Mathematical Programming manuscript No.
(will be inserted by the editor)

Optimality Conditions and Complexity for Non-Lipschitz Constrained Optimization Problems

Wei Bian · Xiaojun Chen

October 26, 2014, Received: date / Accepted: date

Abstract In this paper, we consider a class of nonsmooth, nonconvex constrained optimization problems where the objective function may be not Lipschitz continuous and the feasible set is a general closed convex set. Using the theory of the generalized directional derivative and the Clarke tangent cone, we derive a first order necessary optimality condition for local minimizers of the problem, and define the generalized stationary point of it. The generalized stationary point is the Clarke stationary point when the objective function is Lipschitz continuous at this point, and the scaled stationary point in the existing literature when the objective function is not Lipschitz continuous at this point. We prove the consistency between the generalized directional derivative and the limit of the classic directional derivatives associated with the smoothing function. Moreover we present the computational complexity and lower bound theory of the problem.

Keywords Constrained nonsmooth nonconvex optimization · generalized directional derivative · necessary optimality condition · lower bound theory · consistency

Mathematics Subject Classification (2000) 90C30 · 90C26 · 65K05 · 49M37

The author's work was supported by the NSF foundation (11101107,11471088) of China, HIT.BRETIII.201414, PIRS of HIT No.A201402, and partly by Hong Kong Research Grant Council grant (PolyU5001/12P).

Wei Bian
Department of Mathematics, Harbin Institute of Technology, Harbin, China. E-mail: bianweilvse520@163.com

Xiaojun Chen
Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China. E-mail: maxjchen@polyu.edu.hk

1 Introduction

In this paper, we consider the following constrained optimization problem

$$\min_{x\in\mathcal{X}} \quad f(x) := \Theta(x) + c(h(x)), \tag{1}$$

where $\Theta : R^n \to R$ and $c : R^m \to R$ are continuously differentiable, $h : R^n \to R^m$ is continuous, and $\mathcal{X} \subset R^n$ is a nonempty closed convex set. Of particular interest of this paper is when h is not convex, not differentiable, or even not Lipschitz continuous. Problem (1) includes many problems in practice [5,6,15,16,21,25,27,38]. For instance, the following minimization problem

$$\min_{l\le x\le u, Ax\le b} \quad \Theta(x) + \sum_{i=1}^{m} \varphi(\|D_i^T x\|_p^p) \tag{2}$$

is a special case of (1), where $l \in \{R\cup -\infty\}^n$, $u \in \{R\cup\infty\}^n$, $A \in R^{t\times n}$, $b \in R^t$, $D_i \in R^{n\times r}$, $p \in (0,1)$ and $\varphi : R_+ \to R_+$ is continuous. Such problem arises from image restoration [11,15,16,32], signal processing [8], variable selection [21], etc. Another special case of (1) is the following problem

$$\min_{x\in\mathcal{X}} \quad \Theta(x) + \sum_{i=1}^{m} \max\{\alpha_i - m_i^T x, 0\}^p, \tag{3}$$

where $\alpha_i \in R$ and $m_i \in R^n$, which has attracted much interest in machine learning, wireless communication [31], information theory, data analysis [22, 26], etc. Moreover, a number of constrained optimization problems can be reformulated as problem (1) by using the exact penalty method with nonsmooth or non-Lipschitz continuous penalty functions [3].

When $\mathcal{X} = R^n$ and $c(h(x)) = \|x\|_p^p$ $(0 < p < 1)$, the affine scaled first and second order necessary conditions for local minimizers of (1) are established in [17]. By using subspace techniques, Chen, et al [15] extended the first and second order necessary conditions to $c(h(x)) = \|Dx\|_p^p$ with $D \in R^{m\times n}$. However, the optimality conditions in [15,17] are weaker than the Clarke optimality conditions [18] for $p = 1$, and not applicable to constrained optimization problems. In this paper, we will derive a necessary optimality condition for the non-Lipschitz constrained optimization problem (1), which reduces to the Clarke optimality condition when the objective function in (1) is locally Lipschitz continuous.

A point x^* is called a Clarke stationary point of f if f is locally Lipschitz at x^* and there is $V \in \partial f(x^*)$ such that

$$(V, x - x^*) \ge 0, \quad \forall x \in X, \tag{4}$$

where

$$\partial f(x) = \mathrm{con}\{v \,|\, \nabla f(y) \to v,\ f \text{ is differentiable at } y,\ y \to x\}$$

is the Clarke subdifferential and "con" denotes the convex hull. From Theorem 9.61 and (b) of Corollary 8.47 in the book of Rockafellar and Wets [35], the subdifferential associated with a smoothing function

$$G_{\tilde{f}}(x) = \mathrm{con}\{v \mid \nabla_x \tilde{f}(x^k, \mu_k) \to v, \text{ for } x^k \to x, \ \mu_k \downarrow 0\},$$

is nonempty and bounded, and $\partial f(x) \subseteq G_{\tilde{f}}(x)$. In [9,10,13,35], it is shown that many smoothing functions satisfy the gradient consistency

$$\partial f(x) = G_{\tilde{f}}(x). \tag{5}$$

The gradient consistency is an important property of the smoothing methods, which guarantees the convergence of smoothing methods with adaptive updating schemes of smoothing parameters to a stationary point of the original problem.

Due to the non-Lipschitz continuity of the objective function f, Clarke optimality condition (4) cannot be applied to (1). In [28], Jahn introduced a directional derivative for Lipschitz constrained optimization problems

$$f^{\circ}(\bar{x}; v) = \limsup_{\substack{y \to \bar{x},\, y \in \mathcal{X} \\ t \downarrow 0,\, y + tv \in \mathcal{X}}} \frac{f(y + tv) - f(y)}{t},$$

which is equal to the Clarke generalized directional derivative at the interior points of $\mathcal{X}$. In this paper, we extend the directional derivative in [28] to the non-Lipschitz constrained optimization problem (1). Using the extended directional derivative and the Clarke tangent cone, we derive necessary optimality conditions. The new optimality conditions are equivalent to the optimality conditions in [6,15,17] when the objective function is not Lipschitz continuous, and to the Clarke optimality condition (4) when the objective function is Lipschitz continuous. Moreover, we establish the consistency between the generalized directional derivative and the limit of the classic directional derivatives associated with the smoothing function. The directional derivative consistency guarantees the convergence of smoothing methods to a generalized stationary point of (1).

Problem (1) includes the regularized minimization problem as a special case when $\Theta(x)$ is a data fitting term and $c(h(x))$ is a regularization term (also called a penalty term in some articles). In sparse optimization, nonconvex non-Lipschitz regularization provides more efficient models to extract the essential features of solutions than the convex regularization [5,12,13,16,21, 25,27,32,38]. The SCAD penalty function [21] and the MCP function [38] have various desirable properties in variable selection. Logistic and fraction penalty functions yield edge preservation in image restoration [16,32]. The l_p norm penalty function with $0 < p < 1$ owns the oracle property in statistics [21, 29]. Moreover, the lower bound theory of the l_2-l_p regularized minimization problem [16,17], a special case of (1), states that the absolute value of each component of any local minimizer of the problem is either zero or greater than a positive constant. The lower bound theory not only helps us to distinguish zero and nonzero entries of coefficients in sparse high-dimensional approximation [12, 25], but also brings the restored image closed contours and neat edges [16]. In this paper, we extend the lower bound theory of the l_2-l_p regularization minimization problem to problems (2) and (3) with $0 < p \leq 1$ which include the most widely used models in statistics and sparse reconstruction. Moreover, we extend the complexity results of the l_2-l_p regularization minimization problem [14] to problem (2) with a concave function φ and $0 < p \leq 1$. Such extension of the lower bound theory and complexity is not trivial because of the general constraints and weak conditions on φ.

The rest of the paper is organized as follows. In section 2, we first define a generalized directional derivative and present its properties. Next, we derive necessary optimality conditions for a local minimizer of problem (1), and prove the directional derivative consistency associated with smoothing functions. In section 3, we present the computational complexity and the lower bound theory of problem (2).

In our notation, $R_+ = [0, \infty)$ and $R_{++} = (0, \infty)$. For $x \in R^n$, $0 < p < \infty$ and $\delta > 0$, $\|x\|_p^p = \sum_{i=1}^n |x_i|^p$, $B_\delta(x)$ means the open ball centered at x with radius δ. For a closed convex subset $\Omega \subseteq R^n$, $\mathrm{int}(\Omega)$ means the interior of Ω, $\mathrm{cl}(\Omega)$ means the closure of Ω and $\mathrm{m}(\Omega)$ denotes the element in Ω with the smallest Euclidean norm. $P_{\mathcal{X}}[x] = \arg\min\{\|z - x\|_2 : z \in \mathcal{X}\}$ denotes the orthogonal projection from R^n to $\mathcal{X}$. $\mathbb{N}_{++} = \{1, 2, \ldots\}$.

2 Optimality conditions

Inspired by the generalized directional derivative and the tangent cone, we present a first order necessary optimality condition for local minimizers of the constrained optimization problem (1), which is equivalent to the Clarke necessary condition for locally Lipschitz optimization problems and stronger than the necessary optimality conditions for the non-Lipschitz optimization problems in the existing literature. At the end of this section, we prove the directional directive consistency associated with smoothing functions

We suppose the function h in (1) has the following version

$$h(x) := (h_1(D_1^T x), h_2(D_2^T x), \ldots, h_m(D_m^T x))^T \tag{6}$$

where $D_i \in R^{n \times r}$, $h_i (i = 1, \ldots, m) : R^r \to R$ is continuous, but not necessarily Lipschitz continuous.

2.1 Generalized directional derivative

Definition 1 A function $\phi : R^n \to R$ is said to be Lipschitz continuous at(near) $x \in R^n$ if there exist positive numbers L_x and δ such that

$$|\phi(y) - \phi(z)| \leq L_x \|y - z\|_2, \quad \forall\, y, z \in B_\delta(x).$$

Otherwise, ϕ is said to be not Lipschitz continuous at(near) $x \in R^n$.

For a fixed $\bar{x} \in R^n$, denote

$$\mathcal{I}_{\bar{x}} = \{i \in \{1, 2, \ldots, m\} : h_i \text{ is not Lipschitz continuous at } D_i^T \bar{x}\}, \tag{7}$$

$$\mathcal{V}_{\bar{x}} = \{v : D_i^T v = 0,\ i \in \mathcal{I}_{\bar{x}}\}, \tag{8}$$

and define

$$h_i^{\bar{x}}(D_i^T x) = \begin{cases} h_i(D_i^T x) & i \notin \mathcal{I}_{\bar{x}} \\ h_i(D_i^T \bar{x}) & i \in \mathcal{I}_{\bar{x}}, \end{cases}$$

which is Lipschitz continuous at $D_i^T \bar{x}$, $i = 1, 2, \ldots, m$. Specially, we let $\mathcal{V}_{\bar{x}} = R^n$ when $\mathcal{I}_{\bar{x}} = \emptyset$. And then we let

$$f_{\bar{x}}(x) = \Theta(x) + c(h_{\bar{x}}(x)), \tag{9}$$

with $h_{\bar{x}}(x) := (h_1^{\bar{x}}(D_1^T x), h_2^{\bar{x}}(D_2^T x), \ldots, h_m^{\bar{x}}(D_m^T x))^T$.

The function $f_{\bar{x}}(x)$ is Lipschitz continuous at $\bar{x}$ and $f_{\bar{x}}(\bar{x}) = f(\bar{x})$. The generalized directional derivative [18] of $f_{\bar{x}}$ at $\bar{x}$ in the direction $v \in R^n$ is defined as

$$f_{\bar{x}}^{\circ}(\bar{x}; v) = \limsup_{y \to \bar{x}, t \downarrow 0} \frac{f_{\bar{x}}(y + tv) - f_{\bar{x}}(y)}{t}. \tag{10}$$

Specially, when f is regular,

$$f_{\bar{x}}^{\circ}(\bar{x}; v) = f_{\bar{x}}'(\bar{x}; v) = \lim_{t \downarrow 0} \frac{f_{\bar{x}}(\bar{x} + tv) - f_{\bar{x}}(\bar{x})}{t}.$$

The generalized directional derivative in (10) is generalized in [28] and used in [2,28] for locally Lipschitz constrained optimization. The generalization motives us to use the following generalized directional derivative of $f_{\bar{x}}$ at $\bar{x} \in \mathcal{X}$ in the direction $v \in R^n$

$$f_{\bar{x}}^{\circ}(\bar{x}; v) = \limsup_{\substack{y \to \bar{x},\, y \in \mathcal{X} \\ t \downarrow 0,\, y + tv \in \mathcal{X}}} \frac{f_{\bar{x}}(y + tv) - f_{\bar{x}}(y)}{t}. \tag{11}$$

The definitions in (10) and (11) coincide when $\bar{x} \in \text{int}(\mathcal{X})$.

Proposition 1 *For any $\bar{x} \in \mathcal{X}$ and $v \in \mathcal{V}_{\bar{x}}$,*

$$f^{\circ}(\bar{x}; v) = \limsup_{\substack{y \to \bar{x},\, y \in \mathcal{X} \\ t \downarrow 0,\, y + tv \in \mathcal{X}}} \frac{f(y + tv) - f(y)}{t} \quad \textit{exists} \tag{12}$$

and equals to $f_{\bar{x}}^{\circ}(\bar{x}; v)$ defined in (11).

Proof Fix $\bar{x} \in \mathcal{X}$ and $v \in \mathcal{V}_{\bar{x}}$. For $y \in R^n$ and $t > 0$, there exists $s \in (0, t)$ such that

$$\begin{aligned} c(h(y + tv)) - c(h(y)) &= \nabla c(z)_{z = h(y + sv)}^T (h(y + tv) - h(y)) \\ &= \nabla c(z)_{z = h(y + sv)}^T (h_{\bar{x}}(y + tv) - h_{\bar{x}}(y)). \end{aligned}$$

Then,

$$\frac{f(y+tv)-f(y)}{t}=\frac{\Theta(y+tv)-\Theta(y)+\nabla c(z)^T_{z=h(y+sv)}(h_{\bar{x}}(y+tv)-h_{\bar{x}}(y))}{t}.$$

By the Lipschitz continuity of Θ and $h_i^{\bar{x}}$ at $\bar{x}$, there exist $\delta>0$ and $L>0$ such that $\left|\frac{f(y+tv)-f(y)}{t}\right|\le L$, $\forall y\in B_\delta(\bar{x})$, $t\in(0,\delta)$. Thus, the generalized directional derivative of f at $\bar{x}\in\mathcal{X}$ in the direction $v\in\mathcal{V}_{\bar{x}}$ defined in (12) exists.

Let $\{y_n\}$ and $\{t_n\}$ be the sequences such that $y_n\in\mathcal{X}$, $t_n\downarrow 0$, $y_n\to\bar{x}$, $y_n+t_nv\in\mathcal{X}$ and the upper limit in (12) holds. Using the Lipschitz continuity of $h_i^{\bar{x}}$ at $\bar{x}$ again, we can get the subsequences $\{y_{n_k}\}\subseteq\{y_n\}$ and $\{t_{n_k}\}\subseteq\{t_n\}$ such that

$$\lim_{k\to\infty}\frac{h_{\bar{x}}(y_{n_k}+t_{n_k}v)-h_{\bar{x}}(y_{n_k})}{t_{n_k}}\quad\text{exists.}\tag{13}$$

By the above analysis, then

$$\begin{aligned}f^\circ(\bar{x};v)=&\lim_{k\to\infty}\frac{f(y_{n_k}+t_{n_k}v)-f(y_{n_k})}{t_{n_k}}\\=&\nabla\Theta(\bar{x})+\nabla c(z)_{z=h(\bar{x})}\lim_{k\to\infty}\frac{h_{\bar{x}}(y_{n_k}+t_{n_k}v)-h_{\bar{x}}(y_{n_k})}{t_{n_k}}.\end{aligned}\tag{14}$$

By virtue of (11), we have

$$\begin{aligned}f^\circ_{\bar{x}}(\bar{x};v)\ge&\lim_{k\to\infty}\frac{f_{\bar{x}}(y_{n_k}+t_{n_k}v)-f_{\bar{x}}(y_{n_k})}{t_{n_k}}\\=&\nabla\Theta(\bar{x})+\nabla c(z)_{z=h_{\bar{x}}(\bar{x})}\lim_{k\to\infty}\frac{h_{\bar{x}}(y_{n_k}+t_{n_k}v)-h_{\bar{x}}(y_{n_k})}{t_{n_k}}.\end{aligned}\tag{15}$$

Using $h(\bar{x})=h_{\bar{x}}(\bar{x})$, (14) and (15) we obtain $f^\circ_{\bar{x}}(\bar{x};v)\ge f^\circ(\bar{x};v)$.

On the other hand, by extracting the sequences $\{y_{n_k}\}$ and $\{t_{n_k}\}$ such that the upper limit in (11) holds and the limit in (13) exists with them, similar to the above analysis, we find that $f^\circ(\bar{x};v)\ge f^\circ_{\bar{x}}(\bar{x};v)$.

Therefore, $f^\circ(\bar{x};v)=f^\circ_{\bar{x}}(\bar{x};v)$.

Notice that the generalized directional derivative of f at $\bar{x}\in\mathcal{X}$ in the direction $v\in\mathcal{V}_{\bar{x}}$ defined in (12) involves only the behavior of f at $\bar{x}$ in the hyperplane $\mathcal{V}_{\bar{x}}$.

2.2 Clarke tangent cone

Since $\mathcal{X}$ is a nonempty closed convex subset of R^n, the distance function related to $\mathcal{X}$ is a nonsmooth, Lipschitz continuous function, defined by

$$d_{\mathcal{X}}(x)=\min\{\|x-y\|_2:y\in\mathcal{X}\}.$$

The Clarke tangent cone to $\mathcal{X}$ at $x\in\mathcal{X}$, denoted as $\mathcal{T}_{\mathcal{X}}(x)$, is defined by

$$\mathcal{T}_{\mathcal{X}}(x)=\{v\in R^n:d^\circ_{\mathcal{X}}(x;v)=0\}.$$

Assumption 1 *Assume that* $\mathcal{X} = \mathcal{X}_1 \cap \mathcal{X}_2$ *and* $\mathrm{int}(\mathcal{X}_1) \cap \mathcal{X}_2 \neq \emptyset$, *where* $\mathcal{X}_1 \subseteq R^n$ *is a nonempty closed convex set and* $\mathcal{X}_2 = \{x \mid Ax = b\}$ *with* $A \in R^{t\times n}, b \in R^t$.

Under Assumption 1, we can obtain the following properties of the Clarke tangent cones to $\mathcal{X}_1$, $\mathcal{X}_2$ and $\mathcal{X}$.

Lemma 1 *The following statements hold.*

(1) $\mathrm{int}(\mathcal{T}_{\mathcal{X}_1}(x)) \neq \emptyset$, $\forall x \in \mathcal{X}_1$;
(2) $\mathcal{T}_{\mathcal{X}_2}(x) = \mathrm{cl}\{\lambda(c - x) : c \in \mathcal{X}_2, \lambda \geq 0\}$, $\forall x \in \mathcal{X}_2$;
(3) $\mathcal{T}_{\mathcal{X}_1\cap\mathcal{X}_2}(x) = \mathcal{T}_{\mathcal{X}_1}(x) \cap \mathcal{T}_{\mathcal{X}_2}(x)$, $\forall x \in \mathcal{X}$.

Proof (1) Fix $x \in \mathcal{X}_1$ and denote $\hat{x} \in \mathrm{int}(\mathcal{X}_1)$. Let $\epsilon > 0$ be a constant such that $\hat{x} + B_\epsilon(0) \subseteq \mathrm{int}(\mathcal{X}_1)$. We shall show that $\hat{x} - x + B_\epsilon(0) \subseteq \mathcal{T}_{\mathcal{X}_1}(x)$, and hence $\hat{x} - x \in \mathrm{int}(\mathcal{T}_{\mathcal{X}_1}(x))$.

By the convexity of $\mathcal{X}_1$, $d_{\mathcal{X}_1}(x)$ is a convex function and for $v \in \hat{x} - x + B_\epsilon(0)$, we notice that

$$x + tv \in (1-t)x + t(\hat{x} + B_\epsilon(0)) \subseteq \mathcal{X}_1, \quad \forall x \in \mathcal{X}_1,\ 0 \leq t \leq 1.$$

Then,

$$d^\circ_{\mathcal{X}_1}(x; v) = d'_{\mathcal{X}_1}(x; v) = \lim_{\lambda \downarrow 0} \frac{d_{\mathcal{X}_1}(x + \lambda v) - d_{\mathcal{X}_1}(x)}{\lambda} = 0,$$

which confirms that $v \in \mathcal{T}_{\mathcal{X}_1}(x)$.

(2) Since $\mathcal{X}_2$ is defined by a class of affine equalities, we have $\mathcal{T}_{\mathcal{X}_2}(x) = \mathrm{cl}\{\lambda(c - x) : c \in \mathcal{X}_2, \lambda \geq 0\}$.

(3) By $\mathrm{int}(\mathcal{X}_1) \cap \mathcal{X}_2 \neq \emptyset$, $0 \in \mathrm{int}(\mathcal{X}_1 - \mathcal{X}_2)$, then $\mathcal{T}_{\mathcal{X}_1\cap\mathcal{X}_2}(x) = \mathcal{T}_{\mathcal{X}_1}(x) \cap \mathcal{T}_{\mathcal{X}_2}(x)$ [1, pp.141].

Since $\mathrm{int}(\mathcal{T}_{\mathcal{X}_1}(x)) \neq \emptyset$, for a vector $v \in \mathrm{int}(\mathcal{T}_{\mathcal{X}_1}(x))$, there exists a scalar $\epsilon > 0$ such that

$$y + tw \in \mathcal{T}_{\mathcal{X}_1}(x), \quad \text{for all } y \in \mathcal{T}_{\mathcal{X}_1}(x) \cap B_\epsilon(x),\ w \in B_\epsilon(v) \text{ and } 0 \leq t < \epsilon.$$

We often call $\mathrm{int}(\mathcal{T}_{\mathcal{X}_1}(x))$ the hypertangent cone to $\mathcal{X}_1$ at x.

And by Lemma 1 (2), we have $x + tv \in \mathcal{X}_2$, $\forall x \in \mathcal{X}_2$, $t \geq 0$, $v \in \mathcal{T}_{\mathcal{X}_2}(x)$.

2.3 Necessary optimality condition

Denote

$$\text{r-int}(\mathcal{T}_{\mathcal{X}}(x)) = \mathrm{int}(\mathcal{T}_{\mathcal{X}_1}(x)) \cap \mathcal{T}_{\mathcal{X}_2}(x).$$

Since f is not assumed to be locally Lipschitz continuous, the calculus theory developed in [2] cannot be directly applied to f. The next lemma extends calculus results for the unconstrained case in [18] and the constrained case in [2].

For any $x \in \mathcal{X}$, from $0 \in \text{r-int}(\mathcal{T}_{\mathcal{X}}(x)) \cap \mathcal{V}_x$, we know $\text{r-int}(\mathcal{T}_{\mathcal{X}}(x)) \cap \mathcal{V}_x \neq \emptyset$.

Lemma 2 *For $\bar{x} \in \mathcal{X}$ and $v \in \mathcal{T}_{\mathcal{X}}(\bar{x}) \cap \mathcal{V}_{\bar{x}}$,*

$$f^\circ(\bar{x}; v) = \lim_{\substack{w \to v \\ w \in \text{r-int}(\mathcal{T}_{\mathcal{X}}(\bar{x})) \cap \mathcal{V}_{\bar{x}}}} f^\circ(\bar{x}; w).$$

Proof By the locally Lipschitz continuity of $h_{\bar{x}}$, there are $\epsilon > 0$ and $L_{\bar{x}} > 0$ such that

$$\|h_{\bar{x}}(x) - h_{\bar{x}}(y)\|_2 \leq L_{\bar{x}}\|x - y\|_2, \quad \forall x,\, y \in B_\epsilon(\bar{x}). \tag{16}$$

Let $\{w_k\} \subseteq \text{r-int}(\mathcal{T}_{\mathcal{X}}(\bar{x})) \cap \mathcal{V}_{\bar{x}}$ be a sequence of directions converging to a vector $v \in \mathcal{T}_{\mathcal{X}}(\bar{x}) \cap \mathcal{V}_{\bar{x}}$.

By $\{w_k\} \subseteq \text{r-int}(\mathcal{T}_{\mathcal{X}}(\bar{x}))$, let $\epsilon_k > 0$ be such that $x + tw_k \in \mathcal{X}_1$ whenever $x \in \mathcal{X} \cap B_{\epsilon_k}(\bar{x})$ and $0 \leq t < \epsilon_k$. By Lemma 1 (2), we obtain $x + tv \in \mathcal{X}_2, x + tw_k \in \mathcal{X}_2$, $\forall t \geq 0$, $x \in \mathcal{X}$. Then, for all w_k, it gives

$$\begin{aligned} f^\circ(\bar{x}; v) = & \limsup_{\substack{x \to \bar{x},\, x \in \mathcal{X} \\ t \downarrow 0,\, x + tv \in \mathcal{X}}} \frac{f(x + tv) - f(x)}{t} \\ = & \limsup_{\substack{x \to \bar{x},\, x \in \mathcal{X} \\ t \downarrow 0,\, x + tv \in \mathcal{X} \\ x + tw_k \in \mathcal{X}}} \frac{f(x + tv) - f(x)}{t} \\ = & \limsup_{\substack{x \to \bar{x},\, x \in \mathcal{X} \\ t \downarrow 0,\, x + tv \in \mathcal{X} \\ x + tw_k \in \mathcal{X}}} \frac{f(x + tw_k) - f(x)}{t} + \frac{f(x + tv) - f(x + tw_k)}{t}. \end{aligned} \tag{17}$$

Let $\delta > 0$ be such that $x + tw_k \in B_\epsilon(\bar{x})$ for any $x \in B_\delta(\bar{x})$, $0 \leq t < \delta$ and $k \in \mathbb{N}_{++}$. By the Lipschitz condition in (16), we have

$$\|\frac{h_{\bar{x}}(x + tv) - h_{\bar{x}}(x + tw_k)}{t}\|_2 \leq L_{\bar{x}}\|v - w_k\|_2, \quad \forall x \in B_\delta(\bar{x}), 0 < t < \delta, k \in \mathbb{N}_{++}.$$

From

$$\begin{aligned} & f(x + tv) - f(x + tw_k) \\ = & \Theta(x + tv) - \Theta(x + tw_k) + \nabla c(z)^T_{z \in [h(x+tv), h(x+tw_k)]}(h(x + tv) - h(x + tw_k)) \\ = & \Theta(x + tv) - \Theta(x + tw_k) + \nabla c(z)^T_{z \in [h(x+tv), h(x+tw_k)]}(h_{\bar{x}}(x + tv) - h_{\bar{x}}(x + tw_k)), \end{aligned}$$

for any $x \in B_\delta(\bar{x})$,$0 \leq t < \delta$, we have

$$|\frac{f(x + tv) - f(x + tw_k)}{t}| \leq L_\Theta\|v - w_k\|_2 + L_c L_{\bar{x}}\|v - w_k\|_2,$$

where $L_\Theta = \sup\{\|\nabla\Theta(y)\|_2 : y \in B_\epsilon(\bar{x})\}$ and $L_c = \sup\{\|\nabla c(z)^T_{z=h(y)}\|_2 : y \in B_\epsilon(\bar{x})\}$.

Thus, (17) implies

$$\begin{aligned} f^\circ(\bar{x}; w_k) - & L_\Theta\|v - w_k\|_2 - L_c L_{\bar{x}}\|v - w_k\|_2 \leq f^\circ(\bar{x}; v) \\ & \leq f^\circ(\bar{x}; w_k) + L_\Theta\|v - w_k\|_2 + L_c L_{\bar{x}}\|v - w_k\|_2, \forall k \in \mathbb{N}_{++}. \end{aligned}$$

As k goes to infinity, the above inequality follows $f^\circ(\bar{x}; v) = \lim_{k\to\infty} f^\circ(\bar{x}; w_k)$. Since $\{w_k\}$ is an arbitrary sequence in $\text{r-int}(\mathcal{T}_{\mathcal{X}}(\bar{x})) \cap \mathcal{V}_{\bar{x}}$ converging to v, we obtain the result in this lemma.

Note that the above lemma is not necessarily true when r-int($\mathcal{T}_{\mathcal{X}}(x)$) is empty. A similar example can be given following the idea in [2, Example 3.10]. That is why we put the assumption $\text{int}(\mathcal{X}_1) \cap \mathcal{X}_2 \neq \emptyset$ at the beginning of this section. Based on Lemmas 1-2, the following theorem gives the main theoretical result of this section.

Theorem 1 *If x^* is a local minimizer of (1), then $f^\circ(x^*, v) \geq 0$ for every direction $v \in \mathcal{T}_{\mathcal{X}}(x^*) \cap \mathcal{V}_{x^*}$.*

Proof Suppose x^* is a local minimizer of f over $\mathcal{X}$ and let $w \in \text{r-int}(\mathcal{T}_{\mathcal{X}}(x^*)) \cap \mathcal{V}_{x^*}$.

There exist $\epsilon > 0$ and $L_{x^*} > 0$ such that $f(x^*) \leq f(x)$, and

$$\|h_{x^*}(x) - h_{x^*}(y)\|_2 \leq L_{x^*}\|x - y\|_2, \ \forall x, \, y \in \mathcal{X} \cap B_\epsilon(x^*). \tag{18}$$

Since $w \in \text{int}(\mathcal{T}_{\mathcal{X}_1}(x^*))$, there exists $\bar{\epsilon} \in (0, \epsilon]$ such that

$$x + tw \in \mathcal{X}_1, \forall x \in \mathcal{X}_1 \cap B_{\bar{\epsilon}}(x^*), 0 \leq t \leq \bar{\epsilon}.$$

By Lemma 1 (2), $x + tw \in \mathcal{X}$, $\forall x \in \mathcal{X} \cap B_{\bar{\epsilon}}(x^*)$, $0 \leq t < \bar{\epsilon}$. And then we can choose $\delta \in (0, \bar{\epsilon}]$ such that $x, x + tw, x^* + tw \in B_\epsilon(x^*) \cap \mathcal{X}$,$\forall x \in B_{t^2}(x^*) \cap \mathcal{X}$, $0 \leq t < \delta$.

By (18), for all $x \in B_{t^2}(x^*) \cap \mathcal{X}$, $0 < t < \delta$, we obtain

$$\|\frac{h_{x^*}(x + tw) - h_{x^*}(x^* + tw)}{t} - \frac{h_{x^*}(x) - h_{x^*}(x^*)}{t}\|_2 \leq 2L_{x^*}\frac{\|x - x^*\|_2}{t} \leq 2L_{x^*}t.$$

Thus,

$$\lim_{\substack{x \,\in\, B_{t^2}(x^*) \,\cap\, \mathcal{X} \\ x \,+\, tw \,\in\, \mathcal{X},\, t \,\downarrow\, 0}} \frac{h_{x^*}(x + tw) - h_{x^*}(x^* + tw)}{t} - \frac{h_{x^*}(x) - h_{x^*}(x^*)}{t} = 0. \tag{19}$$

From the mean value theorem, there exist $z_1 = x + s_1 w$ and $z_2 = x^* + s_2 w$ with $s_1, s_2 \in (0, t)$ such that

$$\begin{aligned}
&|\frac{c(h(x + tw)) - c(h(x))}{t} - \frac{c(h(x^* + tw)) - c(h(x^*))}{t}| \\
=&|\frac{\nabla c(h(z_1))^T(h(x + tw) - h(x))}{t} - \frac{\nabla c(h(z_2))^T(h(x^* + tw) - h(x^*))}{t}| \\
=&|\frac{\nabla c(h(z_1))^T(h_{x^*}(x + tw) - h_{x^*}(x))}{t} - \frac{\nabla c(h(z_2))^T(h_{x^*}(x^* + tw) - h_{x^*}(x^*))}{t}|.
\end{aligned} \tag{20}$$

By (19), (20) and the continuous differentiability of Θ, we have

$$\begin{aligned}
&\lim_{\substack{x \,\in\, B_{t^2}(x^*) \,\cap\, \mathcal{X} \\ x \,+\, tw \,\in\, \mathcal{X},\, t \,\downarrow\, 0}} [\frac{f(x + tw) - f(x)}{t} - \frac{f(x^* + tw) - f(x^*)}{t}] \\
=&\nabla c(y)^T_{y=h(x^*)} \lim_{\substack{x \,\in\, B_{t^2}(x^*) \,\cap\, \mathcal{X} \\ x \,+\, tw \,\in\, \mathcal{X},\, t \,\downarrow\, 0}} [\frac{h_{x^*}(x + tw) - h_{x^*}(x)}{t} - \frac{h_{x^*}(x^* + tw) - h_{x^*}(x^*)}{t}] \\
=&0.
\end{aligned}$$

Thus,

$$\limsup_{\substack{x \to x^*,\, x \in \mathcal{X} \\ t \downarrow 0,\, x+tw \in \mathcal{X}}} \left[\frac{f(x+tw)-f(x)}{t} - \frac{f(x^*+tw)-f(x^*)}{t}\right] \geq 0. \tag{21}$$

By $f(x^*+tw)-f(x^*) \geq 0$ for $0 \leq t < \bar{\epsilon}$, (21) implies

$$f^\circ(x^*;w) = \limsup_{\substack{x \to x^*,\, x \in \mathcal{X} \\ t \downarrow 0,\, x+tw \in \mathcal{X}}} \frac{f(x+tw)-f(x)}{t} \geq 0.$$

By Lemma 2, we can give that $f^\circ(x^*;v) \geq 0$ for any $v \in \mathcal{T}_{\mathcal{X}}(x^*) \cap \mathcal{V}_{x^*}$.

Based on Theorem 1, we give a new definition of generalized stationary point of problem (1).

Definition 2 $x^* \in \mathcal{X}$ is said to be a generalized stationary point of (1), if $f^\circ(x^*;v) \geq 0$ for every $v \in \mathcal{T}_{\mathcal{X}}(x^*) \cap \mathcal{V}_{x^*}$.

It is worth noting that the generalized stationary point x^* is a Clarke stationary point of problem (1) when f is Lipschitz continuous at x^*.

Remark 1 Suppose $h_i(D_i^T x)$ is regular in $\mathcal{X} \backslash \mathcal{N}_i$, where

$$\mathcal{N}_i = \{x \in \mathcal{X} : h_i \text{ is not Lipschitz continuous at } D_i^T x\},\ i = 1, 2, \ldots, m.$$

For $\bar{x} \in \mathcal{X}$, the regularity assumption allows us to define $\mathcal{V}_{\bar{x}}$ by

$$\begin{aligned}\mathcal{V}_{\bar{x}} = \{v : & \text{ for any } i \in \mathcal{I}_{\bar{x}}, \text{ there exists } \delta > 0 \text{ such that} \\ & h_i(\bar{x}+tv) = h_i(\bar{x}) \text{ holds for all } 0 \leq t \leq \delta\},\end{aligned}$$

which is a bigger set than $\mathcal{V}_{\bar{x}}$ given in (8). Hence the generalized stationary point defined in Definition 2 can be more robust with this $\mathcal{V}_{\bar{x}}$. For example, if f is defined as in (3), $\mathcal{I}_{\bar{x}} = \{i \in \{1, 2, \ldots, m\} : m_i^T \bar{x} = \alpha_i\}$ and we can let

$$\mathcal{V}_{\bar{x}} = \{v : m_i^T v \geq 0,\ \forall\, i \in \mathcal{I}_{\bar{x}}\},$$

which includes $\{v : m_i^T v = 0,\ \forall i \in \mathcal{I}_{\bar{x}}\}$ as a proper subset.

We notice that a generalized stationary point defined in Definition 2 is a scaled stationary point defined in [5,6,15,24] for the special cases of (2) with $0 < p < 1$. Moreover it is stronger than a scaled stationary point for the Lipschitz case, since it is a Clarke stationary point but a scaled stationary point is not necessarily a Clarke stationary point for the Lipschitz optimization problem.

2.4 Directional derivative consistency

In this subsection, we show that the generalized directional derivative of f defined in (12) can be represented by the limit of a sequence of directional derivatives of a smoothing function of f. This property is important for development of numerical algorithms for nonconvex non-Lipschitz constrained optimization problems.

Definition 3 [13] Let $g : R^n \to R$ be a continuous function. We call $\tilde{g} : R^n \times [0, \infty) \to R$ a smoothing function of g, if $\tilde{g}(\cdot, \mu)$ is continuously differentiable for any fixed $\mu > 0$ and $\lim_{z \to x, \mu \downarrow 0} \tilde{g}(z, \mu) = g(x)$ holds for any $x \in R^n$.

Let $\tilde{h}(x, \mu) = (\tilde{h}_1(D_1^T x, \mu), \tilde{h}_2(D_2^T x, \mu), \ldots, \tilde{h}_m(D_m^T x, \mu))^T$, where $\tilde{h}_i$ is a smoothing function of h_i in (6). Then $\tilde{f}(x, \mu) := \Theta(x) + c(\tilde{h}(x, \mu))$ is a smoothing function of f.

Since $\tilde{f}(x, \mu)$ is continuously differentiable about x for any fixed $\mu > 0$, the generalized directional derivative of it with respect to x can be given by

$$\tilde{f}^\circ(x, \mu; v) = \limsup_{\substack{y \to x,\, y \in \mathcal{X} \\ t \downarrow 0,\, y + tv \in \mathcal{X}}} \frac{\tilde{f}(y + tv, \mu) - \tilde{f}(y)}{t} = \langle \nabla_x \tilde{f}(x, \mu), v \rangle. \tag{22}$$

Theorem 2 *Suppose h_i is continuously differentiable in $\mathcal{X} \backslash \mathcal{N}_i$, $\forall i \in \{1, 2, \ldots, m\}$, where $\mathcal{N}_i = \{x :\ h_i$ is not Lipschitz continuous at $D_i^T x\}$, then*

$$\lim_{\substack{x_k \in \mathcal{X}, \\ x_k \to x,\, \mu_k \downarrow 0}} \langle \nabla_x \tilde{f}(x_k, \mu_k), v \rangle = f^\circ(x; v),\ \forall v \in \mathcal{V}_x. \tag{23}$$

Proof Let x_k be a sequence in $\mathcal{X}$ converging to $\bar{x}$ and $\{\mu_k\}$ be a positive sequence converging to 0. For $w \in \mathcal{V}_{\bar{x}}$, by the closed form of $\nabla_x \tilde{f}(x_k, \mu_k)$, we have

$$\begin{aligned} &\langle \nabla_x \tilde{f}(x_k, \mu_k), w \rangle \\ =&\langle \nabla \Theta(x_k), w \rangle + \langle \nabla_x \tilde{h}(x_k, \mu_k) \nabla c(z)_{z = \tilde{h}(x_k, \mu_k)}, w \rangle \\ =&\langle \nabla \Theta(x_k), w \rangle + \langle \nabla c(z)_{z = \tilde{h}(x_k, \mu_k)}, \nabla_x \tilde{h}(x_k, \mu_k)^T w \rangle, \end{aligned} \tag{24}$$

where

$$\nabla_x \tilde{h}(x_k, \mu_k)^T w = (\nabla_x \tilde{h}_1(D_1^T w, \mu_k)^T w, \ldots, \nabla_x \tilde{h}_m(D_m^T x_k, \mu_k)^T w)^T.$$

For $i \in \mathcal{I}_{\bar{x}}$, by $w \in \mathcal{V}_{\bar{x}}$, we obtain $D_i^T w = 0$, then $\nabla_x \tilde{h}_i(D_i^T x_k, \mu_k)^T w = \nabla_z \tilde{h}_i(z, \mu_k)^T_{z = D_i^T x_k} D_i^T w = 0$.

Define

$$\tilde{h}_i^{\bar{x}}(D_i^T x, \mu) = \begin{cases} \tilde{h}_i(D_i^T x, \mu) & i \notin \mathcal{I}_{\bar{x}}, \\ \tilde{h}_i(D_i^T \bar{x}, \mu) & i \in \mathcal{I}_{\bar{x}}, \end{cases} \quad i = 1, 2, \ldots, m.$$

Denote $\tilde{h}_{\bar{x}}(x, \mu) = (\tilde{h}_1^{\bar{x}}(D_1^T x, \mu), \tilde{h}_2^{\bar{x}}(D_2^T x, \mu), \ldots, \tilde{h}_m^{\bar{x}}(D_m^T x, \mu))^T$. Then,

$$\nabla_x \tilde{h}(x_k, \mu_k)^T w = \nabla_x \tilde{h}_{\bar{x}}(x_k, \mu_k)^T w.$$

Thus, coming back to (24), we obtain

$$
\begin{aligned}
\langle \nabla_x \tilde{f}(x^k, \mu_k), w\rangle =& \langle \nabla \Theta(x_k), w\rangle + \langle \nabla c(z)_{z=\tilde{h}(x_k,\mu_k)}, \nabla_x \tilde{h}_{\bar{x}}(x_k, \mu_k)^T w\rangle \\
=& \langle \nabla \Theta(x_k), w\rangle + \langle \nabla_x \tilde{h}_{\bar{x}}(x_k, \mu_k) \nabla c(z)_{z=\tilde{h}(x_k,\mu_k)}, w\rangle \\
=& \langle \nabla \Theta(x_k) + \nabla_x \tilde{h}_{\bar{x}}(x_k, \mu_k) \nabla c(z)_{z=\tilde{h}(x_k,\mu_k)}, w\rangle.
\end{aligned} \tag{25}
$$

Since h_i is continuously differentiable at $D_i^T \bar{x}$ for $i \notin \mathcal{I}_{\bar{x}}$ and $h_{\bar{x}}(\bar{x}) = h(\bar{x})$, we obtain

$$
\begin{aligned}
&\lim_{k\to\infty} \nabla \Theta(x_k) + \nabla_x \tilde{h}_{\bar{x}}(x_k, \mu_k) \nabla c(z)_{z=\tilde{h}(x_k,\mu_k)} \\
=& \nabla \Theta(\bar{x}) + \nabla h_{\bar{x}}(\bar{x}) \nabla c(z)_{z=h(\bar{x})} = \nabla f_{\bar{x}}(\bar{x}),
\end{aligned} \tag{26}
$$

where $f_{\bar{x}}$ is defined in (9).

Thus,

$$
\begin{aligned}
f^\circ(\bar{x}, w) = f_{\bar{x}}^\circ(\bar{x}, w) &= \langle \nabla f_{\bar{x}}(\bar{x}), w\rangle \\
&= \langle \lim_{k\to\infty} \nabla \Theta(x_k) + \nabla_x \tilde{h}_{\bar{x}}(x_k, \mu_k) \nabla c(z)_{z=\tilde{h}(x_k,\mu_k)}, w\rangle \\
&= \lim_{k\to\infty} \langle \nabla_x \tilde{f}(x^k, \mu_k), w\rangle,
\end{aligned} \tag{27}
$$

where the first equation uses Proposition 1, the third uses (26) and the fourth uses (25).

Now we give another consistency result on subspace $\mathcal{V}_x$.

Lemma 3 *Let x_k be a sequence in $\mathcal{X}$ with a limit point $\bar{x}$. For $w \in \mathcal{V}_{\bar{x}}$, there exists a sequence $\{x_{k_l}\} \subseteq \{x_k\}$ such that $w \in \mathcal{V}_{x_{k_l}}$, $\forall l \in \mathbb{N}_{++}$.*

Proof If not, there is $K \in \mathbb{N}_{++}$ such that

$$w \notin \mathcal{V}_{x_k}, \quad \forall k \geq K.$$

By the definition of $\mathcal{V}_{x_k}$, there exists $i_k \in \mathcal{I}_{x_k}$ such that

$$D_{i_k}^T w \neq 0, \quad \forall k \geq K.$$

By $\mathcal{I}_{x_k} \subseteq \{1, 2, \ldots, m\}$, there exists $j \in \{1, 2, \ldots, m\}$ and a subsequence of $\{x_k\}$, denoted as $\{x_{k_l}\}$, such that $j \in \mathcal{I}_{x_{k_l}}$ and $D_j^T w \neq 0$.

Note that $j \in \mathcal{I}_{x_{k_l}}$ implies h_j is not Lipschitz continuous at $D_j^T x_{k_l}$. Since the non-Lipschitz points of h_j is a closed subset of R^n, h_j is also not Lipschitz continuous at $D_j^T \bar{x}$, which means $j \in \mathcal{I}_{\bar{x}}$. By $w \in \mathcal{V}_{\bar{x}}$, we obtain $D_j^T w = 0$, which leads a contradiction. Therefore, the statement in this lemma holds.

Based on the consistency results given in Theorem 2 and Lemma 3, the next corollary shows the generalized stationary point consistency of the smooth functions.

Corollary 1 *Let $\{\epsilon_k\}$ and $\{\mu_k\}$ be positive sequences converging to 0. With the conditions on h in Theorem 2, if x^k satisfies $\langle \nabla_x \tilde{f}(x^k, \mu_k), v\rangle \geq -\epsilon_k$ for every $v \in \mathcal{T}_{\mathcal{X}}(x^k) \cap \mathcal{V}_{x^k} \cap B_1(0)$, then any accumulation point of $\{x^k\} \subseteq \mathcal{X}$ is a generalized stationary point of (1).*

Proof Let $\bar{x}$ be an accumulation point of $\{x^k\}$. Without loss of generality, we suppose $\lim_{k\to\infty} x_k = \bar{x}$.

For $w \in \text{r-int}(\mathcal{T}_{\mathcal{X}}(\bar{x})) \cap \mathcal{V}_{\bar{x}} \cap B_1(0)$, from Lemma 3, we can suppose

$$w \in \mathcal{V}_{x_k}, \quad \forall k \in \mathbb{N}_{++}.$$

By $w \in \text{r-int}(\mathcal{T}_{\mathcal{X}}(\bar{x}))$, there exists $\epsilon > 0$ such that

$$x + sw \in \mathcal{X}, \quad \forall x \in \mathcal{X} \cap B_\epsilon(\bar{x}),\ 0 \le s \le \epsilon. \tag{28}$$

Since x_k is converging to $\bar{x}$, there exists $K \in \mathbb{N}_{++}$ such that $x_k \in \mathcal{X} \cap B_\epsilon(\bar{x})$, $\forall k \ge K$. By (28), we have $x_k + sw \in \mathcal{X}$, $\forall k \ge K$, $0 \le s \le \epsilon$. From the convexity of $\mathcal{X}$, we obtain $w \in \mathcal{T}_{\mathcal{X}}(x_k)$.

From Theorem 2, we have $f^\circ(\bar{x}, w) \ge 0$. Then, for any $\rho > 0$, we have

$$\begin{aligned} f^\circ(\bar{x};\rho v) = & \limsup_{\substack{y \to \bar{x},\, y \in \mathcal{X} \\ t \downarrow 0,\, y + t\rho v \in \mathcal{X}}} \frac{f(y + t\rho v) - f(y)}{t} \\ =& \rho \limsup_{\substack{y \to \bar{x},\, y \in \mathcal{X} \\ s \downarrow 0,\, y + sv \in \mathcal{X}}} \frac{f(y + sv) - f(y)}{s} = \rho f^\circ(\bar{x}; v) \ge 0. \end{aligned} \tag{29}$$

Thus, $f^\circ(\bar{x}; v) \ge 0$ for every $v \in \text{r-int}(\mathcal{T}_{\mathcal{X}}(\bar{x})) \cap \mathcal{V}_{\bar{x}} \cap B_1(0)$ implies $f^\circ(\bar{x}; v) \ge 0$ for every $v \in \text{r-int}(\mathcal{T}_{\mathcal{X}}(\bar{x})) \cap \mathcal{V}_{\bar{x}}$. By Lemma 2, it is easy to verify that $f^\circ(\bar{x}, v) \ge 0$ holds for any $v \in \mathcal{T}_{\mathcal{X}}(\bar{x}) \cap \mathcal{V}_{\bar{x}}$, which means that $\bar{x}$ is a generalized stationary point of (1).

Remark 2 Suppose the gradient consistency associated with the smoothing function $\tilde{h}_i$ holds at its Lipschitz continuous points, that is

$$\{\lim_{z\to x, \mu\downarrow 0} \nabla_x \tilde{h}_i(D_i^T x, \mu)\} \subseteq \partial h_i(D_i^T x), \quad \forall x \in \mathcal{X},\ i \notin \mathcal{I}_x, \tag{30}$$

then

$$\{\lim_{z\to x, \mu\downarrow 0} \nabla \Theta(z) + \nabla \tilde{h}_x(x,\mu) \nabla c(z)_{z=h_x(x)}\} \subseteq \partial f_x(x),\ \forall x \in \mathcal{X}. \tag{31}$$

Since $f_{\bar{x}}$ is Lipschitz continuous at $\bar{x}$, it gives

$$f_{\bar{x}}^\circ(\bar{x}, v) = \max\{\langle \xi, v\rangle : \xi \in \partial f_{\bar{x}}(\bar{x})\}. \tag{32}$$

Similar to the calculation in (27), by (31) and (32), we obtain

$$\begin{aligned} f^\circ(\bar{x}, w) = f_{\bar{x}}^\circ(\bar{x}, w) =& \max\{\langle \xi, w\rangle : \xi \in \partial f_{\bar{x}}(\bar{x})\} \\ \ge& \limsup_{k\to\infty} \langle \nabla \Theta(x_k) + \nabla_x \tilde{h}_{\bar{x}}(x_k, \mu_k) \nabla c(z)_{z=\tilde{h}(x_k,\mu_k)}, w\rangle \\ =& \limsup_{k\to\infty} \langle \nabla_x \tilde{f}(x^k, \mu_k), w\rangle. \end{aligned}$$

Thus, the conclusion in Corollary 1 can be true with (30), which is weaker than the strict differentiability of h_i in $\mathcal{X}\backslash\mathcal{N}_i$, $i \in \{1,2,\ldots,m\}$. Some conditions can be found in [18] to ensure (30). Specially, when the function h in f is with the form

$$h(x) := (h_1(d_1^T x), h_2(d_2^T x), \ldots, h_m(d_m^T x))^T$$

with $d_i \in R^n$, by [18, Theorem 2.3.9 (i)], the regularity of $h_i(d_i^T x)$ in $\mathcal{X}\backslash\mathcal{N}_i$ is a sufficient condition for the statement in Theorem 2.

Corollary 1 shows that one can find a generalized stationary point of (1) by using the approximate first order optimality condition of $\min_{x\in\mathcal{X}} \tilde{f}(x,\mu)$. Since $\tilde{f}(x,\mu)$ is continuously differentiable for any fixed $\mu > 0$, many numerical algorithms can find a stationary point of $\min_{x\in\mathcal{X}} \tilde{f}(x,\mu)$ [7,19,30,34,37]. We use one example to show the validity of the first order necessary optimality condition and the consistency result given in this section.

Example 1 Consider the following minimization problem

$$\begin{array}{l} \min\ f(x) := (x_1+2x_2-1)^2 + \lambda_1\sqrt{\max\{x_1+x_2+1,0\}} + \lambda_2\sqrt{|x_2|}, \\ \text{s.t. }\ x \in \mathcal{X} = \{x \in R^2 : -1 \le x_1, x_2 \le 1\}. \end{array} \tag{33}$$

This problem is an example of (1) with $\Theta(x) = (x_1+2x_2-1)^2$, $c(y) = \lambda_1 y_1 + \lambda_2 y_2$, $h_1(D_1^T x) = \sqrt{\max\{x_1+x_2+1,0\}}$ and $h_2(D_2^T x) = \sqrt{|x_2|}$, where $D_1 = (1,1)^T$, $D_2 = (0,1)^T$.

Define the smoothing function of f as

$$\tilde{f}(x,\mu) = (x_1+2x_2-1)^2 + \lambda_1\sqrt{\psi(x_1+x_2+1,\mu)} + \lambda_2\sqrt{\theta(x_2,\mu)},$$

with $\psi(s,\mu) = \frac{1}{2}(s+\sqrt{s^2+4\mu^2})$, $\quad \theta(s,\mu) = \begin{cases} |s| & |s| > \mu, \\ \dfrac{s^2}{2\mu} + \dfrac{\mu}{2} & |s| \le \mu. \end{cases}$

Here, we use the classical projected algorithm with Armijo line search to find an approximate generalized stationary point of $\min_{x\in\mathcal{X}} \tilde{f}(x,\mu)$. There exists $\alpha > 0$ such that $\bar{x} - P_{\mathcal{X}}[\bar{x} - \alpha\nabla_x \tilde{f}(\bar{x},\mu)] = 0$ if and only if $\bar{x}$ is a generalized stationary point of $\min_{x\in\mathcal{X}} \tilde{f}(x,\mu)$, which is also a Clarke stationary point of $\min_{x\in\mathcal{X}} \tilde{f}(x,\mu)$ for any fixed $\mu > 0$. We call x^k an approximate stationary point of $\min_{x\in\mathcal{X}} \tilde{f}(x,\mu_k)$, if there exists $\alpha_k > 0$ such that $\|x^k - P_{\mathcal{X}}[x^k - \alpha_k\nabla_x \tilde{f}(x^k,\mu_k)]\|_2 \le \alpha_k\mu_k$, which can be found in finite number of iterations by the analysis in [4].

Choose the initial iterate $x_0 = (0,0)^T$. For different values of λ_1 and λ_2 in (33), the simulation results are listed in Table 1, where f^* indicates the optimal function value of (33), where the iteration is terminated when $\mu_k \le 10^{-6}$.

λ_1	λ_2	accumulation point x^*	$\mathcal{I}_{x^*}$	$\mathcal{V}_{x^*}$	$f(x^*)$	f^*
8	2	$(-1.000, 0.000)^T$	$\{1\}$	$\{v = (a, -a)^T : a \in R\}$	4.000	4.000
0.1	0.2	$(0.982, 0.000)^T$	$\{2\}$	$\{v = (a, 0)^T : a \in R\}$	0.141	0.141
0.5	0.1	$(-1.000, 0.962)^T$	$\emptyset$	R^2	0.594	0.594

Table 1: Simulation results in Example 1

When $\lambda_1 = 8$, $\lambda_2 = 2$, since $h_2(D_2^T x)$ is continuously differentiable at x^*, for $v \in \mathcal{V}_{x^*}$, by $h_1(D_1^T(x^* + tv)) = h_1(D_1^T x^*)$, $\forall t > 0$, we obtain

$$\begin{aligned} f^\circ(x^*; v) = & \limsup_{\substack{y \to x^*,\, y \in \mathcal{X} \\ t \downarrow 0,\, y + tv \in \mathcal{X}}} \frac{\Theta(y + tv) - \Theta(y) + \lambda_2 h_2(D_2^T(y + tv)) - \lambda_2 h_2(D_2^T y)}{t} \\ = & \langle \nabla\Theta(x^*) + \lambda_2 h_2'(D_2^T x^*) D_2, v\rangle = -4v_1 - 550.473 v_2, \end{aligned}$$

where $v_1 = -v_2$ by $v \in \mathcal{V}_{x^*}$, and $v_1 \in R_+$ by $x_1^* = -1.000$ and the condition $x^* + tv \in \mathcal{X}$ in $f^\circ(x^*; v)$. Then, $f^\circ(x^*; v) \geq 0$, $\forall v \in \mathcal{V}_{x^*}$, which means that $(-1.000, 0.000)^T$ is a generalized stationary point of (33). Similarly,

- when $\lambda_1 = 0.1$, $\lambda_2 = 0.2$:

  $$f^\circ(x^*; v) = \langle \nabla\Theta(x^*) + \lambda_1 h_1'(D_1^T x^*) D_1, v\rangle = -0.036 v_2,$$

  where $v_2 = 0$ by $v \in \mathcal{V}_{x^*}$.
- when $\lambda_1 = 0.5$, $\lambda_2 = 0.1$:

  $$f^\circ(x^*; v) = \langle \nabla\Theta(x^*) + \lambda_1 h_1'(D_1^T x^*) D_1 + \lambda_2 h_2'(D_2^T x^*) D_2, v\rangle = 0.102 v_1,$$

  where $v_1 \in R_+$ by $x_1^* = -1.000$,

which gives $f^\circ(x^*; v) \geq 0$, for all $v \in \mathcal{V}_{x^*}$. Thus, the accumulation points in Table 1 are generalized stationary points of (33) with different values of λ_1 and λ_2. Furthermore, the trajectory of x^k of the smoothing algorithm for (33) with $\lambda_1 = 8$, $\lambda_2 = 2$ are pictured in Figure 1 with the isolines of f in $\mathcal{X}$.

3 Nonconvex regularization

In this section, we focus on problem (2) with the function φ satisfying the following assumption.

Assumption 2 *Assume that $\varphi : R_+ \to R_+$ with $\varphi(0) = 0$ is continuously differentiable, non-decreasing and concave on $(0, \infty)$, and φ' is locally Lipschitz continuous on R_{++}.*

The function $\varphi(t) = t$ and $p \in (0, 1)$ satisfies Assumption 2. It is know that problem (2) with $\mathcal{X} = R^n$ and $\varphi(t) = t$ is strongly NP hard but enjoys lower bound theory. However, the complexity and lower bound theory of problem (2) with a general convex set $\mathcal{X}$ and the class of functions φ satisfying Assumption 2 have not been studied. In this section, we show that the key condition for the complexity and lower bound theory is that the function $\varphi(z^p)$ is strictly concave in an open interval.

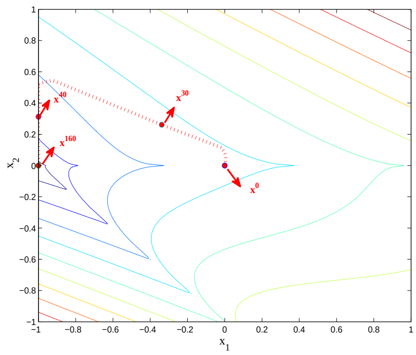

Fig. 1: Trajectory of x^k in Example 1 with $\lambda_1 = 8$ and $\lambda_2 = 2$

3.1 Computational complexity

In this subsection, we will show the strong NP-hardness of the following problem

$$\min \quad \|Hx - c\|_2^2 + \sum_{i=1}^{n} \varphi(|x_i|^p), \tag{34}$$

where $H \in R^{s\times n}$, $c \in R^s$ and $0 < p \leq 1$.

Lemma 4 *$\varphi(|s|^p) + \varphi(|t|^p) \geq \varphi(|s+t|^p)$, $\forall s, t \in R$.*

Proof Define $\psi(\alpha) = \varphi(\alpha + |s|^p) - \varphi(\alpha)$ on $[0, +\infty)$. Then from the concavity of φ, $\psi'(\alpha) = \varphi'(\alpha + |s|^p) - \varphi'(\alpha) \leq 0$, which implies $\psi(|t|^p) \leq \psi(0)$. Thus, $\varphi(|t|^p + |s|^p) \leq \varphi(|t|^p) + \varphi(|s|^p)$. Since $|t + s|^p \leq |t|^p + |s|^p$ and φ is nondecreasing on $[0, +\infty)$, we obtain $\varphi(|t + s|^p) \leq \varphi(|t|^p) + \varphi(|s|^p)$.

First, we give two preliminary results for proving the strong NP-hardness of (34) with $0 < p \leq 1$. The first is for $p = 1$ and the second is for $0 < p < 1$.

Lemma 5 *Suppose φ is strictly concave and twice continuously differentiable on $[\tau_1, \tau_2]$ with $\tau_1 > 0$ and $\tau_2 > \tau_1$. There exists $\bar{\gamma} > 0$ such that when $\gamma > \bar{\gamma}$ and $p = 1$, the minimization problem*

$$\min_{z\in R} \quad g(z) = \gamma|z - \tau_1|^2 + \gamma|z - \tau_2|^2 + \varphi(|z|^p), \tag{35}$$

has a unique solution $z^* \in (\tau_1, \tau_2)$.

Proof Since φ is twice continuously differentiable in $[\tau_1, \tau_2]$, there exists $\alpha > 0$ such that $0 \leq \varphi'(s) \leq \alpha$ and $-\alpha \leq \varphi''(s) \leq 0$, $\forall s \in [\tau_1, \tau_2]$. Let $\bar{\gamma} = \max\{\frac{\alpha}{2(\tau_2-\tau_1)}, \frac{\alpha}{4}\}$ and suppose $\gamma > \bar{\gamma}$.

Note that $g(z) > g(0) = \gamma\tau_1^2 + \gamma\tau_2^2$ for all $z < 0$, and $g(z) > g(\tau_2) = \gamma(\tau_2 - \tau_1)^2 + \varphi(\tau_2)$ for all $z > \tau_2$. Then, the minimum point of $g(z)$ must lie within $[0, \tau_2]$.

To minimize $g(z)$ on $[0, \tau_2]$, we check its first derivative

$$g'(z) = 2\gamma(z - \tau_1) + 2\gamma(z - \tau_2) + \varphi'(z), \quad 0 < z \leq \tau_2.$$

When $0 < z \leq \tau_1$, $g'(z) = 4\gamma z - 2\gamma\tau_1 - 2\gamma\tau_2 + \varphi'(z) \leq 2\gamma\tau_1 - 2\gamma\tau_2 + \alpha < 0$, which means that $g(z)$ is strictly decreasing on $[0, \tau_1]$. Therefore, the minimum point of $g(z)$ must lie within $(\tau_1, \tau_2]$.

Consider solving $g'(z) = 2\gamma(z - \tau_1) + 2\gamma(z - \tau_2) + \varphi'(z) = 0$ on $(\tau_1, \tau_2]$. Calculate $g''(z) = 4\gamma + \varphi''(z) > 0$. And we have $g'(\tau_2) = 2\gamma\tau_2 - 2\gamma\tau_1 + \varphi'(\tau_2) > 0$, $g'(\tau_1) < 0$. Therefore, there exists a unique $\bar{z} \in (\tau_1, \tau_2)$ such that $g'(\bar{z}) = 0$, which is the unique global minimum point of $g(z)$ in R.

For the case that $0 < p < 1$, we need a weaker condition on φ to obtain a similar result as in Lemma 5.

Lemma 6 *Suppose φ is twice continuously differentiable on $[\tau_1^p, \tau_2^p]$ with $\tau_2 > \tau_1 > 0$. There exists $\bar{\gamma} > 0$ such that when $\gamma > \bar{\gamma}$ and $0 < p < 1$, the minimization problem (35) has a unique solution $z^* \in (\tau_1, \tau_2)$.*

Proof First, there exists $\alpha > 0$ such that $0 \leq \varphi'(s) \leq \alpha$ and $-\alpha \leq \varphi''(s) \leq 0$, $\forall s \in [\tau_1^p, \tau_2^p]$. Let $\gamma > \bar{\gamma}$, where

$$\bar{\gamma} = \max\{\frac{2\varphi((\frac{\tau_1+\tau_2}{2})^p)}{(\tau_2 - \tau_1)^2}, \frac{p\alpha\tau_1^{p-1}}{2(\tau_2 - \tau_1)}, \frac{\alpha\tau_1^{2p-2} + \alpha\tau_1^{p-2}}{4}\}.$$

Similar to the analysis in Lemma 5, the minimum point of $g(z)$ must lie within $[0, \tau_2]$. When $z \in [0, \tau_1]$, $g(z) \geq \gamma(\tau_2 - \tau_1)^2$, then by $\gamma > \frac{2\varphi((\frac{\tau_1+\tau_2}{2})^p)}{(\tau_2-\tau_1)^2}$, we have

$$g(z) > g(\frac{\tau_1 + \tau_2}{2}), \, \forall z \in [0, \tau_1].$$

Thus, the minimum point of $g(z)$ must lie with in $(\tau_1, \tau_2]$.

To minimize $g(z)$ on $(\tau_1, \tau_2]$, we check its first derivative. By $\gamma > \frac{p\alpha\tau_1^{p-1}}{2(\tau_2-\tau_1)}$, we have $g'(\tau_1) = 2\gamma(\tau_1 - \tau_2) + p\varphi'(\tau_1^p)\tau_1^{p-1} < 0$, and by $\varphi' \geq 0$, we get $g'(\tau_2) = 2\gamma(\tau_2 - \tau_1) + p\varphi'(\tau_2^p)\tau_2^{p-1} > 0$. Now we consider the solution of the constrained equation

$$g'(z) = 2\gamma(z - \tau_1) + 2\gamma(z - \tau_2) + p\varphi'(z^p)z^{p-1} = 0, \quad z \in (\tau_1, \tau_2].$$

We calculate that $g''(z) = 4\gamma + p^2\varphi''(z^p)z^{2p-2} + p(p-1)\varphi'(z^p)z^{p-2} > 0$ since $\gamma > \frac{\alpha\tau_1^{2p-2}+\alpha\tau_1^{p-2}}{4}$. Combining it with $g'(\tau_1) < 0$ and $g'(\tau_2) > 0$, there exists a unique $\bar{z} \in (\tau_1, \tau_2)$ such that $g'(\bar{z}) = 0$, which is the unique global minimizer of $g(z)$ in R.

Since φ' is locally Lipschitz continuous in R_{++}, φ' is continuously differentiable almost everywhere in R_{++}. If φ is strictly concave in $(\underline{\tau}, \bar{\tau})$ with $\bar{\tau} > \underline{\tau} > 0$, there exist $\tau_1 > 0$ and $\tau_2 > \tau_1$ with $[\tau_1^p, \tau_2^p] \subseteq (\underline{\tau}, \bar{\tau})$ such that φ is strictly concave and twice continuously differentiable on $[\tau_1^p, \tau_2^p]$. Thus, the strict concavity of φ in an open interval of R_+ is sufficient for the existence of $[\tau_1^p, \tau_2^p]$ with $\tau_2 > \tau_1 > 0$ such that φ is strictly concave and twice continuously differentiable on it. And there is no other condition needed to guarantee the supposition of φ in Lemma 6.

Theorem 3 *1. Minimization problem (34) is strongly NP-hard for any given $0 < p < 1$.*
2. If φ is strongly concave in an open interval of R_+, then minimization problem (34) is strongly NP-hard for $p = 1$.

Proof Now we present a polynomial time reduction from the well-known strongly NP-hard partition problem [23] to problem (34). The 3-partition problem can be described as follows: given a multiset S of $n = 3m$ integers $\{a_1, a_2, \dots, a_n\}$ with sum mb, is there a way to partition S into m disjoint subsets $S_1, S_2, \dots, S_m$, such that the sum of the numbers in each subset is equal?

Given an instance of the partition problem with $a = (a_1, a_2, \dots, a_n)^T \in R^n$. We consider the following minimization problem in form (34):

$$\begin{aligned}\min_x \quad P(x) = &\sum_{j=1}^{m} |\sum_{i=1}^{n} \alpha_i x_{ij} - \beta|^2 + \gamma \sum_{i=1}^{n} |\sum_{j=1}^{m} x_{ij} - \tau_1|^2 \\ &+ \gamma \sum_{i=1}^{n} |\sum_{j=1}^{m} x_{ij} - \tau_2|^2 + \sum_{i=1}^{n} (\sum_{j=1}^{m} \varphi(|x_{ij}|^p)),\end{aligned} \tag{36}$$

where the parameters τ_1, τ_2 and γ satisfy the suppositions in Lemma 5 for $p = 1$ and them in Lemma 6 for $0 < p < 1$.

From Lemma 4, we have

$$\begin{aligned}&\min_x P(x) \\ &\geq \min_{x_{ij}} \gamma \sum_{i=1}^{n} |\sum_{j=1}^{m} x_{ij} - \tau_1|^2 + \gamma \sum_{i=1}^{n} |\sum_{j=1}^{m} x_{ij} - \tau_2|^2 + \sum_{i=1}^{n} (\sum_{j=1}^{m} \varphi(|x_{ij}|^p)) \\ &= \sum_{i=1}^{n} \left(\min_{x_{ij}} \gamma |\sum_{j=1}^{m} x_{ij} - \tau_1|^2 + \gamma |\sum_{j=1}^{m} x_{ij} - \tau_2|^2 + \sum_{j=1}^{m} \varphi(|x_{ij}|^p) \right) \\ &\geq \sum_{i=1}^{n} \min_z \gamma |z - \tau_1|^2 + \gamma |z - \tau_2|^2 + \varphi(|z|^p).\end{aligned} \tag{37}$$

By Lemmas 5-6 and the strict concavity of $\varphi(z^p)$ on $[\tau_1, \tau_2]$, we can always choose one of x_{ij} to be $z^*(\neq 0)$ and the others are 0 for any $i = 1, 2, \dots, n$ such that the last inequality in (37) becomes to be an equality and

$$P(x) \geq n g(z^*).$$

		φ_1	φ_4		φ_5			φ_6	
$p=1$	τ_1	none	$(0,\lambda)$		$(\lambda,a\lambda)$			$(0,a\lambda)$	
	τ_2	none	(τ_1,λ)		$(\tau_1,a\lambda)$			$(\tau_1,a\lambda)$	
$0<p<1$	τ_1	$(0,\infty)$	$(0,\lambda)$	(λ,∞)	$(0,\lambda)$	$(\lambda,a\lambda)$	$(a\lambda,\infty)$	$(0,a\lambda)$	$(a\lambda,\infty)$
	τ_2	(τ_1,∞)	(τ_1,λ)	(τ_1,∞)	(τ_1,λ)	$(\tau_1,a\lambda)$	(τ_1,∞)	$(\tau_1,a\lambda)$	(τ_1,∞)

Table 2: Parameters for different potential functions in Remark 3

Now we claim that there exists an equitable partition to the partition problem if and only if the optimal value of (36) equals to $ng(z^*)$. First, if S can be evenly partitioned into m sets, then we define $x_{ik}=z^*$, $x_{ij}=0$ for $j\neq k$ if a_i belongs to S_k. These x_{ij} provide an optimal solution to $P(x)$ with optimal value $ng(z^*)$. On the other hand, if the optimal value of $P(x)$ is $ng(z^*)$, then in the optimal solution, for each i, there is only one element in $\{x_{ij}: 1\leq j\leq m\}$ is nonzero. And we must also have $\sum_{i=1}^n \alpha_i x_{ij}-\beta=0$ holds for any $1\leq j\leq m$, which implies that there exists a partition to set S into m disjoint subsets such that the sum of the numbers in each subset is equal. Thus this theorem is proved.

Remark 3 Many penalty functions satisfy the conditions in Lemma 5 and Lemma 6, such as the logistic penalty function [32], fraction penalty function [32], hard thresholding penalty function [20], SCAD function [20] and MCP function [38]. The soft thresholding penalty function [25,33] only satisfies the conditions in Lemma 6. Here, we list the formulations of these penalty functions below. For φ_2 and φ_3, all choices of τ_1 and τ_2 in R_{++} with $\tau_1<\tau_2$ satisfy the conditions in Lemma 5 and Lemma 6. For the other four penalty functions, the optional parameters of τ_1 and τ_2 are given in Table 2.

- soft thresholding penalty function: $\varphi_1(s)=\lambda s$,
- logistic penalty function : $\varphi_2(s)=\lambda\log(1+as)$,
- fraction penalty function: $\varphi_3(s)=\lambda\frac{as}{1+as}$,
- hard thresholding penalty function: $\varphi_4(s)=\lambda^2-(\lambda-s)_+^2$,
- smoothly clipped absolute deviation (SCAD) penalty function:

$$\varphi_5(s)=\lambda\int_0^s \min\{1,\frac{(a-t/\lambda)_+}{a-1}\}\mathrm{d}t,$$

- minimax concave penalty (MCP) function:

$$\varphi_6(s)=\lambda\int_0^s (1-\frac{t}{a\lambda})_+\mathrm{d}t,$$

with $\lambda>0$ and $a>0$.

3.2 Lower bound theory

In this subsection, we will establish the lower bound theory for the local minimizers of (2) with a special constraint, that is

$$\begin{array}{ll} \min & f(x) := \Theta(x) + \sum_{i=1}^{m} \varphi(\|D_i^T x\|_p^p) \\ \text{s.t.} & x \in \mathcal{X} = \{x : Ax \leq b\}, \end{array} \tag{38}$$

where $D_i = (D_{i1}, \ldots, D_{ir})$ with $D_{ij} \in R^n$, $j = 1, 2, \ldots, r$, $A = (A_1, \ldots, A_q)^T \in R^{q \times n}$ with $A_i \in R^n$, $i = 1, 2, \ldots, q$, and $b = (b_1, b_2, \ldots, b_q)^T \in R^q$.

Denote $\mathcal{M}$ the set of all local minimizers of (38). In this subsection, we suppose that there exists $\beta > 0$ such that $\sup_{x \in \mathcal{M}} \|\nabla^2 \Theta(x)\|_2 \leq \beta$.

For $x \in \mathcal{X}$, let $\mathcal{I}_{ac}(x) = \{i \in \{1, 2, \ldots, q\} : A_i^T x - b_i = 0\}$ be the set of active inequality constraints at x. Then, the Linear Independence Constraint Qualification (LICQ) is satisfied at x, if the set of active constraint gradients $\{A_i : i \in \mathcal{I}_{ac}(x)\}$ is linearly independent.

Theorem 4 *Let $p = 1$ in (38). There exist constants $\theta > 0$ and $\nu_1 > 0$ such that if $|\varphi''(0+)| > \nu_1$, then any local minimizer x^* of (38) meeting the LICQ condition satisfies*

$$\textit{either } \|D_i^T x^*\|_1 = 0 \textit{ or } \|D_i^T x^*\|_1 \geq \theta,\ \forall i \in \{1, 2, \ldots, m\}.$$

Proof We divide $\mathcal{M}$ into the finite disjoint sets $\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_s$ such that all element x in each set have the same following values:
(i) sign values of $\text{sign}(D_{it}^T x)$ for $i = 1, 2, \ldots, m$, $t = 1, 2, \ldots, r$;
(ii) index values of $\mathcal{I}_{ac}(x)$ and $\mathcal{I}_x = \{i \in \{1, 2, \ldots, m\} : D_i^T x = 0\}$.

First, we will prove that there exists $\theta_{1,1} > 0$ and $\kappa_{1,1}$ such that

$$\text{either } \|D_1^T x\|_1 = 0 \text{ or } \|D_1^T x\|_1 \geq \theta_{1,1}, \quad \forall x \in \mathcal{M}_1, \tag{39}$$

when $|\varphi''(0+)| > \beta \kappa_{1,1}$.

Specially, if the values of $\|D_1^T x\|_1$ are same for all $x \in \mathcal{M}_1$, then the statement in (39) holds naturally. In what follows, we suppose that there are at least two elements in $\mathcal{M}_1$ with different values of $\|D_1^T x\|_1$.

Suppose $\bar{x} \in \mathcal{M}_1$ is a local minimizer of minimization problem (38) satisfying $\|D_1^T \bar{x}\| \neq 0$ and meeting the LICQ condition. Then, there exists $\delta > 0$ such that

$$\begin{aligned} f(\bar{x}) &= \min\{\Theta(x) + \sum_{i=1}^{m} \varphi(\|D_i^T x\|_1) : \|x - \bar{x}\|_2 \leq \delta,\ Ax \leq b\} \\ &= \min\{\Theta(x) + \sum_{i \notin \mathcal{I}_{\bar{x}}} \varphi(\|D_i^T x\|_1) : \|x - \bar{x}\|_2 \leq \delta,\ Ax \leq b,\ D_i^T x = 0 \text{ for } i \in \mathcal{I}_{\bar{x}}\}, \end{aligned}$$

which implies that $\bar{x}$ is a local minimizer of the following constrained minimization problem

$$\begin{aligned} \min \quad & f_{\bar{x}}(x) := \Theta(x) + \sum_{i \notin \mathcal{I}_{\bar{x}}} \varphi(\|D_i^T x\|_1) \\ \text{s.t.} \quad & Ax \leq b,\ D_i^T x = 0,\ i \in \mathcal{I}_{\bar{x}}. \end{aligned} \tag{40}$$

Since φ' is locally Lipschitz continuous in R_{++} and the LICQ condition holds at $\bar{x}$, by the second order optimality necessary condition, there exists $\xi_i \in \partial(\varphi'(s))_{s=\|D_i^T \bar{x}\|_1}$ such that

$$v^T \nabla^2 \Theta(\bar{x}) v + \sum_{i \notin \mathcal{I}_{\bar{x}}} \xi_i \left(\sum_{t \in \{1,2,\ldots,r\}} \operatorname{sign}(D_{it}^T \bar{x}) D_{it}^T v \right)^2 \geq 0,\ \forall v \in \mathcal{V}_{\bar{x}}, \tag{41}$$

where

$$\mathcal{V}_{\bar{x}} = \{v : D_j^T v = 0 \,\text{for}\, j \in \mathcal{I}_{\bar{x}} \,\text{and}\, A_k^T v = 0 \,\text{for}\, k \in \mathcal{I}_{ac}(\bar{x})\}. \tag{42}$$

By $\xi_i \leq 0$, $i = 1, 2, \ldots, m$, (41) gives

$$-\xi_1 \left(\sum_{t \in \{1,2,\ldots,r\}} \operatorname{sign}(D_{1t}^T \bar{x}) D_{1t}^T v \right)^2 \leq \|\nabla^2 \Theta(\bar{x})\|_2 \|v\|_2^2, \quad \forall v \in \mathcal{V}_{\bar{x}}. \tag{43}$$

Fix $x \in \mathcal{M}_1$. For $c \in R^r$ with $c_i \in \{-1, 0, 1\}$, consider the following constrained convex minimization problem

$$\begin{aligned} \min \quad & \|v\|_2^2 \\ \text{s.t.} \quad & v \in \mathcal{V}_{x,c} = \{v : \sum_{t \in \{1,\ldots,r\}} c_t D_{1t}^T v = 1 \,\text{and}\, v \in \mathcal{V}_x\}. \end{aligned} \tag{44}$$

When $\mathcal{V}_{x,c} \neq \emptyset$, unique existence of the optimal solution of (44) is guaranteed, denoted by $v_{x,c}$. Take all possible choices of nonzero vector $c \in R^r$ with $c_i \in \{-1, 0, 1\}$ such that $\mathcal{V}_{x,c} \neq \emptyset$, which are finite, and we define

$$\kappa_{1,1} = \max \|v_{x,c}\|_2^2,$$

which is a positive number and same for all elements in $\mathcal{M}_1$ from the decomposition method for $\mathcal{M}$.

Since there is another element in $\mathcal{M}_1$, denoted as $\hat{x}$, such that $\|D_1^T \bar{x}\|_1 \neq \|D_1^T \hat{x}\|_1$, then

$$\tilde{v} = \frac{1}{\|D_1^T \bar{x}\|_1 - \|D_1^T \hat{x}\|_1} (\bar{x} - \hat{x}) \in \mathcal{V}_{\bar{x},c}.$$

Thus, the unique solution of (44) exists in this case and (43) holds with it, which follows

$$-\xi_1 \leq \beta \kappa_{1,1}.$$

If $|\varphi''(0+)| > \beta\kappa_{1,1}$, let

$$\theta_{1,1} = \inf\{t > 0 : \varphi''(t) \text{ exists and } \varphi''(t) \geq \beta\kappa_{1,1}\}, \tag{45}$$

by the upper semicontinuity of $\partial(\varphi'(t)))$ on R_{++}, we obtain that

$$\|D_1^T \bar{x}\|_1 \geq \theta_{1,1}.$$

By the randomicity of $\bar{x} \in \mathcal{M}_1$ satisfying $\|D_1^T \bar{x}\|_1 \neq 0$ in the above analysis, (43) implies

$$-\xi_1 \left(\sum_{t \in \{1,2,\ldots,r\}} \operatorname{sign}(D_{1t}^T x) D_{1t}^T v \right)^2 \leq \|\nabla^2 \Theta(x)\|_2 \|v\|_2^2, \quad \forall v \in \mathcal{V}_x,$$

holds for any $x \in \mathcal{M}_1$ satisfying $\|D_1^T x\|_1 \neq 0$. Since $\kappa_{1,1}$ is same for all elements in $\mathcal{M}_1$, the statement in (39) holds.

Similarly, for any $i = 1, \ldots, m$, $j = 1, \ldots, s$, there exists $\theta_{i,j} > 0$ and $\kappa_{i,j} > 0$ such that

$$\text{either } \|D_i^T x\|_1 = 0 \text{ or } \|D_i^T x\|_1 \geq \theta_{i,j}, \quad \forall x \in \mathcal{M}_j,$$

when $|\varphi''(0+)| > \beta\kappa_{i,j}$.

Therefore, we can complete the proof for this theorem with $\nu_1 = \max\{\beta\kappa_{i,j} : i = 1, \ldots, m, j = 1, \ldots, s\}$ and $\theta = \min\{\theta_{i,j} : i = 1, \ldots, m, j = 1, \ldots, s\}$.

If there exists constant $\nu_1 > 0$ such that $|\varphi''(0+)| \geq \nu_1$, by the concavity of φ and $\varphi' \geq 0$, there must exist $\nu_p > 0$ such that $\varphi'(0+) \geq \nu_p$. However, the converse does not hold. The following theorem presents the lower bound theory for the case that $0 < p < 1$ using the existence of $\nu_p > 0$ such that $\varphi'(0+) \geq \nu_p$.

Theorem 5 *Let $0 < p < 1$ in (38). If there exists $\nu_p > 0$ such that $\varphi'(0+) \geq \nu_p$, then there exists a constant $\theta > 0$ such that any local minimizer x^* of (38) meeting the LICQ condition satisfies*

$$\textit{either } \|D_i^T \bar{x}\|_p = 0 \textit{ or } \|D_i^T \bar{x}\|_p \geq \theta, \ \forall i \in \{1, 2, \ldots, m\}.$$

Proof We divide $\mathcal{M}$ by the method in Theorem 4 and we will also prove that there exists $\theta_{1,1} > 0$ such that

$$\text{either } \|D_1^T x\|_p = 0 \text{ or } \|D_1^T x\|_p \geq \theta_{1,1}, \quad \forall x \in \mathcal{M}_1. \tag{46}$$

Specially, if the values of $\|D_1^T x\|_p$ are same for all $x \in \mathcal{M}_1$, then the statement in (46) holds naturally. In what follows, we also suppose that there are at least two elements in $\mathcal{M}_1$ with different values of $\|D_1^T x\|_p$.

Similar to the analysis in Theorem 4, $\bar{x}$ is a local minimizer of minimization problem (38) satisfying $\|D_1^T\bar{x}\|_p \neq 0$ implies that $\bar{x}$ is a local minimizer of the minimization problem

$$\begin{aligned} \min \quad & f_{\bar{x}}(x) := \Theta(x) + \sum_{i \notin \mathcal{I}_{\bar{x}}} \varphi(\|D_i^T x\|_p^p) \\ \text{s.t.} \quad & Ax \leq b,\ D_i^T x = 0,\ i \in \mathcal{I}_{\bar{x}}. \end{aligned} \tag{47}$$

By the second order optimality necessary condition for the minimizers of (47), there exists $\xi_i \in \partial(\varphi'(s))_{s=\|D_i^T\bar{x}\|_p^p}$ such that

$$\begin{aligned} & v^T\nabla^2\Theta(\bar{x})v + \sum_{i \notin \mathcal{I}_{\bar{x}}} \xi_i \left(\sum_{t \in T_i} p|D_{it}^T\bar{x}|^{p-1}\mathrm{sign}(D_{it}^T\bar{x})D_{it}^T v \right)^2 \\ & + \sum_{i \notin \mathcal{I}_{\bar{x}}} p(p-1)\varphi'(s)_{s=\|D_i^T\bar{x}\|_p^p} \left(\sum_{t \in T_i} |D_{it}^T\bar{x}|^{p-2}(D_{it}^T v)^2 \right) \geq 0,\ \forall v \in \mathcal{V}_{\bar{x}}, \end{aligned}$$

where $\mathcal{V}_{\bar{x}}$ is same as in (42) and $T_i = \{t \in \{1,2,\ldots,r\} : D_{it}^T\bar{x} \neq 0\}$, $i = 1,2,\ldots,m$. Then, by $\xi_i \leq 0$, $\forall i = 1,2,\ldots,m$ and $\|D_1^T\bar{x}\|_p \neq 0$, we obtain

$$p(1-p)\varphi'(s)_{s=\|D_1^T\bar{x}\|_p^p} \left(\sum_{t \in T_i} |D_{1t}^T\bar{x}|^{p-2}(D_{1t}^T v)^2 \right) \leq v^T\nabla^2\Theta(\bar{x})v, \forall v \in \mathcal{V}_{\bar{x}}. \tag{48}$$

Fix $x \in \mathcal{M}_1$. For $t \in T_1$, consider the following constrained convex optimization

$$\begin{aligned} \min \quad & \|v\|_2^2 \\ \text{s.t.} \quad & v \in \mathcal{V}_{x,t} = \{v : D_{1t}^T v = 1 \,\text{and}\, v \in \mathcal{V}_x\}. \end{aligned} \tag{49}$$

When $\mathcal{V}_{x,t} \neq \emptyset$, unique existence of the optimal solution of (49) is guaranteed, denoted by $v_{x,t}$. Take all possible choices of $t \in T_1$ such that $\mathcal{V}_{x,c} \neq \emptyset$, which are finite, and we define

$$\kappa_{1,1} = \max \|v_{x,t}\|_2^2,$$

which is also a positive number same for all elements in $\mathcal{M}_1$.

Since there is another element in $\mathcal{M}_1$, denoted as $\hat{x}$, such that $\|D_1^T\bar{x}\|_p \neq \|D_1^T\hat{x}\|_p$. Then, there exists $t_1 \in \{1,2,\ldots,r\}$ such that $D_{1t_1}^T\bar{x} \neq D_{1t_1}^T\hat{x}$. Thus,

$$\tilde{v} = \frac{1}{D_{1t_1}^T\bar{x} - D_{1t_1}^T\hat{x}}(\bar{x} - \hat{x}) \in \mathcal{V}_{\bar{x},t_1},$$

which follows the existence of the unique solution of (49) exists with $t = t_1$, denoted as $v_{\bar{x},t_1}^*$.

By the decomposition method for $\mathcal{M}$, we have $\mathrm{sign}(D_{1t}^T\bar{x}) = \mathrm{sign}(D_{1t}^T\hat{x})$, which implies $t_1 \in T_1$. Let $v = v_{\bar{x},t_1}^*$ in (48), by $\varphi' \geq 0$, we have

$$p(1-p)\varphi'(s)_{s=\|D_1^T\bar{x}\|_p^p}|D_{1t_1}^T\bar{x}|^{p-2} \leq \beta\kappa_{1,1}. \tag{50}$$

$|D_{1t_1}^T \bar{x}| \leq \|D_1^T \bar{x}\|_p$ implies $|D_{1t_1}^T \bar{x}|^{p-2} \geq \|D_1^T \bar{x}\|_p^{p-2}$, then (50) gives

$$p(1-p)\varphi'(s)_{s=\|D_1^T \bar{x}\|_p^p} \|D_1^T \bar{x}\|_p^{p-2} \leq \beta\kappa_{1,1}. \tag{51}$$

By the concavity of φ, $\lim_{t\to\infty} \varphi'(s)_{s=t^p} t^{p-2} \leq \lim_{t\to\infty} \varphi'(1) t^{p-2} = 0$. From $\varphi'(0+) \geq \nu_2$, $\lim_{t\downarrow 0} \varphi'(t^p)t^{p-2} = +\infty$. Let

$$\theta_{1,1} = \inf\{t > 0 : \varphi'(t^p)t^{p-2} = \frac{\beta\kappa_{1,1}}{p(1-p)}\},$$

which is an existent number larger than 0. Therefore, (51) implies

$$\|D_1^T \bar{x}\|_p \geq \theta_{1,1}.$$

Similar to the analysis in Theorem 4, the statement in this theorem holds.

Remark 4 For the other cases, such as the regularization term is given by $\sum_{i=1}^m \varphi_i(\max\{d_i^T x, 0\}^p)$ with $d_i \in R^n$, the lower bound theory in Theorems 4-5 can also be guaranteed under the same conditions. Moreover, the lower bound theories in Theorems 4-5 can also be extended to the more general case with the objective function $f(x) := \Theta(x) + \sum_{i=1}^m \varphi_i(\|D_i^T x\|_p^p)$.

All the potential functions in Remark 3 satisfy the conditions in Theorem 5, but only φ_2, φ_3, φ_4 and φ_6 may meet the conditions in Theorem 4 under some conditions on the parameters, which shows the superiority of the non-Lipschitz regularization in sparse reconstruction.

4 Final remarks

In Theorem 1, we derive a first order necessary optimality condition for local minimizers of problem (1) based on the new generalized directional derivative (12) and the Clarke tangent cone. The generalized stationary point that satisfies the first order necessary optimality condition is a Clarke stationary point when the objective function f is locally Lipschitz continuous near this point, and a scaled stationary point if f is non-Lipschitz at the point. Moreover, in Theorem 2 we establish the directional derivative consistency associated with smoothing functions and in Corollary 1 we show that the consistency guarantees the convergence of smoothing algorithms to a stationary point of problem (1). Computational complexity and lower bound theory of problem (1) are also studied to illustrate the negative and positive news of the concave penalty function in applications.

References

1. Aubin J.P., Cellina A.: Differential Inclusion: Set-Valued Maps and Viability Theory. Springer-Verlag, Berlin (1984)
2. Audet C., Dennis Jr. J.E.: Mesh adaptive direct search algorithms for constrained optimization. SIAM J. Optim. **17**, 188-217 (2006)

3. Auslender A.: How to deal with the unbounded in optimization: theory and algorithms. Math. Program. **79**, 3-18 (1997)
4. Bertsekas D.P.: On the Goldstein-Levitin-Polyak gradient projection method. IEEE Trans. Autom. Control **AC-21**, 174-184 (1976)
5. Bian W., Chen X.: Worst-case complexity of smoothing quadratic regularization methods for non-Lipschitzian optimization. SIAM J. Optim. **23**, 1718-1741 (2013)
6. Bian W., Chen X., Ye Y.: Complexity analysis of interior point algorithms for non-Lipschitz and nonconvex minimization. to appear in Math. Program. (2014)
7. Borwein J.M., Lewis A.S.: Convex analysis and nonliear optimization. Second edition, Springer, New York (2006)
8. Bruckstein A.M., Donoho D.L., Elad M.: From sparse solutions of systems of equations to sparse modeling of signals and images. SIAM Review **51**, 34-81 (2009)
9. Burke J.V., Hoheisel T.: Epi-convergent smoothing with applications to convex composite functions. SIAM J. Optim. **23**, 1457-1479 (2013)
10. Burke J.V., Hoheisel T., Kanzow C.: Gradient consistency for integral-convolution smoothing functions. Set-Valued Var. Anal. **21**, 359-376 (2013)
11. Chan R.H., Chen K.: A multilevel algorithm for simultaneously denoising and deblurring images. SIAM J. Sci. Comput. **32**, 1043-1063 (2010)
12. Chartrand R., Staneva V.: Restricted isometry properties and nonconvex compressive sensing. Inverse Probl. **24**, 1-14 (2008)
13. Chen X.: Smoothing methods for nonsmooth, nonconvex minimization. Math. Program. **134**, 71-99 (2012)
14. Chen X., Ge D., Wang Z., Ye Y.: Complexity of unconstrained L_2-L_p minimization. Math. Program. **143**, 371-383 (2014)
15. Chen X., Niu L., Yuan Y.: Optimality conditions and smoothing trust region Newton method for non-Lipschitz optimization. SIAM J. Optim. **23**, 1528-1552 (2013)
16. Chen X., Ng M., Zhang C.: Nonconvex l_p regularization and box constrained model for image restoration. IEEE Trans. Image Processing **21**, 4709-4721 (2012)
17. 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**, 2832-2852 (2010)
18. Clarke F.H.: Optimization and Nonsmooth Analysis. John Wiley, New York (1983)
19. Curtis F.E., Overton M.L.: A sequential quadratic programming algorithm for nonconvex, nonsmooth constrained optimization. SIAM J. Optim. **22**, 474-500 (2012)
20. Fan J.: Comments on 'Wavelets in statistics: a review' by A. Antoniadis. Stat. Method. Appl. **6**, 131-138 (1997)
21. Fan J., Li R.: Variable selection via nonconcave penalized likelihood and its oracle properties. J. Amer. Statist. Assoc. **96**, 1348-1360 (2001)
22. Fan J.Q., Peng H.: Nonconcave penalized likelihood with a diverging number of parameters. Ann. Stat. **32**, 928-961 (2004)
23. Garey M.R., Johnson D.S.: Computers and Intractability: A Guide to the Theory of NP-Completeness. W. H. Freeman & Co Ltd, New York (1979)
24. Ge D., Jiang X., Ye Y.: A note on the complexity of L_p minimization. Math. Program. **21**, 1721-1739 (2011)
25. Huang J., Horowitz J.L., Ma S.: Asymptotic properties of bridge estimators in sparse high-dimensional regression models. Ann. Statist. **36**, 587-613 (2008)
26. Huber P.: Robust Estimation. Wiley, New York (1981)
27. Huang J., Ma S., Xue H., Zhang C.: A group bridge approach for variable selection. Biometrika **96**, 339-355 (2009)
28. Jahn J.: Introduction to the Nonlinear Optimization. Springer-Verlag, Berlin (1996)
29. Knight K., Fu W.J.: Asymptotics for lasso-type estimators. Ann. Stat. **28**, 1356-1378 (2000)
30. Levitin E.S., Polyak B.T.: Constrained minimization problems. USSR. Comput. Math. Math. Phys. **6**, 1-50 (1966)
31. Liu Y.F., Dai Y.H., Ma S.: Joint power and admission control via linear programming deflation. IEEE Trans. Signal Processing **61**, 1327-1338 (2013)
32. 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**, 2-25 (2008)

33. Tibshirani R.: Shrinkage and selection via the Lasso. J. Roy. Statist. Soc. Ser. B **58**, 267-288 (1996)
34. Nocedal J., Wright S.J.: Numerical Optimization. Springer, New York (2006)
35. Rockafellar R.T., Wets R.J-B: Variational Analysis. Springer, Berlin (1998)
36. Vazirani V.: Approximation Algorithms. Springer, Berlin (2003)
37. Ye Y.: Interior Point Algorithms: Theory and Analysis. John Wiley & Sons, Inc., New York (1997)
38. Zhang C.-H: Nearly unbiased variable selection under minimax concave penalty. Ann. Statist. **38**, 894-942 (2010)