Optimization Letters (2019) 13:1027–1038
https://doi.org/10.1007/s11590-018-1310-6

ORIGINAL PAPER

Nonsmooth sparsity constrained optimization problems: optimality conditions

N. Movahedian[1] · S. Nobakhtian[1,2] · M. Sarabadan[1]

Received: 16 December 2017 / Accepted: 3 August 2018 / Published online: 9 August 2018

Abstract
This paper concerns a nonsmooth sparsity constrained optimization problem. We present first and second-order necessary and sufficient optimality conditions by using the concept of normal and tangent cones to the sparsity constraint set. Moreover, second-order tangent set to the sparsity constraint is described and then a new second-order necessary optimality condition is established. The results are illustrated by several examples.

Keywords Sparsity constrained optimization · Tangent cone · Normal cone · N-stationary · Optimality condition · Second-order tangent set

1 Introduction

Sparsity constrained optimization (SCO) is to minimize a general nonlinear function subject to a sparsity constraint set. There are numerous applications in which sparse solutions are concerned, for instance in signal and image processing, denoising, model selection, machine learning and more [8,9,13,21]. The SCO is really a combinatorial optimization problem and it is usually NP-hard even for problems with quadratic objective function [14]. Therefore, the classical optimization theory is not effective for SCO and study on the existence of solutions of SCO is difficult. In recent years, sparse optimization problems have drawn significant attentions in the theories and algorithms

✉ N. Movahedian
n.movahedian@sci.ui.ac.ir

S. Nobakhtian
nobakht@math.ui.ac.ir

M. Sarabadan
msarabadan@sci.ui.ac.ir

1 Department of Mathematics, University of Isfahan, P. O. Box: 81745-163, Isfahan, Iran

2 School of Mathematics, Institute for Research in Fundamental Sciences (IPM), P.O. Box: 19395-5746, Tehran, Iran

[2,5,22]. Beck and Eldar in [4] established first-order necessary optimality conditions for smooth SCO. These conditions are then used to derive some numerical algorithms aimed at finding points satisfying the resulting optimality criteria. In [16] the concepts of tangent and normal cones to the sparse set are used to introduce a number of stationary notions for smooth SCO. The first and second-order optimality conditions together with the relation between these stationary notions are also established.

In [11], the authors studied necessary optimality conditions of a general nonlinear sparsity optimization problem under the Robinson's constraint qualification, and proposed penalty decomposition methods for solving this problem. A characterization of the second-order tangent set to the sparsity set is stated in [15] and the second-order optimality conditions are presented in the smooth case.

In this paper, we present necessary and sufficient optimality conditions for a nonsmooth sparsity constrained optimization problem. We first extend some of the stationary notions to the nonsmooth case in terms of the Clarke generalized gradient. Then the relationship between these notions are discussed. These concepts are used to drive the first-order necessary and sufficient optimality conditions of SCO problems. Despite of the involved structure of the sparsity constraint set, only mild generalized convexity assumptions are considered in our first-order sufficient optimality result. We also apply the notions of the first and second-order Dini directional derivatives together with the Bouligand tangent cone to the sparsity constraint set to establish the second-order necessary and sufficient optimality conditions. Furthermore, the second-order tangent set to the sparsity constraint set is characterized and used to drive a new sharp second-order necessary optimality condition for SCO.

This paper is organized as follows. In Sect. 2, we introduce the definitions and notations to be used throughout the paper. Section 3 studies the first and second-order optimality conditions for sparsity constrained optimization problems. Section 4 expresses the second-order tangent set for SCO and gives a second-order necessary optimality condition for sparsity constrained optimization. Moreover, some examples are provided to clarify our results.

2 Preliminaries

All the definitions quoted in this section are taken from [7,18,20], where the reader can find more details, discussions and references.

Throughout this work, $\mathbb{R}^n$ is the usual n-dimensional Euclidean space. Let S be a nonempty subset of $\mathbb{R}^n$, the closure of S is denoted by cl S. For a given subset $J \subseteq \{1, \ldots, n\}$, denote by $\text{span}\{e_i,\ i \in J\}$ the subspace of $\mathbb{R}^n$ spanned by $\{e_i,\ i \in J\}$, where $e_i \in \mathbb{R}^n$ is a vector whose ith component is one and others are zeros.
Let $f : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitzian function (i.e., a function satisfying the Lipschitz condition in a neighbourhood of any point $x \in \mathbb{R}^n$). The Clarke directional derivative of f at x in the direction v, is defined as follows:

$$f^\circ(x; v) := \limsup_{y \to x\, t \downarrow 0} \frac{f(y + tv) - f(y)}{t}. \tag{1}$$

The function $f^\circ(x; .) : \mathbb{R}^n \to \mathbb{R}$ is sublinear. The Clarke generalized gradient of f at x is defined as

$$\partial_c f(x) := \{\xi \in \mathbb{R}^n | \langle \xi; v\rangle \leq f^\circ(x; v) \quad \forall v \in \mathbb{R}^n\}. \tag{2}$$

It is well-known that $\partial_c f(x)$ is a nonempty, convex and compact subset of $\mathbb{R}^n$.
The lower and upper Dini derivatives of f at x in the direction $v \in \mathbb{R}^n$ are given respectively, as

$$D^- f(x; v) := \liminf_{t\downarrow 0} \frac{f(x+tv) - f(x)}{t}, \tag{3}$$

$$D^+ f(x; v) := \limsup_{t\downarrow 0} \frac{f(x+tv) - f(x)}{t}. \tag{4}$$

It is worth mentioning that if f is locally Lipschitzian, then both the lower and upper Dini derivatives exist finitely. To derive our second-order necessary and sufficient optimality results, we need to recall the definition of the second-order Dini directional derivative from [20]. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitzian function and $x, u, v \in \mathbb{R}^n$. Assume that $D^- f(x; u)$ exists. The second-order Dini directional derivative of f at (x, u) in the direction v is defined by

$$D^2 f(x, u, v) := \liminf_{t\downarrow 0} \frac{f(x + tu + t^2 v) - f(x) - tD^- f(x; u)}{t^2}.$$

For stationary results that are given in the next section we need to use the concept of the Clarke partial generalized gradient, which is defined as follows [7].
Let $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be Lipschitzian near $(x_1, x_2) \in \mathbb{R}^n \times \mathbb{R}^m$. The notation $\partial_c^{(1)} f(x_1, x_2)$ denotes the Clarke partial generalized gradient evaluated at x_1, i.e., the generalized gradient of the function $\acute{x_1} \mapsto f(\acute{x_1}, x_2)$ defined on $\mathbb{R}^n$.

Recalling that for any nonempty set $S \subseteq \mathbb{R}^n$, the Bouligand tangent cone $T_S^B(x)$ and its corresponding normal cone $N_S^B(x)$ at $x \in \operatorname{cl} S$ are defined, respectively, by

$$T_S^B(x) := \left\{ \lim_{k\to\infty} \frac{x_k - x}{t_k} : x_k \xrightarrow{S} x, t_k \downarrow 0 \right\},$$
$$N_S^B(x) := \{d \in \mathbb{R}^n : \langle d, z\rangle \leq 0, \forall z \in T_S^B(x)\},$$

where $x_k \xrightarrow{S} x$ means that $x_k \in S$ for each $k = 1, 2, \ldots$, and $\lim_{k\to\infty} x_k = x$.
Also the Clarke tangent cone $T_S^C(x)$ and its corresponding normal cone $N_S^C(x)$ at $x \in \operatorname{cl} S$ are given as:

$$T_S^C(x) = \left\{ d \in \mathbb{R}^n : \forall x_k \xrightarrow{S} x, t_k \downarrow 0, \exists d_k \to d, \text{ such that } x_k + t_k d_k \in S \right\},$$
$$N_S^C(x) := \{d \in \mathbb{R}^n : \langle d, z\rangle \leq 0, \forall z \in T_S^C(x)\}.$$

For any function $f : \mathbb{R}^n \to \mathbb{R}$, the kernel (also called the null space) is defined by $\ker(f) = \{x : x \in \mathbb{R}^n \text{ such that } f(x) = 0\}$.

To establish new sufficient optimality conditions for SCO problems, a suitable generalized convexity notion is required. Thus, following the pattern in [1], we give the definition of ∂_c-pseudoconvex functions.
Let $f : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitzian function, f is said to be ∂_c-pseudoconvex at $\bar{x}$ on $S \subseteq \mathbb{R}^n$ if for each $x \in S$ that $x \neq \bar{x}$ and $f(x) < f(\bar{x})$ we have

$$\langle \xi, x - \bar{x} \rangle < 0 \quad \forall \xi \in \partial_c f(\bar{x}).$$

Note that any convex function is also ∂_c-pseudoconvex. The following theorems from [12], shows that ∂_c-pseudoconvexity is a natural extension of pseudoconvexity.

Theorem 1 *If f is smooth, then f is ∂_c-pseudoconvex, if and only if f is pseudoconvex.*

The important sufficient extremum property of pseudoconvexity remains also for ∂_c-pseudoconvexity.

Theorem 2 *An ∂_c-pseudoconvex f attains its global minimum at x^*, if and only if*

$$0 \in \partial_c f(x^*).$$

The following example from [12], shows that ∂_c-pseudoconvexity is a more general property than pseudoconvexity.

Example 1 Consider $f : \mathbb{R} \to \mathbb{R}$ such that $f(x) := min\{|x|, x^2\}$. Then f is clearly locally Lipschitzian continuous but not convex nor pseudoconvex. However, for all $y > x$ we have

$$f^\circ(x; y - x) = \begin{cases} -1, & x \in (-\infty, -1]; \\ 2x, & x \in (-1, 1]; \\ 1, & x \in (1, \infty), \end{cases}$$

and thus, f is ∂_c-pseudoconvex.

In the sequel, let us give the definition of pseudoconvex set due to [10]. Suppose that $S \subset \mathbb{R}^n$ and $K_S(x)$ is a tangent cone to S at $x \in \text{cl}\, S$. The set S is said to be pseudoconvex with respect to $K_S(x)$ at x if

$$S \subset x + K_S(x).$$

It is obvious that each convex set is pseudoconvex with respect to the classical tangent cone at each of its points.
Let f be a real-valued function defined on the set S. The sublevel set of f at $x \in S$ is given by

$$L(x) = L(f; x; S) := \{y \in S |\ f(y) \leq f(x)\}.$$

We say that f admits pseudoconvex sublevel sets with respect to the tangent cone $K_{L(x)}(x)$ if its sublevel sets are pseudoconvex with respect to $K_{L(x)}(x)$ at each point of $x \in S$, that is

$$L(x) \subset x + K_{L(x)}(x) \quad \forall x \in S.$$

To develop the second-order necessary conditions for SCO problems, we need to use the notion of second-order tangent set from [19].
The vector w is called a second-order tangent to $S \subseteq \mathbb{R}^n$ at $(x^*, d) \in S \times \mathbb{R}^n$ if there exist sequences $x_k \xrightarrow{S} x^*$ and $t_k \downarrow 0$ such that

$$w = \lim_{k\to\infty} \frac{x_k - x^* - t_k d}{\frac{1}{2}(t_k)^2}.$$

The set of all second-order tangents to S at (x^*, d) is denoted by $T_S^2(x^*, d)$. It is easy to see that $T_S^2(x^*, d) = \emptyset$ if $d \notin T_S^B(x^*)$ (see [19]). In general, the second-order tangent set is not a cone, and it is not necessarily convex, even for a convex set S.

3 Optimality conditions

In this section, we study the first and second-order necessary and sufficient optimality conditions of the following sparsity constrained optimization problem

$$\begin{array}{lll} (P) & \min & f(x) \\ & \text{s.t.} & \|x\|_0 \leq s, \end{array}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a locally Lipschitzian function and $\|x\|_0$ is the l_0-norm of $x \in \mathbb{R}^n$, which refers to the number of nonzero elements in the vector x and $s < n$ is a positive integer. Let $S := \{x \in \mathbb{R}^n : \|x\|_0 \leq s\}$ be the feasible region of (P) and $J^* = J(x^*) := \text{supp}\,(x^*) = \{i \in \{1, ..., n\} : x_i^* \neq 0\}$.

For this purpose we first need to recall the following auxiliary results from [16], which compute the Bouligand and Clarke tangent and normal cones of the sparse set S.

Theorem 3 *For any $x^* \in S$, the Bouligand tangent and normal cones of S at x^* are respectively,*

$$\begin{aligned} T_S^B(x^*) &= \{d \in \mathbb{R}^n : \|d\|_0 \leq s, \ \|x^* + \mu d\|_0 \leq s, \ \forall \mu \in \mathbb{R}\} \\ &= \bigcup_{\gamma} \text{span}\,\{e_i, i \in \gamma \supseteq J^*, \ |\gamma| \leq s\}, \end{aligned} \tag{5}$$

$$N_S^B(x^*) = \begin{cases} \{d \in \mathbb{R}^n : d_i = 0, \ i \in J^*\} = \text{span}\,\{e_i, \ i \notin J^*\} & \textit{if } |J^*| = s \\ \{0\} & \textit{if } |J^*| < s. \end{cases} \tag{6}$$

Theorem 4 *For any $x^* \in S$, the Clarke tangent and normal cones of S at x^* are respectively,*

$$T_S^C(x^*) = \{d \in \mathbb{R}^n : supp(d) \subseteq J^*\} = \text{span}\{e_i,\ i \in J^*\}, \tag{7}$$
$$N_S^C(x^*) = \text{span}\{e_i,\ i \notin J^*\}. \tag{8}$$

3.1 First-order optimality conditions

Our first-order optimality conditions are based on the nonsmooth versions of the stationary concepts given in [4,16].

Definition 1 Consider the feasible point $x^* \in S$:

1. x^* is called an $N^\sharp$−stationary point of (P), if

$$0 \in \partial_c f(x^*) + N_S^\sharp(x^*),$$

   where $\sharp \in \{B, C\}$.
2. x^* is said to be a basic feasible (BF) point of (P) if
   a. when $\|x^*\|_0 < s, \quad 0 \in \partial_c f(x^*)$;
   b. when $\|x^*\|_0 = s, \quad 0 \in \partial_c^{J^*} f(x^*)$,

where the meaning of the $\partial_c^{J^*} f(x^*)$, is the Clarke partial generalized gradient subject to the index set J^*.

The first result of this section presents a necessary optimality condition with respect to the N^B−stationary points.

Theorem 5 *Assume that x^* is a local optimal solution of (P) and $\partial_c f(x^*) \subseteq T_S^C(x^*)$. Then x^* is an N^B−stationary point.*

Proof From [6, Corollary, 6.3.9] the local optimality of x^* implies immediately that

$$0 \in \partial_c f(x^*) + N_S^C(x^*). \tag{9}$$

In the case that $\|x^*\|_0 = s$, the result follows from (6) and (8). It remains to consider the other case when $\|x^*\|_0 < s$.
The inclusion in (9) gives us a vector $\bar{\xi} \in \partial_c f(x^*) \cap (-N_S^C(x^*))$. On the other hand, since $\partial_c f(x^*) \subseteq T_S^C(x^*)$, one has $\bar{\xi} \in T_S^C(x^*) \cap (-N_S^C(x^*))$. Applying Theorem 4, we get $\langle \bar{\xi}, \bar{\xi} \rangle = 0$, and thus $\bar{\xi} = 0 \in \partial_c f(x^*)$ which completes the proof of the theorem. □

Next let us prove the equivalence between the N^B-stationary and basic feasibility in the nonsmooth case.

Lemma 1 *Consider the feasible point $x^* \in S$. Then x^* is an N^B−stationary point of problem (P) if and only if x^* is a basic feasible point.*

Proof In the case that $\|x^*\|_0 < s$, the result is trivial. Thus it is sufficient to consider the other case when $\|x^*\|_0 = s$. Assuming that x^* is an N^B-stationary point of (P), then one has by (6)

$$0 \in \partial_c f(x^*) + \text{span}\,\{e_i,\ i \notin J^*\}.$$

Thus there exists $\bar{\xi} \in \partial_c f(x^*)$ such that $\bar{\xi}_i = 0$ for each $i \in J^*$, and consequently, $0 \in \partial_c^{J^*} f(x^*)$.

Conversely, the basic feasibility of x^* implies that $0 \in \partial_c^{J^*} f(x^*)$. Using the definition of the partial generalized gradient subject to the index set J^*, gives us a point $\xi_2 \in \mathbb{R}^{n-|J^*|}$ such that $(0, \xi_2) \in \partial_c f(x^*)$. Putting the above together with (6), we conclude that $0 \in \partial_c f(x^*) + N_S^B(x^*)$ and complete the proof of lemma. □

The next corollary follows immediately from Theorem 5 and Lemma 1.

Corollary 1 *Assume that x^* is a local optimal solution of problem (P) and $\partial_c f(x^*) \subseteq T_S^C(x^*)$. Then x^* is a basic feasible point.*

Now we turn our attention to the nonsmooth first-order sufficient optimality conditions for (P) with respect to the pseudoconvex sets.

Theorem 6 *Let $x^* \in S$ be an N^B-stationary point of problem (P). Suppose also that f is ∂_c-pseudoconvex on S at x^* and $L(f; x^*; S)$ is pseudoconvex with respect to the Bouligand tangent cone. Then x^* is a global minimum of problem (P).*

Proof Suppose on the contrary that for a feasible point $x \neq x^*$, one has $f(x) < f(x^*)$. By the ∂_c-pseudoconvexity of f at x^* we have for each $\xi \in \partial_c f(x^*)$,

$$\langle \xi, x - x^* \rangle < 0. \tag{10}$$

The N^B-stationarity of x^* together with Theorem 3 gives us a vector $\bar{\xi} \in \partial_c f(x^*) \cap (-N_S^B(x^*))$ such that $\langle \bar{\xi}, v \rangle = 0$ for each $v \in T_S^B(x^*)$.

On the other hand, since $f(x) < f(x^*)$ we get $x \in L(f; x^*; S)$ and the pseudoconvexity of $L(f; x^*; S)$, implies that $x - x^* \in T^B_{L(x^*)}(x^*)$. Obviously $L(x^*) \subseteq S$, and thus $T^B_{L(x^*)}(x^*) \subseteq T_S^B(x^*)$ hence $x - x^* \in T_S^B(x^*)$. Therefore we have $\langle \bar{\xi}, x - x^* \rangle = 0$, which contradicts (10). □

The following example illustrates the sufficient optimality result given in Theorem 6.

Example 2 Consider problem (P) where $f(x) = f(x_1, x_2) := |x_2| - |x_1 x_2|$, and $S := \{x = (x_1, x_2) \in \mathbb{R}^2 \mid \|x\|_0 \leq 1\}$. Clearly $S = \mathbb{R} \times \{0\} \cup \{0\} \times \mathbb{R}$ and $x^* = (0, 0)$ is a global optimal solution for this problem. We have $f^\circ(x^*; v) = |v_2|$, where $v \in \mathbb{R}^2$, $\partial_c f(x^*) = \{0\} \times [-1, 1]$ and $N_S^B(x^*) = \{0\}$. Hence x^* is an N^B-stationary point. It is easy to show that f is ∂_c-pseudoconvex on S at x^*. Also one can get

$$L(x^*) = \{y \in S \mid f(y) \leq f(x^*) = 0\} = \mathbb{R} \times \{0\},$$

and therefore $T^B_{L(x^*)}(x^*) = \mathbb{R} \times \{0\}$, which implies that $L(x^*)$ is pseudoconvex with respect to the Bouligand tangent cone.

In the following, we show by two examples that all assumptions of Theorem 6 are essential.

Example 3 Consider the problem (P) where $f(x) = f(x_1, x_2) := -|x_1|$ and S is defined as Example 2. It is clear that f has not any global optimal solution on S.

At any feasible point of the form $\hat{x} = (0, c)$, $c \neq 0$, one has $N_S^B(\hat{x}) = span\{e_1\}$ and $\partial_c f(\hat{x}) = [-1, 1] \times \{0\}$, thus $\hat{x}$ is an N^B-stationary point. It is easy to check that f is not ∂_c-pseudoconvex at $\hat{x}$. Observing also that $L(\hat{x}) = S$ and $T_{L(\hat{x})}^B(\hat{x}) = span\{e_2\}$, the pseudoconvexity of $L(\hat{x})$ is not satisfied.
Considering now the point $\bar{x} = (0, 0)$ and noting that $N_S^B(\bar{x}) = \{0\}$ and $\partial_c f(\bar{x}) = [-1, 1] \times \{0\}$, we see that $\bar{x}$ is an N^B-stationary point. Furthermore, since $L(\bar{x}) = T_{L(\bar{x})}^B(\bar{x}) = S$, we conclude that $L(\bar{x})$ is pseudoconvex. One can easily verify that f is not ∂_c-pseudoconvex at $\bar{x}$.

The final example shows that the pseudoconvexity of the sublevel sets plays a key role in the conclusion of Theorem 6, even if the objective function is convex and differentiable.

Example 4 Consider the problem (P) where $f(x) = f(x_1, x_2) := e^{x_2}$ and S is given as Example 2. It is easy to check that there is no global optimal solution for this problem and the objective function f is convex and differentiable. For the feasible point $\bar{x} := (1, 0) \in S$, one has $\nabla f(\bar{x}) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Obviously, $-\nabla f(\bar{x}) \in \text{span}\{e_2\} = N_S^B(\bar{x})$ and $\bar{x}$ is an N^B-stationary point. However, with a simple calculation we see that $L(\bar{x}) = \mathbb{R} \times \{0\} \cup \{0\} \times \mathbb{R}_-$ and $T_{L(\bar{x})}^B(\bar{x}) = \mathbb{R} \times \{0\}$ and the pseudoconvexity of $L(\bar{x})$ is not satisfied.

3.2 Second-order optimality conditions

This subsection is devoted to the second-order necessary and sufficient optimality conditions for problem (P) in terms of the first and second-order Dini directional derivatives.

The following theorem presents a second-order necessary optimality condition for problem (P) in the framework of the first and second-order Dini directional derivatives.

Theorem 7 *Let $x^* \in S$ be a local optimal solution of problem (P). Then for any $v \in T_S^B(x^*)$, one of the following two conditions hold:*

(i) $D^- f(x^*; v) > 0$;
(ii) $D^- f(x^*; v) = 0$ *and* $D^2 f(x^*; v, v) \geq 0$.

Proof Taking arbitrary $v \in T_S^B(x^*)$ and applying (5), we get $x^* + tv \in S$ for all $t \in \mathbb{R}$. The local optimality of x^* implies that

$$f(x^* + tv) - f(x^*) \geq 0,$$

for sufficiently small $t \in \mathbb{R}$. Therefore,

$$D^- f(x^*; v) = \liminf_{t \downarrow 0} \frac{f(x^* + tv) - f(x^*)}{t} \geq 0.$$

In the case that $D^- f(x^*; v) = 0$, by local optimality of x^* one has

$$\begin{aligned} D^2 f(x^*; v, v) &= \liminf_{t \downarrow 0} \frac{f(x^* + (t + t^2)v) - f(x^*) - tD^- f(x^*; v)}{t^2} \\ &= \liminf_{t \downarrow 0} \frac{f(x^* + (t + t^2)v) - f(x^*)}{t^2} \geq 0, \end{aligned}$$

and the proof is completed. □

In the following let us state our second-order sufficient optimality result.

Theorem 8 *Suppose that $x^* \in S$ and for each $v \in T_S^B(x^*)$ one of the following two conditions hold:*

(i) $D^- f(x^*; v) > 0$;
(ii) $D^- f(x^*; v) = 0$ *and* $D^2 f(x^*; v, v) > 0$.

Then f has a strict local minimum at x^*.

Proof Suppose on the contrary that there exists a sequence $x_k \xrightarrow{S} x^*$ such that for each k, $f(x_k) \leq f(x^*)$.

Denoting $v_k := \frac{x_k - x^*}{\|x_k - x^*\|}$, then passing to a subsequence if necessary, we can assume that $v_k \to v \in T_S^B(x^*)$. On the other hand the Lipschitzness of f at x^* implies that

$$\begin{aligned} D^- f(x^*; v) &= \liminf_{t \downarrow 0} \frac{f(x^* + tv) - f(x^*)}{t} \\ &\leq \liminf_{k \to \infty} \frac{f(x^* + t_k v) - f(x^*)}{t_k} \\ &= \liminf_{k \to \infty} \frac{f(x_k) - f(x^*)}{t_k} \leq 0, \end{aligned}$$

where $t_k = \|x_k - x^*\|$. Putting the above together with (i) and (ii), we deduce that $D^- f(x^*; v) = 0$. We can get $s_k \downarrow 0$ such that $s_k + s_k^2 = t_k$. Therefore, $x^* + (s_k + s_k^2)v = x_k$ and we have

$$\begin{aligned} D^2 f(x^*; v, v) &= \liminf_{t \downarrow 0} \frac{f(x^* + tv + t^2 v) - f(x^*)}{t^2} \\ &\leq \liminf_{k \to \infty} \frac{f(x^* + (s_k + s_k^2)v) - f(x^*)}{s_k^2} \\ &= \liminf_{k \to \infty} \frac{f(x_k) - f(x^*)}{s_k^2} \leq 0. \end{aligned}$$

Thus we arrive at a contradiction which completes the proof of the theorem. □

Finally in this section, we present an example to illustrate the results of Theorems 7 and 8.

Example 5 Consider problem (P) where $f(x) = f(x_1, x_2) := 2|x_2| - x_1 + {x_1}^2$ and S is defined as in Example 2. The feasible set S may be written as $S = \{(0,0), (0,c), (c,0);\ c \in \mathbb{R},\ \ c \neq 0\}$.
In the case that $\acute{x} = (0,c), c \neq 0$, we have $T_S^B(\acute{x}) = span\{e_2\}$ and for any $d \in T_S^B(\acute{x})$,

$$D^- f(\acute{x}; d) = \begin{cases} 2d_2, & \text{if } c > 0; \\ -2d_2, & \text{otherwise}. \end{cases}$$

It is easy to observe that the necessary condition of Theorem 7 is not satisfied.
For $\hat{x} = (0,0)$, $T_S^B(\hat{x}) = \mathbb{R}^2$ and for any $d \in T_S^B(\hat{x})$, $D^- f(\hat{x}, d) = 2|d_2| - d_1$ which is not necessarily nonnegative. Thus $\hat{x}$ is not a local minimizer.
In the case that $\bar{x} = (c, 0)$, $c \neq 0$, since $T_S^B(\bar{x}) = span\{e_1\}$, it follows that for any $d \in T_S^B(\bar{x})$, $D^- f(\bar{x}, d) = (2c - 1)d_1$. In order to establish the necessary condition we need to take $c = \frac{1}{2}$. Hence we consider the point $x^* = (\frac{1}{2}, 0)$. Clearly, $D^- f(x^*; d) = 0$ and $D^2 f(x^*; d, 0) = d_1^2 > 0$, for each $d \in T_S^B(x^*)$. Thus the necessary and sufficient optimality conditions of Theorems 7 and 8 are fulfilled and x^* is a strict local minimizer.

4 Second-order tangent set

In this section we are going to develop the second-order necessary optimality conditions by using the notion of the second-order tangent set. The next theorem gives the expression of the second-order tangent of sparse set S.

Theorem 9 *Let $x^* \in S$ and $d \in T_S^B(x^*)$. Then $T_S^2(x^*, d)$ is given by*

$$T_S^2(x^*, d) = \{w \in \mathbb{R}^n : \|w\|_0 \leq s,\ \|x^* + \mu d + \lambda w\|_0 \leq s, \quad \forall \lambda, \mu \in \mathbb{R}\} \tag{11}$$

$$= \bigcup_{\gamma} \text{span}\,\{e_i, \quad i \in \gamma \supseteq J(d),\ |\gamma| \leq s\}. \tag{12}$$

Proof It is not difficult to show that the sets of the right hands of (11) and (12) are equal. Thus it is sufficient to prove (11). Denote the right hand side of (11) by D. First take arbitrary $w \in T_S^2(x^*, d)$. Then there exist sequences $x_k \xrightarrow{S} x^*$ and $t_k \downarrow 0$ such that $w = \lim_{k\to\infty} \frac{x_k - x^* - t_k d}{\frac{1}{2}(t_k)^2}$. Since $d \in T_S^B(x^*)$, we get from (5) that $\|d\|_0 \leq s$ and $\|x^* + \mu d\|_0 \leq s$ for all $\mu \in \mathbb{R}$. Since $x_k \to x^*$, we can assume without loss of generality that $J(x^*) \subseteq J(x_k)$ for all $k \in \mathbb{N}$. Since $d = \lim_{k\to\infty} \frac{x_k - x^*}{t_k}$, it follows that $J(d) \subseteq J(x_k)$ for all k. All the above together with the definition of w yield $\|w\|_0 \leq s$ and also $\|x^* + \mu d + \lambda w\|_0 \leq s$ for each $\lambda, \mu \in \mathbb{R}$. Thus $T_S^2(x^*, d) \subseteq D$.

Conversely, take arbitrary $w \in D$. Taking a given sequence $t_k \downarrow 0$ and defining $x_k = x^* + t_k d + \frac{1}{2}t_k^2 w$, then obviously $x_k \xrightarrow{S} x^*$ and $w = \lim_{k\to\infty} \frac{x_k - x^* - t_k d}{\frac{1}{2}(t_k)^2}$, which implies that $w \in T_S^2(x^*, d)$ and completes the proof. □

The next theorem establishes a second-order necessary optimality condition for (P) in the framework of the second-order tangent set.

Theorem 10 *Assume that x^* is a local optimal solution of problem (P). Then for each $v \in T_S^B(x^*) \cap \ker D^- f(x^*, .)$, one has*

$$D^2 f(x^*; v, w) \geq 0 \qquad \forall w \in T_S^2(x^*, v).$$

Proof Suppose that for some $v \in T_S^B(x^*) \cap \ker D^- f(x^*, .)$ and $w \in T_S^2(x^*, v)$ one has $D^2 f(x^*; v, w) < 0$. Hence, taking into account that $D^- f(x^*, v) = 0$, the above gives a sequence $t_k \downarrow 0$ such that for all k

$$f(x^* + t_k v + t_k^2 w) < f(x^*).$$

On the other hand from Theorem 9 we have $x^* + t_k v + t_k^2 w \in S$ for all k, which contradicts the local optimality of x^* and completes the proof. □

Eventually we illustrate the result of Theorem 10 in the following example.

Example 6 Consider the problem (P) where f and S are given as Example 5. It was indicated that $x^* = (\frac{1}{2}, 0)$ is a strict local optimal solution for this problem. Let us show that the statement of Theorem 10 is satisfied at x^*.
A direct computation gives $T_S^B(x^*) = span\{e_1\}$ and for each $v \in T_S^B(x^*)$, $T_S^2(x^*, v) = T_S^B(x^*)$. Taking arbitrary point $v \in T_S^B(x^*)$, it is easy to check that $D^- f(x^*, v) = 0$ and $D^2 f(x^*; v, w) = v_1^2 > 0$ for all $w \in T_S^2(x^*, v)$. Therefore, the optimality condition of Theorem 10 holds at x^*.

Acknowledgements The second-named author was partially supported by a Grant from IPM (No. 96900422).

References

1. Aussel, D.: Subdifferential properties of quasiconvex and pseudoconvex functions. J. Optim. Theory Appl. **97**, 29–45 (1998)
2. Bahmani, S., Raj, B., Boufounos, P.T.: Greedy sparsity-constrained optimization. J. Mach. Learn. Res. **14**, 807–841 (2013)
3. Bazaraa, M.S., Sherali, H.D., Shetty, C.M.: Nonlinear Programming: Theory and Algorithms. Wiley, New York (1979)
4. Beck, A., Eldar, Y.C.: Sparsity constrained nonlinear optimization: optimality conditions and algorithms. SIAM J. Optim. **23**, 1480–1509 (2013)
5. Berg, E.V.D., Friedlander, M.P.: Sparse optimization with least-squares constraints. SIAM J. Optim. **21**, 1201–1229 (2011)
6. Borwein, J.M., Lewis, A.S.: Convex Analysis and Nonlinear Optimization, Theoty and Examples. Canadian Mathematical Society, Ottawa (2000)

7. Clarke, F.H., Ledyaev, Y.S., Stern, R.J., Wolenski, P.R.: Nonsmooth Analysis and Control Theory. Springer, New York (1991)
8. Donoho, D.L.: Denoising by soft-thresholding. IEEE Trans. Inform. Theory. **41**, 613–627 (1995)
9. Gorodnitsky, I.F., Rao, B.D.: Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm. IEEE Trans. Signal Process. **45**, 600–616 (1997)
10. Ivanov, V.I.: On the functions with pseudoconvex sublevel sets and optimality conditions. J. Math. Anal. Appl. **345**, 964–974 (2008)
11. Lu, Z., Zhang, Y.: Sparse approximation via penalty decomposition methods. SIAM J. Optim. **23**, 2448–2478 (2013)
12. Makela, M.M., Karmitsa, N., Eronen V.P.: On generalized pseudo- and quasiconvexities for nonsmooth functions. TUCS report, 989, Turku Centre for Computer Science (2011)
13. Mallat, S.: A Wavelet Tour of Signal Processing: The Sparse Way. Academic, New York (2008)
14. Natarajan, B.K.: Sparse approximate solutions to linear systems. SIAM J. Comput. **24**, 227–234 (1995)
15. Pan, L., Luo, Z., Xiu, N.: Restricted Robinson constraint qualification and optimality for cardinality-constrained cone programming. J. Optim. Theory Appl. **175**, 104–118 (2017)
16. Pan, L., Xiu, N., Zhou, S.: On solutions of sparsity constrained optimization. J. Oper. Res. Soc. China. **3**, 421–439 (2015)
17. Penot, J.P.: Generalized convexity in the light of nonsmooth analysis. Recent developments in optimization. Lecture Notes in Economics and Mathematical Systems, Springer, Berlin, Germany, 429, pp. 269–290 (1995)
18. Rockafellar, R.T., Wets, R.J.B.: Variational Analysis. Springer, Berlin (1998)
19. Ruszczynski, A.P.: Nonlinear Optimization. Princeton University Press, Princeton (2006)
20. Studniarski, M.: Second-order necessary conditions for optimality in nonsmooth nonlinear programming. J. Math. Anal. Appli. **154**, 303–317 (1991)
21. Taubman, D., Marcellin, M.W.: Image compression fundamentals, standards and practice. Kluwer, Dordrecht (2001)
22. Tropp, J.A., Wright, S.J.: Computational methods for sparse solution of linear inverse problems. Proc. IEEE. **98**, 948–958 (2010)