A smoothing SQP framework for a class of composite $$L_q$$ L q minimization over polyhedron

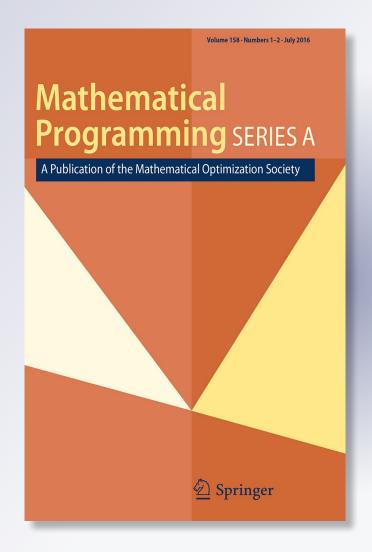## Ya-Feng Liu, Shiqian Ma, Yu-Hong Dai & Shuzhong Zhang

Springer

Springer

CrossMark

A smoothing SQP framework for a class of composite L_q minimization over polyhedron

Ya-Feng Liu[1] · Shiqian Ma[2] · Yu-Hong Dai[1] · Shuzhong Zhang[3]

Abstract The composite L_q $(0 < q < 1)$ minimization problem over a general polyhedron has received various applications in machine learning, wireless communications, image restoration, signal reconstruction, etc. This paper aims to provide a theoretical study on this problem. First, we derive the Karush–Kuhn–Tucker (KKT) optimality conditions for local minimizers of the problem. Second, we propose a smoothing sequential quadratic programming framework for solving this problem.

✉ Ya-Feng Liu
    yafliu@lsec.cc.ac.cn

    Shiqian Ma
    sqma@se.cuhk.edu.hk

    Yu-Hong Dai
    dyh@lsec.cc.ac.cn

    Shuzhong Zhang
    zhangs@umn.edu

1   LSEC, ICMSEC, Academy of Mathematics and Systems Science, Chinese Academy of
    Sciences, Beijing 100190, China

2   Department of Systems Engineering and Engineering Management, The Chinese University of
    Hong Kong, Shatin, New Territories, Hong Kong, China

3   Department of Industrial and Systems Engineering, University of Minnesota, Minneapolis, MN
    55455, USA

⌂ Springer

The framework requires a (approximate) solution of a convex quadratic program at each iteration. Finally, we analyze the worst-case iteration complexity of the framework for returning an ϵ-KKT point; i.e., a feasible point that satisfies a perturbed version of the derived KKT optimality conditions. To the best of our knowledge, the proposed framework is the first one with a worst-case iteration complexity guarantee for solving composite L_q minimization over a general polyhedron.

Keywords Composite L_q minimization · ϵ-KKT point · Nonsmooth nonconvex non-Lipschitzian optimization · Optimality condition · Smoothing approximation · Worst-case iteration complexity

Mathematics Subject Classification 90C26 · 90C30 · 90C46 · 65K05

1 Introduction

In this paper, we consider the following polyhedral constrained composite nonsmooth nonconvex non-Lipschitzian L_q minimization problem

$$
\begin{aligned}
\min_{x} \quad & F(x) := f(x) + h(x) \\
\text{s.t.} \quad & x \in \mathcal{X},
\end{aligned}
\tag{1.1}
$$

where

– $f(x)$ is of the form

$$
f(x) = \|\max\{b - Ax, 0\}\|_q^q = \sum_{m=1}^{M} \max\left\{b_m - a_m^T x, 0\right\}^q
\tag{1.2}
$$

with $A = [a_1, a_2, \ldots, a_M]^T \in \mathbb{R}^{M \times N}$, $b = [b_1, b_2, \ldots, b_M]^T \in \mathbb{R}^M$, and $0 < q < 1$;
– $h(x)$ is a continuously differentiable function with L_h-Lipschitz continuous gradient in $\mathcal{X}$, that is,

$$
\|\nabla h(x) - \nabla h(y)\|_2 \leq L_h \|x - y\|_2, \quad \forall\, x, y \in \mathcal{X};
\tag{1.3}
$$

– and, $\mathcal{X} \subseteq \mathbb{R}^N$ is a polyhedral set.

When $h(x) = 0$ and $\mathcal{X} = \mathbb{R}^N$, problem (1.1) reduces to unconstrained L_q minimization problem

$$
\min_{x} \|\max\{b - Ax, 0\}\|_q^q.
\tag{1.4}
$$

Both problems (1.1) and (1.4) find wide applications in information theory [13], computational biology [60], wireless communications [46,47,49], machine learning [11,28], image restoration [7,23,26,53], signal processing [12,50], and variable selection [31,40]. Three specific applications arising from machine learning, wireless communications, and information theory are given in "Appendix 1".

1.1 Related works

Recently, many algorithms have been proposed to solve problem

$$\min_x h(x) + \|x\|_q^q . \tag{1.5}$$

In particular, when $h(x)$ is a convex quadratic function and $q = 1$, problem (1.5) is shown to be quite effective in finding a sparse vector to minimize $h(x)$ and various efficient algorithms [2,4,39,52,61,63] have been proposed to solve it.

When $q \in (0, 1)$, problem (1.5) is nonsmooth, nonconvex, and even not Lipschitz. Assuming that $h : \mathbb{R}^N \to [0, +\infty)$ is continuously differentiable and its gradient satisfies (1.3), Bian and Chen [6] proposed a smoothing quadratic regularization (SQR) algorithm for problem (1.5) and established the worst-case iteration complexity result $\mathcal{O}(\epsilon^{-2})$ for the SQR algorithm to return an ϵ-KKT point (or ϵ-KKT solution, or ϵ-stationary point, or ϵ-scaled stationary point, or ϵ-scaled first order stationary point) of problem (1.5). In [8], Bian, Chen, and Ye proposed a first order interior-point method (using only the gradient information) and a second order interior-point method (using both the gradient and Hessian information) for problem (1.5) with box constraints. They showed that the iteration complexity of their first order method for returning an ϵ-scaled stationary point is $\mathcal{O}(\epsilon^{-2})$ and the one of their second order method for returning an ϵ-scaled second order stationary point is $\mathcal{O}(\epsilon^{-3/2})$. In [24], Chen, Niu, and Yuan derived affine-scaled second order necessary and sufficient conditions for local minimizers of the problem

$$\min_x h(x) + \sum_{m=1}^{M} \left| a_m^T x \right|^q , \tag{1.6}$$

which includes (1.5) as a special case. Furthermore, they proposed a smoothing trust region Newton (STRN) method and proved that the sequence generated by the STRN algorithm is globally convergent to a point satisfying the affine-scaled second order necessary optimality condition. In [7], Bian and Chen proposed an SQR algorithm for problem (1.6) (possibly with box constraints) and showed that the worst-case iteration complexity of the SQR algorithm for finding an ϵ-stationary point is $\mathcal{O}(\epsilon^{-2})$. Cartis et al. [17] considered problem

$$\min_x h(x) + \varphi(c(x)),$$

where $h : \mathbb{R}^N \to \mathbb{R}$ and $c : \mathbb{R}^N \to \mathbb{R}^M$ are continuously differentiable and $\varphi : \mathbb{R}^M \to \mathbb{R}$ is convex and is globally Lipschitz continuous but possibly nonsmooth. They proved that it takes at most $\mathcal{O}(\epsilon^{-2})$ iterations to obtain an ϵ-KKT point by a first order trust region method or a quadratic regularization method. Ghadimi and Lan [37] generalized Nesterov's accelerated gradient (AG) method [51], originally designed for smooth convex optimization, to solve problem

$$\min_x h(x) + \psi(x),$$

where $h : \mathbb{R}^N \to \mathbb{R}$ is continuous differentiable and $\psi(x) : \mathbb{R}^N \to \mathbb{R}$ is a (simple) convex nonsmooth function with special structures. They showed that it takes at most $\mathcal{O}(\epsilon^{-2})$ iterations to reduce a first order criticality measure below ϵ for the generalized AG method. Jiang and Zhang [43] considered the following block-structured problem

$$\min_x h(x_1, x_2, \ldots, x_M) + \sum_{m=1}^{M} \psi_m(x_m)$$

$$\text{s.t. } x_m \in \mathcal{X}_m, \quad m = 1, 2, \ldots, M,$$

where $h : \mathbb{R}^N \to \mathbb{R}$ is smooth, $\psi_m(x_m) : \mathbb{R}^{N_m} \to \mathbb{R}$ are convex but nonsmooth. They showed that the conditional gradient and gradient projection type of methods can find an ϵ-KKT point of the above problem within $\mathcal{O}(\epsilon^{-2})$ iterations. Here we should notice that the definitions of ϵ-KKT points in the aforementioned works are different and thus are not comparable to each other.

In particular, when $h(x) = \frac{\rho}{2} \|Ax - b\|^2$, problem (1.5) becomes

$$\min_x \frac{\rho}{2} \|Ax - b\|^2 + \|x\|_q^q. \tag{1.7}$$

Chen et al. [22] showed that problem (1.7) is strongly NP-hard. Recently, iterative reweighted L_1 and L_2 minimization algorithms were proposed to (approximately) solve problem (1.7) (see [26,30,32,44,45] and references therein). In [25], Chen et al. derived a lower bound theory for local minimizers of problem (1.7); i.e., each component of any local minimizer of problem (1.7) is either zero or not less than a positive constant which only depends on the problem inputs A, b, ρ, and q. Lu [48] extended the lower bound theory to problem (1.5) with general $h(x)$ satisfying (1.3). Based on the derived lower bound theory, Lu proposed a novel iterative reweighted minimization method for solving problem (1.5) and provided a unified global convergence analysis for the aforementioned iterative reweighted minimization algorithms.

Another problem closely related to problem (1.7) is

$$\min_x \|x\|_q^q, \text{ s.t. } Ax = b. \tag{1.8}$$

In [36], Ge et al. showed that problem (1.8) and its smoothed version are strongly NP-hard. Chartrand [18], Chartrand and Staneva [19], Foucart and Lai [32], and Sun [57] established some sufficient conditions under which problem (1.8) is able to recover the sparsest solution to the undetermined linear system $Ax = b$. Efficient iterative reweighted minimization algorithms were proposed to solve problem (1.8) by Chartrand and Yin [20], Foucart and Lai [32], Daubechies et al. [30], Rao and Kreutz-Delgado [56], and Candès, Wakin, and Boyd [14]. It was shown in [30, Theorem 7.7 (i)] that under suitable conditions, the sequence generated by the iterative reweighted L_2 minimization algorithms converges to the global minimizer of problem (1.8). Moreover, the following related problem

$$\min_x \|x\|_q^q, \text{ s.t. } Ax = b, \quad x \geq 0, \tag{1.9}$$

was also considered in [36], and the authors developed an interior-point potential reduction algorithm for solving problem (1.8), which is guaranteed to return a scaled ϵ-KKT point in no more than $\mathcal{O}(\epsilon^{-1}\log\epsilon^{-1})$ iterations. The similar idea was extended by Ji et al. [41] to solve the matrix counterpart of problem (1.9) where the unknown variable is a positive semidefinite matrix.

Although many algorithms have been mentioned in the above, they cannot be used to solve problem (1.1). For instance, the potential reduction algorithm in [36] cannot be applied to solve problem (1.1) where $h(x)$ is not concave; the SQR algorithms [6,7] and the interior-point algorithms [22] cannot deal with the composite L_q term $f(x)$ and the general polyhedral constraint in problem (1.1); the algorithm proposed in [17] cannot be used to solve problem (1.1) either, since the composite L_q term $f(x)$ in the objective function of (1.1) cannot be expressed as a form of $\varphi(c(x))$. The aforementioned iterative reweighted minimization methods could be modified to solve problem (1.1). However, the worst-case iteration complexity of all existing iterative reweighted minimization methods remains unclear so far and global convergence of some of them are still unknown [14]. The goal of this paper is to develop an algorithmic framework for problem (1.1) with worst-case iteration complexity guarantee.

1.2 Our contribution

In this paper, we consider polyhedral constrained composite nonsmooth nonconvex non-Lipschitzian L_q minimization problem (1.1), which includes problems (1.5), (1.6), (1.7), (1.8), and (1.9) as special cases. A sharp difference between problem (1.1) and the aforementioned problems lies in the composite term in $\|\cdot\|_q^q$, i.e., problem (1.1) tries to find a solution such that the number of positive components of the vector $b - Ax$ is as small as possible. However, problem (1.5), for instance, tries to find a solution such that the number of nonzero entries of x is as small as possible. In other words, problem (1.1) and its special case problem (1.4) considered in this paper are essentially sparse optimization problems with inequality constraints while all previously mentioned problems are sparse optimization with equality constraints.

We propose a smoothing sequential quadratic programming (SSQP) framework for solving problem (1.1) (and problem (1.4)), where a convex quadratic program (QP) is (approximately) solved at each iteration, and analyze the worst-case iteration complexity of the proposed algorithmic framework. One iteration in this paper refers to (approximately) solving one convex QP subproblem. To the best of our knowledge, this is the first algorithm/framework for solving polyhedral constrained composite non-Lipschitzian L_q minimization with worst-case iteration complexity analysis. The main contributions of this paper are summarized as follows.

– Problem (1.4) is shown to be strongly NP-hard (see Theorem 2.1);
– KKT optimality conditions for local minimizers of problem (1.1) are derived (see Theorem 3.3);
– An SSQP algorithmic framework is proposed for solving problem (1.1) and its worst-case iteration complexity is analyzed. In particular, we show in Theorem 5.12 that the SSQP framework can return an ϵ-KKT point of problem (1.1) in Definition 5.7 within $\mathcal{O}\left(\epsilon^{q-4}\right)$ iterations. Here we should notice that the ϵ-KKT

point defined in Definition 5.7 is stronger than the ones used in [6,7,24,25,36] when problem (1.1) reduces to problems (1.5) and (1.6).

The rest of this paper is organized as follows. In Sect. 2, we show that problem (1.4) is strongly NP-hard. In Sect. 3, we show that problem (1.1) and auxiliary problem (3.2) are equivalent in the sense that the two problems share the same local minimizers. This equivalence result further implies the KKT optimality conditions of problem (1.1). In Sect. 4, we give a smoothing approximation for problem (1.1). In Sect. 5, we propose an SSQP algorithmic framework for solving problem (1.1) and analyze the worst-case iteration complexity of the proposed algorithmic framework. Finally, we make some concluding remarks in Sect. 6.

Notations. We denote $\mathcal{M} = \{1, 2, \ldots, M\}$. For any set $\mathcal{K}$, $|\mathcal{K}|$ stands for its cardinality. $\nabla h(x)$ is the gradient of a continuously differentiable function $h(x)$. $P_{\mathcal{X}}(x)$ $\lambda_{\max}(B)$ is the projection of a point x onto the convex set $\mathcal{X}$. I_N is the $N \times N$ identity matrix. We use $\lambda_{\max}(B)$ to denote the largest eigenvalue of a real symmetric matrix B. Throughout this paper, $\|\cdot\|$ denotes the Euclidean norm unless otherwise specified.

2 Intractability analysis

In this section, we show that the unconstrained L_q minimization problem (1.4) with any $q \in (0, 1)$ is strongly NP-hard.

The proof of the NP-hardness of problem (1.4) is based on a polynomial time transformation from the strongly NP-complete 3-partition problem [34, Theorem 4.4]. The 3-partition problem can be described as follows: given a set of positive integers $\{a_i\}_{i \in \mathcal{S}}$ with $\mathcal{S} = \{1, 2, \ldots, 3m\}$ and a positive integer B such that $a_i \in (B/4, B/2)$ for all $i \in \mathcal{S}$ and

$$\sum_{i \in \mathcal{S}} a_i = mB, \tag{2.1}$$

the problem is to check whether there exists a partition $\mathcal{S}_1, \ldots, \mathcal{S}_m$ of $\mathcal{S}$ such that

$$\sum_{i \in \mathcal{S}_j} a_i = B, \quad \forall \, j = 1, 2, \ldots, m. \tag{2.2}$$

Notice that the constraints on $\{a_i\}$ imply that each $\mathcal{S}_j$ in (2.2) must contain exactly three elements.

Theorem 2.1 *For any $q \in (0, 1)$, the unconstrained L_q minimization problem (1.4) is strongly NP-hard.*

Proof We prove the theorem by constructing a polynomial time transformation from the 3-partition problem to the unconstrained L_q minimization problem (1.4) [34,55, 59]. For any given instance of the 3-partition problem with $\{a_i\}_{i \in \mathcal{S}}$, m, and B, we

construct an instance of problem (1.4) with $M = 3m^2 + 4m$ and $N = 3m^2$ as follows:

$$\min_x f(x) := f_1(x) + f_2(x) + f_3(x), \tag{2.3}$$

where

$$f_1(x) = \sum_{i=1}^{3m} \sum_{j=1}^{m} \left(\max\{x_{ij}, 0\}^q + \max\{1 - x_{ij}, 0\}^q \right),$$

$$f_2(x) = \sum_{i=1}^{3m} \max\left\{ \sum_{j=1}^{m} x_{ij} - 1, 0 \right\}^q,$$

$$f_3(x) = \sum_{j=1}^{m} \max\left\{ B - \sum_{i=1}^{3m} a_i x_{ij}, 0 \right\}^q.$$

It is easy to verify that $f_1(x) \geq 3m^2$, $f_2(x) \geq 0$, $f_3(x) \geq 0$. Moreover, $f(x) = 3m^2$ if and only if

$$x_{ij} \in \{0, 1\}, \quad \forall\, i = 1, 2, \ldots, 3m, \quad \forall\, j = 1, 2, \ldots, m, \tag{2.4}$$

$$\sum_{j=1}^{m} x_{ij} \leq 1, \quad \forall\, i = 1, 2, \ldots, 3m, \tag{2.5}$$

$$\sum_{i=1}^{3m} a_i x_{ij} \geq B, \quad \forall\, j = 1, 2, \ldots, m. \tag{2.6}$$

Next, we show that the global minimum of problem (2.3) is not greater than $3m^2$ if and only if the answer to the 3-partition problem is yes. We divide this into two steps. (a) "if" direction. Assuming that there exists a partition of $\mathcal{S}$ such that (2.1) holds true, the system (2.4), (2.5), and (2.6) (with inequalities in (2.5) and (2.6) replaced by equalities) must have a feasible solution x, which further implies $f(x) = 3m^2$. Thus the optimal value of problem (2.3) is not greater than $3m^2$. (b) "only if" direction. Assuming that there exists a point x such that $f(x) = 3m^2$, we know that (2.4), (2.5), and (2.6) hold true at x. In this case, by (2.6), (2.5) and (2.1), we can get that

$$mB \leq \sum_{j=1}^{m} \sum_{i=1}^{3m} a_i x_{ij} = \sum_{i=1}^{3m} a_i \sum_{j=1}^{m} x_{ij} \leq \sum_{i=1}^{3m} a_i = mB.$$

Thus (2.5) and (2.6) must hold with equalities. Combining this with (2.4), (2.5), and (2.6), we can see that x corresponds to a partition of $\mathcal{S}$.

Finally, since this transformation can be done in polynomial time and the 3-partition problem is strongly NP-complete, we conclude that problem (1.4) is strongly NP-hard. $\qquad\square$

Theorem 2.1 indicates that, for any $q \in (0, 1)$, it is computationally intractable to find the global minimizer of problem (1.4).

Remark 2.2 Theorem 2.1 implies that problem (1.1) is strongly NP-hard, since problem (1.4) is a special case of problem (1.1). Note that although the strong NP-hardness of problems (1.7) and (1.8) shown in [22] and [36] can imply the strong NP-hardness of problem (1.1), they do not imply the strong NP-hardness of problem (1.4) as presented in Theorem 2.1.

3 KKT optimality conditions

In this section, we derive the KKT optimality conditions for local minimizers of problem (1.1) (and its special case problem (1.4)). To do so, we introduce an auxiliary problem (3.2) and establish a key one-to-one correspondence of local minimizers of problems (1.1) and (3.2).

For any given $\bar{x} \in \mathcal{X}$, define the sets

$$
\begin{aligned}
\mathcal{I}_{\bar{x}} &= \{m \mid (b - A\bar{x})_m < 0\}, \\
\mathcal{J}_{\bar{x}} &= \{m \mid (b - A\bar{x})_m > 0\}, \\
\mathcal{K}_{\bar{x}} &= \{m \mid (b - A\bar{x})_m = 0\},
\end{aligned}
\tag{3.1}
$$

and the corresponding problem

$$
\begin{aligned}
\min_x \quad & \sum_{m \in \mathcal{J}_{\bar{x}}} (b - Ax)_m^q + h(x) \\
\text{s.t.} \quad & (b - Ax)_m \leq 0, \quad m \in \mathcal{K}_{\bar{x}}, \\
& x \in \mathcal{X}.
\end{aligned}
\tag{3.2}
$$

Notice that the objective value of problem (3.2) is equal to that of problem (1.1) at point $\bar{x}$. Moreover, the objective function of problem (3.2) is continuously differentiable in the neighborhood of point $\bar{x}$.

It is easy to verify the following lemma.

Lemma 3.1 *If $\bar{x}$ is a local minimizer of problem* (1.1)*, then it is also a local minimizer of problem* (3.2) *with $\mathcal{J}_{\bar{x}}$ and $\mathcal{K}_{\bar{x}}$ given in* (3.1)*.*

The following lemma indicates that the converse of Lemma 3.1 is also true.

Lemma 3.2 *If $\bar{x}$ is a local minimizer of problem* (3.2) *with $\mathcal{J}_{\bar{x}}$ and $\mathcal{K}_{\bar{x}}$ given in* (3.1)*, then it is also a local minimizer of problem* (1.1)*.*

Lemma 3.2 can be verified by using the following two facts: the feasible direction cone of problem (1.1) at any feasible point is finitely generated (because $\mathcal{X}$ is a polyhedral set) and the function z^q is non-Lipschitz and concave with respect to $z \geq 0$. Since the detailed proof of Lemma 3.2 is technical, we relegate it to "Appendix 2".

We are now ready to provide the main theorem of this section, which presents the KKT optimality conditions for local minimizers of problem (1.1).

Theorem 3.3 [KKT Optimality Conditions] *If $\bar{x} \in \mathcal{X}$ is a local minimizer of problem* (1.1), *there must exist $\bar{\lambda} \geq 0 \in \mathbb{R}^{|\mathcal{K}_{\bar{x}}|}$ such that*

$$\bar{\lambda}_m (b - A\bar{x})_m = 0, \quad \forall \, m \in \mathcal{K}_{\bar{x}} \tag{3.3}$$

and

$$\bar{x} - P_{\mathcal{X}} \left(\bar{x} - \nabla L(\bar{x}, \bar{\lambda}) \right) = 0, \tag{3.4}$$

where

$$L(x, \lambda) = \sum_{m \in \mathcal{J}_{\bar{x}}} (b - Ax)_m^q + h(x) + \sum_{m \in \mathcal{K}_{\bar{x}}} \lambda_m (b - Ax)_m, \tag{3.5}$$

and $\mathcal{J}_{\bar{x}}$ and $\mathcal{K}_{\bar{x}}$ are defined in (3.1).

Proof By Lemmas 3.1 and 3.2, $\bar{x}$ is a local minimizer of problem (1.1) if and only if it is a local minimizer of problem (3.2) with $\mathcal{J}_{\bar{x}}$ and $\mathcal{K}_{\bar{x}}$ given in (3.1). Combining this equivalence and the fact that $L(x, \lambda)$ in (3.5) is the Lagrangian function of problem (3.2) with λ being the associated Lagrange multiplier, we obtain (3.3) and (3.4) immediately. $\square$

Note that the following versions of the KKT point (or stationary point, or scaled KKT point, or scaled stationary point, or first-order stationary point) for problems (1.5) and (1.6) have been used in many previous works (see, e.g., [6,7,24,25,36]).

Definition 3.4 $\bar{x}$ is called a KKT point of problem (1.5) if it satisfies

$$q |\bar{x}|^q + \bar{X} \nabla h(\bar{x}) = 0, \tag{3.6}$$

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

Definition 3.5 $\bar{x}$ is called a KKT point of problem (1.6) if it satisfies

$$Z_{\bar{x}}^T \nabla F_{\bar{x}}(\bar{x}) = 0, \tag{3.7}$$

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 $\left\{ a_m \mid a_m^T \bar{x} = 0 \right\}$.

In the following, we show that our definition of the KKT point for problem (1.1) in Theorem 3.3 reduces to the ones in Definitions 3.4 and 3.5 when problem (1.1) reduces to problems (1.5) and (1.6), respectively.

Proposition 3.6 *When problem* (1.1) *reduces to problem* (1.6), *there holds*

$$(3.3) \ and \ (3.4) \iff (3.7); \tag{3.8}$$

When problem (1.1) *reduces to problem* (1.5)*, there holds*

$$(3.3) \text{ and } (3.4) \Longleftrightarrow (3.6).$$

Proof For succinctness, we only show the first statement of the proposition. The second one can be shown by using the same arguments. When problem (1.1) reduces to (1.6), problem (3.2) reduces to

$$\min_x \sum_{m \in \hat{\mathcal{I}}_{\bar{x}}} \left(-a_m^T x \right)^q + \sum_{m \in \hat{\mathcal{J}}_{\bar{x}}} \left(a_m^T x \right)^q + h(x)$$
$$\text{s.t. } a_m^T x = 0, \ m \in \hat{\mathcal{K}}_{\bar{x}},$$

with

$$\hat{\mathcal{I}}_{\bar{x}} = \left\{ m \mid a_m^T \bar{x} < 0 \right\}, \ \hat{\mathcal{J}}_{\bar{x}} = \left\{ m \mid a_m^T \bar{x} > 0 \right\}, \text{ and } \hat{\mathcal{K}}_{\bar{x}} = \left\{ m \mid a_m^T \bar{x} = 0 \right\}.$$

Therefore, the KKT optimality conditions (3.3) and (3.4) in Theorem 3.3 reduce to the following: there exists $\bar{\lambda} \in \mathbb{R}^{\left| \hat{\mathcal{K}}_{\bar{x}} \right|}$ such that

$$\bar{\lambda}_m a_m^T \bar{x} = 0, \ \forall \, m \in \hat{\mathcal{K}}_{\bar{x}}, \tag{3.9}$$

and

$$\nabla \hat{L}(\bar{x}, \bar{\lambda}) = 0, \tag{3.10}$$

where

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

Hence, to show (3.8), it suffices to show that

$$(3.9) \text{ and } (3.10) \Longleftrightarrow (3.7).$$

To establish the direction "$\Longrightarrow$", we recall the definitions of $Z_{\bar{x}}$ (see Definition 3.5) and $\hat{L}(x, \lambda)$. By (3.10), we immediately have

$$Z_{\bar{x}}^T \nabla F_{\bar{x}}(\bar{x}) = Z_{\bar{x}}^T \nabla \hat{L}(\bar{x}, \bar{\lambda}) = 0.$$

To establish the direction "$\Longleftarrow$", we have by (3.7) and the defn of $Z_{\bar{x}}$ that

$$\nabla F_{\bar{x}}(\bar{x}) \in Z_{\bar{x}}^{\perp} \Longleftrightarrow \nabla F_{\bar{x}}(\bar{x}) \in \text{span} \left\{ a_m, m \in \hat{\mathcal{K}}_{\bar{x}} \right\}$$
$$\Longleftrightarrow \nabla F_{\bar{x}}(\bar{x}) = - \sum_{m \in \hat{\mathcal{K}}_{\bar{x}}} \bar{\lambda}_m a_m \text{ for some } \left\{ \bar{\lambda}_m \right\}_{m \in \hat{\mathcal{K}}_{\bar{x}}},$$

which implies (3.10) in turn. Due to the definition of $\hat{\mathcal{K}}_{\bar{x}}$, (3.9) holds true trivially. The proof is completed. □

4 Smoothing approximation

Smoothing approximations for nonsmooth minimization have been extensively studied in [6,9,21,52] and references therein. In this section, we propose to use the smooth function

$$\theta(t, \mu) = \begin{cases} t, & \text{if } t > \mu; \\ \frac{t^2}{2\mu} + \frac{\mu}{2}, & \text{if } 0 \leq t \leq \mu; \\ \frac{\mu}{2}, & \text{if } t < 0 \end{cases} \tag{4.1}$$

to approximate the max function

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

Based on (4.1), we can construct a smoothing function $\tilde{F}$ of F and thus a smoothing approximation problem of nonsmooth problem (1.1).

We first summarize some useful properties of $\theta(t, \mu)$. Clearly, for any fixed $\mu > 0$, we have

$$\theta(t, \mu) = \theta(t), \quad \forall t \geq \mu,$$

and

$$\theta(t, \mu) \geq \frac{\mu}{2}, \quad \forall t. \tag{4.2}$$

In addition, $\theta^q(t, \mu)$ is continuously differentiable and twice continuously differentiable everywhere except at the points $t = 0$ and $t = \mu$. The first and second order derivatives of $\theta^q(t, \mu)$ with respect to t are given as follows:

$$\left[\theta^q(t, \mu)\right]' = \begin{cases} qt^{q-1}, & \text{if } t > \mu; \\ q\theta^{q-1}(t, \mu)\frac{t}{\mu}, & \text{if } 0 \leq t \leq \mu; \\ 0, & \text{if } t < 0, \end{cases} \tag{4.3}$$

$$\left[\theta^q(t, \mu)\right]'' = \begin{cases} q(q-1)t^{q-2}, & \text{if } t > \mu; \\ q(q-1)\theta^{q-2}(t, \mu)\frac{t^2}{\mu^2} + q\theta^{q-1}(t, \mu)\frac{1}{\mu}, & \text{if } 0 < t < \mu; \\ 0, & \text{if } t < 0. \end{cases} \tag{4.4}$$

Lemma 4.1 *For any $q \in (0, 1)$ and $\mu \in (0, +\infty)$, the following statements hold true.*

(i) $0 \leq \theta^q(t, \mu) - \theta^q(t) \leq \left(\frac{\mu}{2}\right)^q$, $\forall t \in (-\infty, \mu]$;

(ii) $\max\left\{|v| \mid v \in \partial_t\left(\left[\theta^q(t, \mu)\right]'\right)\right\} \leq 4q\mu^{q-2}$, $\forall t \in \mathbb{R}$, *where ∂_t denotes the Clarke generalized gradient with respect to t [27];*

(iii) Define

$$\kappa(t, \mu) = \begin{cases} 4q\mu^{q-2}, & \text{if } -\mu \le t \le 2\mu; \\ 0, & \text{otherwise.} \end{cases} \tag{4.5}$$

Then

$$\theta^q(t, \mu) \le \theta^q(\hat{t}, \mu) + \left[\theta^q(\hat{t}, \mu)\right]'(t - \hat{t}) + \frac{\kappa(\hat{t}, \mu)}{2}(t - \hat{t})^2 \tag{4.6}$$

for any t and $\hat{t}$ such that $t - \hat{t} \ge -\hat{t}/2$ if $\hat{t} > 2\mu$, or $t \in (-\infty, +\infty)$ if $-\mu \le \hat{t} \le 2\mu$, or $t - \hat{t} \le \mu$ if $\hat{t} < -\mu$.

Proof See "Appendix 3".   □

Define

$$\tilde{F}(x, \mu) = \tilde{f}(x, \mu) + h(x), \tag{4.7}$$

where

$$\tilde{f}(x, \mu) = \sum_{m \in \mathcal{M}} \theta^q((b - Ax)_m, \mu). \tag{4.8}$$

Based on Lemma 4.1 and the discussions beforehand, we know that $\tilde{F}(x, \mu)$ is a smoothing function of $F(x)$ and satisfies

$$F(x) \le \tilde{F}(x, \mu) \le F(x) + M\left(\frac{\mu}{2}\right)^q, \quad \forall x, \tag{4.9}$$

and

$$\nabla \tilde{F}(x, \mu) = \nabla \tilde{f}(x, \mu) + \nabla h(x) = -\sum_{m \in \mathcal{M}} \left[\theta^q(t, \mu)\right]'_{t=(b-Ax)_m} a_m + \nabla h(x). \tag{4.10}$$

Therefore,

$$\begin{aligned} \min_{x} \quad & \tilde{F}(x, \mu) \\ \text{s.t.} \quad & x \in \chi \end{aligned} \tag{4.11}$$

is a smoothing approximation to problem (1.1).

5 An SSQP framework and worst-case iteration complexity analysis

In this section, we propose a smoothing SQP (SSQP) algorithmic framework for solving problem (1.1) (and its special case problem (1.4)). The proposed algorithmic framework (approximately) solves a convex QP at each iteration. The objective function of the QP subproblem is constructed as a local upper bound of the smoothing

function $\tilde{F}(x, \mu)$ in (4.7). In the proposed SSQP framework, the smoothing parameter is updated if the residual of the smoothing problem (4.11) is not greater than some constant (depending on the current smoothing parameter). We shall also analyze the worst-case iteration complexity of the proposed framework.

Specifically, we construct a local convex quadratic upper bound of $\tilde{F}(x, \mu)$ and present the SSQP algorithmic framework for problem (1.1) in Sect. 5.1. Then we define the ϵ-KKT point of problem (1.1) and analyze the worst-case iteration complexity of the proposed algorithm/framework for obtaining an ϵ-KKT point in Sect. 5.2. Finally, we compare the proposed SSQP algorithm/framework with some existing algorithms in Sect. 5.3.

5.1 An SSQP algorithmic framework for problem (1.1)

For any fixed $\mu > 0$, define the quadratic approximation of $\tilde{f}(\cdot, \mu)$ around x_k as

$$Q_1(x, x_k, \mu) = \tilde{f}(x_k, \mu) + \nabla \tilde{f}(x_k, \mu)^T (x - x_k) + \frac{1}{2}(x - x_k)^T \tilde{B}(x_k, \mu)(x - x_k),$$
(5.1)

where $\tilde{f}(x, \mu)$ is given in (4.8),

$$\tilde{B}(x, \mu) = A^T \text{Diag}\left(\kappa((b - Ax)_1, \mu), \ldots, \kappa((b - Ax)_M, \mu)\right) A$$
$$= \sum_{m \in \mathcal{M}} \kappa((b - Ax)_m, \mu) a_m a_m^T,$$

and $\kappa(\cdot, \mu)$ is given in (4.5). By the definition of $\kappa(\cdot, \mu)$, we have

$$\lambda_{\max}\left(\tilde{B}(x, \mu)\right) \leq \lambda_{\max}\left(\sum_{m \in \mathcal{M}} 4q\mu^{q-2} a_m a_m^T\right) \leq 4q\mu^{q-2} \sum_{m \in \mathcal{M}} \|a_m\|^2. \quad (5.2)$$

Similarly, define the quadratic approximation of $h(\cdot)$ around x_k as

$$Q_2(x, x_k) = h(x_k) + \nabla h(x_k)^T (x - x_k) + \frac{1}{2} L_h^k \|x - x_k\|^2,$$

where $L_h^k > 0$ is an estimation of L_h in (1.3). Define

$$Q(x, x_k, \mu) = Q_1(x, x_k, \mu) + Q_2(x, x_k). \quad (5.3)$$

The following lemma indicates that the convex quadratic function $Q(x, x_k, \mu)$ in (5.3) is a local upper bound of the smoothing function $\tilde{F}(x, \mu)$ defined in (4.7) around point x_k as long as

$$h(x) \leq Q_2(x, x_k) \quad (5.4)$$

holds true with $L_h^k > 0$.

Lemma 5.1 *For any x_k and x such that*

$$(A(x_k - x))_m \leq \mu, \ m \in \mathcal{I}_{x_k}^\mu, \tag{5.5}$$

$$(A(x_k - x))_m \geq \frac{-(b - Ax_k)_m}{2}, \ m \in \mathcal{J}_{x_k}^\mu, \tag{5.6}$$

where

$$\begin{aligned}
\mathcal{I}_{x_k}^\mu &= \left\{ m \mid (b - Ax_k)_m < -\mu \right\}, \\
\mathcal{J}_{x_k}^\mu &= \left\{ m \mid (b - Ax_k)_m > 2\mu \right\},
\end{aligned} \tag{5.7}$$

if (5.4) *holds true with* $L_h^k > 0$, *then*

$$\tilde{F}(x, \mu) \leq Q(x, x_k, \mu), \tag{5.8}$$

where $Q(x, x_k, \mu)$ is defined in (5.3).

Proof Recalling the definition of $Q_1(x, x_k, \mu)$ in (5.1) and treating $(b - Ax)_m$ and $(b - Ax_k)_m$ as t and $\hat{t}$ in (iii) of Lemma 4.1, respectively, we get $\tilde{f}(x, \mu) \leq Q_1(x, x_k, \mu)$. Combining this, (5.3), and (5.4), we immediately obtain the desired result (5.8). □

Based on Lemma 5.1, we propose our SSQP algorithmic framework for solving problem (1.1). Some remarks on the proposed SSQP algorithmic framework are in order.

First, to solve the nonsmooth problem (1.1), the proposed SSQP framework approximately solves a series of smoothing approximation problems (4.11) with decreasing smoothing parameters. The solution accuracy of the smoothing approximation problem (4.11) is adaptively controlled by (5.12).

Second, the convex QP problem (5.13) can be efficiently solved (in an exact manner) by the active-set method or the interior-point method [3,38,54,58,62]. In fact, performing a simple shrink projection gradient step for solving problem (5.13) in an inexact fashion is sufficient to guarantee (5.14) (see Lemma 5.4) and hence the worst-case iteration complexity of the proposed framework (see Theorem 5.6).

Third, the Lipschitz constant L_h, when is unknown, is adaptively updated in **Step 4**, which is also used in [1,10,29,42,48,61]. If L_h is known, we can set $L_h^{\max} = L_h^{\min} = L_h^0 = L_h$ in the proposed algorithm, and the r_k in (5.15) satisfies $r_k \leq 1$ at each iteration. Other adaptive ways of updating L_h can also be found in [6,15–17]. However, this will not affect the worst-case iteration complexity order of the proposed framework.

Finally, the parameter μ_0 in (5.9) is chosen such that the final smoothing parameter μ is equal to ϵ once the framework is terminated. This simplifies the worst-case iteration complexity analysis, but does not affect the worst-case iteration complexity order.

In the following analysis, we assume, without loss of generality, that $F(x) \geq 0$ for all $x \in \mathcal{X}$. This, together with (4.9), immediately implies that $\tilde{F}(x, \mu)$ in (4.7) with any $\mu \geq 0$ satisfies

---

An SSQP Algorithmic Framework for Problem (1.1)

Step 1. Initialization. Choose the initial feasible point x_0 and the parameters $\epsilon \in (0, 1]$, $0 < \sigma < 1$, $\eta > 1$, and $L_h^{\max} \geq L_h^0 \geq L_h^{\min} > 0$ with $L_h^{\max} \geq L_h$. Set $i = 0$ (outer iteration index), $k = 0$ (inner iteration index),

$$\mu_0 = \frac{\epsilon}{\sigma^{\lfloor \log_\sigma \epsilon \rfloor}} \in (\sigma, 1], \tag{5.9}$$

and

$$J_0 = \max\left\{ 8q \sum_m \|a_m\|^2 + 2\bar{L}_h, 2 \max_m \{\|a_m\|\} + 2 \right\}, \tag{5.10}$$

where

$$\bar{L}_h \geq \max\left\{ L_h^0, L_h^{\max}, \eta L_h \right\}. \tag{5.11}$$

Step 2. Termination of the inner iteration. Set $\mu = \mu_i$. If x_k satisfies

$$\left\| P_{\mathcal{X}}\left(x_k - \nabla \tilde{F}(x_k, \mu) \right) - x_k \right\| \leq \mu, \tag{5.12}$$

go to **Step 5**; else go to **Step 3**.

Step 3. Calculating the new iterate. Let x_{k+1} be an (approximate) solution of the following convex QP

$$\begin{aligned}
\min_{x \in \mathcal{X}} \quad & Q(x, x_k, \mu) \\
\text{s.t.} \quad & (A(x_k - x))_m \leq \mu, \ m \in \mathcal{I}_{x_k}^\mu, \\
& (A(x_k - x))_m \geq -\frac{(b - Ax_k)_m}{2}, \ m \in \mathcal{J}_{x_k}^\mu
\end{aligned} \tag{5.13}$$

such that

$$\tilde{F}(x_k, \mu) - \tilde{F}(x_{k+1}, \mu) \geq \frac{\mu^{4-q}}{J_0}, \tag{5.14}$$

where $Q(x, x_k, \mu)$ is defined in (5.3) and $\mathcal{I}_{x_k}^\mu$, $\mathcal{J}_{x_k}^\mu$ are defined in (5.7), respectively. Compute $s_{k+1} = x_{k+1} - x_k$, $y_{k+1} = \nabla h(x_{k+1}) - \nabla h(x_k)$, and

$$r_k = \frac{h(x_{k+1}) - h(x_k) - \nabla h(x_k)^T (x_{k+1} - x_k)}{\frac{1}{2} L_h^k \|x_{k+1} - x_k\|^2}. \tag{5.15}$$

Step 4. Updating the estimated Lipschitz constant. If $r_k \leq 1$, compute L_h^{k+1} by

$$L_h^{k+1} = \max\left\{ \min\left\{ L_h^{\max}, \frac{s_{k+1}^T y_{k+1}}{\|s_{k+1}\|^2} \right\}, L_h^{\min} \right\}, \tag{5.16}$$

set $k = k + 1$, and go to **Step 2**; else set

$$L_h^k = \eta L_h^k,$$

and go to **Step 3**.

Step 5. Termination of the outer iteration. If $\mu \leq \epsilon$, terminate the algorithm; else go to **Step 6**.

Step 6. Updating the smoothing parameter. Set

$$\mu_{i+1} = \sigma \mu_i,$$

$i = i + 1$, $x_0 = x_k$, $k = 0$, and go to **Step 2**.

$$\tilde{F}(x, \mu) \geq 0, \quad \forall x \in \mathcal{X}, \quad \forall \mu \geq 0. \tag{5.17}$$

Next, we show that the proposed SSQP framework is well defined and will terminate after finitely many iterations.

Lemma 5.2 *For any $\mu > 0$ and $k \geq 0$,* **Step 3** *in the proposed SSQP framework will be executed at most*

$$K_0 := \left\lceil \log_\eta \frac{L_h}{L_h^{\min}} \right\rceil + 1 \tag{5.18}$$

times, i.e., the convex QP in the form of (5.13) with any $\mu > 0$ and $k \geq 0$ will be (approximately) solved at most K_0 times.

Proof It follows from (5.16) that $L_h^k \geq L_h^{\min}$ for any $k \geq 0$. Since

$$L_h^k \eta^{K_0-1} \geq L_h^{\min} \eta^{K_0-1} \geq L_h,$$

it follows from (1.3) that

$$
\begin{aligned}
h(x_{k+1}) &\leq h(x_k) + \nabla h(x_k)^T (x_{k+1} - x_k) + \frac{1}{2} L_h \|x_{k+1} - x_k\|^2 \\
&\leq h(x_k) + \nabla h(x_k)^T (x_{k+1} - x_k) + \frac{1}{2} L_h^k \eta^{K_0-1} \|x_{k+1} - x_k\|^2,
\end{aligned}
$$

which further implies that r_k in (5.15) satisfies $r_k \leq 1$. According to **Step 4** of the SSQP framework, the inner iteration index k will be incremented after solving the convex QP in the form of (5.13) at most K_0 times.   □

Lemma 5.3 *For any $k \geq 0$ in the SSQP framework, we have*

$$L_h^k \leq \bar{L}_h, \tag{5.19}$$

where $\bar{L}_h$ is given in (5.11).

Proof From (1.3), for any $x \in \mathcal{X}$, we have

$$\frac{h(x) - h(x_k) - \nabla h(x_k)^T (x - x_k)}{\frac{1}{2} L_h \|x - x_k\|^2} \leq 1.$$

According to **Step 4** of the SSQP framework, L_h^k is set to be ηL_h^k only when $r_k > 1$, and in this case there must hold $L_h^k < L_h$. Hence, (5.11) is true.

The following Lemma 5.4 guarantees the existence of x_{k+1} satisfying the relation (5.14).

Lemma 5.4 *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 (5.13),
- x_{k+1}^{snorm} is the solution of the following problem

$$\min_{x \in \mathcal{X}} \quad Q(x, x_k, \mu) \tag{5.20}$$
$$s.t. \quad \|A(x - x_k)\|_\infty \leq \mu,$$

- *and*

$$x_{k+1}^{proj} = x_k + \xi_k \tau_k d_k, \tag{5.21}$$

where

$$\tau_k = \frac{\mu}{(\max_m \{\|a_m\|\} + 1) \|d_k\|}, \tag{5.22}$$

$$\xi_k = \min \left\{ \frac{-d_k^T \nabla \tilde{F}(x_k, \mu)}{\tau_k d_k^T \left(\tilde{B}_k + L_h I_N \right) d_k}, 1 \right\}, \tag{5.23}$$

and

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

If (5.12) is not satisfied, then

$$\tilde{F}(x_k, \mu) - \tilde{F}(x_{k+1}^{exact}, \mu) \geq \tilde{F}(x_k, \mu) - \tilde{F}(x_{k+1}^{snorm}, \mu)$$
$$\geq \tilde{F}(x_k, \mu) - \tilde{F}(x_{k+1}^{proj}, \mu) \geq \frac{\mu^{4-q}}{J_0}, \tag{5.24}$$

where J_0 and $\bar{L}_h$ are given in (5.10) and (5.11), respectively.

Proof For simplicity, denote $\tilde{B}(x_k, \mu)$ and $\nabla \tilde{F}(x_k, \mu)$ by $\tilde{B}_k$ and $\nabla \tilde{F}_k$, respectively in the proof. Since $d_k = P_{\mathcal{X}}(x_k - \nabla \tilde{F}_k) - x_k$, it follows from the property of projection that

$$- \nabla \tilde{F}_k^T d_k \geq \|d_k\|^2 \geq 0. \tag{5.25}$$

This implies that ξ_k in (5.23) satisfies $\xi_k \in [0, 1]$.

Next, we first show the last inequality $\tilde{F}(x_k, \mu) - \tilde{F}(x_{k+1}^{proj}, \mu) \geq \mu^{4-q}/J_0$ in (5.24) holds true.

We claim that $x_k(\xi) := x_k + \xi \tau_k d_k$ is feasible to problem (5.13) for all $\xi \in [0, 1]$. First of all, since (5.12) is not satisfied, we have

$$\|d_k\| > \mu. \tag{5.26}$$

By this and (5.22), we get

$$\tau_k \leq \frac{\mu}{(\max_m \{\|a_m\|\} + 1) \mu} = \frac{1}{\max_m \{\|a_m\|\} + 1} \leq 1.$$

Hence, $x_k(\xi) = (1 - \xi\tau_k)\,x_k + \xi\tau_k\,P_{\mathcal{X}}(x_k - \nabla\tilde{F}_k)$ is a convex combination of $x_k \in \mathcal{X}$ and $P_{\mathcal{X}}(x_k - \nabla\tilde{F}_k) \in \mathcal{X}$. By the convexity of $\mathcal{X}$, we have

$$x_k(\xi) \in \mathcal{X}, \quad \forall\, \xi \in [0, 1]. \tag{5.27}$$

Moreover, by the definition (5.22) of τ_k, we have

$$\left| (A(x_k - x_k(\xi)))_m \right| = \left| a_m^T(x_k - x_k(\xi)) \right| = \xi\tau_k |a_m^T d_k| \leq \xi\tau_k \|a_m\| \|d_k\| \leq \mu \tag{5.28}$$

for all $\xi \in [0, 1]$ and $m \in \mathcal{M}$. This shows that $x_k(\xi)$ satisfies (5.5) and (5.6) for all $\xi \in [0, 1]$. Hence, $x_k(\xi)$ is feasible to problem (5.13) for all $\xi \in [0, 1]$ and hence x_{k+1}^{proj} in (5.21) (due to $\xi_k \in [0, 1]$).

Now, we consider the univariate box constrained QP problem

$$\xi_k = \arg\min_{0 \leq \xi \leq 1} Q(x_k + \xi\tau_k d_k, x_k, \mu), \tag{5.29}$$

which admits a closed-form solution (5.23). We first consider the case $\xi_k = 1$, which implies that $-\nabla\tilde{F}_k^T d_k \geq \tau_k d_k^T\left(\tilde{B}_k + L_h I_N\right) d_k$. Here, we have

$$\tau_k \nabla\tilde{F}_k^T d_k + \frac{1}{2}\tau_k^2 d_k(\tilde{B}_k + L_h I_N)d_k \leq \frac{\tau_k}{2}\nabla\tilde{F}_k^T d_k \leq -\frac{\tau_k}{2}\|d_k\|^2 = \frac{-\mu\|d_k\|}{2\,(\max\{\|a_m\|\}+1)}, \tag{5.30}$$

where the second inequality is due to (5.25). For the other case where $\xi_k = \dfrac{-\nabla\tilde{F}_k^T d_k}{\tau_k d_k^T\left(\tilde{B}_k + L_h I_N\right)d_k}$, we have

$$
\begin{aligned}
\xi_k\tau_k\nabla\tilde{F}_k^T d_k + \frac{1}{2}\xi_k^2\tau_k^2 d_k(\tilde{B}_k + L_h I_N)d_k &= -\frac{\left(\nabla\tilde{F}_k^T d_k\right)^2}{2 d_k^T\left(\tilde{B}_k + L_h I_N\right)d_k} \\
&\leq -\frac{\|d_k\|^4}{2\lambda_{\max}\left(\tilde{B}_k + L_h I_N\right)\|d_k\|^2} \\
&\leq -\frac{\|d_k\|^2\mu^{2-q}}{2\left(4q\sum_m\|a_m\|^2 + L_h\right)},
\end{aligned}
\tag{5.31}
$$

where the first inequality is due to (5.25) and the last inequality is due to (5.2). Combining (5.8), (5.30), and (5.31), we obtain

$$
\begin{aligned}
\tilde{F}(x_k, \mu) - \tilde{F}(x_{k+1}^{\text{proj}}, \mu) &= \tilde{F}(x_k, \mu) - \tilde{F}(x_k + \xi_k\tau_k d_k, \mu) \\
&\geq \tilde{F}(x_k, \mu) - Q(x_k + \xi_k\tau_k d_k, \mu)
\end{aligned}
$$

$$= -\xi_k \tau_k \nabla \tilde{F}_k^T d_k - \frac{1}{2} \xi_k^2 \tau_k^2 d_k (\tilde{B}_k + L_h) d_k$$

$$\geq \min \left\{ \frac{\mu \|d_k\|}{2 \left(\max \{\|a_m\|\} + 1 \right)}, \frac{\|d_k\|^2 \mu^{2-q}}{2 \left(4q \sum_m \|a_m\|^2 + L_h \right)} \right\}.$$

This, together with (5.10), (5.11), and (5.26), immediately implies the desired result

$$\tilde{F}(x_k, \mu) - \tilde{F}(x_{k+1}^{\text{proj}}, \mu) \geq \frac{\mu^{4-q}}{J_0}.$$

Now we show the first two inequalities in (5.24). From the above analysis (cf. (5.27) and (5.28)), we know that x_{k+1}^{proj} is feasible to problem (5.20). Since x_{k+1}^{snorm} is the solution of problem (5.20), it follows that

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

Moreover, since the feasible region of problem (5.20) is a subset of the one of problem (5.13), we immediately get

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

The proof is completed. $\qquad\square$

As shown in Lemma 5.4, for the next iterate x_{k+1} to achieve a decrease of μ^{4-q}/J_0 as required in (5.14), problem (5.13) is not necessarily to be solved in an exact manner; a simple shrink projection gradient step, i.e., $x_{k+1}^{\text{proj}} = x_k + \xi_k \tau_k d_k$, suffices to satisfy (5.24). This gives the flexibility to choose subroutines for solving problem (5.13) inexactly.

It is also worthwhile remarking that Lemma 5.4 holds true for any convex set $\mathcal{X}$, which is not necessarily a polyhedron. If $\mathcal{X} = \mathbb{R}^N$, then problem (5.20) is a trust region subproblem with a scaled infinity norm constraint. The infinity norm in (5.20) could be replaced by the Euclidean norm, and the solution to the corresponding counterpart still satisfies (5.14).

Without loss of generality, we focus on analyzing the SSQP framework when x_{k+1} in **Step 3** is chosen to be x_{k+1}^{proj} in (5.21) in the rest part of this section. The same results hold true if x_{k+1} in **Step 3** is chosen such that (5.14) is satisfied. For instance, x_{k+1} can also be chosen to be x_{k+1}^{exact} or x_{k+1}^{snorm} in Lemma 5.4.

The following lemma states that the inner loop termination criterion (5.12) of the SSQP framework can be satisfied after finite number of iterations.

Lemma 5.5 *Let* $x_{k+1} = x_{k+1}^{\text{proj}}$ *in the proposed SSQP framework. Then, for any* $\mu \in (0, 1]$, *Step 2 of the proposed SSQP framework will be executed at most*

$$\left\lceil \tilde{F}(x_0, 1) J_0 \mu^{q-4} \right\rceil$$

times, and the inner termination criterion (5.12) is satisfied after at most

$$\left\lceil \tilde{F}(x_0, 1) J_0 K_0 \mu^{q-4} \right\rceil$$

iterations, where J_0 and K_0 are given in (5.10) and (5.18), respectively.

Proof For any $\mu \in (0, 1]$, we use $\hat{k}$ to denote the first inner iteration index such that $x_{\hat{k}}$ satisfies (5.12). Then (5.26) holds true for all $k \leq \hat{k} - 1$. Combining this and Lemma 5.4, we obtain

$$\tilde{F}(x_0, \mu) - \tilde{F}(x_{\hat{k}}, \mu) = \sum_{k=1}^{\hat{k}} \left(\tilde{F}(x_{k-1}, \mu) - \tilde{F}(x_k, \mu) \right) \geq \hat{k} \frac{\mu^{4-q}}{J_0}.$$

By using (5.17), we conclude

$$\hat{k} \leq \tilde{F}(x_0, \mu) J_0 \mu^{q-4} \leq \tilde{F}(x_0, 1) J_0 \mu^{q-4}.$$

This, together with Lemma 5.2, immediately implies the second statement of Lemma 5.5.   □

Now, we are ready to show that the proposed SSQP framework terminates after finite number of iterations.

Theorem 5.6 *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*

$$\left\lceil J_T^q \epsilon^{q-4} \right\rceil \tag{5.32}$$

iterations, where

$$J_T^q = \frac{\sigma^{q-4} \left(\tilde{F}(x_0, 1) J_0 K_0 + 1 \right)}{\sigma^{q-4} - 1}, \tag{5.33}$$

and K_0 and J_0 are defined in (5.18) and (5.10), respectively.

Proof Define $I_0 = \lfloor \log_\sigma \epsilon \rfloor$. According to the SSQP framework, we have

$$\mu_i = \mu_0 \sigma^i \geq \mu_0 \sigma^{I_0} = \epsilon, \quad \forall\, i = 0, 1, \ldots, I_0. \tag{5.34}$$

In particular, we have

$$\mu_{I_0} = \mu_0 \sigma^{I_0} = \epsilon. \tag{5.35}$$

By (5.34) and Lemma 5.5, for any fixed $\mu > 0$, the number of iterations that the SSQP framework takes to return a point satisfying (5.12) is at most

$$\left\lceil \tilde{F}(x_0, 1) J_0 K_0 \mu^{q-4} \right\rceil = \left\lceil \tilde{F}(x_0, 1) J_0 K_0 \left(\mu_0 \sigma^i \right)^{q-4} \right\rceil.$$

Therefore, the total number of iterations for the proposed framework to terminate is at most

$$\sum_{i=0}^{I_0} \left\lceil \tilde{F}(x_0, 1) J_0 K_0 \left(\mu_0 \sigma^i\right)^{q-4} \right\rceil$$

$$\leq \left(\tilde{F}(x_0, 1) J_0 K_0 + 1\right) \mu_0^{q-4} \frac{\sigma^{(q-4)(I_0+1)} - 1}{\sigma^{q-4} - 1} \leq J_T^q \epsilon^{q-4},$$

where the last inequality is due to (5.33) and (5.35). □

The worst-case iteration complexity function in (5.32) is a strictly decreasing function with respect to $q \in (0, 1)$ for fixed $\epsilon \in (0, 1)$. This is intuitive because problem (1.1) becomes more difficult to solve as q decreases.

5.2 Worst-case iteration complexity analysis

In this subsection, we show that the point returned by the proposed SSQP framework is an ϵ-KKT point of problem (1.1). To do this, we need to give the definition of the ϵ-KKT point first. Our definition of the ϵ-KKT point of problem (1.1) is given as follows, which is a perturbation of the KKT optimality conditions in Theorem 3.3.

Definition 5.7 (ϵ-*KKT Point*) For any given $\epsilon > 0$, $\bar{x} \in \mathcal{X}$ is called an ϵ-KKT point of problem (1.1) if there exists $\bar{\lambda} \geq 0 \in \mathbb{R}^{|\mathcal{K}_{\bar{x}}^\epsilon|}$ such that

$$\left|\bar{\lambda}_m (b - A\bar{x})_m\right| \leq \epsilon^q, \quad m \in \mathcal{K}_{\bar{x}}^\epsilon \tag{5.36}$$

and

$$\left\|\bar{x} - P_{\mathcal{X}}\left(\bar{x} - \nabla L^\epsilon(\bar{x}, \bar{\lambda})\right)\right\| \leq \epsilon, \tag{5.37}$$

where

$$L^\epsilon(x, \lambda) = \sum_{m \in \mathcal{J}_{\bar{x}}^\epsilon} (b - Ax)_m^q + h(x) + \sum_{m \in \mathcal{K}_{\bar{x}}^\epsilon} \lambda_m (b - Ax)_m \tag{5.38}$$

with

$$\begin{aligned}
\mathcal{I}_{\bar{x}}^\epsilon &= \{m \mid (b - A\bar{x})_m < -\epsilon\}, \\
\mathcal{J}_{\bar{x}}^\epsilon &= \{m \mid (b - A\bar{x})_m > \epsilon\}, \\
\mathcal{K}_{\bar{x}}^\epsilon &= \{m \mid -\epsilon \leq (b - A\bar{x})_m \leq \epsilon\}.
\end{aligned} \tag{5.39}$$

Notice that if $\epsilon = 0$ in (5.36), (5.37), and (5.39), then the ϵ-KKT point in Definition 5.7 reduces to the KKT point of problem (1.1) (cf. Theorem 3.3).

The following definition of the ϵ-KKT point for problem (1.6) has been used in [7,24].

Definition 5.8 For any $\epsilon \in (0, 1]$, $\bar{x}$ is called an ϵ-KKT point of problem (1.6) if it satisfies

$$\left\| \left(Z_{\bar{x}}^{\epsilon}\right)^T \nabla F_{\bar{x}}^{\epsilon}(\bar{x}) \right\|_{\infty} \leq \epsilon, \tag{5.40}$$

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 $\left\{ a_m \mid \left| a_m^T \bar{x} \right| \leq \epsilon \right\}$.

Our definition of the ϵ-KKT point in Definition 5.7 reduces to Definition 5.9 when problem (1.1) reduces to problem (1.6).

Definition 5.9 For any given $\epsilon > 0$, $\bar{x}$ is called an ϵ-KKT point of problem (1.6) if there exists $\bar{\lambda} \in \mathbb{R}^{|\hat{\mathcal{K}}_{\bar{x}}^{\epsilon}|}$ such that

$$\left| \bar{\lambda}_m a_m^T \bar{x} \right| \leq \epsilon^q, \quad m \in \hat{\mathcal{K}}_{\bar{x}}^{\epsilon} \tag{5.41}$$

and

$$\left\| \nabla \hat{L}^{\epsilon}(\bar{x}, \bar{\lambda}) \right\| \leq \epsilon, \tag{5.42}$$

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

$$\begin{aligned}
\hat{\mathcal{I}}_{\bar{x}}^{\epsilon} &= \left\{ m \mid a_m^T \bar{x} < -\epsilon \right\}, \\
\hat{\mathcal{J}}_{\bar{x}}^{\epsilon} &= \left\{ m \mid a_m^T \bar{x} > \epsilon \right\}, \\
\hat{\mathcal{K}}_{\bar{x}}^{\epsilon} &= \left\{ m \mid -\epsilon \leq a_m^T \bar{x} \leq \epsilon \right\}.
\end{aligned} \tag{5.43}$$

Remark 5.10 The ϵ-KKT point for problem (1.6) in Definition 5.9 is stronger than the one in Definition 5.8. On one hand, it is clear that (5.42) implies (5.40). On the other hand, if (5.40) is true, then there must exist $\bar{\lambda}$ such that (5.42) is satisfied.[1] However, this $\bar{\lambda}$ does not necessarily satisfy (5.41).

In the next, we show that the point returned by the proposed SSQP framework is an ϵ-KKT point of problem (1.1) defined in Definition 5.7.

For any given $\epsilon \in (0, 1]$, let $\bar{x}$ be the point returned by the proposed SSQP framework. When the framework is terminated, there holds $\mu = \epsilon$ (cf. (5.35)). Then, it

---

[1] Here, the differences between two norms ($\| \cdot \|_{\infty}$ in (5.40) and $\| \cdot \|$ in (5.42)) are neglected.

follows from (5.12) that $\bar{x} \in \mathcal{X}$ satisfies

$$\left\| \bar{x} - P_{\mathcal{X}}(\bar{x} - \nabla \tilde{F}(\bar{x}, \epsilon)) \right\| \leq \epsilon. \tag{5.44}$$

Define $\mathcal{I}_{\bar{x}}^{\epsilon}$, $\mathcal{J}_{\bar{x}}^{\epsilon}$, and $\mathcal{K}_{\bar{x}}^{\epsilon}$ as in (5.39), and

$$\bar{\lambda}_m = \left[\theta^q(t, \epsilon) \right]'_{t=(b-A\bar{x})_m}, \quad m \in \mathcal{K}_{\bar{x}}^{\epsilon}. \tag{5.45}$$

It is obvious that $\bar{\lambda}_m \geq 0$ for all $m \in \mathcal{K}_{\bar{x}}^{\epsilon}$.

Theorem 5.11 *For any $\epsilon \in (0, 1]$, let $\bar{x}$ be the point returned by the proposed SSQP framework and $\bar{\lambda}$ be defined in (5.45). Then $\bar{x}$ and $\bar{\lambda}$ satisfy (5.36) and (5.37).*

Proof Let us first show that $\bar{x}$ and $\bar{\lambda}$ satisfy (5.36).

- For any $m \in \mathcal{K}_{\bar{x}}^{\epsilon}$ with $-\epsilon \leq (b - A\bar{x})_m \leq 0$, it follows from (4.3) and (5.45) that $\bar{\lambda}_m = 0$, and thus $\left| \bar{\lambda}_m (b - A\bar{x})_m \right| = 0 \leq \epsilon^q$;
- For any $m \in \mathcal{K}_{\bar{x}}^{\epsilon}$ with $0 < (b - A\bar{x})_m \leq \epsilon$, we have

$$\left| \bar{\lambda}_m (b - A\bar{x})_m \right| = q\theta^{q-1}((b-A\bar{x})_m, \epsilon) \frac{(b-A\bar{x})_m}{\epsilon} (b-A\bar{x})_m \leq q \left(\frac{\epsilon}{2} \right)^{q-1} \epsilon \leq \epsilon^q,$$

where the equality comes from (4.3) and (5.45), the first inequality is due to (4.2) and $0 < (b-A\bar{x})_m \leq \epsilon$, and the second inequality is due to the fact that $q2^{1-q} \leq 1$ for all $q \in (0, 1)$.

Now we show that $\bar{x}$ and $\bar{\lambda}$ satisfy (5.37). By (5.44) and the nonexpansive property of the projection operator $P_{\mathcal{X}}(\cdot)$, we get

$$\left\| \bar{x} - P_{\mathcal{X}}(\bar{x} - \nabla L^{\epsilon}(\bar{x}, \bar{\lambda})) \right\|$$
$$\leq \left\| \bar{x} - P_{\mathcal{X}}(\bar{x} - \nabla \tilde{F}(\bar{x}, \epsilon)) \right\| + \left\| P_{\mathcal{X}}(\bar{x} - \nabla \tilde{F}(\bar{x}, \epsilon)) - P_{\mathcal{X}}(\bar{x} - \nabla L^{\epsilon}(\bar{x}, \bar{\lambda})) \right\|$$
$$\leq \epsilon + \| \nabla \tilde{F}(\bar{x}, \epsilon) - \nabla L^{\epsilon}(\bar{x}, \bar{\lambda}) \|. \tag{5.46}$$

From (4.10), (5.38), and (5.45), we have $\nabla \tilde{F}(\bar{x}, \epsilon) - \nabla L^{\epsilon}(\bar{x}, \bar{\lambda}) = 0$. Combining this with (5.46) immediately yields (5.37). The proof is completed. □

By combining Theorems 5.6 and 5.11, we obtain the following worst-case iteration complexity result.

Theorem 5.12 *For any $\epsilon \in (0, 1]$, the total number of iterations for the SSQP framework to return an ϵ-KKT point of problem (1.1) satisfying (5.36) and (5.37) is at most*

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

where J_T^q is given in (5.33).

As a direct consequence of Theorem 5.12, we have the following corollary.

Corollary 5.13 *For any $\epsilon \in (0, 1]$, let $\bar{x}$ be the point returned by the SSQP framework when applied to solve problem (1.6), and define*

$$\bar{\lambda}_m = sgn\left(a_m^T \bar{x}\right) \left[\theta^q(t, \epsilon)\right]'_{t=|a_m^T \bar{x}|}, \quad m \in \hat{\mathcal{K}}_{\bar{x}}^{\epsilon},$$

where $\hat{\mathcal{K}}_{\bar{x}}^{\epsilon}$ is given in (5.43). Then, $\bar{x}$ and $\bar{\lambda}$ satisfy (5.41) and (5.42). Moreover, the total number of iterations for the SSQP framework to return the ϵ-KKT point $\bar{x}$ of problem (1.6) is at most

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

5.3 Comparisons of SSQP and SQR

In this subsection, we compare the proposed SSQP framework with the SQR algorithms proposed in [7] and [24] for solving problem (1.6) (possibly with box constraints) and problem (1.5), respectively.

First of all, the SSQP algorithmic framework is designed for solving a more difficult problem, i.e., problem (1.1) with a composite non-Lipschitzian objective and a general polyhedral constraint, which includes problems (1.5) and (1.6) as special cases.

Now, we give a detailed comparison of the SQR algorithm in [7] and the proposed SSQP framework with x_{k+1} chosen to be x_{k+1}^{proj} at each iteration from the perspective of iteration complexity and solution quality when both of them are applied to solve the unconstrained problem (1.6); see Table 1, where $\tilde{F}(x, \epsilon)$ reduces to

$$\tilde{F}(x, \epsilon) = \sum_{m \in \mathcal{M}} \left(\theta^q\left(a_m^T x, \epsilon\right) + \theta^q\left(-a_m^T x, \epsilon\right)\right) + h(x).$$

It is shown in [7] that the SQR algorithm returns an ϵ-KKT point $\bar{x}$ satisfying (5.40) within $\mathcal{O}(\epsilon^{-2})$ iterations. The SSQP framework with $x_{k+1} = x_{k+1}^{proj}$, when applied to

Table 1 Comparisons of the SQR algorithm in [7] and the proposed SSQP framework with $x_{k+1} = x_{k+1}^{proj}$ in (5.21) for solving problem (1.6)

	SQR [7]	SSQP
Complexity		
Iteration number	$\mathcal{O}(\epsilon^{-2})$	$\mathcal{O}(\epsilon^{q-4})$
Subproblem per iteration	n-Dimensional QP (2.9) in [7]	Univariate QP (5.29)
Quality		
Optimality residual I	$\left\|\left(Z_{\bar{x}}^{\epsilon}\right)^T \nabla F_{\bar{x}}^{\epsilon}(\bar{x})\right\|_{\infty} \le \epsilon$	$\left\|\nabla \hat{L}^{\epsilon}(\bar{x}, \bar{\lambda})\right\| \le \epsilon$
Optimality residual II	$\left\|\nabla \tilde{F}(\bar{x}, \epsilon)\right\| \le \mathcal{O}\left(\epsilon^{2-2/q}\right)$	$\left\|\nabla \tilde{F}(\bar{x}, \epsilon)\right\| \le \epsilon$
Complementary violation	Not guaranteed	$\left|\bar{\lambda}_m a_m^T \bar{x}\right| \le \epsilon^q, \ m \in \hat{\mathcal{K}}_{\bar{x}}^{\epsilon}$

solve problem (1.6), can return an ϵ-KKT point $\bar{x}$ satisfying (5.41) and (5.42) in no more than $\mathcal{O}(\epsilon^{q-4})$ iterations (see Corollary 5.13). Here, one iteration in the SQR algorithm needs solving exactly an n-dimensional box constrained QP (problem (2.9) in [7]), and the exact solution of the QP subproblem is necessary for Lemma 2.3 there to hold true. Since the condition number of the quadratic objective in (2.9) in [7] increases asymptotically with $\mathcal{O}(\mu^{q-2})$ as the smoothing parameter μ decreases, the n-dimensional box constrained QP (2.9) becomes more and more difficult to solve. In contrast, one iteration in the SSQP framework with $x_{k+1} = x_{k+1}^{\mathrm{proj}}$ only needs solving approximately the QP problem (5.13). As shown in Lemma 5.4, a good approximate solution (5.21) of problem (5.13) can be obtained by solving an univariate box constrained QP (5.29), which admits a closed-form solution (5.23).

From the perspective of solution quality, the SSQP framework (with $x_{k+1} = x_{k+1}^{\mathrm{proj}}$) actually returns a better solution compared to the SQR algorithm. As discussed in Remark 5.10, the ϵ-KKT point returned by the SSQP framework is stronger than the one returned by the SQR algorithm. In addition, in terms of the residual of smoothing problem, the SSQP framework actually returns an ϵ-KKT point $\bar{x}$ satisfying $\left\| \nabla \tilde{F}(\bar{x}, \epsilon) \right\| \leq \epsilon$, while the SQR algorithm outputs an ϵ-KKT point $\bar{x}$ with $\left\| \nabla \tilde{F}(\bar{x}, \epsilon) \right\| \leq \mathcal{O}\left(\epsilon^{2-2/q}\right)$. To derive this bound, it suffices to use (2.13) and (2.19) of [7]. Notice that both of the parameters γ and ρ in (2.13) are zero in the unconstrained case.

Finally, we remark that the proposed SSQP framework can be directly applied to solve problem (1.1) with $q = 1$ and is guaranteed to return an ϵ-Clarke KKT point of problem (1.1) within $\mathcal{O}(\epsilon^{-3})$ iterations. The worst-case iteration complexity of the proposed SSQP framework for computing an ϵ-Clarke KKT point of problem (1.1) with $q = 1$ is thus the same as the one of the SQR$_1$ algorithm for problem (1.5) with $q = 1$ in [6] and better than $\mathcal{O}(\epsilon^{-3} \log \epsilon^{-1})$ of the smoothing direct search algorithm for unconstrained Lipschitzian minimization problems in [35].

In the following, we first extend the definition of ϵ-Clarke KKT point for unconstrained locally Lipschitz continuous optimization problem in [35] to constrained locally Lipschitz continuous optimization problem (1.1) with $q = 1$, and then present the worst-case iteration complexity of the proposed SSQP framework for obtaining such an ϵ-Clarke KKT point.

Definition 5.14 The point $\bar{x}$ is called an ϵ-Clarke KKT point of problem (1.1) with $q = 1$ if it satisfies

$$\left\| P_{\mathcal{X}}\left(\bar{x} - \nabla \tilde{F}(\bar{x}, \mu)\right) - \bar{x} \right\| \leq \epsilon \quad \text{and} \quad \mu \leq \epsilon.$$

Using the same argument as in the proof of Theorem 5.6, we can show the following iteration complexity result.

Theorem 5.15 *For any $\epsilon \in (0, 1]$, the total number of iterations for the SSQP framework to return an ϵ-Clarke KKT point of problem (1.1) with $q = 1$ is at most $\mathcal{O}\left(\epsilon^{-3}\right)$.*

6 Concluding remarks

In this paper, we considered the composite nonsmooth nonconvex non-Lipschitzian L_q $(0 < q < 1)$ minimization problem (1.1) over a general polyhedral set. We derived KKT optimality conditions for problem (1.1). These conditions unify various optimality conditions for non-Lipschitzian optimization problems developed in [6,7,24,25,36]. Moreover, we proposed an SSQP framework for solving problem (1.1) and showed that the proposed framework is guaranteed to return an ϵ-KKT point of problem (1.1) satisfying (5.36) and (5.37) within $\mathcal{O}(\epsilon^{q-4})$ iterations. To the best of our knowledge, this is the first algorithmic framework for the polyhedral constrained composite L_q minimization with worst-case iteration complexity analysis. The proposed SSQP framework can directly be applied to solve problem (1.1) with $q = 1$ and its worst-case iteration complexity for returning an ϵ-Clarke KKT point is $\mathcal{O}(\epsilon^{-3})$.

Although we focused on the L_q minimization problem (1.1) in this paper, the techniques developed here can be useful for developing and analyzing algorithms for problems with other regularizers such as the ones given in "Appendix 1" of [6]. Moreover, most of the results presented in this paper can be easily generalized to problem (1.1) where the unknown variable is a positive semidefinite matrix.

Appendix 1: Three motivating applications

Support Vector Machine [11,28]. The support vector machine (SVM) is a state-of-the-art classification method introduced by Boser, Guyon, and Vapnik in 1992 in [11]. 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 pattern or example and y_m is the label associated with s_m. For convenience, we assume the labels are $+1$ for positive examples and -1 for negative examples. If the data are linearly separable, the task of SVM is to find a linear discriminant function of the form $\ell(s) = \hat{s}^T x$ with $\hat{s} = [s^T, 1]^T \in \mathbb{R}^N$ such that all data are correctly classified and at the same time the margin of the hyperplane ℓ that separates the two classes of examples is maximized. Mathematically, the above problem can be formulated as

$$\min_{x} \quad \frac{1}{2} \sum_{n=1}^{N-1} x_n^2 \qquad (6.1)$$
$$\text{s.t.} \quad y_m \hat{s}_m^T x \geq 1, \quad m = 1, 2, \ldots, M.$$

In practice, data are often not linearly separable. In this case, problem (6.1) is not feasible, and the following problem can be solved instead:

$$\min_x \sum_{m=1}^{M} \max\left\{1 - y_m \hat{s}_m^T x, 0\right\}^q + \frac{\rho}{2} \sum_{n=1}^{N-1} x_n^2, \tag{6.2}$$

where the constant $\rho \geq 0$ balances the relative importance of minimizing the classification errors and maximizing the margin. Problem (6.2) with $q = 1$ is called the soft-margin SVM in [28]. It is clear that problem (6.2) is a special instance of (1.1) with

$$A = \begin{bmatrix} y_1 \hat{s}_1^T \\ \vdots \\ y_M \hat{s}_M^T \end{bmatrix}, \quad b = e, \quad h(x) = \frac{\rho}{2} \sum_{n=1}^{N-1} x_n^2, \quad \text{and } \mathcal{X} = \mathbb{R}^N.$$

Here, e is the all-one vector of dimension M.

Joint Power and Admission Control [46,49]. Consider a wireless network consisting of K interfering links (a link corresponds to a transmitter/receiver pair) with channel gains $g_{kj} \geq 0$ (from the transmitter of link j to the receiver of link k), noise power $\eta_k > 0$, signal-to-interference-plus-noise-ratio (SINR) target $\gamma_k > 0$, and power budget $\bar{p}_k > 0$ for $k, j = 1, 2, \ldots, K$. Denoting the transmission power of transmitter k by x_k, the SINR at the kth receiver can be expressed as

$$\text{SINR}_k = \frac{g_{kk} x_k}{\eta_k + \sum_{j \neq k} g_{kj} x_j}, \quad k = 1, 2, \ldots, K. \tag{6.3}$$

Due to the existence of mutual interferences among different links [which correspond to the term $\sum_{j \neq k} g_{kj} x_j$ in (6.3)], the linear system

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

may not be feasible. The joint power and admission control problem aims at supporting a maximum number of links at their specified SINR targets while using a minimum total transmission power. Assuming without loss of generality that $g_{kk} = \gamma_k = \bar{p}_k = 1$ for all $k = 1, 2, \ldots, K$, the joint power and admission control problem can be formulated as follows (see [46])

$$\begin{aligned} \min_x \quad & \|\max\{b - Ax, 0\}\|_q^q + \rho e^T x \\ \text{s.t.} \quad & 0 \leq x \leq e, \end{aligned} \tag{6.4}$$

where $\rho > 0$ is a parameter, $b = [\eta_1, \eta_2, \ldots, \eta_K]^T$, and $A = [a_{kj}] \in \mathbb{R}^{K \times K}$ with

$$a_{kj} = \begin{cases} 1, & \text{if } k = j; \\ -g_{kj}, & \text{if } k \neq j. \end{cases}$$

By utilizing the special structure of A, i.e., all of its diagonal entries are positive and off-diagonal entries are nonpositive, it is shown in [47, Theorem 1] that the solution of problem (6.4) can maximize the number of supported links using a minimum total

transmission power as long as q is chosen to be sufficiently small (but not necessarily to be zero). Clearly, (6.4) is a special case of (1.1) with

$$M = K, \ N = K, \ h(x) = \rho e^T x, \quad \text{and } \mathcal{X} = \{x \mid 0 \le x \le e\} \subseteq \mathbb{R}^N.$$

Linear Decoding Problem [13]. Given the coding matrix $C \in \mathbb{R}^{K_1 \times K_2}$ and corrupted measurement $c = Cx + e_u \in \mathbb{R}^{K_1}$, where e_u is an unknown vector of errors, the linear decoding problem is to recover x from c. It is shown in [13] that, if C satisfies the restricted isometry property, x can be exactly recovered by solving the convex minimization problem

$$\min_x \|c - Cx\|_1$$

provided that e_u is sparse. By [33, Theorem 4.10], L_q ($q \in (0, 1)$) minimization

$$\min_x \|c - Cx\|_q^q \tag{6.5}$$

has a better capability of recovering x than L_1 minimization. By using the equation $|a| = \max\{a, 0\} + \max\{-a, 0\}$, it is simple to see problem (6.5) is a special case of (1.1) with

$$M = 2K_1, \quad N = K_2, \quad A = \begin{bmatrix} C \\ -C \end{bmatrix}, \quad b = \begin{bmatrix} c \\ -c \end{bmatrix}, \quad h(x) = 0, \quad \text{and } \mathcal{X} = \mathbb{R}^N.$$

Appendix 2: Proof of Lemma 3.2

Let $\bar{x}$ be any local minimizer of problem (3.2) with $\mathcal{I}_{\bar{x}}$, $\mathcal{J}_{\bar{x}}$, and $\mathcal{K}_{\bar{x}}$ given in (3.1). For convenience, we denote $\mathcal{I}_{\bar{x}}$, $\mathcal{J}_{\bar{x}}$, $\mathcal{K}_{\bar{x}}$ as $\mathcal{I}$, $\mathcal{J}$, $\mathcal{K}$ in this proof. We prove that $\bar{x}$ is a local minimizer of problem (1.1) by dividing the proof into two parts. The first one is the easy case where $\mathcal{K} = \emptyset$ and the second one deals with the complicated case where $\mathcal{K} \neq \emptyset$.

Part 1: $\mathcal{K} = \emptyset$. In this case, $\bar{x}$ is a local minimizer of problem

$$\begin{aligned} \min_x \quad & \|(b - Ax)_{\mathcal{J}}\|_q^q + h(x) \\ \text{s.t.} \quad & x \in \mathcal{X}. \end{aligned}$$

By the definition, $\bar{x}$ is a local minimizer of problem (1.1).

Part 2: $\mathcal{K} \neq \emptyset$. Consider the feasible direction cone $\mathcal{D}_{\bar{x}}$ of problem (1.1) at point $\bar{x}$, i.e.,

$$\mathcal{D}_{\bar{x}} = \{d \mid \bar{x} + \alpha d \in \mathcal{X} \text{ for some } \alpha > 0\}.$$

For simplicity, we use $\mathcal{D}$ to denote $\mathcal{D}_{\bar{x}}$ in the subsequent proof. For any subset $\mathcal{K}_p$ of $\mathcal{K}$ indexed by $p = 1, 2, \ldots, P := 2^{|\mathcal{K}|}$, let $\mathcal{K}_p^c = \mathcal{K} \setminus \mathcal{K}_p$, and define

$$\mathcal{D}_p = \left\{ d \mid (Ad)_{\mathcal{K}_p} \le 0, \ (Ad)_{\mathcal{K}_p^c} \ge 0 \right\} \bigcap \mathcal{D}.$$

By the Minkowski–Weyl Theorem [5, Proposition 3.2.1], there exist $d_p^1, d_p^2, \ldots, d_p^{g_p} \in \mathcal{D}_p$, such that

$$\mathcal{D}_p = \text{Cone} \left\{ d_p^1, d_p^2, \ldots, d_p^{g_p} \right\},$$

and thus

$$\mathcal{D} = \bigcup_{p=1}^{P} \text{Cone} \left\{ d_p^1, d_p^2, \ldots, d_p^{g_p} \right\}.$$

Without loss of generality, assume $\|d_p^j\| = 1$ for all $j = 1, 2, \ldots, g_p$, $p = 1, 2, \ldots, P$. For any $d \in \bigcup_{p=1}^{P} \left\{ d_p^1, d_p^2, \ldots, d_p^{g_p} \right\} \subseteq \mathcal{D}$, define

$$\overleftarrow{\mathcal{K}}^d = \{ m \in \mathcal{K} \mid (Ad)_m < 0 \}. \tag{6.6}$$

Next, we consider the two cases where $\overleftarrow{\mathcal{K}}^d$ is nonempty and empty, respectively. The former happens when d is not a feasible direction of problem (3.2) at point $\bar{x}$; while the latter happens when d is a feasible direction of problem (3.2) at point $\bar{x}$.

Case 1: $\overleftarrow{\mathcal{K}}^d \neq \emptyset$. Since $d \in \mathcal{D}$, there must exist ϵ_0^d so that $\bar{x} + \epsilon d \in \mathcal{X}$ holds for all $0 \leq \epsilon \leq \epsilon_0^d$. Define

$$\overrightarrow{\mathcal{J}}^d = \{ m \in \mathcal{J} \mid (Ad)_m > 0 \}. \tag{6.7}$$

Choose ϵ_1^d small enough such that

$$\left(b - A\bar{x} - \epsilon_1^d Ad \right)_m \leq 0, \quad \forall\, m \in \mathcal{I} \tag{6.8}$$

and

$$\left(b - A\bar{x} - \epsilon_1^d Ad \right)_m \geq \frac{(b - A\bar{x})_m}{2} > 0, \quad \forall\, m \in \overrightarrow{\mathcal{J}}^d. \tag{6.9}$$

Therefore, for $0 \leq \epsilon \leq \min \left\{ \epsilon_0^d, \epsilon_1^d \right\}$, we obtain

$$
\begin{aligned}
f(\bar{x} + \epsilon d) &- f(\bar{x}) \\
&= \| \max \{ b - A(\bar{x} + \epsilon d), 0 \} \|_q^q - \| \max \{ b - A\bar{x}, 0 \} \|_q^q \\
&= \sum_{m \in \overleftarrow{\mathcal{K}}^d \cup \mathcal{J}} (b - A\bar{x} - \epsilon Ad)_m^q - \sum_{m \in \mathcal{J}} (b - A\bar{x})_m^q \tag{6.10} \\
&\geq \sum_{m \in \overleftarrow{\mathcal{K}}^d} (-Ad)_m^q \epsilon^q + \sum_{m \in \overrightarrow{\mathcal{J}}^d} \left((b - A\bar{x} - \epsilon Ad)_m^q - (b - A\bar{x})_m^q \right) \tag{6.11} \\
&\geq \sum_{m \in \overleftarrow{\mathcal{K}}^d} (-Ad)_m^q \epsilon^q + \sum_{m \in \overrightarrow{\mathcal{J}}^d} q(b - A\bar{x} - \epsilon Ad)_m^{q-1} (-\epsilon(Ad)_m) \tag{6.12} \\
&\geq \sum_{m \in \overleftarrow{\mathcal{K}}^d} (-Ad)_m^q \epsilon^q + \sum_{m \in \overrightarrow{\mathcal{J}}^d} q \left(\frac{(b - A\bar{x})_m}{2} \right)^{q-1} (-\epsilon(Ad)_m), \tag{6.13}
\end{aligned}
$$

where (6.10) is due to (3.1), (6.6), and (6.8); (6.11) is due to (6.7); (6.12) is due to the concavity of the function z^q with respect to $z > 0$; (6.13) is due to (6.9) and the definition of $\overrightarrow{\mathcal{J}}^d$ in (6.7). Moreover, by (1.3) and the Taylor's expansion, for any $0 \leq \epsilon \leq 1$, there exists $\xi \in (0, 1)$ such that

$$
\begin{aligned}
h(\bar{x} + \epsilon d) - h(\bar{x}) &= \epsilon \nabla h(\bar{x} + \xi \epsilon d)^T d \\
&\geq -\epsilon \|\nabla h(\bar{x} + \xi \epsilon d)\| \\
&\geq -\epsilon \left(\|\nabla h(\bar{x})\| + \epsilon L_h \right) \\
&\geq -\epsilon \left(\|\nabla h(\bar{x})\| + L_h \right).
\end{aligned} \tag{6.14}
$$

Combining (6.13) with (6.14), for any $0 \leq \epsilon \leq \min \left\{ \epsilon_0^d, \epsilon_1^d, 1 \right\}$, we obtain

$$
F(\bar{x} + \epsilon d) - F(\bar{x}) \geq \lambda_1^d \epsilon^q - \lambda_2^d \epsilon,
$$

where

$$
\lambda_1^d := \sum_{m \in \overleftarrow{\mathcal{K}}^d} (-Ad)_m^q \epsilon^q > 0,
$$

$$
\lambda_2^d := \sum_{m \in \overrightarrow{\mathcal{J}}^d} q \left(\frac{(b - A\bar{x})_m}{2} \right)^{q-1} (Ad)_m + \|\nabla h(\bar{x})\| + L_h > 0.
$$

Define

$$
\epsilon_2^d := \left(\frac{\lambda_1^d}{\lambda_2^d} \right)^{\frac{1}{1-q}}
$$

and

$$
\bar{\epsilon}^d := \min \left\{ \epsilon_0^d, \epsilon_1^d, \epsilon_2^d, 1 \right\} > 0.
$$

From the above analysis, we can conclude that, for any $d \in \bigcup_{p=1}^P \left\{ d_p^1, d_p^2, \ldots, d_p^{g_p} \right\}$ with $\overleftarrow{\mathcal{K}}^d \neq \emptyset$, $F(\bar{x} + \epsilon d) \geq F(\bar{x})$ holds for all $\epsilon \in [0, \bar{\epsilon}^d]$.

Case 2: $\overleftarrow{\mathcal{K}}^d = \emptyset$. Recall the definition of $\overleftarrow{\mathcal{K}}^d$ (cf. (6.6)). $\overleftarrow{\mathcal{K}}^d = \emptyset$ implies that d is a feasible direction of problem (3.2) at point $\bar{x}$. From the assumption that $\bar{x}$ is a local minimizer of problem (3.2), we know that there exists an $\tilde{\epsilon} > 0$ such that for all $d \in \bigcup_{p=1}^P \left\{ d_p^1, d_p^2, \ldots, d_p^{g_p} \right\}$ with $\overleftarrow{\mathcal{K}}^d = \emptyset$, there holds $F(\bar{x} + \epsilon d) \geq F(\bar{x})$ for all $\epsilon \in [0, \tilde{\epsilon}]$.

We now combine the above two cases: **Case 1** and **Case 2**. Since there are finitely many directions $\bigcup_{p=1}^{P} \left\{ d_p^1, d_p^2, \ldots, d_p^{g_p} \right\}$, it follows that

$$\bar{\epsilon} := \min \left\{ \min_{\mathcal{K}^d \neq \emptyset, \, j=1,\ldots,g_p, \, p=1,\ldots,P} \left\{ \bar{\epsilon}^{d_p^j} \right\}, \tilde{\epsilon} \right\} > 0$$

and

$$\bar{x} + \epsilon d_p^j \in \mathcal{X}, \ F(\bar{x} + \epsilon d_p^j) \geq F(\bar{x}), \ \forall \ j = 1, 2, \ldots, g_p, \ p = 1, 2, \ldots, P \tag{6.15}$$

hold true for all $\epsilon \in [0, \bar{\epsilon}]$.

Let $\mathrm{Conv}_p(\bar{x}, \bar{\epsilon})$ denote the convex hull spanned by points $\bar{x}$ and $\bar{x} + \bar{\epsilon} d_p^j$, $j = 1, 2, \ldots, g_p$. Then, for any $x \in \bigcup_{p=1}^{P} \mathrm{Conv}_p(\bar{x}, \bar{\epsilon})$, we have $F(x) \geq F(\bar{x})$ by (6.15) and the fact that $f(x)$ is concave in $\mathrm{Conv}_p(\bar{x}, \bar{\epsilon})$. Furthermore, one can always choose a sufficiently small but fixed $\epsilon > 0$ such that $B(\bar{x}, \epsilon) \cap \mathcal{X} \subseteq \bigcup_{p=1}^{P} \mathrm{Conv}_p(\bar{x}, \bar{\epsilon})$. Therefore, $\bar{x}$ is a local minimizer of problem (1.1). $\qquad \square$

Appendix 3: Proof of Lemma 4.1

We show the three items of Lemma 4.1 separately.

(i) of Lemma 4.1: it follows directly from the inequality

$$\theta^q(t) \leq \theta^q(t, \mu) \leq \left(\theta(t) + \frac{\mu}{2} \right)^q \leq \theta^q(t) + \left(\frac{\mu}{2} \right)^q, \quad \forall \, t \leq \mu.$$

(ii) of Lemma 4.1: When $t \neq 0$ and $t \neq \mu$, $\theta^q(t, \mu)$ is twice continuously differentiable with respect to t. Recall $\theta^q(t, \mu) \geq (\mu/2)^q$ for all t (cf. (4.2)). Then it follows from (4.4) that

$$\left| \left[\theta^q(t, \mu) \right]'' \right| \leq 4q\mu^{q-2}, \quad \forall \, t \notin \{0, \mu\}.$$

This further implies (ii) of Lemma 4.1.

(iii) of Lemma 4.1: By the mean-value theorem [27, Theorem 2.3.7], we have

$$\theta^q(t, \mu) = \theta^q(\hat{t}, \mu) + \left[\theta^q(\hat{t}, \mu) \right]' (t - \hat{t}) + \frac{\upsilon}{2} (t - \hat{t})^2, \tag{6.16}$$

where $\upsilon \in \partial_t \left(\left[\theta^q(\xi \hat{t} + (1 - \xi)t, \mu) \right]' \right)$ and $\xi \in [0, 1]$. We consider the following three cases.

– Case $\hat{t} > 2\mu$: Since $t - \hat{t} \geq -\hat{t}/2$, it follows for any $\xi \in [0, 1]$ that

$$\xi t + (1 - \xi)\hat{t} = \hat{t} + \xi(t - \hat{t}) \geq \hat{t}/2 > \mu.$$

This, together with (4.4) and (4.5), implies that the υ in (6.16) satisfies

$$\upsilon \leq 0 = \kappa(\hat{t}, \mu).$$

From this and (6.16), we obtain (4.6).

– Case $\hat{t} \in [-\mu, 2\mu]$: From (ii) of Lemma 4.1, $|\upsilon|$ is uniformly bounded by $\kappa(\hat{t}, \mu) = 4q\mu^{q-2}$. Combining this with (6.16) yields (4.6).
– Case $\hat{t} < -\mu$: Since $t - \hat{t} \leq \mu$, for any $\xi \in [0, 1]$, it follows

$$\xi t + (1 - \xi)\hat{t} = \hat{t} + \xi(t - \hat{t}) < 0.$$

From this, (4.4), (4.5), and (6.16), we can obtain (4.6).

This completes the proof of Lemma 4.1.   □

References

1. Barzilai, J., Borwein, J.M.: Two-point step size gradient methods. IMA J. Numer. Anal. **8**(1), 141–148 (1988)
2. Beck, A., Teboulle, M.: A fast iterative shrinkage–thresholding algorithm for linear inverse problems. SIAM J. Imaging Sci. **2**(1), 183–202 (2009)
3. Ben-Tal, A., Nemirovski, A.: Lectures on Modern Convex Optimization. MPS–SIAM Series on Optimization. SIAM, Philadelphia (2001)
4. Berg, E.V.D., Friedlander, M.P.: Probing the pareto frontier for basis pursuit solutions. SIAM J. Sci. Comput. **31**(2), 890–912 (2008)
5. Bertsekas, D.P.: Convex Analyis and Optimization. Athena Scientific, Massachusetts (2003)
6. Bian, W., Chen, X.: Worst-case complexity of smoothing quadratic regularization methods for non-Lipschitzian optimization. SIAM J. Optim. **23**(3), 1718–1741 (2013)
7. Bian, W., Chen, X.: Smoothing quadratic regularization methods for box constrained non-Lipschitz optimization in image restoration. Technical report, Hong Kong Polytechnic University (2014)
8. Bian, W., Chen, X., Ye, Y.: Complexity analysis of interior point algorithms for non-Lipschitz and nonconvex minimization. Math. Program. **149**(1–2), 301–327 (2015)
9. Birbil, S.I., Fang, S.C., Frenk, J.B.G., Zhang, S.: Recursive approximation of the high dimensional max function. Oper. Res. Lett. **33**(5), 450–458 (2005)
10. Birgin, E.G., Martínez, J.M., Raydan, M.: Nonmonotone spectral projected gradient methods on convex sets. SIAM J. Optim. **10**(4), 1196–1211 (2000)
11. Boser, B.E., Guyon, I., Vapnik, V.: A training algorithm for optimal margin classifiers. In: The Fifth Annual Workshop of Computational Learning Theory, pp. 144–152 (1992)
12. Bruckstein, A.M., Donoho, D.L., Elad, M.: From sparse solutions of systems of equations to sparse modeling of signals and images. SIAM Rev. **51**(1), 34–81 (2009)
13. Candès, E.J., Tao, T.: Decoding by linear programming. IEEE Trans. Inf. Theory **51**(12), 4203–4215 (2005)
14. Candès, E.J., Wakin, M.B., Boyd, S.P.: Enhancing sparsity by reweighted ℓ_1 minimization. J. Fourier Anal. Appl. **14**(5), 877–905 (2008)
15. Cartis, C., Gould, N.I.M., Toint, P.L.: Adaptive cubic regularisation methods for unconstrained optimization. Part I: motivation, convergence and numerical results. Math. Program. **127**(2), 245–295 (2011)
16. Cartis, C., Gould, N.I.M., Toint, P.L.: Adaptive cubic regularisation methods for unconstrained optimization. Part II: worst-case function- and derivative-evaluation complexity. Math. Program. **130**(2), 295–319 (2011)
17. Cartis, C., Gould, N.I.M., Toint, P.L.: On the evaluation complexity of composite function minimization with applications to nonconvex nonlinear programming. SIAM J. Optim. **21**(4), 1721–1739 (2011)

18. Chartrand, R.: Exact reconstruction of sparse signals via nonconvex minimization. IEEE Signal Process. Lett. **14**(10), 707–710 (2007)
19. Chartrand, R., Staneva, V.: Restricted isometry properties and nonconvex compressive sensing. Inverse Probl. **24**(3), 1–14 (2008)
20. Chartrand, R., Yin, W.: Iteratively reweighted algorithms for compressive sensing. In: Internal Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 3869–3872 (2008)
21. Chen, X.: Smoothing methods for nonsmooth, nonconvex minimization. Math. Program. **134**(1), 71–99 (2012)
22. Chen, X., Ge, D., Wang, Z., Ye, Y.: Complexity of unconstrained l_2-l_p minimization. Math. Program. **143**(1–2), 371–383 (2014)
23. Chen, X., Ng, M.K., Zhang, C.: Non-Lipschitz l_p-regularization and box constrained model for image restoration. IEEE Trans. Image Process. **21**(12), 4709–4721 (2012)
24. Chen, X., Niu, L., Yuan, Y.: Optimality conditions and a smoothing trust region newton method for nonLipschitz optimization. SIAM J. Optim. **23**(3), 1528–1552 (2013)
25. Chen, X., Xu, F., Ye, Y.: Lower bound theory of nonzero entries in solutions of ℓ_2-ℓ_p minimization. SIAM J. Sci. Comput. **32**(5), 2832–2852 (2010)
26. Chen, X., Zhou, W.: Smoothing nonlinear conjugate gradient method for image restoration using nonsmooth nonconvex minimization. SIAM J. Imaging Sci. **3**(4), 765–790 (2010)
27. Clarke, F.H.: Optimization and Nonsmooth Analysis. John Wiley, New York (1983)
28. Cortes, C., Vapnik, V.: Support-vector networks. Mach. Learn. **20**(3), 273–297 (1995)
29. Dai, Y.H., Liao, L.Z.: R-linear convergence of the Barzilai and Borwein gradient method. IMA J. Numer. Anal. **22**(1), 1–10 (2002)
30. Daubechies, I., DeVore, R., Fornasier, M., Güntürk, C.S.: Iteratively reweighted least squares minimization for sparse recovery. Commun. Pure Appl. Math. **63**(1), 1–38 (2010)
31. Fan, J., Li, R.: Variable selection via nonconcave penalized likelihood and its oracle properties. J. Am. Stat. Assoc. **96**(456), 1348–1359 (2001)
32. Foucart, S., Lai, M.J.: Sparsest solutions of underdetermined linear systems via ℓ_q-minimization for $0 < q \leq 1$. Appl. Comput. Harmon. Anal. **26**(3), 395–407 (2009)
33. Foucart, S., Rauhut, H.: A Mathematical Introduction to Compressive Sensing. Springer, New York (2013)
34. Garey, M.R., Johnson, D.S.: Computers and Intractability: A Guide to the Theory of NP-Completeness. W. H. Freeman, New York (1979)
35. Garmanjani, R., Vicente, L.N.: Smoothing and worst case complexity for direct-search methods in non-smooth optimization. IMA J. Numer. Anal. **33**(3), 1008–1028 (2013)
36. Ge, D., Jiang, X., Ye, Y.: A note on the complexity of l_p minimization. Math. Program. **129**(2), 285–299 (2011)
37. Ghadimi, S., Lan, G.: Accelerated gradient methods for nonconvex nonlinear and stochastic programming. Math. Progam. (2015). doi:10.1007/s10107-015-0871-8
38. Gould, N.I.M., Toint, P.L.: Preprocessing for quadratic programming. Math. Program. **100**(1), 95–132 (2004)
39. Hale, E.T., Yin, W., Zhang, Y.: Fixed-point continuation applied to compressed sensing: implemetation and numerical experiments. J. Comput. Math. **28**(2), 170–194 (2010)
40. Huang, J., Ma, S., Xie, H., Zhang, C.H.: A group bridge approach for variable selection. Biometrika **96**(2), 339–355 (2009)
41. Ji, S., Sze, K.F., Zhou, Z., So, A.M.C., Ye, Y.: Beyond convex relaxation: A polynomial-time non-convex optimization approach to network localization. In: IEEE Conference on Computer Communications (INFOCOM), pp. 2499–2507 (2013)
42. Jiang, B., Dai, Y.H.: A framework of constraint preserving update schemes for optimization on stiefel manifold. Math. Program. (2015). doi:10.1007/s10107-014-0816-7
43. Jiang, B., Zhang, S.: Iteration bounds for finding ϵ-stationary points of structured nonconvex optimization. Technical report, University of Minnesota (2014)
44. Lai, M.J., Wang, J.: An unconstrained ℓ_q minimization with $0 < q \leq 1$ for sparse solution of underdetermined linear systems. SIAM J. Optim. **21**(1), 82–101 (2011)
45. Lai, M.J., Xu, Y., Yin, W.: Improved iteratively reweighted least squares for unconstrained smoothed l_q minimization. SIAM J. Numer. Anal. **51**(2), 927–957 (2013)
46. Liu, Y.F., Dai, Y.H., Luo, Z.Q.: Joint power and admission control via linear programming deflation. IEEE Trans. Signal Process. **61**(6), 1327–1338 (2013)

47. Liu, Y.F., Dai, Y.H., Ma, S.: Joint power and admission control: non-convex l_q approximation and an effective polynomial time deflation approach. IEEE Trans. Signal Process. **63**(14), 3641–3656 (2015)
48. Lu, Z.: Iterative reweighted minimization methods for l_p regularized unconstrained nonlinear programming. Math. Program. **147**(1–2), 277–307 (2014)
49. Mitliagkas, I., Sidiropoulos, N.D., Swami, A.: Joint power and admission control for ad-hoc and cognitive underlay networks: convex approximation and distributed implementation. IEEE Trans. Wireless Commun. **10**(12), 4110–4121 (2011)
50. Mourad, N., Reilly, J.P.: Minimizing nonconvex functions for sparse vector reconstruction. IEEE Trans. Signal Process. **58**(7), 3485–3496 (2010)
51. Nesterov, Y.: A method of solving a convex programming problem with convergence rate O($1/k^2$). Sov. Math. Dokl. **27**(2), 372–376 (1983)
52. Nesterov, Y.: Smooth minimization of non-smooth functions. Math. Program. **103**(1), 127–152 (2005)
53. Nikolova, M., Ng, M.K., Zhang, S., Ching, W.K.: Efficient reconstruction of piecewise constant images using nonsmooth nonconvex minimization. SIAM J. Imaging Sci. **1**(1), 2–25 (2008)
54. Nocedal, J., Wright, S.J.: Numerical Optimization, 2nd edn. Springer, New York (2006)
55. Papadimitriou, C.H.: Computational Complexity. Addison-Wesley, Massachusetts (1994)
56. Rao, B.D., Kreutz-delgado, K.: An affine scaling methodology for best basis selection. IEEE Trans. Signal Process. **47**(1), 187–200 (1999)
57. Sun, Q.: Recovery of sparsest signals via ℓ_q-minimization. Appl. Comput. Harmon. Anal. **32**(3), 329–341 (2012)
58. Sun, W., Yuan, Y.: Optimization Theory and Methods: Nonlinear Programming. Springer, New York (2006)
59. Vazirani, V.V.: Approximation Algorithms. Springer, New York (2001)
60. Wagner, M., Meller, J., Elber, R.: Large-scale linear programming techniques for the design of protein folding potentials. Math. Program. **101**(2), 301–318 (2004)
61. Wright, S.J., Nowak, R.D., Figueiredo, M.A.T.: Sparse reconstruction by separable approximation. IEEE Trans. Signal Process. **57**(7), 2479–2493 (2009)
62. Ye, Y.: Interior Point Algorithms-Theory and Analysis. Wiley, New York (1997)
63. Yun, S., Toh, K.C.: A coordinate gradient descent method for l_1-regularized convex minimization. Comput. Optim. Appl. **48**(2), 273–307 (2011)