Strong NP-Hardness Result for Regularized $L_q$-Minimization Problems with Concave Penalty Functions

Dongdong Ge∗ Zizhuo Wang † Yinyu Ye ‡ Hao Yin §

Abstract

We show that finding a global optimal solution for the regularized $L_q$-minimization problem ($q \geq 1$) is strongly NP-hard if the penalty function is concave but not linear in a neighborhood of zero and satisfies a very mild technical condition. This implies that it is impossible to have a fully polynomial-time approximation scheme (FPTAS) for such problems unless P = NP. This result clarifies the complexity for a large class of regularized optimization problems recently studied in the literature.

Keywords: Nonconvex optimization · Computational complexity · Regularized $L_q$-minimization · Concave penalty

1 Main Result

In this paper, we consider the following minimization problem:

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_q^q + \lambda \sum_{j=1}^{n} p(|x_j|)$$

where $q \geq 1$ and $\lambda > 0$ are parameters, $p(\cdot)$ is a concave penalty function defined on $[0, +\infty)$, and $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ are input data. When $q = 2$, this is the penalized least squares problem and has been studied extensively in the statistics literature in the past decade, especially in variable selection and sparse regression for high dimensional data. For a review of recent advances in these studies, we refer the readers to Fan and Lv (2010) and Fan et al. (2014).

Two mainstream penalty functions used in those problems are the LASSO (or the $L_1$ penalty, see Tibshirani (1996)) and the folded concave penalty, such as the smoothly clipped absolute deviation (SCAD, see Fan and Li (2001)) and the minimized concave penalty (MCP, see Zhang

∗Research Center for Management Science and Information Analytics, Shanghai University of Finance and Economics, Shanghai, China. E-mail: ge.dongdong@mail.shufe.edu.cn.
†Department of Industrial and Systems Engineering, University of Minnesota, Minneapolis, MN 55455. E-mail: zwang@umn.edu.
‡Department of Management Science and Engineering, Stanford University, Stanford, CA 94305. E-mail: yinyu-ye@stanford.edu.
§Institute for Computational and Mathematical Engineering, Stanford University, Stanford, CA 94305. E-mail: yinh@stanford.edu.
It was shown by Fan et al. (2014) that the folded concave penalization estimator owns superior statistical properties while the LASSO enjoys a higher computational efficiency.

Minimizing a folded concave penalized problem is complicated due to its intrinsic nonconvex structure. Many approximate approaches have been developed, such as the local quadratic approximation (LQA, see Fan and Li (2001)) and the local linear approximation (LLA, see Zou and Li (2008)). However, the computational complexity of the regularized Lq-minimization problem (q ≥ 1) with a general concave penalty function has not been widely discussed. We will present a detailed literature review in Section 2.

In this paper, we present a general condition on the penalty function p(·) such that problem (1) is strongly NP-hard. In our discussion, without loss of generality, we assume p(0) = 0. Below is our main result:

**Theorem 1** For any given q ≥ 1, λ > 0, if p(·) satisfies the following conditions:

(a) (Monotonicity) p(·) is non-decreasing on [0, +∞);
(b) (Concavity and differentiability) There exists τ > 0 such that p(·) is concave but not linear on [0, τ], and p(·) is twice continuously differentiable in a neighborhood of t = τ;

then the minimization problem (1) is strongly NP-hard.

Note that the concavity condition is weaker than strict concavity. It only requires the penalty function be concave while ruling out the possibility of linear penalty function (i.e., the LASSO) which is concave but also convex. As an example, this condition is satisfied by the L0 penalty and the piecewise linear penalty functions (Examples 1 and 5 below), which is not strictly concave in any interval. In Section 2, we compare our conditions to those that have been studied in the literature.

Many classes of penalty functions in the literature satisfy the above conditions. Below we present a few such examples. For the penalty functions in these examples, finding the global optimal solution for the corresponding Lq-minimization problem is strongly NP-hard by Theorem 1.

1. In variable selection problems, the L0 penalization p(t) = I_{t\neq 0} arises naturally as a penalty for the number of factors selected.
2. A natural generalization of the L0 penalization is the Lp penalization p(t) = tp where (0 < p < 1). The corresponding minimization problem is called the bridge regression problem (Frank and Freidman (1993)).
3. To obtain a hard-thresholding estimator, Antoniadis and Fan (2001) use the penalty functions pγ(t) = γ² - ((γ - t)+)² with γ > 0, where (x)+ := max{0, x} denotes the positive part of x.
4. Any penalty function that belongs to the folded concave penalty family (Fan et al. (2014)) satisfies the conditions in Theorem 1. Examples include the SCAD (Fan and Li (2001)) and the MCP (Zhang (2010a)), whose derivatives on (0, +∞) are

\[ p'_{\gamma}(t) = \gamma I_{t\leq \gamma} + \frac{(a \gamma - t)+}{a - 1} I_{t > \gamma} \quad \text{and} \quad p'_{\gamma}(t) = (\gamma - \frac{t}{b})+, \]

respectively, where γ > 0, a > 2 and b > 1.
5. The conditions in Theorem are also satisfied by the clipped $L_1$ penalty function (Antoniadis and Fan (2001), Zhang (2010b)) $p_\gamma(t) = \gamma \cdot \min(t, \gamma)$ with $\gamma > 0$. This is a special case of the piecewise linear penalty function:

$$p(t) = \begin{cases} k_1 t & \text{if } 0 \leq t \leq a \\ k_2 t + (k_1 - k_2)a & \text{if } t > a \end{cases}$$

where $0 \leq k_2 < k_1$ and $a > 0$.

6. Another family of penalty functions which bridges the $L_0$ and $L_1$ penalties are the fraction penalty functions $p_\gamma(t) = \frac{(\gamma + 1)t}{\gamma + t}$ with $\gamma > 0$ (Lv and Fan (2009)).

7. The family of log-penalty functions:

$$p(t) = \frac{1}{\log(1 + \gamma)} \log(1 + \gamma t)$$

with $\gamma > 0$, also bridges the $L_0$ and $L_1$ penalties (Candes et al. (2008)).

2 Related Literature

In this section, we review closely related works and point out the differences between our results and those in previous works.

The first closely related work is Huo and Chen (2010), in which the authors proved the following results:

**Theorem 2 (Theorem 3.1 in Huo and Chen (2010))** If $q = 2$ and $p(\cdot)$ satisfies the following conditions:

1. $p(\cdot)$ is non-decreasing on $[0, +\infty)$;
2. There exists $\tau > 0$ and a constant $C_0 > 0$ such that $p(t) \geq p(\tau) - C_0(\tau - t)^2$ holds for any $0 \leq t < \tau$;
3. For the aforementioned $\tau$, $p(t_1) + p(t_2) \geq p(t_1 + t_2)$ holds for any $t_1, t_2 \in [0, \tau)$;
4. $p(t) + p(\tau - t) > p(\tau)$ holds for any $0 < t < \tau$,

then the optimization problem (1) is NP-hard.

**Theorem 3 (Theorem 3.3 in Huo and Chen (2010))** If $q = 2$, $A$ is of full row rank, and $p(\cdot)$ satisfies the following conditions:

1. $p(\cdot)$ is non-decreasing on $[0, +\infty)$;
2. $p(\cdot)$ is strictly concave on $[0, +\infty)$;
3. $p(\cdot)$ satisfies the Lipschitz condition: there exists a constant $C_1 > 0$ such that $|p(t_1) - p(t_2)| \leq C_1|t_1 - t_2|$ holds for any $t_1, t_2 > 0$,

then the optimization problem (1) is NP-hard.

**Theorem 4 (Theorem 4.1 in Huo and Chen (2010))** If $q = 1$ and $p(\cdot)$ satisfies the following conditions:

1. $p(\cdot)$ is non-decreasing on $[0, +\infty)$;
2. There exists $\tau > 0$ such that $p(\cdot)$ is concave on $[0, 2\tau]$;
3. For the aforementioned $\tau$, $p(t) + p(\tau - t) > p(\tau)$ holds for any $0 < t < \tau$,
then the optimization problem (1) is NP-hard.

Now we compare our results to theirs. First we comment that we do have an additional condition than theirs that $p(\cdot)$ has a smooth part (second-order continuously differentiable) around $\tau$. However, in practice, $p(\cdot)$ usually satisfies such a condition, e.g., in all of our examples shown in Section 1. Except for this, our Theorem 1 is equivalent to their Theorem 4 in the case of $q = 1$. However, for the case of $q = 2$, our Theorem 1 requires weaker conditions on $p(\cdot)$ than theirs. In particular, among the examples listed in the end of last section, Examples 2 and 5 do not satisfy the conditions in either of their theorems. Moreover, our result applies to any $q \geq 1$ and obtains strong NP-hardness. It is worth pointing out that strong NP-hardness is meaningful since it rules out the possibility of having a fully polynomial-time approximation scheme (FPTAS) unless $P = NP$. Also, we use a different reduction technique in our proof.

The second related work is Chen et al. (2014) in which the authors proved that problem (1) with $p(t) = t^p$ is strongly NP-hard for any given $0 \leq p < 1$, $q \geq 1$, and $\lambda > 0$. To show the difference between our work and theirs, we note that Theorem 1 is much more general, with the result in Chen et al. (2014) only applies to Example 2 listed in the last section.

Finally, in a concurrent work by Bian and Chen (2014), the authors considered the problem (1) with $q = 2$ and $p(t) = \phi(t^\gamma)$, where $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$. They proved the following result:

**Theorem 5 (Theorem 3 in Bian and Chen (2014))** If $q = 2$, and $\phi(\cdot)$ satisfies the following conditions:
1. $\phi(\cdot)$ is non-decreasing on $[0, +\infty)$;
2. $\phi(\cdot)$ is concave on $[0, +\infty)$;
3. $\phi(\cdot)$ is continuously differentiable and $\phi'(\cdot)$ is locally Lipschitz continuous on $(0, +\infty)$,
then for any given $0 < \gamma < 1$, the minimization problem (1) is strongly NP-hard; when $\gamma = 1$, if further that
4. $\phi(\cdot)$ is strictly concave in an open interval on $(0, +\infty)$,
then the minimization problem (1) is strongly NP-hard.

For any given penalty function $p(t) = \phi(t^\gamma)$ such that the above result applies, if $0 < \gamma < 1$, then since $\phi(\cdot)$ is concave on $(0, +\infty)$, $p(\cdot)$ is strictly concave on $(0, +\infty)$; if $\gamma = 1$, then $p(t) = \phi(t)$, and thus $p(\cdot)$ is strictly concave in an open interval. Therefore, their result requires $p(t)$ to be at least locally strictly concave, while ours does not. In particular, among the examples listed in the last section, their results do not apply to Examples 1 and 5. Moreover, our result applies to any $q \geq 1$.

To summarize, our result provides the most general conditions to date for the strong NP-hardness of the regularized $L_q$-minimization problem with concave penalty function. In addition, the conditions in our result are easy to verify thus it might be handy to use in practice.

---

1. This condition is used to guarantee that the reduction from a strongly NP-hard problem to our problem is a polynomial-time reduction.
3 Proof of Theorem 1

In this section, we prove Theorem 1. The proofs of all the lemmas can be found at the end of this section.

We first illustrate several properties of the penalty function if it satisfies the conditions in Theorem 1.

Lemma 6 If $p(t)$ satisfies the conditions in Theorem 1, then for any $l \geq 2$, and any $t_1, t_2, \ldots, t_l \in \mathbb{R}$, we have $p(|t_1|) + \cdots + p(|t_l|) \leq \min\{p(|t_1 + \cdots + t_l|), p(\tau)\}$.

Lemma 7 If $p(t)$ satisfies the conditions in Theorem 1, then there exists $\tau_0 \in (0, \tau)$ such that $p(\cdot)$ is concave but not linear on $[0, \tau_0]$ and is twice continuously differentiable on $[\tau_0, \tau]$. Furthermore, for any $t \in (\tau_0, \tau)$, let $\delta = \min\{\tau_0/3, \hat{\tau} - \tau_0, \tau - \hat{\tau}\}$. Then for any $\delta \in (0, \delta)$ $l \geq 2$, and any $t_1, t_2, \ldots, t_l$ such that $t_1 + \cdots + t_l = \hat{\tau}$, we have

$$p(|t_1|) + \cdots + p(|t_l|) < p(\hat{\tau}) + C_1 \delta$$

only if $|t_i - \hat{\tau}| < \delta$ for some $i$ while $|t_j| < \delta$ for all $j \neq i$, where $C_1 = \frac{p(\tau_0/3) + p(2\tau_0/3) - p(\tau_0)}{\tau_0/3} > 0$.

In our proof of Theorem 1 we will consider the following function $g_{\theta, \mu}(t) := p(|t|) + \theta \cdot |t|^q + \mu \cdot |\hat{\tau} - t|^q$ with $\theta, \mu > 0$, where $\hat{\tau}$ is a fixed rational number in $(\tau_0, \tau)$. We have the following lemma about $g_{\theta, \mu}(t)$.

Lemma 8 If $p(t)$ satisfies the conditions in Theorem 1, $q > 1$, and $\tau_0$ satisfies the properties in Lemma 4, then there exist $\theta > 0$ and $\mu > 0$ such that for any $\theta \geq \theta_0$ and $\mu \geq \mu_0$, the following properties are satisfied:

1. $g_{\theta, \mu}''(t) \geq 1$ for any $t \in [\tau_0, \tau]$;
2. $g_{\theta, \mu}(t)$ has a unique global minimizer $t^*(\theta, \mu) \in (\tau_0, \tau)$;
3. Let $\hat{\tau} = \min\{t^*(\theta, \mu) - \tau_0, \tau - t^*(\theta, \mu), 1\}$, then for any $\delta \in (0, \delta)$, we have $g_{\theta, \mu}(t) < h(\theta, \mu) + \delta^2$ only if $|t - t^*(\theta, \mu)| < \delta$, where $h(\theta, \mu)$ is the minimal value of $g_{\theta, \mu}(t)$.

Lemma 9 If $p(t)$ satisfies the conditions in Theorem 1, $q = 1$, and $\tau_0$ satisfies the properties in Lemma 7, then there exist $\mu > 0$ such that for any $\mu \geq \mu_0$, the following properties are satisfied:

1. $g_{0, \mu}'(t) < -1$ for any $t \in [\tau_0, \hat{\tau})$ and $g_{0, \mu}'(t) > 1$ for any $t \in (\hat{\tau}, \tau]$;
2. $g_{0, \mu}(t)$ has a unique global minimizer $t^*(0, \mu) = \hat{\tau} \in (\tau_0, \tau)$;
3. Let $\hat{\tau} = \min\{\hat{\tau} - \tau_0, \tau - \hat{\tau}, 1\}$, then for any $\delta \in (0, \hat{\delta})$, we have $g_{0, \mu}(t) < h(0, \mu) + \delta^2$ only if $|t - \hat{\tau}| < \delta$.

By combining the above results, we have the following lemma, which is useful in our proof of Theorem 1.

Lemma 10 Suppose $p(t)$ satisfies the conditions in Theorem 1 and $\tau_0$ satisfies the properties in Lemma 7. Let $h(\theta, \mu)$ and $t^*(\theta, \mu)$ be as defined in Lemma 8 and Lemma 9 respectively for the case $q > 1$ and $q = 1$. Then we can find $\theta$ and $\mu$ such that for any $l \geq 2$, $t_1, \ldots, t_l \in \mathbb{R}$,

$$\sum_{j=1}^l p(|t_j|) + \theta \cdot \sum_{j=1}^l t_j^q + \mu \cdot \left| \sum_{j=1}^l t_j - \hat{\tau} \right|^q \geq h(\theta, \mu).$$
Moreover, let \( \delta = \min \left\{ \frac{\tau}{2}, \frac{r^*(\theta, \mu) - \tau}{2}, \frac{\tau - r^*(\theta, \mu)}{2}, 1, C_1 \right\} \) where \( C_1 \) is defined in Lemma 7 then for any \( \delta \in (0, \delta) \), we have

\[
\sum_{j=1}^{l} p(|t_j|) + \theta \cdot \left\{ \sum_{j=1}^{l} t_j \right\}^q + \mu \cdot \left\{ \sum_{j=1}^{l} t_j - \hat{\tau} \right\}^q < h(\theta, \mu) + \delta^2
\]

holds only if \(|t_i - t^*(\theta, \mu)| < 2\delta\) for some \( i \) while \(|t_j| \leq \delta\) for all \( j \neq i \).

**Proof of Theorem 4.** We present a polynomial time reduction to problem (1) from the 3-partition problem, which is known to be strongly NP-hard (Garey and Johnson (1978, 1979)). The 3-partition problem can be described as follows:

- Given a multiset \( S \) of \( n = 3m \) integers \( \{b_1, \ldots, b_n\} \) with sum \( mB \), determine whether \( S \) can be partitioned into \( m \) subsets, such that the sum of numbers in each subset is equal to \( B \).

For any given instance of the 3-partition problem with \( b = (b_1, \ldots, b_n) \), we consider the minimization problem \( \min_{\bar{x}} F(\bar{x}) \) in the form of (1) with \( \bar{x} = [x_{ij}], 1 \leq i \leq n, 1 \leq j \leq m \), where

\[
F(\bar{x}) := \sum_{j=2}^{m} \sum_{i=1}^{n} b_i x_{ij} - \sum_{i=1}^{n} b_i x_{i1} \left\{ \sum_{i=1}^{n} p(|x_{ij}|) + \theta \cdot \left\{ \sum_{j=1}^{m} x_{ij} \right\}^q + \mu \cdot \left\{ \sum_{j=1}^{m} x_{ij} - \hat{\tau} \right\}^q \right\} + \lambda \sum_{i=1}^{n} \sum_{j=1}^{m} p(|x_{ij}|).
\]

Note that the lower bounds \( \theta, \mu \), and \( \mu \) only depend on the penalty function \( p(\cdot) \), we can choose \( \theta \geq \hat{\theta} \) and \( \mu \geq \hat{\mu} \) if \( q > 1 \), or \( \theta = 0 \) and \( \mu \geq \hat{\mu} \) if \( q = 1 \), such that \((\lambda \theta)^{1/q} \) and \((\lambda \mu)^{1/q} \) are both rational numbers. Since \( \hat{\tau} \) is also rational, all the coefficients of \( F(\bar{x}) \) are of finite size and independent of the input size of the given 3-partition instance. Therefore, the minimization problem \( \min_{\bar{x}} F(\bar{x}) \) has polynomial size with respect to the given 3-partition instance.

For any \( \bar{x} \), by Lemma 10

\[
F(\bar{x}) \geq 0 + \lambda \cdot \sum_{i=1}^{n} \left\{ \sum_{j=1}^{m} p(|x_{ij}|) + \theta \cdot \left\{ \sum_{j=1}^{m} x_{ij} \right\}^q + \mu \cdot \left\{ \sum_{j=1}^{m} x_{ij} - \hat{\tau} \right\}^q \right\} \geq n\lambda \cdot h(\theta, \mu). \quad (3)
\]

Now we claim that there exists an equitable partition to the 3-partition problem if and only if the optimal value of \( F(\bar{x}) \) equals \( n\lambda \cdot h(\theta, \mu) \). On one hand, if \( S \) can be equally partitioned into \( m \) subsets, then we define

\[
x_{ij} = \begin{cases} t^*(\theta, \mu) & \text{if } b_i \text{ belongs to the } j \text{th subset in the equal partition;} \\ 0 & \text{otherwise.} \end{cases}
\]

It can be easily verified that these \( x_{ij} \)'s satisfy \( F(\bar{x}) = n\lambda \cdot h(\theta, \mu) \). Then due to (3), we know that these \( x_{ij} \)'s provide an optimal solution to \( F(\bar{x}) \) with optimal value \( n\lambda \cdot h(\theta, \mu) \).

On the other hand, suppose the optimal value of \( F(\bar{x}) \) is \( n\lambda \cdot h(\theta, \mu) \), and there is a polynomial-time algorithm that solves (1). Then for

\[
\delta = \min \left\{ \frac{\tau_0}{2 \sum_{i=1}^{n} b_i}, \delta \right\} \quad \text{and} \quad \epsilon = \min\{\lambda \delta^2, (\tau_0/2)^q\}
\]
where
\[ \bar{\tau} = \min \left\{ \frac{\tau_0}{3}, \frac{t^*(\theta, \mu) - \tau_0}{2}, \frac{\tau - t^*(\theta, \mu)}{2}, p(\tau_0/3) + p(2\tau_0/3) - p(\tau_0) \right\}, \]
we are able to find a near-optimal solution \( \tilde{x} \) such that \( F(\tilde{x}) < n\lambda \cdot h(\theta, \mu) + \epsilon \) within a polynomial time of \( \log(1/\epsilon) \) and the size of \( F(\tilde{x}) \), which is polynomial with respect to the size of the given 3-partition instance. Now we show that we can find an equitable partition based on this near-optimal solution. By the definition of \( \epsilon \), \( F(\tilde{x}) < n\lambda \cdot h(\theta, \mu) + \epsilon \) implies
\[
\sum_{j=1}^{m} p(|x_{ij}|) + \theta \left| \sum_{j=1}^{m} x_{ij} \right|^q + \mu \left| \sum_{j=1}^{m} x_{ij} - \tau \right|^q < h(\theta, \mu) + \delta^2, \quad \forall i = 1, \ldots, n. \tag{4}
\]
According to Lemma\[10\] for each \( i = 1, \ldots, n \), (4) implies that there exists \( k \) such that \( |x_{ik} - t^*(\theta, \mu)| < 2\delta \) and \( |x_{ij}| < \delta \) for any \( j \neq k \). Now let
\[ y_{ij} = \begin{cases} t^*(\theta, \mu) & \text{if } |x_{ik} - t^*(\theta, \mu)| < 2\delta \\ 0 & \text{if } |x_{ij}| < \delta \end{cases}. \]
We define a partition by assigning \( b_i \) to the \( j \)-th subset \( S_j \) if \( y_{ij} = t^*(\theta, \mu) \). Note that this partition is well-defined since for each \( i \), by the definition of \( \delta \), there exists one and only one \( y_{i*} = t^*(\theta, \mu) \) while the others equal 0. Now we show that this is an equitable partition.

Note that for any \( j = 1, \ldots, m \), the difference between the sum of the \( j \)-th subset and the first subset is
\[
\left| \sum_{S_j} b_i - \sum_{S_1} b_i \right| = \left| \sum_{i=1}^{n} \frac{y_{ij}}{t^*(\theta, \mu)} \cdot b_i - \sum_{i=1}^{n} \frac{y_{i1}}{t^*(\theta, \mu)} \cdot b_i \right|
\]
\[
= \frac{1}{t^*(\theta, \mu)} \left| \sum_{i=1}^{n} b_i y_{ij} - \sum_{i=1}^{n} b_i y_{i1} \right|
\]
\[
\leq \frac{1}{t^*(\theta, \mu)} \left( \sum_{i=1}^{n} b_i \cdot |y_{ij} - x_{ij}| + \sum_{i=1}^{n} b_i \cdot |y_{i1} - x_{i1}| + \sum_{i=1}^{n} b_i x_{ij} - \sum_{i=1}^{n} b_i x_{i1} \right) \right). \]

By the definition of \( y_{ij} \), we have \( |y_{ij} - x_{ij}| < 2\delta \) for any \( i, j \). For the last term, since \( F(\tilde{x}) < n\lambda \cdot h(\theta, \mu) + \epsilon \), we know that
\[
\sum_{i=1}^{n} b_i x_{ij} - \sum_{i=1}^{n} b_i x_{i1} < \epsilon^{1/q} \leq \tau_0/2.
\]
Therefore, we have
\[
\left| \sum_{S_j} b_i - \sum_{S_1} b_i \right| < \frac{1}{t^*(\theta, \mu)} \left( 4\delta \sum_{i=1}^{n} b_i + \frac{\tau_0}{2} \right) \leq 1.
\]
Now since \( b_i \)'s are all integers, we must have \( \sum_{S_j} b_i = \sum_{S_1} b_i \), which means that the partition is equitable.

**Proof of Lemma**\[6\]

Let \( t = \min\{|t_1| + \cdots + t_i|, \tau\} \in [0, \tau] \), then it suffices to show that \( p(|t_1|) + \cdots + p(|t_i|) \geq p(t) \). Note that we have \(|t_1| + \cdots |t_i| \geq |t_1 + \cdots + t_i| \geq t \). Moreover,
since \( p(0) = 0 \) and \( p(\cdot) \) is concave on \([0, \tau]\), we must have \( p(\cdot) \) being subadditive, i.e., for any \( s_1, \ldots, s_l \geq 0 \) such that \( s_1 + \cdots + s_l \leq \tau \), we have \( p(s_1) + \cdots + p(s_l) \geq p(s_1 + \cdots + s_l) \). Combining both facts, we have

\[
\sum_{i=1}^{l} p(|t_i|) \geq \sum_{i=1}^{l} \left( \frac{t}{|t_1| + \cdots + |t_l|} \cdot |t_i| \right) \geq p(\sum_{i=1}^{l} \frac{t}{|t_1| + \cdots + |t_l|} \cdot |t_i|) = p(t),
\]

where the first inequality is due to monotonicity and the second is due to subadditivity of \( p(\cdot) \).

**Proof of Lemma\(^7\)** According to the conditions for \( p(\cdot) \), there exists \( \tau_2 < \tau \) such that \( p(\cdot) \) is twice continuously differentiable on \([\tau_2, \tau]\). We first show that there exists \( \tau_0 \in (\tau_2, \tau) \) such that \( p(\cdot) \) is concave but not linear on \([0, \tau_0]\). If otherwise, \( p(\cdot) \) must be a linear function on \([0, \tau]\), then since \( p(\cdot) \) is continuous at \( t = \tau \) where continuity follows from concavity, we must have \( p(\cdot) \) is a linear function on \([0, \tau]\), which contradicts with that \( p(\cdot) \) is not linear on \([0, \tau]\). In the following, we show that this \( \tau_0 \) satisfies the conditions in the lemma.

We first show that \( C_1 > 0 \). If otherwise, we have \( \frac{p(\tau_0/3) - p(0)}{\tau_0/3} \leq \frac{p(\tau_0) - p(2\tau_0/3)}{\tau_0/3} \). Since \( p(t) \) is concave, this must imply that \( p(t) \) is linear on \([0, \tau_0]\), which contradicts with that \( p(\cdot) \) is not linear on \([0, \tau_0]\).

Before proving the result, we first introduce two auxiliary functions. For any \( s \in [0, \tau_0] \), define \( \bar{\epsilon}(s) := p(t - s) + p(s) - p(t) \) and \( \epsilon(s) := p(\tau_0 - s) + p(s) - p(\tau_0) \). Note that they have the following properties:

(i) \( C_1 = \frac{\epsilon(\tau_0/3)}{\tau_0/3} \);

(ii) \( \epsilon(s) = \bar{\epsilon}(s) \): this is due to \( \bar{\epsilon}(s) - \epsilon(s) = (p(\tau_0) - p(\tau_0 - s)) - (p(t) - p(t - s)) \geq 0 \);

(iii) \( \epsilon(s)/s \) is non-decreasing in \( s \): this is due to

\[
\frac{\epsilon(s)}{s} = \frac{p(s) - p(0)}{s} - \frac{p(\tau_0) - p(\tau_0 - s)}{s}
\]

where \( \frac{p(s) - p(0)}{s} \) is non-increasing while \( \frac{p(\tau_0) - p(\tau_0 - s)}{s} \) is non-decreasing;

(iv) Combining (i) – (iii) above, for any \( s \in (0, \tau_0/3] \), we have

\[
p(s) \geq p(s) + p(t - s) - p(t) = \bar{\epsilon}(s) = \epsilon(s) \geq C_1 s.
\]

When \( s = \tau_0/3 \), this implies that \( p(\tau_0/3) + p(t - \tau_0/3) - p(t) \geq C_1 \cdot \tau_0/3 > C_1 \delta \).

Now we prove the last statement of Lemma\(^7\). Suppose \( t_1 + \cdots + t_l = \bar{t} \), and \( p(|t_i|) + \cdots + p(|t_l|) - p(\bar{t}) < C_1 \delta \). Without loss of generality, we assume \( t_1 \geq t_2 \geq \cdots \geq t_l \). Now it suffices to show that \( |\bar{t} - t_1| < \delta, t_2 < \delta, \) and \( t_l > -\delta \).

Denote \( T = \{t_1, \ldots, t_l\} \). For any \( S \subseteq T \), we use \( \sigma(S) \) to denote the sum of all the elements of \( S \). Now we show that \( \sigma(S) > -\delta \) for any \( S \). If otherwise, then \( \sum_{S \subseteq T} t_i \geq \bar{t} + \delta \geq \bar{t} \), and we have

\[
C_1 \delta > \sum_{S} p(|t_i|) + \sum_{S \subseteq T} p(|t_i|) - p(\bar{t}) \geq p(\delta) + p(\bar{t}) - p(\bar{t}) \geq C_1 \delta,
\]

where the second inequality is due to Lemma\(^6\) and the monotonicity of \( p(\cdot) \), and the third one is due to (iv) above. This is a contradiction. Note that by having \( S = \{t_1\} \), this result implies that \( t_1 > -\delta \). Also, by considering the complement of a subset, we have \( \sigma(S) = \sigma(T) - \sigma(S^c) < \bar{t} + \delta < \tau \) for any \( S \subseteq T \). This has two implications. First, according to Lemma\(^6\) we have \( \sum_{S} p(|t_i|) \geq p(\sum_{S} t_i) \); second, by letting \( S = \{t_1\} \), we have \( t_1 < \bar{t} + \delta \).
Now we show that $t_1 > \tilde{t} - \delta$, by sequentially showing that $t_1 > \tau_0/3$, $t_1 > \tilde{t} - \tau_0/3$, and then $t_1 > \tilde{t} - \delta$. If $t_1 \leq \tau_0/3$, then we have $|t_i| \leq \tau_0/3$ for any $i$. Then we can divide $T$ into two sets $T_1$ and $T_2$ such that $|\sigma(T_1) - \sigma(T_2)| \leq \tau_0/3$, thus $\sigma(T_1), \sigma(T_2) \in (t/2 - \tau_0/6, t/2 + \tau_0/6) \subseteq (\tau_0/3, \tilde{t} - \tau_0/3)$. Now we have

$$C_1 \delta > p \left( \left| \sum_{t_i \in T_1} t_i \right| \right) + p \left( \left| \sum_{t_i \in T_2} t_i \right| \right) - p(\tilde{t}) - p(\tilde{t} - \delta/3) > C_1 \delta,$$

which is a contradiction. Note that here the first inequality is due to Lemma 3 and the second one is due to the concavity of $p(\cdot)$.

Now we show that $t_1 > \tilde{t} - \tau_0/3$. If otherwise, since we have proved that $t_1 \geq \tau_0/3$, we have $t_1 \in [\tau_0/3, \tilde{t} - \tau_0/3]$. Now by letting $T_1 = \{t_1\}$ and $T_2 = T - T_1$, we have $\sigma(T_1), \sigma(T_2) \in (\tau_0/3, \tilde{t} - \tau_0/3)$, and contradiction arises in the same way as in the previous case.

Now we show that $t_1 > \tilde{t} - \delta$, which is equivalent to showing that $t_2 = t_2 + \cdots + t_i = \tilde{t} - t_1 < \delta$. If $\tilde{t}_2 \geq \delta$, then due to subadditivity, concavity, and (iv) above, we have

$$C_1 \delta > p(|t_1|) + p(|\tilde{t}_2|) - p(\tilde{t}) \geq p(\tilde{t} - \delta) + p(\delta) - p(\tilde{t}) \geq C_1 \delta,$$

which is a contradiction.

Now to complete the proof, the only last thing we need to show is that $t_2 < \delta$. If $t_2 \geq \delta$, then due to subadditivity and concavity, we have

$$C_1 \delta > p(|t_2|) + p(|\tilde{t} - t_2|) - p(\tilde{t}) \geq p(\delta) + p(\tilde{t} - \delta) - p(\tilde{t}) \geq C_1 \delta,$$

which is a contradiction. 

\[ \square \]

**Proof of Lemma 8.** According to Lemma 7, $p(\cdot)$ is twice continuously differentiable on $[\tau_0, \tau]$, thus there exists $K > 0$ such that $p''(t) \geq -K$ for any $t \in [\tau_0, \tau]$. Now we take $\theta = \frac{q}{q(q-1)\min_{\tilde{t}} \frac{\tilde{t}^{-q}}{(\tilde{t}^{-q} - p')^{1+K}}, \mu = \frac{\tilde{t}^{-q} + \tilde{t}^{-q}}{q(q-1)}},$ and verify the results in the lemma.

For the first result, we have for any $t \in [\tau_0, \tau]$,

$$g_{\theta, \mu}(t) = p''(t) + \theta q(q - 1)t^{q-2} + \mu q(q - 1)|\tilde{t} - t|^q \geq -K + \tilde{t}^{-q} + 0 \geq 1,$$

thus $g''_{\theta, \mu}(t) \geq 1$ for any $t \in [\tau_0, \tau]$.

Now we show the result of unique minimizer. Since $g_{\theta, \mu}(t)$ is strictly increasing on $[\tilde{t}, +\infty)$, any global minimizer must lie in $(-\infty, \tilde{t})$. Moreover, for any $t \in (\infty, \tau_0)$, we have

$$g_{\theta, \mu}(t) > 0 + \theta \mu \cdot |\tau_0 - \tilde{t}|^q = \theta \tilde{t}^q,$$

thus any global minimizer must lie within $(\tau_0, \tilde{t})$. Now since $g''(t) \geq 1$ for any $t \in (\tau_0, \tau)$, we know that $g(\cdot)$ is strictly convex thereon, thus the global minimizer of $g_{\theta, \mu}(t)$ on $[\tau_0, \tau]$ exists and is unique. Denote the minimizer on $[\tau_0, \tau]$ by $t^*(\theta, \mu)$, then according to the previous discussion, $t^*(\theta, \mu)$ must also be the global minimizer of $g_{\theta, \mu}(t)$ on $\mathbb{R}$.

Now we show the last statement. Suppose that $g_{\theta, \mu}(\tilde{t}) < h(\theta, \mu) + \delta^2$ for some $\delta \in (0, \delta)$. We first consider the case where $\tilde{t} \in [\tau_0, \tau]$. According to the mean-value theorem, there exists $\tilde{t}$ between $\tilde{t}$ and $t^*(\theta, \mu)$ such that

$$g_{\theta, \mu}(\tilde{t}) = g_{\theta, \mu}(t^*(\theta, \mu)) + g''(\tilde{t})(\tilde{t} - t^*(\theta, \mu))^2 \geq h(\theta, \mu) + (\tilde{t} - t^*(\theta, \mu))^2.$$
Therefore, a necessary condition for \( g_{\theta, \mu}(\hat{t}) < h(\theta, \mu) + \delta^2 \) is that \(|\hat{t} - t^*(\theta, \mu)| < \delta\). Note that this implies \( g_{\theta, \mu}(\tau) \geq h(\theta, \mu) + \delta^2 \). Now to complete the proof, we only need to show that \( g_{\theta, \mu}(\tau) \geq h(\theta, \mu) + \delta^2 \) for any \( t \in (-\infty, \tau_0) \cup [\tau, +\infty) \). The inequality with \( t \in (-\infty, \tau_0) \) has been proved in (5). And for any \( t \in [\tau, +\infty) \), we have \( g_{\theta, \mu}(t) \geq g_{\theta, \mu}(\tau) \geq h(\theta, \mu) + \delta^2 \). Therefore, the proof is complete.

**Proof of Lemma 9.** We take \( \hat{\mu} = \max \left\{ 1 + p'(\tau_0), \frac{p'(\tau) + 1}{\tau - \tau_0} \right\} \) and verify the results in the lemma. Note that we have \( p(\cdot) \) being twice continuously differentiable on \([\tau_0, \tau]\) thus \( p'(\tau) \) is well-defined.

For any \( t \in [\tau_0, \hat{\tau}] \), we have \( g_{\theta, \mu}'(t) = p'(t) - \mu \leq p'(\tau_0) - \hat{\mu} \leq -1 \); and for any \( t \in (\hat{\tau}, \tau) \), we have \( g_{\theta, \mu}'(t) = p'(t) + \mu \geq 0 + \hat{\mu} \geq 1 \). Therefore, the first property in Lemma 9 holds.

Now we show the result of unique minimizer. Since \( g_{\theta, \mu}(t) \) is strictly increasing on \([\hat{\tau}, +\infty)\), any global minimizer must lie in \((-\infty, \hat{\tau})\). Moreover, for any \( t \in (-\infty, \tau_0) \), we have

\[
g_{\theta, \mu}(t) \geq 0 + \hat{\mu} \cdot |\tau_0 - \hat{\tau}| \geq p(\hat{\tau}) + 1 = g_{\theta, \mu}(\hat{\tau}) + 1,
\]

(6)

Thus any global minimizer must lie within \((\tau_0, \hat{\tau})\). Now since \( g_{\theta, \mu}'(t) < -1 \) for any \( t \in [\tau_0, \hat{\tau}) \), the global minimizer of \( g_{\theta, \mu}(\cdot) \) is \( t^*(0, \mu) = \hat{\tau} \) and is unique.

Now we show the last statement. Suppose that \( g_{\theta, \mu}(\hat{t}) < h(0, \mu) + \delta^2 \) for some \( \delta \in (0, \hat{\delta}) \). Again we first consider the case where \( \hat{t} \in [\tau_0, \tau] \). When \( \hat{t} \in [\hat{\tau}, \tau] \), since \( g_{\theta, \mu}'(t) > 1 \), we have \( 0 + \hat{\mu} \geq 1 \). Therefore, a necessary condition for \( g_{\theta, \mu}(\hat{t}) < h(0, \mu) + \delta^2 \) is that \(|\hat{t} - \hat{\tau}| < \delta^2 < \delta\). Note that this implies \( g_{\theta, \mu}(\tau) \geq h(0, \mu) + \delta^2 \). Now to complete the proof, we only need to show that \( g_{\theta, \mu}(t) > h(0, \mu) + \delta^2 \) for any \( t \in (-\infty, \tau_0) \cup [\tau, +\infty) \). The inequality with \( t \in (-\infty, \tau_0) \) has been proved in (5). And for any \( t \in [\tau, +\infty) \), we have \( g_{\theta, \mu}(t) \geq g_{\theta, \mu}(\tau) \geq h(0, \mu) + \delta^2 \). Therefore, the proof is complete.

**Proof of Lemma 10.** If \( q > 1 \), then we can find \( \theta \) and \( \mu \) such that the properties in Lemma 8 is satisfied; if \( q = 1 \), then we can set \( \theta = 0 \) and find \( \mu \) such that the properties in Lemma 9 is satisfied.

Now we first prove the desired inequality in two cases. In the first case, we suppose that \( |\sum_{j=1}^{l} t_j| > \tau \). Then due to Lemma 8 we have \( \sum_{j=1}^{l} p(|t_j|) \geq p(\tau) \), thus

\[
\sum_{j=1}^{l} p(|t_j|) + \theta \cdot \sum_{j=1}^{l} t_j \cdot \left| \sum_{j=1}^{l} t_j - \hat{\tau} \right|^{q} + \mu \cdot \left| \sum_{j=1}^{l} t_j - \hat{\tau} \right|^{q} > p(\tau) + \theta \tau^q + \mu |\tau - \hat{\tau}|^q = g_{\theta, \mu}(\tau) > h(\theta, \mu) + \delta^2 \quad (7)
\]

where the last inequality is proved in Lemmas 8 and 9. In the second case, we suppose that \( |\sum_{j=1}^{l} t_j| \leq \tau \). Then according to Lemma 6 we have \( \sum_{j=1}^{l} p(|t_j|) \geq p \left( \left| \sum_{j=1}^{l} t_j \right| \right) \), thus

\[
\sum_{j=1}^{l} p(|t_j|) + \theta \cdot \sum_{j=1}^{l} t_j \cdot \left| \sum_{j=1}^{l} t_j - \hat{\tau} \right|^{q} + \mu \cdot \left| \sum_{j=1}^{l} t_j - \hat{\tau} \right|^{q} \geq g_{\theta, \mu} \left( \sum_{j=1}^{l} t_j \right) \geq h(\theta, \mu), \quad (8)
\]

where the second inequality is due to Lemmas 8 and 9.

Now we prove the “only if” statement. Suppose we have \( t_1, \ldots, t_l \in \mathbb{R} \) such that (2) holds. Now according to (7), we must have \( |\sum_{j=1}^{l} t_j| \leq \tau \), and combining (8), we have \( g_{\theta, \mu} \left( \sum_{j=1}^{l} t_j \right) < h(\theta, \mu) + \delta^2 \). Then we have \( |\sum_{j=1}^{l} t_j - t^*(\theta, \mu)| < \delta \) according to Lemmas 8 and 9 thus \( \hat{t} := \)
\[ \sum_{j=1}^{l} t_j \in [\tau_0, \tau]. \]
Moreover, in order for (2) to hold, we must also have \[ \sum_{j=1}^{l} p(|t_j|) - p(\tilde{t}) \leq \delta^2 \leq C_1 \delta. \] Then according to Lemma 7 we must have \[ |t_i - \tilde{t}| < \delta \] for some \( i \) while \[ |t_j| < \delta \] for all \( j \neq i \). Now since \( |\tilde{t} - t^*(\theta, \mu)| < \delta \), we have \[ |t_i - t^*(\theta, \mu)| < 2\delta, \] which completes the proof. \( \Box \)

Acknowledgements
We thank Jianqing Fan, Mengdi Wang, and Yuan Yao for their valuable comments and discussions. We also thank the referee and the associate editor for their insightful comments.

References

Antoniadis, A., J. Fan. 2001. Regularization of wavelet approximations. *Journal of the American Statistical Association* 96(455) 939–967.

Bian, W., X. Chen. 2014. Optimality conditions and complexity for non-Lipschitz constrained optimization problems. *Preprint*.

Candes, E., M. Wakin, S. Boyd. 2008. Enhancing sparsity by reweighted \( L_1 \) minimization. *Journal of Fourier Analysis and Applications* 14(5-6) 877–905.

Chen, X., D. Ge, Z. Wang, Y. Ye. 2014. Complexity of unconstrained \( L_2 - L_p \) minimization. *Mathematical Programming* 143(1-2) 371–383.

Fan, J., R. Li. 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* 96(456) 1348–1360.

Fan, J., J. Lv. 2010. A selective overview of variable selection in high dimensional feature space. *Statistica Sinica* 20(1) 101–148.

Fan, J., L. Xue, H. Zou. 2014. Strong oracle optimality of folded concave penalized estimation. *The Annals of Statistics* 42(3) 819–849.

Frank, I. E., J. H. Friedman. 1993. A statistical view of some chemometrics regression tools (with discussion). *Technometrics* 35(2) 109–148.

Garey, M. R., D. S. Johnson. 1978. “Strong” NP-Completeness results: Motivation, examples, and implications. *Journal of the ACM* 25(3) 499–508.

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

Huo, X., J. Chen. 2010. Complexity of penalized likelihood estimation. *Journal of Statistical Computation and Simulations* 80(7) 747–759.

Lv, J., Y. Fan. 2009. A unified approach to model selection and sparse recovery using regularized least squares. *The Annals of Statistics* 37(6A) 3498–3528.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B* 58(1) 267–288.

Zhang, C.-H. 2010a. Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics* 38(2) 894–942.

Zhang, T. 2010b. Analysis of multi-stage convex relaxation for sparse regularization. *Journal of Machine Learning Research* 11 1081–1107.

Zou, H., R. Li. 2008. One-step sparse estimates in nonconcave penalized likelihood models. *The Annals of Statistics* 36(4) 1509–1533.