A Note on Complexity of L_p Minimization

Xiaoye Jiang* Yinyu Ye†

September 3, 2009

Abstract

We show that the L_p $(0 \leq p < 1)$ minimization problem arising from sparse solution construction and compressed sensing is both hard and easy. More precisely, for any fixed $0 < p < 1$, we prove that checking the global minimal value of the problem is NP-Hard; but computing a local minimizer of the problem is polynomial-time doable.

1 Short Introduction

In this note, we consider the following optimization problem:

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

and

$$\begin{array}{ll} \text{Minimize} & \sum_{1\leq j\leq n} |x_j|^p \\ \text{Subject to} & Ax = b; \end{array} \tag{2}$$

where data $A \in R^{m\times n}, b \in R^m$, and $0 < p < 1$.

Sparse signal or solution reconstruction by solving optimization problem (1) or (2), especially for the cases of $0 \leq p \leq 1$, recently received great attentions. In signal reconstruction, one typically has linear measurements $b = Ax^*$ where x^* is a sparse signal, and the sparse signal would be recovered by solving inverse problem (1) or (2) with $p = 0$, that is, to find the sparsest or smallest support cardinality solution of a linear system (here $|x|^0 = 1$ if $x \neq 0$ and 0 otherwise). From the computational complexity point of view, when $p = 0$, problem (1) or (2) is shown to be NP-hard [6] to solve; when $p = 1$, both problems are linear programs, hence they are polynomial-time solvable.

*Institute for Computational and Mathematical Engineering, Stanford University, Stanford, CA 94305. E–mail: xiaoyej@stanford.edu

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

In [1, 4], it was shown that if certain *restricted isometry property* holds for A, then the solutions of (2) for $p = 0$ and $p = 1$ are identical. Hence, problem (2) with $p = 0$ can be relaxed to problem (2) with $p = 1$. However, *restricted isometry property* may be too strong for practical basis design matrices A to hold. Thus, one may consider sparse recovery by solving relaxation problem (1) or (2) for a fixed p, $0 < p < 1$. Recently, this approach has attracted a lot of research efforts in variable selection and sparse reconstruction, e.g., [5]. It exhibits desired threshold bounds on any non-zero entry of a computed solution [3], and computational experiences show that by replacing $p = 1$ with a $p < 1$, reconstruction can be done equally fast with many fewer measurements while being more robust to noise and signal nonsparsity, e.g., [2].

In this note, we show that the L_p ($0 \le p < 1$) minimization problem is both hard and easy. More precisely, for a given real number v, the question, "is there a feasible solution to (1) or (2) such that its objective value less than or equal to v", is NP-Hard to answer. On the other hand, any basic (feasible) solution of (1) or (2) is a local minimizer, so that computing a local minimizer of the problem is polynomial-time doable.

2 The Hardness

Theorem 1. *For a given real number v, it is NP-hard to decide if the minimal objective value of problem (1) is less than or equal to v.*

Proof. We present a poly-time reduction from the well known NP-complete partition problem. An instance of 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?

Let vector $a = (a_1, a_2, \ldots, a_n) \in R^n$. Then, we consider the following minimization problem in form (1):

$$\begin{array}{ll} \text{Minimize} & P(x,y) = \sum_{1\le j\le n} (x_j^p + y_j^p) \\ \text{Subject to} & a^T(x-y) = 0, \\ & x_j + y_j = 1, \ \forall j, \\ & x, y \ge 0. \end{array} \tag{3}$$

From the strict concavity of the objective function,

$$x_j^p + y_j^p \ge x_j + y_j = 1, \ \forall j,$$

and they are equal if and only if $(x_j = 1, y_j = 0)$ or $(x_j = 0, y_j = 1)$. Thus, $P(x, y) \ge n$ for any (continuous) feasible solution of (3); and if there is a feasible solution pair (x, y) such that $P(x, y) \le n$, it must be true $x_j^p + y_j^p = 1 = x_j + y_j$ 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 partition of the entries of a.

On the other hand, if the entries of a has an equitable partition, then (3) must have a binary solution pair (x, y) such that $P(x, y) = n$. Therefore, it is NP-hard to decide if there is a feasible solution (x, y) such that its objective value $P(x, y) \le n$. $\square$

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

$$\begin{array}{ll} \text{Minimize} & \sum_{1\le j\le n} (|x_j|^p + |y_j|^p) \\ \text{Subject to} & a^T(x-y) = 0, \\ & x_j + y_j = 1, \ \forall j. \end{array} \tag{4}$$

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

$$|x_j|^p + |y_j|^p \ge x_j + y_j = 1, \ \forall j.$$

This is because when $x_j + y_j = 1$, the minimal value of $|x_j|^p + |y_j|^p$ is 1, and it equals 1 if and only if $(x_j = 1, y_j = 0)$ or $(x_j = 0, y_j = 1)$. Thus, it remains NP-hard to decide if there is a feasible solution (x, y) such that the objective value of (4) is less than or equal to n. This leads to:

Theorem 2. *For a given real number v, it is NP-hard to decide if the minimal objective value of problem (2) is less than or equal to v.*

Note that the L_1 minimization of the reduced problem does not reveal much information sparsity of the solution set, since any feasible solution is a (global) minimizer.

3 The Easiness

The above discussion reveals that finding a global minimizer for the L_p norm optimization problem is NP-hard as long as $p < 1$. 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. What about local minimizers of problems (1) and (2)? The answer is that they are easy to find.

Theorem 3. *The set of all basic feasible solutions of (1) is exactly the set of its all local minimizers.*

Proof. Observe that the objective function of (1) is strictly concave and its feasible region is a convex polyhedral set.

If x is a basic feasible solution (or extreme point), then consider its $\epsilon(> 0)$ neighborhood in the feasible region. Note that any other feasible solution in the neighborhood must have one variable having a positive value less than ϵ and it is zero in x. However, the derivative of ϵ^p can be arbitrarily large if ϵ is sufficiently small enough. This implies that the value of the objective must be increased no matter which feasible direction one follows when it starts from a basic feasible solution. Thus, x must be a strict local minimizer.

On the other hand, let x be a local minimizer but not a basic feasible solution (extreme point). Then, x must be in the interior of a face of the convex polyhedral set. Thus, there is a feasible direction $d \neq 0$ such that both $x + \epsilon d$ and $x - \epsilon d$ are feasible for sufficiently small but positive ϵ. Since either d or $-d$ will be a descent direction of the strict concave objective function, x cannot be a local minimizer. □

Similarly, we can prove

Theorem 4. *The set of all basic solutions of (2) is exactly the set of its all local minimizers.*

4 Interior-Point Algorithm

These local minimizer results show that there is little hope to solve (1) starting from a basic feasible solution. Naturally, one would start from an interior-point feasible solution such as the analytic center x^0 of the feasible polytope (if it is bounded and has an interior feasible point). Similar to potential reduction algorithms for linear programming, one could consider the potential function

$$\phi(x) = \rho \log(\sum_{j=1}^{n} x_j^p - \bar{z}) - p\sum_{j=1}^{n} \log x_j = \rho \log(p(x) - \bar{z}) - p\sum_{j=1}^{n} \log x_j, \tag{5}$$

where $\bar{z}$ is a lower bound on the global minimal objective value of (1) and parameter $\rho > n$. For simplicity, we set $\bar{z} = 0$ in the rest of discussion. Note now that

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

so that

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

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

In a manner similar to the potential reduction algorithm discussed in [7] for non-convex quadratic minimization, one can consider one-iteration update from x to x^+. Let d_x, $Ad_x = 0$, 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, if $\|X^{-1}d_x\| \leq \beta < 1$, where $X = \mathrm{Diag}(x)$,

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

where e is the vector of all ones.

Let $d' = X^{-1}d_x$. Then, to achieve a potential function, one can minimize an affine-scaled linear function subject to a ball constraint as it is done for linear programming:

$$\begin{array}{ll} \text{Minimize} & \left(\frac{\rho}{p(x)}\nabla p(x)^T - pe^T X^{-1}\right) X d' \\ \text{Subject to} & AXd' = 0 \\ & \|d'\|^2 \leq \beta^2. \end{array} \tag{6}$$

This is simply a linear projection problem. If the minimal objective value of the subproblem is less than $-\beta$, then

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

where the potential value is reduced by a constant for setting $\beta = 1/2$. On the other hand, if the minimal objective value of the subproblem is greater than or equal to $-\beta$, then one can show that we must have an ϵ-stationary point after setting $\rho = \frac{n}{\epsilon}$. The algorithm then will provably return an ϵ-stationary point of (1) in no more than $O(\frac{n}{\epsilon}\log\frac{1}{\epsilon})$ iterations. A more careful computation will make the stationary point satisfy the second order optimality condition; see [7]. Therefore, interior-point algorithms, including the simple affine-scaling algorithm, can be effective in solving the L_p minimization problem as well.

References

[1] Emmanuel. J. Candès and Terence Tao, "Decoding by linear programming", *IEEE Transaction of Information Theory*, 51(2005), 4203-4215.

[2] Rick Chartrand "Fast algorithms for nonconvex compressive sensing: MRI reconstruction from very few data", in *IEEE International Symposium on Biomedical Imaging (ISBI)*, 2009.

[3] Xiaojun Chen, Fengmin Xu and Yinyu Ye, "Lower Bound Theory of Nonzero Entries in Solutions of ℓ_2-ℓ_p Minimization", Technical Report, The Hong Kong Polytechnic University, 2009.

[4] David 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.

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

[6] Balas K. Natarajan, "Sparse approximate solutions to linear systems", *SIAM Journal on Computing*, 24(1995), 227-234.

[7] Yinyu Ye, "On the complexity of approximating a KKT point of quadratic programming", *Mathematical Programming*, 80(1998), 195-211.