Complexity of Unconstrained L_2-L_p Minimization

Xiaojun Chen* Dongdong Ge† Zizhuo Wang‡ Yinyu Ye§

11 May, 2011, Revised 9 December,2011, 6 July 2012, Accepted 10 November 2012

Abstract

We consider the unconstrained L_q-L_p minimization: find a minimizer of $\|Ax-b\|_q^q+\lambda\|x\|_p^p$ for given $A \in R^{m\times n}$, $b \in R^m$ and parameters $\lambda > 0$, $p \in [0, 1)$ and $q \geq 1$. This problem has been studied extensively in many areas. Especially, for the case when $q = 2$, this problem is known as the L_2-L_p minimization problem and has found its applications in variable selection problems and sparse least squares fitting for high dimensional data. Theoretical results show that the minimizers of the L_q-L_p problem have various attractive features due to the concavity and non-Lipschitzian property of the regularization function $\|\cdot\|_p^p$. In this paper, we show that the L_q-L_p minimization problem is strongly NP-hard for any $p \in [0, 1)$ and $q \geq 1$, including its smoothed version. On the other hand, we show that, by choosing parameters (p, λ) carefully, a minimizer, global or local, will have certain desired sparsity. We believe that these results provide new theoretical insights to the studies and applications of the concave regularized optimization problems.

Keywords. Nonsmooth optimization, nonconvex optimization, variable selection, sparse solution reconstruction, bridge estimator.

MSC2010 Classification. 90C26, 90C51

1 Introduction

In this paper, we consider the following L_2-L_p minimization problem:

$$\text{Minimize}_x \quad f_p(x) := \|Ax-b\|_2^2 + \lambda\|x\|_p^p \tag{1}$$

where data and parameter $A = (a_1, ..., a_n) \in R^{m\times n}, 0 \neq b \in R^m$, $\lambda > 0$ and $0 \leq p < 1$, and variables $x \in R^n$. This regularized formulation has been studied extensively in variable selection and sparse least squares fitting for high dimensional data, see [1, 2, 3, 5, 6, 7, 10, 11, 12, 13, 14] and references therein. Here, when $p = 0$,

$$\|x\|_0^0 = \|x\|_0 = |\{i : x_i \neq 0\}|$$

*Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China. E-mail:maxjchen@polyu.edu.hk

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

‡Department of Management Science and Engineering, Stanford University, Stanford, CA 94305. E-mail: zzwang@stanford.edu

§Department of Management Science and Engineering, Stanford University, Stanford, CA 94305; and Visiting Professor of Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China. E-mail: yinyu-ye@stanford.edu

that is, the number of nonzero entries in x.

The original goal of the model was to find a least squares solution with fewer nonzero entries for an under-determined linear system that has more variables than the data measurements. For this purpose, people considered the regularized L_2-L_0 problem. For instance, the variable subset selection method can be viewed as the L_2-L_0 problem, which is the most popular method of regression regularization used in statistics [7].

However, the L_0 regularized problem is difficult to deal with because of the discrete structure of the 0-norm, while the solvability of the L_2-L_p problem for $p \in (0, 1)$ can be derived from the continuity and level boundedness of $f_p(x)$ defined in (1). A global minimizer of the L_2-L_p problem is also called a bridge estimator in statistical literature [7] and has various nice properties including the oracle property [5, 11, 12]. Moreover, theoretical results show that in distinguishing zero and nonzero entries of coefficients in sparse high-dimensional approximation, the bridge estimators have advantages over the Lasso estimators that minimize the following convex L_2-L_1 minimization problem:

$$\text{Minimize}_x \quad f_1(x) := \|Ax - b\|_2^2 + \lambda \|x\|_1. \tag{2}$$

Due to these advantages, researchers have been interested in the L_p regularization problem for $0 < p < 1$. However, the L_2-L_p problem (1) is a nonconvex, non-Lipschitz optimization problem. There are not many optimization theories on analyzing this type of problems. Many practical approaches have been developed to tackle the problem (1), see, e.g., [1, 2, 3, 11, 13]; but there is no globally convergent algorithm that guarantees to find a global minimizer or bridge estimator.

To the best of our knowledge, the computational complexity of the L_2-L_p minimization problem remains an open problem. It is also not clear if one can directly draw a hardness result from the following problem:

$$\begin{array}{ll} \text{Minimize} & \|x\|_p^p \\ \text{Subject to} & Ax = b, \end{array} \tag{3}$$

which is shown in [10] to be strongly NP-hard for $p \in [0, 1)$; or the problem

$$\begin{array}{ll} \text{Minimize} & \|x\|_0 \\ \text{Subject to} & \|Ax - b\|_2 \le \epsilon, \end{array} \tag{4}$$

which is shown in [14] to be NP-hard for certain ϵ. From complexity theory perspective, an NP-hard optimization problem with a polynomially bounded objective function does not admit a polynomial-time algorithm, and a strongly NP-hard optimization problem with a polynomially bounded objective function does not even admit a fully-polynomial-time approximation scheme (FPTAS), unless P=NP [17].

Indeed, the L_2-L_p problem (1) can be viewed as a quadratic penalty problem of problem (3). Intuitively, solving an unconstrained penalty optimization problem is easier than solving the constrained optimization problem. Unfortunately, we show that this is not true for solving (1). More precisely, we show that finding a global minimizer of L_2-L_p problem (1) remains strongly NP-hard for all $0 \le p < 1$ and $\lambda > 0$, including its smoothed version. We also extend the strong NP-hardness result to the L_q-L_p minimization problem for any $q \ge 1$.

On the positive side, we present a sufficient condition on the choice of λ for the desired sparsity of all minimizers, global or local, of the L_2-L_p problem for given (A, b, p), as long as their objective value is below that of the all-zero solution. Under this condition, any such a local optimal solution of problem (1) is a sparse estimator to the original problem. This may explain why many methods, e.g., [1, 2, 3, 11, 13], have reported encouraging computational results, although what they calculate may not be a global minimizer.

The remainder of this paper is organized as follows: in Section 2, we present sufficient conditions on the choice of λ to meet the sparsity requirement of global or local minimizers of the L_2-L_p minimization problem. In general, when λ is sufficiently large with respect to data (A, b) and p, the number of nonzero entries in any minimizer of the problem must be small. In Section 3, we prove that the L_q-L_p minimization problem:

$$\text{Minimize}_x \quad f_{q,p}(x) := \|Ax - b\|_q^q + \lambda \|x\|_p^p \tag{5}$$

is strongly NP-hard for any given $0 \le p < 1$, $q \ge 1$ and $\lambda > 0$. We then extend our hardness result to its smoothed version:

$$\text{Minimize}_x \quad f_{q,p,\epsilon}(x) := \|Ax - b\|_q^q + \lambda \textstyle\sum_{i=1}^n (|x_i| + \epsilon)^p \tag{6}$$

for any given $0 < p < 1$, $q \ge 1$, $\lambda > 0$ and $\epsilon > 0$, even though the objective function in this case is Lipschitz continuous. Thus, changing the non-Lipschitz regularization model (5) to a Lipschitz continuous model (6) gains no advantage in terms of computational complexity. Finally, we show that our results are consistent with the existing findings from statistical literature, but give more specific bounds on choosing regularization parameters. We also illustrate that for the purpose of finding a least squares solution with a targeted number of nonzero entries, finding a local minimizer of problem (1) is likely to accomplish the same objective as finding a global minimizer does.

In the rest of the paper, we define $z^0 = 0$ if $z = 0$ and $z^0 = 1$ if $z \neq 0$. We use $(x \cdot y)$ to represent the vector $(x_1 y_1, \dots, x_n y_n)^T \in R^n$ and $\|\cdot\|$ to denote the L_2 norm.

2 Choosing the parameter λ for sparsity

In applications like variable selection and sparse solution reconstruction, one wants to find least square estimators with no more than k nonzero entries. On the other hand, one obviously wants to avoid the all-zero solution. The L_2-L_p regularized approach is to first solve L_2-L_p problem (1) to find a minimizer. Then, eliminate all variables who have zero values in the minimizer, and solve the least square problem using only remaining variables. Thus, the key is to control the support size of minimizers of problem (1) such that it does not exceed k, and this is typically accomplished by selecting a suitable λ. We now give a sufficient condition on λ for global minimizers of the L_2-L_p problem to have desirable sparsity.

Theorem 1. *Let*

$$\beta(k) = k^{p/2-1} \left(\frac{2\alpha}{p(1-p)} \right)^{p/2} \|b\|^{2-p}, \quad \alpha = \max_{1 \le i \le n} \|a_i\|^2, \quad 1 \le k \le n. \tag{7}$$

The following statements hold.

(1) If $\lambda \ge \beta(k)$, any global minimizer x^* of L_2-L_p problem (1) satisfies $\|x^*\|_0 < k$ for $k \ge 2$.

(2) If $\lambda \ge \beta(1)$, $x^* = 0$ is the unique global minimizer of L_2-L_p problem (1).

(3) Suppose that set $C := \{ x \mid Ax = b \}$ is non-empty. Then, if $\lambda \leq \frac{\|b\|^2}{\|x_c\|_p^p}$ for some $x_c \in C$, any global minimizer x^* of L_2-L_p problem (1) satisfies $\|x^*\|_0 \geq 1$.

Proof. Suppose that $x^* \neq 0$ is a global minimizer of the L_2-L_p problem (1). Let $B = A_T \in R^{m\times|T|}$, where T =support(x^*) and $|T| = \|x^*\|_0$ is the cardinality of the set T. Then by considering the first order necessary condition, x^* must satisfy

$$2B^T(Bx_T^* - b) + p\lambda(|x_T^*|^{p-2} \cdot (x_T^*)) = 0. \tag{8}$$

This implies $Ax^* - b = Bx_T^* - b \neq 0$. Hence we have

$$f_p(x^*) = \|Ax^* - b\|^2 + \lambda\|x^*\|_p^p > \lambda \sum_{i\in T} |x_i^*|^p. \tag{9}$$

In Theorem 2.2 of [3], it is shown that $|x_i^*| \geq \left(\frac{\lambda p(1-p)}{2\alpha}\right)^{1/(2-p)}$, $i = 1, \ldots, n$. Therefore, we have $f_p(x^*) \geq \lambda|T|\left(\frac{\lambda p(1-p)}{2\alpha}\right)^{p/(2-p)}$. Now we discuss by different cases:

(1) Suppose that $\lambda \geq \beta(k)$. If $\|x^*\|_0 \geq k \geq 1$, then from (9) and the definition of $\beta(k)$ in (7), we have

$$f_p(x^*) > k\lambda^{2/(2-p)}\left(\frac{p(1-p)}{2\alpha}\right)^{p/(2-p)} \geq \|b\|^2 = f_p(0).$$

This contradicts to that x^* is a global minimizer of (1). Hence $\|x^*\|_0 < k$.

(2) Suppose $\lambda \geq \beta(1)$. If $x^* \neq 0$, then there is i such that $x_i^* \neq 0$ and

$$f_p(x^*) = \|Ax^* - b\| + \lambda\|x^*\|_p^p > \lambda|x_i^*|^p \geq \lambda\left(\frac{\lambda p(1-p)}{2\alpha}\right)^{p/(2-p)} \geq \|b\|^2 = f(0).$$

This contradicts to that x^* is a global minimizer of (1). Hence, $x = 0$ is the unique global minimizer of (1).

(3) Note that $f_p(0) = \|b\|^2$ and $f_p(x_c) = \lambda\|x_c\|_p^p$ for $x_c \in C$. Therefore, if

$$\lambda \leq \frac{\|b\|^2}{\|x_c\|_p^p} \quad \text{for some} \quad x_c \in C \tag{10}$$

then $f_p(0) \geq f_p(x_c)$. Since x_c is not a stationary point of L_2-L_p problem [3], there is $\tilde{x}$ near x_c such that $f_p(x_c) > f_p(\tilde{x})$. Hence $x = 0$ cannot be a global minimizer of (1). □

Remark 1 It was known that $x = 0$ is always a local minimizer of the L_2-L_p problem (1) for any value of $\lambda > 0$ [3], and $x = 0$ is a global minimizer of (1) for a "sufficiently large" λ [11]. Theorem 1, for the first time, establishes a specific bound $\beta(1)$, such that $x = 0$ is the unique global minimizer of (1) for $\lambda \geq \beta(1)$. An important algorithmic implication of Theorem 1 is that, for given data (A, b) and p, choosing $\lambda \geq \beta(k)$ for a small constant k does not help to solve the original sparse least squares problem. For a small constant k, say from 1 to 3, one might be better off to enumerate all combinations of solutions, each with no more than k nonzero entries, to find a minimizer. This can be done in a strongly polynomial time of the problem dimensions. Moreover, the three statements of Theorem 1 can be extended to problem (6) with $q = 2$ and $\epsilon < \left(\frac{\lambda p(1-p)}{2\alpha}\right)^{1/(2-p)}$, since from Theorem 2.2 in [4], we know that any local minimizer x^* of problem (6) satisfies $|x_i^*| \geq \left(\frac{\lambda p(1-p)}{2\alpha}\right)^{1/(2-p)} - \epsilon$, $i = 1, \ldots, n$. In fact, we only need to replace $\left(\frac{\lambda p(1-p)}{2\alpha}\right)^{1/(2-p)}$ by $\left(\frac{\lambda p(1-p)}{2\alpha}\right)^{1/(2-p)} - \epsilon$ in (7).

One may be also interested in the relation of λ and the support sizes of local minimizers of L_2-L_p problem (1). We present the following result for the sparsity of certain local minimizers of (1).

Theorem 2. *Let*

$$\gamma(k) = k^{p-1} \left(\frac{2\|A\|}{p}\right)^p \|b\|^{2-p}. \tag{11}$$

If $\lambda \geq \gamma(k)$, then any local minimizer x^* of problem (1), with $f_p(x^*) \leq f_p(0) = \|b\|^2$, satisfies $\|x^*\|_0 < k$ for $k \geq 2$.

Proof. Note that (8) holds for any local minimizer of L_2-L_p problem (1). By Theorem 2.3 in [3], for any local minimizer x^* of L_2-L_p problem (1) in the level set $\{x : \ f_p(x) \leq f_p(0)\}$, we have

$$f_p(x^*) = \|Ax^* - b\|^2 + \lambda\|x^*\|_p^p > \lambda \sum_{i \in T} |x_i^*|^p \geq \lambda |T| \left(\frac{\lambda p}{2\|A\|\|b\|}\right)^{p/(1-p)}, \tag{12}$$

where T =support(x^*). If $|T| = \|x^*\|_0 \geq k \geq 1$, then

$$f_p(x^*) > \lambda k \left(\frac{\lambda p}{2\|A\|\|b\|}\right)^{p/(1-p)} = \lambda^{1/(1-p)} k \left(\frac{p}{2\|A\|}\right)^{p/(1-p)} \|b\|^{p/(p-1)} \geq \|b\|^2 = f_p(0),$$

which is a contradiction. □

Theorem 1 concerns global minimizers of L_2-L_p problem (1) while Theorem 2 concerns its local minimizers in the level set $\{x : \ f_p(x) \leq f_p(0)\}$. Since $x = 0$ is a trivial local minimizer for problem (1), we believe any good method would likely find a minimizer that at least is better than $x = 0$. Below, we use an example to illustrate the bounds presented in Theorems 1 and 2.
Example 2.1 Consider the following L_2-$L_{1/2}$ minimization problem

$$\text{Minimize} \quad f(x) := (x_1 + x_2 - 1)^2 + \lambda(\sqrt{|x_1|} + \sqrt{|x_2|}). \tag{13}$$

From $A = (1, 1)$, $b = 1$ and $x_c = (1, 0)$, we easily find these data in Theorem 1 and Theorem 2,

$$\alpha = 1, \quad \|b\| = 1, \quad \beta(k) = 8^{1/4} k^{-3/4}, \quad \frac{\|b\|^2}{\|x_c\|_p^p} = 1, \quad \gamma(k) = 32^{1/4} k^{-1/2}.$$

For $k = 2$, we have $\beta(2) = 1$. Using parts 1 and 3 of Theorem 1, we can claim that any minimizer x^* of (13) with $\lambda = 1$ satisfies $\|x^*\|_0 = 1$. Using part 2 of Theorem 1, we can claim that $x = 0$ is the unique minimizer of (13) with $\lambda \geq \beta(1) = 8^{1/4}$. The lower bound $\beta(1)$ can be improved further. In fact, we can give a number $\beta^* \leq \beta(1)$ such that $x = 0$ is the unique minimizer of (13) with $\lambda \geq \beta^*$ by using the first and second order necessary conditions [3] for (1).

For $\lambda = \frac{8}{3\sqrt{3}} < 8^{1/4}$, it is easy to see that $(x_1, x_2) = (1/3, 0)$ and $(x_1, x_2) = (0, 1/3)$ are two vectors satisfying

$$2x_1(x_1 + x_2 - 1) + \frac{\lambda}{2}\sqrt{|x_1|} = 0, \qquad 2x_2(x_1 + x_2 - 1) + \frac{\lambda}{2}\sqrt{|x_2|} = 0,$$

and

$$H(x) = 2\begin{pmatrix} x_1^2 & x_1x_2 \\ x_1x_2 & x_2^2 \end{pmatrix} - \frac{\lambda}{4}\begin{pmatrix} \sqrt{|x_1|} & 0 \\ 0 & \sqrt{|x_2|} \end{pmatrix} = 0.$$

However, since the third order derivative of $g(t) := f((1/3 + t)e_1)$ (or $g(t) := f((1/3 + t)e_2)$) is strictly positive on both side of $t = 0$, $(x_1, x_2) = (1/3, 0)$ and $(x_1, x_2) = (0, 1/3)$ are not local minimizers. Moreover, these two vectors are the only nonzero vectors satisfying both first and second order necessary conditions. We can claim that $x = 0$ is the unique global minimizer of (13).

Our theorems reinforce the findings from statistical literature that global minimizers of the L_2-L_p regularization problem may have many advantages over those from other convex regularization problems. Furthermore, our new results actually give precise bounds on how to choose λ for desirable sparsity on all local minimizer in a certain level set. The remaining question: is the L_2-L_p regularization problem (1) tractable for given $\lambda > 0$ and $0 \le p < 1$? Or more specifically, is there an efficient or polynomial-time algorithm to find a global minimizer of problem (1)? Unfortunately, we prove a strong negative result in the next section.

3 The L_2-L_p problem is strongly NP-hard

As we mentioned earlier, one may attempt to draw a negative result directly from constrained L_p problem (3) or (4). However, it is well known that the quadratic penalty function is not exact because its minimizer is generally not the same as the solution of the corresponding constrained optimization; see, e.g., [15]. For example, the all-zero vector is a local minimizer of the L_2-L_p problem (1), but it may not even be feasible for the L_p problem (3). On the other hand, the set of all basic feasible solutions of (3) is exactly the set of its local minimizer [10], but such a local minimizer of (3) may not even be a stationary point of problem (1). In fact, there is no $\lambda > 0$ such that $\bar{x}$, any feasible solution of problem (3), satisfies the first order necessary condition of L_2-L_p problem (1). Another difference between (3) and (1) is the following: it has been shown in [10] that any solution is a local minimizer of (3) as long as it satisfies the first and second order necessary optimality conditions of (3). However, Example 2.1 shows that this fact is not true for L_2-L_p problem (1).

Thus, we need somewhat new proofs for the hardness result. To facilitate the new proof, we first prove that problem (5) is NP-hard, and then extend to the strongly NP-hard result.

Theorem 3. *Minimization problem (5) is NP-hard for any given $0 \le p < 1$, $q \ge 1$ and $\lambda > 0$.*

We first provide a useful technical lemma. The proof is given in the appendix.

Lemma 4. *Consider the problem*

$$\textit{Minimize}_{z \in R} \quad g(z) := |1 - z|^q + \frac{1}{2}|z|^p \tag{14}$$

for some given $0 \le p < 1$ and $q \ge 1$. It is minimized at a unique point (denoted by $z^*(p, q)$) on $(0, 1]$. And the optimal value $c(p, q)$ is less than $\frac{1}{2}$.

Proof of Theorem 3. First we claim that without loss of generality we only need to consider the problem with $\lambda = \frac{1}{2}$. This is because given any problem of form (5), we can make the following transformation:

$$\tilde{x} = (2\lambda)^{1/p} x \ , \ \tilde{A} = (2\lambda)^{-1/p} A \text{ and } \tilde{b} = b$$

and scale this problem to:

$$\text{Minimize}_{\tilde{x}} \quad \|\tilde{A}\tilde{x} - \tilde{b}\|_q^q + \frac{1}{2}\|\tilde{x}\|_p^p. \tag{15}$$

Note that this transformation is invertible, i.e., for any given λ_0, one can transform an instance with $\lambda = \lambda_0$ to one with $\lambda = \frac{1}{2}$ and vice versa. Therefore, we only need to consider the case when $\lambda = \frac{1}{2}$.

Now we present a polynomial time reduction from the well known NP-complete *partition problem* [9] to problem (15). The partition problem can be described as follows: given a set S of rational numbers $\{a_1, a_2, \ldots, a_n\}$ with sum $2b$, 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 and the sum of the numbers in S_2 both equal to b?

Given an instance of the partition problem with $a = (a_1, a_2, \ldots, a_n)^T \in R^n$. We consider the following minimization problem in form (15):

$$\text{Minimize}_{x,y} \quad P(x,y) = |a^T(x-y)|^q + \sum_{1 \le j \le n} |x_j + y_j - 1|^q + \frac{1}{2} \sum_{1 \le j \le n} (|x_j|^p + |y_j|^p). \tag{16}$$

We have

$$\begin{aligned}
\text{Minimize}_{x,y} P(x,y) &\ge \text{Minimize}_{x_j,y_j} \sum_{1 \le j \le n} |x_j + y_j - 1|^q + \frac{1}{2} \sum_{1 \le j \le n} (|x_j|^p + |y_j|^p) \\
&= \sum_{1 \le j \le n} \text{Minimize}_{x_j,y_j} \ |x_j + y_j - 1|^q + \frac{1}{2}(|x_j|^p + |y_j|^p) \\
&= n \cdot \text{Minimize}_z \ |1 - z|^q + \frac{1}{2}|z|^p,
\end{aligned}$$

where the last equality is from the facts that $|x_j|^p + |y_j|^p \ge |x_j + y_j|^p$ and that we can always choose one of them to be 0 such that the equality holds.

By applying Lemma 4, we have

$$P(x,y) \ge nc(p,q).$$

Now we claim that there exists an equitable partition to the partition problem if and only if the optimal value of (16) equals to $nc(p,q)$. First, if S can be evenly partitioned into two sets S_1 and S_2, then we define $(x_i = z^*(p,q), y_i = 0)$ if a_i belongs to S_1 and define $(x_i = 0, y_i = z^*(p,q))$ otherwise. These (x_j, y_j) provide an optimal solution to $P(x,y)$ with optimal value $nc(p,q)$. On the other hand, if the optimal value of (16) is $nc(p,q)$, then in the optimal solution, for each i, we must have either $(x_i = z^*(p,q), y_i = 0)$ or $(x_i = 0, y_i = z^*(p,q))$. And we must also have $a^T(x-y) = 0$, which implies that there exists an equitable partition to set S. Thus Theorem 3 is proved. $\square$

Remark 2. The concavity $p < 1$ is necessary in Theorem 3. The proof cannot go through for the case $p = 1$ because even if the optimal value of (16) equals to $nc(p, q)$, it only implies that for each i, $x_i + y_i = z^*(p, q)$. For such a pair, we can let $x_i = y_i = \frac{z^*(p,q)}{2}$ to achieve that optimal value. However, it cannot guarantee an equatable partition.

In the following, using the similar idea, we prove a stronger result:

Theorem 5. *Minimization problem (5) is strongly NP-hard for any given* $0 \le p < 1$*,* $q \ge 1$ *and* $\lambda > 0$.

Proof. We present a polynomial time reduction from the well known strongly NP-hard 3-partition problem [8, 9]. The 3-partition problem can be described as follows: given a multiset S of $n = 3m$ integers $\{a_1, a_2, \ldots, a_n\}$ with sum mB, can S be partitioned into m subsets, such that the sum of the numbers in each subset is equal?

We consider the following minimization problem in the form (15):

$$\text{Minimize} \quad P(x) = \sum_{j=1}^{m} |\sum_{i=1}^{n} a_i x_{ij} - B|^q + \sum_{i=1}^{n} |\sum_{j=1}^{m} x_{ij} - 1|^q + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m} |x_{ij}|^p. \tag{17}$$

The remaining argument will be the same as the proof for Theorem 3. □

Theorem 5 implies that the L_2-L_p minimization problem is strongly NP-hard. Next we generalize the NP-hardness result to the smoothed version of this problem in (6).

Theorem 6. *Minimization problem (6) is strongly NP-hard for any give* $0 < p < 1$*,* $q \ge 1$*,* $\lambda > 0$ *and* $\epsilon > 0$.

The proof is similar to the previous one and we leave it to the Appendix.

The above results reveal that finding a global minimizer for the L_q-L_p minimization problem is strongly NP-hard, or the original sparse least squares problem is intrinsically hard, and no regularized optimization models/methods could help much in the worst case. That is, relaxing L_0 to L_p for some $0 < p < 1$ in the regularization gains no significant advantage in terms of the (worst-case) computational complexity.

4 Bounds $\beta(k)$ and $\gamma(k)$ for asymptotic properties

Given the strong negative result for computing a global minimizer, our hope now is to find a local minimizer of problem (1), still good enough for the desired sparsity – say no more than k nonzero entries. This is indeed guaranteed by Theorem 2 if one chooses $\lambda \ge \gamma(k)$ of (11), instead of $\lambda \ge \beta(k)$ of (7). In the following, we present a positive result in the bridge estimator model considered by [5, 11, 12].

Consider asymptotic properties of the L_2-L_p minimization (1) where the sample size m tends to infinity in the model of [5, 11, 12]. Suppose that the true estimator x^* has no more than k nonzero entries. One expects that there is a sequence of bridge estimators, i.e. solutions x_m^* of

$$\text{Minimize} \, \|Ax - b\|^2 + \lambda_m \|x\|_p^p$$

such that dist(support$\{x_m^*\}$, support$\{x^*\}$) $\to 0$, as $m \to \infty$, with probability 1.

In applications of variable selection, the design matrix is typically standardized so that

$$\|a_i\|^2 = m \quad \text{for} \quad i = 1, \dots, n.$$

Moreover, the smallest and largest eigenvalues ρ_1 and ρ_2 of the covariate matrix $\sum_m = \frac{1}{m}A^TA$ satisfy $0 < c_1 \le \rho_1 \le \rho_2 < c_2$ for some constants c_1 and c_2, see [11]. This assumption implies that $\sqrt{c_1 m} \le \|A\| \le \sqrt{c_2 m}$. For simplicity, let us fix $\|A\| = \sqrt{m}$ and $p = 1/2$. Then we have

$$\beta(k) = k^{-3/4}(8m)^{1/4}\|b\|^{3/2} \quad \text{and} \quad \gamma(k) = k^{-1/2}(16m)^{1/4}\|b\|^{3/2}.$$

One can see that $\gamma(k) > \beta(k)$ for all $k \ge 1$.

If k is a constant, we see that $\beta(k)$ and $\gamma(k)$ are in the same order of m and $\|b\|$. Thus, finding any local minimizer of problem (1) in the objective level set $f_p(0)$ is sufficient to guarantee desired sparsity when $\lambda_m = \beta(k)$. That is, there is no significant guaranteed sparsity difference between global and local minimizers of problem (1). This seems also observed in computational experiments when the true estimator is extremely sparse. Of course, when k increases as $m \to \infty$, a global minimizer of problem (1) would likely become sparser than its local minimizer, since $\beta(k)/\gamma(k) = O(k^{-1/4})$.

In general, both $\beta(k)$ and $\gamma(k)$ meet the conditions in the analysis of consistency and oracle efficiency of bridge estimators of [11, 12]. In their model, the parameter λ_m is required to satisfy certain conditions. For instances,

$$([12, Theorem\, 3]) \qquad \lambda_m m^{-p/2} \quad \to \quad \lambda_0 \ge 0 \quad \text{as} \quad m \to \infty \tag{18}$$

$$([11, A3, (a)]) \qquad \lambda_m m^{-1/2} \quad \to \quad 0 \quad \text{as} \quad m \to \infty. \tag{19}$$

With $\|a_i\|^2 = m$ for $i = 1, \dots, n$ and $\|A\| = \sqrt{m}$ in their model, we have

$$\beta(k)m^{-p/2} = k^{p/2-1}\left(\frac{2}{p(1-p)}\right)^{p/2}\|b\|^{2-p} \quad \to \quad \lambda_0 \ge 0 \quad \text{as} \quad m \to \infty$$

and

$$\beta(k)m^{-1/2} = k^{p/2-1}\left(\frac{2}{p(1-p)}\right)^{p/2}\|b\|^{2-p}m^{(p-1)/2} \quad \to \quad 0 \quad \text{as} \quad m \to \infty.$$

For $\gamma(k)$, we have

$$\gamma(k)m^{-p/2} = k^{p-1}\left(\frac{2}{p}\right)^{p}\|b\|^{2-p} \quad \to \quad \lambda_0 \ge 0 \quad \text{as} \quad m \to \infty$$

and

$$\gamma(k)m^{-1/2} = k^{p-1}\left(\frac{2}{p}\right)^{p}\|b\|^{2-p}m^{(p-1)/2} \quad \to \quad 0 \quad \text{as} \quad m \to \infty.$$

Hence, both $\lambda_m = \beta(k)$ and $\lambda_m = \gamma(k)$ satisfy (18) and (19). Moreover, by Theorem 1 and Theorem 2, any minimizer of L_2-L_p problem (1) with $\lambda = \lambda_m$ is likely to have less than k nonzero entries. Hence each of them could be a good choice for consistency and oracle efficiency of bridge estimators via solving the unconstrained L_2-L_p minimization problem (1).

Acknowledgements

The authors would like to thank anonymous referees for their helpful comments. The authors acknowledge the support of the Hong Kong Polytechnic University and Hong Kong Research Grant Council. Research by the second author is supported by National Natural Science Foundation of China under Grants 71001062.

References

[1] R. Chartrand, Exact reconstruction of sparse signals via nonconvex minimization, *IEEE Signal Processing Letters*, 14 (2007), 707-710.

[2] R. Chartrand and V. Staneva, Restricted isometry properties and nonconvex compressive sensing, *Inverse Problem*, 24 (2008), 1-14.

[3] X. Chen, F. Xu and Y. Ye, Lower bound theory of nonzero entries in solutions of ℓ_2-ℓ_p minimization, *SIAM J. Scientific Computing*, 32 (2010), 2832-2852.

[4] X. Chen and W. Zhou, Convergence of reweighted l_1 minimization algorithms and unique solution of truncated l_p minimization, Preprint, Department of Applied Mathematics, The Hong Kong Polytechnic University, 2010.

[5] J. Fan and R. Li, Variable selection via nonconcave penalized likelihood and its oracle properties, *Journal of American Statistical Society*, 96 (2001), 1348-1360.

[6] S. Foucart and M. J. Lai, Sparsest solutions of under-determined Linear Systems via l_q minimization for $0 < q \leq 1$, *Applied and Computational Harmonic Analysis*, 26 (2009), 395-407.

[7] I. E. Frank and J. H. Freidman, A statistical view of some chemometrics regression tools (with discussion), *Technometrics*, 35(1993), 109-148.

[8] 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.

[9] M. R. Garey and D. S. Johnson, Computers and Intractability; A Guide to the Theory of NP-Completeness, W. H. Freeman, New York, 1979.

[10] D. Ge, X. Jiang and Y. Ye, A note on the complexity of L_p minimization, *Mathematical Programming*, 129(2011), 285-299.

[11] J. Huang, J. L. Horowitz and S. Ma, Asymptotic properties of bridge estimators in sparse high-dimensional regression models, *The Annals of Statistics*, 36 (2008), 587-613.

[12] K. Knight and W.J. Fu, Asymptotics for lasso-type estimators, *The Annals of Statistics*, 28 (2000), 1356-1378.

[13] M. Lai and Y. Wang, An unconstrained l_q minimization with $0 < q < 1$ for sparse solution of under-determined linear systems, *SIAM J. Optimization*, 21 (2011), 82-101.

[14] B. K. Natarajan, Sparse approximate solutions to linear systems, *SIAM J. Computing*, 24 (1995), 227-234.

[15] J. Nocedal and S.J. Wright, Numerical Optimization, 2nd Edition, Springer, New York, 2006.

[16] R. Tibshirani, Regression shrinkage and selection via the Lasso, *J Royal Statistical Society B*, 58 (1996), 267-288.

[17] V. Vazirani, Approximation Algorithms, Springer, Berlin, (2003).

Appendix

Proof of Lemma 4

First it is easy to see that when $p = 0$, $g(z)$ has a unique minimizer at $z = 1$, and the optimal value is $\frac{1}{2}$. Now we consider the case when $p \neq 0$. Note that $g(z) > g(0) = 1$ for all $z < 0$, and $g(z) > g(1) = \frac{1}{2}$ for all $z > 1$. Therefore the minimum point must lie within $[0, 1]$.

To optimize $g(z)$ on $[0, 1]$, we check its first derivative

$$g'(z) = -q(1-z)^{q-1} + \frac{pz^{p-1}}{2}. \tag{20}$$

We have $g'(0^+) = +\infty$ and $g'(1) = \frac{p}{2} > 0$. Therefore, if function $g(z)$ has at most two stationary points in (0,1), the first one must be a local maximum and the second one must be the unique global minimum and the minimum value $c(p, q)$ must be less than $\frac{1}{2}$.

Now we check the possible stationary points of $g(z)$. Consider solving $g'(z) = -q(1-z)^{q-1} + \frac{pz^{p-1}}{2} = 0$. We get $z^{1-p}(1-z)^{q-1} = \frac{p}{2q}$.

Define $h(z) = z^{1-p}(1-z)^{q-1}$. We have

$$h'(z) = h(z)\left(\frac{1-p}{z} - \frac{q-1}{1-z}\right).$$

Note that $\frac{1-p}{z} - \frac{q-1}{1-z}$ is decreasing in z and must have a root on $(0, 1)$. Therefore, there exists a point $\bar{z} \in (0, 1)$ such that $h'(z) > 0$ for $z < \bar{z}$ and $h'(z) < 0$ for $z > \bar{z}$. This implies that $h(z) = \frac{p}{2q}$ can have at most two solutions in $(0, 1)$, i.e., $g(z)$ can have at most two stationary points. By the previous discussions, the lemma holds. $\square$

Proof of Theorem 6

We again consider the same 3-partition problem, we claim that it can be reduced to a minimization problem in form (6). Again, it suffices to only consider the case when $\lambda = \frac{1}{2}$ (Here we consider the hardness result for any given $\epsilon > 0$. Note that after the scaling, ϵ may have changed). Consider:

$$\text{Minimize}_x \quad P_\epsilon(x) = \sum_{j=1}^{m} |\sum_{i=1}^{n} a_i x_{ij} - B|^q + \sum_{i=1}^{n} |\sum_{j=1}^{m} x_{ij} - 1|^q + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}(|x_{ij}| + \epsilon)^p. \tag{21}$$

We have

$$\begin{aligned}
\text{Minimize}_x P_\epsilon(x) &\geq \text{Minimize}_x \sum_{i=1}^{n} |\sum_{j=1}^{m} x_{ij} - 1|^q + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}(|x_{ij}| + \epsilon)^p \\
&= \sum_{i=1}^{n} \text{Minimize}_x |\sum_{j=1}^{m} x_{ij} - 1|^q + \frac{1}{2}\sum_{j=1}^{m}(|x_{ij}| + \epsilon)^p \\
&= n \cdot \text{Minimize}_z |1 - z|^q + \frac{1}{2}(|z| + \epsilon)^p + \frac{(m-1)}{2}\epsilon^p.
\end{aligned}$$

The last equality comes from the submodularity of the function $(x+\epsilon)^p$ and the fact that one can always choose only one of x_{ij} to be nonzero in each set such that the equality holds. Consider function $g_\epsilon(z) = |1-z|^q + \frac{1}{2}(|z|+\epsilon)^p$. Similar to Lemma 4, one can prove that $g_\epsilon(z)$ has a unique minimizer in $[0, 1]$. Denote this minimum value by $c(p, q, \epsilon)$, we know that $P_\epsilon(x) \geq nc(p, q, \epsilon)$. Then we can argue that the 3-partition problem has a solution if and only if $P_\epsilon(x) = nc(p, q, \epsilon)$. Therefore Theorem 6 holds. $\square$