Mathematical Programming manuscript No.
(will be inserted by the editor)

A Note on the Complexity of L_p Minimization

Dongdong Ge · Xiaoye Jiang · Yinyu Ye

Abstract We discuss the L_p $(0 \leq p < 1)$ minimization problem arising from sparse solution construction and compressed sensing. For any fixed $0 < p < 1$, we prove that finding the global minimal value of the problem is strongly NP-Hard, but computing a local minimizer of the problem can be done in polynomial time. We also develop an interior-point potential reduction algorithm with a provable complexity bound and demonstrate preliminary computational results of effectiveness of the algorithm.

Keywords nonconvex programming · global optimization · interior-point method · sparse solution reconstruction
MSC2010 Classification Code: 90C26, 90C51

1 Introduction

In this note, we consider the following optimization problems:

$$\begin{array}{lll} \text{(P)} & \text{Minimize} & p(x) := \sum_{j=1}^{n} x_j^p \\ & \text{Subject to} & x \in \mathcal{F} := \{x : Ax = b, x \geq 0\}, \end{array} \tag{1}$$

and

$$\begin{array}{ll} \text{Minimize} & \sum_{j=1}^{n} |x_j|^p \\ \text{Subject to} & x \in \mathcal{F}' := \{x : Ax = b\}, \end{array} \tag{2}$$

where the problem inputs consist of $A \in R^{m \times n}, b \in R^m$, and $0 < p < 1$.

Sparse signal or solution reconstruction by solving optimization problem (1) or (2), especially for the cases of $0 < p \leq 1$, has recently received considerable attention; for example, see [6, 22] and references therein. In signal reconstruction, one typically has linear measurements $b = Ax$ where x is a sparse signal, i.e., the sparsest or smallest support cardinality solution of the linear system. This sparse signal is recovered by solving the inverse problem (1) or (2) with the objective function $\|x\|_0$ that is the L_0 norm of x and is defined as the number of nonzero components of x [7]. In this note we essentially consider the L_p norm functional $L_p(x) = \left(\sum_{j=1}^{n} |x_j|^p\right)^{1/p}$ where $0 < p \leq 1$ as our optimization objective function. It is easy to see that $L_p^p(x)$ converges to the L_0 norm functional on a bounded set $\Omega \subset R^n$. Thus, as a potential approach to the sparse signal problem, minimizing the L_p norm function of x, or simply $p(x)$, naturally arises.

Dongdong Ge
Antai School of Economics and Management, Shanghai Jiao Tong University, Shanghai, China 200052.
E-mail: ddge@sjtu.edu.cn

Xiaoye Jiang
Institute for Computational and Mathematical Engineering, Stanford University, Stanford, CA 94305
E-mail: xiaoye@stanford.edu

Yinyu Ye
Department of Management Science and Engineering, Stanford University, Stanford, CA 94305
E-mail: yinyu-ye@stanford.edu

In terms of computational complexity, the problem with L_0 norm is shown to be NP-hard [19]. On the other hand, when $p = 1$, the problem (1) or (2), which is a relaxation problem for the L_0 norm problem, is a linear program, and hence it is solvable in polynomial time. For more general convex programs involving the L_p norm ($p > 1$) in the objective (or constraints) and in conic representation, readers are referred to [20] and [25].

In [7,12], it was shown that if a certain *restricted isometry property*(RIP) holds for A, then the solutions of L_p norm minimization for $p = 0$ and $p = 1$ are identical. Hence, problem L_0 minimization can be relaxed to problem (2) with $p = 1$. However, this property may be too strong for practical basis design matrices A. Instead, one may consider the sparse recovery problem by solving relaxation problem (1) or (2) for a fixed p with $0 < p < 1$. Recently, this approach has attracted many research efforts in variable selection and sparse reconstruction, e.g., [9,10,13,17]. This approach exhibits desired threshold bounds on any non-zero entry of a computed solution [11], and computational experiments show that by replacing $p = 1$ with $p < 1$, reconstruction can be done equally fast with many fewer measurements while maintaining robustness to noise and signal non-sparsity [10].

In this paper we present several interesting properties of the L_p ($0 < p < 1$) minimization problem. Minimization problems (1) and (2) are both *strongly* NP-Hard. However, any basic feasible solution of (1) or (2) is a local minimizer. Moreover, a feasible point satisfying the first order and second order necessary optimality conditions is always a local minimizer. This motivates us to design interior-point algorithms, which iterate within the *interior* of the feasible region, to generate a local minimizer (hopefully a sparse local minimizer) satisfying the Karush-Kuhn-Tucker (KKT) optimality conditions .

1.1 Notation and Preliminaries

For the simplicity of our analysis, throughout this paper we assume that the feasible set $\mathcal{F}$ is bounded and A is full-ranked.

A feasible point x is called a *local minimum point* or *local minimizer* of problem (1) if there exists $\epsilon > 0$ such that $p(x') \geq p(x)$ for any $x' \in B(x, \epsilon) \bigcap \mathcal{F}$ where $B(x, \epsilon) = \{x' : \| x' - x \|_2 \leq \epsilon\}$.

For any feasible solution x of (1), let $S(x)$ be the support or active set of x, that is,

$$S(x) = \{j : \ x_j > 0\}.$$

Thus, $|S| \leq n$. Let x^* be a local minimum point of problem (1). Although $p(x^*)$ is not differentiable when x^* is on the boundary of the feasible region, it is not difficult to verify that its non-zero components must be local minimizers in the active set of variables, that is, for the problem

$$\min p_{S(x)}(z) := \sum_{j \in S(x)} z_j^p, \ \text{s.t.} \ A_{S(x)} z = b, \ z \geq 0,$$

where $A_S \in R^{m \times |S|}$ is the submatrix of A exactly consisting of the columns of A according to index set S (see [11], for example). Thus, $x^* \geq 0$ satisfies the following necessary conditions ([2]).

- The *first order necessary condition* or *KKT condition*: there exists a Lagrange multiplier vector $y^* \in R^m$, such that

$$p(x^*)_j^{p-1} - (A^T y^*)_j \geq 0, \ \forall j \in S(x^*), \tag{3}$$

  and the *complementarity* condition holds:

$$p(x^*)_j^p - x_j^* (A^T y^*)_j = 0, \ \forall j. \tag{4}$$

- The *second order necessary condition*:

$$\lambda^T \nabla_{zz}^2 p_{S(x^*)}(z^*) \lambda \geq 0, \tag{5}$$

  for all $\lambda \in R^{|S(x^*)|}$ such that $A_{S(x^*)}\lambda = 0$, where z^* is the vector of all non-zero components of x^*.

The fact that $\mathcal{F}$ is bounded and nonempty implies that $p(x)$ has a minimum value (denoted $\underline{z}$) and a maximum value (denoted $\bar{z}$). An ϵ-minimal solution or ϵ-minimizer is defined as a feasible solution x such that

$$\frac{p(x) - \underline{z}}{\bar{z} - \underline{z}} \leq \epsilon. \tag{6}$$

Vavasis [23] demonstrated the importance of the term $\bar{z} - \underline{z}$ in the criterion for continuous optimization. Similarly Ye [26] defined an ϵ-KKT (or ϵ-stationary) point as an (x, y) that satisfies (3) in the active set $S(x)$ of x, and the complementarity gap

$$\frac{\sum_{j=1}^{n}\left(p(x^*)_j^p - x_j^*(A^Ty)_j\right)}{\bar{z} - \underline{z}} \leq \epsilon. \tag{7}$$

This note is organized as follows: in section 2, we show that the L_p $(0 < p < 1)$ minimization problem (1) or (2) is *strongly* NP-Hard. In section 3 we prove that the set of all basic feasible solutions of (1) and (2) are identical with the set of all local minimizers. In section 4 we present our FPTAS (Fully Polynomial Time Approximation Scheme) interior-point potential reduction algorithm to find an ϵ-KKT point of problem (1). Numerical experiments are conducted to test its efficiency in section 5.

2 Hardness

To help the reader understand the basic idea of the hardness proof, we start by showing that the L_p $(0 < p < 1)$ minimization problems (1) and (2) are both NP-hard before we prove their *strong* NP-hardness. To prove HP-hardness, we employ a polynomial time reduction from the well known NP-complete *partition problem* [16]. The partition problem can be described as follows: given a set S of integers or rational numbers $\{a_1, a_2, \ldots, a_n\}$, is there a way to partition S into two disjoint subsets S_1 and S_2 such that the sum of the numbers in S_1 equals the sum of the numbers in S_2?

An instance of the partition problem can be reduced to an instance of the L_p $(0 < p < 1)$ minimization problem (1) that has the optimal value n if and only if the former has an equitable bipartition. Consider the following minimization problem:

$$\begin{array}{ll} \text{Minimize} & P(x,y) = \sum_{1\leq j\leq n} (x_j^p + y_j^p) \\ \text{Subject to} & a^T(x-y) = 0, \\ & x_j + y_j = 1, \ \forall j, \\ & x, y \geq 0. \end{array} \tag{8}$$

From the strict concavity of the objective function,

$$x_j^p + y_j^p \geq x_j + y_j (= 1), \ \forall j,$$

and the equality holds if and only if $(x_j = 1, y_j = 0)$ or $(x_j = 0, y_j = 1)$. Thus, $P(x, y) \geq n$ for any (continuous) feasible solution of problem (8).

If there is a feasible (optimal) solution pair (x, y) such that $P(x, y) = n$, it must be true that $x_j^p + y_j^p = 1 = x_j + y_j$ for all j so that (x, y) must be a binary solution $((x_j = 1, y_j = 0)$ or $(x_j = 0, y_j = 1))$, which generates an equitable bipartition of the entries of a. On the other hand, if the entries of a have an equitable bipartition, then the problem must have a binary solution pair (x, y) such that $P(x, y) = n$. Thus we prove the NP-hardness of problem (1).

Note that the objective value of (8) is a constant n when $p = 1$, so that any feasible solution is a (global) minimizer. However, the sparsest solution of (8) has cardinality n if the problem admits an equitable bipartition.

The same instance of the partition problem can also be reduced to the following minimization problem in form (2):

$$\begin{array}{ll} \text{Minimize} & \sum_{1\leq j\leq n} (|x_j|^p + |y_j|^p) \\ \text{Subject to} & a^T(x-y) = 0, \\ & x_j + y_j = 1, \ \forall j. \end{array} \tag{9}$$

Note that this problem has no non-negativity constraints on variables (x, y). However, for any feasible solution (x, y) of the problem, we still have

$$|x_j|^p + |y_j|^p \geq x_j + y_j (= 1), \ \forall j.$$

This is because the minimal value of $|x_j|^p + |y_j|^p$ is 1 if $x_j + y_j = 1$, and the quality holds if and only if $(x_j = 1, y_j = 0)$ or $(x_j = 0, y_j = 1)$. Therefore, the instance of the partition problem has an equitable bipartition if and only if the objective value of (9) is n. This leads to the NP-hardness of (2).

Now we prove a stronger result:

Theorem 1 *The L_p ($0 < p < 1$) minimization problems (1) and (2) are both strongly NP-hard.*

Proof We present a polynomial time reduction from the well known strongly NP-hard 3-partition problem [15,16]. The 3-partition problem can be described as follows: given a multiset S of $n = 3m$ integers $\{a_1, a_2, \dots, a_n\}$. The sum of S is equal to mB and each integer in S is strictly between $B/4$ and $B/2$. Can S be partitioned into m subsets, such that the sum of the numbers in each subset is equal to each other, i.e., B, which implies each subset has exactly three elements?

We describe a reduction from an instance of the 3-partition problem to an instance of the L_p ($0 < p < 1$) minimization problem (1) that has the optimal value n if and only if the former has an equitable 3-partition. Given an instance of the partition problem, let vector $a = (a_1, a_2, \dots, a_n) \in R^n$ and $n = 3m$. Let the sum of a be mB and each $a_i \in (B/4, B/2)$. Consider the following minimization problem in form (1):

$$\begin{array}{lll} \text{Minimize} & P(x) = \sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij}^p & \\ \text{Subject to} & \sum_{j=1}^m x_{ij} = 1, & i = 1, 2, \cdots, n, \\ & \sum_{i=1}^n a_i x_{ij} = B, & j = 1, 2, \cdots, m, \\ & x_{ij} \geq 0, & i = 1, 2, \cdots, n; j = 1, 2, \cdots, m, \end{array} \tag{10}$$

From the strict concavity of the objective function, and $x_{ij} \in [0, 1]$,

$$\sum_{j=1}^{m} x_{ij}^p \geq \sum_{j=1}^{m} x_{ij}(= 1), \ i = 1, 2, \cdots, n.$$

The equality holds if and only if $x_{ij_0} = 1$ for some j_0 and other $x_{ij} = 0$, $\forall j \neq j_0$. Thus, $P(x) \geq n$ for any feasible solution of (10).

If there is a feasible (optimal) solution x such that $P(x) = n$, it must be true that $\sum_{j=1}^m x_{ij}^p = 1 = \sum_{j=1}^m x_{ij}$ for all i so that for any i, $x_{ij_0} = 1$ for some j_0 and other $x_{ij} = 0$, $\forall j \neq j_0$. This generates an equitable 3-partition of the entries of a. On the other hand, if the entries of a have an equitable 3-partition, then (10) must have a binary solution x such that $P(x) = n$. Thus we prove the strong NP-hardness of problem (1) according to [15].

For the same instance of the 3-partition problem, we consider the following minimization problem in form (2):

$$\begin{array}{lll} \text{Minimize} & \sum_{i=1}^{n}\sum_{j=1}^{m} |x_{ij}|^p & \\ \text{Subject to} & \sum_{j=1}^m x_{ij} = 1, & i = 1, 2, \cdots, n, \\ & \sum_{i=1}^n a_i x_{ij} = B, & j = 1, 2, \cdots, m, \end{array} \tag{11}$$

Note that this problem has no non-negativity constraints on the variables x. However, for any feasible solution x of the problem, we still have

$$\sum_{j=1}^{m} |x_{ij}|^p \geq \sum_{j=1}^{m} x_{ij}(= 1), \ i = 1, 2, \cdots, n.$$

This is because the minimal value of $\sum_{j=1}^m |x_{ij}|^p$ is 1 if $\sum_{j=1}^m x_{ij} = 1$, and the equality holds if and only if for any i, $x_{ij_0} = 1$ for some j_0 and other $x_{ij} = 0$, $\forall j \neq j_0$.

Therefore, we can similarly prove that this instance of the partition problem has an equitable 3-partition if and only if the objective value of (11) is n, which leads to the strong NP-hardness of problem (2).

2.1 The Hardness of Smoothed L_p Minimization

To avoid the non-differentiability of $p(x)$, smooth versions of L_p minimization were developed:

$$\begin{array}{ll} \text{(P) Minimize} & \sum_{j=1}^{n} (x_j + \epsilon)^p \\ \text{Subject to} & x \in \mathcal{G} := \{x : Ax = b, x \geq 0\}, \end{array} \tag{12}$$

and

$$\begin{array}{ll} \text{Minimize} & \sum_{j=1}^{n} (|x_j| + \epsilon)^p \\ \text{Subject to} & x \in \mathcal{G}' := \{x : Ax = b\}, \end{array} \tag{13}$$

for a fixed positive constant ϵ; see [5,9–11].

A similar reduction can be used to derive the NP-hardness of the smoothed versions.

Theorem 2 *The smoothed L_p $(0 < p < 1)$ minimization problems (12) and (13) are both strongly NP-hard.*

Proof For problem (12), we also form a reduced minimization problem similar to the form of (10) for the 3-partition problem with the objective function defined as:

$$\begin{array}{lll} \text{Minimize} & P_\epsilon(x) = \sum_{i=1}^{n}\sum_{j=1}^{m}(x_{ij}+\epsilon)^p & \\ \text{Subject to} & \sum_{j=1}^{m} x_{ij} = 1, & i = 1, 2, \cdots, n, \\ & \sum_{i=1}^{n} a_i x_{ij} = B, & j = 1, 2, \cdots, m, \\ & x_{ij} \geq 0, & i = 1, 2, \cdots, n; j = 1, 2, \cdots, m. \end{array} \tag{14}$$

From the concavity of the objective function, for all i, we have

$$\begin{aligned} \sum_{j=1}^{m}(x_{ij}+\epsilon)^p &= \sum_{j=1}^{m}(x_{ij}+\epsilon\sum_{k=1}^{m} x_{ik})^p \\ &= \sum_{j=1}^{m}(x_{ij}(1+\epsilon)+\epsilon(1-x_{ij}))^p \\ &\geq \sum_{j=1}^{m}(x_{ij}(1+\epsilon)^p+\epsilon^p(1-x_{ij})) \\ &= (1+\epsilon)^p+(m-1)\epsilon^p \end{aligned}$$

The equality holds if and only if $x_{ij_0} = 1$ for some j_0 and $x_{ij} = 0$, $\forall j \neq j_0$. Thus, $P_\epsilon(x) \geq n((1+\epsilon)^p + (m-1)\epsilon^p)$ for any feasible solution of (14).

If there is a feasible (optimal) solution x such that $P_\epsilon(x) = n((1+\epsilon)^p+(m-1)\epsilon^p)$, it must be true that $\sum_{j=1}^{m}(x_{ij}+\epsilon)^p = (1+\epsilon)^p+(m-1)\epsilon^p$ for all i so that for any i, $x_{ij_0} = 1$ for some j_0 and $x_{ij} = 0$, $\forall j \neq j_0$. This generates an equitable 3-partition of the entries of a. On the other hand, if the entries of a have an equitable 3-partition, then (14) must have a binary solution x such that $P_\epsilon(x) = n((1+\epsilon)^p+(m-1)\epsilon^p)$. Thus we prove the strong NP-hardness of problem (14) and thereby (12). Similar arguments can also be developed to prove the strong NP-hardness of problem (13).

3 The Easiness

The above discussion reveals that finding a global minimizer for the L_p norm minimization problem is *strongly* NP-hard as long as $p < 1$. From a theoretical perspective a strongly NP-hard optimization problem with a polynomially bounded objective function does not admit an FPTAS unless P=NP [24]. Thus, relaxing $p = 0$ to some $p < 1$ gains no advantage in terms of the (worst-case) computational complexity. We now turn our attention to local minimizers. Note that, for many optimization problems, finding a local minimizer, or checking if a solution is a local minimizer, remains NP-hard. Here we show that local minimizers of problems (1) and (2) are easy to certify and compute.

Theorem 3 *The set of all basic feasible solutions of (1) or (2) is exactly the set of local minimizers. Thus, computing a local minimizer of (1) or (2) can be done in polynomial time.*

Proof Let $\bar{x}$ be a basic feasible solution (or extreme point) of the feasible polytope of (1), where, without loss of generality, the basic variables are $\bar{x}_B = (\bar{x}_1, \cdots, \bar{x}_m) > 0$ and $\bar{x}_j = 0,\ j = m+1, \cdots n$. The feasible directions form a polyhedral cone pointed at $\bar{x}$. Consider a directional edge of the feasible direction cone $d = d^{m+1} := (d_1, d_2, \cdots, d_{m+1}, 0, \cdots 0)$ from $\bar{x}$ to an adjacent extreme point. Let $\|d\|_2 = 1$. Since d is a feasible direction, there exists an appropriate and fixed $\epsilon_{m+1} > 0$, such that $\bar{x}+\epsilon d$ is feasible for any $0 < \epsilon < \epsilon_{m+1}$, which implies $d_{m+1} > 0$. Thus

$$p(\bar{x}+\epsilon d) - p(\bar{x}) = \sum_{i=1}^{m}\left((\bar{x}_i+\epsilon d_i)^p - \bar{x}_i^p\right) + (\epsilon d_{m+1})^p.$$

Define the index set $I^- = \{i : d_i < 0\}$. If I^- is empty, then $p(\bar{x} + \epsilon d) > p(\bar{x})$ for any $\epsilon > 0$ so that d is a strictly increasing feasible direction. If I^- is nonempty, one can choose sufficiently small but fixed ϵ'_{m+1} such that $\bar{x}_i + \epsilon d_i \geq \frac{\bar{x}_i}{2}$ for any $i \in I^-$ and $0 \leq \epsilon \leq \epsilon'_{m+1}$. Then, for such $0 < \epsilon \leq \epsilon'_{m+1}$

$$p(\bar{x} + \epsilon d) - p(\bar{x}) \geq \sum_{i \in I^-} ((\bar{x}_i + \epsilon d_i))^p - \bar{x}_i^p) + (\epsilon d_{m+1})^p > \sum_{i \in I^-} (\epsilon d_i) p(\frac{\bar{x}_i}{2})^{p-1} + (\epsilon d_{m+1})^p,$$

where the last inequality comes from the strict concavity of $p(x)$. Note that

$$\sum_{i \in I^-} (\epsilon d_i) p(\frac{\bar{x}_i}{2})^{p-1} + (\epsilon d_{m+1})^p > 0$$

if

$$\epsilon < \epsilon''_{m+1} := d_{m+1}^{\frac{p}{1-p}} \left(p \sum_{i \in I^-} (-d_i)(\frac{x_i}{2})^{p-1} \right)^{\frac{1}{p-1}}. \tag{15}$$

This again shows that the edge direction, $d = d^{m+1}$, is a strictly increasing direction within a fixed step size $\min\{\epsilon_{m+1}, \epsilon'_{m+1}, \epsilon''_{m+1}\} > 0$.

There are at most $(n - m)$ edge directions of the feasible direction cone, say d^j, $j = m + 1, \cdots n$. Thus, there exists a fixed $\epsilon_{\bar{x}} > 0$ such that $\bar{x} + \epsilon d^j$ is feasible and

$$p(\bar{x} + \epsilon d^j) - p(\bar{x}) > 0, \ \forall j = m + 1, \cdots n,$$

for all $0 < \epsilon \leq \epsilon_{\bar{x}}$. Let $\text{Conv}(\bar{x}, \epsilon_{\bar{x}})$ denote the convex hull spanned by points $\bar{x}$ and $\bar{x} + \epsilon_{\bar{x}} d^j$, $j = m+1, \cdots, n$. For any $x \in \text{Conv}(\bar{x}, \epsilon_{\bar{x}})$ and $x \neq \bar{x}$, we have $p(x) > p(\bar{x})$ by the strict concavity of $p(x)$ and that x can be represented as a convex combination of the corner points of the convex hull. Furthermore, one can always choose a sufficiently small but fixed $\epsilon > 0$, such that $B(\bar{x}, \epsilon) \bigcap \mathcal{F} \subset \text{Conv}(\bar{x}, \epsilon_{\bar{x}})$. Therefore, by the definition, $\bar{x}$ is a (strictly) local minimizer.

On the other hand, if $\bar{x}$ is a local minimizer but not a basic feasible solution or extreme point of the feasible polytope, then for any $\epsilon > 0$, consider the neighborhood $B(\bar{x}, \epsilon) \bigcap \mathcal{F}$. There must exist a feasible direction d with $\|d\|_2 = 1$ such that both $\bar{x} + \epsilon' d$ and $\bar{x} - \epsilon' d$ are feasible for sufficiently small $0 < \epsilon' < \epsilon$, and they both belong to $B(\bar{x}, \epsilon) \bigcap \mathcal{F}$. The strict concavity of $p(x)$ implies that one of two must be smaller than $p(\bar{x})$. Thus, $\bar{x}$ cannot be a local minimizer.

Similarly, we can prove that the set of all basic solutions of (2) is exactly the set of all of its local minimizers.

It is well known that computing a basic feasible solution of (1) or (2) can be done in polynomial time as solving a linear programming feasibility problem (e.g., [27]), which completes the proof.

Furthermore, one can observe the following property of a local minimizer.

Theorem 4 *If the first order necessary condition (3) and the second order necessary condition (5) hold at x and $x \in \mathcal{F}$, then x is a local minimizer of problem (1).*

Proof If x is in the interior of the feasible region $\mathcal{F}$ and satisfies the first order condition (3), the strict concavity of $p(x)$ implies x must be the unique maximum point.

If x lies on the boundary, define its support set S and positive vector z as in the second order necessary condition (5). Since $\nabla^2_{zz} p_S(z)$ is negative definite, we know that $\lambda^T \nabla^2_{zz}(z) \lambda$ cannot be nonnegative in the null space of A_S unless $\{\lambda : A_S \lambda = 0\} = \{0\}$. Thus, the second order necessary condition (5) holds only when x is a basic feasible solution. By Theorem 3, x is a local minimizer.

Example 1 This example shows that not all basic feasible solutions have a good sparse structure and the L_1 minimization does not always work. Let

$$A = \begin{pmatrix} 2 & 4 & 2 & 2 & 2 & 4 \\ 2 & 2 & 4 & 5 & 4 & 4 \\ 1 & 2 & 2 & 0 & 6 & 4 \end{pmatrix}, \quad b = \begin{pmatrix} 8 \\ 8 \\ 4 \end{pmatrix}.$$

It is not difficult to verify that the only optimal basic feasible solution for the L_1 minimization is $x = (0, 1.2, 0, 0.8, 0, 0.4)'$ and the optimal value is 2.4. However, it is not the sparsest solution. A sparsest solution can be given by $x = (4, 0, 0, 0, 0, 0)'$. One can also observe that any basic feasible solution containing column 1 is the sparsest solution, but other basic feasible solutions are all dense with cardinality 3.

4 Interior-Point Algorithm

From a complexity point of view, Theorem 3 implies that a local minimizer of the L_p minimization can be computed in polynomial time as a linear program to find a basic feasible solution. Of course, we are really interested in finding a sparse basic feasible solution; if we start from non-sparse basic feasible solution, there is little hope to find a sparser one by solving problem (1) since the starting point is already a strict local minimizer.

Naturally, one would start from an interior-point feasible solution such as an (approximate) analytic center x^0 of the feasible polytope (if it is bounded and has an interior feasible point), and consider an interior-point algorithm for approximately solving (1). Moreover, $p(x)$ is differentiable in the interior feasible region. Hopefully, the interior-point algorithm would generate a sequence of interior points that bypasses non-sparse basic feasible solutions and converges to a sparse one.

This is exactly the idea of the potential reduction algorithm developed in [26] for nonconvex quadratic programming. The algorithm starts from the initial point, follows an interior feasible path and finally converges to either a global minimizer or a KKT point or local minimizer. At each step, it chooses a new interior point which produces the maximal potential reduction to a potential function by an affine-scaling operation. See [18] or [27] for an extensive introduction of interior-point algorithms.

We now extend the potential reduction algorithm for the L_p minimization. The algorithm starts from an interior-point feasible solution such as the analytic center x^0 of the feasible polytope. As in linear programming, one could consider the potential function

$$\phi(x) = \rho \log\left(\sum_{j=1}^{n} x_j^p - \underline{z}\right) - \sum_{j=1}^{n} \log x_j = \rho \log(p(x) - \underline{z}) - \sum_{j=1}^{n} \log x_j, \tag{16}$$

where $\underline{z}$ is a lower bound on the global minimal objective value of (1) and ρ satisfies parameter $\rho > n$. For simplicity, we set $\underline{z} = 0$ in this paper, since the L_p minimization objective function is always nonnegative. Note that

$$\frac{\sum_{j=1}^{n} x_j^p}{n} \geq \left(\prod_{j=1}^{n} x_j^p\right)^{1/n}$$

and therefore

$$\frac{n}{p}\log(p(x)) - \sum_{j=1}^{n} \log x_j \geq \frac{n}{p}\log n.$$

Thus, if $\phi(x) \leq (\rho - n/p)\log(\epsilon) + (n/p)\log n$, we must have $p(x) \leq \epsilon$, which implies that x must be an ϵ-global minimizer.

Give an interior point x, the algorithm looks for the best possible potential reduction from x. In a manner similar to the algorithm discussed in [26], one can consider a potential reduction $\phi(x^+) - \phi(x)$ by one-iteration update from x to x^+.

Note that

$$\phi(x^+) - \phi(x) = \rho(\log(p(x^+)) - \log(p(x))) + (-\sum_{j=1}^{n} \log(x_j^+) + \sum_{j=1}^{n} \log(x_j)).$$

Let d_x with $Ad_x = 0$ be a vector such that $x^+ = x + d_x > 0$. Then, from the concavity of $\log(p(x))$, we have

$$\log(p(x^+)) - \log(p(x)) \leq \frac{1}{p(x)} \nabla p(x)^T d_x.$$

On the other hand, by restricting $\|X^{-1}d_x\| \leq \beta < 1$, where $X = \mathrm{Diag}(x)$, we have (see Section 9.3 in [3] for a detailed discussion)

$$-\sum_{j=1}^{n} \log(x_j^+) + \sum_{j=1}^{n} \log(x_j) \leq -e^T X^{-1} d_x + \frac{\beta^2}{2(1-\beta)}.$$

From the analysis above, if $\|X^{-1}d_x\| \leq \beta < 1$, then $x^+ = x + d_x > 0$, and

$$\phi(x^+) - \phi(x) \leq (\frac{\rho}{p(x)} \nabla p(x)^T X - e^T) X^{-1} d_x + \frac{\beta^2}{2(1-\beta)}.$$

Let $d' = X^{-1}d_x$. Then, to achieve a potential reduction, one can minimize an affine-scaled linear function subject to a ball constraint as it is done for linear programming (see Chapter 1 and 4 in [27] for more details):

$$\begin{aligned} Z(d') := \text{Minimize} \quad & \left(\tfrac{\rho}{p(x)}\nabla p(x)^T X - e^T\right) d' \\ \text{Subject to} \quad & AXd' = 0 \\ & \|d'\|^2 \le \beta^2. \end{aligned} \tag{17}$$

This is simply a linear projection problem. The minimal value is $Z(d') = -\beta \cdot \|g(x)\|$ and the optimal direction is given by $d' = \frac{\beta}{\|g(x)\|} g(x)$. Here

$$\begin{aligned} g(x) &= -(I - XA^T(AX^2A^T)^{-1}AX)(\frac{\rho}{p(x)}X\nabla p(x) - e) \\ &= e - \frac{\rho}{p(x)}X(\nabla p(x) - A^T y), \end{aligned}$$

where $y = (AX^2A^T)^{-1}AX(X^p - \frac{p(x)}{\rho}e)$.

If $\|g(x)\| \ge 1$, then the minimal objective value of the subproblem is less than $-\beta$ so that

$$\phi(x^+) - \phi(x) < -\beta + \frac{\beta^2}{2(1-\beta)}.$$

Thus, the potential value is reduced by a constant if we set $\beta = 1/2$. If this case would hold for $O((\rho - n/p)\log\frac{1}{\epsilon})$ iterations, we would have produced an ϵ-global minimizer of (1).

On the other hand, if $\|g(x)\| \le 1$, from $g(x) = e - \frac{\rho}{p(x)}X(\nabla p(x) - A^T y)$, we must have

$$\frac{\rho}{p(x)}X(\nabla p(x) - A^T y) \ge 0, \quad \frac{\rho}{p(x)}X(\nabla p(x) - A^T y) \le 2e, \forall j.$$

In other words,

$$\left(\nabla p(x) - A^T y\right)_j \ge 0, \quad \frac{x_j}{p(x)}\left(\nabla p(x) - A^T y\right)_j \le \frac{2}{\rho}, \ \forall j.$$

The first condition indicates that the Lagrange multiplier y is feasible. For the second inequality, by choosing $\rho \ge \frac{2n}{\epsilon}$ we have $\frac{1}{p(x)}x^T(\nabla p(x) - A^T y) \le \epsilon$. Therefore,

$$\frac{x^T(\nabla p(x) - A^T y)}{\bar{z} - \underline{z}} \le \frac{x^T(\nabla p(x) - A^T y)}{p(x)} \le \epsilon,$$

which implies that x is an ϵ-KKT point of (1); see (7).

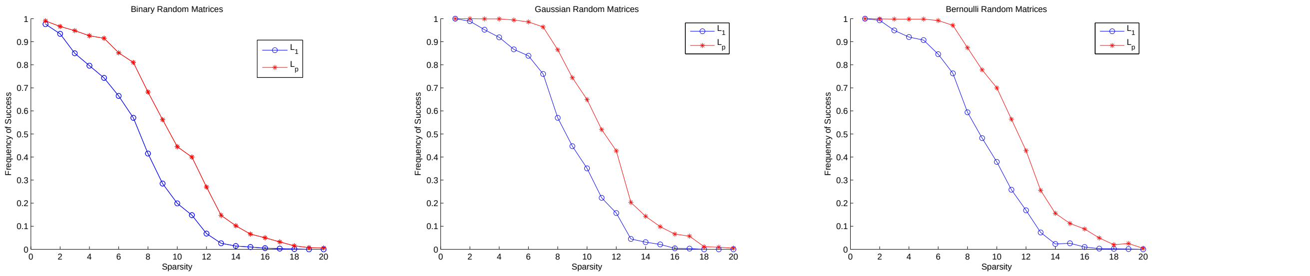
Concluding the analysis above, we have the following lemma.

Lemma 1 *The interior-point algorithm returns an ϵ-KKT or ϵ-global solution of (1) in no more than $O(\frac{n}{\epsilon}\log\frac{1}{\epsilon})$ iterations.*

A more careful computation will make the complementarity point satisfy the second order necessary condition; see [26]. By combining these observations with Theorem 4, we immediately conclude the convergence of our interior-point algorithm.

Corollary 1 *The potential reduction algorithm that starts from the (approximate) analytic center of the polytope and generates a sequence of interior points converging to a local minimizer is an FPTAS to compute an approximate local minimizer for the L_p minimization problem.*

To conclude, some interior-point algorithms, including the simple affine-scaling algorithm that always goes along a descent direction, can be effective (with fully polynomial time) in tackling the L_p minimization problem as well.



(a) (b) (c)

Fig. 1 Successful sparse recovery rates of L_p and L_1 solutions, with matrix A constructed from (a) a binary matrix; (b) a Gaussian random matrix; and (c) a Bernoulli random matrix.

5 Computational Experiments of the Potential Reduction Algorithm

In this section, we compute a solution of (1) using the interior-point algorithm above and compare it with the solution of the L_1 problem, i.e., the solution of (1) with $p = 1$, which is also computed by an interior-point algorithm for linear programming. Our preliminary results reinforce our reasoning that interior-point algorithms likely avoid some local minimizers on the boundary and recover the sparse solution. A more extensive computation is in working process.

We construct 1000 random pairs (A, x) with matrices A size of 30 by 120 and vectors $x \in R^{120}$ for sparsity $\|x\|_0 = s$ with $s = 1, 2, \cdots, 20$. With the basis matrix given, the vector $b = Ax$ is known. We use several basis design matrices A to test our algorithms. In particular, let $A = [M, -M]$ where M is one of the following matrices: (1) Sparse binary random matrices where there are only a small number of ones in each column; (2) Gaussian random matrices whose entries are i.i.d. Gaussian random variables; (3) Bernoulli random matrices whose entries are i.i.d. Bernouli random variables. We note that Gaussian or Bernoulli random matrices satisfy the *restricted isometry property* [7,8], i.e., columns of the basis matrix are nearly orthogonal, while sparse binary random matrices only satisfy a weaker form of that property [1]. Because two copies of the same random matrices are concatenated in A, column orthogonality will not be maintained, which suggests that sparse recovery by the L_1 problem will fail. We solve (1) with $p = 1/2$ by the interior-point algorithm developed above and compare the successful sparse recovery rate with the solutions of the L_1 problem. A solution is considered to successfully recovering x if the L_2 distance between the solution and x is less than 10^{-3}.

The phase transitions of successful sparse recovery rates for L_p and L_1 problem for three cases of basis matrices are plotted in figure 5. We observe that the $L_{0.5}$ interior-point algorithm performs better in successfully identifying the sparse solution x than the L_1 algorithm does. Moreover, we note that when the basis matrices are binary, we have comparatively lower rates of successful sparse recovery (when sparsity $s = 4$, for example, successful recovery rate for L_p solutions is about 95% for binary random matrices, compared with almost 100% for Gaussian or Bernoulli random matrices; for L_1 solutions it is about 80% for binary random matrices compared with about 90% for Gaussian or Bernoulli random matrices). This may be supported by the fact that sparse binary random matrices have even worse column orthogonality than Gaussian or Bernoulli cases. Our simulation results also show that the interior-point algorithm for the L_p problem with $0 < p < 1$ runs as fast as the interior-point method for the L_1 problem, which makes the interior-point method competitive for large scale sparse recovery problems.

As a final remark, we aim to compare the solution sparsity between the L_p and the L_1 minimization models but not the algorithms. Thus, the L_1 minimizer we used in our experiments is independent of the specific choice of computational techniques. We note that many classes of computational techniques have arisen in recent years for solving sparse approximation problems; see [22] for a state-of-the-art survey. Some of those methods such as the Orthogonal Matching Pursuit [14,21], iterative thresholding [4], etc, can be shown to be able to approximately find the L_1 minimizer under good RIP conditions on the basis matrix A. In our case, when the basis matrix A does not have a good RIP property, these computational techniques have frequently failed. For example, the orthogonal matching pursuit algorithm exhibited difficulties in identifying which bases maximize the absolute value of the correlation with the residual due to the fact that there are two copies of the same basis that only differ in signs; the iterative thresholding approach could not converge to a sparse solution with our basis matrix. However, when A has a good RIP property, we see no significant recovering difference among all these methods.

Acknowledgements The authors would like to thank Zizhuo Wang, Jiawei Zhang and John Gunnar Carlsson for their helpful discussions, as well as the editor and anonymous referees for their insightful comments and suggestions. Research by the first author is supported by National Natural Science Foundation of China under Grants 71001062.

References

1. R. Berinde, A. C. Gilbert, P. Indyk, H. J. Karloff and M. J. Strauss, "Combining geometry and combinatorics: A unified approach to sparse signal recovery", preprint, 2008.
2. D. P. Bertsekas, "Nonlinear programming", Athena Scientific, (1999).
3. D. Bertsimas and J. Tsitsiklis, "Introduction to linear Optimization", Athena Scientific, (1997), 414.
4. T. Blumensath and M. Davies, "Iterative hard thresholding for compressed sensing", *Appl. Comp. Harmonic Anal.*, 27(2009), 265-274.
5. J. M. Borwein and D. R. Luke, "Entropic Regularization of the l_0 function", in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering in Springer Optimization and Its Applications*, (2010).
6. A.M. Bruckstein, D.L. Donoho, M. Elad, From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Review* 51-1(2009), 34C81.
7. E. J. Candès and T. Tao, "Decoding by linear programming", *IEEE Transaction of Information Theory*, 51(2005), 4203-4215.
8. E. J. Candès and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies," *IEEE Transaction of Information Theory*, 52(2006), 5406-5425.
9. R. Chartrand "Exact Reconstruction of sparse signals via nonconvex minimization", *IEEE Signal Processing Letters*, 14-10(2009).
10. R. Chartrand "Fast algorithms for nonconvex compressive sensing: MRI reconstruction from very few data", in *IEEE International Symposium on Biomedical Imaging (ISBI)*, (2009).
11. X. Chen, F. Xu and Y. Ye, "Lower Bound Theory of Nonzero Entries in Solutions of ℓ_2-ℓ_p Minimization", Technical Report, The Hong Kong Polytechnic University, (2009).
12. D. Donoho, "For most large underdetermined systems of linear equations the minimal l_1-norm solution is also the sparsest Solution", Technical Report, Stanford University, (2004).
13. J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties", *Journal of American Statistical Society*, 96(2001), 1348-1360.
14. A. C. Gilbert and M. Muthukrishnan and M. J. Strauss, "Approximation of functions over redundant dictionaries using coherence", in *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms(SODA)*, (2003).
15. M. R. Garey and D. S. Johnson, "Strong" NP-Completeness Results: Motivation, Examples, and Implications, *Journal of the Association of Computing Machinery*, 25(1978), 499-508.
16. M. R. Garey and D. S. Johnson, "Computers and Intractability; A Guide to the Theory of NP-Completeness", W. H. Freeman & Company, (1979), 96-105.
17. N. Mourad and P. Reilly, "Minimizing nonconvex functions for sparse vector reconstruction", *IEEE Transactions on Signal Processing*, 58(2010), 3485–3496.
18. Y. Nesterov and A. Nemirovski, Interior Point Polynomial Methods in Convex Programming, SIAM, Philadelphia, PA, (1994).
19. B. K. Natarajan, "Sparse approximate solutions to linear systems", *SIAM Journal on Computing*, 24(1995), 227-234.
20. T. Terlaky, "On ℓ^p programming", *European Journal of Operational Research*, 22(1985), 70- 100.
21. J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation", *IEEE Trans. Info. Theory*, 50(2004), 22312242.
22. J. Tropp and S.J. Wright, "Computational methods for sparse solutions of linear inverse problems", to appear in *Proceedings of the IEEE*, (2010).
23. S. A. Vavasis, "Polynomial time weak approximation algorithms for quadratic programming", in C.A. Floudas and P.M. Pardalos: *Complexity in Numerical Optimization*, World Scientific, New Jersey, (1993).
24. V. Vazirani, "Approximation Algorithms", Springer, Berlin, (2003).
25. G. Xue and Y. Ye, "An efficient algorithm for minimizing a sum of P-norms," *SIAM J. Optimization* 10 (2000) 551-579.
26. Y. Ye, "On the complexity of approximating a KKT point of quadratic programming", *Mathematical Programming*, 80(1998), 195-211.
27. Y. Ye, "Interior point algorithms: theory and analysis", John Wiley & Sons, Inc., New York, (1997).