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:

$$\begin{aligned} \text{Minimize} \quad & h(x) = \frac{1}{2}x^T Qx + a^T x + c + \lambda \sum_i x_i^p \\ \text{Subject to} \quad & x \geq 0 \end{aligned} \tag{1}$$

- $Q \in R^{n \times n}, 0 \preceq Q \prec \beta I, 0 < p < 1.$
- A generalization of the $L_2 - L_p$ minimization problem:

$$\begin{aligned} \text{Minimize} \quad & \frac{1}{2}\|Ax - b\|^2 + \lambda \sum_i x_i^p \\ \text{Subject to} \quad & x \geq 0 \end{aligned} \tag{2}$$

Main Theorem

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 Recontruction
- Influence Maximization in Social Network
- Customer Behavior Study: Products Assortment
- Financial Engineering
- Flexible Supply Chain
- Military; Game Theory...

L_p Minimization

- Consider the problem:

$$\begin{aligned} \text{Minimize} \quad & p(x) = \sum_{1 \le j \le n} x_j^p \\ \text{Subject to} \quad & Ax = b, \\ & x \ge 0, \end{aligned} \tag{3}$$

- NP-hard when $p = 0$
- Strongly NP-Hard when $0 < p < 1$ [5]
- $\exists$ an FPTAS in $O(\frac{n}{\epsilon} \log \frac{1}{\epsilon})$ iterations to approach ϵ-stationary point [5]

$L_2 - L_p$ Minimization Model

- $min_x f(x) = \|Ax - b\|_2^2 + \lambda \|x\|_p^p,$
- Lasso Regression when $p = 1$.
- Bridge Regression when $0 < p < 1$; Strong NP-Hard [4]

Theorem

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

$$\text{Minimize}_x \quad f_{q,p}(x) := \|Ax - b\|_q^q + \lambda \|x\|_p^p \tag{4}$$

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

$$\text{Minimize}_x \quad f_{q,p,\epsilon}(x) := \|Ax - b\|_q^q + \lambda \sum_{i=1}^{n} (|x_i| + \epsilon)^p \tag{5}$$

  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 ϵ-KKT point in $O(\epsilon^{-2})$ steps.
- The second order approximation: $O(\epsilon^{-\frac{3}{2}})$; a higher computational complexity at each iteration.
- Bian et al. [2] present an smoothing quadratic regularization algorithm for solving a class of unconstrained non-smooth non-convex problems.
- They show that their method takes at most $O(\epsilon^{-2})$ steps to find an ϵ-KKT solution.

ϵ-KKT Condition

$$\text{Minimize} \quad h(x) = \frac{1}{2}x^T Qx + a^T x + c + \lambda \sum_i x_i^p$$

$$\text{Subject to} \quad x \geq 0$$

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

$$\begin{aligned} x^* &\in F_p \\ \|[\nabla h(x^*) - y^*]_i\| &\leq \epsilon, \quad x_i \neq 0 \\ y^* &\geq 0 \\ (y^*)^T x^* &\leq \epsilon \end{aligned} \tag{6}$$

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 γ 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

- $$(MQP): \quad \min \quad L_z(x) = f(x) + \nabla g(z)(x - z) \\ x \geq 0 \tag{7}$$
- Let $\bar{z}$ be the minimizer of (MQP), then the potential function is
$$\Delta L(z) = L_z(z) - L_z(\bar{z}) \tag{8}$$

Lemma

For any $z \geq 0$, if $\Delta L(z) \leq \frac{\epsilon^2}{2\|Q^{\frac{1}{2}}\|^2}$, 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 $x_i^k \le L$ 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 $(d^k)^T \nabla^2 h(x^k) d^k \le \tau \|d^k\|^2$ **then**

$t_k = max\{t | x^k + td^k \ge 0, x^k - td^k \ge 0\}$

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

end if

end while

A 3-Criteria Algorithm: Continued

while Not Stop **do**

Case 3:

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

$x^{k+1} = x^k + sd^k$

end if

if $x^k = 0$ or $\Delta L_k \leq \frac{\epsilon^2}{2\|Q^{\frac{1}{2}}\|^2}$ **then**

$x^* = x^k$; Stop;

Stop

else

$k = k + 1$

end if

end while

Complexity Analysis

Table: Summary of 3-Criteria Algorithm

	Objective $h(x^k)$	Potential func. ΔL_k	$\|x^k\|_0$
C 1	nonincreasing	$\leq h(x_0)$	decreased by 1
C 2	$h(x^k) - h(x^{k+1}) \geq M$	$\leq h(x_0)$	nonincreasing
C 3	nonincreasing	Shrink at $(1 - s\delta)$	nonincreasing

- 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: $\leq \lfloor \frac{h(x^0)}{M} \rfloor$ times.
- Case 3: strongly convex. The value of potential function, $\leq O(\log \frac{1}{\epsilon})$ steps.

The cardinality Decrement

Lemma

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

$$\begin{cases} x_j^{k+1} = x_j^k, & j \neq i \\ x_j^{k+1} = 0, & j = i, \end{cases} \tag{9}$$

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* $x^{k+1} = argmin_{x \in \{x^k + t_k d^k \geq 0, x^k - t_k d^k \geq 0\}} h(x)$,
- *Then*
$$h(x^k) - h(x^{k+1}) \geq M > 0,$$
- *where* $M = \frac{1}{4!} p(1-p)(2-p)(3-p)L^p - \frac{1}{2}\tau n \gamma^2$.

Case 3: Strongly Convex

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* $x^{k+1} = x^k + sd_k$
- *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* $0 < s \leq min\{\frac{\alpha}{u}(\frac{\tau}{\beta} - \frac{\delta}{2}), w, 1\}(0 < w < 1,$ $u = \frac{\beta}{2} + \frac{1}{\alpha}[[p(1-p)(L(1-w))^{p-2}]^2 + \alpha^2]$

Convergence Theorem

Theorem

For any $\epsilon \in (0,1)$, *the algorithm obtains an* $\epsilon - KKT$ *point of (1) in no more than* $O((n + \frac{h(x_0)}{M}) \log \frac{1}{\epsilon})$ *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 $O((n + \frac{h(x_0)}{M}))$ times.
- Using Pigeonhole theorem, easy to prove $O((n + \frac{h(x_0)}{M}) \log \frac{1}{\epsilon})$ iterations.

□

W. Bian, X. Chen and Y. Ye,
Complexity Analysis of Interior Point Algorithms for Non-Lipschitz and Nonconvex Minimization, *Preprint*, 2012.

W. Bian, X. Chen,
Worst-Case Complexity of Smoothing Quadratic Regularization Methods for Non-Lipschitzian Optimization, *SIAM Journal on Optimization*, 2013.

X. Chen, F. Xu and Y. Ye,
Lower Bound Theory of Nonzero Entries in Solutions of ℓ_2-ℓ_p Minimization, *SIAM Journal on Scientific Computing*, 32(5), 2832-2852, 2010.

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

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