary and sufficient condition.

Theorem 1. Assume $\mathbf{x}^*$ is K-sparse and $\mathbf{y} = \mathbf{A}\mathbf{x}^*$ with $\mathbf{A}$ satisfying that $\mathrm{Spark}(\mathbf{A}) \ge K + 1$. Then for any $J(\cdot) \in \mathcal{J}$, $\mathbf{x}^*$ is a local optimum to problem (5).

Proof. See Section 4.1. □

Theorem 2. Assume that the sensing matrix $\mathbf{A}$ satisfies $\mathrm{Spark}(\mathbf{A}) = K$. Then there exists a K-sparse signal $\mathbf{x}^*$, if $\mathbf{y} = \mathbf{A}\mathbf{x}^*$, then for any $J(\cdot)$ formed by $F(\cdot)$ satisfying Definition 1, $\mathbf{x}^*$ is not a local optimum to problem (5).

Proof. See Section 4.2. □

Remark 5. According to Theorem 1 and Theorem 2, $\mathrm{Spark}(\mathbf{A}) \ge K + 1$ is a necessary and sufficient condition for the local optimality of both ℓ_0 and ℓ_p minimizations $(0 < p < 1)$. An intuitive explanation for this is that ℓ_p "norm" looks locally similar to ℓ_0 "norm" at the origin.

Remark 6. According to Proposition 1.1)-2), the necessary and sufficient condition for the global optimality of ℓ_0 minimization is $\gamma(\ell_0, \mathbf{A}, K) < 1$, which is equivalent to $\mathrm{Spark}(\mathbf{A}) \ge 2K + 1$[1]. Therefore, for non-convex ℓ_0 and ℓ_p minimizations, local optimality is strictly weaker than global optimality, which means that even when the sparse signal is not the global optimum, you still get a chance to find it as long as you start with a sufficiently good initialization. This differs from convex ℓ_1 minimization obviously.

Though the global optimality of ℓ_p minimization is guaranteed under weaker conditions than ℓ_1 minimization [17], the infinite derivative of ℓ_p "norm" around the origin may cause trouble in designing effective and robust algorithms. To address this problem, smoothed version of ℓ_p "norm" is adopted, for example,

$$F_{p,\sigma}(|x|) = \frac{|x|}{(|x| + \sigma)^{1-p}} \xrightarrow{\sigma \to 0^+} |x|^p, \quad p \in [0, 1). \tag{7}$$

When $p = 0$, this approximation is related with reweighted ℓ_1 minimization [24] where $|x|$ in the denominator of (7) is approximated by the previous estimate. When $0 < p < 1$, this approximation is connected to the reweighted algorithm introduced in [15] where $|x|$ in the denominator is also approximated as aforementioned. The global optimality of the corresponding approximated optimization can be revealed by the following theorem.

Theorem 3. Assume $\mathbf{x}^*$ is K-sparse and $\mathbf{y} = \mathbf{A}\mathbf{x}^*$ with $\mathbf{A}$ satisfying $\gamma(\ell_p, \mathbf{A}, K) < 1$, $p \in [0, 1)$. Then the global optimum $\mathbf{x}^\sigma$ of the problem

$$\underset{\mathbf{x}}{\mathrm{argmin}} J_{p,\sigma}(\mathbf{x}) := \sum_{i=1}^{N} F_{p,\sigma}(|x_i|) \text{ subject to } \mathbf{A}\mathbf{x} = \mathbf{y} \tag{8}$$

satisfies $\lim_{\sigma \to 0^+} \|\mathbf{x}^\sigma - \mathbf{x}^*\|_2 = 0$.

Proof. See Section 4.3. □

Remark 7. Theorem 2 in [21] and Theorem 3 above are two important results in exhibiting the performance of problem (8) from two different perspective of views: the former indicates the performance of (8) for $(\mathbf{A}, K)$, while the latter is for $(\mathbf{A}, \mathbf{x}^*)$. Specifically, Theorem 2 in [21] reveals that for *fixed* $\sigma > 0$, $\gamma(J_{p,\sigma}, \mathbf{A}, K) = \gamma(\ell_1, \mathbf{A}, K)$. This means that guaranteeing *any* K-sparse signal to

[1]The condition $\mathrm{Spark}(\mathbf{A}) \ge 2K + 1$ has been shown to imply the global optimality of ℓ_0 minimization in many works such as [7, 22, 23].

be the global optimum to (8) is equivalent to guaranteeing *any* K-sparse signal to be the global optimum to ℓ_1 minimization. But notice that the condition in Theorem 3 is $\gamma(\ell_p, \mathbf{A}, K) < 1$, which reveals that for fixed $\mathbf{x}^*$, decreasing σ would indeed improve the recovery performance, which approaches the performance of ℓ_p minimization in the limiting scenario.

Remark 8. A similar result is presented in Proposition 4.3 of [15] with two main differences. First, they consider a different approximated penalty $\|\mathbf{x}\|_p^p \approx \sum_{i=1}^N (|x_i| + \epsilon)^p$. Second, their condition is based on restricted isometry constant which is a sufficient condition, while in our result the condition is based on null space constant which is a tight quantity in characterizing the global optimality.

4. PROOF

4.1. Proof of Theorem 1

Proof. We need to prove that there exists a neighborhood $\mathcal{B}(\mathbf{x}^*)$ of $\mathbf{x}^*$ such that for any $\mathbf{x} \in \mathcal{B}(\mathbf{x}^*) \setminus \{\mathbf{x}^*\}$ satisfying $\mathbf{y} = \mathbf{A}\mathbf{x}$, $J(\mathbf{x}) > J(\mathbf{x}^*)$. Define $m = \min\{|x_i^*| : x_i^* \neq 0\}$ as the smallest nonzero magnitude of the entries of $\mathbf{x}^*$, and $\sigma_{\min}^K(\mathbf{A})$ as the smallest singular value taken over all K-column submatrices of $\mathbf{A}$. Since $\mathrm{Spark}(\mathbf{A}) \geq K + 1$, any K columns of $\mathbf{A}$ are linearly independent, and $\sigma_{\min}^K(\mathbf{A}) > 0$. Since $F(x)/x$ is non-increasing in $x \in (0, +\infty)$ and tends to positive infinity as x approaches zero, there exists $r \in (0, m)$ such that

$$F(r)/r > (F(m)/m)(\sqrt{K}\|\mathbf{A}\|_2/\sigma_{\min}^K(\mathbf{A})). \tag{9}$$

We choose $\mathcal{B}(\mathbf{x}^*)$ as the ℓ_2-ball centered at $\mathbf{x}^*$ with radius r.

For any $\mathbf{x} \in \mathcal{B}(\mathbf{x}^*) \setminus \{\mathbf{x}^*\}$ satisfying $\mathbf{y} = \mathbf{A}\mathbf{x}$, let $\mathbf{z} = \mathbf{x} - \mathbf{x}^*$. Then $\|\mathbf{z}\|_2 \leq r$ and $\mathbf{z} \in \mathcal{N}(\mathbf{A}) \setminus \{\mathbf{0}\}$. Defining set $T = \{i : x_i^* \neq 0\}$ as the support set of $\mathbf{x}^*$, it's easy to see that

$$J(\mathbf{x}) - J(\mathbf{x}^*) = \sum_{i \notin T} F(|z_i|) + \sum_{i \in T} (F(|x_i^* + z_i|) - F(|x_i^*|)). \tag{10}$$

First, since $|z_i| \leq \|\mathbf{z}\|_2 \leq r$ and $F(x)/x$ is non-increasing in $x \in (0, +\infty)$, $F(|z_i|) \geq (F(r)/r)|z_i|$, therefore

$$\sum_{i \notin T} F(|z_i|) \geq (F(r)/r)\|\mathbf{z}_{T^c}\|_1. \tag{11}$$

Second, for $i \in T$, since $|z_i| \leq r < m \leq |x_i^*|$, the second term on the right hand side of (10) is no less than

$$\begin{aligned} -\sum_{i \in T}(F(|x_i^*|) - F(|x_i^*| - |z_i|)) &\geq -\sum_{i \in T}(F(|x_i^*|)/|x_i^*|)|z_i| \\ &\geq -(F(m)/m)\|\mathbf{z}_T\|_1. \end{aligned} \tag{12}$$

Substituting (11) and (12) into (10) yields

$$J(\mathbf{x}) - J(\mathbf{x}^*) \geq (F(r)/r)\|\mathbf{z}_{T^c}\|_1 - (F(m)/m)\|\mathbf{z}_T\|_1. \tag{13}$$

Since $\mathbf{A}\mathbf{z} = \mathbf{0}$, the equality $\mathbf{A}\mathbf{z}_T = -\mathbf{A}\mathbf{z}_{T^c}$ holds and hence $\|\mathbf{A}\mathbf{z}_T\|_2 = \|\mathbf{A}\mathbf{z}_{T^c}\|_2$. It can be calculated that

$$\|\mathbf{A}\mathbf{z}_T\|_2 \geq \sigma_{\min}^K(\mathbf{A})\|\mathbf{z}_T\|_2 \geq \sigma_{\min}^K(\mathbf{A})\|\mathbf{z}_T\|_1/\sqrt{K}, \tag{14}$$

$$\|\mathbf{A}\mathbf{z}_{T^c}\|_2 \leq \|\mathbf{A}\|_2\|\mathbf{z}_{T^c}\|_2 \leq \|\mathbf{A}\|_2\|\mathbf{z}_{T^c}\|_1. \tag{15}$$

Since any K columns of $\mathbf{A}$ are linearly independent, $\mathbf{z}$ is at least $(K+1)$-sparse, and $\mathbf{z}_{T^c} \neq \mathbf{0}$. Therefore, according to (13), (14), (15), and the definition of r, it can be derived that

$$\begin{aligned} &J(\mathbf{x}) - J(\mathbf{x}^*) \\ &\geq \|\mathbf{z}_{T^c}\|_1 \left(F(r)/r - (F(m)/m)(\|\mathbf{z}_T\|_1/\|\mathbf{z}_{T^c}\|_1)\right) \\ &\geq \|\mathbf{z}_{T^c}\|_1 \left(F(r)/r - (F(m)/m)(\sqrt{K}\|\mathbf{A}\|_2/\sigma_{\min}^K(\mathbf{A}))\right) \\ &> 0. \end{aligned}$$

To sum up, we have proved that $\mathbf{x}^*$ is a local optimum to (5). □

4.2. Proof of Theorem 2

Proof. Since $\mathrm{Spark}(\mathbf{A}) = K$, $\mathbf{A}$ has K linearly dependent columns, and there exists a K-sparse vector $\mathbf{z} \in \mathcal{N}(\mathbf{A}) \setminus \{\mathbf{0}\}$. Set $\mathbf{x}^* = \mathbf{z}$, then $\mathbf{y} = \mathbf{A}\mathbf{x}^* = \mathbf{0}$. For any $\lambda \in (0, 1)$, $\mathbf{y} = \mathbf{A}(\mathbf{x}^* - \lambda\mathbf{z})$, and

$$J(\mathbf{x}^* - \lambda\mathbf{z}) = \sum_{i=1}^N F((1-\lambda)|x_i^*|) \leq \sum_{i=1}^N F(|x_i^*|) = J(\mathbf{x}^*). \tag{16}$$

It is obvious that $\mathbf{x}^*$ is not a local optimum to (5). □

4.3. Proof of Theorem 3

Proof. First, we prove the result for $p \in (0, 1)$. For any $\varepsilon > 0$, define

$$\sigma_0 = \frac{\varepsilon}{\left(\frac{1+\gamma(\ell_p, \mathbf{A}, K)}{1-\gamma(\ell_p, \mathbf{A}, K)} N\right)^{1/p}} > 0. \tag{17}$$

Then for any $\sigma \leq \sigma_0$, let $\mathbf{x}^\sigma$ be the global optimum of problem (8), and define I^σ as the set of index i satisfying $|x_i^\sigma| \geq \sigma_0$. Due to the global optimality of $\mathbf{x}^\sigma$,

$$J_{p,\sigma}(\mathbf{x}^\sigma) \leq J_{p,\sigma}(\mathbf{x}^*) \leq \|\mathbf{x}^*\|_p^p. \tag{18}$$

It is easy to see that

$$J_{p,\sigma}(\mathbf{x}^\sigma) = \sum_{i=1}^N F_{p,\sigma}(|x_i^\sigma|) \geq \sum_{i \in I^\sigma} F_{p,\sigma_0}(|x_i^\sigma|). \tag{19}$$

For any $i \in I^\sigma$, since $|x_i^\sigma| \geq \sigma_0$,

$$\frac{|x_i^\sigma|}{(|x_i^\sigma| + \sigma_0)^{1-p}} = \frac{|x_i^\sigma|^p}{(1 + \frac{\sigma_0}{|x_i^\sigma|})^{1-p}} \geq \left(1 - \frac{\sigma_0}{|x_i^\sigma|}\right)|x_i^\sigma|^p \tag{20}$$

holds, therefore,

$$\sum_{i \in I^\sigma} F_{p,\sigma_0}(|x_i^\sigma|) \geq \sum_{i \in I^\sigma} |x_i^\sigma|^p - \sigma_0 \sum_{i \in I^\sigma} |x_i^\sigma|^{p-1}. \tag{21}$$

Combining (18), (19), and (21), we can derive that

$$\|\mathbf{x}^\sigma\|_p^p - \|\mathbf{x}^*\|_p^p \leq \sum_{i \notin I^\sigma} |x_i^\sigma|^p + \sigma_0 \sum_{i \in I^\sigma} |x_i^\sigma|^{p-1}. \tag{22}$$

According to the definition of I^σ and $p \in (0, 1)$, the right hand side of (22) is less than or equal to

$$(N - \#I^\sigma)\sigma_0^p + \sigma_0 \#I^\sigma \sigma_0^{p-1} = N\sigma_0^p. \tag{23}$$

As for the left hand side of (22), define T as the support set of $\mathbf{x}^*$ and $\mathbf{z}^\sigma = \mathbf{x}^\sigma - \mathbf{x}^* \in \mathcal{N}(\mathbf{A})$. Then the left hand side of (22) equals

$$\begin{aligned} \|\mathbf{x}_{T^c}^\sigma\|_p^p + \left(\|\mathbf{x}_T^\sigma\|_p^p - \|\mathbf{x}_T^*\|_p^p\right) &\geq \|\mathbf{z}_{T^c}^\sigma\|_p^p - \|\mathbf{z}_T^\sigma\|_p^p \\ &\geq \frac{1 - \gamma(\ell_p, \mathbf{A}, K)}{1 + \gamma(\ell_p, \mathbf{A}, K)}\|\mathbf{z}^\sigma\|_p^p \end{aligned} \tag{24}$$

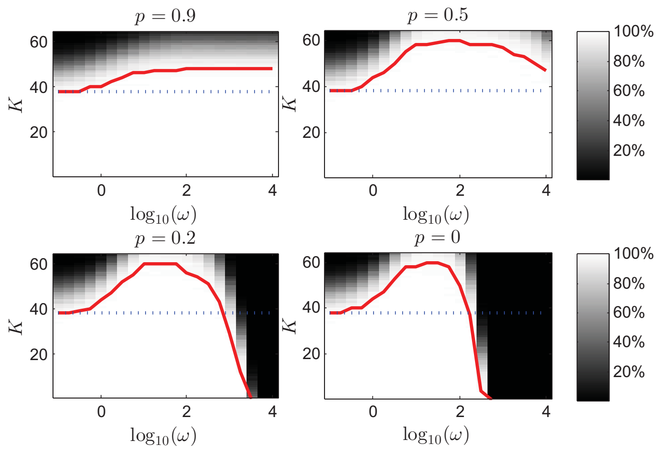

Fig. 1. The figure shows the successful recovery probability of (8) using PGG [21] with different p, non-convexity $\omega = (1-p)/\sigma$, and sparsity level K. The solid line represents the maximum sparsity level that guarantees 99% successful recovery versus ω. The dotted line indicates the maximum sparsity level that guarantees 99% successful recovery of ℓ_1 minimization, which serves as a benchmark.

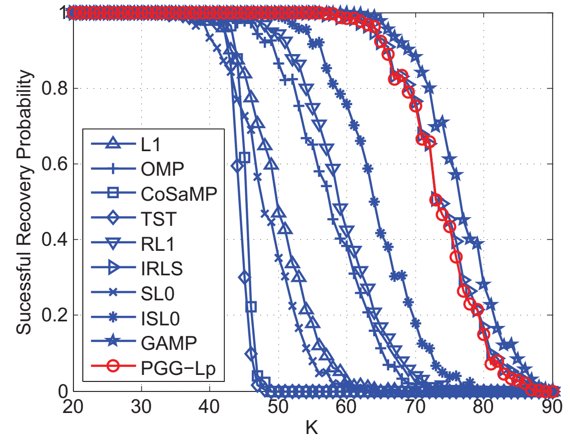

Fig. 2. The figure demonstrates the recovery probability of different algorithms versus sparsity level K.

where the last inequality is due to the definition of null space constant $\gamma(\ell_p, \mathbf{A}, K)$. Therefore,

$$\|\mathbf{z}^\sigma\|_p^p \leq \frac{1+\gamma(\ell_p, \mathbf{A}, K)}{1-\gamma(\ell_p, \mathbf{A}, K)} N\sigma_0^p. \quad (25)$$

According to Lemma 4.5 in [17], since $0 < p < 2$,

$$\|\mathbf{z}^\sigma\|_2 \leq \|\mathbf{z}^\sigma\|_p \leq \left(\frac{1+\gamma(\ell_p, \mathbf{A}, K)}{1-\gamma(\ell_p, \mathbf{A}, K)} N\right)^{1/p} \sigma_0 = \varepsilon. \quad (26)$$

Therefore, we have proved that when $p \in (0, 1)$, for any $\varepsilon > 0$, there exists $\sigma_0 > 0$ such that for any $\sigma \leq \sigma_0$, $\|\mathbf{x}^\sigma - \mathbf{x}^*\|_2 \leq \varepsilon$, which implies the result. As for the case when $p = 0$, similar argument as the proof of Theorem 1 in [21] also leads to the desired result. □

5. NUMERICAL SIMULATION

We demonstrate two experiments to test the recovery performance of optimization (8). The algorithm we adopt is the projected generalized gradient (PGG) method proposed in [21]. Initialized as least squares solution, in each iteration PGG first update along the negative gradient of penalty with step size κ, and then projected to the affine subspace $\{\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{y}\}$. To accelerate convergence with favourable accuracy, we vary step size κ dynamically: initialized as 10^{-2}, reduced by a half every 100 iterations, until less than 10^{-6}. The sensing matrix $\mathbf{A}$ is of size 128×256, with i.i.d. Gaussian entries of zero mean and variance $1/128$. The nonzero entries of sparse signal $\mathbf{x}^*$ are i.i.d. Gaussian with zero mean, and $\mathbf{x}^*$ is normalized to have unit ℓ_2 norm.

The first experiment tests the recovery performance of optimization (8) using PGG with different choices of p and σ. Following [21], the non-convexity of the penalty is defined as $\omega = (1-p)/\sigma$. Therefore, with fixed p, $\sigma \to 0^+$ is equivalent to $\omega \to +\infty$. Fig. 1 shows the successful recovery probability. The non-convexity ω increases from 10^{-1} to 10^4 with common ratio $10^{0.25}$, and the sparsity level K increases from 1 to 64 with step size 1. If the recovery SNR is higher than 40dB, this recovery is regarded as a success. The simulation is repeated 500 times to calculate the recovery probability. The solid line represents the maximum sparsity level that guarantees 99% successful recovery versus ω (or equivalently, σ). The dotted line indicates the maximum sparsity level that guarantees 99% successful recovery of ℓ_1 minimization, which serves as a benchmark. As can be seen, when non-convexity increases, its recovery performance increases from that of ℓ_1 minimization, and decreases when the non-convexity is sufficiently large. This coincides with Theorem 3 in [21] that the non-convexity should be smaller than a threshold to guarantee the convergence of algorithm. Among all the four choices of p, $p = 0.5$ is less sensitive to the choice of ω while still maintains good recovery performance.

In the second experiment, the recovery performance of optimization (8) using PGG is compared with some reference algorithms, including ℓ_1-magic [6], OMP [25], CoSaMP [26], TST [27], RL1 [24], IRLS [14], SL0 [28], ISL0 [29], and GAMP [30]. For PGG solving (8), set $p = 0.5$ and non-convexity $\omega = (1-p)/\sigma = 10^{1.75}$. The parameters of reference algorithms are set as recommended, while CoSaMP and GAMP further require the true sparsity level K, and GAMP additionally needs the distributions of both sensing matrix and nonzero entries of $\mathbf{x}^*$. The recovery probability of different algorithms versus K is demonstrated in Fig. 2. IRLS and PGG share similar performance since they both set $p = 0.5$, though different approximations and updating strategies are adopted. It's not surprising that GAMP is the best since it needs the most information.

6. CONCLUSION

For the rarely studied local optimality of ℓ_p minimization, this paper derives a tight condition in guaranteeing any K-sparse signal to be a local optimum of ℓ_p minimization, which differs significantly from that of convex ℓ_1 minimization. For the approximated version of ℓ_p minimization, its recovery performance is shown to approach that of ℓ_p minimization in the limiting scenario. Experiments are implemented to test the recovery performance of the approximated optimization under different parameter settings, and it's among the best compared with other reference algorithms.

7. REFERENCES

[1] D. Malioutov, M. Cetin, and A. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3010-3022, Aug. 2005.

[2] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3736-3745, Dec. 2006.

[3] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210-227, Feb. 2009.

[4] B. Natarajan, "Sparse approximate solutions to linear systems," *SIAM Journal on Computing*, vol. 24, no. 2, pp. 227-234, 1995.

[5] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33-61, 1998.

[6] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207-1223, Aug. 2006.

[7] D. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization," *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197-2202, Mar. 2003.

[8] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Transactions on Information Theory*, vol. 49, no. 12, pp. 3320-3325, Dec. 2003.

[9] E. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 9-10, pp. 589-592, May 2008.

[10] S. Boyd and L. Vandenberghe, Convex Optimization, 2009: Cambridge University Press.

[11] R. Gribonval and M. Nielsen, "Highly sparse representations from dictionaries are unique and independent of the sparseness measure," *Applied and Computational Harmonic Analysis*, vol. 22, no. 3, pp. 335-355, May 2007.

[12] R. Chartrand, "Nonconvex compressed sensing and error correction," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. III-889-III-892, 2007.

[13] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Processing Letters*, vol. 14, no. 10, pp. 707-710, Oct. 2007.

[14] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 3869-3872, 2008.

[15] S. Foucart and M. Lai, "Sparsest solutions of underdetermined linear systems via ℓ_q-minimization for $0 < q \leq 1$," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 395-407, May 2009.

[16] R. Saab and Ö. Yilmaz, "Sparse recovery by non-convex optimization - instance optimality," *Applied and Computational Harmonic Analysis*, vol. 29, no. 1, pp. 30-48, July 2010.

[17] S. Foucart. Notes on compressed sensing for Math 394, Spring 2009[Online].

[18] D. Ge, X. Jiang, and Y. Ye, "A note on the complexity of L_p minimiation," *Mathematical Programming*, vol. 129, no. 2, pp. 285-299, June 2011.

[19] D. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 2845-2862, Nov. 2001.

[20] E. Candès and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203-4215, Dec. 2005.

[21] L. Chen and Y. Gu, "The convergence guarantees of a nonconvex approach for sparse recovery," *IEEE Transactions on Signal Processing*, vol. 62, no. 15, pp. 3754-3767, Aug. 2014.

[22] I. Gorodnitsky and B. Rao, "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 600-616, Mar. 1997.

[23] A. Bruckstein, D. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Review*, vol. 51, no. 1, pp. 34-81, 2009.

[24] E. Candès, M. Wakin, and S. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5-6, pp. 877-905, Dec. 2008.

[25] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655-4666, Dec. 2007.

[26] D. Needell and J. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301-321, May 2009.

[27] A. Maleki and D. Donoho, "Optimally tuned iterative reconstruction algorithms for compressed sensing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 330-341, Apr. 2010.

[28] H. Mohimani, M. Babaie-Zadeh, and C. Jutten, "A fast approach for overcomplete sparse decomposition based on smoothed ℓ^0 norm," *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 289-301, Jan. 2009.

[29] M. Hyder and K. Mahata, "An improved smoothed ℓ^0 approximation algorithm for sparse representation," *IEEE Transactions on Signal Processing*, vol. 58, no. 4, pp. 2194-2205, Apr. 2010.

[30] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," *IEEE International Symposium on Information Theory Proceedings*, pp. 2168-2172, 2011.