Analysis of a Class of Minimization Problems Lacking Lower Semicontinuity

Shaoning Han
Daniel J. Epstein Department of Industrial and Systems Engineering, University of Southern California, Los Angeles 90089
shaoning@usc.edu

Ying Cui
Department of Industrial Engineering and Operations Research, University of California at Berkeley, 94720
yingcui@berkeley.edu

Jong-Shi Pang
Daniel J. Epstein Department of Industrial and Systems Engineering, University of Southern California, Los Angeles 90089
jongship@usc.edu

The minimization of non-lower 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 pseudo stationarity 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 pseudo stationarity is implied by, thus weaker than, a sharper subdifferential based stationarity condition which we term epi-stationarity; (b) introducing a set-theoretic sufficient condition, which we term local convexity-like property, under which an epi-stationary point of a possibly non-lower 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 epi-stationary solution of the given Heaviside composite program with broad classes of B-differentiable component functions can in principle be approximately computed by surrogation methods.

Key words: Heaviside functions; lower semicontinuity; nonsmooth analysis; 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 composite Heaviside problem:

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

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 functions (which are usually continuous). While 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 non-negative. We refer to the reference [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 [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 [12, 22]. As a unification of these special cases, the class of additive Heaviside composite optimization problems along with the concept of *pseudo stationarity* were introduced in [7]. The latter concept has its origin in [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 [15], a special case of the composite Heaviside optimization problem is the problem with affine sparsity constraints (ASCs) that was introduced in [10] as a computational framework for rigorously solving estimation problems with structured sparsity (i.e., logical sparsity conditions). Such constraints define the following set:

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

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:

$$|\, x_3 \,|_0 \leq |\, x_1 \,|_0 + |\, x_2 \,|_0.$$

An optimization problem over ASCs is a generalization of cardinality constrained problems, whose continuous relaxations have been extensively studied in the existing literature [2, 17, 3, 18]. It is known from [10] 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 [10, 7]. 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 non-lower semicontinuous functions within objectives and constraints that lead in particular to non-closed 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 epi-limit [23, Definition 7.1] (see also [25, 24]) of a function sequence is always lower semicontinuous, it is thus not possible to construct approximating functions that exhibit epi-convergence to the original non-lower semicontinuous functions. This absence of epi-convergence 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 epi-limit 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 [10, Example 1])

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

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: $x_1 \neq 0 \Longrightarrow x_2 \neq 0$. 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 $\bigcap_{i=1}^{m} C_m$ when $m > 1$ and at least one C_i is non-closed. This closure can be significantly smaller than $\bigcap_{i=1}^{m}\text{closure}\{C_i\}$.

Since there is a simple linear structure in the ASC constraint set X_{ASC} and the only combinatorial aspect of the set is due to the ℓ_0-function that has a well-known integer description, a natural question is whether the non-closedness of X_{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 of X_{ASC}, obtaining in particular a representation of the closure of X_{ASC} that extends [10, Proposition 4], from which we can deduce a "big-M" integer description of the closure. Deepening the analysis in this reference where the side set P in X_{ASC} is the entire space, i.e., for an unconstrained ASC system, the extended analysis elucidates the general difficulty associated with a mixed-signed combination of multiple ℓ_0-functions in that while an integral formulation of the closure of the set X_{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.

For a general optimization problem, stationary conditions are necessary for local optimality (often requiring constraint qualifications). 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 [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 [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, while 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 due to the potential failure of the chain rule. In contrast, by exposing the constraints as given, tangents to the constraint set, even if the latter is not closed, can often be more easily described and lead to constructive approaches to compute sharp stationary solutions. This approach of treating the constraints as they appear is inspired by the motivation of practical computation that dictates its focus and sets the framework for subsequent algorithmic developments elsewhere.

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 pseudo-stationarity [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 [9, 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 pseudo-stationarity is a weaker notion than the subdifferential based stationarity, which we term "epi-stationarity" for reasons to be made clear later and will formally define in Section 4. A follow-up question is whether there are classes of problems whose epi-stationary points are local minimizers. This question has its origin in differentiable problems (extendable to B-differentiable problems) for which the class of pseudo-convex functions introduced by Mangasarian [19, 20] provides an answer. Specifically, for a convex-constrained optimization problem with a differentiable pseudo-convex 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 [6] and subsequently expanded in [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 quasi-convex 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 epi-stationary point of a non-lower 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 [9, Chapter 7] when the functions in the lifted program are "surrogatable" (difference-of-convex, e.g.). 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 the problem (1), we organize the rest of this paper along with the main contributions as follows.

(A) In Section 3, we first extend [10, Proposition 4] that addresses an unconstrained ASC system (i.e., the set X_{ASC} with the polyhedron P being the entire space) to the constrained case (i.e., when P is a proper polyhedral subset of $\mathbb{R}^n$). Specifically, Proposition 1 provides an algebraic description of the closure of X_{ASC}. Based on a classical result, we derive a necessary and sufficient condition for X_{ASC} to have a mixed integer linear representation, enhanced by a more detailed description of the representation by exploiting the structure of X_{ASC}; see Proposition 2. It follows from the latter result that a closed ASC system has a conceptual "big M" mixed-integer representation (Corollary 1) whose practical utilisation remains to be further investigated.

(B) We formally define epi-stationarity for an optimization problem lacking lower semicontinuity in Section 4 (see Definition 1) and establish several important properties of an epi-stationary solution. First, epi-stationarity is a necessary condition for local minimization (Proposition 3); epi-stationarity has an equivalent description in terms of a suitable subderivative (Proposition 4); for a B-differentiable problem, epi-stationariy recovers B-stationarity (Proposition 5); finally, for the HSC-constrained optimization problem (1), epi-stationarity is sharper than the pulled-out pseudo-stationarity (Proposition 6).

(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 epi-stationary point (Proposition 10). Being the local version of the classical result of pseudo convexity implying global optimality for a differentiable problem, our result is for a non-lower semicontinuous program with a possibly nonconvex feasible set. The terminology of epi-stationarity sufficiency is borrowed from "minimum principle sufficiency" [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 the algebraic descriptions of tangent vectors of various cases of an HSC set, we summarize in Theorem 2 when such a set has the local convexity-like property, thereby obtaining the equivalence of epi-stationarity 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 the problem (1) and show that the B-stationary points of this lifted formulation, where all functions in the lifted space are B-differentiable, yield pseudo stationary points of (1), through projecting the B-stationarity points from the lifted domain onto the original space; see Proposition 16. Bouligand stationarity can be obtained from the lifting under a further assumption; see Proposition 17. 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 to 1. 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 $\mathrm{dist}(x,S) \triangleq \inf\{\|x-y\|_\infty : y \in S\}$. 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:

$$\underset{x \in X}{\textbf{minimize}} \; f(x), \tag{3}$$

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):

$$\underset{x \in \mathbb{R}^n}{\textbf{minimize}} \; f_X(x) \triangleq f(x) + \delta_X(x)$$

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 [23, Theorem 10.1] that if $\bar{x} \in X$ is a local minimizer of (3), then $0 \in \widehat{\partial} f_X(\bar{x})$, where

$$\begin{aligned}
\widehat{\partial} f_X(\bar{x}) &\triangleq \left\{ v \in \mathbb{R}^n \mid \liminf_{x(\neq \bar{x}) \to \bar{x}} \frac{f_X(x) - f_X(\bar{x}) - v^\top (x - \bar{x})}{\|\, x - \bar{x} \,\|} \geq 0 \right\} \qquad \text{[23, Definition 8.3]} \\
&= \left\{ v \in \mathbb{R}^n \mid v^\top w \leq \mathrm{d}f_X(\bar{x})(w) \;\; \text{for all} \;\; w \in \mathbb{R}^n \right\} \qquad \text{[23, Exercise 8.4]}
\end{aligned}$$

$$\begin{aligned}
\text{with} \;\; \mathrm{d}f_X(\bar{x})(v) &\triangleq \liminf_{\substack{w \to v; \; \tau \downarrow 0 \\ \bar{x} + \tau w \in X}} \frac{f(\bar{x} + \tau w) - f(\bar{x})}{\tau} \qquad \text{[23, Definition 8.1]} \\
&= \liminf_{\substack{\tau^{-1}(x' - \bar{x}) \to 0; \; \tau \downarrow 0 \\ x' + \tau v \in X}} \frac{f(x' + \tau v) - f(\bar{x})}{\tau} \quad \text{under the equality: } \bar{x} + \tau w = x' + \tau v.
\end{aligned}$$

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

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

It is easy to see that $\mathrm{d}f_X(\bar{x})(v) = -\infty$ if $v \notin \mathcal{T}(\bar{x};X)$, i.e., the domain of the subderivative $\mathrm{d}f_X(\bar{x})(\bullet)$ is a subset of the tangent cone of of X at $\bar{x}$. According to the cited reference, $\widehat{\partial} f_X(\bar{x})$ 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 $\bar{x}$. While the subdifferential $\widehat{\partial} f_X(\bar{x})$ 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 there properties.

When f is a B-differentiable function [9, 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) [9, Definition 6.1.1] if

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

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 [9, 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 $0 \in \widehat{\partial} f_X(\bar{x})$; 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 $0 \in \widehat{\partial} f(\bar{x}) + \mathcal{N}(\bar{x};X)$, 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 [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}. In this section, we discuss sufficient and necessary conditions for the ASC constraint set X_{ASC} to be *mixed-integer linear representable* (MILR). Needless to say, the challenge in dealing with this set 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., [1, 11]). 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):

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

where $M > 0$ is chosen to be sufficiently large to ensure $X_{\mathrm{ASC}} \subseteq \{x \in \mathbb{R}^n \mid \|x\|_\infty \leq M\}$. 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. A set S is termed MILR if there exist rational matrices A, B and C, and a rational vector d, all of appropriate dimensions, such that

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

The following classical result [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 the 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 $P_1, \cdots, P_k \subseteq \mathbb{R}^n$ and vectors $r^1, \cdots, r^m \in \mathbb{Z}^n$ such that

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

where $\text{intcone}\{\, r^1, \cdots, r^m \,\} \triangleq \left\{ \sum_{i=1}^{m} \lambda_i r^i \,:\, \lambda \in \mathbb{Z}^m_+ \right\}$. $\square$

Note that a MILR set must be closed. Indeed, a set S is closed if and only if $S \cap \{x : \|x\|_2 \leq \tau\}$ is closed for any scalar $\tau > 0$. If S is MILR, then by Theorem 1, the set $S \cap \{x : \|x\|_2 \leq \tau\}$ is a finite union of compact sets and is thus closed. This implies that X_{ASC} is not MILR in general. In [10, Section 3], the issue of closedness of X_{ASC} with $P = \mathbb{R}^n$ and the identification of its closure have been studied; the result below generalizes this previous study to the case where the polyhedron P is a proper subset of $\mathbb{R}^n$.

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

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 $\widehat{S} \triangleq \bigcup_{z \in S} \{y \in \{0,1\}^n \mid y \leq z\}$ be the downward closure generated by S. Since $\widehat{S} \subseteq \{0,1\}^n$, one has $\widehat{S} = \text{conv}(\widehat{S}) \cap \{0,1\}^n$. We claim that $\text{conv}(\widehat{S}) = \{\, z \geq 0 \mid \widetilde{A} z \leq \widetilde{b} \,\}$ for some matrix $\widetilde{A} \geq 0$ and $\{0,1\}$-vector $\widetilde{b}$. For this purpose, we first show that $y \in \text{conv}(\widehat{S})$ if and only if $y \geq 0$ and $y^\top u \leq \max_{z \in \widehat{S}} u^\top z$ for all $u \geq 0$. The "only if" assertion is obvious. For the "if" assertion, suppose that $0 \leq y \notin \text{conv}(\widehat{S})$ is such that $y^\top u \leq \max_{z \in \widehat{S}} u^\top z$ for all $u \geq 0$. Since $\text{conv}(\widehat{S})$ is a polytope, by separation, there exist a vector $\widetilde{u}$ and a scalar γ such that $y^\top \widetilde{u} > \gamma \geq \max_{z \in \text{conv}(\widehat{S})} \widetilde{u}^\top z$. For any vector $z \in \widehat{S}$, the vector $\widetilde{z}$ obtained by zeroing out the components of z corresponding to the negative components of $\widetilde{u}$ remains an element of $\widehat{S}$. Thus, with $\widetilde{u}^+$ denoting the nonnegative part of the vector $\widetilde{u}$, we have

$$y^\top \widetilde{u}^+ \geq y^\top \widetilde{u} > \widetilde{z}^\top \widetilde{u} = z^\top \widetilde{u}^+,$$

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

$$y^\top u \leq \alpha \text{ for all } u \geq 0, \alpha \geq \max_{z \in \widehat{S}} u^\top z \iff \begin{cases} y^\top u \leq 1 & \text{for all } u \geq 0 \text{ such that } \max_{z \in \widehat{S}} u^\top z \leq 1 \\ y^\top u \leq 0 & \text{for all } u \geq 0 \text{ such that } \max_{z \in \widehat{S}} u^\top z \leq 0 \end{cases}$$

$$\iff \begin{cases} y^\top u \leq 1 & \forall u \in P_1 \triangleq \{\, u \mid u \geq 0, u^\top z \leq 1, \forall z \in \widehat{S} \,\} \\ y^\top u \leq 0 & \forall u \in P_0 \triangleq \{\, u \mid u \geq 0, u^\top z \leq 0, \forall z \in \widehat{S} \,\}. \end{cases}$$

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

$$\{\, y \mid y \geq\, 0,\, (u^{1j})^\top y \leq 1,\, (u^{2\ell})^\top y \leq 0,\ \forall j = 1, \cdots, k_1,\, \ell = 1, \cdots, k_2 \},$$

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

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

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 $\widetilde{A}$ in Proposition 1 is known. Take any point $\bar{x} \in \operatorname{cl}(X_{\text{ASC}})$ and let $\bar{z} = |\,\bar{x}\,|_0 \in \{0,1\}^n$. Given $\widetilde{A} \geq 0$, it is easy to identify a maximal element $\widehat{z} \in \{0,1\}^n$ in the support set such that $\bar{z} \leq \widehat{z}$. Following this, one can determine a point $\widehat{x} \in \operatorname{cl}(X_{\text{ASC}})$ such that $|\,\widehat{x}\,|_0 = \widehat{z}$ by solving linear programs over $\{x \in P \mid x_i(1-\widehat{z}_i) = 0,\, i = 1, \cdots, n\}$. Consequently, we have $\widehat{x} \in X_{\text{ASC}}$, which implies that $\varepsilon\widehat{x} + (1-\varepsilon)\bar{x} \in X_{\text{ASC}}$ for almost all $\varepsilon \in (0,1]$.

However, it is worth noting that while the existence of $\widetilde{A}$ is guaranteed by Proposition 1, unfortunately, the effective construction of $\widetilde{A}$ remains unclear. Consequently, this proposition is primarily of conceptual significance. In the following, we show that for the set X_{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. Since X_{ASC} is closed, by Proposition 1, one can assume $A \geq 0$. Observe that there exists $t_0 > 0$, such that as $t > t_0$, $\operatorname{supp}(\bar{x} + tr) = \operatorname{supp}(\bar{x}) \cup \operatorname{supp}(r)$. 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; two, its generators are recession vectors of the base polyhedron P whose nonzero components correspond to those of a maximal element of the set $\widehat{S}$ in the proof of Proposition 1.

Proposition 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 $X_{\text{ASC}} = \bigcup_{i=1}^k P_i + R$. Furthermore, the recession cone R takes the form $\{r \in P_\infty \mid r_i = 0,\ \forall\, i \notin \operatorname{supp}(z^{\max})\}$, where $z^{\max}$ is any maximal element of $\{z \in \{0,1\}^n \mid z = |\,x\,|_0,\, x \in X_{\text{ASC}}\}$.

Proof. Thanks to Proposition 1, we can assume that matrix, denoted as A, of the coefficients a_{ij} in the definition of X_{ASC} are 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 $X_{\text{ASC}} = \bigcup_{i=1}^k P_i + \operatorname{intcone}\{r^1, \ldots, r^m\}$. 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 $x \in X_{\text{ASC}}$ since we can always scale r by a positive integer to make μ integral. Since 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. Since P_i are polytopes, $\bigcup_{i=1}^{k} P_i$ is MILR by Theorem 1. Since a polyhedral set is always MILR and the Minkowski sum of two MILR sets is MILR, we can deduce that $X_{\text{ASC}} = \bigcup_{i=1}^{k} P_i + R$ is MILR.

It remains to prove the representation of the cone R. Let $\widetilde{R} \triangleq \{r \in P_\infty \mid r_i = 0,\ \forall\, i \notin \text{supp}(z^{\max})\}$. 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 $r \in \widetilde{R}$, $\text{supp}(r) \subseteq \text{supp}(\bar{x})$. 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, $\widetilde{R} \subseteq R$. If there exists $r \in R \setminus \widetilde{R}$, 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 $\widetilde{R} = R$. $\square$

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

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

Proof. Take M large enough such that in the statement of Proposition 2, it holds that $\widehat{P} \triangleq \bigcup_{i=1}^{k} P_i$ is contained in $\{x : \|x\|_\infty \leq M\}$. If $x \in X_{\text{ASC}}$, then there exists y and r such that $y \in \widehat{P} \subseteq X_{\text{ASC}}$ 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 $y \in X_{\text{ASC}}$. Since $X_{\text{ASC}} + R \subseteq X_{\text{ASC}}$, the conclusion follows. $\square$

4. Epi-stationarity. It is trivial to cast the 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 $\text{epi}(f)$ is not closed. Nevertheless, we can formally introduce the following concept.

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

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 epi-stationary 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 epi-stationarity 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 3. If $\bar{x}$ is a local minimizer of (3), then $\bar{x}$ is an epi-stationary point of (3).

Proof. We need to verify the implication (6). With the conditions beneath the limit, we have $t_k \geq f(x^k) \geq f(\bar{x})$. Thus it follows readily that $dt \geq 0$ as desired. □

For the purpose to connect epi-stationarity 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 ∞.

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 $\left\{ w^k \triangleq \dfrac{x^k - \bar{x}}{\|x^k - \bar{x}\|} \right\}$ 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 epi-stationarity with the nonnegativity of the subderivative $\mathrm{d}f_X(\bar{x})$ on $\mathcal{T}(\bar{x}; X)$, provided that the latter subderivative is finite valued, and with regular subdifferential based stationarity.

PROPOSITION 4. 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 statements hold:

(a) If $\mathrm{d}f_X(\bar{x})(v)\geq 0$ for all $v\in\mathcal{T}(\bar{x};X)$, then $\bar{x}$ is an epi-stationary point of f on X.

(b) Conversely, if $\mathrm{d}f_X(\bar{x})$ is finite valued (e.g., if f is locally Lipschitz continuous near $\bar{x}$) and if $\bar{x}$ is an epi-stationary point of f on X, then $\mathrm{d}f_X(\bar{x})(v)\geq 0$ for all $v\in\mathcal{T}(\bar{x};X)$.

(c) If $0\in\widehat{\partial}f_X(\bar{x})$, then $\bar{x}$ is an epi-stationary point of f on X.

(d) Conversely, if $\mathrm{d}f_X(\bar{x})$ is finite valued on $\mathcal{T}(\bar{x};X)$ and $\bar{x}$ is an epi-stationary point of f on X, then $0\in\widehat{\partial}f_X(\bar{x})$.

Proof. To prove (a), 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. Conversely, suppose $\bar{x}\in X$ is an epi-stationary point of f on X. Let $v\in\mathcal{T}(\bar{x};X)$ be arbitrary. Let $\{w^k\}\to v$ and $\{\tau_k\}\downarrow 0$ be sequences 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 $t_k\triangleq f(x^k)$. Provided that $\mathrm{d}f_X(\bar{x})(v)$ is finite, 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.

Statement (c) follows readily by combining the definition of $\widehat{\partial}f_X(\bar{x})$, the equalities (7), and part (a). Finally, statement (d) follows similarly by invoking part (b) instead of (a). □

REMARK 1. While the proof of Proposition 4 is closely related to [23, Proposition 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 $\mathrm{d}\bar{g}_X(\bar{x})$ of $\bar{g}_X\triangleq g+\delta_X$, the main point of the proposition is on the restatement of epi-stationarity in terms of subderivatives. □

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

PROPOSITION 5. 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 epi-stationary.

Proof. "Only if". The proof is the same as that of part (a) of Proposition 4. Indeed, with the previous set-up, we have

$$dt \geq \lim_{k\to\infty}\frac{f(x^k)-f(\bar{x})}{\tau_k} = f'(\bar{x};v)\geq 0,$$

where the equality holds by the B-differentiability of f at $\bar{x}$.

"If". We need to show 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 $v=\lim_{k\to\infty}\dfrac{x^k-\bar{x}}{\tau_k}$. 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 epi-stationarity 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 $X_{\rm HSC}$ is a *pseudo stationary point* of this problem if $\bar{x}$ is an epi-stationary point of the "pulled out" problem:

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

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 $i = 0,1,\cdots,m$. We also define $\mathcal{J}_{i,0}(\bar{x}) \triangleq \{\, j \mid \phi_{ij}(\bar{x}) = 0 \,\}$. We remark that in [7, Definition 3], the definition of pseudo-stationarity assumes that all the functions $\left\{ \{\psi_{ij}, \phi_{ij}\}_{j=1}^{J_i} \right\}_{i=0}^{m}$ are B-differentiable; here these functions can be arbitrary.

The pseudo-stationarity 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 $\left\{ \{\psi_{ij}, \phi_{ij}\}_{j=1}^{J_i} \right\}_{i=0}^{m}$ have favorable properties (e.g., difference-of-convexity), the resulting problem (8) is computationally tractable [21] and enables the verification of the stipulated fixed-point condition on the candidate solution $\bar{x}$. The paper [7] has provided constructive ways to approximately compute an B-stationary point of (8) under some sign conditions on the functions ψ_{ij}; see also [14, 16] for a special quadratic sparse optimization problem involving the ℓ_0-function.

In the following, we show that for the problem (1), epi-stationarity is sharper than pseudo-stationarity. Note that (8) is a restriction of the original problem (1) around $\bar{x}$. Thus, the global/local optimality of (8) are necessary conditions for the respective optimality of (1).

PROPOSITION 6. If $\bar{x}$ is an epi-stationary point of (1), then it is pseudo stationary.

Proof. Letting $\psi_{\rm HSC}^{\rm ps}$ and $X_{\rm HSC}^{\rm ps}$ denote the objective function and constraint set of (8), respectively, and $Z_{\rm HSC}^{\rm ps} \triangleq {\rm epi}(\psi_{\rm HSC}^{\rm ps}) \cap (X_{\rm HSC}^{\rm ps} \times \mathbb{R})$, and recalling the epigraphical set Z (see (5)) of the problem (1), it suffices to show that if x is sufficiently close to $\bar{x}$ and if (x,t) belongs to $Z_{\rm HSC}^{\rm ps}$, 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:

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

In particular, $f_{\rm HSC}(x) = \psi_{\rm HSC}^{\rm ps}(x)$. $\square$

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 [8, Section 4.2] (see the earlier reference [6, Proposition 4.1] for a special case of this property) if there exists a neighborhood $\mathcal{N}$ of $\bar{x}$ such that

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

Slightly generalizing the family of functions in [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., $\varphi(z) \geq \varphi(z')$ for any two L-dimensional vectors $z \geq z'$), $\Theta : \mathbb{R}^m \to \mathbb{R}^L$ is a vector-valued function such that each of its component functions $\theta_\ell : \mathbb{R}^m \to \mathbb{R}$ for $\ell = 1, \cdots, L$ 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

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

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

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

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 proved in [8, Proposition 9(ii)].

Proposition 7. Let P be a polyhedron. Suppose that f and each f_k for $k = 1, \cdots, K$ 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 minimum. We next introduce an important geometric property of an arbitrary set that allows us to establish epi-stationarity sufficiency; i.e., the question of when an epi-stationary 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 8. Let $X \triangleq \{ x \in \mathbb{R}^n \mid f_k(x) \leq 0, \ k = 1, \cdots, K \}$ 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

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

Hence if $x \in X \cap \mathcal{N}$, the above inequalities imply that $f_k'(\bar{x}; x - \bar{x}) \leq 0$ for all $k \in \mathcal{A}(\bar{x})$. Hence $x - \bar{x} \in \mathcal{T}(\bar{x}; X)$ under the ACQ. Since $x \in X \cap \mathcal{N}$ 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 9. 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 $\text{epi}(f)$ is locally convex-like at $(\bar{x}, f(\bar{x}))$.

Proof. By [23, Proposition 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 $f(x) \geq h(x)$ 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 $\text{epi}(f)$ at $(\bar{x}, f(\bar{x}))$. $\square$

The next result establishes the promised epi-stationarity 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 10. 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 epi-stationary 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 $f(x') \geq f(\bar{x})$ 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 epi-stationarity, we have $dt \geq 0$. But then $t' = f(\bar{x}) + dt \geq f(\bar{x})$, which is a contradiction. $\square$

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

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

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

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

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.

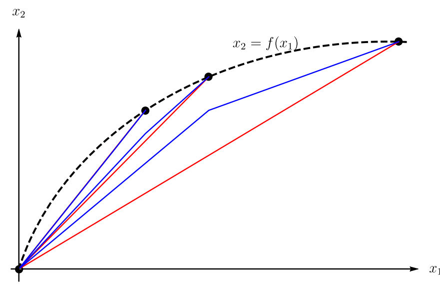

FIGURE 1. Intersection of two locally convex-like sets: X_1 and X_2 consist of the red and blue line segments respectively; their intersection is represented by the black points.

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 $f_i(x) = \operatorname{dist}(x, Y_i)$ 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 $X_1 = \bigcup_i L_i$, where each L_i is a line segment as shown in Figure 2. Thus, $f_1(x) = \min\left\{\min_i h_i(x), r_0(x), r_1(x)\right\}$ where $h_i(x) \triangleq \operatorname{dist}(x, L_i)$, $i = 1,2,\ldots$, and $r_j = \operatorname{dist}(x, R_j)$, $j = 1,2$. Let $S = \operatorname{conv}(Y_1)$ and $r(x) = \operatorname{dist}(x, S)$. Consider an arbitrary $\bar{x} \in \mathbb{R}^n$. There are four cases.

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

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 $\bar{x}$, which implies f_1 locally convex-like and B-differentiable at $\bar{x}$; see Figure 2 for illustration. It remains to show that f_1 is convex-like and directionally differentiable at $\bar{x} = 0$. Define a closed set $R = \{r : r = tx,\ t \geq 0, x \in Y_1\}$ as the cone generated by Y_1. Note that $f'(0; d) = \operatorname{dist}(d, R)$. Indeed, if $d = (1,1)$, then $f'(0; d) = \operatorname{dist}(d, R) = 0$. If d is not a scalar multiple of $(1,1)$, then since Y_1 is locally a finite union of line segments near td for $t > 0$ small enough, $f'(0; d) = \operatorname{dist}(d, R)$. Since $X_1 \subseteq R$, we have $f_1'(0; d) \leq f_1(d)$ and thus, f_1 is locally convex-like at 0. 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. □

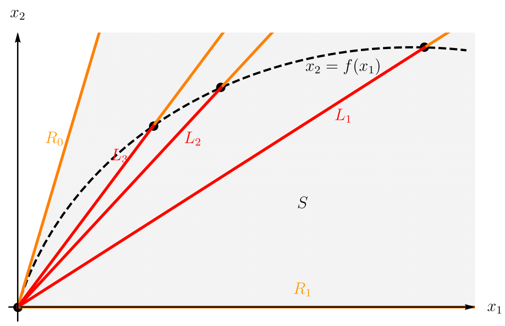

FIGURE 2. Illustration of Example 2. X_1 and R consists of the red line segments and orange rays respectively. The set S is represented by the shaded region.

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 11. 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 $g(t) \geq g(\bar{t}) + g'(\bar{t}; t - \bar{t})$ 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 $S = \arg\max\{h(t) : t_0 \leq t \leq t_1\}$ which is a compact set. Let $t_* = \max\{t : t \in S\}$. Then by construction $h(t) < h(t_*)$ for $t_* < t \leq t_1$. Since $h(t_0) = 0 > h(t_1)$, 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 $t_* \leq t < t_* + \varepsilon$, one has $h(t) \geq h(t_*) + h'(t_*; t - t_*) = h(t_*)$. However, this contradicts $h(t) < h(t_*)$ for all $t_* < t \leq t_1$.

• *Case 2:* $t_0 < t^* < t_1$. Since $h(t) < h(t_*)$ for all $t_* < t \leq t_1$, one can deduce that $h'(t_*; 1) \leq 0$. If $h'(t^*; 1) = 0$, we can repeat the same argument in the first case to draw a contradiction. For this reason, we assume $h'(t^*; 1) < 0$. Since $h(t) = g(t) - (t - t_0)g'(t_0; 1) - g(t_0)$ for $t \geq t_0$ by the Clarke regularity of g, it follows that $h'(t_*; \bullet)$ is convex, thus $h'(t_*; 1) + h'(t_*; -1) \geq h'(t^*; 0) = 0$, 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 $t_* \in S$. $\square$

Assume $X \triangleq \{\, x \in \mathbb{R}^n \mid f_k(x) \leq 0,\ k = 1, \cdots, K \,\}$, where each f_k is a locally convex-like function. Proposition 11 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 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 $\bar{x} \in X_{\mathrm{HSC}}$, 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 [10, Proposition 10]. Specifically, we have

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

where $\bar{\beta} \triangleq \{ i \,|\, \bar{x}_i \neq 0 \} \triangleq \text{supp}(\bar{x})$ is the *support* of the vector $\bar{x}$. [Remark: 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 12. Let P be a polyhedron. The set X_{ASC} is locally convex-like at every $\bar{x}$ in X_{ASC}.

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

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

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

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 $\bar{x} \in X_{\text{HSC}}$ and use this expression to show: (a) X_{HSC} is locally convex-like at $\bar{x}$ and (b) epi-stationarity of a B-differentiable objective function on the set X_{HSC} is sharper than pseudo B-stationarity.

Proposition 13. Let P be a polyhedron. Let each ψ_{ij} be an affine function and ϕ_{ij} be a piecewise affine function for all $j = 1, \cdots J_i$ and $i = 1, \cdots, m$. For $\bar{x} \in X_{\text{HSC}}$, it holds that

$\mathcal{T}(\bar{x}; X_{\text{HSC}}) \supseteq$ closure of

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

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

(A) for all $j \in \mathcal{J}_{i,0}(\bar{x})$,

$$[\, v \in \mathcal{T}(\bar{x}; P) \text{ and } \phi'_{ij}(\bar{x}; v) > 0\,] \;\Rightarrow\; \nabla\psi_{ij}(\bar{x})^\top v \geq 0;$$

(B) for all $j \in \mathcal{J}_{i,+}(\bar{x})$, $\nabla\psi_{ij}(\bar{x}) \in \mathcal{T}(\bar{x}; P)^*$, where $\mathcal{T}(\bar{x}; P)^*$ is the dual of $\mathcal{T}(\bar{x}; P)$,

then equality holds in (12).

Proof. Let $v \in \mathcal{T}(\bar{x}; P)$ satisfy the functional conditions in the right-hand set. We claim that v belongs to $\mathcal{T}(\bar{x}; X_{\text{HSC}})$ by showing that $\bar{x}^\tau \triangleq \bar{x} + \tau v \in X_{\text{HSC}}$ 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. Since P is a polyhedron, we have $x^\tau \in P$ for all $\tau > 0$ sufficiently small. Moreover, by continuity of ϕ_{ij}, we have

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

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

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

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

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

Therefore, we can further derive that

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

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

Conversely, let $v \in \mathcal{T}(\bar{x}; X_{\text{HSC}})$. Let $\{x^\nu\} \subset X_{\text{HSC}}$ be a sequence converging to $\bar{x}$ and $\{\tau_\nu\} \downarrow 0$ such that $v = \lim_{\nu \to \infty} w^\nu$, where $w^\nu \triangleq \dfrac{x^\nu - \bar{x}}{\tau_\nu}$ 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 $i = 1, \cdots, m$. We have for all $i = 1, \cdots, m$,

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

Hence, we obtain that

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

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

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

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

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

Hence w^ν belongs to the right-hand set in (12) without the closure. Since 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 13 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. Hence, for the set X_{AHC} with piecewise affine ϕ_{ij}, the result in Proposition 13 directly extends the tangent cone of X_{ASC} with the ℓ_0-function replaced by the Heaviside function composed with a piecewise affine function. We have the representation

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

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 13; 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 $j \in [J_i] \triangleq \{1, \cdots, J_i\}$ and $i = 1, \cdots, m$. Let $\boldsymbol{\Xi}(\bar{x})$ and $\boldsymbol{\Xi}^c(\bar{x})$ 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 $i = 1, \cdots, m$ and α_i^c is the complement of α_i in $\mathcal{J}_{i,0}(\bar{x})$. For each tuple $\boldsymbol{\alpha} \in \boldsymbol{\Xi}(\bar{x})$ with complement $\boldsymbol{\alpha^c} \in \boldsymbol{\Xi}^c(\bar{x})$, define the set

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

which may or may not contain the vector $\bar{x}$. Let $\overline{\boldsymbol{\Xi}}(\bar{x})$ be the subfamily of $\boldsymbol{\Xi}(\bar{x})$ 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 $\mathcal{T}(\bar{x}; X_{\rm HSC})$ in terms of the sets $\mathcal{S}_{\boldsymbol{\alpha}}(\bar{x})$ for all tuples $\boldsymbol{\alpha} \in \boldsymbol{\Xi}(\bar{x})$; in turn, this can be used to obtain a characterization of epi-stationarity of (3) without f being B-differentiable.

PROPOSITION 14. 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

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

Hence, if for all $\boldsymbol{\alpha} \in \overline{\boldsymbol{\Xi}}(\bar{x})$, the set $\mathcal{S}_{\boldsymbol{\alpha}}(\bar{x})$ is locally convex-like at $\bar{x}$, then so is $X_{\rm HSC}$. 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

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

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

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

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

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

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

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

showing that $x \in X_{\rm HSC}$. 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 $\mathcal{T}(\bar{x}; X_{\rm HSC})$, let $\{x^k\} \subset X_{\rm HSC}$ be a sequence converging to $\bar{x}$ and $\{\tau_k\} \downarrow 0$ such that $v = \lim_{k \to \infty} \dfrac{x^k - \bar{x}}{\tau_k}$. 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 $\overline{\boldsymbol{\Xi}}(\bar{x})$, by continuity of ψ_{ij}. This shows that $v \in \mathcal{T}(\bar{x}; \mathcal{S}_{\boldsymbol{\alpha}}(\bar{x}))$ for an index tuple $\boldsymbol{\alpha} \in \overline{\boldsymbol{\Xi}}(\bar{x})$, completing the proof of (13). The next-to-last statement of the proposition is clear because the union of 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. The expression (14) shows that for any closed set $S \subseteq \mathcal{N}$, the set $X_{\rm HSC} \cap S$ is closed, provided that the functions ψ_{ij} and ϕ_{ij} are continuous and ψ_{ij} is nonnegative. □

The example below shows that the nonnegativity assumption on ψ_{ij} is essential for the equality (13) to hold and that the piecewise affine property of the ϕ_{ij} functions is essential for the validity of Proposition 15.

EXAMPLE 3. Let

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

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 the 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.

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

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

PROPOSITION 15. 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 $\bar{x} \in X_{\text{HSC}}$ be arbitrary. With the same neighborhood $\mathcal{N}$ of $\bar{x}$ as defined in the proof of Proposition 14, it can similarly be proved that (no sign restriction on ψ_{ij} is needed)

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

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 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 $\widehat{\mathcal{S}}_{\boldsymbol{\alpha}}(\bar{x}) = \bigcup_{i\in I_{\boldsymbol{\alpha}}} S_{\boldsymbol{\alpha}}^i$, where $I_{\boldsymbol{\alpha}}$ is a certain finite index set, and each $S_{\boldsymbol{\alpha}}^i$ is a certain nonempty convex (not necessarily closed) set. Thus,

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

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

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

which yields,

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

Combining the last two expressions, we deduce

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

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

The discussion of the section is summarized in Table 6. Each entry is indexed by a combination of convexity and piecewise affinity imposed over the functions ψ_{ij} and ϕ_{ij} and 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 14. Proposition 15 illustrates the entry (1,2). The conclusion corresponding to the last entry can be proved using similar polyhedral decomposition techniques as in the proof of Proposition 15.

THEOREM 2. The set X_{HSC} is locally convex-like at every one of its elements if the assumptions given by any entry of Table 6 are true. In particular, X_{ASC} is locally convex-like and X_{AHC} is locally convex-like if each ϕ_{ij} is piecewise affine. $\square$

ψ_{ij} \ ϕ_{ij}	convex	piecewise affine
convex	$\psi_{ij} \geq 0$	free
piecewise affine	$\psi_{ij} \geq 0$	free

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

Combining Theorem 2 with Proposition 10, 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 6 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 epi-stationary point. $\square$

7. Computation of Pseudo- and Epi-Stationary 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 epi-stationary points of the 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 [7, Section 6] for the constraint $\sum_{j=1}^{J_i} \psi_{ij}(x)\,\mathbf{1}_{(0,\infty)}(\phi_{ij}(x)) \leq b_i$ 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 pseudo-stationary solution of the given HSC-constrained problem when the functions $\left\{ \{\psi_{ij}, \phi_{ij}\}_{j=1}^{J_i} \right\}_{i=1}^{m}$ 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 epi-stationary 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: $\psi_{ij}^{\pm} \triangleq \max(\pm\psi_{ij}, 0)$. Introducing an arbitrary scalar $\varepsilon \geq 0$, we note that

$$X_{\text{HSC}} = \left\{ x \in P \;\middle|\; \sum_{j=1}^{J_i} \psi_{ij}^{+;\varepsilon}(x)\,\mathbf{1}_{(0,\infty)}(\phi_{ij}(x)) \leq \sum_{j=1}^{J_i} \psi_{ij}^{-;\varepsilon}(x)\,\mathbf{1}_{(0,\infty)}(\phi_{ij}(x)) + b_i,\; i = 1, \cdots, m \right\},$$

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

- **t-lifting:**

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

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

Next, we note that,

$$t_{ij} \leq \mathbf{1}_{(0,\infty)}(\phi_{ij}(x)) \ \Leftrightarrow\ \exists\, y_{ij} \geq 0 \text{ such that } t_{ij} \leq \min\left(\phi_{ij}^{+}(x) y_{ij},\, 1\right). \tag{15}$$

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

- **(t,y)-lifting:**

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

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 [7, Section 7], particularly Proposition 7 therein, by considering the relaxation $u_{ij} \geq \psi_{ij}^{+;\varepsilon}(x)\mathbf{1}_{(0,\infty)}(\phi_{ij}(x))$ and replacing it using a piecewise composite function; this leads to

- **(t,y,u)-lifting:**

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

which is also closed if ϕ_{ij} and ψ_{ij} are continuous; moreover if these functions are B-differentiable, then all the inequalities in $\widehat{X_{\rm HSC}^{t,y,u;\varepsilon}}$ 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 $\widehat{X_{\rm HSC}^{t,y,u;\varepsilon}}$ 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 [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:**

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

In the case of an affine sparsity constraint system:

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

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

$$\widetilde{X_{\rm ASC}^{t,y,s;\varepsilon}} \triangleq \left\{ \begin{array}{l|l} x \in P & \displaystyle\sum_{j=1}^{n} (a_{ij}^{+} + \varepsilon)\,s_j \leq \sum_{j=1}^{n} (a_{ij}^{-} + \varepsilon)\,t_j + b_i, \quad i = 1,\cdots,m \\ t_j,\, s_j \in [0,1], \text{ all } j & t_j \leq |x_j|\,y_j, \qquad j = 1,\cdots,n \\ y_j \geq 0, \text{ all } j & x_j\,(1-s_j) = 0, \; j = 1,\cdots,n \end{array} \right\},$$

where the only nonlinear functional constraints are defined by products of two variables. A noteworthy remark about $\widetilde{X_{\rm HSC}^{t,y,s;\varepsilon}}$ 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 due to their associations with the respective signed functions $\psi_{ij}^{\pm}$.

The two lifted sets $\widehat{X_{\rm HSC}^{t,y,u;\varepsilon}}$ and $\widetilde{X_{\rm HSC}^{t,y,s;\varepsilon}}$ offer a computationally tractable venue for the minimization of a wide class of nonconvex nondifferentiable objective functions f over the non-closed set $X_{\rm HSC}$, provided that f and all the functions ϕ_{ij} and ψ_{ij} are surrogatable by pointwise minima of convex differentiable functions; see [9, Chapter 7]. We omit the algorithmic details.

7.2. Recovering pseudo stationarity. 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 $\left\{ \{\phi_{ij}, \psi_{ij}\}_{j=1}^{J_i} \right\}_{i=1}^{m}$ are B-differentiable (which do not imply that the set $X_{\rm HSC}$ 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 $\widetilde{X^{t,y,s;\varepsilon}_{\rm HSC}}$, then $\bar{x}$ is pseudo B-stationary of f on $X_{\rm HSC}$; i.e., $\bar{x}$ is a B-stationary point of the problem

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

whose feasible set is a subset of $X_{\rm HSC}$. Note that for x sufficiently close to $\bar{x}$, we must have $\phi_{ij}(x) \geq 0$ for all $j \in \mathcal{J}_{i,+}(\bar{x})$; cf. the constraints in (8). By the definition of the pseudo B-stationarity of f on $X_{\rm HSC}$, we know that $\bar{x}$ must be feasible to $X_{\rm HSC}$ that is possibly non-closed. We omit the analysis for the set $\widehat{X^{t,y,u;\varepsilon}_{\rm HSC}}$ 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 $\widetilde{X^{t,y,s;\varepsilon}_{\rm HSC}}$. The scalar ε plays an important role for the validity of the result.

Proposition 16. 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 $\widetilde{X^{t,y,s;\varepsilon}_{\rm HSC}}$, then $\bar{x}$ is a pseudo stationary point of f on $X_{\rm HSC}$.

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

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

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

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

which is equivalent to $\displaystyle\sum_{j \in \mathcal{J}_{i,+}(\bar{x})} \psi_{ij}(x) \leq b_i$. 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 $\displaystyle\lim_{k \to \infty} \frac{x^k - \bar{x}}{\tau_k} = v$. We need to show that $f'(\bar{x}; v) \geq 0$. 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 $\widetilde{X^{t,y,s;\varepsilon}_{\rm HSC}}$ for all k sufficiently large and the three sequences:

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

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, since x^k is feasible to (16), we must have $\phi_{ij}(x^k) > 0$ implies $\phi_{ij}(\bar{x}) > 0$. Hence,

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

Since 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 $\Delta(\bar{x};\bar{x}) \leq 0$. Let

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

We have

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

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

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

Case I: Suppose that $S \cup T \neq \emptyset$. Since $\varepsilon > 0$, the above inequality implies that

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

We can write

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

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

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

Since $\Delta(\bullet;\bar{x})$ is continuous, it follows that $\lim\limits_{k\to\infty}\Delta_j^s(x^k) = \Delta_j^s$ and $\lim\limits_{k\to\infty}\Delta_j^t(x^k) = \Delta_j^t$. Next, we construct a sequence (t^k, y^k, s^k) such that $(x^k, t^k, y^k, s^k) \in \widetilde{X^{t,y,s;\varepsilon}_{\text{HSC}}}$. Let

$$\begin{aligned}
s_j^k &\triangleq \begin{cases} \min\left\{\dfrac{\Delta_j^s(x^k)}{\psi_j^+(x^k)+\varepsilon},\,\bar{s}_j\right\} & \text{if } j\in S \\ \bar{s}_j & \text{otherwise;} \end{cases} \\
1 - t_j^k &\triangleq \begin{cases} \min\left\{\dfrac{\Delta_j^t(x^k)}{\psi_j^-(x^k)+\varepsilon},\,1-\bar{t}_j\right\} & \text{if } j\in T \\ 1-\bar{t}_j & \text{otherwise;} \end{cases} \qquad (22) \\
y_j^k &\triangleq \begin{cases} \max\left\{\dfrac{t_j^k}{\phi_j(x^k)},\,\bar{y}_j\right\} & \text{if } \phi_j(x^k) > 0 \; (\Leftrightarrow\; \phi_j(\bar{x}) > 0\,) \\ \bar{y}_j & \text{if } \phi_j(x^k) \leq 0. \end{cases}
\end{aligned}$$

We need to verify the functional inequalities in $\widetilde{X_{\text{HSC}}^{t,y,s;\varepsilon}}$. These are done in the following 3 steps.

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

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

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 $(x^k,t^k,y^k,s^k)\in\widetilde{X_{\text{HSC}}^{t,y,s;\varepsilon}}$ 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 $\lim\limits_{k\to\infty}s_j^k=\bar{s}_j$ for all j. Similarly, we can show that $\lim\limits_{k\to\infty}t_j^k=\bar{t}_j$ and $\lim\limits_{k\to\infty}y_j^k=\bar{y}_j$ for all j for all j. Since 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 $\dfrac{t_j^k-\bar{t}_j}{\tau_k}$, $\dfrac{s_j^k-\bar{s}_j}{\tau_k}$ and $\dfrac{y_j^k-\bar{y}_j}{\tau_k}$ are bounded.

Case II:: 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 16 to the stronger property of B-stationarity of f

on X_{HSC}. Specifically, consider the set with an additional scalar $\eta > 0$:

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

We have the following result.

Proposition 17. 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 $(\bar{x}, \bar{t}, \bar{y}, \bar{s}) \in \widetilde{X^{t,y,s;\varepsilon}_{\text{HSC}}}$ is a B-stationary point of f on $\widetilde{X^{t,y,s;\varepsilon}_{\text{HSC};\eta}}$, then $\bar{x}$ is a B-stationary point of f on X_{HSC}.

Proof. We proceed as in the proof of Proposition 16. Let $\{x^k\}$ be a sequence in X_{HSC} converging to $\bar{x}$ and $\{\tau_k\} \downarrow 0$ such that $\displaystyle\lim_{k\to\infty} \frac{x^k - \bar{x}}{\tau_k} = v$. 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 $\widetilde{X^{t,y,s;\varepsilon}_{\text{HSC};\eta}}$ for all k sufficiently large and the three sequences:

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

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^k_j, t^k_j, y^k_j) by (22), we see that the proof of Steps 1 and 2 in Proposition 16 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 $s^k_j = \bar{s}_j = 1$. If $\phi_j(\bar{x}) = 0$, then $\phi^{+}_j(x^k)\,(1 - s^k_j) \le \eta$ for k sufficiently large.

Summarizing the 3 steps, we have established $(x^k, t^k, y^k, s^k) \in \widetilde{X^{t,y,s;\varepsilon}_{\text{HSC};\eta}}$ 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 work of Cui was based on research supported by the U.S. National Science Foundation under grants CCF-2153352 and DMS-2309729, and National Institutes of Health under grant 1R01CA287413-01. The work of Pang was based on research supported by the U.S. Air Force Office of Scientific Research under grant FA9550-22-1-0045.

References

[1] Atamtürk, A., Gómez, A., and Han, S. (2021). Sparse and smooth signal estimation: Convexification of ℓ_0-formulations. *Journal of Machine Learning Research*, 22(52):1–43.

[2] Bian, W. and Chen, X. (2020). A smoothing proximal gradient algorithm for nonsmooth convex regression with cardinality penalty. *SIAM Journal on Numerical Analysis*, 58(1):858–883.

[3] Chen, X., Pan, L., and Xiu, N. (2022). Solution sets of three sparse optimization problems for multivariate regression. *Journal of Global Optimization*, pages 1–25.

[4] Clarke, F. H. (1990). *Optimization and nonsmooth analysis*, volume 5. SIAM.

[5] Conforti, M., Cornuéjols, G., and Zambelli, G. (2014). *Integer programming*, volume 271. Springer.

[6] Cui, Y., Chang, T.-H., Hong, M., and Pang, J.-S. (2020). A study of piecewise linear-quadratic programs. *Journal of Optimization Theory and Applications*, 186:523–553.

[7] Cui, Y., Liu, J., and Pang, J.-S. (2022a). The minimization of piecewise functions: Pseudo stationarity. *Journal of Convex Analysis (accepted).*

[8] Cui, Y., Liu, J., and Pang, J.-S. (2022b). Nonconvex and nonsmooth approaches for affine chance-constrained stochastic programs. *Set-Valued and Variational Analysis*, 30(3):1149–1211.

[9] Cui, Y. and Pang, J.-S. (2021). *Modern nonconvex nondifferentiable optimization*. SIAM Publications.

[10] Dong, H., Ahn, M., and Pang, J.-S. (2019). Structural properties of affine sparsity constraints. *Mathematical Programming*, 176:95–135.

[11] Dong, H. and Linderoth, J. (2013). On valid inequalities for quadratic programming with continuous variables and binary indicators. In *Integer Programming and Combinatorial Optimization: 16th International Conference, IPCO 2013, Valparaíso, Chile, March 18-20, 2013. Proceedings 16*, pages 169–180. Springer.

[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, CA.

[13] Ferris, M. C. and Mangasarian, O. L. (1992). Minimum principle sufficiency. *Mathematical programming*, 57(1-3):1–14.

[14] Gómez, A., He, Z., and Pang, J.-S. (2023). Linear-step solvability of some folded concave and singly-parametric sparse optimization problems. *Mathematical Programming*, 198(2):1339–1380.

[15] Hastie, T., Tibshirani, R., and Wainwright, M. (2015). *Statistical learning with sparsity: the lasso and generalizations*. CRC press.

[16] He, Z., Han, S., Gómez, A., Cui, Y., and Pang, J.-S. (2023). Comparing solution paths of sparse quadratic minimization with a stieltjes matrix. *Mathematical Programming*.

[17] Kanzow, C., Raharja, A. B., and Schwartz, A. (2021). Sequential optimality conditions for cardinality-constrained optimization problems with applications. *Computational Optimization and Applications*, 80:185–211.

[18] Kanzow, C., Schwarz, A., and Weiß, F. (2022). The sparse (st) optimization problem: Reformulations, optimality, stationarity, and numerical results. *arXiv preprint arXiv:2210.09589*.

[19] Mangasarian, O. L. (1975). Pseudo-convex functions. In *Stochastic optimization models in finance*, pages 23–32. Elsevier.

[20] Mangasarian, O. L. (1994). *Nonlinear programming*. SIAM Publications.

[21] Pang, J.-S., Razaviyayn, M., and Alvarado, A. (2017). Computing b-stationary points of nonsmooth dc programs. *Mathematics of Operations Research*, 42(1):95–118.

[22] Qi, Z., Cui, Y., Liu, Y., and Pang, J.-S. (2019). Estimation of individualized decision rules based on an optimized covariate-dependent equivalent of random outcomes. *SIAM Journal on Optimization*, 29(3):2337–2362.

[23] Rockafellar, R. T. and Wets, R. J.-B. (2009). *Variational analysis*, volume 317. Springer Science & Business Media.

[24] Royset, J. O. (2023). Consistent approximations in composite optimization. *Mathematical Programming*, 201(1-2):339–372.

[25] Royset, J. O. and Wets, R. J.-B. (2022). *An optimization primer*. Springer.