Composite L_q $(0 < q < 1)$ Minimization over Polyhedron

Ya-Feng Liu

State Key Laboratory of Scientific/Engineering Computing,
Institute of Computational Mathematics and Scientific/Engineering Computing,
Academy of Mathematics and Systems Science,
Chinese Academy of Sciences, Beijing, China
Email: yafliu@lsec.cc.ac.cn

Joint Work with Shiqian Ma, Yu-Hong Dai, and Shuzhong Zhang

2014 Workshop on Optimization for Modern Computation

Polyhedral Constrained Composite L_q Minimization

- Polyhedral constrained composite L_q $(0 < q < 1)$ minimization problem

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^N} \quad & F(x) := \|\max\{b - Ax, 0\}\|_q^q + h(x) \\ \text{s.t.} \quad & x \in \mathcal{X}. \end{aligned} \tag{1}$$

  - $A = [a_1, a_2, ..., a_M]^T \in \mathbb{R}^{M\times N}$, $b = [b_1, b_2, ..., b_M]^T \in \mathbb{R}^M$;
  - $h(x)$: continuously differentiable satisfying

$$\|\nabla h(x) - \nabla h(y)\|_2 \le L_h \|x - y\|_2, \ \forall \ x, y \in \mathcal{X};$$

  - $\mathcal{X} \subseteq \mathbb{R}^N$: a general polyhedral set.

Two Extremes

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^N} \quad & \|\max\{b - Ax, 0\}\|_q^q + h(x) \\ \text{s.t.} \quad & x \in \mathcal{X}. \end{aligned}$$

- as $q \to 0$, the above L_q minimization problem approaches

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^N} \quad & \|\max\{b - Ax, 0\}\|_0 + h(x) \\ \text{s.t.} \quad & x \in \mathcal{X}. \end{aligned}$$

- as $q \to 1$, the above L_q minimization problem approaches

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^N} \quad & \|\max\{b - Ax, 0\}\|_1 + h(x) \\ \text{s.t.} \quad & x \in \mathcal{X}. \end{aligned}$$

Outline

- Motivated Applications
- Related Works
- Exact Recovery
- Computational Complexity
- Optimality Conditions
- Algorithmic Framework & Analysis
- Simulation Results (NOT Covered)

Two Motivated Applications

Wireless Communications

- SINR at receiver k in the K-link SISO interference channel:

$$\text{SINR}_k := \frac{g_{kk} p_k}{\sum_{j \neq k} g_{kj} p_j + \eta_k} \geq \gamma_k, \;\; k = 1, 2, ..., K$$

$$\bar{p}_k \geq p_k \geq 0, \;\; k = 1, 2, ..., K$$

  - p_k : transmission power at transmitter k
  - $g_{kj} \geq 0$: channel gain from transmitter j to receiver k
  - $\eta_k > 0$: noise power of link k
  - $\gamma_k > 0$: SINR target of link k
  - $\bar{p}_k > 0$: power budget at transmitter k

Joint Power and Admission Control

- Infeasibility issues of the linear system

$$\mathrm{SINR}_k \geq \gamma_k,\ \ \bar{p}_k \geq p_k \geq 0,\ \ k = 1, 2, \ldots, K$$

  - mutual interference among different links
  - individual power budget constraints

Joint Power and Admission Control

- Infeasibility issues of the linear system

$$\mathrm{SINR}_k \geq \gamma_k,\ \ \bar{p}_k \geq p_k \geq 0,\ \ k = 1, 2, \ldots, K$$

  - mutual interference among different links
  - individual power budget constraints

- The admission control is necessary to determine the connections to be rejected.

Joint Power and Admission Control

- Infeasibility issues of the linear system

$$\text{SINR}_k \geq \gamma_k,\ \bar{p}_k \geq p_k \geq 0,\ k = 1, 2, \ldots, K$$

  - mutual interference among different links
  - individual power budget constraints

- The admission control is necessary to determine the connections to be rejected.

- Joint power and admission control (JPAC):

  - the admitted links should be satisfied with their required SINR targets
  - the number of admitted (removed) links should be maximized (minimized)
  - the total transmission power to support the admitted links should be minimized

Normalized Channel

- Two equivalent equations:
  - power constraint: $0 \le p_k \le \bar{p}_k \Leftrightarrow 0 \le x_k := \frac{p_k}{\bar{p}_k} \le 1$
  - SINR constraint: $\frac{g_{kk} p_k}{\sum_{j \neq k} g_{kj} p_j + \eta_k} \ge \gamma_k \Leftrightarrow \frac{x_k}{\sum_{j \neq k} \frac{\gamma_k g_{kj} \bar{p}_j}{g_{kk} \bar{p}_k} x_j + \frac{\gamma_k \eta_k}{g_{kk} \bar{p}_k}} \ge 1$
- Normalized channel:
  - noise vector $b = \left(\frac{\gamma_1 \eta_1}{g_{11} \bar{p}_1}, \frac{\gamma_2 \eta_2}{g_{22} \bar{p}_2}, \cdots, \frac{\gamma_K \eta_K}{g_{KK} \bar{p}_K} \right)^T > \mathbf{0}$
  - power allocation vector $x = \left(\frac{p_1}{\bar{p}_1}, \frac{p_2}{\bar{p}_2}, \cdots, \frac{p_K}{\bar{p}_K} \right)^T$
  - channel gain matrix A with its (k,j)-th entry

$$a_{kj} = \begin{cases} -\frac{\gamma_k g_{kj} \bar{p}_j}{g_{kk} \bar{p}_k}, & \text{if } k \neq j; \\ 1, & \text{if } k = j. \end{cases}$$

Composite L_q Minimization Formulation

- Simple to check

$$\frac{g_{kk}p_k}{\sum_{j\neq k} g_{kj}p_j + \eta_k} \geq \gamma_k \Longleftrightarrow (b - Ax)_k \leq 0$$

Composite L_q Minimization Formulation

- Simple to check

$$\frac{g_{kk}p_k}{\sum_{j\neq k} g_{kj}p_j + \eta_k} \geq \gamma_k \Longleftrightarrow (b - Ax)_k \leq 0$$

- The JPAC problem can be formulated as [L.-Dai-Luo, 2013]

$$\begin{aligned} \min_{\mathbf{x}} \quad & \|\max\{b - Ax, 0\}\|_q^q + \rho \bar{p}^T x \\ \text{s.t.} \quad & 0 \leq x \leq e. \end{aligned} \tag{2}$$

Support Vector Machine: Linearly Separable Data

- Given a database $\left\{s_m \in \mathbb{R}^{N-1},\ y_m \in \mathbb{R}\right\}_{m=1}^{M}$, where s_m is called example and y_m is the label associated with s_m.
- Find a linear discriminant function $\ell(s) = \hat{s}^T x$ with $\hat{s} = [s^T, 1]^T \in \mathbb{R}^N$
  - all data are correctly classified
  - the margin of the hyperplane ℓ that separates the two classes is maximized
- If the data are linearly separable, the above problem can be formulated as

$$
\begin{aligned}
\min_{x} \quad & \frac{1}{2}\sum_{n=1}^{N-1} x_n^2 \\
\text{s.t.} \quad & y_m \hat{s}_m^T x \geq 1,\ \ m = 1, 2, \ldots, M.
\end{aligned}
$$

Support Vector Machine: Not Linearly Separable Data

- Data are often NOT linearly separable in practice, and thus the above problem is not feasible.

- For the not linearly separable data, we can solve the following model instead:

$$\min_{x} \quad \sum_{m=1}^{M} \max\left\{1 - y_m \hat{s}_m^T x, 0\right\}^q + \frac{\rho}{2} \sum_{n=1}^{N-1} x_n^2.$$

- The above problem with $q = 1$ is called the soft-margin SVM in [Cortes-Vapnik, 1995].

Related Works

Existing Works I

$$\min_{x} \frac{\rho}{2}\|Ax - b\|^2 + \|x\|_q^q$$

- Lower bound theory [Chen-Xu-Ye, 2010]
- Strong NP-hardness [Chen-Ge-Wang-Ye, 2014]
- Iterative reweighted L_1 and L_2 minimization algorithms [Xu-Chang-Xu-Zhang, 2012; Lai-Xu-Yin, 2013;...]

Existing Works I

$$\min_{x} \frac{\rho}{2}\|Ax - b\|^2 + \|x\|_q^q$$

- Lower bound theory [Chen-Xu-Ye, 2010]
- Strong NP-hardness [Chen-Ge-Wang-Ye, 2014]
- Iterative reweighted L_1 and L_2 minimization algorithms [Xu-Chang-Xu-Zhang, 2012; Lai-Xu-Yin, 2013;...]

$$\min_{x} \|x\|_q^q \text{ s.t. } \quad Ax = b$$

- Sufficient conditions in recovering the sparsest solution [Chartrand, 2007; Chartrand-Staneva, 2008; Foucart-Lai, 2009]
- Strong NP-hardness and a potential reduction algorithm [Ge-Jiang-Ye, 2011]
- Iterative reweighted minimization methods [Chartrand-Yin, 2008; Daubechies et al., 2010; ...]
- Extend to the matrix case [Ji-Sze-Zhou-So-Ye, 2013]

Existing Works II

$$\min_{x} h(x) + \|x\|_q^q \tag{3}$$

- Smoothing quadratic regularization (SQR) algorithm and $O(\epsilon^{-2})$ worst-case iteration complexity analysis [Bian-Chen, 2013]
- First and second order interior-point methods, $O(\epsilon^{-2})$ and $O(\epsilon^{-3/2})$ iteration complexity results [Bian-Chen-Ye, 2014]
- Lower bound theory, iterative reweighted minimization methods, unified global convergence analysis [Lv, 2012]

Existing Works III

$$\min_x h(x) + \sum_{m=1}^{M} |a_m^T x|^q \tag{4}$$

- Second order necessary and sufficient conditions [Chen-Niu-Yuan, 2013]
- Smoothing trust region Newton (STRN) method [Chen-Niu-Yuan, 2013]
- An SQR algorithm and $O(\epsilon^{-2})$ iteration complexity analysis [Bian-Chen, 2014]

There Are More in This Workshop!

- "A Smoothing Majorization Method for ℓ_2-ℓ_p Matrix Minimization" [Zhang]
- "An Improved Algorithm for the L_2-L_p Minimization Problem [Ge]
- "p-Norm Constrained Quadratic Programming: Conic Approximation Methods" [Xing]
- ......

Some Remarks

- The definitions of ϵ-KKT points in the aforementioned works are different and thus are not comparable to each other.

Some Remarks

- The definitions of ϵ-KKT points in the aforementioned works are different and thus are not comparable to each other.
- All of the aforementioned problems are special cases of problem (1).

Some Remarks

- The definitions of ϵ-KKT points in the aforementioned works are different and thus are not comparable to each other.
- All of the aforementioned problems are special cases of problem (1).
- All of the aforementioned problems are sparse optimization problem with "equality constraints".

Some Remarks

- The definitions of ϵ-KKT points in the aforementioned works are different and thus are not comparable to each other.
- All of the aforementioned problems are special cases of problem (1).
- All of the aforementioned problems are sparse optimization problem with "equality constraints".
- Problem (1) is essentially a sparse optimization problem with "inequality constraints".

Some Remarks

- The definitions of ϵ-KKT points in the aforementioned works are different and thus are not comparable to each other.
- All of the aforementioned problems are special cases of problem (1).
- All of the aforementioned problems are sparse optimization problem with "equality constraints".
- Problem (1) is essentially a sparse optimization problem with "inequality constraints".
- Many of the aforementioned algorithms cannot be used to solve problem (1).

Some Remarks

- The definitions of ϵ-KKT points in the aforementioned works are different and thus are not comparable to each other.
- All of the aforementioned problems are special cases of problem (1).
- All of the aforementioned problems are sparse optimization problem with "equality constraints".
- Problem (1) is essentially a sparse optimization problem with "inequality constraints".
- Many of the aforementioned algorithms cannot be used to solve problem (1).
- Iterative reweighted minimization methods can be modified to solve problem (1).

Some Remarks

- The definitions of ϵ-KKT points in the aforementioned works are different and thus are not comparable to each other.
- All of the aforementioned problems are special cases of problem (1).
- All of the aforementioned problems are sparse optimization problem with "equality constraints".
- Problem (1) is essentially a sparse optimization problem with "inequality constraints".
- Many of the aforementioned algorithms cannot be used to solve problem (1).
- Iterative reweighted minimization methods can be modified to solve problem (1).
- However, the worst-case iteration complexity of all existing iterative reweighted minimization methods remains unclear.

Some Fundamental Questions

- Polyhedral constrained composite L_q minimization:

$$\begin{aligned} \min_{x \in \mathbb{R}^N} \quad & \|\max\{b - Ax, 0\}\|_q^q + h(x) \\ \text{s.t.} \quad & x \in \mathcal{X}. \end{aligned}$$

- Some fundamental questions that will be addressed in this talk:

Q1: Why use the non-convex L_q minimization formulation? Is it better than the corresponding convex L_1 counterpart? Can the solution of the L_q minimization solve the original L_0 minimization problem?

Q2: Is it easy to solve? Is there any polynomial time algorithm which can solve it to global optimality?

Q3: How to check a given point is a local minimizer or a stationary point of the problem? What is the KKT optimality conditions?

Q4: Since the problem is non-convex, nonsmooth, and non-Lipschitz, how to solve it efficiently with a worst-case iteration complexity guarantee?

Q1: Why use the non-convex L_q minimization formulation? Is it better than the corresponding convex L_1 counterpart? Can the solution of the L_q minimization solve the original L_0 minimization problem?

Exact Recovery

L_1 vs L_q: A Toy Example

- Let $A, b, \bar{p}$ in the JPAC problem (2) be

$$A = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ -1 & -1 & 1 \end{pmatrix}, \ b = 0.5e, \ \bar{p} = e.$$

- The optimal solution to problem (2) with $q = 0$ is

$$x^* = (0.5, 0.5, 0)^T.$$

- For any $\rho \geq 0$, $x = 0$ is the unique global minimizer of the L_1 minimization problem.
- For any given $q \in (0, 1)$, if ρ satisfies

$$0 < \rho < \bar{\rho}_q := \min\left\{1 + (0.5)^q, 2^q\right\} - (1.5)^q,$$

then the unique global minimizer of the L_q minimization problem (2) is x^*.

Why L_1 Does Not Work Well?

- The problem of minimizing $\|Ax - b\|_1$ is equivalent to the problem of minimizing $\|Ax - b\|_0$ with high probability under the assumptions that [Candes-Tao, 2005]

  1) the vector $Ax - b$ at the true solution x^* is sparse, where $A \in \mathbb{R}^{m \times n}$ and $m > n$; and

  2) the entries of the matrix A is independent and identically distributed (i.i.d.) Gaussian.

- However, these two assumptions often do not hold true.

- For instance, A in the JPAC problem has a special structure, i.e., all diagonal entries are one and all non-diagonal entries are non-positive.

Exact Recovery Result

Theorem (L.-Ma-Dai, 2013)

For any given instance of the JPAC problem (2)*, there exists* $\bar{q} > 0$ *such that when* $q \in (0, \bar{q}]$*, the global solution to the* L_q *minimization problem is one of the optimal solutions to problem* (2) *with* $q = 0$*.*

- This result depends on the special structure of A and b.
- Does this result hold true generally?
- More works along this direction need to be done.

Q2: Is the non-convex L_q minimization problem easy to solve? Is there any polynomial time algorithm which can solve it to global optimality?

Computational Complexity

Convexity vs Non-Convexity

Convexity vs Non-Convexity

- Two "easy" non-convex problems:
  - ratio of quadratic functions over an ellipsoid [Beck-Teboulle, 2009; Xia, 2013]

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & \frac{x^T A_1 x + b_1^T x + c_1}{x^T A_2 x + b_2^T x + c_2} \\ \text{s.t.} \quad & \|A_3 x\|_2 \le \rho. \end{aligned}$$

  - max-min fairness linear transceiver design for the SIMO interference channel [L.-Hong-Dai, 2013]

$$\begin{aligned} \max_{\{u_k,\, p_k\}} \quad & \min_k \left\{ \frac{|u_k^\dagger h_{kk}|^2 p_k}{\sigma_k^2 \|u_k\|^2 + \sum_{j\neq k} |u_k^\dagger h_{kj}|^2 p_j} \right\} \\ \text{s.t.} \quad & 0 \le p_k \le \bar{p}_k,\ k = 1, 2, ..., K. \end{aligned}$$

Convexity vs Non-Convexity

- Two "easy" non-convex problems:
  - ratio of quadratic functions over an ellipsoid [Beck-Teboulle, 2009; Xia, 2013]

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & \frac{x^T A_1 x + b_1^T x + c_1}{x^T A_2 x + b_2^T x + c_2} \\ \text{s.t.} \quad & \|A_3 x\|_2 \le \rho. \end{aligned}$$

  - max-min fairness linear transceiver design for the SIMO interference channel [L.-Hong-Dai, 2013]

$$\begin{aligned} \max_{\{u_k,\, p_k\}} \quad & \min_k \left\{ \frac{|u_k^\dagger h_{kk}|^2 p_k}{\sigma_k^2 \|u_k\|^2 + \sum_{j\neq k} |u_k^\dagger h_{kj}|^2 p_j} \right\} \\ \text{s.t.} \quad & 0 \le p_k \le \bar{p}_k,\ k = 1, 2, ..., K. \end{aligned}$$

- Complexity theory: a robust tool to characterize the computational tractability of an optimization problem

Complexity Results

Theorem (L.-Ma-Dai-Zhang, 2014)

For any given $0 < q < 1$, the unconstrained minimization

$$\min_x \| \max \{b - Ax, 0\} \|_q^q$$

is strongly NP-hard, and hence so is the polyhedral constrained L_q minimization problem (1).

Complexity Results

Theorem (L.-Ma-Dai-Zhang, 2014)

For any given $0 < q < 1$, *the unconstrained minimization*

$$\min_{x} \| \max\{b - Ax, 0\} \|_q^q$$

is strongly NP-hard, and hence so is the polyhedral constrained L_q *minimization problem* (1).

$\Longrightarrow$ Find high quality approximate solutions or locally optimal solutions in polynomial time

Q3: How to check a given point is a local minimizer or a stationary point of the composite L_q minimization problem? What is the KKT optimality conditions?

Optimality Conditions

An Auxiliary Smooth Problem

- Original nonsmooth non-Lipschitzian problem

$$\begin{array}{ll} \min_{x} & \|\max\{b - Ax, 0\}\|_q^q + h(x) \\ \text{s.t.} & x \in \mathcal{X}. \end{array}$$

- For any given $\bar{x}$, construct an auxiliary smooth problem

$$\begin{array}{ll} \min_{x} & \sum_{m \in \mathcal{J}_{\bar{x}}} (b - Ax)_m^q + h(x) \\ \text{s.t.} & (b - Ax)_m \leq 0, \ m \in \mathcal{K}_{\bar{x}}, \\ & x \in \mathcal{X}. \end{array} \tag{5}$$

with

$$\begin{aligned} \mathcal{I}_{\bar{x}} &= \{m \,|\, (b - A\bar{x})_m < 0\}, \\ \mathcal{J}_{\bar{x}} &= \{m \,|\, (b - A\bar{x})_m > 0\}, \\ \mathcal{K}_{\bar{x}} &= \{m \,|\, (b - A\bar{x})_m = 0\}. \end{aligned} \tag{6}$$

Key Connections

- Some observations
  - The objective value of problem (5) is equal to that of problem (1) at point $\bar{x}$.
  - The objective function of problem (5) is continuously differentiable in the neighborhood of point $\bar{x}$.

Key Connections

- Some observations
  - The objective value of problem (5) is equal to that of problem (1) at point $\bar{x}$.
  - The objective function of problem (5) is continuously differentiable in the neighborhood of point $\bar{x}$.
- Equivalence of problems (1) and (5) in the sense of sharing the same local minimizers

Lemma

$\bar{x}$ is a local minimizer of problem (1) *if any only if it is a local minimizer of problem* (5) *with $\mathcal{J}_{\bar{x}}$ and $\mathcal{K}_{\bar{x}}$ given in* (6).

Optimality Conditions

- First order optimality conditions

Theorem (L.-Ma-Dai-Zhang, 2014)

If $\bar{x} \in \mathcal{X}$ *is a local minimizer of problem* (1), *there must exist* $\bar{\lambda} \geq 0 \in \mathbb{R}^{|\mathcal{K}_{\bar{x}}|}$ *such that*

$$\bar{\lambda}_m (b - A\bar{x})_m = 0, \ \forall \ m \in \mathcal{K}_{\bar{x}} \tag{7}$$

and

$$\bar{x} - P_{\mathcal{X}}\left(\bar{x} - \nabla L(\bar{x}, \bar{\lambda})\right) = 0, \tag{8}$$

where

$$L(x, \lambda) = \sum_{m \in \mathcal{J}_{\bar{x}}} (b - Ax)_m^q + h(x) + \sum_{m \in \mathcal{K}_{\bar{x}}} \lambda_m (b - Ax)_m,$$

and $\mathcal{J}_{\bar{x}}$ *and* $\mathcal{K}_{\bar{x}}$ *are defined in* (6).

- Second order optimality conditions (skipped)

KKT Condition of Problem (3)

Definition (Chen-Xu-Ye, 2010; Ge-Jiang-Ye, 2011; Bian-Chen, 2013, 2014)

$\bar{x}$ is called a KKT point of problem

$$\min_x h(x) + \|x\|_q^q$$

if it satisfies

$$q|\bar{x}|^q + \bar{X}\nabla h(\bar{x}) = 0, \tag{9}$$

where $|\bar{x}|^q = \left(|\bar{x}_1|^q, \ldots, |\bar{x}_N|^q\right)^T$ *and* $\bar{X} = diag\left(\bar{x}_1, \ldots, \bar{x}_N\right)$.

KKT Condition of Problem (4)

Definition (Chen-Niu-Yuan, 2013)

$\bar{x}$ is called a KKT point of problem

$$\min_x h(x) + \sum_{m=1}^{M} |a_m^T x|^q$$

if it satisfies

$$Z_{\bar{x}}^T \nabla F_{\bar{x}}(\bar{x}) = 0, \tag{10}$$

where

$$F_{\bar{x}}(x) = \sum_{a_m^T \bar{x} \neq 0} \left| a_m^T x \right|^q + h(x)$$

and $Z_{\bar{x}}$ is the matrix whose columns form an orthogonal basis for the null space of $\left\{ a_m \mid a_m^T \bar{x} = 0 \right\}$.

Equivalence of Different Definitions

Proposition

When problem (1) *reduces to problem* (4), *there holds*

$$(7) \textit{ and } (8) \Longleftrightarrow (10);$$

When problem (1) *reduces to problem* (3), *there holds*

$$(7) \textit{ and } (8) \Longleftrightarrow (9).$$

Q4: Since problem (1) is non-convex, nonsmooth, and non-Lipschitz, how to solve it efficiently with a worst-case iteration complexity guarantee?

An SSQP Framework & Analysis

Challenges

- Two challenges of smoothing algorithms
  - How to choose a smoothing function and an algorithm for the smoothing problem?
  - How to update the smoothing parameter?
- Both the choice of smoothing functions and the updating rule of the smoothing parameter play a key role in convergence and iteration complexity analysis of the smoothing algorithms.

Smoothing Approximation[1]

- Use

$$\theta(t,\mu) = \begin{cases} t, & \text{if } t > \mu; \\ \dfrac{t^2}{2\mu} + \dfrac{\mu}{2}, & \text{if } 0 \le t \le \mu; \\ \dfrac{\mu}{2}, & \text{if } t < 0 \end{cases}$$

  to approximate

$$\theta(t) = \max\{t, 0\}.$$

- Approximation properties
  - $\theta(t,\mu) = \theta(t), \ \forall \ t \ge \mu$
  - $\theta(t,\mu) \ge \dfrac{\mu}{2}, \ \forall \ t$
  - $\theta^q(t,\mu)$ is continuously differentiable

[1] Thanks Prof. Xiaojun Chen for the discussion on the choice of the smoothing function.

Smoothing Problem

- Define $\tilde{F}(x,\mu) = \tilde{f}(x,\mu) + h(x)$, where $\tilde{f}(x,\mu) = \sum_{m\in\mathcal{M}} \theta^q((b-Ax)_m,\mu)$, then

$$F(x) \le \tilde{F}(x,\mu) \le F(x) + \sum_{(b-Ax)_m \le \mu} \left(\frac{\mu}{2}\right)^q, \ \forall\ x.$$

- Smoothing problem:

$$\begin{aligned} \min_x \quad & \tilde{F}(x,\mu) := \sum_{m\in\mathcal{M}} \theta^q((b-Ax)_m,\mu) + h(x) \\ \text{s.t.} \quad & x \in \mathcal{X} \end{aligned} \tag{11}$$

Smoothing Problem

- Define $\tilde{F}(x,\mu) = \tilde{f}(x,\mu) + h(x)$, where $\tilde{f}(x,\mu) = \sum_{m\in\mathcal{M}} \theta^q((b-Ax)_m,\mu)$, then

$$F(x) \le \tilde{F}(x,\mu) \le F(x) + \sum_{(b-Ax)_m \le \mu} \left(\frac{\mu}{2}\right)^q, \ \forall\ x.$$

- Smoothing problem:

$$\begin{aligned} \min_x \quad & \tilde{F}(x,\mu) := \sum_{m\in\mathcal{M}} \theta^q((b-Ax)_m,\mu) + h(x) \\ \text{s.t.} \quad & x \in \mathcal{X} \end{aligned} \tag{11}$$

Theorem

For any $q \in (0,1)$ and $\mu > 0$, the smoothing approximation problem (11) is strongly NP-hard (even for the special case when $h(x) = 0$ and $\mathcal{X} = \mathbb{R}^N$).

Local Convex Quadratic Upper Bound

- A local convex quadratic upper bound at point x_k

$$Q(x, x_k, \mu) = Q_1(x, x_k, \mu) + Q_2(x, x_k) \tag{12}$$

- $Q_1(x, x_k, \mu) = \tilde{f}(x_k, \mu) + \nabla \tilde{f}(x_k, \mu)^T (x - x_k) + \frac{1}{2}(x - x_k)^T \tilde{B}(x_k, \mu)(x - x_k)$
- $\tilde{B}(x, \mu) = \sum_{m \in \mathcal{M}} \kappa((b - Ax)_m, \mu) a_m a_m^T$
- $Q_2(x, x_k) = h(x_k) + \nabla h(x_k)^T (x - x_k) + \frac{1}{2} L_h \|x - x_k\|^2$

Local QP Upper Bound of $\tilde{F}(x,\mu)$

Lemma (A Local QP Upper Bound of Smoothing Function $\tilde{F}(x,\mu)$)

For any x_k and x such that

$$(A(x_k - x))_m \leq \mu, \;\; m \in \mathcal{I}^{\mu}_{x_k},$$
$$(A(x_k - x))_m \geq \frac{-(b - Ax_k)_m}{2}, \;\; m \in \mathcal{J}^{\mu}_{x_k},$$

where

$$\begin{aligned} \mathcal{I}^{\mu}_{x_k} &= \{m \,|\, (b - Ax_k)_m < -\mu\}, \\ \mathcal{J}^{\mu}_{x_k} &= \{m \,|\, (b - Ax_k)_m > 2\mu\}, \end{aligned}$$

then

$$\tilde{F}(x,\mu) \leq Q(x, x_k, \mu),$$

where $Q(x, x_k, \mu)$ is defined in (12).

An SSQP Framework

- Update rule of the smoothing parameter: if x_k satisfies

$$\left\| P_{\mathcal{X}} \left(x_k - \nabla \tilde{F}(x_k, \mu) \right) - x_k \right\| \le \mu, \tag{13}$$

set

$$\mu = \sigma \mu, \; x_0 = x_k, \; k = 0;$$

else compute the next point x_{k+1}.

- Algorithmic framework for solving the smoothing problem: let x_{k+1} be an (approximate) solution of the following convex QP

$$\begin{array}{ll} \min\limits_{x \in \mathcal{X}} & Q(x, x_k, \mu) \\ \text{s.t.} & (A(x_k - x))_m \le \mu, \; m \in \mathcal{I}^{\mu}_{x_k}, \\ & (A(x_k - x))_m \ge -\dfrac{(b - Ax_k)_m}{2}, \; m \in \mathcal{J}^{\mu}_{x_k} \end{array} \tag{14}$$

such that

$$\tilde{F}(x_k, \mu) - \tilde{F}(x_{k+1}, \mu) \ge O(\mu^{4-q}).$$

- Termination criterion: the above procedure is repeated until $\mu \le \epsilon$ and (13) is satisfied.

Some Remarks

- Flexible to choose subroutines for solving problem (14)
- Can deal with the case where L_h is unknown

Existence of x_{k+1} : A Shrink Projection Gradient Step

Lemma

For any $\mu \in (0,1]$ *and* $k \geq 0$ *in the proposed SSQP framework,*

$$x_{k+1}^{proj} = x_k + \xi_k \tau_k d_k, \tag{15}$$

where

$$\xi_k = \min\left\{ \frac{-d_k^T \nabla \tilde{F}(x_k, \mu)}{\tau_k d_k^T \left(\tilde{B}_k + L_h I_N \right) d_k}, 1 \right\}, \ \tau_k = \frac{\mu}{\left(\max_m \{\|a_m\|\} + 1\right) \|d_k\|} < 1,$$

and

$$d_k = P_{\mathcal{X}}(x_k - \nabla \tilde{F}(x_k, \mu)) - x_k.$$

If (13) *is not satisfied, then*

$$\tilde{F}(x_k, \mu) - \tilde{F}(x_{k+1}^{proj}, \mu) \geq \mu^{4-q} / J_0, \tag{16}$$

where $J_0 = \max\left\{8q \sum_m \|a_m\|^2 + 2L_h, 2\max_m\{\|a_m\|\} + 2\right\}$.

Existence of x_{k+1} : Many Other Choices

Lemma

For any $\mu \in (0,1]$ *and* $k \geq 0$ *in the proposed SSQP framework, suppose that*

- x_{k+1}^{exact} *is the solution of problem* (14),
- x_{k+1}^{snorm} *is the solution of the following problem*

$$\begin{array}{ll} \min\limits_{x \in \mathcal{X}} & Q(x, x_k, \mu) \\ s.t. & \|A(x - x_k)\|_\infty \leq \mu. \end{array}$$

If (13) *is not satisfied, then*

$$\tilde{F}(x_k, \mu) - \tilde{F}(x_{k+1}^{exact}, \mu) \geq \tilde{F}(x_k, \mu) - \tilde{F}(x_{k+1}^{snorm}, \mu) \geq \tilde{F}(x_k, \mu) - \tilde{F}(x_{k+1}^{proj}, \mu). \quad (17)$$

Iteration Complexity

Theorem

Let $x_{k+1} = x_{k+1}^{proj}$ in the proposed SSQP framework. Then, for any $\epsilon \in (0, 1]$, the framework will terminate within at most

$$\left\lceil J_T^q \epsilon^{q-4} \right\rceil \tag{18}$$

iterations, where

$$J_T^q = \frac{\sigma^{q-4}\left(\tilde{F}(x_0, 1)J_0 + 1\right)}{\sigma^{q-4} - 1}. \tag{19}$$

Iteration Complexity

Theorem

Let $x_{k+1} = x_{k+1}^{proj}$ in the proposed SSQP framework. Then, for any $\epsilon \in (0, 1]$, the framework will terminate within at most

$$\left\lceil J_T^q \epsilon^{q-4} \right\rceil \tag{18}$$

iterations, where

$$J_T^q = \frac{\sigma^{q-4}\left(\tilde{F}(x_0, 1)J_0 + 1\right)}{\sigma^{q-4} - 1}. \tag{19}$$

- The worst-case iteration complexity function in (18) is a strictly decreasing function with respect to $q \in (0, 1)$ for fixed $\epsilon \in (0, 1)$.
- This is consistent with the intuition that problem (1) becomes more difficult to solve as q decreases.

ϵ-KKT Point: A Perturbation of the KKT Point

Definition (L.-Ma-Dai-Zhang, 2014)

For any given $\epsilon > 0$, $\bar{x} \in \mathcal{X}$ *is called an* ϵ*-KKT point of problem* (1) *if there exists* $\bar{\lambda} \geq 0 \in \mathbb{R}^{|\mathcal{K}_{\bar{x}}^{\epsilon}|}$ *such that*

$$\left|\bar{\lambda}_m (b - A\bar{x})_m\right| \leq \epsilon^q, \;\; m \in \mathcal{K}_{\bar{x}}^{\epsilon} \tag{20}$$

and

$$\left\|\bar{x} - P_{\mathcal{X}}\left(\bar{x} - \nabla L^{\epsilon}(\bar{x}, \bar{\lambda})\right)\right\| \leq \epsilon, \tag{21}$$

where

$$L^{\epsilon}(x, \lambda) = \sum_{m \in \mathcal{J}_{\bar{x}}^{\epsilon}} (b - Ax)_m^q + h(x) + \sum_{m \in \mathcal{K}_{\bar{x}}^{\epsilon}} \lambda_m (b - Ax)_m$$

with

$$\begin{aligned} \mathcal{I}_{\bar{x}}^{\epsilon} &= \left\{m \,|\, (b - A\bar{x})_m < -\epsilon\right\}, \\ \mathcal{J}_{\bar{x}}^{\epsilon} &= \left\{m \,|\, (b - A\bar{x})_m > \epsilon\right\}, \\ \mathcal{K}_{\bar{x}}^{\epsilon} &= \left\{m \,|\, -\epsilon \leq (b - A\bar{x})_m \leq \epsilon\right\}. \end{aligned} \tag{22}$$

When Problem (1) Reduces to Problem (4)

Definition

For any given $\epsilon > 0$, $\bar{x}$ *is called an* ϵ*-KKT point of problem* (4) *if there exists* $\bar{\lambda} \in \mathbb{R}^{|\hat{\mathcal{K}}_{\bar{x}}^{\epsilon}|}$ *such that*

$$\left|\bar{\lambda}_m a_m^T \bar{x}\right| \leq \epsilon^q, \ \ m \in \hat{\mathcal{K}}_{\bar{x}}^{\epsilon} \tag{23}$$

and

$$\left\|\nabla \hat{L}^{\epsilon}(\bar{x}, \bar{\lambda})\right\| \leq \epsilon, \tag{24}$$

where

$$\hat{L}^{\epsilon}(x,\lambda) = \sum_{m \in \hat{\mathcal{I}}_{\bar{x}}^{\epsilon}} (-a_m^T x)^q + \sum_{m \in \hat{\mathcal{J}}_{\bar{x}}^{\epsilon}} (a_m^T x)^q + h(x) + \sum_{m \in \hat{\mathcal{K}}_{\bar{x}}^{\epsilon}} \lambda_m (b - Ax)_m$$

with

$$\begin{aligned} \hat{\mathcal{I}}_{\bar{x}}^{\epsilon} &= \left\{m \,|\, a_m^T \bar{x} < -\epsilon\right\}, \\ \hat{\mathcal{J}}_{\bar{x}}^{\epsilon} &= \left\{m \,|\, a_m^T \bar{x} > \epsilon\right\}, \\ \hat{\mathcal{K}}_{\bar{x}}^{\epsilon} &= \left\{m \,|\, -\epsilon \leq a_m^T \bar{x} \leq \epsilon\right\}. \end{aligned}$$

Definition of ϵ-KKT Point for Problem (4)

Definition (Bian-Chen, 2014)

For any $\epsilon \in (0,1]$, $\bar{x}$ *is called an* ϵ*-KKT point of problem* (4) *if it satisfies*

$$\left\| \left(Z_{\bar{x}}^{\epsilon}\right)^T \nabla F_{\bar{x}}^{\epsilon}(\bar{x}) \right\|_{\infty} \leq \epsilon, \tag{25}$$

where

$$F_{\bar{x}}^{\epsilon}(x) = \sum_{|a_m^T \bar{x}| > \epsilon} \left| a_m^T x \right|^q + h(x)$$

and $Z_{\bar{x}}^{\epsilon}$ *is the matrix whose columns form an orthogonal basis for the null space of* $\left\{ a_m \mid \left| a_m^T \bar{x} \right| \leq \epsilon \right\}$.

- (23) and (24) $\Longrightarrow$ (25)
- Shall talk more about the comparison later

The SSQP Framework Returns An ϵ-KKT Point

- Define

$$\begin{aligned}\mathcal{I}_{\bar{x}}^{\epsilon} &= \{m \,|\, (b - A\bar{x})_m < -\epsilon\} \\ \mathcal{J}_{\bar{x}}^{\epsilon} &= \{m \,|\, (b - A\bar{x})_m > \epsilon\} \\ \mathcal{K}_{\bar{x}}^{\epsilon} &= \{m \,|\, -\epsilon \leq (b - A\bar{x})_m \leq \epsilon\}\end{aligned}$$

as in (22), and

$$\bar{\lambda}_m = [\theta^q(t,\epsilon)]'_{t=(b-A\bar{x})_m}, \;\; m \in \mathcal{K}_{\bar{x}}^{\epsilon} \tag{26}$$

Theorem

For any $\epsilon \in (0,1]$, let $\bar{x}$ be the point returned by the proposed SSQP framework and $\bar{\lambda}$ be defined in (26). Then $\bar{x}$ and $\bar{\lambda}$ satisfy (20) and (21).

Worst-Case Iteration Complexity

Theorem (L.-Ma-Dai-Zhang, 2014)

For any $\epsilon \in (0,1]$*, the total number of iterations for the SSQP framework to return an* ϵ*-KKT point of problem* (1) *satisfying* (20) *and* (21) *is at most*

$$O\left(\epsilon^{q-4}\right).$$

In particular, letting x_{k+1} *be* x_{k+1}^{proj}, x_{k+1}^{snorm}, *or* x_{k+1}^{exact} *in the proposed SSQP framework, the total number of iterations for the framework to return an* ϵ*-KKT point of problem* (1) *satisfying* (20) *and* (21) *is at most*

$$\left\lceil J_T^q \epsilon^{q-4} \right\rceil,$$

where J_T^q *is given in* (19).

SSQP vs Existing Works

- The SSQP algorithmic framework is designed for solving a more general and difficult problem.

SSQP vs Existing Works

- The SSQP algorithmic framework is designed for solving a more general and difficult problem.
- SSQP with $x_{k+1} = x_{k+1}^{\text{proj}}$ vs SQR when applied to solve problem (4)

		SQR [Bian-Chen, 2014]	SSQP
complexity	iteration number	$O(\epsilon^{-2})$	$O(\epsilon^{q-4})$
	subproblem per iteration	n-dimensional QP	univariate QP
quality	optimality residual I	$\left\|(Z_{\bar{x}}^{\epsilon})^T \nabla F_{\bar{x}}^{\epsilon}(\bar{x})\right\|_{\infty} \leq \epsilon$	$\left\|\nabla \hat{L}^{\epsilon}(\bar{x}, \bar{\lambda})\right\| \leq \epsilon$
	optimality residual II	$\left\|\nabla \tilde{F}(\bar{x}, \epsilon)\right\| = O\left(\epsilon^{2-2/q}\right)$	$\left\|\nabla \tilde{F}(\bar{x}, \epsilon)\right\| \leq \epsilon$
	complementary violation	not guaranteed	$\left\|\bar{\lambda}_m a_m^T \bar{x}\right\| \leq \epsilon^q, \ m \in \hat{\mathcal{K}}_{\bar{x}}^{\epsilon}$

Concluding Remarks

- Polyhedral constrained composite L_q minimization

Concluding Remarks

- Polyhedral constrained composite L_q minimization
- Applications from wireless communications and machine learning

Concluding Remarks

- Polyhedral constrained composite L_q minimization
- Applications from wireless communications and machine learning
- Exact recovery result for JPAC

Concluding Remarks

- Polyhedral constrained composite L_q minimization
- Applications from wireless communications and machine learning
- Exact recovery result for JPAC
- Computational intractbility

Concluding Remarks

- Polyhedral constrained composite L_q minimization
- Applications from wireless communications and machine learning
- Exact recovery result for JPAC
- Computational intractbility
- Optimality conditions

Concluding Remarks

- Polyhedral constrained composite L_q minimization
- Applications from wireless communications and machine learning
- Exact recovery result for JPAC
- Computational intractbility
- Optimality conditions
- SSQP framework and iteration complexity analysis

Concluding Remarks

- Polyhedral constrained composite L_q minimization
- Applications from wireless communications and machine learning
- Exact recovery result for JPAC
- Computational intractbility
- Optimality conditions
- SSQP framework and iteration complexity analysis
- Extend to matrix case

Thank You!

Email: yafliu@lsec.cc.ac.cn