informs
https://pubsonline.informs.org/journal/moor

MATHEMATICS OF OPERATIONS RESEARCH
Articles in Advance, pp. 1–24
ISSN 0364-765X (print), ISSN 1526-5471 (online)

Analysis of a Class of Minimization Problems Lacking Lower Semicontinuity

Shaoning Han,[a,*] Ying Cui,[b] Jong-Shi Pang[c]

[a] Department of Mathematics, National University of Singapore, Singapore 119076; [b] Department of Industrial Engineering and Operations Research, University of California, Berkeley, California 94720; [c] Daniel J. Epstein Department of Industrial and Systems Engineering, University of Southern California, Los Angeles, California 90089

*Corresponding author

Contact: shaoninghan@nus.edu.sg, https://orcid.org/0000-0003-1145-7981 (SH); yingcui@berkeley.edu, https://orcid.org/0000-0003-4173-5647 (YC); jongship@usc.edu, https://orcid.org/0000-0002-2224-6787 (J-SP)

Received: September 21, 2023
Revised: April 9, 2024
Accepted: June 28, 2024
Published Online in Articles in Advance: August 27, 2024

MSC2020 Subject Classifications: Primary: 90C26, 90C30

https://doi.org/10.1287/moor.2023.0295

Abstract. The minimization of nonlower semicontinuous functions is a difficult topic that has been minimally studied. Among such functions is a Heaviside composite function that is the composition of a Heaviside function with a possibly nonsmooth multivariate function. Unifying a statistical estimation problem with hierarchical selection of variables and a sample average approximation of composite chance constrained stochastic programs, a Heaviside composite optimization problem is one whose objective and constraints are defined by sums of possibly nonlinear multiples of such composite functions. Via a pulled-out formulation, a pseudostationarity concept for a feasible point was introduced in an earlier work as a necessary condition for a local minimizer of a Heaviside composite optimization problem. The present paper extends this previous study in several directions: (a) showing that pseudostationarity is implied by (and thus, weaker than) a sharper subdifferential-based stationarity condition that we term epistationarity; (b) introducing a set-theoretic sufficient condition, which we term a local convexity-like property, under which an epistationary point of a possibly nonlower semicontinuous optimization problem is a local minimizer; (c) providing several classes of Heaviside composite functions satisfying this local convexity-like property; (d) extending the epigraphical formulation of a nonnegative multiple of a Heaviside composite function to a lifted formulation for arbitrarily signed multiples of the Heaviside composite function, based on which we show that an epistationary solution of the given Heaviside composite program with broad classes of B-differentiable component functions can in principle be approximately computed by surrogation methods.

Funding: The work of Y. Cui was based on research supported by the National Science Foundation [Grants CCF-2153352, DMS-2309729, and CCF-2416172] and the National Institutes of Health [Grant 1R01CA287413-01]. The work of J.-S. Pang was based on research supported by the Air Force Office of Scientific Research [Grant FA9550-22-1-0045].

Keywords: nonsmooth analysis • lower semicontinuity • Heaviside functions • local convexity-like property

1. Introduction

In this work, we examine a class of minimization problems featured by objective and/or constraint functions that do not exhibit lower semicontinuity. Analyzing and solving such problems present considerable challenges because the desirable points, such as global/local solutions, stationary points, or even feasible solutions, might not be easily accomplished. A broad class of such problems is the following Heaviside composite (HSC) problem:

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\textbf{minimize}} & f_{\text{HSC}}(x) \triangleq \displaystyle\sum_{j=1}^{J_0} \psi_{0j}(x)\,\mathbf{1}_{(0,\infty)}(\phi_{0j}(x)), \\ \textbf{subject to} & x \in X_{\text{HSC}} \triangleq \left\{ x \in P \;\middle|\; \displaystyle\sum_{j=1}^{J_i} \psi_{ij}(x)\,\mathbf{1}_{(0,\infty)}(\phi_{ij}(x)) \le b_i,\ i = 1,\ldots,m \right\}, \end{array} \tag{1}$$

where the (open) Heaviside function $\mathbf{1}_{(0,\infty)}(t) \triangleq \begin{cases} 1 & \text{if } t > 0 \\ 0 & \text{otherwise} \end{cases}$ is the indicator function of the (open) interval $(0,\infty)$; $P \subseteq \mathbb{R}^n$ is a given polyhedron, m and $\{J_i\}_{i=0}^m$ are positive integers, $\{b_i\}_{i=1}^m$ are scalars, and ψ_{ij} and $\phi_{ij} : \mathcal{O} \subseteq$

$\mathbb{R}^n \to \mathbb{R}$ are some given continuous functions defined on an open set $\mathcal{O}$ containing P. Although the Heaviside function $\mathbf{1}_{(0,\infty)}(t)$ exhibits lower semicontinuity for $t \in \mathbb{R}$, the lower semicontinuity can be destroyed when multiplied by functions ψ_{ij} that are not consistently nonnegative. We refer to the reference Cui et al. [7], which has presented the modeling breadth of the HSC constraint set X_{HSC}. In particular, the Heaviside function is central to the treatment of chance constraints in stochastic programming; see Cui et al. [8] for a comprehensive study of such a treatment. In turn, to model conjunctive/disjunctive events, the random functionals in the chance constraints involve pointwise minimum/maximum operations that render them nondifferentiable. Furthermore, decision-dependent multiples of the Heaviside composite functions are used in treatment problems to describe rewards conditional on variable outcomes (Fang [12], Qi et al. [22]). As a unification of these special cases, the class of additive Heaviside composite optimization problems and the concept of *pseudostationarity* were introduced in Cui et al. [7]. The latter concept has its origin in Gómez et al. [14] for the sparse optimization problem and is defined by a fixed-point property of a "pulled-out" formulation.

Originated from a statistical estimation problem with sparsity (Hastie et al. [15]), a special case of the composite Heaviside optimization problem is the problem with affine sparsity constraints (ASCs) that was introduced in Dong et al. [11] as a computational framework for rigorously solving estimation problems with structured sparsity (i.e., logical sparsity conditions). Such constraints define the following set:

$$X_{\text{ASC}} \triangleq \left\{ x \in P \;\middle|\; \sum_{j=1}^{n} a_{ij} |x_j|_0 \le b_i,\ i = 1, \ldots, m \right\}, \tag{2}$$

where $|t|_0 \triangleq \begin{cases} 1 & \text{if } t \neq 0 \\ 0 & \text{otherwise} \end{cases}$ is the sparsity function that is closely related to the Heaviside function(s). For example, to model the hierarchical selection among three variables such that x_3 can only be selected if at least one of x_1 or x_2 is chosen, the following inequality can be employed:

$$|x_3|_0 \le |x_1|_0 + |x_2|_0.$$

An optimization problem over ASCs is a generalization of cardinality constrained problems, whose continuous relaxations have been extensively studied in the existing literature (Bian and Chen [2], Chen et al. [3], Kanzow et al. [17], Kanzow et al. [18]). It is known from Dong et al. [11] that X_{ASC} may not be a closed set when the coefficients a_{ij} have negative signs, such as in the above example. When it comes to optimization problems over these sets, a sign restriction on the multiplier functions is a key requirement in their study (Cui et al. [7], Dong et al. [11]). A main contribution of our work herein is to address problems not satisfying such a sign condition for both sets X_{HSC} and X_{ASC}.

In addressing nonlower semicontinuous functions within objectives and constraints that lead in particular to nonclosed feasible sets, an immediate strategy is to consider the closures of these sets. However, this approach might not be ideal for the following reasons.

- Given that the epilimit (Rockafellar and Wets [23, definition 7.1]; see also Royset [24] and Royset and Wets [25]) of a function sequence is always lower semicontinuous, it is thus not possible to construct approximating functions that exhibit epiconvergence to the original nonlower semicontinuous functions. This absence of epiconvergence in the approximating functions, either within the objective or constraints, can impede the convergence of the global minimizers, let alone stationary solutions, for the approximating problems, among many difficulties.
- The best convergence in terms of the epilimit one can achieve from the approximating functions is to the closure of the lower semicontinuous function. However, in the realm of logical implications and structured variable selections, the closure of a given constraint can potentially compromise its expressiveness. Consider, for instance, the constraint (see Dong et al. [11, example 1])

$$|x_1|_0 \le |x_2|_0 \Longleftrightarrow \mathbf{1}_{(0,+\infty)}(|x_1|) \le \mathbf{1}_{(0,+\infty)}(|x_2|).$$

The feasible set for this constraint is $((0,+\infty) \times \mathbb{R}) \cup ((-\infty, 0) \times \mathbb{R}) \cup \{(0,0)\}$. This constraint expresses the logical implication: $x_1 \neq 0 \Rightarrow x_2 \neq 0$. Yet, the closure of this set is equal to the entire space $\mathbb{R}^2$, which clearly does not (even approximately) model the desired logical conditions accurately.

- On top of the difficulties mentioned above, when there are multiple constraints, it is a demanding task to construct the closure of $\cap_{i=1}^{m} C_m$ when $m > 1$ and at least one C_i is nonclosed. This closure can be significantly smaller than $\cap_{i=1}^{m}$ closure$\{C_i\}$.

Because there is a simple linear structure in the ASC constraint set X_{ASC} and the only combinatorial aspect of the set is because of the ℓ_0-function that has a well-known integer description, a natural question is whether the nonclosedness of $X_{\rm ASC}$ is indeed a challenging issue to deal with per the advances in integer optimization. We approach this question from the perspective of mixed integer linear representability (MILP) of $X_{\rm ASC}$, obtaining in particular a representation of the closure of $X_{\rm ASC}$ that complements the results in Dong et al. [11, section 3] from which we can deduce a "big-M" integer description of the closure. Deepening the analysis in this reference, the extended analysis elucidates the general difficulty associated with a mixed signed combination of multiple ℓ_0-functions in that although an integral formulation of the closure of the set $X_{\rm ASC}$ aids the understanding of its structure, an integer approach for dealing with this set is primarily of conceptual value at the present time; the efficient solution of an optimization problem over this set would require much further research for the approach to be practically viable. Because of the difficulty of the global solution of (1), per the integer programming analysis, an in-depth understanding of the local properties of the feasible set $X_{\rm HSC}$ and of the optimization problem itself is, therefore, imperative; the study of such properties is the focus of the remaining sections of the paper.

The study of the local properties of (1) begins with its stationarity conditions, which for a general optimization problem, are necessary for local optimality. For problems where a minimizer, local or global, is impractical to be computed, a stationary solution is a realistic goal one can hope to obtain in practical computation. The advances in variational analysis (Rockafellar and Wets [23]) have led to the definitions of many notions of subgradients of extended-valued functions, each of which can be used to define a stationarity concept. Among these, the regular subgradients (Rockafellar and Wets [23, definition 8.3]) lead to a sharp stationarity concept that in principle, is applicable to a general constrained optimization problem without regard to the properties of the defining functions and constraints. However, although offering convenience for mathematical analysis, such an extended-valued, subdifferential-based stationarity concept has a major drawback. Namely, it hides the constraints in the objective, rendering the identification of a subgradient a very difficult task. In contrast, by exposing the constraints as given, tangents to the constraint set can often be more easily described and lead to constructive approaches to compute sharp stationary solutions. Indeed, practical computation provides a strong motivation for treating the constraints as they appear.

There are several fundamental issues associated with the stationarity concepts of a minimization problem lacking lower semicontinuity. Foremost is the question of how the previously defined pulled-out-based pseudostationarity (Cui et al. [7]) is related to regular subdifferential-based stationarity as the latter is known to be the sharpest among many stationarity concepts for the very broad class of "Bouligand differentiable" (abbreviated as B-differentiable) problems; see Cui and Pang [6, proposition 6.1.8], where the term Bouligand stationarity was used. Although a Heaviside composite function is not B-differentiable, we are able to demonstrate that pseudostationarity is a weaker notion than the subdifferential-based stationarity, which we term "epistationarity" for reasons to be made clear later and we will formally define in Section 4. A follow-up question is whether there are classes of problems whose epistationary points are local minimizers. This question has its origin in differentiable problems (extendable to B-differentiable problems) for which the class of pseudoconvex functions introduced by Mangasarian [19] and Mangasarian [20] provides an answer. Specifically, for a convex-constrained optimization problem with a differentiable pseudoconvex objective function, a first-order stationary point must be a global minimizer. As an extension to nonsmooth functions, the property of (local) convexity like of a B-differentiable function at the given point, initially defined in the study of piecewise quadratic programming (Cui et al. [9]) and subsequently expanded in Cui et al. [8, section 4.2], provides a sufficient condition for a B-stationary solution of a Bouligand differentiable problem to be a local minimizer. It should be noted, however, that unlike the well-known quasiconvex functions, which yield convex level sets, the level set of a locally convex-like function may not be convex. A further question is whether there are constructive procedures to (approximately) compute an epistationary point of a nonlower semicontinuous Heaviside composite program. We answer this question via lifting the problem to one with additional variables and then resorting to the family of surrogation methods (Cui and Pang [6, chapter 7]) when the functions in the lifted program are "surrogatable" (e.g., difference of convex). Details of such an algorithmic development are not addressed in the present work; these are best left for a separate computational study.

1.1. Organization and Contributions

After a brief summary of the notations and some relevant background materials in Section 2 for the study of Problem (1), we organize the rest of this paper along with the main contributions as follows.

a. In Section 3, we provide an algebraic description of the closure of the set $X_{\rm ASC}$ that complements the results in Dong et al. [11, section 3]. Based on Theorem 1, which is a restatement of a classical result, we derive a necessary and sufficient condition for $X_{\rm ASC}$ to have a mixed integer linear representation. This result is enhanced by a more detailed description of the representation by exploiting the structure of $X_{\rm ASC}$; see Theorem 2 and Corollary 1.

b. We formally define epistationarity for an optimization problem lacking lower semicontinuity in Section 4 (see Definition 1) and establish several important properties of an epistationary solution. First, epistationarity is a necessary condition for local minimization (Proposition 2). Epistationarity has an equivalent description in terms of a suitable subderivative (Proposition 3). For a B-differentiable problem, epistationariy recovers B-stationarity (Proposition 4). Finally, for the HSC-constrained optimization Problem (1), epistationarity is sharper than the pulled-out pseudostationarity (Proposition 5).

c. In Section 5, we generalize the functional convexity-like condition to a set-theoretic local convexity-like property and establish its sufficiency for local minimization of an epistationary point (Proposition 9). Being the local version of the classical result of pseudoconvexity implying global optimality for a differentiable problem, our result is for a nonlower semicontinuous program with a possibly nonconvex feasible set. The terminology of epistationarity sufficiency is borrowed from "minimum principle sufficiency" (Ferris and Mangasarian [13]), which aims to answer a related but different question pertaining to the characterization of the set of optimal solutions of a convex differentiable program in terms of the minimum principle of the program at a given optimal solution.

d. In Section 6, based on the algebraic descriptions of tangent vectors of various cases of an HSC set, we summarize in Theorem 3 when such a set has the local convexity-like property. With this property, we obtain the equivalence of epistationarity with local optimality for these classes of Heaviside-defined optimization problems.

e. In Section 7, where we assume, for simplicity, that the objective function is B-differentiable, we introduce through several steps a lifted formulation of Problem (1) and show that the B-stationary points of this lifted formulation, where all functions in the lifted space are B-differentiable, yield pseudostationary points of (1) through projecting the B-stationarity points from the lifted domain onto the original space; see Proposition 15. Bouligand stationarity can be obtained from the lifting under a further assumption; see Proposition 16. Both results are established without any sign condition on the multiplier functions $\{\psi_{ij}\}$.

2. Notations and Background

Parallel to the notation $\mathbb{R}^n$ for the n-dimensional Euclidean space of real numbers, we denote the set of n-dimensional integers and positive integers by $\mathbb{Z}^n$ and $\mathbb{Z}^n_+$, respectively. The superscript n is omitted if it equals one. For a given set S, we denote its closure by $\mathrm{cl}(S)$, convex hull by $\mathrm{conv}(S)$, recession cone by S_∞, and distance to a point $x \in \mathbb{R}^n$ by $\mathrm{dist}(x, S) \triangleq \inf\{\|x-y\|_\infty : y \in S\}$, where $\|a\|_\infty \triangleq \max_i |a_i|$ is the infinity norm of a vector. For any vector $x \in \mathbb{R}^n$, we write its support as $\mathrm{supp}(x)$, and $|x|_0$ for the vector whose components are $|x_i|_0$ for $i = 1, \cdots, n$.

To prepare for the analysis of the Heaviside-defined optimization Problem (1), we review some background pertaining to a general constrained optimization problem in finite dimensions:

$$\operatorname*{\mathbf{minimize}}_{x\in X} f(x), \tag{3}$$

where X is a nonempty subset of $\mathbb{R}^n$ (which is not necessarily closed) and $f : \mathcal{O} \to \mathbb{R}$ is a function defined on the open set $\mathcal{O}$ that contains X. It is common in variational analysis to consider the unconstrained formulation of (3),

$$\operatorname*{\mathbf{minimize}}_{x\in\mathbb{R}^n} f_X(x) \triangleq f(x) + \delta_X(x),$$

by hiding the constraint set X using the extended-valued indicator function: $\delta_X(x) \triangleq \begin{cases} \infty & \text{if } x \notin X \\ 0 & \text{if } x \in X. \end{cases}$ It is known from Royset and Wets [23, theorem 10.1] that if $\bar{x} \in X$ is a local minimizer of (3), then $0 \in \hat{\partial} f_X(\bar{x})$, where

$$\begin{aligned}\hat{\partial} f_X(\bar{x}) &\triangleq \left\{ v \in \mathbb{R}^n \,|\, \liminf_{x(\neq\bar{x})\to\bar{x}} \frac{f_X(x) - f_X(\bar{x}) - v^\top(x-\bar{x})}{\|x-\bar{x}\|} \geq 0 \right\} \quad \text{Royset and Wets [23, definition 8.3]} \\ &= \{ v \in \mathbb{R}^n \,|\, v^\top w \leq \mathrm{d}f_X(\bar{x})(w) \ \text{ for all } \ w \in \mathbb{R}^n \} \quad \text{Royset and Wets [23, exercise 8.4]}\end{aligned}$$

and

$$\begin{aligned}\mathrm{d}f_X(\bar{x})(v) &\triangleq \begin{cases} \displaystyle\liminf_{\substack{w\to v;\,\tau\downarrow 0\\ \bar{x}+\tau w\in X}} \frac{f(\bar{x}+\tau w) - f(\bar{x})}{\tau} & \text{Royset and Wets [23, definition 8.1]} \\ \infty & \text{if no such } w \text{ exists} \end{cases} \\ &= \liminf_{\substack{\tau^{-1}(x'-\bar{x})\to 0;\,\tau\downarrow 0\\ x'+\tau v\in X}} \frac{f(x'+\tau v) - f(\bar{x})}{\tau} \quad \text{under identification} : x' = \bar{x} + \tau(w - v).\end{aligned}$$

Following Royset and Wets [23, definition 6.1], we define the tangent cone of X at $\bar{x} \in X$ as

$$\mathcal{T}(\bar{x}; X) \triangleq \left\{ v \in \mathbb{R}^n \,|\, \exists \{x^\nu\} \subset X \text{ converging to } \bar{x} \text{ and } \{\tau_\nu\} \downarrow 0 \text{ such that } v = \lim_{\nu \to \infty} \frac{x^\nu - \bar{x}}{\tau_\nu} \right\}.$$

Note that the domain of the subderivative $\mathrm{d}f_X(\bar{x})(\bullet)$ is a subset of the tangent cone of X at $\bar{x}$. According to the cited reference, $\hat{\partial} f_X(\bar{x})$ and $\mathrm{d}f_X(\bar{x})(v)$ are, respectively, the *constrained* regular subdifferential and the subderivative of the pair (f, X) at the vector $\bar{x} \in X$. The difference between the two limit infima in $\mathrm{d}f_X(\bar{x})(v)$ is that in the first liminf, the vector $\bar{x}$ is fixed in the first term $f(\bar{x} + \tau w)$, and the direction w is allowed to vary near the given direction v, whereas in the second, the direction v is fixed in the same term $f(x' + \tau v)$, and the vector x' is allowed to vary near $\bar{x}$. Although the subdifferential $\hat{\partial} f_X(\bar{x})$ is very convenient for analysis, the fact that the set X is hidden in the extended-valued function f_X complicates the design of solution methods; indeed, unwrapping the elements therein to expose the set X is invariably needed to take advantage of these properties.

When f is a B-differentiable function (Cui and Pang [6, definition 4.1.1] at $\bar{x} \in X$ (i.e., f is locally Lipschitz continuous near $\bar{x}$ and directionally differentiable there) so that the one-sided directional derivatives

$$f'(\bar{x}; v) \triangleq \lim_{\tau \downarrow 0} \frac{f(\bar{x} + \tau v) - f(\bar{x})}{\tau}$$

exist for all $v \in \mathbb{R}^n$, the vector $\bar{x}$ is said to be a *B-stationary point* of (3) (Cui and Pang [6, definition 6.1.1] if

$$f'(\bar{x}; v) \geq 0, \; \forall v \in \mathcal{T}(\bar{x}; X).$$

The closedness of the set X is not needed for the definition of the tangent cone or for B-stationarity; nevertheless, the directional differentiability of the objective is needed for the latter. It is clear that B-stationarity is a necessary condition for a local minimizer. Moreover, it is shown in Cui and Pang [6, proposition 6.1.8] that if f is B-differentiable at $\bar{x}$ and X is a closed convex set, then $\bar{x}$ is a B-stationary point of f on X if and only if $0 \in \hat{\partial} f_X(\bar{x})$; additionally, if $f'(\bar{x}; \bullet)$ is a convex function and X is a convex set, then these stationarity properties are further equivalent to the condition that $0 \in \hat{\partial} f(\bar{x}) + \mathcal{N}(\bar{x}; X)$, where $\mathcal{N}(\bar{x}; X)$ is the normal cone of the convex set X at $\bar{x}$ as in classical convex analysis. A B-differentiable function f is said to be *Clarke regular* at a point $\bar{x}$ in its domain (Clarke [4, definition 2.3.4]) if

$$f'(\bar{x}; v) = f^\circ(\bar{x}; v) \triangleq \limsup_{\substack{x \to \bar{x} \\ \tau \downarrow 0}} \frac{f(x + \tau v) - f(x)}{\tau}, \quad v \in \mathbb{R}^n,$$

where $f^\circ(\bar{x}; v)$ is the Clarke directional derivative of f at $\bar{x}$ along the direction v.

3. Mixed Integer Linear Representability of X_{ASC}.

To provide a strong motivation for the remainder of the paper, this section validates the computational difficulty of Problem (1) by providing sufficient and necessary conditions for the ASC constraint set X_{ASC} to be *mixed integer linear representable* (MILR). Specifically, a subset S of $\mathbb{R}^n$ is termed MILR if there exist rational matrices A, B, and C and a rational vector d, all of appropriate dimensions, such that

$$S = \{x \in \mathbb{R}^n \,|\, \exists (y, z) \in \mathbb{R}^n \times \mathbb{Z}^q \text{ such that } Ax + By + Cz \leq d\}.$$

As we will see, even obtaining such a representation for X_{ASC} is not a trivial task, which would be a reasonable first step in attempting to solve an associated optimization problem to global optimality.

Needless to say, the challenge in dealing with the set X_{ASC} is the ℓ_0 function $|\bullet|_0$. To address this function, the integer programming community often employs an indicator variable $z \in \{0,1\}^n$ to represent the support of the continuous variable $x \in \mathbb{R}^n$ (see, e.g., Atamtürk et al. [1] and Dong and Linderoth [10]). The constraint $z = |x|_0$ is further relaxed to $-Mz \leq x \leq Mz$ via the standard big-M technique, enabling a more tractable formulation. This yields the following mixed integer set that contains X_{ASC} (assumed bounded):

$$\left\{ x \in P \,\middle|\, \exists z \in \{0,1\}^n \text{ such that } -Mz \leq x \leq Mz \text{ and } \sum_{j=1}^n a_{ij} z_j \leq b_i, \; i = 1, \ldots, m \right\},$$

where $M > 0$ is chosen to be sufficiently large to ensure $X_{\text{ASC}} \subseteq \{x \in \mathbb{R}^n \,|\, \|x\|_\infty \leq M\}$. It is known that such a relaxation is exact provided that all the coefficients a_{ij} are nonnegative. However, complexities arise when A does not meet the sign condition.

The following classical result (Conforti et al. [5, theorem 4.47]) provides geometric conditions under which a subset S of $\mathbb{R}^n$ is MILR. It is important to note that the intcone in Expression (4) is an "integer cone" that consists of nonnegative integral combinations of integer vectors; in particular, this cone is not necessarily polyhedral.

Theorem 1. *A set $S \subseteq \mathbb{R}^n$ is MILR if and only if there exist rational polytopes $P_1, \ldots, P_k \subseteq \mathbb{R}^n$ and vectors $r^1, \ldots, r^m \in \mathbb{Z}^n$ such that*

$$S = \bigcup_{i=1}^{k} P_i + \operatorname{intcone}\{r^1, \ldots, r^m\}, \tag{4}$$

where $\operatorname{intcone}\{r^1, \ldots, r^m\} \triangleq \{\sum_{i=1}^m \lambda_i r^i \mid \lambda \in \mathbb{Z}_+^m\}$.

Note that an MILR set must be closed but not necessarily bounded. Indeed, a set S is closed if and only if $S \cap \{x : \|x\|_2 \le \tau\}$ is closed for any scalar $\tau > 0$. If S is MILR, then by Theorem 1, the set $S \cap \{x : \|x\|_2 \le \tau\}$ is a finite union of compact sets and is thus closed. This implies that X_{ASC} is not MILR in general. In Dong et al. [11, section 3], the issue of closedness of X_{ASC} and the identification of its closure have been studied under a key assumption. The result below generalizes this previous study without such an assumption; besides the improved identification, which is seemingly conceptual, its proof provides a constructive pathway to the subsequent result, Theorem 2, that provides a full characterization of the MILR property of the set X_{ASC}.

Proposition 1. *Let $P \subseteq \mathbb{R}^n$ be a polyhedron. There exist a matrix $\tilde{A} \ge 0$ and a $\{0,1\}$-vector $\tilde{b}$ such that* $\operatorname{cl}(X_{\text{ASC}}) = \{x \in P \mid \tilde{A}|x|_0 \le \tilde{b}\}$.

Proof. Let $S \triangleq \{z \in \{0,1\}^n \mid z = |x|_0,\ x \in X_{\text{ASC}}\}$ be the set of possible supports of the feasible region. Let $\hat{S} \triangleq \cup_{z \in S}\{y \in \{0,1\}^n \mid y \le z\}$ be the downward closure generated by S. Because $\hat{S} \subseteq \{0,1\}^n$, one has $\hat{S} = \operatorname{conv}(\hat{S}) \cap \{0,1\}^n$. We claim that $\operatorname{conv}(\hat{S}) = \{z \ge 0 \mid \tilde{A}z \le \tilde{b}\}$ for some matrix $\tilde{A} \ge 0$ and $\{0,1\}$-vector $\tilde{b}$. For this purpose, we first show that $y \in \operatorname{conv}(\hat{S})$ if and only if $y \ge 0$ and $y^\top u \le \max_{z \in \hat{S}} u^\top z$ for all $u \ge 0$. The "only if" assertion is obvious. For the "if" assertion, suppose that $0 \le y \notin \operatorname{conv}(\hat{S})$ is such that $y^\top u \le \max_{z \in \hat{S}} u^\top z$ for all $u \ge 0$. Because $\operatorname{conv}(\hat{S})$ is a polytope, by separation, there exist a vector $\tilde{u}$ and a scalar γ such that $y^\top \tilde{u} > \gamma \ge \max_{z \in \operatorname{conv}(\hat{S})} \tilde{u}^\top z$. For any vector $z \in \widehat{S}$, the vector $\tilde{z}$ obtained by zeroing out the components of z corresponding to the negative components of $\tilde{u}$ remains an element of $\hat{S}$. Thus, with $\tilde{u}^+$ denoting the nonnegative part of the vector $\tilde{u}$, we have

$$y^\top \tilde{u}^+ \ge y^\top \tilde{u} > \tilde{z}^\top \tilde{u} = z^\top \tilde{u}^+,$$

which is a contradiction. This completes the proof of the description of a vector $y \in \operatorname{conv}(\hat{S})$. Next, we note that $y^\top u \le \max_{z \in \hat{S}} u^\top z$ for all $u \ge 0$ is equivalent to

$$y^\top u \le \alpha \text{ for all } u \ge 0, \alpha \ge \max_{z \in \hat{S}} u^\top z \Longleftrightarrow \begin{cases} y^\top u \le 1 & \text{for all } u \ge 0 \text{ such that } \max\limits_{z \in \hat{S}} u^\top z \le 1 \\ y^\top u \le 0 & \text{for all } u \ge 0 \text{ such that } \max\limits_{z \in \hat{S}} u^\top z \le 0 \end{cases}$$

$$\Longleftrightarrow \begin{cases} y^\top u \le 1 & \forall u \in P_1 \triangleq \{u \mid u \ge 0, u^\top z \le 1,\ \forall z \in \hat{S}\} \\ y^\top u \le 0 & \forall u \in P_0 \triangleq \{u \mid u \ge 0, u^\top z \le 0,\ \forall z \in \hat{S}\}. \end{cases}$$

Because P_0 and P_1 are polytopes, one has $P_i = \operatorname{conv}\{u^{ij} \mid j = 1, \ldots, k_i\}$ for certain finite families of vectors $\{u^{ij}\}_{j=1}^{k_i} \subseteq \mathbb{R}_+^n$, for $i = 1, 2$. Therefore, $y \in \operatorname{conv}(\hat{S})$ if and only if y belongs to the set

$$\{y \mid y \ge 0,\ (u^{1j})^\top y \le 1,\ (u^{2\ell})^\top y \le 0,\ \forall j = 1, \ldots, k_1,\ \ell = 1, \ldots, k_2\},$$

completing the proof of the claimed polyhedral representation of $\operatorname{conv}(\hat{S})$.

It remains to show that $\operatorname{cl}(X_{\text{ASC}}) = \tilde{X} \triangleq \{x \in P \mid \tilde{A}|x|_0 \le \tilde{b}\}$. Note that $\tilde{X}$ is a closed set because of $\tilde{A} \ge 0$. It is evident that $\operatorname{cl}(X_{\text{ASC}}) \subseteq \tilde{X}$. To prove the converse inclusion, consider an arbitrary $x \in \tilde{X}$. One has $|x|_0 \in \hat{S}$, which implies that there exists $\hat{x} \in X_{\text{ASC}}$ such that $|\hat{x}|_0 \ge |x|_0$ by the construction of $\hat{S}$. Let $x(\varepsilon) = \varepsilon\hat{x} + (1 - \varepsilon)x$ for $\varepsilon \in [0,1]$. Clearly, $x(\varepsilon)$ belongs to P, and for almost all $\varepsilon \in (0,1]$, $|x(\varepsilon)|_0 = |\hat{x}|_0$. Because $A|\hat{x}|_0 \le b$, one can deduce that for almost all $\varepsilon \in (0,1]$, $x(\epsilon) \in X_{\text{ASC}}$. The proof is now complete because $\lim_{\varepsilon \downarrow 0} x(\varepsilon) = x$. □

A point $x \in S$ is called a *maximal element* in S if there does not exist a point $y \ne x \in S$ such that $y \ge x$. The proof of Proposition 1 indicates that if $x \in \operatorname{cl}(X_{\text{ASC}})$ and $|x|_0$ is the maximal element in the support set $\{z \in \{0,1\}^n \mid z = |x|_0, x \in X_{\text{ASC}}\}$, then $x \in X_{\text{ASC}}$. This fact is useful when searching for a point in X_{ASC} to approximate elements in $\operatorname{cl}(X_{\text{ASC}})$. Specifically, consider the case where the matrix $\tilde{A}$ in Proposition 1 is known. Take any point $\overline{x} \in \operatorname{cl}(X_{\text{ASC}})$, and let $\overline{z} = |\overline{x}|_0 \in \{0,1\}^n$. Given $\tilde{A} \ge 0$, it is easy to identify a maximal element $\hat{z} \in \{0,1\}^n$ in the support set such that $\bar{z} \leq \hat{z}$. Following this, one can determine a point $\hat{x} \in \text{cl}(X_{\text{ASC}})$ such that $|\hat{x}|_0 = \hat{z}$ by solving linear programs over $\{x \in P \,|\, x_i(1-\hat{z}_i) = 0,\ i = 1,\ldots,n\}$. Consequently, we have $\hat{x} \in X_{\text{ASC}}$, which implies that $\varepsilon\hat{x} + (1-\varepsilon)\bar{x} \in X_{\text{ASC}}$ for almost all $\varepsilon \in (0,1]$.

However, it is worth noting that although the existence of $\tilde{A}$ is guaranteed by Proposition 1, unfortunately, the effective construction of $\tilde{A}$ remains unclear. Consequently, this proposition is primarily of conceptual significance. In the following, we show that for the set $X_{\text{ASC},}$ the integer cone in Theorem 1 can be replaced with a polyhedral cone that is given by any maximal element from the support set. We start with a technical lemma.

Lemma 1. *Let $r \in \mathbb{R}^n$ and $\bar{x} \in X_{\text{ASC}}$. If X_{ASC} is closed and there exists a nonnegative sequence $\{t_k\} \to \infty$ such that $\bar{x} + t_k r \in X_{\text{ASC}}$ for all k, then the ray $\{\bar{x} + tr : t \geq 0\} \subseteq X_{\text{ASC}}$.*

Proof. Because X_{ASC} is closed, by Proposition 1, one can assume $A \geq 0$. Observe that there exists $t_0 > 0$ such that as $t > t_0$, $\text{supp}(\bar{x} + tr) = \text{supp}(\bar{x}) \cup \text{supp}(r)$. If in addition, $x(t) \triangleq \bar{x} + tr \in X_{\text{ASC}}$, then for any $y = \lambda\bar{x} + (1-\lambda)x(t)$ and $\lambda \in [0,1]$, one has $y \in P$ and $|y|_0 \leq |x(t)|_0$, which implies that $A|y|_0 \leq A|x(t)|_0 \leq b$. Therefore, we have $y \in X_{\text{ASC}}$. The conclusion follows from the assumption that $t \to \infty$. □

The noteworthy point of the MILR of X_{ASC} in the result below is twofold; one, the cone in (4) can be made polyhedral, and two, its generators are recession vectors of the base polyhedron P whose nonzero components correspond to those of a maximal element of the set $\hat{S}$ in the proof of Proposition 1.

Theorem 2. *Assume P is a polyhedron defined by rational data. Then, X_{ASC} is MILR if and only if there exist nonempty rational polytopes $\{P_i\}_{i=1}^k$ and a polyhedral cone R such that $X_{\text{ASC}} = \cup_{i=1}^k P_i + R$. Furthermore, the recession cone R takes the form $\{r \in P_\infty \,|\, r_i = 0,\ \forall i \notin \text{supp}(z^{\max})\}$, where $z^{\max}$ is any maximal element of $\{z \in \{0,1\}^n \,|\, z = |x|_0,\ x \in X_{\text{ASC}}\}$.*

Proof. Thanks to Proposition 1, we can assume that the matrix, denoted as A, of the coefficients a_{ij} in the definition of X_{ASC} is all nonnegative, without loss of generality.

Necessity. Suppose X_{ASC} is MILR. By Theorem 1, there exist rational polytopes $P_1,\ldots,P_k \subseteq \mathbb{R}^n$ and vectors $r^1,\ldots,r^m \in \mathbb{Z}^n$ such that $X_{\text{ASC}} = \cup_{i=1}^k P_i + \text{intcone}\{r^1,\ldots,r^m\}$. For an arbitrary vector $r = \sum_{i=1}^m \lambda_i r^i$ with $\lambda \in \mathbb{Z}_+^m$ and an arbitrary point $x \in X_{\text{ASC}}$, it holds that $x + tr \in X_{\text{ASC}}$ for all $t \in \mathbb{Z}_+$. Thus, one can deduce from Lemma 1 that $x + tr \in X_{\text{ASC}}$ for all $t \geq 0$. This further implies that $x + t\sum_{i=1}^m \mu_i r^i \in X_{\text{ASC}}$ for all $\mu \in \mathbb{R}_+^m$ that is rational, all $t \in \mathbb{R}_+$, and $x \in X_{\text{ASC}}$ because we can always scale r by a positive integer to make μ integral. Because X_{ASC} is closed, it follows that $x + r \in X_{\text{ASC}}$ for all r in the cone generated by the vectors $\{r^i\}_{i=1}^m$, which we denote by R.

Sufficiency. Because P_i are polytopes, $\cup_{i=1}^k P_i$ is MILR by Theorem 1. Because a polyhedral set is always MILR and the Minkowski sum of two MILR sets is MILR, we can deduce that $X_{\text{ASC}} = \cup_{i=1}^k P_i + R$ is MILR.

It remains to prove the representation of the cone R. Let $\tilde{R} \triangleq \{r \in P_\infty \,|\, r_i = 0,\ \forall i \notin \text{supp}(z^{\max})\}$. By the definition of $z^{\max}$, there exists $\bar{x} \in X_{\text{ASC}}$ such that $\text{supp}(z^{\max}) = \text{supp}(\bar{x})$. Note that for any $r \in \tilde{R}$, $\text{supp}(r) \subseteq \text{supp}(\bar{x})$. Thus, for any $t \geq 0$, $A|\bar{x} + tr|_0 \leq A|\bar{x}|_0 \leq b$. Hence, $\bar{x} + tr \in X_{\text{ASC}}$ for any $t \geq 0$; thus, $r \in R$ by the above proof for the first statement of this proposition. Hence, $\tilde{R} \subseteq R$. If there exists $r \in R \setminus \tilde{R}$, then $\bar{x} + tr \in X_{\text{ASC}}$ and $|\bar{x} + tr|_0 > z^{\max}$ for t large enough, contradicting the maximality of $z^{\max}$. This proves $\tilde{R} = R$. □

If $X_{\text{ASC}} \triangleq \{x \in P \,|\, A|x|_0 \leq b\}$ is MILR with A nonnegative, one can readily obtain a maximal element $z^{\max}$ in the support set and the resulting recession cone R. In this favorable case, X_{ASC} admits a big-M extended reformulation. The result is formally stated below.

Corollary 1. *Assume that $A \geq 0$ is rational and $X_{\text{ASC}} \triangleq \{x \in P \,|\, A|x|_0 \leq b\}$ is MILR with the recession cone R. Then , there exists $M \geq 0$ such that*

$$X_{\text{ASC}} = \{x \in \mathbb{R}^n \,|\, \exists (y,z,r) \in P \times \{0,1\}^n \times R \text{ s.t. } Az \leq b; -Mz \leq y \leq Mz, \text{ and } x = y + r\}.$$

Proof. Take M large enough such that in the statement of Theorem 2, it holds that $\hat{P} \triangleq \cup_{i=1}^k P_i$ is contained in $\{x : \|x\|_\infty \leq M\}$. If $x \in X_{\text{ASC}}$, then there exists y and r such that $y \in \hat{P} \subseteq X_{\text{ASC}}$ and $r \in R$. Thus, $y \in P$ and $A|y|_0 \leq b$. This shows that the X_{ASC} is a subset of the right-hand set in the claim. Conversely, suppose (x,y,z,r) satisfies the inequality system in the right-hand set in the claim. Then, $z \geq |y|_0$ and $A \geq 0$ imply that $A|y|_0 \leq b$, from which we can deduce that $y \in X_{\text{ASC}}$. Because $X_{\text{ASC}} + R \subseteq X_{\text{ASC}}$, the conclusion follows. □

4. Epistationarity

It is trivial to cast Problem (3) as one with a B-differentiable objective function by "epigraphicalizing" the function *f*; this maneuver leads to the lifted problem with an auxiliary variable:

$$\underset{(x,t) \in \mathbb{R}^{n+1}}{\textbf{minimize}}\ t \ \textbf{subject to}\ (x,t) \in Z \triangleq \text{epi}(f) \cap (X \times \mathbb{R}), \tag{5}$$

where $\text{epi}(f) \triangleq \{(x,t) \in \mathcal{O} \times \mathbb{R} : f(x) \leq t\}$ is the *epigraph* of f. In this form, we can speak of a pair $(\bar{x},\bar{t}) \in Z$ with $\bar{t} \triangleq f(\bar{x})$ as being a B-stationary point of (5). When f is not lower semicontinuous, its epigraph epi(f) is not closed. Nevertheless, we can formally introduce the following concept.

Definition 1. A vector $\bar{x} \in X$ is an *epistationary solution* of (3) if the pair $(\bar{x}, f(\bar{x}))$ is a B-stationary solution of the lifted Problem (5).

Unwrapping the B-stationarity condition in the lifted formulation based on the tangent cone of Z, we remark that $\bar{x} \in X$ is an epistationary point of f on X if the following implication holds:

$$\left[\lim_{\substack{(x^k,t_k)\in\text{epi}(f)\cap(X\times\mathbb{R})\\(x^k,t_k)\to(\bar{x},\bar{t}),\ \tau_k\downarrow 0}} \frac{(x^k,t_k)-(\bar{x},\bar{t})}{\tau_k} = (v,dt)\right] \Rightarrow dt \geq 0. \tag{6}$$

The following simple result shows that epistationarity is a necessary condition for locally minimizing; the noteworthy point of the result is that no assumption is required of the pair (f, X).

Proposition 2. *Let f be continuous. A vector $\bar{x} \in X$ is a local minimizer of* (3) *if and only if the pair $(\bar{x}, f(\bar{x}))$ is a local minimizer of* (5)*. Thus, if $\bar{x} \in X$ is a local minimizer of* (3)*, then $\bar{x}$ is an epistationary point of* (3).

Proof. "Only if." Suppose $\bar{x} \in X$ is a local minimizer of (3). Let $\mathcal{N}$ be a neighborhood of $\bar{x}$ such that $f(x) \geq f(\bar{x})$ for all $x \in X \cap \mathcal{N}$. Thus, if $(x,t) \in Z \cap (\mathcal{N} \times \mathbb{R})$, then $t \geq f(x) \geq f(\bar{x})$, showing that $(\bar{x}, f(\bar{x}))$ is a local minimizer of (5).

"If." Conversely, suppose $(\bar{x}, f(\bar{x}))$ is a local minimizer of (5). Let $\mathcal{N}_x \times \mathcal{N}_t$ be a neighborhood of $(\bar{x}, f(\bar{x}))$ such that $t \geq f(\bar{x})$ for all $(x,t) \in \mathcal{N}_x \times \mathcal{N}_t$. Let $\mathcal{N}_x^{'} \subseteq \mathcal{N}_x$ be a neighborhood of $\bar{x}$ such that $f(x) \in \mathcal{N}_t$ for all $x \in \mathcal{N}_x^{'}$. It then follows that for $x \in \mathcal{N}_x^{'}$, the pair $(x, f(x))$ belongs to $\mathcal{N}_x \times \mathcal{N}_t$; thus, $f(x) \geq f(\bar{x})$, showing that $\bar{x}$ is a local minimizer of (3). The last statement of the proposition does not require proof. □

For the purpose to connect epistationarity with regular subdifferential-based stationarity, we first establish a lemma.

Lemma 2. *Let $\bar{x} \in X$. It holds that*

$$\liminf_{\bar{x}\neq x(\in X)\to\bar{x}} \frac{f(x)-f(\bar{x})}{\|x-\bar{x}\|} = \inf_{v\in\mathcal{T}(\bar{x};X);\,\|v\|=1} \mathrm{d}f_X(\bar{x})(v) \tag{7}$$

with the values $\pm\infty$ allowed. In particular, if $\bar{x}$ is an isolated vector in X, then the two values are both equal to ∞.

Proof. Let $\{x^k\} \subset X \setminus \{\bar{x}\}$ be a sequence converging to $\bar{x}$ such that

$$\liminf_{\bar{x}\neq x(\in X)\to\bar{x}} \frac{f(x)-f(\bar{x})}{\|x-\bar{x}\|} = \lim_{k\to\infty} \frac{f(x^k)-f(\bar{x})}{\|x^k-\bar{x}\|}.$$

Without loss of generality, we may assume that the normalized sequence $\left\{w^k \triangleq \frac{x^k-\bar{x}}{\|x^k-\bar{x}\|}\right\}$ converges to a tangent vector $v^\infty \in \mathcal{T}(\bar{x};X)$, which must have unit norm. Letting $\tau_k \triangleq \|x^k-\bar{x}\|$, we have $x^k = \bar{x} + \tau_k w^k$; hence,

$$\begin{aligned}\liminf_{\bar{x}\neq x(\in X)\to\bar{x}} \frac{f(x)-f(\bar{x})}{\|x-\bar{x}\|} &= \lim_{k\to\infty} \frac{f(\bar{x}+\tau_k w^k)-f(\bar{x})}{\tau_k} \\ &\geq \liminf_{\substack{w\to v^\infty;\,\tau\downarrow 0\\ \bar{x}+\tau w\in X}} \frac{f(\bar{x}+\tau w)-f(\bar{x})}{\tau} \\ &= \mathrm{d}f_X(\bar{x})(v^\infty) \geq \inf_{v\in\mathcal{T}(\bar{x};X);\,\|v\|=1} \mathrm{d}f_X(\bar{x})(v).\end{aligned}$$

Conversely, let $v \in \mathcal{T}(\bar{x};X)$ be an arbitrary vector with unit norm. We have

$$\mathrm{d}f_X(\bar{x})(v) = \liminf_{\substack{w\to v;\,\tau\downarrow 0\\ \bar{x}+\tau w\in X}} \frac{f(\bar{x}+\tau w)-f(\bar{x})}{\tau\|w\|} \geq \liminf_{\bar{x}\neq x(\in X)\to\bar{x}} \frac{f(x)-f(\bar{x})}{\|x-\bar{x}\|}.$$

Hence, the equalities in (7) hold. □

The following result establishes the equivalence of epistationarity with the nonnegativity of the subderivative $\mathrm{d}f_X(\bar{x})$ on $\mathcal{T}(\bar{x};X)$ and with regular subdifferential-based stationarity.

Proposition 3. *Let $f:\mathbb{R}^n \to \mathbb{R}$ be an arbitrary function and X be an arbitrary subset of $\mathbb{R}^n$. Let $\bar{x} \in X$ be given. The following three statements are equivalent:*
a. $\mathrm{d}f_X(\bar{x})(v) \geq 0$ *for all* $v \in \mathcal{T}(\bar{x}; X)$,
b. $\bar{x}$ *is an epistationary point of f on X, and*
c. $0 \in \hat{\partial} f_X(\bar{x})$.

Proof. (a) ⇒ (b). Let $\{x^k\}$, $\{t_k\}$, $\{\tau_k\}$, and dt satisfy the conditions in the left-hand limit of (6). Then, $v \in \mathcal{T}(\bar{x}; X)$; furthermore,

$$dt \geq \limsup_{k\to\infty} \frac{f(x^k) - f(\bar{x})}{\tau_k} \geq \liminf_{\substack{w\to v;\, \tau\downarrow 0 \\ \bar{x}+\tau w \in X}} \frac{f(\bar{x}+\tau w) - f(\bar{x})}{\tau} = \mathrm{d}f_X(\bar{x})(v) \geq 0,$$

where the last inequality holds because $\mathrm{d}f_X(\bar{x})(v) \geq 0$ by assumption.

(b) ⇒ (c). Suppose $\bar{x} \in X$ is an epistationary point of f on X. If suffices to show that $\mathrm{d}f_X(\bar{x})(v) \geq 0$ for all $v \in \mathbb{R}^n$. This is clearly true if $\mathrm{d}f_X(\bar{x})(v) = \infty$. Suppose that $\mathrm{d}f_X(\bar{x})(v)$ is finite. Then, there exist $\{w^k\} \to v$ and $\{\tau_k\} \downarrow 0$ such that $\bar{x} + \tau_k w^k \in X$ for all k and

$$\mathrm{d}f_X(\bar{x})(v) = \lim_{k\to\infty} \frac{f(\bar{x}+\tau_k w^k) - f(\bar{x})}{\tau_k}.$$

Let $x^k \triangleq \bar{x} + \tau_k w^k$ and $t_k \triangleq f(x^k)$. It follows that the sequences $\{x^k\}$, $\{t_k\}$, and $\{\tau_k\}$ satisfy the conditions in the left-hand limit of (6) with $dt = \mathrm{d}f_X(\bar{x})(v)$. Thus, this subderivative is nonnegative.

Lastly, suppose that $\mathrm{d}f_X(\bar{x})(v) = -\infty$. Then, there exists $\{w^k\} \to v$ such that $\bar{x} + \tau_k w^k \in X$ for some $\tau_k \downarrow 0$, and

$$\lim_{k\to\infty} \frac{f(\bar{x}+\tau_k w^k) - f(\bar{x})}{\tau_k} = -\infty.$$

Thus, there exists a positive integer K such that

$$f(\bar{x}+\tau_k w^k) - f(\bar{x}) \leq -\tau_k, \;\; \forall k \geq K.$$

Let $x^k \triangleq \bar{x} + \tau_k w^k$ and $t_k \triangleq f(\bar{x}) - \tau_k$. It follows that the sequences $\{x^k\}$, $\{t_k\}$, and $\{\tau_k\}$ satisfy the conditions in the left-hand limit of (6) with $dt = -1$. This contradicts the epistationarity of $\bar{x}$.

(c) ⇒ (a). This is obvious by the definition of $\hat{\partial} f_X(\bar{x})$. □

Remark 1. Although the proof of Proposition 3 is closely related to Royset and Wets [23, theorem 8.2], which asserts that the tangent cone of the epigraph of an extended-valued function g at the pair $(\bar{x}, g(\bar{x}))$ with $g(\bar{x})$ finite is equal to the epigraph of the (unconstrained) subderivative $\mathrm{d}g_X(\bar{x})$ of $g_X \triangleq g + \delta_X$, the main point of the proposition is on the restatement of epistationarity in terms of subderivatives.

We next show that the new concept of epistationarity coincides with the old concept of B-stationarity when the objective function f is B-differentiable.

Proposition 4. *Let f be B-differentiable near $\bar{x} \in X$. Then, $\bar{x}$ is a B-stationary point of* (3) *if and only if $\bar{x}$ is epistationary.*

Proof. The "only if" part is the same as that of part (a) of Proposition 3 and straightforward. It remains to prove the "if" part by showing that $f'(\bar{x}; v) \geq 0$ for all $v \in \mathcal{T}(\bar{x}; X)$. There exist sequences $\{x^k\} \subset X$ converging to $\bar{x}$ and $\{\tau_k\} \downarrow 0$ such that $v = \lim_{k\to\infty} \frac{x^k - \bar{x}}{\tau_k}$. Let $t_k \triangleq f(x^k)$ and $dt \triangleq f'(\bar{x}; dx)$. Then, $(v, f'(\bar{x}; v)) \in \mathcal{T}(\bar{z}; Z)$, where $\bar{z} \triangleq (\bar{x}, f(\bar{x}))$, and Z is given in (5). By the epistationarity of $\bar{x}$, it follows that $f'(\bar{x}; v) \geq 0$. □

Referring to the HSC-constrained optimization Problem (1), we say that a vector $\bar{x}$ in X_{HSC} is a *pseudostationary point* of this problem if $\bar{x}$ is an epistationary point of the "pulled-out" problem:

$$\begin{aligned}
&\underset{x \in P}{\textbf{minimize}} && \sum_{j \in \mathcal{J}_{0,+}(\bar{x})} \psi_{0j}(x) \\
&\textbf{subject to} && \text{for all } i = 0, 1, \ldots, m \\
& && \sum_{j \in \mathcal{J}_{i,+}(\bar{x})} \psi_{ij}(x) \leq b_i \;\; (\text{with } b_0 = \infty) \\
& && \phi_{ij}(x) \geq 0, \text{for all } j \in \mathcal{J}_{i,+}(\bar{x}) \\
&\textbf{and} && \phi_{ij}(x) \leq 0, \text{for all } j \in \mathcal{J}_{i,\leq}(\bar{x}),
\end{aligned} \tag{8}$$

where $\mathcal{J}_{i,\leq}(\bar{x}) \triangleq \{j \mid \phi_{ij}(\bar{x}) \leq 0\}$ and $\mathcal{J}_{i,+}(\bar{x}) \triangleq \{j \mid \phi_{ij}(\bar{x}) > 0\}$ for $i = 0, 1, \ldots, m$. We also define $\mathcal{J}_{i,0}(\bar{x}) \triangleq \{j \mid \phi_{ij}(\bar{x}) = 0\}$.

We remark that in Cui et al. [7, definition 3], the definition of pseudostationarity assumes that all the functions $\{\{\psi_{ij}, \phi_{ij}\}_{j=1}^{J_i}\}_{i=0}^{m}$ are B-differentiable; here, these continuous functions can be arbitrary.

The pseudostationarity definition provides one way to resolve the challenge caused by the Heaviside composite functions $\mathbf{1}_{(0,\infty)}(\phi_{ij}(x))$ by exposing the inner functions relative to the reference vector $\bar{x}$ instead of at the variable vector x. Provided that the functions $\{\{\psi_{ij}, \phi_{ij}\}_{j=1}^{J_i}\}_{i=0}^{m}$ have favorable properties (e.g., difference of convexity), the resulting Problem (8) is computationally tractable (Pang et al. [21]) and enables the verification of the stipulated fixed-point condition on the candidate solution $\bar{x}$. The paper by Cui et al. [7] has provided constructive ways to approximately compute a B-stationary point of (8) under some sign conditions on the functions ψ_{ij}; see also Gómez et al. [14] and He et al. [16] for a special quadratic sparse optimization problem involving the ℓ_0-function.

In the following, we show that for Problem (1), epistationarity is sharper than pseudostationarity. Note that (8) is a restriction of the original Problem (1) around $\bar{x}$. Thus, the global optimality and the local optimality of (8) are necessary conditions for the respective optimality of (1).

Proposition 5. *If $\bar{x}$ is an epistationary point of* (1), *then it is pseudostationary.*

Proof. Let $\psi_{\text{HSC}}^{\text{ps}}$ and $X_{\text{HSC}}^{\text{ps}}$ denote the objective function and constraint set of (8), respectively, and $Z_{\text{HSC}}^{\text{ps}} \triangleq \text{epi}(\psi_{\text{HSC}}^{\text{ps}}) \cap (X_{\text{HSC}}^{\text{ps}} \times \mathbb{R})$. Recalling the epigraphical set Z (see (5)) of Problem (1), we first show that if x is sufficiently close to $\bar{x}$ and if (x, t) belongs to $Z_{\text{HSC}}^{\text{ps}}$, then $(x, t) \in Z$. This is indeed true because for such an x, it holds that

$$\mathcal{J}_{i,+}(x) = \{j \mid \phi_{ij}(x) > 0\} = \mathcal{J}_{i,+}(\bar{x})$$

for all $i = 0, 1, \cdots, m$, which implies

$$\sum_{j=1}^{J_i} \psi_{ij}(x)\mathbf{1}_{(0,\infty)}(\phi_{ij}(x)) = \sum_{j \in \mathcal{J}_{i,+}(x)} \psi_{ij}(x) = \sum_{j \in \mathcal{J}_{i,+}(\bar{x})} \psi_{ij}(x) \le b_i.$$

In particular, $f_{\text{HSC}}(x) = \psi_{\text{HSC}}^{\text{ps}}(x)$. Because $\bar{x}$ is an epistationary point of (1), one has (6). Moreover, because of the inclusive relationship of Z and $Z_{\text{HSC}}^{\text{ps}}$ at $\bar{x}$, one can deduce that the implication (6) still holds true if $Z = \text{epi}(f) \cap (X \times \mathbb{R})$ is replaced with $Z_{\text{HSC}}^{\text{ps}}$. The conclusion follows by the definition of epistationarity. □

5. The Set-Theoretic Local Convexity-Like Property

To motivate the subsequent definition, we recall that a B-differentiable function f is *(locally) convex like* at a point $\bar{x}$ in its domain (Cui et al. [8, section 4.2]; see the earlier reference (Cui et al. [9, proposition 4.1]) for a special case of this property) if there exists a neighborhood $\mathcal{N}$ of $\bar{x}$ such that

$$f(x) \ge f(\bar{x}) + f'(\bar{x}; x - \bar{x}), \;\; \forall x \in \mathcal{N}. \tag{9}$$

Slightly generalizing the family of functions in Cui et al. [8, display (25)], a large class of convex-like functions is given by the composition of convex functions and piecewise affine functions:

$$f = \varphi \circ \Theta \circ \psi,$$

where $\varphi : \mathbb{R}^L \to \mathbb{R}$ is (multivariate) piecewise affine and *isotone* (i.e., $\varphi(z) \ge \varphi(z')$ for any two L-dimensional vectors $z \ge z'$); $\Theta : \mathbb{R}^m \to \mathbb{R}^L$ is a vector-valued function such that each of its component functions $\theta_\ell : \mathbb{R}^m \to \mathbb{R}$ for $\ell = 1, \ldots, L$ is convex; and $\psi : \mathbb{R}^n \to \mathbb{R}^m$ is a piecewise affine function. In classical nonlinear programming problems, the set X is often closed and takes the form

$$X \triangleq \{x \in P \mid f_k(x) \le 0, \, k = 1, \ldots, K\} \tag{10}$$

for some integer $K > 0$, where P is a polyhedron and each $f_k : \mathcal{O} \to \mathbb{R}$ is a B-differentiable function near a given $\bar{x} \in X$. We say that the *Abadie constraint qualification* (ACQ) holds at $\bar{x}$ if

$$\mathcal{T}(\bar{x}; X) = \{v \in \mathcal{T}(\bar{x}; P) \mid f_k'(\bar{x}; v) \le 0, \, k \in \mathcal{A}(\bar{x})\} \triangleq \mathcal{L}(\bar{x}; X),$$

where $\mathcal{A}(\bar{x}) \triangleq \{k \mid f_k(\bar{x}) = 0\}$ is the index set of the active constraints at $\bar{x}$. The following is proven in Cui et al. [8, proposition 9(ii)].

Proposition 6. *Let P be a polyhedron. Suppose that f and each f_k for $k = 1, \ldots, K$ are locally convex like near a B-stationary point $\bar{x}$ of* (3) *with X given by* (10). *If the ACQ holds at $\bar{x}$, then $\bar{x}$ is a local minimizer of f on X.*

The above is a B-stationarity sufficiency result, meaning that sufficient conditions are provided under which a B-stationary point is a local minimizer. We next introduce an important geometric property of an arbitrary set that allows us to establish epistationarity sufficiency (i.e., the question of when an epistationary point is a local minimizer).

Definition 2. A subset $S \subseteq \mathbb{R}^N$ is said to be *locally convex like* at a vector $\bar{z} \in S$ if there exists a neighborhood $\mathcal{N}$ of $\bar{z}$ such that $S \cap \mathcal{N} \subseteq \bar{z} + \mathcal{T}(\bar{z}; S)$.

Without involving stationarity, the next result shows that the functional convexity-like property implies the set-theoretic convexity-like property, under a suitable constraint qualification.

Proposition 7. *Let $X \triangleq \{x \in \mathbb{R}^n \,|\, f_k(x) \le 0,\ k = 1, \ldots, K\}$,where each f_k is B-differentiable near $\bar{x} \in X$. If each f_k for $k \in \mathcal{A}(\bar{x})$ is locally convex like near $\bar{x}$ and the ACQ holds at $\bar{x}$ for the set X, then the set X is locally convex-like near $\bar{x}$.*

Proof. By the local convexity like of each f_k near $\bar{x}$ for $k \in \mathcal{A}(\bar{x})$, there exists a neighborhood $\mathcal{N}$ of $\bar{x}$ such that

$$f_k(x) \ge f_k(\bar{x}) + f_k^{'}(\bar{x}; x - \bar{x}),\ \forall k \in \mathcal{A}(\bar{x}) \text{ and } \forall x \in \mathcal{N}.$$

Hence, if $x \in X \cap \mathcal{N}$, the above inequalities imply that $f_k^{'}(\bar{x}; x - \bar{x}) \le 0$ for all $k \in \mathcal{A}(\bar{x})$. Hence, $x - \bar{x} \in \mathcal{T}(\bar{x}; X)$ under the ACQ. Because $x \in X \cap \mathcal{N}$ is arbitrary, it follows that $X \cap \mathcal{N} \subseteq \bar{x} + \mathcal{T}(\bar{x}; X)$, establishing the local convexity like of the set X near $\bar{x}$. □

A further connection between locally convex-like functions and locally convex-like sets is presented in the next result.

Proposition 8. *A B-differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ near $\bar{x}$ is locally convex like at $\bar{x}$ if and only if its epigraph* epi(f) *is locally convex like at $(\bar{x}, f(\bar{x}))$.*

Proof. By Rockafellar and Wets [23, theorem 8.2], it holds that $\mathcal{T}((\bar{x}, f(\bar{x})); \text{epi}(f)) = \text{epi}(f'(\bar{x}; \bullet))$. Hence, with $h(x) \triangleq f(\bar{x}) + f'(\bar{x}; x - \bar{x})$, it follows that $\mathcal{T}((\bar{x}, f(\bar{x})); \text{epi}(f)) + \{(\bar{x}, f(\bar{x}))\} = \text{epi}(h)$. By definition, f is locally convex like at $\bar{x}$ if and only if there exists a neighborhood $\mathcal{N}$ of $\bar{x}$ such that $f(x) \ge h(x)$ for all $x \in \mathcal{N}$; equivalently, $\text{epi}(f) \cap \widehat{\mathcal{N}} \subseteq \text{epi}(h)$, where $\widehat{\mathcal{N}} \triangleq \mathcal{N} \times \mathbb{R}$. Hence, f is locally convex like at $\bar{x}$ if and only if there exists a neighborhood $\mathcal{N}$ of $\bar{x}$ such that

$$\text{epi}(f) \cap \widehat{\mathcal{N}} \subseteq \mathcal{T}((\bar{x}, f(\bar{x})); \text{epi}(f)) + \{(\bar{x}, f(\bar{x}))\},$$

which is the local convexity-like property of epi(f) at $(\bar{x}, f(\bar{x}))$. □

The next result establishes the promised epistationarity sufficiency under the set-theoretic local convexity-like property; it highlights the fundamental role of the latter property in the local optimality theory of optimization problems lacking lower semicontinuity.

Proposition 9. *If the set Z defined in (5) is locally convex like at $\bar{z} \triangleq (\bar{x}, f(\bar{x}))$ and $\bar{x}$ is an epistationary point of (3), then $\bar{x}$ is a local minimizer of f on X.*

Proof. Let $\mathcal{N} = \mathcal{N}_x \times \mathcal{N}_t$ be a neighborhood of $\bar{z}$ such that $Z \cap \mathcal{N} \subseteq \bar{z} + \mathcal{T}(\bar{z}; Z)$. It suffices to show that $f(x') \ge f(\bar{x})$ for all $x' \in X \cap \mathcal{N}_x$. By way of contradiction, assume that there exists $x' \in X \cap \mathcal{N}_x$ such that $f(x') < f(\bar{x})$. Let $t' \in \mathcal{N}_t$ be such that $f(x') < t' < f(\bar{x})$. Then, $(x', t') \in Z \cap \mathcal{N}$. Thus, there exists $(dx, dt) \in \mathcal{T}(\bar{z}; Z)$ such that $(x', t') = (\bar{x}, f(\bar{x})) + (dx, dt)$. By epistationarity, we have $dt \ge 0$. However, then $t' = f(\bar{x}) + dt \ge f(\bar{x})$, which is a contradiction. □

Clearly, convex sets are locally convex like; although it is not too interesting from an optimization perspective, we remark that open sets are always locally convex like. The union of finitely many locally convex-like sets at a common vector is locally convex like at the vector; the Cartesian product of finitely many locally convex-like sets is locally convex like. In general, the intersection of locally convex-like sets is not necessarily locally convex like unless a suitable constraint qualification holds so that the tangent cone of the intersection of these sets is equal to the intersection of the respective tangent cones of the sets. This is illustrated in the following example.

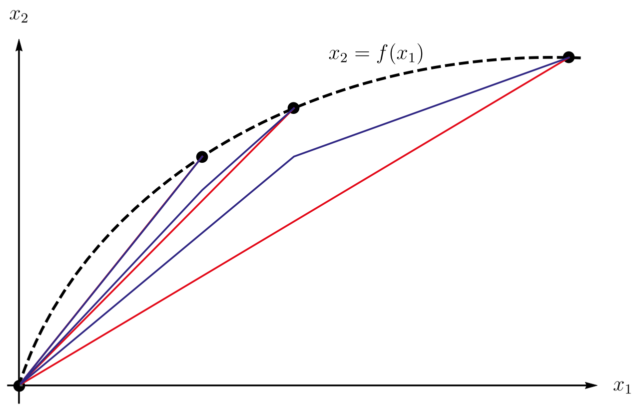
Example 1. Define $f(t) = \log(t + 1)$. Consider

$$X_1 = \bigcup_{n \in [N]} \left\{ (x, y) \,|\, y = 2nf\left(\frac{1}{2n}\right)x,\ x \in \left[0, \frac{1}{2n}\right] \right\}$$

$$X_2 = \bigcup_{n \in [N]} \left\{ (x, y) \,|\, y = (2n+1)f\left(\frac{1}{2n+2}\right)x,\ x \in \left[0, \frac{1}{2n+1}\right] \right\} \bigcup$$

$$\bigcup_{n \in [N]} \left\{ \begin{array}{l} (x, y) \,|\, y = f\left(\frac{1}{2n+2}\right) + \frac{f\left(\frac{1}{2n}\right) - f\left(\frac{1}{2n+2}\right)}{\frac{1}{2n} - \frac{1}{2n+1}} \left(x - \frac{1}{2n+1}\right) \\ x \in \left[\frac{1}{2n+1}, \frac{1}{2n}\right] \end{array} \right\}.$$

Figure 1. (Color online) Intersection of two locally convex-like sets. X_1 and X_2 consist of the red and blue line segments, respectively; their intersection is represented by the black points.



Then,

$$X_1 \cap X_2 = \bigcup_{n \in [N]} \left\{ \left(\frac{1}{2n}, f\left(\frac{1}{2n}\right) \right) \right\} \bigcup \{(0,0)\}$$

is a closed set but not locally convex like at (0, 0). See Figure 1 for the illustration.

The following example shows that unlike (quasi-)convex functions, the sublevel set of a locally convex-like function is generally not locally convex like.

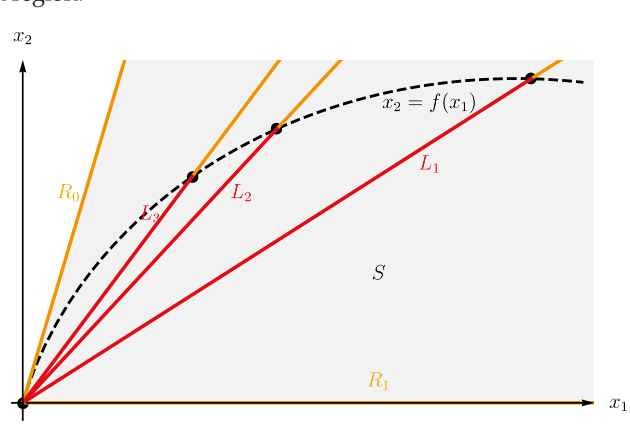
Example 2. Consider the two sets X_1 and X_2 given in Example 1. Let $R_0 = \{(t,t) : t \geq 0\}$ and $R_1 = \{(t,0), t \geq 0\}$ be two rays. Define $Y_i = X_i \cup R_0 \cup R_1$, $i = 1, 2$. Note that Y_1 and Y_2 are two closed convex-like sets. Define $f_i(x) = \mathrm{dist}(x, Y_i)$ for $i = 1, 2$. If f_1 and f_2 are locally convex like and B-differentiable, then so is $\max\{f_1, f_2\}$. However, the sublevel set $\{x : \max\{f_1, f_2\}(x) \leq 0\} = Y_1 \cap Y_2$ is not locally convex like for the similar reason as in Example 1.

Next, we prove that f_1 is indeed locally convex like and B-differentiable. Note that $X_1 = \cup_i L_i$, where each L_i is a line segment as shown in Figure 2. Thus, $f_1(x) = \min\{\min_i h_i(x), r_0(x), r_1(x)\}$, where $h_i(x) \triangleq \mathrm{dist}(x, L_i)$, $i = 1, 2, \ldots,$ and $r_j = \mathrm{dist}(x, R_j)$, $j = 1, 2$. Let $S = \mathrm{conv}(Y_1)$ and $r(x) = \mathrm{dist}(x, S)$. Consider an arbitrary $\overline{x} \in \mathbb{R}^n$. There are four cases.

- $\overline{x} \in \mathbb{R}^n \setminus S$. In this case, $f_1(\overline{x}) = r(\overline{x})$.
- $\overline{x}$ is an inner point of S. In this case, the set of active pieces $\{i : f(x) = h_i(x)\} \subseteq \{i : f(\overline{x}) = (\overline{x})\}$ is finite near $\overline{x}$.
- $\overline{x} \in (R_0 \cup R_1) \setminus \{0\}$. In this case, $f_1(x) = r_0(x)$ or $r_1(x)$ near $\overline{x}$.
- $\overline{x} = 0$.

In the first three cases, it can be seen easily that f_1 is a pointwise minimum of a finite number of convex functions near $\overline{x}$, which implies f_1 locally convex like and B-differentiable at $\overline{x}$; see Figure 2 for illustration. It remains

Figure 2. (Color online) Illustration of Example 2. X_1 and R consist of the red line segments and orange rays, respectively. The set S is represented by the shaded region.



to show that f_1 is convex like and directionally differentiable at $\bar{x} = 0$. Define a closed set $R = \{r : r = tx, t \ge 0, x \in Y_1\}$ as the cone generated by Y_1. Note that $f'(0; d) = \text{dist}(d, R)$. Indeed, if $d = (1, 1)$, then $f'(0; d) = \text{dist}(d, R) = 0$. If d is not a scalar multiple of (1, 1), then because Y_1 is locally a finite union of line segments near td for $t > 0$ small enough, $f'(0; d) = \text{dist}(d, R)$. Because $X_1 \subseteq R$, we have $f_1'(0; d) \le f_1(d)$, and thus, f_1 is locally convex like at zero. The above arguments can be extended to prove that f_2 is a locally convex-like function in a similar way. We omit the details.

It turns out that the gap between the everywhere local convexity-like property and the global convexity is the Clarke regularity, as can be seen from the following proposition.

Proposition 10. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be locally convex like at every point in $\mathbb{R}^n$. Then, f is Clarke regular at every point in $\mathbb{R}^n$ if and only if it is convex on $\mathbb{R}^n$.*

Proof. Taking an arbitrary reference point $\bar{x} \in \mathbb{R}^n$ and an arbitrary direction $d \in \mathbb{R}^n$, we define a univariate function $g(t) = f(\bar{x} + td)$. Note that g is convex like by definition. It suffices to prove that g is convex, which amounts to $g(t) \ge g(\bar{t}) + g'(\bar{t}; t - \bar{t})$ for all $t, \bar{t} \in \mathbb{R}$. Assume for contradiction that there exist t_1 and t_0 such that $g(t_1) < g(t_0) + g'(t_0; t_1 - t_0)$. Let $h(t) = g(t) - g'(t_0; t - t_0) - g(t_0)$. Without loss of generality, we also assume $t_1 > t_0$. Define $S = \arg\max\{h(t) : t_0 \le t \le t_1\}$, which is a compact set. Let $t_* = \max\{t : t \in S\}$. Then, by construction, $h(t) < h(t_*)$ for $t_* < t \le t_1$. Because $h(t_0) = 0 > h(t_1)$, one has $t_* < t_1$. Thus, we have either $t_* = t_0$ or $t_0 < t_* < t_1$. These two cases are addressed below.

- Case 1. $t_* = t_0$. In this case, $h'(t_*; 1) = h'(t_0; 1) = 0$. By the local convexity-like property of h over (t_0, ∞), an $\varepsilon > 0$ exists such that for $t_* \le t < t_* + \varepsilon$, one has $h(t) \ge h(t_*) + h'(t_*; t - t_*) = h(t_*)$. However, this contradicts $h(t) < h(t_*)$ for all $t_* < t \le t_1$.
- Case 2. $t_0 < t^* < t_1$. Because $h(t) < h(t_*)$ for all $t_* < t \le t_1$, one can deduce that $h'(t_*; 1) \le 0$. If $h'(t^*; 1) = 0$, we can repeat the same argument in the first case to draw a contradiction. For this reason, we assume $h'(t^*; 1) < 0$. Because $h(t) = g(t) - (t - t_0)g'(t_0; 1) - g(t_0)$ for $t \ge t_0$ by the Clarke regularity of g, it follows that $h'(t_*; \bullet)$ is convex; thus, $h'(t_*; 1) + h'(t_*; -1) \ge h'(t^*; 0) = 0$, which implies $h'(t_*; -1) > 0$. However, this indicates that $h(t) > h(t^*)$ for all t smaller than but close enough to t_*, contradicting the fact that $t_* \in S$. □

Assume $X \triangleq \{x \in \mathbb{R}^n \,|\, f_k(x) \le 0, k = 1, \ldots, K\}$, where each f_k is a locally convex-like function. Proposition 10 implies that if X is a locally convex like but not convex set, then at least one f_k is nondifferentiable. Another immediate consequence of this proposition is that if f is a PC^1 (piecewise continuously differentiable) function with convex element functions (i.e., if f is continuous and there exist finitely many convex differentiable functions $\{f_i\}_{i=1}^I$ such that $f(x) \in \{f_i(x)\}_{i=1}^I$ for all $x \in \mathbb{R}^n$), then f is convex if and only if it is Clarke regular. This is because such a function f must be locally convex like at every point in $\mathbb{R}^n$.

6. Tangents of Heaviside Composite Constraints

As the tangent cone plays an important role in the local convexity-like property and is of independent interest, it would be useful to describe the tangent vectors of the set X_{HSC}. Such descriptions will be instrumental to demonstrate the local convexity-like property of X_{HSC} at $\bar{x} \in X_{\text{HSC}}$ under appropriate assumptions of the defining functions; see Table 1. We start with the ASC set X_{ASC} whose tangent cone at a vector $\bar{x} \in X_{\text{ASC}}$ is known (Dong et al. [11, proposition 10]). Specifically, we have

$$\mathcal{T}(\bar{x}; X_{\text{ASC}}) = \text{cl}\left\{ v \in \mathcal{T}(\bar{x}; P) \;\middle|\; \sum_{j \notin \bar{\beta}} a_{ij} |v_j|_0 \le \underbrace{b_i - \sum_{j=1}^{n} a_{ij} |\bar{x}_j|_0}_{= b_i - \sum_{j \in \bar{\beta}} a_{ij} \ge 0}, \; i = 1, \ldots, m \right\}, \tag{11}$$

Table 1. Conditions for the local convexity-like property of X_{HSC}.

ψ_{ij} \ ϕ_{ij}	Convex	Piecewise affine
Convex	$\psi_{ij} \ge 0$	Free
Piecewise affine	$\psi_{ij} \ge 0$	Free

where $\overline{\beta} \triangleq \{i | \overline{x}_i \neq 0\} \triangleq \text{supp}(\overline{x})$ is the *support* of the vector $\overline{x}$. (We remark that although the proof of this representation in the reference has the side polyhedron P being the entire space, the proof therein applies to P being a proper polyhedral set.) The closure on the right-hand cone in (11) can be removed if all the coefficients a_{ij} are nonnegative as this cone itself is closed in this case. Based on the above representation, the following result is easy to prove.

Proposition 11. *Let P be a polyhedron. The set X_{ASC} is locally convex like at every $\overline{x}$ in X_{ASC}.*

Proof. Let $\mathcal{N}$ be a neighborhood of $\overline{x}$ such that $x_j \neq 0$ for all $j \in \overline{\beta}$ and all $x \in \mathcal{N}$. Let $x \in X_{\text{ASC}} \cap \mathcal{N}$. Then, we have

$$\begin{aligned} b_i \geq \sum_{j=1}^{n} a_{ij} |x_j|_0 &= \sum_{j \notin \overline{\beta}} a_{ij} |x_j|_0 + \sum_{j \in \overline{\beta}} a_{ij} |x_j|_0 \\ &= \sum_{j \notin \overline{\beta}} a_{ij} |x_j - \overline{x}_j|_0 + \sum_{j \in \overline{\beta}} a_{ij}. \end{aligned}$$

Thus, $b_i - \sum_{j \in \overline{\beta}} a_{ij} \geq \sum_{j \notin \overline{\beta}} a_{ij} |x_j - \overline{x}_j|_0$ for all $i = 1, \ldots, m$. Hence, $x - \overline{x} \in \mathcal{T}(\overline{x}; X_{\text{ASC}})$. □

As a preliminary result for the set X_{HSC}, we consider the case where each function ψ_{ij} is affine and ϕ_{ij} is piecewise affine. First, we derive an explicit expression of the tangent cone of X_{HSC} at an arbitrary vector $\overline{x} \in X_{\text{HSC}}$ and use this expression to show that (a) X_{HSC} is locally convex like at $\overline{x}$ and that (b) epistationarity of a B-differentiable objective function on the set X_{HSC} is sharper than pseudo-B-stationarity.

Proposition 12. *Let P be a polyhedron. Let each ψ_{ij} be an affine function and ϕ_{ij} be a piecewise affine function for all $j = 1, \ldots J_i$ and $i = 1, \ldots, m$. For $\overline{x} \in X_{\text{HSC}}$, it holds that*

$$\mathcal{T}(\overline{x}; X_{\text{HSC}}) \supseteq \textit{closure of} \left\{ v \in \mathcal{T}(\overline{x}; P) \;\middle|\; \begin{array}{l} \textit{for all } i = 1, \ldots, m: \\ \displaystyle\sum_{j \in \mathcal{J}_{i,0}(\overline{x})} \psi_{ij}(\overline{x}) \mathbf{1}_{(0,\infty)}(\phi'_{ij}(\overline{x}; v)) + \sum_{j \in \mathcal{J}_{i,+}(\overline{x})} \psi_{ij}(\overline{x}) \leq b_i \\ \textit{and if } \displaystyle\sum_{j \in \mathcal{J}_{i,0}(\overline{x})} \psi_{ij}(\overline{x}) \mathbf{1}_{(0,\infty)}(\phi'_{ij}(\overline{x}; v)) + \sum_{j \in \mathcal{J}_{i,+}(\overline{x})} \psi_{ij}(\overline{x}) = b_i, \textit{ then} \\ \displaystyle\sum_{j \in \mathcal{J}_{i,0}(\overline{x})} [\nabla \psi_{ij}(\overline{x})^\top v] \mathbf{1}_{(0,\infty)}(\phi'_{ij}(\overline{x}; v)) + \sum_{j \in \mathcal{J}_{i,+}(\overline{x})} \nabla \psi_{ij}(\overline{x})^\top v \leq 0 \end{array} \right\}. \tag{12}$$

Conversely, if the following two conditions hold for all $i = 1, \ldots, m$:

a. *for all $j \in \mathcal{J}_{i,0}(\overline{x})$,*

$$[v \in \mathcal{T}(\overline{x}; P) \textit{ and } \phi'_{ij}(\overline{x}; v) > 0] \Rightarrow \nabla \psi_{ij}(\overline{x})^\top v \leq 0; \quad \textit{and}$$

b. *for all $j \in \mathcal{J}_{i,+}(\overline{x})$, $\nabla \psi_{ij}(\overline{x}) \in \mathcal{T}(\overline{x}; P)^*$, where $\mathcal{T}(\overline{x}; P)^*$ is the dual of $\mathcal{T}(\overline{x}; P)$,*
then equality holds in (12).

Proof. Let $v \in \mathcal{T}(\overline{x}; P)$ satisfy the functional conditions in the right-hand set. We claim that v belongs to $\mathcal{T}(\overline{x}; X_{\text{HSC}})$ by showing that $\overline{x}^\tau \triangleq \overline{x} + \tau v \in X_{\text{HSC}}$ for all $\tau > 0$ sufficiently small that depends on v. Once this is shown, the one-side inclusion $\supseteq$ of the two cones in (12) follows. Because P is a polyhedron, we have $x^\tau \in P$ for all $\tau > 0$ sufficiently small. Moreover, by continuity of ϕ_{ij}, we have

$$[\phi_{ij}(\overline{x}) > 0 \Rightarrow \phi_{ij}(\overline{x}^\tau) > 0] \text{ and } [\phi_{ij}(\overline{x}) < 0 \Rightarrow \phi_{ij}(\overline{x}^\tau) < 0]$$

for all $\tau > 0$ sufficiently small. Hence,

$$\sum_{j=1}^{J_i} \psi_{ij}(\overline{x}^\tau) \mathbf{1}_{(0,\infty)}(\phi_{ij}(\overline{x}^\tau)) = \sum_{j \in \mathcal{J}_{i,0}(\overline{x})} \psi_{ij}(\overline{x}^\tau) \mathbf{1}_{(0,\infty)}(\phi_{ij}(\overline{x}^\tau)) + \sum_{j \in \mathcal{J}_{i,+}(\overline{x})} \psi_{ij}(\overline{x}^\tau).$$

Because ϕ_{ij} is piecewise affine, it follows that if $\tau > 0$ is sufficiently small, we have

$$\phi_{ij}(\overline{x}^\tau) = \phi_{ij}(\overline{x}) + \tau \phi'_{ij}(\overline{x}; v) = \tau \phi'_{ij}(\overline{x}; v), \text{ if } j \in \mathcal{J}_{i,0}(\overline{x}).$$

Therefore, we can further derive that

$$\sum_{j=1}^{J_i} \psi_{ij}(\bar{x}^\tau)\,\mathbf{1}_{(0,\infty)}(\phi_{ij}(\bar{x}^\tau)) = \sum_{j\in\mathcal{J}_{i,0}(\bar{x})} \psi_{ij}(\bar{x})\,\mathbf{1}_{(0,\infty)}(\phi'_{ij}(\bar{x};v)) + \sum_{j\in\mathcal{J}_{i,+}(\bar{x})} \psi_{ij}(\bar{x})$$
$$+\ \tau\left\{\sum_{j\in\mathcal{J}_{i,0}(\bar{x})} [\nabla\psi_{ij}(\bar{x})^\top v]\,\mathbf{1}_{(0,\infty)}(\phi'_{ij}(\bar{x};v)) + \sum_{j\in\mathcal{J}_{i,+}(\bar{x})} \nabla\psi_{ij}(\bar{x})^\top v\right\}.$$

Hence, with v as specified, it follows that for $\tau > 0$ sufficiently small, which depends on v, we have $\sum_{j=1}^{J_i}\psi_{ij}(\bar{x}^\tau)\,\mathbf{1}_{(0,\infty)}(\phi_{ij}(\bar{x}^\tau)) \le b_i$ for all i. Thus, $\bar{x}^\tau \in X_{\mathrm{HSC}}$.

Conversely, let $v \in \mathcal{T}(\bar{x}; X_{\mathrm{HSC}})$. Let $\{x^\nu\} \subset X_{\mathrm{HSC}}$ be a sequence converging to $\bar{x}$ and $\{\tau_\nu\} \downarrow 0$ such that $v = \lim_{\nu\to\infty} w^\nu$, where $w^\nu \triangleq \frac{x^\nu - \bar{x}}{\tau_\nu}$ clearly belongs to $\mathcal{T}(\bar{x}; P)$. Moreover, we have $\phi_{ij}(x^\nu) > 0$ for all ν sufficiently large, all $j \in \mathcal{J}_{i,+}(\bar{x})$, all $i = 1, \ldots, m$. We have for all $i = 1, \ldots, m$,

$$\begin{aligned} b_i &\ge \sum_{j=1}^{J_i} \psi_{ij}(x^\nu)\,\mathbf{1}_{(0,\infty)}(\phi_{ij}(x^\nu)) \\ &= \sum_{j\in\mathcal{J}_{i,0}(\bar{x})} \psi_{ij}(x^\nu)\,\mathbf{1}_{(0,\infty)}(\phi_{ij}(x^\nu)) + \sum_{j\in\mathcal{J}_{i,+}(\bar{x})} \psi_{ij}(x^\nu) \\ &= \sum_{j\in\mathcal{J}_{i,0}(\bar{x})} [\psi_{ij}(\bar{x}) + \nabla\psi_{ij}(\bar{x})^\top (x^\nu - \bar{x})]\mathbf{1}_{(0,\infty)}(\phi'_{ij}(\bar{x}; x^\nu - \bar{x})) \\ &\quad + \sum_{j\in\mathcal{J}_{i,+}(\bar{x})} [\psi_{ij}(\bar{x}) + \nabla\psi_{ij}(\bar{x})^\top (x^\nu - \bar{x})]. \end{aligned}$$

Hence, we obtain that

$$b_i - \sum_{j\in\mathcal{J}_{i,0}(\bar{x})} \psi_{ij}(\bar{x})\,\mathbf{1}_{(0,\infty)}(\phi'_{ij}(\bar{x};w^\nu)) - \sum_{j\in\mathcal{J}_{i,+}(\bar{x})} \psi_{ij}(\bar{x})$$
$$\ge\ \tau_\nu\left\{\sum_{j\in\mathcal{J}_{i,0}(\bar{x})} [\nabla\psi_{ij}(\bar{x})^\top w^\nu]\,\mathbf{1}_{(0,\infty)}(\phi'_{ij}(\bar{x};w^\nu)) + \sum_{j\in\mathcal{J}_{i,+}(\bar{x})} \nabla\psi_{ij}(\bar{x})^\top w^\nu\right\}.$$

Under the two assumed conditions (a) and (b), the right side of the above expression is nonnegative because $w^\nu \in \mathcal{T}(\bar{x}; P)$; hence, so is the left-hand side, which shows that w^ν satisfies

$$\sum_{j\in\mathcal{J}_{i,0}(\bar{x})} \psi_{ij}(\bar{x})\,\mathbf{1}_{(0,\infty)}(\phi'_{ij}(\bar{x};w^\nu)) + \sum_{j\in\mathcal{J}_{i,+}(\bar{x})} \psi_{ij}(\bar{x}) \le b_i,\ \forall i = 1, \ldots, m.$$

Moreover, if for some i, it holds that

$$\sum_{j\in\mathcal{J}_{i,0}(\bar{x})} \psi_{ij}(\bar{x})\,\mathbf{1}_{(0,\infty)}(\phi'_{ij}(\bar{x};w^\nu)) + \sum_{j\in\mathcal{J}_{i,+}(\bar{x})} \psi_{ij}(\bar{x}) = b_i,$$

then

$$\sum_{j\in\mathcal{J}_{i,0}(\bar{x})} [\nabla\psi_{ij}(\bar{x})^\top w^\nu]\,\mathbf{1}_{(0,\infty)}(\phi'_{ij}(\bar{x};w^\nu)) + \sum_{j\in\mathcal{J}_{i,+}(\bar{x})} \nabla\psi_{ij}(\bar{x})^\top w^\nu = 0.$$

Hence, w^ν belongs to the right-hand set in (12) without the closure. Because v is the limit of $\{w^\nu\}$, it follows that v belongs to the closure of this set. Hence, equality holds in (12). □

Remark 2. In fact, the piecewise affinity assumption of each ϕ_{ij} in Proposition 12 can be relaxed to the local convexity-like property at $\bar{x}$ in a straightforward manner.

Clearly, conditions (a) and (b) hold trivially if each ψ_{ij} is a constant function: that is, for the set

$$X_{\mathrm{AHC}} \triangleq \left\{ x \in P \,\middle|\, \sum_{j=1}^{J_i} a_{ij}\,\mathbf{1}_{(0,\infty)}(\phi_{ij}(x)) \le b_i,\ i = 1, \ldots, m \right\}.$$

If each ϕ_{ij} is piecewise affine, the result in Proposition 12 directly extends the tangent cone expression of X_{ASC} with the ℓ_0-function replaced by the Heaviside function composed with a piecewise affine function. This leads to

$$\mathcal{T}(\bar{x}; X_{\text{AHC}}) = \text{cl}\left\{ v \in \mathcal{T}(\bar{x}; P) \,\middle|\, \begin{array}{l} \displaystyle\sum_{j \in \mathcal{J}_{i,0}(\bar{x})} a_{ij}\, \mathbf{1}_{(0,\infty)}(\phi'_{ij}(\bar{x}; v)) \\ \displaystyle + \sum_{j \in \mathcal{J}_{i,+}(\bar{x})} a_{ij} \le b_i, \quad i = 1, \cdots, m \end{array} \right\}.$$

Like $\mathcal{T}(\bar{x}; X_{\text{ASC}})$ in (11), the closure operation can be dropped if the coefficients a_{ij} are all nonnegative. Note also that the above representation of $\mathcal{T}(\bar{x}; X_{\text{AHC}})$ and that of the $\mathcal{T}(\bar{x}; X_{\text{ASC}})$ require no "constraint qualifications," although both the ℓ_0 function and the Heaviside function are discontinuous. The local convexity-like property of the set X_{AHC} follows readily from its tangent cone representation and the proof of the converse part of Proposition 12; no proof is needed.

Corollary 2. *Let P be a polyhedron. If each function ϕ_{ij} is piecewise affine, then the set X_{AHC} is locally convex like near every $\bar{x} \in X_{\text{AHC}}$.*

We next give a full description of the tangent cone $\mathcal{T}(\bar{x}; X_{\text{HSC}})$ under a sign restriction on the functions $\{\psi_{ij}\}$ for $j \in [J_i] \triangleq \{1, \ldots, J_i\}$ and $i = 1, \ldots, m$. Let $\Xi(\bar{x})$ and $\Xi^c(\bar{x})$ be families of complementary index tuples $\boldsymbol{\alpha} \triangleq (\alpha_i)_{i=1}^m$ and $\boldsymbol{\alpha}^c \triangleq (\alpha_i^c)_{i=1}^m$, respectively, where each $\alpha_i \subseteq \mathcal{J}_{i,0}(\bar{x})$ for $i = 1, \ldots, m$ and α_i^c is the complement of α_i in $\mathcal{J}_{i,0}(\bar{x})$. For each tuple $\boldsymbol{\alpha} \in \Xi(\bar{x})$ with complement $\boldsymbol{\alpha}^c \in \Xi^c(\bar{x})$, define the set

$$\mathcal{S}_{\boldsymbol{\alpha}}(\bar{x}) \triangleq \left\{ x \in P \,\middle|\, \begin{array}{l} \displaystyle\sum_{j \in \alpha_i} \psi_{ij}(x) + \sum_{j \in \mathcal{J}_{i,+}(\bar{x})} \psi_{ij}(x) \le b_i, \quad i = ,1, \ldots, m \\ \phi_{ij}(x) \le 0, j \in \alpha_i^c, i = 1, \ldots, m \end{array} \right\},$$

which may or may not contain the vector $\bar{x}$. Let $\bar{\Xi}(\bar{x})$ be the subfamily of $\Xi(\bar{x})$ consisting of tuples $\boldsymbol{\alpha}$ for which $\bar{x} \in \mathcal{S}_{\boldsymbol{\alpha}}(\bar{x})$. Under a nonnegativity condition on the functions ψ_{ij}, the following result gives a complete description of $\mathcal{T}(\bar{x}; X_{\text{HSC}})$ in terms of the sets $\mathcal{S}_{\boldsymbol{\alpha}}(\bar{x})$ for all tuples $\boldsymbol{\alpha} \in \bar{\Xi}(\bar{x})$; in turn, this can be used to obtain a characterization of epistationarity of (3) without f being B-differentiable.

Proposition 13. *Let each ϕ_{ij} and ψ_{ij} be continuous near $\bar{x}$. If ψ_{ij} is nonnegative in a neighborhood of $\bar{x}$ for all $j \in \mathcal{J}_{i,0}(\bar{x})$, then*

$$\mathcal{T}(\bar{x}; X_{\text{HSC}}) = \bigcup_{\boldsymbol{\alpha} \in \bar{\Xi}(\bar{x})} \mathcal{T}(\bar{x}; \mathcal{S}_{\boldsymbol{\alpha}}(\bar{x})). \tag{13}$$

Hence, if for all $\boldsymbol{\alpha} \in \bar{\Xi}(\bar{x})$, the set $\mathcal{S}_{\boldsymbol{\alpha}}(\bar{x})$ is locally convex like at $\bar{x}$, then so is X_{HSC}. In particular, this holds if all ϕ_{ij} and ψ_{ij} are convex, with the latter being nonnegative also.

Proof. We first show that there exists a neighborhood $\mathcal{N}$ of $\bar{x}$ such that

$$X_{\text{HSC}} \cap \mathcal{N} = \left(\bigcup_{\boldsymbol{\alpha} \in \Xi(\bar{x})} \mathcal{S}_{\boldsymbol{\alpha}}(\bar{x}) \right) \cap \mathcal{N}. \tag{14}$$

We choose $\mathcal{N}$ to be such that ψ_{ij} is nonnegative in $\mathcal{N}$ and

$$[\phi_{ij}(\bar{x}) > 0 \Rightarrow \phi_{ij}(x) > 0] \text{ and } [\phi_{ij}(\bar{x}) < 0 \Rightarrow \phi_{ij}(x) < 0], \ \forall x \in \mathcal{N}.$$

For a vector x in the left-hand intersection of (14), it is clear that $x \in \mathcal{S}_{\boldsymbol{\alpha}}(\bar{x})$, where

$$\alpha_i \triangleq \{ j \in \mathcal{J}_{i,0}(\bar{x}) \,|\, \phi_{ij}(x) > 0 \}, i = 1, \ldots, m.$$

Conversely, suppose $x \in \mathcal{S}_{\boldsymbol{\alpha}}(\bar{x}) \cap \mathcal{N}$ for some tuple $\boldsymbol{\alpha} \in \Xi(\bar{x})$; then, by the nonnegativity of ψ_{ij} in $\mathcal{N}$, we have because $\mathcal{J}_{i,+}(x) \subseteq \alpha_i \cup \mathcal{J}_{i,+}(\bar{x})$,

$$b_i \ge \sum_{j \in \alpha_i} \psi_{ij}(x) + \sum_{j \in \mathcal{J}_{i,+}(\bar{x})} \psi_{ij}(x) \ge \sum_{j \in \mathcal{J}_{i,+}(x)} \psi_{ij}(x), \quad i = 1, \ldots, m,$$

showing that $x \in X_{\text{HSC}}$. Thus, (14) holds. To see how (14) implies (13), we note that the right-hand union of tangent cones in (13) is necessarily a subcone of the left-hand cone. Conversely, for a vector v in $\mathcal{T}(\bar{x}; X_{\text{HSC}})$, let $\{x^k\} \subset X_{\text{HSC}}$ be a sequence converging to $\bar{x}$ and $\{\tau_k\} \downarrow 0$ such that $v = \lim_{k\to\infty} \frac{x^k - \bar{x}}{\tau_k}$. By (14), we may assume with no loss of generality that there exists $\boldsymbol{\alpha} \in \boldsymbol{\Xi}(\bar{x})$ such that $x^k \in \mathcal{S}_{\boldsymbol{\alpha}}(\bar{x})$ for all k. Such an index tuple $\boldsymbol{\alpha}$ must necessarily be an element of $\bar{\boldsymbol{\Xi}}(\bar{x})$ by continuity of ψ_{ij}. This shows that $v \in \mathcal{T}(\bar{x}; \mathcal{S}_{\boldsymbol{\alpha}}(\bar{x}))$ for an index tuple $\boldsymbol{\alpha} \in \bar{\boldsymbol{\Xi}}(\bar{x})$, completing the proof of (13). The next-to-last statement of the proposition is clear because the union of a finite number of convex-like sets each containing a common vector (which in this case, is $\bar{x}$) is locally convex like near the vector. □

Remark 3. Expression (14) shows that for any closed set $S \subseteq \mathcal{N}$, the set $X_{\mathrm{HSC}} \cap S$ is closed, provided that the functions ψ_{ij} and ϕ_{ij} are continuous and that ψ_{ij} is nonnegative.

The example below shows that the nonnegativity assumption on ψ_{ij} is essential for Equality (13) to hold and that the piecewise affine property of the ϕ_{ij} functions is essential for the validity of Proposition 14.

Example 3. Let

$$X = \{(x_1, x_2) \in \mathbb{R}^2 \mid -x_1 - \mathbf{1}_{(0,\infty)}(x_1^2 + x_2^2 - 1) \le -1\}.$$

Then, $X = \{(1,0)\} \cup \{(x_1, x_2) \in \mathbb{R}_+ \times \mathbb{R} \mid x_1^2 + x_2^2 > 1\}$. With $\bar{x} = (1,0)$, we have $\mathcal{T}(\bar{x}; X) = \mathbb{R}_+ \times \mathbb{R}$; it is easy to see that X is not convex like near $\bar{x}$. Thus, Equality (13) cannot hold.

We next give a different set of assumptions of the component functions ψ_{ij} and ϕ_{ij} for the set X_{HSC} to be locally convex like. On one hand, we replace the nonnegativity of ψ_{ij} by its convexity; on the other hand, we restrict ϕ_{ij} to be piecewise affine. This combination, therefore, generalizes the setting of Corollary 2; the proof employs a subset of each $\mathcal{S}_{\boldsymbol{\alpha}}(\bar{x})$ in which the piecewise structure of each ϕ_{ij} can be easily exposed:

$$\widehat{\mathcal{S}}_{\boldsymbol{\alpha}}(\bar{x}) \triangleq \bigcap_{i=1}^{m} \left\{ x \in P \;\middle|\; \begin{array}{l} \displaystyle\sum_{j \in \alpha_i} \psi_{ij}(x) + \sum_{j \in \mathcal{J}_{i,+}(\bar{x})} \psi_{ij}(x) \le b_i \\ \phi_{ij}(x) > 0, \ \forall j \in \alpha_i \\ \phi_{ij}(x) \le 0, \ \forall j \in \alpha_i^c \end{array} \right\}, \quad \begin{array}{l} \boldsymbol{\alpha} \in \boldsymbol{\Xi}(\bar{x}), \text{ where the} \\ \text{pair } (\alpha_i, \alpha_i^c) \text{ partitions the} \\ \text{index set } \mathcal{J}_{i,0}(\bar{x}) \text{ for } i \in [m]. \end{array}$$

We note that $\widehat{\mathcal{S}}_{\boldsymbol{\alpha}}(\bar{x}) \subseteq X_{\mathrm{HSC}} \cap \mathcal{S}_{\boldsymbol{\alpha}}(\bar{x})$; yet, $\bar{x} \notin \widehat{\mathcal{S}}_{\boldsymbol{\alpha}}(\bar{x})$ as long as α_i is nonempty for some i.

Proposition 14. *Let P be a polyhedron. If each function ψ_{ij} is convex and each function ϕ_{ij} is piecewise affine, then the set X_{HSC} is locally convex like at every one of its elements.*

Proof. Let $\bar{x} \in X_{\mathrm{HSC}}$ be arbitrary. With the same neighborhood $\mathcal{N}$ of $\bar{x}$ as defined in the proof of Proposition 13, it can similarly be proved that (no sign restriction on ψ_{ij} is needed)

$$X_{\mathrm{HSC}} \cap \mathcal{N} = \left(\bigcup_{\boldsymbol{\alpha} \in \boldsymbol{\Xi}(\bar{x})} \widehat{\mathcal{S}}_{\boldsymbol{\alpha}}(\bar{x}) \right) \cap \mathcal{N}.$$

Without loss of generality, we may assume that $\widehat{\mathcal{S}}_{\boldsymbol{\alpha}}(\bar{x}) \neq \emptyset$ for all $\boldsymbol{\alpha} \in \boldsymbol{\Xi}(\bar{x})$. By the distributive laws of unions and intersections and by the piecewise affinity of the functions ϕ_{ij}, each $\widehat{\mathcal{S}}_{\boldsymbol{\alpha}}(\bar{x})$ is the finite union of nonempty convex (albeit not necessarily closed) sets, which we write as $\widehat{\mathcal{S}}_{\boldsymbol{\alpha}}(\bar{x}) = \cup_{i \in I_{\boldsymbol{\alpha}}} S_{\boldsymbol{\alpha}}^i$, where $I_{\boldsymbol{\alpha}}$ is a certain finite index set and where each $S_{\boldsymbol{\alpha}}^i$ is a certain nonempty convex (not necessarily closed) set. Thus,

$$X_{\mathrm{HSC}} \cap \mathcal{N} = \left(\bigcup_{\boldsymbol{\alpha} \in \boldsymbol{\Xi}(\bar{x})} \bigcup_{i \in I_{\boldsymbol{\alpha}}} S_{\boldsymbol{\alpha}}^i \right) \cap \mathcal{N}.$$

The convexity of S_{α}^i implies $\mathrm{cl}\, S_{\boldsymbol{\alpha}}^i \subseteq \bar{x} + \mathcal{T}(\bar{x}; \mathrm{cl}\, S_{\boldsymbol{\alpha}}^i)$, provided that $\bar{x} \in \mathrm{cl}\, S_{\boldsymbol{\alpha}}^i$. We may restrict the neighborhood $\mathcal{N}$ such that $\mathcal{N} \cap \mathrm{cl}\, S_{\boldsymbol{\alpha}}^i = \emptyset$ for all $i \in I_{\boldsymbol{\alpha}}$ and all $\boldsymbol{\alpha} \in \boldsymbol{\Xi}(\bar{x})$ such that $\bar{x} \notin \mathrm{cl}\, S_{\boldsymbol{\alpha}}^i$. Letting $\boldsymbol{\mathcal{I}}(\bar{x})$ be the collection of pairs $(i, \boldsymbol{\alpha})$ such that $\bar{x} \in \mathrm{cl}\, S_{\boldsymbol{\alpha}}^i$, we deduce

$$X_{\mathrm{HSC}} \cap \mathcal{N} = \left(\bigcup_{(i,\boldsymbol{\alpha}) \in \boldsymbol{\mathcal{I}}(\bar{x})} S_{\boldsymbol{\alpha}}^i \right) \cap \mathcal{N},$$

which yields

$$\mathcal{T}(\bar{x}; X_{\mathrm{HSC}}) = \bigcup_{(i,\boldsymbol{\alpha}) \in \boldsymbol{\mathcal{I}}(\bar{x})} \mathcal{T}(\bar{x}; \mathrm{cl}\, S_{\boldsymbol{\alpha}}^i).$$

Combining the last two expressions, we deduce

$$X_{\mathrm{HSC}} \cap \mathcal{N} \subseteq \bigcup_{(i,\boldsymbol{\alpha})\in\mathcal{I}(\bar{x})} [\bar{x} + \mathcal{T}(\bar{x}; \mathrm{cl}\ S^i_{\boldsymbol{\alpha}})] = \bar{x} + \bigcup_{(i,\boldsymbol{\alpha})\in\mathcal{I}(\bar{x})} \mathcal{T}(\bar{x}; \mathrm{cl}\ S^i_{\boldsymbol{\alpha}}).$$

Thus, $X_{\mathrm{HSC}} \cap \mathcal{N} \subseteq \bar{x} + \mathcal{T}(\bar{x}; X_{\mathrm{HSC}})$; hence, X_{HSC} is locally convex like at $\bar{x}$. □

The discussion of the section is summarized in Table 1. Each entry is indexed by a combination of convexity and piecewise affinity imposed over the functions ψ_{ij} and ϕ_{ij}, and it indicates whether the nonnegativity of the latter functions is needed to ensure the local convexity-like property of X_{HSC}. For example, the first entry implies that if each ϕ_{ij} is convex and each ψ_{ij} is nonnegative and convex, then X_{HSC} is locally convex like. The conclusion of the first column is given by Proposition 13. Proposition 14 illustrates the entry (1, 2). The conclusion corresponding to the last entry can be proven using similar polyhedral decomposition techniques as in the proof of Proposition 14.

Theorem 3. *The set X_{HSC} is locally convex like at every one of its elements if the assumptions given by any entry of Table 1 are true. In particular, X_{ASC} is locally convex like, and X_{AHC} is locally convex like if each ϕ_{ij} is piecewise affine.*

Combining Theorem 3 with Proposition 9, we obtain the following result for the Heaviside constrained optimization Problem (1).

Corollary 3. *Let P be a polyhedron. If the assumptions given by any entry of Table 1 hold for the functions ψ_{ij} and ϕ_{ij}, then a point is a local minimizer of (1) if and only if it is an epistationary point.*

7. Computation of Pseudo- and Epistationary Points via Lifting

The results in the last section are all derived under certain convexity/sign/piecewise affinity restrictions under which tangents of the set X_{HSC} are identified and its local convexity-like property is established. There has been no discussion, however, about how pseudo- or epistationary points of Problem (1) can potentially be computed. In this section, via lifting, we present formulations that make such computation possible. One such lifted formulation was provided in a previous work (Cui et al. [7, section 6]) for the constraint $\sum_{j=1}^{J_i} \psi_{ij}(x)\mathbf{1}_{(0,\infty)}(\phi_{ij}(x)) \le b_i$ and under a sign restriction of the function ψ_{ij} on the zero set of ϕ_{ij}. It was shown therein that a B-stationary solution of the lifted problem would yield a pseudostationary solution of the given HSC-constrained problem when the functions $\{\{\psi_{ij}, \phi_{ij}\}_{j=1}^{J_i}\}_{i=1}^m$ are B-differentiable. The significance of the results in this section is twofold; (a) the sign restriction can be removed via an alternative lifted formulation, and (b) a relaxation of the latter formulation provides a constructive pathway to compute an epistationary solution.

7.1. Derivation of the Lifted Formulations

The derivation of the lifted formulations consists of several steps, beginning with the expression of each function $\psi_{ij} = \psi^+_{ij} - \psi^-_{ij}$ as the difference of its nonnegative and nonpositive parts, respectively: $\psi^{\pm}_{ij} \triangleq \max(\pm\psi_{ij}, 0)$. Introducing an arbitrary scalar $\varepsilon \ge 0$, we note that

$$X_{\mathrm{HSC}} = \left\{ x \in P \;\middle|\; \begin{array}{l} \displaystyle\sum_{j=1}^{J_i} \psi^{+;\varepsilon}_{ij}(x)\mathbf{1}_{(0,\infty)}(\phi_{ij}(x)) \\ \displaystyle\le \sum_{j=1}^{J_i} \psi^{-;\varepsilon}_{ij}(x)\mathbf{1}_{(0,\infty)}(\phi_{ij}(x)) + b_i,\ i = 1,\ldots,m \end{array} \right\},$$

where $\psi^{\pm;\varepsilon}_{ij} \triangleq \psi^{\pm}_{ij} + \varepsilon$. The first lifting of the set X_{HSC} exploits the property that the function $x \longmapsto \psi^{+;\varepsilon}_{ij}(x)\mathbf{1}_{(0,\infty)}(\phi_{ij}(x))$ is lower semicontinuous if both ψ_{ij} and ϕ_{ij} are lower semicontinuous. Thus, we have the option of not lifting the sum $\sum_{j=1}^{J_i} \psi^{+;\varepsilon}_{ij}(x)\mathbf{1}_{(0,\infty)}(\phi_{ij}(x))$ and lifting only the products $\psi^{-;\varepsilon}_{ij}(x)\mathbf{1}_{(0,\infty)}(\phi_{ij}(x))$. This leads to the following lifting scheme:

- t-lifting:

$$\widehat{X^{t;\varepsilon}_{\mathrm{HSC}}} \triangleq \left\{ \begin{array}{l} x \in P \\ t_{ij} \in [0,1] \\ \text{all } i,j \end{array} \;\middle|\; \left. \begin{array}{l} \displaystyle\sum_{j=1}^{J_i} \psi^{+;\varepsilon}_{ij}(x)\mathbf{1}_{(0,\infty)}(\phi_{ij}(x)) \le \sum_{j=1}^{J_i} \psi^{-;\varepsilon}_{ij}(x)\, t_{ij} + b_i \\ t_{ij} \le \mathbf{1}_{(0,\infty)}(\phi_{ij}(x)),\ j = 1,\ldots,J_i; \end{array} \right\} i = 1,\ldots,m \right\},$$

which is connected to X_{HSC} via the equivalence: $x \in X_{\mathrm{HSC}}$ if and only if there exists t such that $(x,t) \in \widehat{X^{t;\varepsilon}_{\mathrm{HSC}}}$.

Next, we note that

$$t_{ij} \le \mathbf{1}_{(0,\infty)}(\phi_{ij}(x)) \Longleftrightarrow \exists\, y_{ij} \ge 0 \text{ such that } t_{ij} \le \min(\phi_{ij}^{+}(x) y_{ij}, 1). \tag{15}$$

Indeed, if the left-hand inequality holds, then we may let $y_{ij} \begin{cases} \ge \phi_{ij}(x)^{-1} & \text{if } \phi_{ij}(x) > 0 \\ = 0 & \text{otherwise.} \end{cases}$ Conversely, suppose there is y_{ij} such that the right-hand conditions are satisfied. If $\phi_{ij}(x) \le 0$, then the left-hand inequality implies $t_{ij} \le 0$, which is the same as the right-hand inequality in this case. If $\phi_{ij}(x) > 0$, then the left-hand inequality yields $t_{ij} \le 1$, which is the right-hand inequality in this case. Substituting the right-hand conditions in (15) to replace the left-hand conditions for all (i, j) in the set $\widehat{X_{\text{HSC}}^{t;\varepsilon}}$, we obtain the next level of lifting:

- (t, y)-lifting:

$$\widehat{X_{\text{HSC}}^{t,y;\varepsilon}} \triangleq \left\{ \begin{array}{l|l} x \in P & \text{for all } i = 1, \ldots, m: \\ t_{ij} \in [0,1], \text{ all } i,j & \displaystyle\sum_{j=1}^{J_i} \psi_{ij}^{+;\varepsilon}(x)\, \mathbf{1}_{(0,\infty)}(\phi_{ij}(x)) \le \sum_{j=1}^{J_i} \psi_{ij}^{-;\varepsilon}(x)\, t_{ij} + b_i \\ y_{ij} \ge 0, \text{ all } i,j & t_{ij} \le \phi_{ij}^{+}(x)\, y_{ij},\ j = 1, \ldots, J_i \end{array} \right\},$$

which is a closed set in the lifted (x, t, y)-space, provided that the functions ϕ_{ij} and ψ_{ij} are continuous.

The last lifting is the product $u_{ij} \triangleq \psi_{ij}^{+;\varepsilon}(x)\,\mathbf{1}_{(0,\infty)}(\phi_{ij}(x))$. There are two ways to do this; one is to apply the epigraphical approach (Cui et al. [7, section 7]), particularly Cui et al. [7, proposition 7], by considering the relaxation $u_{ij} \ge \psi_{ij}^{+;\varepsilon}(x)\mathbf{1}_{(0,\infty)}(\phi_{ij}(x))$ and replacing it using a piecewise composite function; this leads to

- (t, y, u)-lifting:

$$\widehat{X_{\text{HSC}}^{t,y,u;\varepsilon}} \triangleq \left\{ \begin{array}{l|l} & \displaystyle\sum_{j=1}^{J_i} u_{ij} \le \sum_{j=1}^{J_i} \psi_{ij}^{-;\varepsilon}(x)\, t_{ij} + b_i, \quad i = 1, \ldots, m \\ x \in P & \\ t_{ij} \in [0,1], \text{ all } i,j & \text{and for all } j = 1, \ldots, J_i \text{ and } i = 1, \ldots, m: \\ y_{ij} \ge 0, \text{ all } i,j & t_{ij} \le \phi_{ij}^{+}(x)\, y_{ij}, \text{ and} \\ u_{ij} \ge 0, \text{ all } i,j & \underbrace{\min\{\psi_{ij}^{+;\varepsilon}(x) - u_{ij}, \phi_{ij}(x)\} \le 0}_{\text{equivalent to } u_{ij} \ge \psi_{ij}^{+;\varepsilon}(x)\mathbf{1}_{(0,\infty)}(\phi_{ij}(x)),\ \text{given } u_{ij} \ge 0} \end{array} \right\},$$

which is also closed if ϕ_{ij} and ψ_{ij} are continuous; moreover, if these functions are B-differentiable, then all the inequalities in $\widehat{X_{\text{HSC}}^{t,y,u;\varepsilon}}$ are defined by B-differentiable function. Furthermore, if ϕ_{ij} and ψ_{ij} are difference-of-convex or piecewise affine functions, then the constraints in $\widehat{X_{\text{HSC}}^{t,y,u;\varepsilon}}$ are of the difference-of-convex kind; thus, optimization over this set can in principle be solved by the difference-of-convex methods described in Pang et al. [21].

An alternative to the piecewise min/max lifting of $\psi_{ij}^{+;\varepsilon}(x)\mathbf{1}_{(0,\infty)}(\phi_{ij}(x))$ is derived from the observation that

$$\sum_{j=1}^{J_i} \psi_{ij}^{+;\varepsilon}(x)\, \mathbf{1}_{(0,\infty)}(\phi_{ij}(x)) = \sum_{j=1}^{J_i} \psi_{ij}^{+;\varepsilon}(x)\, s_{ij},$$

where $s_{ij} \in [0,1]$ satisfies $\phi_{ij}^{+}(x)(1 - s_{ij}) = 0$. This leads to

- (t, y, s)-lifting:

$$\widetilde{X_{\text{HSC}}^{t,y,s;\varepsilon}} \triangleq \left\{ \begin{array}{l|l} x \in P & \displaystyle\sum_{j=1}^{J_i} \psi_{ij}^{+;\varepsilon}(x)\, s_{ij} \le \sum_{j=1}^{J_i} \psi_{ij}^{-;\varepsilon}(x)\, t_{ij} + b_i, \quad i = 1, \ldots, m \\ t_{ij},\, s_{ij} \in [0,1] \ \ \forall i, j & \text{and for all } j = 1, \ldots, J_i \text{ and } i = 1, \ldots, m: \\ y_{ij} \ge 0, \text{ all } i,j & t_{ij} \le \phi_{ij}^{+}(x)\, y_{ij}, \phi_{ij}^{+}(x)\,(1 - s_{ij}) = 0 \end{array} \right\}.$$

In the case of an affine sparsity constraint system

$$X_{\mathrm{ASC}} \triangleq \left\{ x \in P \, : \, \sum_{j=1}^{n} a_{ij} |x_j|_0 \, \le \, b_i, \, i = 1, \ldots, m \right\},$$

the resulting representations of the sets $\widehat{X^{t,y,u;\varepsilon}_{\mathrm{ASC}}}$ and $\widetilde{X^{t,y,s;\varepsilon}_{\mathrm{ASC}}}$ simplify somewhat; for simplicity, we give only the latter:

$$\widetilde{X^{t,y,s;\varepsilon}_{\mathrm{ASC}}} \triangleq \left\{ \begin{array}{l|l} x \in P & \displaystyle\sum_{j=1}^{n} (a^+_{ij} + \varepsilon) s_j \le \sum_{j=1}^{n} (a^-_{ij} + \varepsilon) t_j + b_i, \quad i = 1, \ldots, m \\ t_j, \, s_j \in [0,1], \text{ all } j & t_j \, \le \, |x_j| y_j, \quad j = 1, \ldots, n \\ y_j \ge 0, \text{ all } j & x_j (1 - s_j) = 0, \quad j = 1, \ldots, n \end{array} \right\},$$

where the only nonlinear functional constraints are defined by products of two variables. A noteworthy remark about $\widetilde{X^{t,y,s;\varepsilon}_{\mathrm{HSC}}}$ is that both auxiliary variables s_{ij} and t_{ij} are introduced as a surrogate for the same Heaviside composite term $\mathbf{1}_{(0,\infty)}(\phi_{ij}(x))$; their roles and constraints differ because of their associations with the respective signed functions $\psi^{\pm}_{ij}$.

The two lifted sets $\widehat{X^{t,y,u;\varepsilon}_{\mathrm{HSC}}}$ and $\widetilde{X^{t,y,s;\varepsilon}_{\mathrm{HSC}}}$ offer a computationally tractable venue for the minimization of a wide class of nonconvex nondifferentiable objective functions f over the nonclosed set X_{HSC}, provided that f and all the functions ϕ_{ij} and ψ_{ij} are surrogatable by pointwise minima of convex differentiable functions; see Cui and Pang [6, chapter 7]. We omit the algorithmic details.

7.2. Recovering Pseudostationarity

For simplicity, we assume that the objective function f (omitting the subscript HSC) in (1) is B-differentiable so that it is not necessary to work with the epigraphical formulation (5). We further assume that all the functions $\{\{\phi_{ij}, \psi_{ij}\}_{j=1}^{J_i}\}_{i=1}^{m}$ are B-differentiable (which does not imply that the set X_{HSC} is closed). In this subsection, we show that if $(\bar{x}, \bar{t}, \bar{y}, \bar{s})$ is any B-stationary tuple of f on $\widetilde{X^{t,y,s;\varepsilon}_{\mathrm{HSC}}}$, then $\bar{x}$ is pseudo-B-stationary of f on X_{HSC}; that is, $\bar{x}$ is a B-stationary point of the problem

$$\begin{array}{ll} \underset{x \in P}{\textbf{minimize}} & f(x) \\ \textbf{subject to} & \displaystyle\sum_{j \in \mathcal{J}_{i,+}(\bar{x})} \psi_{ij}(x) \le b_i, \quad i = 1, \ldots, m \\ \textbf{and} & \phi_{ij}(x) \le 0, \; \forall j \in \mathcal{J}_{i,\le}(\bar{x}), \, i = 1, \ldots, m. \end{array} \tag{16}$$

Note that for x sufficiently close to $\bar{x}$, we must have $\phi_{ij}(x) > 0$ for all $j \in \mathcal{J}_{i,+}(\bar{x})$ (confer (cf.) the constraints in (8)). Thus, the feasible region of (16) is "locally" a subset of X_{HSC} at $\bar{x}$ (i.e., there exists an open neighborhood $\mathcal{O}_{\bar{x}}$ of $\bar{x}$ such that if $x \in \mathcal{O}_{\bar{x}}$ is feasible to (16), then $x \in X_{\mathrm{HSC}}$). By the definition of the pseudo-B-stationarity of f on X_{HSC}, we know that $\bar{x}$ must be feasible to X_{HSC} that is possibly nonclosed. We omit the analysis for the set $\widehat{X^{t,y,u;\varepsilon}_{\mathrm{HSC}}}$ that involves pointwise minimum constraints. The proof of the proposition below is not straightforward as it requires the verification of significant details. Part of the challenge is that the triple $(\bar{t}, \bar{y}, \bar{s})$ is quite arbitrary and is related to $\bar{x}$ only through the constraints in $\widetilde{X^{t,y,s;\varepsilon}_{\mathrm{HSC}}}$. The scalar ε plays an important role for the validity of the result.

Proposition 15. *Let P be a polyhedron and $\varepsilon > 0$ be arbitrary. Let the functions f, ϕ_{ij}, and ψ_{ij} be B-differentiable near $\bar{x} \in P$. If the tuple $(\bar{x}, \bar{t}, \bar{y}, \bar{s})$ is a B-stationary point of f on $\widetilde{X^{t,y,s;\varepsilon}_{\mathrm{HSC}}}$, then $\bar{x}$ is a pseudostationary point of f on X_{HSC}.*

Proof. We first show that $\bar{x}$ is feasible to (16) by verifying

$$\psi^{+;\varepsilon}_{ij}(\bar{x}) \mathbf{1}_{(0,\infty)}(\phi_{ij}(\bar{x})) \le \psi^{+;\varepsilon}_{ij}(\bar{x}) \bar{s}_{ij} \text{ and } \psi^{-;\varepsilon}_{ij}(\bar{x}) \mathbf{1}_{(0,\infty)}(\phi_{ij}(\bar{x})) \ge \psi^{-;\varepsilon}_{ij}(\bar{x}) \, \bar{t}_{ij}.$$

Indeed, if $\phi_{ij}(\bar{x}) > 0$, then $\bar{s}_{ij} = 1$, and the first inequality holds; the second inequality also holds because $\bar{t}_{ij} \le 1$. If $\phi_{ij}(\bar{x}) \le 0$, then the first inequality clearly holds; moreover, we must have $\bar{t}_{ij} = 0$. It ,therefore, follows that

$$\sum_{j=1}^{J_i} \psi^{+;\varepsilon}_{ij}(x) \, \mathbf{1}_{(0,\infty)}(\phi_{ij}(\bar{x})) \le \sum_{j=1}^{J_i} \psi^{-;\varepsilon}_{ij}(x) \, \mathbf{1}_{(0,\infty)}(\phi_{ij}(\bar{x})) + b_i,$$

which is equivalent to $\sum_{j\in\mathcal{J}_{i,+}(\bar{x})}\psi_{ij}(x)\le b_i$. Thus, $\bar{x}$ is feasible to (16). It remains to show that $\bar{x}$ is a B-stationary point of (16). For this purpose, let $\{x^k\}$ be a sequence converging to $\bar{x}$ and $\{\tau_k\}\downarrow 0$ such that each x^k is feasible to (16) and $\lim_{k\to\infty}\frac{x^k-\bar{x}}{\tau_k}=v$. We need to show that $f'(\bar{x};v)\ge 0$. It turn, it suffices to show the existence of a corresponding sequence $\{(t^k,y^k,s^k)\}$ converging to $(\bar{t},\bar{y},\bar{s})$ such that (x^k,t^k,y^k,s^k) belongs to $\widetilde{X^{t,y,s;\varepsilon}_{\mathrm{HSC}}}$ for all k sufficiently large, and the three sequences

$$\left\{\frac{t^k-\bar{t}}{\tau_k}\right\};\left\{\frac{y^k-\bar{y}}{\tau_k}\right\};\text{ and }\left\{\frac{s^k-\bar{s}}{\tau_k}\right\} \tag{17}$$

are bounded. Without loss of generality, we may assume that for all (i,j,k), $\phi_{ij}(x^k)$ has the same sign as $\phi_{ij}(\bar{x})$ if the latter is nonzero. Furthermore, because x^k is feasible to (16), we must have that $\phi_{ij}(x^k)>0$ implies $\phi_{ij}(\bar{x})>0$. Hence,

$$\mathbf{1}_{(0,\infty)}(\phi_{ij}(\bar{x}))=\mathbf{1}_{(0,\infty)}(\phi_{ij}(x^k)),\ \forall k. \tag{18}$$

Because the constraints in (16) are separable in i, for notational simplicity, we drop the index i in the rest of the proof. Let

$$\Delta(\bullet;\bar{x})\triangleq\sum_{j=1}^{J}\psi_j(\bullet)\,\mathbf{1}_{(0,\infty)}(\phi_j(\bar{x}))-b, \tag{19}$$

which is a B-differentiable function. Note that $\Delta(\bar{x};\bar{x})\le 0$. Let

$$S\triangleq\{j\,|\,\phi_j(\bar{x})\le 0<\bar{s}_j\}\ \text{ and }\ T\triangleq\{j\,|\,\phi_j(\bar{x})>0>\bar{t}_j-1\}.$$

We have

$$\begin{aligned}
&\sum_{j\in S}(\psi_j^+(\bar{x})+\varepsilon)\bar{s}_j+\sum_{j\in T}(\psi_j^-(\bar{x})+\varepsilon)(1-\bar{t}_j)\\
&=\sum_{j=1}^{J}(\psi_j^+(\bar{x})+\varepsilon)\bar{s}_j-\sum_{j=1}^{J}(\psi_j^-(\bar{x})+\varepsilon)\bar{t}_j-\sum_{j:\phi_j(\bar{x})>0}[(\psi_j^+(\bar{x})+\varepsilon)\bar{s}_j-(\psi_j^-(\bar{x})+\varepsilon)]\\
&\le b-\sum_{j:\phi_j(\bar{x})>0}[(\psi_j^+(\bar{x})+\varepsilon)-(\psi_j^-(\bar{x})+\varepsilon)]=b-\sum_{j:\phi_j(\bar{x})>0}\psi_j(\bar{x})=-\Delta(\bar{x};\bar{x}),
\end{aligned} \tag{20}$$

where the last inequality holds because $(\bar{x},\bar{t},\bar{s},\bar{y})\in\widetilde{X^{t,y,s;\varepsilon}_{\mathrm{HSC}}}$ and $\bar{s}_j=1$ for j such that $\phi_j(\bar{x})>0$. Hence,

$$\Delta(\bar{x};\bar{x})+\sum_{j\in S}(\psi_j^+(\bar{x})+\varepsilon)\bar{s}_j+\sum_{j\in T}(\psi_j^-(\bar{x})+\varepsilon)(1-\bar{t}_j)\le 0.$$

Case 1. Suppose that $S\cup T\ne\emptyset$. Because $\varepsilon>0$, the above inequality implies that

$$\Delta(\bar{x};\bar{x})+\sum_{j\in S}(\psi_j^+(\bar{x})+\varepsilon)\bar{s}_j+\sum_{j\in T}(\psi_j^-(\bar{x})+\varepsilon)(1-\bar{t}_j)<0. \tag{21}$$

We can write

$$\Delta(\bar{x};\bar{x})=-\left\{\sum_{j\in S}\left[\underbrace{(\psi_j^+(\bar{x})+\varepsilon)\bar{s}_j+\delta_j^s}_{\text{denoted }\Delta_j^s\ge 0}\right]+\sum_{j\in T}\left[\underbrace{(\psi_j^-(\bar{x})+\varepsilon)(1-t_j)+\delta_j^t}_{\text{denoted }\Delta_j^t\ge 0}\right]\right\}$$

for some nonnegative scalars δ_j^s and δ_j^t. Define the nonnegative scalars:

$$\Delta_j^s(x^k)\triangleq\frac{\Delta(x^k;\bar{x})}{\Delta(\bar{x};\bar{x})}\,\Delta_j^s\ \text{ and }\ \Delta_j^s(x^k)\triangleq\frac{\Delta(x^k;\bar{x})}{\Delta(\bar{x};\bar{x})}\,\Delta_j^t.$$

Because $\Delta(\bullet;\bar{x})$ is continuous, it follows that $\lim_{k\to\infty}\Delta_j^s(x^k)=\Delta_j^s$ and $\lim_{k\to\infty}\Delta_j^t(x^k)=\Delta_j^t$. Next, we construct a sequence (t^k,y^k,s^k) such that $(x^k,t^k,y^k,s^k)\in\widetilde{X_{\rm HSC}^{t,y,s;\varepsilon}}$. Let

$$s_j^k \triangleq \begin{cases} \min\left\{\dfrac{\Delta_j^s(x^k)}{\psi_j^+(x^k)+\varepsilon},\,\bar{s}_j\right\} & \text{if } j\in S \\ \bar{s}_j & \text{otherwise;} \end{cases}$$

$$1-t_j^k \triangleq \begin{cases} \min\left\{\dfrac{\Delta_j^t(x^k)}{\psi_j^-(x^k)+\varepsilon},\,1-\bar{t}_j\right\} & \text{if } j\in T \\ 1-\bar{t}_j & \text{otherwise;} \end{cases}$$

$$y_j^k \triangleq \begin{cases} \max\left\{\dfrac{t_j^k}{\phi_j(x^k)},\,\bar{y}_j\right\} & \text{if } \phi_j(x^k)>0\ (\Longleftrightarrow \phi_j(\bar{x})>0) \\ \bar{y}_j & \text{if } \phi_j(x^k)\le 0. \end{cases} \tag{22}$$

We need to verify the functional inequalities in $\widetilde{X_{\rm HSC}^{t,y,s;\varepsilon}}$. These are done in the following three steps.

Step 1. By a derivation similar to (20), we can verify the first equality in the following string of derivations:

$$\begin{aligned}
&\sum_{j=1}^{J}(\psi_j^+(x^k)+\varepsilon)\,s_j^k-\sum_{j=1}^{J_i}(\psi_j^-(x^k)+\varepsilon)\,t_j^k\\
&=\sum_{j\in S}(\psi_j^+(x^k)+\varepsilon)\,s_j^k+\sum_{j\notin S}(\psi_j^+(x^k)+\varepsilon)\,\bar{s}_j-\sum_{j\in T}(\psi_j^-(x^k)+\varepsilon)\,t_j^k-\sum_{j\notin T}(\psi_j^-(x^k)+\varepsilon)\,\bar{t}_j\\
&=\sum_{j\in S}(\psi_j^+(x^k)+\varepsilon)\,s_j^k+\sum_{j\in T}(\psi_j^-(x^k)+\varepsilon)\,(1-t_j^k)+\sum_{j:\phi_j(\bar{x})>0}(\psi_j^+(x^k)+\varepsilon)\\
&\quad-\sum_{j:\phi_j(\bar{x})>0}(\psi_j^-(x^k)+\varepsilon)\ \text{by properties of } \bar{s}_j(\bar{t}_j) \text{ for } j\notin S (j\notin T)\\
&=\sum_{j\in S}(\psi_j^+(x^k)+\varepsilon)\,s_j^k+\sum_{j\in T}(\psi_j^-(x^k)+\varepsilon)\,(1-t_j^k)+\sum_{j=1}^{J}\psi_j(x^k)\mathbf{1}_{(0,\infty)}(\phi_j(\bar{x}))\\
&\le\sum_{j\in S}\Delta_k^s(x^k)+\sum_{j\in T}\Delta_k^t(x^k)+\Delta(x^k;\bar{x})+b,\ \text{by the definitions of the } \Delta\text{'s; see (19) and (22)}\\
&\le b,\qquad \text{by (21) and the continuity of } \sum_{j\in S}\Delta_k^s(\bullet)+\sum_{j\in T}\Delta_k^t(\bullet)+\Delta(\bullet;\bar{x}).
\end{aligned}$$

Step 2. If $\phi_j(x^k)>0$, we clearly have $\phi_j(x^k)y_j^k\ge t_j^k$ by the definition of y_j^k. If $\phi_j(x^k)\le 0$, then $\phi_j(\bar{x})\le 0$; thus, $\bar{t}_j=0$. Moreover, $j\notin T$; hence, $t_j^k=\bar{t}_j=0$, and $t_j^k\le\phi_j^+(x^k)y_j^k$ holds.

Step 3. If $\phi_j(x^k)>0$, then $\phi_j(\bar{x})>0$ by (18), and $j\notin S$; hence, $s_j^k=\bar{s}_j=1$. It follows that $\phi_j^+(x^k)(1-s_j^k)=0$. The latter clearly holds if $\phi_j(x^k)\le 0$.

We have, therefore, shown that $(x^k,t^k,y^k,s^k)\in\widetilde{X_{\rm HSC}^{t,y,s;\varepsilon}}$ for all k. Next, for $j\in S$, we have

$$\lim_{k\to\infty}s_j^k=\min\left\{\lim_{k\to\infty}\frac{\Delta_j^s(x^k)}{\psi_j^+(x^k)+\varepsilon},\,\bar{s}_j\right\}=\min\left\{\frac{\Delta_j^s}{\psi_j^+(\bar{x})+\varepsilon},\,\bar{s}_j\right\}=\bar{s}_j,$$

where the last equality holds because $\Delta_j^s\ge\psi_j^+(\bar{x})+\varepsilon$ by the definition of Δ_j^s. Hence, $\lim_{k\to\infty}s_j^k=\bar{s}_j$ for all j. Similarly, we can show that $\lim_{k\to\infty}t_j^k=\bar{t}_j$ and $\lim_{k\to\infty}y_j^k=\bar{y}_j$ for all j for all j. Because s_j^k and t_j^k are either constants ($\bar{s}_j$ or $\bar{t}_j$, respectively) or the pointwise minima of a B-differentiable fraction of x^k and a constant ($\bar{s}_j$ or $\bar{t}_j$), they are, therefore, B-differentiable functions of x^k, and hence, so is y_j^k. Therefore, the fractions $\frac{t_j^k-\bar{t}_j}{\tau_k}$, $\frac{s_j^k-\bar{s}_j}{\tau_k}$, and $\frac{y_j^k-\bar{y}_j}{\tau_k}$ are bounded.

Case 2. If $S\cup T=\emptyset$, then define $s^k=\bar{s}$ and $t^k=\bar{t}$ for all k and y_j^k as above. A similar proof applies. □

7.3. Recovering Bouligand Stationarity

It turns out that by requiring the tuple $(\bar{x},\bar{t},\bar{y},\bar{s}) \in \widetilde{X_{\text{HSC}}^{t,y,s;\varepsilon}}$ to be a B-stationary point of an enlargement of the lifted set $\widetilde{X_{\text{HSC}}^{t,y,s;\varepsilon}}$, it is possible to sharpen the conclusion of Proposition 15 to the stronger property of B-stationarity of f on X_{HSC}. Specifically, consider the set with an additional scalar $\eta > 0$:

$$\widetilde{X_{\text{HSC};\eta}^{t,y,s;\varepsilon}} \triangleq \left\{ \begin{array}{l|l} x \in P & \displaystyle\sum_{j=1}^{J_i} \psi_{ij}^{+;\varepsilon}(x)\, s_{ij} \le \sum_{j=1}^{J_i} \psi_{ij}^{-;\varepsilon}(x)\, t_{ij} + b_i, \quad i = 1,\ldots,m \\ t_{ij},\, s_{ij} \in [0,1], \text{ all } i,j & \\ y_{ij} \ge 0, \text{ all } i,j & \text{and for all } j = 1,\ldots,J_i \text{ and } i = 1,\ldots,m: \\ & t_{ij} \le \phi_{ij}^{+}(x)\, y_{ij}, \quad \phi_{ij}^{+}(x)\,(1 - s_{ij}) \le \eta \end{array} \right\}.$$

We have the following result.

Proposition 16. *Let the functions f, ϕ_{ij}, and ψ_{ij} be B-differentiable near $\bar{x} \in P$. For an arbitrary pair $(\varepsilon,\eta) > 0$, if $(\bar{x},\bar{t},\bar{y},\bar{s}) \in \widetilde{X_{\text{HSC}}^{t,y,s;\varepsilon}}$ is a B-stationary point of f on $\widetilde{X_{\text{HSC};\eta}^{t,y,s;\varepsilon}}$, then $\bar{x}$ is a B-stationary point of f on X_{HSC}.*

Proof. We proceed as in the proof of Proposition 15. Let $\{x^k\}$ be a sequence in X_{HSC} converging to $\bar{x}$ and $\{\tau_k\} \downarrow 0$ such that $\lim_{k\to\infty} \frac{x^k - \bar{x}}{\tau_k} = v$. We need to show that $f'(\bar{x};v) \ge 0$. It suffices to show the existence of a corresponding sequence $\{(t^k,y^k,s^k)\}$ converging to $(\bar{t},\bar{y},\bar{s})$ such that (x^k,t^k,y^k,s^k) belongs to $\widetilde{X_{\text{HSC};\eta}^{t,y,s;\varepsilon}}$ for all k sufficiently large and the three sequences

$$\left\{\frac{t^k - \bar{t}}{\tau_k}\right\};\left\{\frac{y^k - \bar{y}}{\tau_k}\right\}; \text{ and } \left\{\frac{s^k - \bar{s}}{\tau_k}\right\} \tag{23}$$

are bounded. As before, we may assume that for all (i,j,k), $\phi_{ij}(x^k)$ has the same sign as $\phi_{ij}(\bar{x})$ if the latter is nonzero. Furthermore, (18) is valid for all k except for a k such that $\phi_{ij}(\bar{x}) = 0 < \phi_{ij}(x^k)$. Defining (s_j^k, t_j^k, y_j^k) by (22), we see that the proof of steps 1 and 2 in Proposition 15 is valid as (18) is not used until the last step 3, which we analyze below.

Step 3. If $\phi_j(x^k) > 0$, then $\phi_j(\bar{x}) \ge 0$. If $\phi_j(\bar{x}) > 0$, then $j \notin S$ and $s_j^k = \bar{s}_j = 1$. If $\phi_j(\bar{x}) = 0$, then $\phi_j^{+}(x^k)\,(1 - s_j^k) \le \eta$ for k sufficiently large.

Summarizing the three steps, we have established $(x^k,t^k,y^k,s^k) \in \widetilde{X_{\text{HSC};\eta}^{t,y,s;\varepsilon}}$ for all k sufficiently large. The proof of the convergence of $\{(t^k,y^k,s^k)\}$ to $(\bar{t},\bar{y},\bar{s})$ and that of the boundedness of the sequences in (23) are the same as before. □

Acknowledgments

The authors are grateful for the referees' constructive comments, which improved the quality of the paper.

References

[1] Atamtürk A, Gómez A, Han S (2021) Sparse and smooth signal estimation: Convexification of ℓ_0-formulations. *J. Machine Learn. Res.* 22(52):1–43.
[2] Bian W, Chen X (2020) A smoothing proximal gradient algorithm for nonsmooth convex regression with cardinality penalty. *SIAM J. Numer. Anal.* 58(1):858–883.
[3] Chen X, Pan L, Xiu N (2022) Solution sets of three sparse optimization problems for multivariate regression. *J. Global Optim.* 87:347–371.
[4] Clarke FH (1990) *Optimization and Nonsmooth Analysis*, vol. 5 (SIAM Publications, Philadelphia).
[5] Conforti M, Cornuéjols G, Zambelli G (2014) *Integer Programming*, vol. 271 (Springer, Cham, Switzerland).
[6] Cui Y, Pang J-S (2021) *Modern Nonconvex Nondifferentiable Optimization* (SIAM Publications, Philadelphia).
[7] Cui Y, Liu J, Pang J-S (2023) The minimization of piecewise functions: Pseudo stationarity. *J. Convex Anal.* 30(3):793–834.
[8] Cui Y, Liu J, Pang J-S (2022) Nonconvex and nonsmooth approaches for affine chance-constrained stochastic programs. *Set-Valued Variational Anal.* 30(3):1149–1211.
[9] Cui Y, Chang T-H, Hong M, Pang J-S (2020) A study of piecewise linear-quadratic programs. *J. Optim. Theory Appl.* 186:523–553.
[10] Dong H, Linderoth J (2013) On valid inequalities for quadratic programming with continuous variables and binary indicators. Goemans M, Correa J, eds. *Integer Programming Combinatorial Optim. 16th Internat. Conf. IPCO 2013*, vol. 16 (Springer, Cham, Switzerland), 169–180.
[11] Dong H, Ahn M, Pang J-S (2019) Structural properties of affine sparsity constraints. *Math. Programming* 176:95–135.
[12] Fang Y (2023) Essays on treatment effect and policy learning under different settings. PhD thesis, Department of Economics, University of Southern California, Los Angeles.
[13] Ferris MC, Mangasarian OL (1992) Minimum principle sufficiency. *Math. Programming* 57(1–3):1–14.
[14] Gómez A, He Z, Pang J-S (2023) Linear-step solvability of some folded concave and singly-parametric sparse optimization problems. *Math. Programming* 198(2):1339–1380.
[15] Hastie T, Tibshirani R, Wainwright M (2015) *Statistical Learning with Sparsity: The Lasso and Generalizations* (CRC Press, Boca Raton, FL).

[16] He Z, Han S, Gómez A, Cui Y, Pang J-S (2024) Comparing solution paths of sparse quadratic minimization with a Stieltjes matrix. *Math. Programming* 204:517–566.
[17] Kanzow C, Raharja AB, Schwartz A (2021) Sequential optimality conditions for cardinality-constrained optimization problems with applications. *Comput. Optim. Appl.* 80:185–211.
[18] Kanzow C, Schwarz A, Weiß F (2022) The sparse (st) optimization problem: Reformulations, optimality, stationarity, and numerical results. Preprint, submitted October 18, https://arxiv.org/abs/2210.09589.
[19] Mangasarian OL (1975) Pseudo-convex functions. Ziemba WT, Vickson RG, eds. *Stochastic Optimization Models in Finance* (Elsevier, Amsterdam), 23–32.
[20] Mangasarian OL (1994) *Nonlinear Programming* (SIAM Publications, Philadelphia).
[21] Pang J-S, Razaviyayn M, Alvarado A (2017) Computing b-stationary points of nonsmooth dc programs. *Math. Oper. Res.* 42(1):95–118.
[22] Qi Z, Cui Y, Liu Y, Pang J-S (2019) Estimation of individualized decision rules based on an optimized covariate-dependent equivalent of random outcomes. *SIAM J. Optim.* 29(3):2337–2362.
[23] Rockafellar RT, Wets RJ-B (2009) *Variational Analysis*, vol. 317 (Springer Science & Business Media, New York).
[24] Royset JO (2023) Consistent approximations in composite optimization. *Math. Programming* 201(1–2):339–372.
[25] Royset JO, Wets RJ-B (2022) *An Optimization Primer* (Springer, Cham, Switzerland).