An improved Algorithm for the $L_2 - L_p$ Minimization

Dongdong Ge

Research Center for Management Science and Information Analytics
Shanghai University of Finance and Economics, China

Joint work with Rongchuan He, Simai He

September 3, 2014
Outline

▶ Problem Formulation
▶ Background & Applications
▶ Previous Work
▶ Algorithm & Analysis
The Basic Model

Consider a non-Lipschitz and nonconvex problem:

Minimize \( h(x) = \frac{1}{2}x^T Q x + a^T x + c + \lambda \sum_i x_i^p \)  \hspace{1cm} (1)

Subject to \( x \geq 0 \)

\( Q \in \mathbb{R}^{n \times n}, 0 \preceq Q \prec \beta I, 0 < p < 1. \)

A generalization of the \( L_2 - L_p \) minimization problem:

Minimize \( \frac{1}{2} \| Ax - b \|^2 + \lambda \sum_i x_i^p \)  \hspace{1cm} (2)

Subject to \( x \geq 0 \)
Theorem
For any $\epsilon \in (0, 1)$, the algorithm obtains an $\epsilon - KKT$ point of (1) in no more than $O((n + \frac{h(x_0)}{M}) \log \frac{1}{\epsilon})$ steps.
Applications

- Signal Processing, Image Reconstruction
- Influence Maximization in Social Network
- Customer Behavior Study: Products Assortment
- Financial Engineering
- Flexible Supply Chain
- Military; Game Theory...
Consider the problem:

\[
\text{Minimize } \quad p(x) = \sum_{1 \leq j \leq n} x_j^p \\
\text{Subject to } \quad Ax = b, \quad x \geq 0,
\]

- NP-hard when \( p = 0 \)
- Strongly NP-Hard when \( 0 < p < 1 \) [5]
- \( \exists \) an FPTAS in \( O\left(\frac{n}{\epsilon} \log \frac{1}{\epsilon}\right) \) iterations to approach \( \epsilon \)-stationary point [5]
\[ \min_x f(x) = \|Ax - b\|_2^2 + \lambda \|x\|_p^p, \]

- Lasso Regression when \( p = 1 \).
- Bridge Regression when \( 0 < p < 1 \); Strong NP-Hard [4]

**Theorem**

[3] (Chen et al. 2009) Let \( \beta \) be a positive constant such that for a local minimizer \( x^* \) : \( \|A^T(Ax^* - b)\| < \beta \), and let \( L = \left( \frac{\lambda p}{2\beta} \right)^{\frac{1}{1-p}} \).

Then, the local minimizer \( x^* \) possesses the property

\[ x^*_j \in (-L, L) \Rightarrow x^*_j = 0, j \in \mathcal{N}. \]
The Hardness Results

- the $L_q$-$L_p$ minimization problem:

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

(4)

is strongly NP-hard for any given $0 \leq p < 1$, $q \geq 1$ and $\lambda > 0$.

- 

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

(5)

is strongly NP-hard for any given $0 < p < 1$, $q \geq 1$, $\lambda > 0$ and $\epsilon > 0$. 
Previous Work

- Bian et al. [1]: non-Lipschitz and non-convex minimization with box constraints by affine scaling.
- The first order approximation: obtain an $\epsilon$-KKT point in $O(\epsilon^{-2})$ steps.
- The second order approximation: $O(\epsilon^{-\frac{3}{2}})$; a higher computational complexity at each iteration.
- They show that their method takes at most $O(\epsilon^{-2})$ steps to find an $\epsilon$-KKT solution.
Minimize \[ h(x) = \frac{1}{2} x^T Q x + a^T x + c + \lambda \sum_i x_i^p \]

Subject to \[ x \geq 0 \]

\textbf{Definition}

For a given \( \epsilon \in (0, 1) \), we call \( x^* \in F_p \) an \( \epsilon - KKT \) point of (1), if there is \( y^* \geq 0 \), such that

\[
\begin{align*}
  &x^* \in F_p \\
  &\|[\nabla h(x^*) - y^*]_i\| \leq \epsilon, \quad x_i \neq 0 \\
  &y^* \geq 0 \\
  &(y^*)^T x^* \leq \epsilon
\end{align*}
\]
Assumptions & Notations

- Assumption 1: The optimal value of problem (1) is lower bounded by 0.
- Assumption 2: For any $x^0 \geq 0$, there exists $\gamma$ such that $\sup\{\|x\|_\infty | h(x) \leq h(x^0)\} \leq \gamma$.
- $h(x) = f(x) + g(x)$
- $f(x) = \frac{1}{2}\beta x^T x + a^T x + c$, $g(x) = \lambda \sum_i x_i^p + \frac{1}{2}x^T (Q - \beta I)x$.
- $d_z = \bar{z} - z$
Potential Function

(MQP) : \( \min \ L_z(x) = f(x) + \nabla g(z)(x - z) \)
\[ x \geq 0 \] (7)

Let \( \bar{z} \) be the minimizer of (MQP), then the potential function is

\[ \Delta L(z) = L_z(z) - L_z(\bar{z}) \] (8)

Lemma

For any \( z \geq 0 \), if \( \Delta L(z) \leq \frac{\epsilon^2}{2\|Q^{1/2}\|^2} \), then \( z \) is an \( \epsilon - KKT \) point of (1).
A 3-Criteria Algorithm

**Require:** \( \epsilon \in (0, 1), \ x^0 \in F_p \)

Fix \( s > 0, \tau > 0 \) and \( L > 0 \) (will specify later)

\( k = 0 \)

**while** Not Stop **do**

Case 1:

**if** \( x_i^k \leq L \) for an index \( i \), **then**

Update \( x^{k+1} \) by Removing \( x_i^k \) from (1)

**end if**

Case 2:

**if** \( x^k > L \) and \( (d^k)^T \nabla^2 h(x^k) d^k \leq \tau \|d^k\|^2 \) **then**

\[ t_k = \max \{ t | x^k + td^k \geq 0, x^k - td^k \geq 0 \} \]

\( x^{k+1} = \arg\min_{x \in \{x^k + t_k d^k, x^k - t_k d^k\}} h(x) \)

**end if**

**end while**
A 3-Criteria Algorithm: Continued

\textbf{while} Not Stop \textbf{do}
\begin{enumerate}
    \item [Case 3:]
        \begin{enumerate}
            \item \textbf{if} $x^k > L$ and $(d^k)^T \nabla^2 h(x^k) d^k > \tau \|d^k\|^2$ \textbf{then}
                \begin{equation*}
                    x^{k+1} = x^k + sd^k
                \end{equation*}
            \end{enumerate}
    \end{enumerate}
\begin{enumerate}
    \item \textbf{if} $x^k = 0$ or $\Delta L_k \leq \frac{\epsilon^2}{2\|Q^2\|^2}$ \textbf{then}
                \begin{equation*}
                    x^* = x^k; \text{ Stop; Stop}
                \end{equation*}
    \end{enumerate}
\begin{enumerate}
    \item \textbf{else}
                \begin{equation*}
                    k = k + 1
                \end{equation*}
    \end{enumerate}
\textbf{end if}
\textbf{end while}
Complexity Analysis

Table: Summary of 3-Criteria Algorithm

<table>
<thead>
<tr>
<th></th>
<th>Objective $h(x^k)$</th>
<th>Potential func. $\Delta L_k$</th>
<th>$|x^k|_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>C 1</td>
<td>nonincreasing</td>
<td>$\leq h(x_0)$</td>
<td>decreased by 1</td>
</tr>
<tr>
<td>C 2</td>
<td>$h(x^k) - h(x^{k+1}) \geq M$</td>
<td>$\leq h(x_0)$</td>
<td>nonincreasing</td>
</tr>
<tr>
<td>C 3</td>
<td>nonincreasing</td>
<td>Shrink at $(1 - s\delta)$</td>
<td>nonincreasing</td>
</tr>
</tbody>
</table>

- Case 1: nearly zero component. The cardinality of the solution is decreased. $\leq n$ times.
- Case 2: non-strongly convex. The decrement of objective value: $\leq \left\lfloor \frac{h(x^0)}{M} \right\rfloor$ times.
- Case 3: strongly convex. The value of potential function, $\leq O(\log \frac{1}{\epsilon})$ steps.
Lemma

Case 1: For any \( k \geq 0 \), if

\[
0 < L < \min\{ (n\|Q_i\|\gamma + \frac{\alpha}{2} - \frac{Q_{ii}}{2} - a_i)^{\frac{1}{p-1}}, \forall i \}, \|x^k\|_{\infty} \leq \gamma, \text{ and there exists } i \text{ such that } x_i \text{ is in } (1) \text{ and } x_i^k \leq L, \text{ then let }
\]

\[
\begin{align*}
x_{j}^{k+1} &= x_j^k, \quad j \neq i \\
x_{i}^{k+1} &= 0, \quad j = i,
\end{align*}
\]

(9)

and we have \( h(x^k) - h(x^{k+1}) > 0 \).
Non-Strongly Convex

Lemma

Case 2: For any $k \geq 0$ and $L > 0$, if $x^k > L$, and

$0 < \tau < \frac{2p(1-p)(2-p)(3-p)L^p}{4!n\gamma^2}$,

$(d^k)^T \nabla^2 h(x^k)d^k \leq \tau \|d^k\|^2$, $\|x^k\|_\infty \leq \gamma$,

let $x^{k+1} = \text{argmin}_{x \in \{x^k + t_k d^k \geq 0, x^k - t_k d^k \geq 0\}} h(x)$,

Then

$h(x^k) - h(x^{k+1}) \geq M > 0$,

where $M = \frac{1}{4!} p(1-p)(2-p)(3-p)L^p - \frac{1}{2} \tau n\gamma^2$. 

Dongdong Ge
An improved Algorithm for the $L_2 - L_p$ Minimization
Lemma

Case 3: For any \( k \geq 0, \tau > 0 \) and \( L > 0 \), if \( x^k > L \), and

\[ (d^k)^T \nabla^2 h(x^k) d^k > \tau \| d^k \|^2 \]

\[ \| x^k \|_\infty \leq \gamma, \]

let \( x^{k+1} = x^k + s d_k \)

we have

\[ h(x^k) - h(x^{k+1}) \geq 0, \]

\[ \Delta L_{k+1} \leq (1 - s \delta) \Delta L_k, \]

where \( 0 < \delta < \min\{\frac{2 \tau}{\beta}, 1\} \), and

\[ 0 < s \leq \min\{\frac{\alpha}{u} (\frac{\tau}{\beta} - \frac{\delta}{2}), w, 1\} (0 < w < 1, \]

\[ u = \frac{\beta}{2} + \frac{1}{\alpha} \left[ p(1 - p)(L(1 - w))^{p-2}\right]^2 + \alpha^2 \]
Convergence Theorem

Theorem

For any $\epsilon \in (0, 1)$, the algorithm obtains an $\epsilon - KKT$ point of (1) in no more than $O\left((n + \frac{h(x_0)}{M}) \log \frac{1}{\epsilon}\right)$ steps.

Proof.

- During the process, the objective function and the cardinality of the solution keep decreasing.
- The potential function value may come back in Case 1 and 2.
- But Case 1 and 2 only happen at most $O\left((n + \frac{h(x_0)}{M})\right)$ times.
- Using Pigeonhole theorem, easy to prove $O\left((n + \frac{h(x_0)}{M}) \log \frac{1}{\epsilon}\right)$ iterations.
W. Bian, X. Chen and Y. Ye,

W. Bian, X. Chen,

X. Chen, F. Xu and Y. Ye,

X. Chen, D. Ge, Z. Wang and Y. Ye,

D. Ge, X. Jiang and Y. Ye,