

# Variational Theory and Algorithms for a Class of Asymptotically Approachable Nonconvex Problems

Hanyang Li <sup>\*</sup>      Ying Cui <sup>†</sup>

November 20, 2024

## Abstract

We investigate a class of composite nonconvex functions, where the outer function is the sum of univariate extended-real-valued convex functions and the inner function is the limit of difference-of-convex functions. A notable feature of this class is that the inner function may fail to be locally Lipschitz continuous. It covers a range of important yet challenging applications, including inverse optimal value optimization and problems under value-at-risk constraints. We propose an asymptotic decomposition of the composite function that guarantees epi-convergence to the original function, leading to necessary optimality conditions for the corresponding minimization problem. The proposed decomposition also enables us to design a numerical algorithm such that any accumulation point of the generated sequence, if exists, satisfies the newly introduced optimality conditions. These results expand on the study of so-called amenable functions introduced by Poliquin and Rockafellar in 1992, which are compositions of convex functions with smooth maps, and the prox-linear methods for their minimization. To demonstrate that our algorithmic framework is practically implementable, we further present verifiable termination criteria and preliminary numerical results.

**Keywords:** epi-convergence; optimality conditions; nonsmooth analysis; difference-of-convex functions

## 1 Introduction.

We consider a class of composite optimization problems of the form:

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \sum_{p=1}^m \left[ F_p(x) \triangleq \varphi_p(f_p(x)) \right], \quad (\text{CP}_0)$$

where for each  $p = 1, \dots, m$ , the outer function  $\varphi_p : \mathbb{R} \rightarrow \mathbb{R} \cup \{+\infty\}$  is proper, convex, lower semicontinuous (lsc), and the inner function  $f_p : \mathbb{R}^n \rightarrow \mathbb{R}$  is not necessarily locally Lipschitz continuous.

---

<sup>\*</sup>Department of Industrial Engineering and Operations Research, University of California, Berkeley, Berkeley 94720 (hanyang.li@berkeley.edu)

<sup>†</sup>Department of Industrial Engineering and Operations Research, University of California, Berkeley, Berkeley 94720 (yingcui@berkeley.edu)

If each inner function  $f_p$  is continuously differentiable, then the objective in  $(\text{CP}_0)$  belongs to the family of *amenable functions* under a constraint qualification [25, 26]. For a thorough exploration of the variational theory of amenable functions, readers are referred to [30, Chapter 10(F)]. The properties of amenable functions have also led to the development of prox-linear algorithms, where convex subproblems are constructed through the linearization of the inner smooth mapping [16, 4, 5, 19, 14].

However, there are various applications of composite optimization problem in the form of  $(\text{CP}_0)$  where the inner function  $f_p$  is nondifferentiable. In the following, we provide two such examples.

**Example 1.1** (The inverse optimal value optimization). For  $p = 1, \dots, m$ , consider the optimal value function

$$f_p(x) \triangleq \inf_{y \in \mathbb{R}^d} \left\{ (c^p + C^p x)^\top y + \frac{1}{2} y^\top Q^p y \mid A^p x + B^p y \leq b^p \right\} \quad x \in \mathbb{R}^n \quad (1)$$

with appropriate dimensional vectors  $b^p$  and  $c^p$ , and matrices  $A^p, B^p, C^p$  and  $Q^p$ . The function  $f_p$  is not smooth in general. The inverse (multi) optimal value problem [2, 24] finds a vector  $x \in \mathbb{R}^n$  that minimizes the discrepancy between observed optimal values  $\{\nu_p\}_{p=1}^m$  and true optimal values  $\{f_p(x)\}_{p=1}^m$  based on a prescribed metric, such as the  $\ell_1$ -error:

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \sum_{p=1}^m |\nu_p - f_p(x)|. \quad (2)$$

If  $f_p$  is real-valued for  $p = 1, \dots, m$ , one can express problem (2) in the form of  $(\text{CP}_0)$  by defining the outer function  $\varphi_p(t) = |\nu_p - t|$ .

**Example 1.2** (The portfolio optimization under a value-at-risk constraint). Given a random variable  $Y$ , the Value-at-risk (VaR) of  $Y$  at a confidence level  $\alpha \in (0, 1)$  is defined as  $\text{VaR}_\alpha(Y) \triangleq \min \{\gamma \in \mathbb{R} \mid \mathbb{P}(Y \leq \gamma) \geq \alpha\}$ . Let  $Z$  be the random return of investments and  $c(\cdot, \cdot)$  be a lsc function representing the profit of  $Z$  parameterized by  $x \in \mathbb{R}^n$ . An agent's goal is to maximize the expected profit, denoted by  $\mathbb{E}[c(x, Z)]$ , while also controlling the risk via a constraint on  $\text{VaR}_\alpha[c(x, Z)]$  under a prescribed level  $r$ . The model can be written as

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad -\mathbb{E}[c(x, Z)] \quad \text{subject to} \quad \text{VaR}_\alpha[c(x, Z)] \geq r. \quad (3)$$

Define  $\delta_A$  as the indicator function of a set  $A$ , where  $\delta_A(t) = 0$  for  $t \in A$  and  $\delta_A(t) = +\infty$  for  $t \notin A$ . Problem (3) can then be put into the framework  $(\text{CP}_0)$  by setting  $\varphi_1(t) = -t$ ,  $f_1(x) = \mathbb{E}[c(x, Z)]$ ,  $\varphi_2(t) = \delta_{[r, +\infty)}(t)$ , and  $f_2(x) = \text{VaR}_\alpha[c(x, Z)]$ . We note that the function  $\text{VaR}_\alpha[c(\cdot, Z)]$  can be nondifferentiable even if the function  $c(\cdot, z)$  is differentiable for every  $z$ .

Due to the nondifferentiability of the inner function  $f_p$  in  $(\text{CP}_0)$ , the overall objective is not amenable and the prox-linear algorithm [16] is not applicable to solve this composite optimization problem. In this paper, we develop an algorithmic framework for a subclass of  $(\text{CP}_0)$ , where each inner function  $f_p$ , although nondifferentiable, can be derived from DC functions through a limiting process. We refer to this class of functions as *approachable difference-of-convex (ADC) functions* (see section 2.1 for the formal definition). It is important to note that ADC functions are ubiquitous. In particular, we will show that the optimal value function  $f_p$  in (1) and  $\text{VaR}_\alpha[c(\cdot, Z)]$  in (3) are

instances of ADC functions under mild conditions. In fact, based on the result recently shown in [31], any lsc function is the epi-limit of piecewise affine DC functions.

With this new class of functions in hand, we have made a first step to understand the variational properties of the composite ADC minimization problem  $(\text{CP}_0)$ , including an in-depth analysis of its necessary optimality conditions. The novel optimality conditions are defined through a handy approximation of the subdifferential mapping  $\partial f_p$  that explores the ADC structure of  $f_p$ . Using the notion of *epi-convergence*, we further show that these optimality conditions are necessary conditions for any local solution of  $(\text{CP}_0)$ . Additionally, we propose a double-loop algorithm to solve  $(\text{CP}_0)$ , where the outer loop dynamically updates the DC functions approximating each  $f_p$ , and the inner loop finds an approximate stationary point of the resulting composite DC problem through successive convex approximations. It can be shown that any accumulation point of the sequence generated by our algorithm satisfies the newly introduced optimality conditions.

Our strategy to handle the nondifferentiable and possibly discontinuous inner function  $f_p$  through a sequence of DC functions shares certain similarities with the approximation frameworks in the existing literature. For instance, Ermoliev et al. [15] have designed smoothing approximations for lsc functions utilizing convolutions with bounded mollifier sequences, a technique akin to local “averaging”. Research has sought to identify conditions that ensure gradient consistency for the smoothing approximation of composite nonconvex functions [10, 8, 6, 7]. Notably, Burke and Hoheisel [6] have emphasized the importance of epi-convergence for the approximating sequence, a less stringent requirement than the continuous convergence assumed in earlier works [10, 3]. In recent work, Royset [32] has studied the consistent approximation of the composite optimization in terms of the global minimizers and stationary solutions, where the inner function is assumed to be locally Lipschitz continuous. Our notion of subdifferentials and optimality conditions for  $(\text{CP}_0)$  takes inspiration from these works but adapts to accommodate nonsmooth approximating sequences that exhibit the advantageous property of being DC.

The rest of the paper is organized as follows. Section 2 presents a class of ADC functions and introduces a new associated notion of subdifferential. In section 3, we investigate the necessary optimality conditions for problem  $(\text{CP}_0)$ . Section 4 is devoted to an algorithmic framework for solving  $(\text{CP}_0)$  and its convergence analysis to the newly introduced optimality conditions. We also discuss termination criteria for practical implementation in section 4.3. Preliminary numerical experiments on the inverse optimal value problems are presented in the last section.

**Notation and Terminology.** Let  $\|\cdot\|$  denote the Euclidean norm in  $\mathbb{R}^n$ . We use the symbol  $\mathbb{B}(\bar{x}, \delta)$  to denote the Euclidean ball  $\{x \in \mathbb{R}^n \mid \|x - \bar{x}\| \leq \delta\}$ . The set of nonpositive and nonnegative are denoted by  $\mathbb{R}_-$  and  $\mathbb{R}_+$ , respectively, and the set of nonnegative integers is denoted by  $\mathbb{N}$ . We write  $\mathbb{N}_\infty^\# \triangleq \{N \subset \mathbb{N} \mid N \text{ infinite}\}$  and  $\mathbb{N}_\infty \triangleq \{N \mid \mathbb{N} \setminus N \text{ finite}\}$ . Notation  $\{t^k\}$  is used to simplify the expression of any sequence  $\{t^k\}_{k \in \mathbb{N}}$ , where the elements can be points, sets, or functions. By  $t^k \rightarrow t$  and  $t^k \rightarrow_N t$ , we mean that the sequence  $\{t^k\}$  and the subsequence  $\{t^k\}_{k \in N}$  indexed by  $N \in \mathbb{N}_\infty^\#$  converge to  $t$ , respectively.

Given two sets  $A$  and  $B$  in  $\mathbb{R}^n$  and a scalar  $\lambda \in \mathbb{R}$ , the Minkowski sum and the scalar multiple are defined as  $A + B \triangleq \{a + b \mid a \in A, b \in B\}$  and  $\lambda A \triangleq \{\lambda a \mid a \in A\}$ . We also define  $0 \cdot \emptyset = \{0\}$  and  $\lambda \cdot \emptyset = \emptyset$  whenever  $\lambda \neq 0$ . When  $A$  and  $B$  are nonempty and closed, we define the one-sided deviation of  $A$  from  $B$  as  $\mathbb{D}(A, B) \triangleq \sup_{x \in A} \text{dist}(x, B)$ , where  $\text{dist}(x, B) \triangleq \inf_{y \in B} \|y - x\|$ . The Hausdorff distance between  $A$  and  $B$  is given by  $\mathbb{H}(A, B) \triangleq \max\{\mathbb{D}(A, B), \mathbb{D}(B, A)\}$ . The boundary and interior of  $A$  are denoted by  $\text{bdry}(A)$  and  $\text{int}(A)$ . The topological closure and the convex hull

of  $A$  are indicated by  $\text{cl}(A)$  and  $\text{con } A$ .

For a sequence of sets  $\{C^k\}$ , we define its outer limit as

$$\text{Lim sup}_{k \rightarrow +\infty} C^k \triangleq \{u \mid \exists N \in \mathbb{N}_{\infty}^{\sharp}, u^k \rightarrow_N u \text{ with } u^k \in C^k\},$$

and the horizon outer limit as

$$\text{Lim sup}_{k \rightarrow +\infty}^{\infty} C^k \triangleq \{0\} \cup \left\{ u \mid \exists N \in \mathbb{N}_{\infty}^{\sharp}, \lambda_k \downarrow 0, \lambda_k u^k \rightarrow_N u \text{ with } u^k \in C^k \right\}.$$

The outer limit of a set-valued mapping  $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$  is defined as

$$\text{Lim sup}_{x \rightarrow \bar{x}} S(x) \triangleq \bigcup_{x^k \rightarrow \bar{x}} \text{Lim sup}_{k \rightarrow +\infty} S(x^k) = \{u \mid \exists x^k \rightarrow \bar{x}, u^k \rightarrow u \text{ with } u^k \in S(x^k)\} \quad \bar{x} \in \mathbb{R}^n.$$

We say  $S$  is outer semicontinuous (osc) at  $\bar{x} \in \mathbb{R}^n$  if  $\text{Lim sup}_{x \rightarrow \bar{x}} S(x) \subset S(\bar{x})$ . Consider some index set  $N \in \mathbb{N}_{\infty}^{\sharp}$ . A sequence of sets  $\{C^k\}_{k \in N}$  is equi-bounded if there exists a bounded set  $B$  such that  $C^k \subset B$  for all  $k \in N$ . Otherwise, the sequence is unbounded. If there is an integer  $K \in N$  such that  $\{C^k\}_{k \in N, k \geq K}$  is equi-bounded, then the sequence  $\{C^k\}_{k \in N}$  is said to be eventually bounded. Interested readers are referred to [30, Chapter 4] for a comprehensive study of set convergence.

The regular normal cone and the limiting normal cone of a set  $C \subset \mathbb{R}^n$  at  $\bar{x} \in C$  are given by

$$\widehat{\mathcal{N}}_C(\bar{x}) \triangleq \left\{ v \mid v^{\top}(x - \bar{x}) \leq o(\|x - \bar{x}\|) \text{ for all } x \in C \right\} \quad \text{and} \quad \mathcal{N}_C(\bar{x}) \triangleq \text{Lim sup}_{x \in C \rightarrow \bar{x}} \widehat{\mathcal{N}}_C(x).$$

The proximal normal cone of a set  $C$  at  $\bar{x} \in C$  is defined as  $\mathcal{N}_C^p(\bar{x}) \triangleq \{\lambda(x - \bar{x}) \mid \bar{x} \in P_C(x), \lambda \geq 0\}$ , where  $P_C$  is the projection onto  $C$  that maps any  $x$  to the set of points in  $C$  that are closest to  $x$ .

For an extended-real-valued function  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}} \triangleq \mathbb{R} \cup \{\pm\infty\}$ , we write its effective domain as  $\text{dom } f \triangleq \{x \in \mathbb{R}^n \mid f(x) < +\infty\}$ , and the epigraph as  $\text{epi } f \triangleq \{(x, \alpha) \in \mathbb{R}^{n+1} \mid \alpha \geq f(x)\}$ . We say  $f$  is proper if  $\text{dom } f$  is nonempty and  $f(x) > -\infty$  for all  $x \in \mathbb{R}^n$ . We adopt the common rules for extended arithmetic operations, and the lower and upper limits of a sequence of scalars in  $\overline{\mathbb{R}}$  (cf. [30, Chapter 1(E)]).

Let  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  be a proper function. We write  $x \rightarrow_f \bar{x}$ , if  $x \rightarrow \bar{x}$  and  $f(x) \rightarrow f(\bar{x})$ . The regular subdifferential and the limiting subdifferential of  $f$  at  $\bar{x} \in \text{dom } f$  are respectively defined as

$$\widehat{\partial}f(\bar{x}) \triangleq \{v \mid f(x) \geq f(\bar{x}) + v^{\top}(x - \bar{x}) + o(\|x - \bar{x}\|) \text{ for all } x\} \quad \text{and} \quad \partial f(\bar{x}) \triangleq \text{Lim sup}_{x \rightarrow_f \bar{x}} \widehat{\partial}f(x).$$

For any  $\bar{x} \notin \text{dom } f$ , we set  $\widehat{\partial}f(\bar{x}) = \partial f(\bar{x}) = \emptyset$ . When  $f$  is locally Lipschitz continuous at  $\bar{x}$ ,  $\text{con } \partial f(\bar{x})$  equals to the Clarke subdifferential  $\partial_C f(\bar{x})$ . We further say  $f$  is subdifferentially regular at  $\bar{x} \in \text{dom } f$  if  $f$  is lsc at  $\bar{x}$  and  $\widehat{\partial}f(\bar{x}) = \partial f(\bar{x})$ . When  $f$  is proper and convex,  $\widehat{\partial}f$ ,  $\partial f$ , and  $\partial_C f$  coincide with the concept of the subdifferential in convex analysis.

Finally, we introduce the notion of function convergence. A sequence of functions  $\{f^k : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}\}$  is said to converge pointwise to  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ , written  $f^k \xrightarrow{p} f$ , if  $\lim_{k \rightarrow +\infty} f^k(x) = f(x)$  for any  $x \in \mathbb{R}^n$ . The sequence  $\{f^k\}$  is said to epi-converge to  $f$ , written  $f^k \xrightarrow{e} f$ , if for any  $x$ , it holds

$$\begin{cases} \liminf_{k \rightarrow +\infty} f^k(x^k) \geq f(x) & \text{for every sequence } x^k \rightarrow x, \\ \limsup_{k \rightarrow +\infty} f^k(x^k) \leq f(x) & \text{for some sequence } x^k \rightarrow x. \end{cases}$$

The sequence  $\{f^k\}$  is said to converge continuously to  $f$ , written  $f^k \xrightarrow{c} f$ , if  $\lim_{k \rightarrow +\infty} f^k(x^k) = f(x)$  for any  $x$  and any sequence  $x^k \rightarrow x$ .

## 2 Approachable difference-of-convex functions.

In this section, we formally introduce a class of functions that can be asymptotically approximated by DC functions. A new concept of subdifferential that is defined through the approximating functions is proposed. At the end of this section, we provide several examples that demonstrate the introduced concepts.

### 2.1 Definitions and properties.

An extended-real-valued function can be approximated by a sequence of functions in various notions of convergence, as comprehensively investigated in [30, Chapter 7(A-C)]. Among these approaches, epi-convergence has a notable advantage in its ability to preserve the global minimizers [30, Theorem 7.31]. Our focus lies on a particular class of approximating functions, wherein each function exhibits a DC structure.

**Definition 1.** *A function  $f$  is said to be DC on its domain if there exist proper, lsc and convex functions  $g, h : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  such that  $\text{dom } f = [\text{dom } g \cap \text{dom } h]$  and  $f(x) = g(x) - h(x)$  for any  $x \in \text{dom } f$ .*

With this definition, we introduce the concept of ADC functions.

**Definition 2** (ADC functions). *Let  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  be a proper function.*

(a)  *$f$  is said to be pointwise approachable DC (p-ADC) if there exist proper functions  $\{f^k : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}\}$ , DC on their respective domains, such that  $f^k \xrightarrow{p} f$ .*

(b)  *$f$  is said to be epigraphically approachable DC (e-ADC) if there exist proper functions  $\{f^k : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}\}$ , DC on their respective domains, such that  $f^k \xrightarrow{e} f$ .*

(c)  *$f$  is said to be continuously approachable DC (c-ADC) if there exist proper functions  $\{f^k : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}\}$ , DC on their respective domains, such that  $f^k \xrightarrow{c} f$ .*

*A function  $f$  is said to be ADC associated with  $\{f^k\}$  if  $\{f^k\}$  confirms one of these convergence properties. By a slight abuse of notation, we denote the DC decomposition of each  $f^k$  as  $f^k = g^k - h^k$ , although the equality may only hold for  $x \in \text{dom } f^k$ .*

A p-ADC function may not be lsc. An example is given by  $f(x) = \mathbf{1}_{\{0\}}(x) + 2 \cdot \mathbf{1}_{(0,+\infty)}(x)$ , where for a set  $C \subset \mathbb{R}^n$ , we write  $\mathbf{1}_C(x) = 1$  if  $x \in C$  and  $\mathbf{1}_C(x) = 0$  if  $x \notin C$ . In this case,  $f$  is not lsc at  $x = 0$ . However,  $f$  is p-ADC associated with  $f^k(x) = \max(0, 2kx + 1) - \max(0, 2kx - 1)$ . In contrast, any e-ADC function must be lsc [30, Proposition 7.4(a)], and any c-ADC function is continuous [30, Theorem 7.14].

The relationships among different notions of function convergence, including the unaddressed uniform convergence, have been thoroughly examined in [30]. Generally, pointwise convergence and epi-convergence do not imply one another, but they coincide when the sequence  $\{f^k\}$  is asymptotically equi-lsc everywhere [30, Theorem 7.10]. In addition,  $\{f^k\}$  converges continuously to  $f$  if and only if both  $f^k \xrightarrow{e} f$  and  $(-f^k) \xrightarrow{e} (-f)$  are satisfied [30, Theorem 7.11]. While verifying epi-convergence is often challenging, it becomes simpler for a monotonic sequence  $\{f^k\}$  that converges pointwise to  $f$  [30, Proposition 7.4(c-d)].

## 2.2 Subdifferentials of ADC functions.

Characterizing the limiting and Clarke subdifferentials can be challenging when dealing with functions that exhibit complex composite structures. Our focus in this subsection is on numerically computable approximations of the limiting subdifferentials. We begin with the definitions.

**Definition 3** (approximate subdifferentials). *Consider an ADC function  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  associated with  $\{f^k = g^k - h^k\}$ . The approximate subdifferential of  $f$  (associated with  $\{f^k = g^k - h^k\}$ ) at  $\bar{x} \in \mathbb{R}^n$  is defined as*

$$\partial_A f(\bar{x}) \triangleq \bigcup_{x^k \rightarrow \bar{x}} \limsup_{k \rightarrow +\infty} [\partial g^k(x^k) - \partial h^k(x^k)].$$

The approximate horizon subdifferential of  $f$  (associated with  $\{f^k = g^k - h^k\}$ ) at  $\bar{x} \in \mathbb{R}^n$  is defined as

$$\partial_A^\infty f(\bar{x}) \triangleq \bigcup_{x^k \rightarrow \bar{x}} \limsup_{k \rightarrow +\infty}^\infty [\partial g^k(x^k) - \partial h^k(x^k)].$$

Unlike the limiting subdifferential which requires  $x^k \rightarrow_f \bar{x}$ ,  $\partial_A f(x)$  is defined using all the sequences  $x^k \rightarrow \bar{x}$  without necessitating the convergence of function values. It follows directly from the definitions that the mappings  $x \mapsto \partial_A f(x)$  and  $x \mapsto \partial_A^\infty f(x)$  are osc. The following proposition presents a sufficient condition for  $\partial_A f(\bar{x}) = \partial f(\bar{x}) = \emptyset$  at any  $\bar{x} \notin \text{dom } f$ .

**Proposition 1.** *Let  $\bar{x} \notin \text{dom } f$ . Then  $\partial_A f(\bar{x}) = \emptyset$  if for any sequence  $x^k \rightarrow \bar{x}$ , we have  $x^k \notin \text{dom } f^k$  for all sufficiently large  $k$ . The latter condition is particularly satisfied whenever  $\text{dom } f$  is closed and  $\text{dom } f^k \subset \text{dom } f$  for all sufficiently large  $k$ .*

*Proof.* Note that for any  $x^k \rightarrow \bar{x} \notin \text{dom } f$ , we have  $[\partial g^k(x^k) - \partial h^k(x^k)] = \emptyset$  for all sufficiently large  $k$  due to  $x^k \notin \text{dom } f^k = [\text{dom } g^k \cap \text{dom } h^k]$ . Thus,  $\partial_A f(\bar{x}) = \emptyset$  for any  $\bar{x} \notin \text{dom } f$ .  $\square$

In the subsequent analysis, we restrict our attention to  $\bar{x} \in \text{dom } f$ . Admittedly, the set  $\partial_A f(\bar{x})$  depends on the approximating sequence  $\{f^k\}$  and the DC decomposition of each  $f^k$ , which may contain irrelevant information concerning the local geometry of  $\text{epi } f$ . In fact, for a given ADC function  $f$ , we can make the set  $\partial_A f(\bar{x})$  arbitrarily large by adding the same nonsmooth functions to both  $g^k$  and  $h^k$ . By Attouch's theorem (see for example [30, Theorem 12.35]), for proper, lsc, convex functions  $f$  and  $\{f^k\}$ , if  $f^k \xrightarrow{e} f$ , we immediately have  $\partial_A f = \partial f$  when taking  $g^k = f^k$  and  $h^k = 0$ . In what follows, we further explore the relationships among  $\partial_A f$  and other commonly employed subdifferentials in the literature beyond the convex setting. As it turns out, with respect to an arbitrary DC function  $f^k$  that is lsc,  $\partial_A f(\bar{x})$  contains the limiting subdifferential of  $f$  at any  $\bar{x} \in \text{dom } f$  whenever  $f^k \xrightarrow{e} f$ .

**Theorem 1** (subdifferentials relationships). *Consider an ADC function  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ . The following statements hold for any  $\bar{x} \in \text{dom } f$ .*

- (a) *If  $f$  is e-ADC associated with  $\{f^k\}$  and  $f^k$  is lsc, then  $\partial f(\bar{x}) \subset \partial_A f(\bar{x})$  and  $\partial^\infty f(\bar{x}) \subset \partial_A^\infty f(\bar{x})$ .*
- (b) *If  $f$  is locally Lipschitz continuous and bounded from below, then there exists a sequence of DC functions  $\{f^k\}$  such that  $f^k \xrightarrow{e} f$ ,  $\partial f(\bar{x}) \subset \partial_A f(\bar{x}) \subset \partial_C f(\bar{x})$ , and  $\partial_A^\infty f(\bar{x}) = \{0\}$ . Consequently,  $\text{con } \partial_A f(\bar{x}) = \partial_C f(\bar{x})$ , the set  $\partial_A f(\bar{x})$  is nonempty and bounded, and  $\partial f(\bar{x}) = \partial_A f(\bar{x})$  when  $f$  is subdifferentially regular at  $\bar{x}$ .*

*Proof.* (a) Let  $g^k - h^k$  be a DC decomposition of  $f^k$ . Since  $f$  is e-ADC, it must be lsc [30, Proposition 7.4(a)]. Using epi-convergence of  $\{f^k\}$  to  $f$ , we know from [30, corollary 8.47(b)] and [30, Proposition 8.46(e)] that any element of  $\partial f(\bar{x})$  can be generated as a limit of regular subgradients at  $x^k$  with  $x^k \rightarrow_N \bar{x}$  and  $f^k(x^k) \rightarrow_N f(\bar{x})$  for some  $N \in \mathbb{N}_\infty$ . Indeed, we can further restrict  $x^k \in \text{dom } f^k$  since  $f^k(x^k) \rightarrow_N f(\bar{x})$  and  $\bar{x} \in \text{dom } f$ . Then, we have

$$\partial f(\bar{x}) \subset \bigcup_{x^k \in \text{dom } f^k \rightarrow \bar{x}} \text{Lim sup}_{k \rightarrow +\infty} \widehat{\partial} f^k(x^k) \subset \bigcup_{x^k \in \text{dom } f^k \rightarrow \bar{x}} \text{Lim sup}_{k \rightarrow +\infty} [\partial g^k(x^k) - \partial h^k(x^k)] \subset \partial_A f(\bar{x}),$$

where the second inclusion can be verified as follows: Firstly, due to the lower semicontinuity of  $f^k$  and  $h^k$ , and  $x^k \in \text{dom } f^k \subset \text{dom } g^k$ , it follows from the sum rule of regular subdifferentials [30, corollary 10.9] that  $\widehat{\partial} g^k(x^k) \supset \widehat{\partial} f^k(x^k) + \widehat{\partial} h^k(x^k)$ . Consequently,  $\widehat{\partial} f^k(x^k) \subset \widehat{\partial} g^k(x^k) - \widehat{\partial} h^k(x^k) = \partial g^k(x^k) - \partial h^k(x^k)$  since  $g^k$  and  $h^k$  are proper and convex [30, Proposition 8.12]. Similarly, by [30, corollary 8.47(b)], we have

$$\partial^\infty f(\bar{x}) \subset \bigcup_{x^k \in \text{dom } f^k \rightarrow \bar{x}} \text{Lim sup}^\infty \widehat{\partial} f^k(x^k) \subset \bigcup_{x^k \in \text{dom } f^k \rightarrow \bar{x}} \text{Lim sup}^\infty [\partial g^k(x^k) - \partial h^k(x^k)] \subset \partial_A^\infty f(\bar{x}).$$

(b) For a locally Lipschitz continuous function  $f$ , consider its Moreau envelope  $e_{\gamma} f(x) \triangleq \inf_z \{f(z) + \|z - x\|^2 / (2\gamma)\}$  and the set-valued mapping  $P_{\gamma} f(x) \triangleq \text{argmin}_z \{f(z) + \|z - x\|^2 / (2\gamma)\}$ . For any sequence  $\gamma_k \downarrow 0$ , we demonstrate in the following that  $\{f^k \triangleq e_{\gamma_k} f\}$  is the desired sequence of approximating functions. Firstly, since  $f$  is bounded from below, it must be prox-bounded and, thus, each  $f^k$  is continuous and  $f^k(\bar{x}) \uparrow f(\bar{x})$  for all  $\bar{x}$  (cf. [30, Theorem 1.25]). By the continuity of  $f$  and  $f^k$ , we have  $f^k \xrightarrow{c} f$  from [30, Proposition 7.4(c-d)]. It then follows from part (a) that  $\partial f(\bar{x}) \subset \partial_A f(\bar{x})$ . Consider the following DC decomposition of each  $f^k$ :

$$f^k(x) = \underbrace{\frac{\|x\|^2}{2\gamma_k}}_{\triangleq g^k(x)} - \underbrace{\sup_{z \in \mathbb{R}^n} \left\{ -f(z) - \frac{\|z\|^2}{2\gamma_k} + \frac{z^\top x}{\gamma_k} \right\}}_{\triangleq h^k(x)} \quad x \in \mathbb{R}^n.$$

It is clear that  $f(z) + \|z\|^2 / (2\gamma_k) + z^\top x / \gamma_k$  is level-bounded in  $z$  locally uniformly in  $x$ , since for any  $r \in \mathbb{R}$  and any bounded set  $X \subset \mathbb{R}^n$ , the set

$$\left\{ z \in \mathbb{R}^n \mid x \in X, f(z) + \frac{\|z\|^2}{2\gamma_k} - \frac{z^\top x}{\gamma_k} \leq r \right\} \subset \left\{ z \in \mathbb{R}^n \mid x \in X, \|z - x\|^2 \leq \|x\|^2 + 2\gamma_k \left[ r - \inf_z f(z) \right] \right\}$$

is bounded. Due to the level-boundedness condition, we can apply the subdifferential formula of the parametric minimization [30, Theorem 10.13] to get

$$\partial(-h^k)(x) \subset \bigcup_{z \in P_{\gamma_k} f(x)} \left\{ y \mid (0, y) \in \partial_{(z,x)} \left( f(z) + \frac{\|z\|^2}{2\gamma_k} - \frac{z^\top x}{\gamma_k} \right) \right\} \subset \bigcup_{z \in P_{\gamma_k} f(x)} \left\{ \partial f(z) - \frac{x}{\gamma_k} \right\},$$

where the last inclusion is due to the calculus rules [30, Proposition 10.5 and exercise 8.8(c)]. Since  $h^k$  is convex, we have  $-\partial h^k(x) = \partial_C(-h^k)(x) = \text{con } \partial(-h^k)(x)$  by [30, Theorem 9.61], which further yields that

$$\left[ \partial g^k(x) - \partial h^k(x) \right] \subset \text{con } \bigcup \{ \partial f(z) \mid z \in P_{\gamma_k} f(x) \} \quad \forall x \in \mathbb{R}^n, k \in \mathbb{N}. \quad (4)$$

For any  $x^k \rightarrow \bar{x}$  and any  $z^k \in P_{\gamma_k f}(x^k)$ , we have

$$\frac{1}{2\gamma_k} \|z^k - x^k\|^2 + \inf_x f(x) \leq \frac{1}{2\gamma_k} \|z^k - x^k\|^2 + f(z^k) \leq \frac{1}{2\gamma_k} \|\bar{x} - x^k\|^2 + f(\bar{x}).$$

Then,  $\|z^k - x^k\| \leq \sqrt{\|\bar{x} - x^k\|^2 + 2\gamma_k[f(\bar{x}) - \inf_x f(x)]} \rightarrow 0$  due to the assumption that  $f$  is bounded from below and therefore  $z^k \rightarrow \bar{x}$ . By the local Lipschitz continuity of  $f$ , it follows from [30, Theorem 9.13] that the mapping  $\partial f : x \mapsto \partial f(x)$  is locally bounded at  $\bar{x}$ . Thus, there is a bounded set  $S$  such that  $\bigcup\{\partial f(z^k) \mid z^k \in P_{\gamma_k f}(x^k)\} \subset S$  for all sufficiently large  $k$ . It follows directly from [30, Example 4.22] and the definition of the approximate horizon subdifferential that  $\partial_A^\infty f(\bar{x}) = \{0\}$ .

Next, we will prove  $\partial_A f(\bar{x}) \subset \partial_C f(\bar{x})$ . For any  $u \in \partial_A f(\bar{x})$ , from (4), there exist sequences of vectors  $x^k \rightarrow \bar{x}$  and  $u^k \rightarrow u$  with each  $u^k$  taken from the convex hull of a bounded set  $\bigcup\{\partial f(z^k) \mid z^k \in P_{\gamma_k f}(x^k)\}$ . By Carathéodory's Theorem (see, e.g. [27, Theorem 17.1]), for each  $k$ , we have  $u^k = \sum_{i=1}^{n+1} \lambda_{k,i} v^{k,i}$  for some nonnegative scalars  $\{\lambda_{k,i}\}_{i=1}^{n+1}$  with  $\sum_{i=1}^{n+1} \lambda_{k,i} = 1$  and a sequence  $\{v^{k,i} \in \partial f(z^{k,i})\}_{i=1}^{n+1}$  with  $\{z^{k,i} \in P_{\gamma_k f}(x^k)\}_{i=1}^{n+1}$ . It is easy to see that the sequences  $\{\lambda_{k,i}\}_{k \in \mathbb{N}}$  and  $\{v^{k,i}\}_{k \in \mathbb{N}}$  are bounded for each  $i$ . We can then obtain convergent subsequences  $\lambda_{k,i} \rightarrow_N \bar{\lambda}_i \geq 0$  with  $\sum_{i=1}^{n+1} \bar{\lambda}_i = 1$  and  $v^{k,i} \rightarrow_N \bar{v}^i$  for each  $i$ . Since  $z^{k,i} \rightarrow \bar{x}$ , we have  $\bar{v}^i \in \partial f(\bar{x})$  by using the outer semicontinuity of  $\partial f$ . Thus,  $u^k \rightarrow_N u = \sum_{i=1}^{n+1} \bar{\lambda}_i \bar{v}^i \in \text{con } \partial f(\bar{x}) = \partial_C f(\bar{x})$ . This implies that  $\partial_A f(\bar{x}) \subset \partial_C f(\bar{x})$ . The rest of the statements in (b) follows from the fact that  $\partial_C f(\bar{x})$  is nonempty and bounded whenever  $f$  is locally Lipschitz continuous [30, Theorem 9.61].  $\square$

Under suitable assumptions, Theorem 1(b) guarantees the existence of an ADC decomposition that has its approximate subdifferential contained in the Clarke subdifferential of the original function. Notably, this decomposition may not always be practically useful due to the necessity of computing the Moreau envelope for a generally nonconvex function. Another noteworthy remark is that the assumptions and results of Theorem 1 can be localized to any specific point  $\bar{x}$ . This can be accomplished by defining a notion of ‘‘local epi-convergence’’ at  $\bar{x}$  and extending the result of [30, corollary 8.47] accordingly.

### 2.3 Examples of ADC functions.

In this subsection, we provide examples of ADC functions, including functions that are discontinuous relative to their domains, with explicit and computationally tractable approximating sequences. Moreover, we undertake an investigation into the approximate subdifferentials of these ADC functions.

**Example 2.1** (implicitly convex-concave functions). The concept of *implicitly convex-concave (icc) functions* is introduced in the monograph [13], and is further generalized to extended-real-valued functions in [20]. A proper function  $f : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$  is icc if there exists a lifted function  $\bar{f} : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$  such that the following three conditions hold:

- (i)  $\bar{f}(z, x) = +\infty$  if  $z \notin \text{dom } f, x \in \mathbb{R}^n$ , and  $\bar{f}(z, x) = -\infty$  if  $z \in \text{dom } f, x \notin \text{dom } f$ ;
- (ii)  $\bar{f}(\cdot, x)$  is convex for any fixed  $x \in \text{dom } f$ , and  $\bar{f}(z, \cdot)$  is concave for any fixed  $z \in \text{dom } f$ ;
- (iii)  $f(x) = \bar{f}(x, x)$  for any  $x \in \text{dom } f$ .



A notable example of icc functions is the optimal value function  $f_p$  in (1), which is associated with the lifted function defined by (the subscripts/superscripts  $p$  are omitted for brevity):

$$\bar{f}(z, x) \triangleq \inf_{y \in \mathbb{R}^d} \left\{ (c + Cx)^\top y + \frac{1}{2} y^\top Q y \mid Az + By \leq b \right\} \quad (x, z) \in \text{dom } f \times \text{dom } f. \quad (5)$$

Let  $\partial_1 \bar{f}(\cdot, x)$  and  $\partial_2(-\bar{f})(z, \cdot)$  denote the subdifferentials of the convex functions  $\bar{f}(\cdot, x)$  and  $(-\bar{f})(z, \cdot)$ , respectively, for any  $(x, z) \in \text{dom } f \times \text{dom } f$ . For any  $\gamma > 0$ , the *partial Moreau envelope* of an icc function  $f$  associated with  $\bar{f}$  is given by

$$\inf_{z \in \mathbb{R}^n} \left\{ \bar{f}(z, x) + \frac{1}{2\gamma} \|z - x\|^2 \right\} = \underbrace{\frac{\|x\|^2}{2\gamma}}_{\triangleq g_\gamma(x)} - \underbrace{\sup_{z \in \mathbb{R}^n} \left\{ -\bar{f}(z, x) - \frac{\|z\|^2}{2\gamma} + \frac{z^\top x}{\gamma} \right\}}_{\triangleq h_\gamma(x)} \quad x \in \text{dom } f. \quad (6)$$

This decomposition, established in [20], offers computational advantages compared to the standard Moreau envelope, as the maximization problem defining  $h_\gamma$  is concave in  $z$  for any fixed  $x$ . In what follows, we present new results on the conditions under which the icc function  $f$  is e-ADC and c-ADC based on the partial Moreau envelope. Additionally, we explore a relationship between  $\partial_A f(\bar{x})$  and  $\partial_1 \bar{f}(\bar{x}, \bar{x}) - \partial_2(-\bar{f})(\bar{x}, \bar{x})$ , where the latter is known to be an outer estimate of  $\partial_C f(\bar{x})$  [13, Proposition 4.4.26]. The proof is deferred to Appendix A.

**Proposition 2.** *Let  $f : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$  be a proper, lsc, icc function associated with  $\bar{f}$ , where  $\text{dom } f$  is closed and  $\bar{f}$  is lsc on  $\mathbb{R}^n \times \text{dom } f$ , bounded below on  $\text{dom } f \times \text{dom } f$ , and continuous relative to  $\text{int}(\text{dom } f) \times \text{int}(\text{dom } f)$ . Given a sequence of scalars  $\gamma_k \downarrow 0$ , we have:*

- (a)  *$f$  is e-ADC associated with  $\{f^k\}$ , where each  $f^k(x) \triangleq g_{\gamma_k}(x) - h_{\gamma_k}(x) + \delta_{\text{dom } f}(x)$ . In addition, if  $\text{dom } f = \mathbb{R}^n$ , then  $f$  is c-ADC associated with  $\{f^k\}$ .*
- (b)  *$\partial_A f(\bar{x}) \subset \partial_1 \bar{f}(\bar{x}, \bar{x}) - \partial_2(-\bar{f})(\bar{x}, \bar{x})$  and  $\partial_A^\infty f(\bar{x}) = \{0\}$  for any  $\bar{x} \in \text{int}(\text{dom } f)$ .*

**Example 2.2** (VaR for continuous random variables). Given a continuous random variable  $Y : \Omega \rightarrow \mathbb{R}$ , its conditional value-at-risk (CVaR) at a confidence level  $\alpha \in (0, 1)$  is defined as  $\text{CVaR}_\alpha(Y) \triangleq \mathbb{E}[Y \mid Y \geq \text{VaR}_\alpha(Y)]$ , where  $\text{VaR}_\alpha$  is the value-at-risk given in Example 1.2 (see, e.g., [29]). For any  $\alpha \in (0, 1)$  and  $k > 1/\alpha$ , we define

$$g^k(x) \triangleq [k(1 - \alpha) + 1] \text{CVaR}_{\alpha - 1/k}[c(x, Z)], \quad h^k(x) \triangleq k(1 - \alpha) \text{CVaR}_\alpha[c(x, Z)] \quad x \in \mathbb{R}^n. \quad (7)$$

The following properties of VaR for continuous random variables hold, with proofs provided in Appendix A.

**Proposition 3.** *Let  $c : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$  be a lsc function and  $Z : \Omega \rightarrow \mathbb{R}^m$  be a random vector. Suppose that  $c(\cdot, z)$  is convex for any fixed  $z \in \mathbb{R}^m$ , and  $c(x, Z)$  is a random variable having a continuous distribution induced by that of  $Z$  for any fixed  $x \in \mathbb{R}^n$ . Additionally, assume that  $\mathbb{E}[|c(x, Z)|] < +\infty$  for any  $x \in \mathbb{R}^n$ . For any given constant  $\alpha \in (0, 1)$ , the following properties hold.*

- (a)  *$\text{VaR}_\alpha[c(\cdot, Z)]$  is lsc and e-ADC associated with  $\{g^k - h^k\}$  (with the definitions of  $g^k$  and  $h^k$  in (7)). Additionally, if  $c(\cdot, \cdot)$  is continuous, then  $\text{VaR}_\alpha[c(\cdot, Z)]$  is continuous and c-ADC associated with  $\{g^k - h^k\}$ .*

(b) If there exists a measurable function  $\kappa : \mathbb{R}^m \rightarrow \mathbb{R}_+$  such that  $\mathbb{E}[\kappa(Z)] < +\infty$  and  $|c(x, z) - c(x', z)| \leq \kappa(z)\|x - x'\|$  for all  $x, x' \in \mathbb{R}^n$  and  $z \in \mathbb{R}^m$ , then for any  $\bar{x} \in \mathbb{R}^n$ ,

$$\partial_A \text{VaR}_\alpha[c(\cdot, Z)](\bar{x}) = \bigcup_{x^k \rightarrow \bar{x}} \limsup_{k \rightarrow +\infty} \mathbb{E} \left[ \partial_1 c(x^k, Z) \mid \text{VaR}_{\alpha-1/k}[c(x^k, Z)] < c(x^k, Z) < \text{VaR}_\alpha[c(x^k, Z)] \right],$$

where  $\mathbb{E}[\mathcal{A}(Z) \mid B]$  for a random set-valued mapping  $\mathcal{A}$  and an event  $B$  is defined as the set of conditional expectations  $\mathbb{E}[a(Z) \mid B]$  for all measurable selections  $a(Z) \in \mathcal{A}(Z)$ .

### 3 The convex composite ADC functions and minimization.

This section aims to derive necessary optimality conditions for  $(\text{CP}_0)$ , particularly focusing on the inner function  $f_p$  that lacks local Lipschitz continuity. Throughout the rest of this paper, we assume that  $\varphi_p : \mathbb{R} \rightarrow \mathbb{R} \cup \{+\infty\}$  is proper, convex, lsc and  $f_p : \mathbb{R}^n \rightarrow \mathbb{R}$  is real-valued for all  $p = 1, \dots, m$ . Depending on whether  $\varphi_p$  is nondecreasing or not, we partition  $\{1, \dots, m\}$  into two categories:

$$I_1 \triangleq \{p \in \{1, \dots, m\} \mid \varphi_p \text{ nondecreasing}\} \quad \text{and} \quad I_2 \triangleq \{1, \dots, m\} \setminus I_1. \quad (8)$$

We do not specifically address the case where  $\varphi_p$  is nonincreasing, as one can always redefine  $\tilde{\varphi}_p(t) = \varphi_p(-t)$  and  $\tilde{f}_p(x) = -f_p(x)$ , enabling the treatment of these indices in the same manner as those in  $I_1$ . Therefore, the set  $I_2$  should be viewed as the collection of indices  $p$  where  $\varphi_p$  is not monotone. We further make the following assumptions on the functions  $\varphi_p$  and  $f_p$ .

**Assumption 1** For each  $p$ , we have

- (a)  $f_p$  is e-ADC associated with  $\{f_p^k = g_p^k - h_p^k\}_{k \in \mathbb{N}}$ , and  $\text{dom } g_p^k = \text{dom } h_p^k = \mathbb{R}^n$ ;
- (b)  $-\infty < \liminf_{x' \rightarrow x, k \rightarrow +\infty} f_p^k(x') \leq \limsup_{x' \rightarrow x, k \rightarrow +\infty} f_p^k(x') < +\infty$  for all  $x \in \mathbb{R}^n$ ;
- (c)  $[F_p^k \triangleq \varphi_p \circ f_p^k] \xrightarrow{e} F_p$ .

From Assumption 1(a), each  $f_p^k$  is locally Lipschitz continuous since any real-valued convex function is locally Lipschitz continuous. Obviously,  $f_p^k \xrightarrow{c} f_p$  is sufficient for Assumption 1(b) to hold. Since  $f_p^k \xrightarrow{e} f_p$ , we have  $\liminf_{x' \rightarrow x, k \rightarrow +\infty} f_p^k(x') \geq f_p(x) > -\infty$  for each  $p$  at any  $x \in \mathbb{R}^n$ . However,  $\limsup_{x' \rightarrow x, k \rightarrow +\infty} f_p^k(x') < +\infty$  does not hold trivially. For example, consider a continuous function  $f$  and

$$f^k(x) = \begin{cases} f(x) + k^2x + k & \text{if } x \in [-1/k, 0] \\ f(x) - k^2x + k & \text{if } x \in (0, 1/k] \\ f(x) & \text{otherwise} \end{cases},$$

which results in  $f^k \xrightarrow{c} f$  but  $\limsup_{k \rightarrow +\infty} f^k(0) = +\infty$ . Additionally, Assumption 1(b) ensures that at each point  $x$  and for any sequence  $x^k \rightarrow x$ , the sequence  $\{f_p^k(x^k)\}_{k \in \mathbb{N}}$  must be bounded.

It follows from [30, Exercise 7.8(c)] and [32, Theorem 2.4] that there are several sufficient conditions for Assumption 1(c) to hold, which differ based on the monotonicity of each  $\varphi_p$ : (i)

For  $p \in I_1$ , either  $\varphi_p$  is real-valued or  $f_p^k \leq f_p$ ; (ii) For  $p \in I_2$ ,  $f_p$  is c-ADC and for all  $x$  with  $f_p(x) \in \text{bdry}(\text{dom } \varphi_p)$ , there exists a sequence  $x^k \rightarrow x$  with  $f(x^k) \in \text{int}(\text{dom } \varphi_p)$ . In addition, according to [30, Proposition 7.4(a)], Assumption 1(c) implies that  $F_p = \varphi_p \circ f_p$  is lsc. We also note that Assumption 1(c) doesn't necessarily imply  $\sum_{p=1}^m F_p^k \xrightarrow{e} \sum_{p=1}^m F_p$ . To maintain epi-convergence under addition of functions, one may refer to the sufficient conditions in [30, Theorem 7.46].

### 3.1 Asymptotic stationarity under epi-convergence.

In this subsection, we introduce a novel stationarity concept for problem  $(\text{CP}_0)$ , grounded in a monotonic decomposition of univariate convex functions. We demonstrate that under certain constraint qualifications, epi-convergence of approximating functions ensures this stationarity concept as a necessary optimality condition. Alongside the known fact that epi-convergence also ensures the consistency of global optimal solutions [30, Theorem 7.31(b)], this highlights the usefulness of epi-convergence as a tool for studying the approximation of problem  $(\text{CP}_0)$ .

The following lemma is an extension of [13, Lemma 6.1.1] from real-valued univariate convex functions to extended-real-valued univariate convex functions.

**Lemma 1** (a monotonic decomposition of univariate convex functions). *Let  $\varphi : \mathbb{R} \rightarrow \overline{\mathbb{R}}$  be a proper, lsc and convex function. Then there exist a proper, lsc, convex and nondecreasing function  $\varphi^\uparrow$ , as well as a proper, lsc, convex and nonincreasing function  $\varphi^\downarrow$ , such that  $\varphi = \varphi^\uparrow + \varphi^\downarrow$ . In addition, if  $\text{int}(\text{dom } \varphi) \neq \emptyset$ , then  $\partial\varphi(z) = \partial\varphi^\uparrow(z) + \partial\varphi^\downarrow(z)$  for any  $z \in \text{dom } \varphi$ .*

*Proof.* From the convexity of  $\varphi$ ,  $\text{dom } \varphi$  is an interval on  $\mathbb{R}$ , possibly unbounded. In fact, we can explicitly construct  $\varphi^\uparrow$  and  $\varphi^\downarrow$  in following two cases.

Case 1. If  $\varphi$  has no direction of recession, i.e., there does not exist  $d \neq 0$  such that for any  $z$ ,  $\varphi(z + \lambda d)$  is a nonincreasing function of  $\lambda > 0$ , it follows from [27, Theorem 27.2] that  $\varphi$  attains its minimum at some  $z^* \in \text{dom } \varphi$ . Define

$$\varphi^\uparrow(z) = \begin{cases} \varphi(z^*) & \text{if } z \leq z^* \\ \varphi(z) & \text{if } z > z^* \end{cases} \quad \text{and} \quad \varphi^\downarrow(z) = \begin{cases} \varphi(z) - \varphi(z^*) & \text{if } z \leq z^* \\ 0 & \text{if } z > z^* \end{cases}.$$

Observe that  $\emptyset \neq \text{int}(\text{dom } \varphi) \subset [\text{int}(\text{dom } \varphi^\uparrow) \cap \text{int}(\text{dom } \varphi^\downarrow)]$ . Consequently, from [27, Theorem 23.8], we have  $\partial\varphi(z) = \partial\varphi^\uparrow(z) + \partial\varphi^\downarrow(z)$  for any  $z \in \mathbb{R}$ .

Case 2. Otherwise, there exists  $d \neq 0$  such that for any  $z \in \mathbb{R}$ ,  $\varphi(z + \lambda d)$  is a nonincreasing function of  $\lambda > 0$ . Consequently,  $\text{dom } \varphi$  must be an unbounded interval on  $\mathbb{R}$ . Let  $d = 1$  (or  $-1$ ) be such a recession direction, then  $\varphi$  is nonincreasing (or nondecreasing) on  $\mathbb{R}$ . We can set  $\varphi^\uparrow = 0$  and  $\varphi^\downarrow = \varphi$  (or  $\varphi^\uparrow = \varphi$  and  $\varphi^\downarrow = 0$ ). In this case, it is obvious that  $\partial\varphi(z) = \partial\varphi^\uparrow(z) + \partial\varphi^\downarrow(z)$  for any  $z \in \mathbb{R}$ . The proof is thus completed.  $\square$

In the subsequent analysis, we use  $\varphi^\uparrow$  and  $\varphi^\downarrow$  to denote the monotonic decomposition of any univariate, proper, lsc, and convex function  $\varphi$  constructed in the proof of Lemma 1 and, in particular, we take  $\varphi^\downarrow = 0$  whenever  $\varphi$  is nondecreasing. We are now ready to present the definition of asymptotically stationary points.

**Definition 4** (asymptotically stationary points). *Let each  $f_p$  be an ADC function associated with  $\{f_p^k = g_p^k - h_p^k\}_{k \in \mathbb{N}}$ . For each  $p$ , define*

$$T_p(x) \triangleq \left\{ t_p \mid \exists N \in \mathbb{N}_\infty^\#, x^k \rightarrow x \text{ with } f_p^k(x^k) \rightarrow_N t_p \right\} \quad x \in \mathbb{R}^n. \quad (9)$$

We say that  $\bar{x}$  is an asymptotically stationary (A-stationary) point of problem (CP<sub>0</sub>) if for each  $p$ , there exists  $y_p \in \bigcup\{\partial\varphi_p(t_p) \mid t_p \in T_p(\bar{x})\}$  such that

$$0 \in \sum_{p=1}^m (\{y_p \partial_A f_p(\bar{x})\} \cup [\pm \partial_A^\infty f_p(\bar{x}) \setminus \{0\}]). \quad (10)$$

We say that  $\bar{x}$  is a weakly asymptotically stationary (weakly A-stationary) point of problem (CP<sub>0</sub>) if for each  $p$ , there exist  $\bar{t}_p \in T_p(\bar{x})$ ,  $y_{p,1} \in \partial\varphi_p^\uparrow(\bar{t}_p)$  and  $y_{p,2} \in \partial\varphi_p^\downarrow(\bar{t}_p)$  such that

$$0 \in \sum_{p=1}^m (\{y_{p,1} \partial_A f_p(\bar{x}) + y_{p,2} \partial_A f_p(\bar{x})\} \cup [\pm \partial_A^\infty f_p(\bar{x}) \setminus \{0\}]).$$

**Remark 1.** (i) Given that the approximate subdifferential  $\partial_A f_p$  is determined by the approximating sequence  $\{f_p^k\}_{k \in \mathbb{N}}$  and their corresponding DC decompositions, the notion of (weak) A-stationarity also depends on these sequences and decompositions. (ii) It follows directly from Lemma 1 that an A-stationary point must be a weakly A-stationary point if  $\text{int}(\text{dom } \varphi_p) \neq \emptyset$  for each  $p = 1, \dots, m$ . (iii) When each  $\varphi_p$  is nondecreasing or nonincreasing, the concepts of weak A-stationarity and A-stationarity coincide. (iv) Given a point  $\bar{x}$ , we can rewrite (10) as

$$0 \in \sum_{p \in I} [\pm \partial_A^\infty f_p(\bar{x}) \setminus \{0\}] + \sum_{p \in \{1, \dots, m\} \setminus I} \{y_p \partial_A f_p(\bar{x})\}$$

for some index set  $I \subset \{1, \dots, m\}$  that is potentially empty. For each  $p \in I$ , although the scalar  $y_p$  does not explicitly appear in this inclusion, its existence implies that  $\bigcup\{\partial\varphi_p(t_p) \mid t_p \in T_p(\bar{x})\} \neq \emptyset$ , which plays a role in ensuring  $\bar{x} \in \text{dom}(\varphi_p \circ f_p)$ . For instance, if  $f_p^k \xrightarrow{c} f_p$  for some  $p \in I$ , then  $T_p(\bar{x}) = \{f_p(\bar{x})\}$ , and the existence of  $y_p \in \bigcup\{\partial\varphi_p(t_p) \mid t_p \in T_p(\bar{x})\} = \partial\varphi_p(f_p(\bar{x}))$  yields  $\bar{x} \in \text{dom}(\varphi_p \circ f_p)$ .

In the following, we take a detour to compare the A-stationarity with the stationarity defined in [32], where the author has focused on a more general composite problem

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \varphi(f(x)),$$

where  $\varphi : \mathbb{R}^m \rightarrow \bar{\mathbb{R}}$  is proper, lsc, convex and  $f \triangleq (f_1, \dots, f_m) : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is a locally Lipschitz continuous mapping. Consider the special case where  $\varphi(z) = \sum_{p=1}^m \varphi_p(z_p)$  with  $z = (z_1, \dots, z_m)$ . Under this setting, a vector  $\bar{x}$  is called a stationary point in [32] if there exist  $\bar{y}$  and  $\bar{z}$  such that

$$0 \in S(\bar{x}, \bar{y}, \bar{z}) \triangleq \{(f_1(\bar{x}), \dots, f_m(\bar{x})) - \bar{z}\} \times \{\partial\varphi_1(\bar{z}_1) \times \dots \times \partial\varphi_m(\bar{z}_m) - \bar{y}\} \times \left( \sum_{p=1}^m \bar{y}_p \partial_C f_p(\bar{x}) \right), \quad (11)$$

which can be equivalently written as

$$0 \in \sum_{p=1}^m \bar{y}_p \partial_C f_p(\bar{x}) \quad \text{for some } \bar{y}_p \in \partial\varphi_p(f_p(\bar{x})) \quad p = 1, \dots, m. \quad (12)$$

For any fixed  $k \in \mathbb{N}$ , a surrogate set-valued mapping  $S^k$  can be defined similarly as  $S$  in (11) by substituting  $f_p$  and  $\varphi_p$  with  $f_p^k$  and  $\varphi_p^k$  for each  $p$ . The cited paper provides sufficient conditions to ensure  $\text{Lim sup}_{k \rightarrow +\infty} (\text{gph } S^k) \subset \text{gph } S$ , which asserts that any accumulation point  $(\bar{x}, \bar{y}, \bar{z})$  of a sequence  $\{(x^k, y^k, z^k)\}$  with  $0 \in S^k(x^k, y^k, z^k)$  yields a stationary point  $\bar{x}$ . Our study on the asymptotic stationarity differs from [32] in the following aspects:

1. Our outer convex function  $\varphi$  is assumed to have the separable form  $\sum_{p=1}^m \varphi_p$ , while [32] allows a general proper, lsc, convex function. In addition, each  $\varphi_p$  is fixed in our approximating problem while [32] considers a sequence of convex functions  $\{\varphi_p^k\}_{k \in \mathbb{N}}$  that epi-converges to  $\varphi_p$ .
2. We do not require the inner function  $f_p$  to be locally Lipschitz continuous.

If each  $f_p$  is locally Lipschitz continuous and bounded from below, it then follows from Theorem 1 that  $f_p$  is c-ADC associated with  $\{f_p^k = g_p^k - h_p^k\}_{k \in \mathbb{N}}$  such that  $\partial f_p(x) \subset \partial_A f_p(x) \subset \partial_C f_p(x)$  and  $\partial_A^\infty f_p(x) = \{0\}$  for any  $x$ . Moreover, by  $f_p^k \xrightarrow{c} f_p$ , one has  $T_p(x) = \{f_p(x)\}$ . Thus, for any A-stationary point  $\bar{x}$  induced by these ADC decompositions, there exists  $\bar{y}_p \in \partial \varphi_p(f_p(\bar{x}))$  for each  $p$  such that

$$0 \in \sum_{p=1}^m \{\bar{y}_p \partial_A f_p(\bar{x})\} \subset \sum_{p=1}^m \{\bar{y}_p \partial_C f_p(\bar{x})\}. \quad (13)$$

Hence,  $\bar{x}$  is also a stationary point defined in (12). Indeed, A-stationarity here can be sharper than the latter one as the last inclusion in (13) may not hold with equality.

When  $f_p$  fails to be locally Lipschitz continuous for some  $p$ , it is not known if (11) is still a necessary condition for a local solution of  $(\text{CP}_0)$ . This situation further complicates the fulfillment of conditions outlined in [32, Theorem 2.4], especially the requirement of  $f_p^k \xrightarrow{c} f_p$ , due to the potential discontinuity of  $f_p$ . As will be shown in Theorem 2 below, despite these challenges, weak A-stationarity continues to be a necessary optimality condition under Assumption 1.

To proceed, for each  $p$  and any  $x \in \text{dom}(\varphi_p \circ f_p)$ , we define  $S_p(x)$  to be a collection of sequences:

$$S_p(x) \triangleq \left\{ \{x_p^k\}_{k \in \mathbb{N}} \mid x_p^k \rightarrow x \text{ with } \varphi_p(f_p^k(x_p^k)) \rightarrow \varphi_p(f_p(x)) \right\}. \quad (14)$$

**Theorem 2** (necessary conditions for optimality). *Let  $\bar{x} \in \bigcap_{p=1}^m \text{dom } F_p$  be a local minimizer of problem  $(\text{CP}_0)$ . Suppose that Assumption 1 and the following two conditions hold:*

(i) *For each  $p$  and any sequence  $\{x_p^k\}_{k \in \mathbb{N}} \in S_p(\bar{x})$ , there is a positive integer  $K$  such that*

$$0 \notin \partial_C f_p^k(x_p^k) \quad \text{or} \quad \mathcal{N}_{\text{dom } \varphi_p}(f_p^k(x_p^k)) = \{0\} \quad \forall k \geq K, \quad (15)$$

and

$$\left[ 0 \in y_p \partial_A f_p(\bar{x}), y_p \in \bigcup \{ \mathcal{N}_{\text{dom } \varphi_p}(t_p) \mid t_p \in T_p(\bar{x}) \} \right] \implies y_p = 0, \quad p = 1, \dots, m. \quad (16)$$

(ii) *One has*

$$\left[ \sum_{p=1}^m w_p = 0, w_p \in \partial^\infty(\varphi_p \circ f_p)(\bar{x}) \right] \implies w_1 = \dots = w_m = 0. \quad (17)$$

Then  $\bar{x}$  is an  $A$ -stationary point of  $(\text{CP}_0)$ . Additionally,  $\bar{x}$  is a weakly  $A$ -stationary point of  $(\text{CP}_0)$  if  $\text{int}(\text{dom } \varphi_p) \neq \emptyset$  for each  $p = 1, \dots, m$ .

*Proof.* By using Fermat's rule [30, Theorem 10.1] and the sum rule of the limiting subdifferentials [30, Corollary 10.9] due to the condition (17), we have

$$\begin{aligned}
0 \in \partial \left[ \sum_{p=1}^m (\varphi_p \circ f_p)(\bar{x}) \right] &\subset \sum_{p=1}^m \partial(\varphi_p \circ f_p)(\bar{x}) \stackrel{(i)}{\subset} \sum_{p=1}^m \bigcup_{\{x_p^k\}_{k \in \mathbb{N}} \in S_p(\bar{x})} \text{Lim sup}_{k \rightarrow +\infty} \partial(\varphi_p \circ f_p^k)(x_p^k) \\
&\stackrel{(ii)}{\subset} \sum_{p=1}^m \bigcup_{\{x_p^k\}_{k \in \mathbb{N}} \in S_p(\bar{x})} \text{Lim sup}_{k \rightarrow +\infty} \bigcup \left\{ \partial(y_p^k f_p^k)(x_p^k) \mid y_p^k \in \partial\varphi_p(f_p^k(x_p^k)) \right\} \\
&\stackrel{(iii)}{\subset} \sum_{p=1}^m \bigcup_{\{x_p^k\}_{k \in \mathbb{N}} \in S_p(\bar{x})} \text{Lim sup}_{k \rightarrow +\infty} \left\{ y_p^k v_p^k \mid y_p^k \in \partial\varphi_p(f_p^k(x_p^k)), v_p^k \in \partial_C f_p^k(x_p^k) \right\} \\
&\stackrel{(iv)}{\subset} \sum_{p=1}^m \bigcup_{\{x_p^k\}_{k \in \mathbb{N}} \in S_p(\bar{x})} \text{Lim sup}_{k \rightarrow +\infty} \left\{ y_p^k v_p^k \mid y_p^k \in \partial\varphi_p(f_p^k(x_p^k)), v_p^k \in [\partial g_p^k(x_p^k) - \partial h_p^k(x_p^k)] \right\}.
\end{aligned} \tag{18}$$

The inclusion (i) is due to  $\varphi_p \circ f_p^k \xrightarrow{e} \varphi_p \circ f_p$  in Assumption 1(c) and approximation of subgradients under epi-convergence [30, corollary 8.47] and [30, Proposition 8.46(e)]; (ii) follows from the nonsmooth Lagrange multiplier rule [30, Exercise 10.52] due to the local Lipschitz continuity of  $f_p^k$  [30, Example 9.14] and the condition (15); (iii) and (iv) use the calculus rules of the Clarke subdifferential [12, Chapter 2.3]. For each  $p$ , any sequence  $\{x_p^k\}_{k \in \mathbb{N}} \in S_p(\bar{x})$  and any element

$$\bar{w}_p \in \text{Lim sup}_{k \rightarrow +\infty} \left\{ y_p^k v_p^k \mid y_p^k \in \partial\varphi_p(f_p^k(x_p^k)), v_p^k \in [\partial g_p^k(x_p^k) - \partial h_p^k(x_p^k)] \right\},$$

there is a subsequence  $w_p^k \rightarrow_N \bar{w}_p$  with  $w_p^k = y_p^k v_p^k$  for some  $N \in \mathbb{N}_\infty^\sharp$ . Next, we show the existence of  $\bar{y}_p \in \bigcup \{ \partial\varphi_p(t_p) \mid t_p \in T_p(\bar{x}) \}$  for each  $p$  such that

$$\bar{w}_p \in \{ \bar{y}_p \partial_A f_p(\bar{x}) \} \cup [ \pm \partial_A^\infty f_p(\bar{x}) \setminus \{0\} ]. \tag{19}$$

By Assumption 1(b), the subsequence  $\{f_p^k(x_p^k)\}_{k \in \mathbb{N}}$  is bounded. Taking a subsequence if necessary, we can suppose that  $f_p^k(x_p^k) \rightarrow_N \bar{z}_p \in T_p(\bar{x})$ . If  $\{y_p^k\}_{k \in \mathbb{N}}$  is unbounded, then  $\{v_p^k\}_{k \in \mathbb{N}}$  has a subsequence converging to 0 and, thus,  $0 \in \partial_A f_p(\bar{x})$ . Additionally, there exists  $\tilde{y}_p \neq 0$  such that

$$\frac{y_p^k}{|y_p^k|} \rightarrow_N \tilde{y}_p \in \text{Lim sup}_{k \in \mathbb{N} \rightarrow +\infty}^\infty \partial\varphi_p(f_p^k(x_p^k)) \stackrel{(v)}{=} \text{Lim sup}_{k \in \mathbb{N} \rightarrow +\infty}^\infty \widehat{\partial}\varphi_p(f_p^k(x_p^k)) \stackrel{(vi)}{\subset} \partial^\infty \varphi_p(\bar{z}_p) \stackrel{(vii)}{=} \mathcal{N}_{\text{dom } \varphi_p}(\bar{z}_p). \tag{20}$$

The equation (v) follows from [30, Proposition 8.12] by the convexity of  $\varphi_p$ . From  $\{x_p^k\}_{k \in \mathbb{N}} \in S_p(\bar{x})$  and  $\bar{x} \in \text{dom } F_p$ , we must have  $f_p^k(x_p^k) \in \text{dom } \varphi_p$  for sufficiently large  $k \in \mathbb{N}$ . Since  $\varphi_p$  is lsc, it holds that  $\varphi_p(\bar{z}_p) \leq \liminf_{k \in \mathbb{N} \rightarrow +\infty} \varphi_p(f_p^k(x_p^k)) = \varphi_p(f_p(\bar{x}))$  and, thus,  $\bar{z}_p \in \text{dom } \varphi_p$ . Also, notice that  $\varphi_p$  is continuous relative to its domain as it is univariate convex and lsc [27, Theorem 10.2]. This continuity implies  $\varphi_p(f_p^k(x_p^k)) \rightarrow_N \varphi_p(\bar{z}_p)$ . The inclusion (vi) follows directly from the definition of the horizon subdifferential. Lastly, (vii) is due to the lower semicontinuity of the proper convex

function  $\varphi$  and [30, Proposition 8.12]. Therefore, we have  $(0 \neq) \tilde{y}_p \in \bigcup \{ \mathcal{N}_{\text{dom } \varphi_p}(t_p) \mid t_p \in T_p(\bar{x}) \}$  with  $0 \in \tilde{y}_p \partial_A f_p(\bar{x})$  due to  $0 \in \partial_A f_p(\bar{x})$ , contradicting (16). So far, we conclude that  $\{y_p^k\}_{k \in N}$  is a bounded sequence. Suppose that  $y_p^k \rightarrow_N \bar{y}_p$  and, thus,  $\bar{y}_p \in \partial \varphi_p(\bar{z}_p)$  by the outer semicontinuity of  $\partial \varphi_p$  [30, Proposition 8.7].

Case 1. If  $\bar{y}_p = 0$ , inclusion (19) holds trivially for  $\bar{w}_p = 0$ , and for  $\bar{w}_p \neq 0$  we can find a subsequence  $\{|y_p^k|\}_{k \in N'} \downarrow 0$  such that  $\{|y_p^k| v_p^k\}_{k \in N'}$  converges to  $\bar{w}_p$  or  $-\bar{w}_p (\neq 0)$  with  $v_p^k \in [\partial g_p^k(x_p^k) - \partial h_p^k(x_p^k)]$  for all  $k \in N'$ . Therefore, (19) follows from

$$\bar{w}_p \in \left[ \left( \pm \text{Lim sup}_{k \rightarrow +\infty}^\infty [\partial g_p^k(x_p^k) - \partial h_p^k(x_p^k)] \right) \setminus \{0\} \right] \subset [\pm \partial_A^\infty f_p(\bar{x}) \setminus \{0\}].$$

Case 2. Otherwise,  $\|v_p^k\| \rightarrow_N \|\bar{w}_p\|/|\bar{y}_p|$ . This means that  $\{v_p^k\}_{k \in N}$  is bounded. Suppose  $v_p^k \rightarrow_N \bar{v}_p$ . Then,  $\bar{v}_p \in \text{Lim sup}_{k \rightarrow +\infty} [\partial g_p^k(x_p^k) - \partial h_p^k(x_p^k)] \subset \partial_A f_p(\bar{x})$ , and (19) is evident from  $\bar{w}_p = \bar{y}_p \bar{v}_p$ .

In either case, we have proved (19). Combining (18) with (19), for some  $\bar{y}_p \in \bigcup \{ \partial \varphi_p(t_p) \mid t_p \in T_p(\bar{x}) \}$ , we know that  $\bar{x}$  is an A-stationary point of  $(\text{CP}_0)$ . The final assertion follows from Remark 1(ii).  $\square$

### 3.2 An example of A-stationarity.

We present an example to illustrate the concept of A-stationarity and to study its relationship with other known optimality conditions.

**Example 3.1** (bi-parametrized two-stage stochastic programs). Consider the following bi-parametrized two-stage stochastic program with fixed scenarios described in [21]:

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \theta(x) + \frac{1}{m_1} \sum_{p=1}^{m_1} f_p(x) \quad \text{subject to} \quad \phi_p(x) \leq 0, \quad p = 1, \dots, m_2, \quad (21)$$

where  $\theta, \phi_p : \mathbb{R}^n \rightarrow \mathbb{R}$  are convex, continuously differentiable for  $p = 1, \dots, m_2$ , and  $f_p$ , as defined in (1), is real-valued for  $p = 1, \dots, m_1$ . At  $x = \bar{x}$ , let  $Y_p(\bar{x})$  and  $\Lambda_p(\bar{x})$  represent the optimal solutions and multipliers for each second-stage problem (1). Suppose that  $Y_p(\bar{x})$  and  $\Lambda_p(\bar{x})$  are bounded. Note that  $\theta$  and  $\phi_p$  are ADC functions since they are convex. Example 2.1 shows that  $f_p$  is an ADC function, and therefore, problem (21) is a specific case of the composite model  $(\text{CP}_0)$ . Given an A-stationary point  $\bar{x}$  of (21), under the assumptions of Example 2.1, we have

$$\begin{aligned} 0 &\in \nabla \theta(\bar{x}) + \frac{1}{m_1} \sum_{p=1}^{m_1} \left( \{ \partial_A f_p(\bar{x}) \} \cup [\pm \partial_A^\infty f_p(\bar{x}) \setminus \{0\}] \right) + \sum_{p=1}^{m_2} \bar{\mu}^{m_1+p} \nabla \phi_p(\bar{x}) \\ &\subset \nabla \theta(\bar{x}) + \frac{1}{m_1} \sum_{p=1}^{m_1} \{ \partial_1 \bar{f}_p(\bar{x}, \bar{x}) - \partial_2(-\bar{f}_p)(\bar{x}, \bar{x}) \} + \sum_{p=1}^{m_2} \bar{\mu}^{m_1+p} \nabla \phi_p(\bar{x}), \end{aligned} \quad (22)$$

where  $\bar{\mu}^{m_1+p} \in \mathcal{N}_{(-\infty, 0]}(\phi_p(\bar{x}))$  for  $p = 1, \dots, m_2$  and  $\bar{f}_p$  is defined in (5) for  $p = 1, \dots, m_1$ . By assumptions, both  $\Lambda_p(\bar{x})$  and  $Y_p(\bar{x})$  are nonempty, bounded, and

$$\Lambda_p(\bar{x}) \times Y_p(\bar{x}) = \left\{ (\bar{y}^p, \bar{\mu}^p) \mid c^p + C^p \bar{x} + Q^p \bar{y}^p + (B^p)^\top \bar{\mu}^p = 0, 0 \leq b^p - A^p \bar{x} - B^p \bar{y}^p \perp \bar{\mu}^p \geq 0 \right\}.$$

It then follows from Danskin's Theorem [11, Theorem 2.1] that

$$\begin{aligned}\partial_1 \bar{f}_p(\bar{x}, \bar{x}) &= \text{con} \left\{ (A^p)^\top \bar{\mu}^p \mid \bar{\mu}^p \in \Lambda_p(\bar{x}) \right\} = \left\{ (A^p)^\top \bar{\mu}^p \mid \bar{\mu}^p \in \Lambda_p(\bar{x}) \right\}, \\ \partial_2(-\bar{f}_p)(\bar{x}, \bar{x}) &= \text{con} \left\{ -(C^p)^\top \bar{y}^p \mid \bar{y}^p \in Y_p(\bar{x}) \right\} = \left\{ -(C^p)^\top \bar{y}^p \mid \bar{y}^p \in Y_p(\bar{x}) \right\}.\end{aligned}$$

Combining these expressions with (22), we obtain

$$\begin{cases} 0 = \nabla \theta(\bar{x}) + \frac{1}{m_1} \sum_{p=1}^{m_1} \left[ (C^p)^\top \bar{y}^p + (A^p)^\top \bar{\mu}^p \right] + \sum_{p=1}^{m_2} \bar{\mu}^{m_1+p} \nabla \phi_p(\bar{x}), \\ c^p + C^p \bar{x} + Q^p \bar{y}^p + (B^p)^\top \bar{\mu}^p = 0, \quad 0 \leq b^p - A^p \bar{x} - B^p \bar{y}^p \perp \bar{\mu}^p \geq 0, \quad p = 1, \dots, m_1, \\ 0 \leq \phi_p(\bar{x}) \perp \bar{\mu}^{m_1+p} \geq 0, \quad p = 1, \dots, m_2, \end{cases}$$

which are the Karush-Kuhn-Tucker (KKT) conditions for the deterministic equivalent of (21).

## 4 A computational algorithm.

In this section, we consider a double-loop algorithm for solving problem  $(\text{CP}_0)$ . The inner loop finds an approximate stationary point of the perturbed composite optimization problem

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \sum_{p=1}^m \left[ F_p^k(x) \triangleq \varphi_p(f_p^k(x)) \right] \quad (23)$$

by solving a sequence of convex subproblems, while the outer loop drives  $k \rightarrow +\infty$ . It is important to note the potential infeasibility in (23) because  $[F_p^k = \varphi_p \circ f_p^k] \xrightarrow{c} F_p$  in Assumption 1(c), together with  $\text{dom}(\varphi_p \circ f_p) \neq \emptyset$ , does not guarantee  $\text{dom}(\varphi_p \circ f_p^k) \neq \emptyset$  for all  $k \in \mathbb{N}$ . This can be seen from the example of  $\varphi(t) = \delta_{(-\infty, 0]}(t)$ ,  $f(x) = \max\{x, 0\} - 1/10$  and  $f^k(x) = \max\{x, 0\} + 1/k - 1/10$ . Obviously  $\text{dom}(\varphi \circ f) = (-\infty, 1/10]$  and  $\varphi \circ f^k \xrightarrow{c} \varphi \circ f$  by [32, Theorem 2.4(d)], but we have  $\text{dom}(\varphi \circ f^k) = \emptyset$  for  $k = 1, \dots, 9$ . Even though  $\text{dom}(\varphi_p \circ f_p^k) \neq \emptyset$  for all  $k \in \mathbb{N}$  and each  $p$ , this does not imply the feasibility of convex subproblems used in the inner loop to approximate (23).

For simplicity of the analysis, we assume that in problem  $(\text{CP}_0)$ ,  $\varphi_p$  is real-valued for  $p = 1, \dots, m_1$ , and  $\varphi_p = \delta_{(-\infty, 0]}$  for  $p = m_1 + 1, \dots, m$ . Namely, the problem takes the following form:

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \sum_{p=1}^{m_1} [F_p(x) = \varphi_p(f_p(x))] \quad \text{subject to } f_p(x) \leq 0, \quad p = m_1 + 1, \dots, m. \quad (\text{CP}_1)$$

For  $p = 1, \dots, m_1$ , the convexity of each real-valued function  $\varphi_p$  implies its continuity by [27, corollary 10.1.1]. Consequently, the composite function  $F_p^k = \varphi_p \circ f_p^k$  is also continuous for  $p = 1, \dots, m_1$  and  $k \in \mathbb{N}$  due to the continuity of each approximating function  $f_p^k$ . It is important to note that model  $(\text{CP}_1)$  still covers discontinuous objective functions since each  $f_p$  can be discontinuous, even though the approximating sequence  $\{f_p^k\}_{k \in \mathbb{N}}$  only consists of locally Lipschitz continuous functions.



## 4.1 Assumptions and examples.

Firstly, we make an assumption to address the feasibility issue outlined at the start of this section. For all  $k \in \mathbb{N}$  and  $p = m_1 + 1, \dots, m$ , define

$$\alpha_p^k \triangleq \sup_{x \in X^k} \left[ f_p^{k+1}(x) - f_p^k(x) \right]_+ \quad \text{with } X^k \triangleq \left\{ x \in \mathbb{R}^n \mid f_p^k(x) \leq 0, p = m_1 + 1, \dots, m \right\}.$$

Based on these auxiliary sequences, we need an initial point  $x^0$  that is strictly feasible to the constraints  $f_p^0(x) \leq 0$  for each  $p = m_1 + 1, \dots, m$ .

**Assumption 2 (strict feasibility)** There exist  $x^0$  and nonnegative sequences  $\{\widehat{\alpha}_p^k\}_{k \in \mathbb{N}}$  for  $p = m_1 + 1, \dots, m$ , such that  $\alpha_p^k \leq \widehat{\alpha}_p^k$  for all  $k \in \mathbb{N}$  and

$$\sum_{k'=0}^{+\infty} \widehat{\alpha}_p^{k'} < +\infty, \quad f_p^0(x^0) \leq - \sum_{k'=0}^{+\infty} \widehat{\alpha}_p^{k'}, \quad p = m_1 + 1, \dots, m.$$

To streamline our notation and analysis, we extend the definitions of  $\alpha_p^k$  and introduce  $\widehat{\alpha}_p^k$  for  $p = 1, \dots, m_1$  by setting  $\alpha_p^k = \widehat{\alpha}_p^k = 0$  for all  $k \in \mathbb{N}$  and  $p = 1, \dots, m_1$ . Since the quantity  $\alpha_p^k$  depends on the sequence  $\{f_p^k\}_{k \in \mathbb{N}}$ , Assumption 2 poses a condition for this approximating sequence. Consider a fixed index  $p \in \{m_1 + 1, \dots, m\}$ . One can use the following way to construct  $\{\alpha_p^k\}_{k \in \mathbb{N}}$ . Suppose that there exist a function  $\widetilde{f}_p : \mathbb{R}^n \times [0, 1] \rightarrow \mathbb{R}$  and a nonnegative sequence  $\gamma_k \downarrow 0$  such that

$$\widetilde{f}_p(x, \gamma_k) = f_p^k(x) \quad \text{and} \quad \widetilde{f}_p(x, 0) = f_p(x).$$

Additionally, assume that for any fixed  $x$ , the function  $\widetilde{f}_p(x, \cdot)$  is continuous on  $[0, 1]$  and differentiable on  $(0, 1)$ , and there exists a constant  $C_p$  such that  $|\nabla_\gamma \widetilde{f}_p(x, \gamma)| \leq C_p$  for any  $x$  and  $\gamma \in (0, 1)$ . For any fixed  $x$ , by the mean value theorem, there exists a point  $\bar{\gamma}_k \in (\gamma_{k+1}, \gamma_k)$  such that  $f_p^{k+1}(x) - f_p^k(x) = (\gamma_{k+1} - \gamma_k) \nabla_\gamma \widetilde{f}_p(x, \bar{\gamma}_k)$ . Thus,

$$\sum_{k'=0}^{+\infty} \alpha_p^{k'} \leq \sum_{k'=0}^{+\infty} (\gamma_{k'} - \gamma_{k'+1}) \sup_{x \in \mathbb{R}^n} |\nabla_\gamma \widetilde{f}_p(x, \bar{\gamma}_{k'})| \leq \sum_{k'=0}^{+\infty} \left[ \widehat{\alpha}_p^{k'} \triangleq C_p (\gamma_{k'} - \gamma_{k'+1}) \right] = C_p \gamma_0 < +\infty.$$

Two more assumptions on the approximating sequences  $\{f_p^k\}_{k \in \mathbb{N}}$  are needed.

**Assumption 3 (smoothness of  $g_p^k$  or  $h_p^k$ )** For each  $k \in \mathbb{N}$ , there exists  $\ell_k > 0$  such that

$$\min \left\{ \mathbb{H}(\partial g_p^k(x), \partial g_p^k(x')), \mathbb{H}(\partial h_p^k(x), \partial h_p^k(x')) \right\} \leq \ell_k \|x' - x\| \quad \forall x, x' \in \mathbb{R}^n, p = 1, \dots, m.$$

**Assumption 4 (level-boundedness)** For each  $k \in \mathbb{N}$ , the function  $H^k \triangleq \sum_{p=1}^m F_p^k$  is level-bounded, i.e., for any  $r \in \mathbb{R}$ , the set

$$\left\{ x \in \mathbb{R}^n \mid \sum_{p=1}^{m_1} \varphi_p(f_p^k(x)) + \sum_{p=m_1+1}^m \delta_{(-\infty, 0]}(f_p^k(x)) \leq r \right\} = \left\{ x \in \mathbb{R}^n \mid \sum_{p=1}^{m_1} \varphi_p(f_p^k(x)) \leq r \right\} \cap X^k$$

is bounded.

Assumption 3 imposes conditions on the Lipschitz continuity of the subdifferential mapping  $\partial g_p^k$  or  $\partial h_p^k$ , which will be used to determine the termination rule of the inner loop. A straightforward sufficient condition for this assumption is that, for each  $p$  and  $k$ , at least one of the functions  $g_p^k$  and  $h_p^k$  is  $\ell_k$ -smooth, i.e.,  $\|\nabla g_p^k(x) - \nabla g_p^k(x')\| \leq \ell_k \|x - x'\|$  or  $\|\nabla h_p^k(x) - \nabla h_p^k(x')\| \leq \ell_k \|x - x'\|$  for any  $x, x' \in \mathbb{R}^n$ . We also remark that Assumption 3 can hold even though both  $g_p^k$  and  $h_p^k$  are nondifferentiable. This can be seen from the following univariate example:  $g_p^k(x) = |x|$  and  $h_p^k(x) = |x - 1|$  for any  $x \in \mathbb{R}$ . It is not difficult to verify that Assumption 3 holds for  $\ell_k = 2$ . Assumption 4 is a standard condition to ensure the boundedness of the generated sequences for each  $k \in \mathbb{N}$ .

In addition, we need a technical assumption to ensure the boundedness of the multiplier sequences in our algorithm.

**Assumption 5 (an asymptotic constraint qualification)** For any  $\bar{x} \in \bigcap_{p=1}^m \text{dom } F_p$ , if there exists  $\{y_p\}_{p=1}^m$  satisfying  $0 = \sum_{p=1}^m y_p v_p$  where for each  $p$  (with the definition of  $T_p(\bar{x})$  in (9)),

$$(y_p, v_p) \in \left( \bigcup \{ \mathcal{N}_{\text{dom } \varphi_p}(t_p) \mid t_p \in T_p(\bar{x}) \} \times \text{con } \partial_A f_p(\bar{x}) \right) \cup \left( \mathbb{R} \times [\partial_A^\infty f_p(\bar{x}) \setminus \{0\}] \right), \quad (24)$$

then we must have  $y_1 = \dots = y_m = 0$ .

The normal cone  $\mathcal{N}_{\text{dom } \varphi_p}(t_p)$  in (24) reduces to  $\{0\}$  for  $p = 1, \dots, m_1$  and  $\mathcal{N}_{(-\infty, 0]}(t_p)$  for  $p = m_1 + 1, \dots, m$ . According to the definitions of  $\partial_A f_p(\bar{x})$  and  $\partial_A^\infty f_p(\bar{x})$ , Assumption 5 depends on the approximating sequences  $\{f_p^k\}_{k \in \mathbb{N}}$  for  $p = 1, \dots, m$ . It holds trivially if each  $\varphi_p$  is real-valued and  $\partial_A^\infty f_p(\bar{x}) = \{0\}$ . By Theorem 1(b), the condition  $\partial_A^\infty f_p(\bar{x}) = \{0\}$  holds when the ADC decompositions are constructed using the Moreau envelope, provided that  $f_p$  is locally Lipschitz continuous and bounded from below. However, in general, Assumption 5 is not easy to verify. For Example 3.1, the assumption translates into

$$\left[ \sum_{p=1}^{m_2} \lambda_p \nabla \phi_p(\bar{x}) = 0, \lambda_p \in \mathcal{N}_{(-\infty, 0]}(\phi_p(\bar{x})), p = 1, \dots, m_2 \right] \implies \lambda_1 = \dots = \lambda_{m_2} = 0.$$

This is equivalent to the Mangasarian-Fromovitz constraint qualification (MFCQ) for problem (21) by [30, Example 6.40]; see also [28].

Furthermore, if each  $f_p$  is c-ADC associated with  $\{f_p^k = g_p^k - h_p^k\}_{k \in \mathbb{N}}$  such that  $\text{con } \partial_A f_p(\bar{x}) = \partial_C f_p(\bar{x})$ , and  $\partial_A^\infty f_p(\bar{x}) = \{0\}$ , Assumption 5 states that

$$\left[ 0 \in \sum_{p=1}^m y_p \partial_C f_p(\bar{x}), y_p \in \mathcal{N}_{\text{dom } \varphi_p}(f_p(\bar{x})), p = 1, \dots, m \right] \implies y_1 = \dots = y_m = 0.$$

This condition aligns with the constraint qualification for the composite optimization problem in [32, Proposition 2.1], and is stronger than the condition in the nonsmooth Lagrange multiplier rule [30, Exercise 10.52]. Finally, Assumption 5 implies the constraint qualifications (15)-(17) in Theorem 2. We formally present this conclusion in the following proposition. The proof of Proposition 4 is given in Appendix B.

**Proposition 4** (consequences of Assumption 5). *Suppose that Assumptions 1 and 5 hold, and  $f_p^k \xrightarrow{c} f_p$  for each  $p$ . If  $\sup \varphi_p = +\infty$  for  $p \in I_1$ , and  $f_p$  is locally Lipschitz continuous for  $p \in I_2$*

(with the definitions of  $I_1$  and  $I_2$  in (8)), then conditions (15), (16), and (17) hold at any feasible point  $\bar{x}$  of (CP<sub>1</sub>). Consequently, any local solution of (CP<sub>1</sub>) is a (weakly)  $A$ -stationary point of (CP<sub>1</sub>).

In the following, we use two examples to further illustrate Assumption 3 and the computation of  $\{\widehat{\alpha}_p^k\}_{k \in \mathbb{N}}$  in Assumption 2.

**Example 4.1** (icc constraints). Let  $f_p$  be real-valued and icc associated with  $\bar{f}_p$ , where  $\bar{f}_p(\cdot, x)$  is Lipschitz continuous with modulus  $L$  for any  $x$ . For the sequence  $\{f_p^k\}_{k \in \mathbb{N}}$  in Example 2.1, it follows from  $g_p^k(x) = \|x\|^2/(2\gamma_k)$  that Assumption 3 holds for  $\ell_k = 1/\gamma_k$ . To construct the quantities  $\widehat{\alpha}_p^k$  in Assumption 2, we notice that

$$\alpha_p^k \leq \sup_{x \in \mathbb{R}^n} \left[ f_p^{k+1}(x) - f_p^k(x) \right]_+ \leq \sup_{x \in \mathbb{R}^n} \left[ f_p(x) - f_p^k(x) \right]_+ \leq \frac{\gamma_k L^2}{2} \triangleq \widehat{\alpha}_p^k \quad \forall k \in \mathbb{N}, \quad (25)$$

where the second inequality is due to  $f_p^{k+1}(x) \leq f_p(x)$  for any  $x$ , and the last one uses the bound between the partial Moreau envelope and the original function [20, Lemma 3]. Thus, the sequence  $\{\widehat{\alpha}_p^k\}_{k \in \mathbb{N}}$  satisfies  $\sum_{k'=0}^{+\infty} \widehat{\alpha}_p^{k'} < +\infty$  if  $\{\gamma_k\}$  is summable.

Alternatively, we can construct the quantities  $\widehat{\alpha}_p^k$  as follows. Let the partial Moreau envelope in (6) be the function  $\widetilde{f}_p$  jointly defined for  $(x, \gamma) \in \mathbb{R}^n \times (0, 1]$ , and  $\widetilde{f}_p(x, 0) = f_p(x)$  for any  $x$ . We claim that  $\widetilde{f}_p(x, \cdot)$  is continuous on  $[0, 1]$  and differentiable on  $(0, 1)$  for any fixed  $x$ . Continuity in  $\gamma$  can be simply checked by a standard argument [30, Theorem 1.17(c)], noting that the optimal value is achieved at a unique point as the function  $\bar{f}_p(\cdot, x) + \|\cdot - x\|^2/(2\gamma)$  is strongly convex for any fixed  $x$ . Differentiability follows from the Danskin's Theorem [11, Theorem 2.1] that  $\nabla_\gamma \widetilde{f}_p(x, \gamma) = -\|z - x\|^2/(2\gamma^2)$  with  $z$  satisfying  $(x - z)/\gamma \in \partial_1 \bar{f}(z, x)$  for any  $(x, \gamma) \in \mathbb{R}^n \times (0, 1]$ . It then follows from the Lipschitz continuity of  $\bar{f}_p(\cdot, x)$  that  $|\nabla_\gamma \widetilde{f}_p(x, \gamma)| \leq L^2/2 \triangleq C_p$  for any  $(x, \gamma) \in \mathbb{R}^n \times (0, 1]$ . Therefore,  $\alpha_p^k \leq C_p(\gamma_k - \gamma_{k+1}) \triangleq \widehat{\alpha}_p^k$  and  $\sum_{k'=0}^{+\infty} \widehat{\alpha}_p^{k'} = C_p \gamma_0 < +\infty$  for any sequence  $\{\widetilde{f}_p(\cdot, \gamma_k)\}_{k \in \mathbb{N}}$  defined by the partial Moreau envelope with  $\gamma_k \downarrow 0$ .

**Example 4.2** (VaR constraints for log-normal distributions). Consider  $f_p(x) = \text{VaR}_\alpha[c(x, Z)]$  with  $c(x, Z) = \exp(x^\top Z)$  for some random vector  $Z \sim \text{Normal}(\mu, \Sigma)$ , where  $\Sigma$  is a positive definite covariance matrix. We have  $c(x, Z) \sim \text{Lognormal}(x^\top \mu, \sqrt{x^\top \Sigma x})$ . The variable  $x$  is restricted to a compact set  $X \subset \mathbb{R}^n$ . Denote the  $\alpha$ -quantile of the standard normal distribution by  $q_\alpha$  and the cumulative distribution function of the standard normal distribution by  $\Phi(\cdot)$ . By direct calculation (cf. [23, Section 3.2]), we have

$$\text{VaR}_\alpha[c(x, Z)] = \exp\left(x^\top \mu + \sqrt{x^\top \Sigma x} q_\alpha\right), \quad \text{CVaR}_\alpha[c(x, Z)] = \exp\left(x^\top \mu + \frac{x^\top \Sigma x}{2}\right) \frac{\Phi(\sqrt{x^\top \Sigma x} - q_\alpha)}{1 - \alpha}.$$

Hence,  $f_p(x) = \text{VaR}_\alpha[c(x, Z)]$  is neither convex nor concave if  $q_\alpha < 0$ . For the sequence  $\{f_p^k\}_{k \in \mathbb{N}}$  in Example 2.2, we can derive that

$$h_p^k(x) = k(1 - \alpha) \text{CVaR}_\alpha[c(x, Z)] = k \exp\left(x^\top \mu + x^\top \Sigma x / 2\right) \Phi(\sqrt{x^\top \Sigma x} - q_\alpha).$$

Since  $\Sigma$  is positive definite, it is easy to see that  $h_p^k$  is twice continuously differentiable. Consequently,  $h_p^k$  is  $\ell_k$ -smooth relative to the compact set  $X$  for some  $\ell_k$ , and Assumption 3 holds

(relative to  $X$ ). Next, we define  $\tilde{f}_p(x, \gamma) = \frac{1}{\gamma} \int_{\alpha-\gamma}^{\alpha} \text{VaR}_t[c(x, Z)] dt$  for any  $(x, \gamma) \in \mathbb{R}^n \times (0, \frac{\alpha}{2}]$  and  $\tilde{f}_p(x, 0) = f_p(x)$  for any  $x$ . Obviously,  $\tilde{f}_p(x, \cdot)$  is continuous on  $[0, \frac{\alpha}{2}]$  and differentiable on  $(0, \frac{\alpha}{2})$  for any fixed  $x$ . By using the Leibniz rule for differentiating the parametric integral, for  $\gamma \in (0, \frac{\alpha}{2})$ , we have

$$\begin{aligned} \left| \nabla_{\gamma} \tilde{f}_p(x, \gamma) \right| &= \frac{1}{\gamma^2} \int_{\alpha-\gamma}^{\alpha} (\text{VaR}_t[c(x, Z)] - \text{VaR}_{\alpha-\gamma}[c(x, Z)]) dt \\ &\leq \frac{1}{\gamma} (\text{VaR}_{\alpha}[c(x, Z)] - \text{VaR}_{\alpha-\gamma}[c(x, Z)]) \\ &= \exp(x^{\top} \mu) \frac{\exp(\sqrt{x^{\top} \Sigma x} q_{\alpha}) - \exp(\sqrt{x^{\top} \Sigma x} q_{\alpha-\gamma})}{\gamma} \\ &= \exp(x^{\top} \mu) \cdot \left[ \exp(\sqrt{x^{\top} \Sigma x} q_{\alpha'}) \sqrt{x^{\top} \Sigma x} \nabla_{\alpha} q_{\alpha'} \right] \end{aligned}$$

for some  $\alpha' \in (\alpha - \gamma, \alpha)$  by the mean-value theorem. By using the fact  $\nabla_{\alpha} q_{\alpha} = \sqrt{2\pi} \exp(q_{\alpha}^2/2)$ , the monotonicity  $q_{\alpha/2} < q_{\alpha'} < q_{\alpha}$ , and the compactness of  $X$ , we further have

$$\sup_{x \in X} \left| \nabla_{\gamma} \tilde{f}_p(x, \gamma) \right| \leq \exp\left(\frac{\max\{q_{\alpha}^2, q_{\alpha/2}^2\}}{2}\right) \sup_{x \in X} \left\{ \exp(x^{\top} \mu + \sqrt{x^{\top} \Sigma x} q_{\alpha}) \sqrt{(2\pi) x^{\top} \Sigma x} \right\} \triangleq C_p < +\infty.$$

Therefore,  $\alpha_p^k \leq C_p(\gamma_k - \gamma_{k+1}) \triangleq \widehat{\alpha}_p^k$  and  $\sum_{k'=0}^{\infty} \widehat{\alpha}_p^k = C_p \gamma_0 < +\infty$  for any sequence  $\{\tilde{f}_p(\cdot, \gamma_k)\}_{k \in \mathbb{N}}$  with  $\gamma_k \downarrow 0$ .

## 4.2 The algorithmic framework and convergence analysis.

We now formalize the algorithm for solving (CP<sub>1</sub>). For  $p = m_1 + 1, \dots, m$ , recall the nonnegative sequences  $\{\widehat{\alpha}_p^k\}_{k \in \mathbb{N}}$  introduced in Assumption 2, and observe that  $\sum_{k'=k}^{+\infty} \widehat{\alpha}_p^{k'} \rightarrow 0$  as  $k \rightarrow +\infty$ . For consistency of our notation, we also set  $\widehat{\alpha}_p^k \equiv 0$  for all  $k \in \mathbb{N}$  and  $p = 1, \dots, m_1$ . At the  $k$ -th outer iteration and for  $p = 1, \dots, m$ , consider the upper and lower approximation of  $f_p^k$  at a point  $y$  by taking some  $a_p^k \in \partial h_p^k(y)$ ,  $b_p^k \in \partial g_p^k(y)$  and incorporating sequences  $\{\widehat{\alpha}_p^k\}_{k \in \mathbb{N}}$ :

$$\begin{aligned} f_p^{k, \text{upper}}(x; y) &\triangleq g_p^k(x) - h_p^k(y) - (a_p^k)^{\top} (x - y) + \sum_{k'=k}^{+\infty} \widehat{\alpha}_p^{k'}, \\ f_p^{k, \text{lower}}(x; y) &\triangleq g_p^k(y) + (b_p^k)^{\top} (x - y) - h_p^k(x). \end{aligned} \tag{26}$$

Observe that, for fixed  $y$ , the upper approximation  $f_p^{k, \text{upper}}(\cdot; y)$  is convex while the lower approximation  $f_p^{k, \text{lower}}(\cdot; y)$  is concave. For  $p = 1, \dots, m_1$ , consider the following function

$$\widehat{F}_p^k(x; y) \triangleq \varphi_p^{\uparrow} \left( f_p^{k, \text{upper}}(x; y) \right) + \varphi_p^{\downarrow} \left( f_p^{k, \text{lower}}(x; y) \right), \tag{27}$$

which is a convex majorization of  $F_p^k$  at a point  $y$  by the fact that  $\varphi_p^{\uparrow}$  is nondecreasing and  $\varphi_p^{\downarrow}$  is nonincreasing. For  $p = m_1 + 1, \dots, m$ , consider the convex constraint  $f_p^{k, \text{upper}}(x; y) \leq 0$  as an approximation for  $f_p^k(x) \leq 0$ .

We summarize the properties of all the surrogate functions as follows. Note that (28a) and (28b) hold for  $p = 1, \dots, m$ , while (28c) holds only for  $p = 1, \dots, m_1$ .

$$f_p^{k,\text{upper}}(x; y) \geq f_p^k(x) + \sum_{k'=k}^{\infty} \widehat{\alpha}_p^{k'} \geq f_p^k(x), \quad f_p^{k,\text{upper}}(x; x) = f_p^k(x) + \sum_{k'=k}^{\infty} \widehat{\alpha}_p^{k'}, \quad (28a)$$

$$f_p^{k,\text{lower}}(x; y) \leq f_p^k(x), \quad f_p^{k,\text{lower}}(x; x) = f_p^k(x), \quad (28b)$$

$$\widehat{F}_p^k(x; y) \geq F_p^k(x), \quad \widehat{F}_p^k(x; x) = F_p^k(x). \quad (28c)$$

The proposed method for solving problem (CP<sub>1</sub>) is outlined in Algorithm 1. The inner loop of the algorithm (indexed by  $i$ ) is terminated when the following conditions are satisfied:

$$\begin{cases} f_p^{k,\text{upper}}(x^{k,i+1}; x^{k,i}) \leq f_p^k(x^{k,i+1}) + \sum_{k'=k}^{+\infty} \widehat{\alpha}_p^{k'} + \epsilon_k, & p = 1, \dots, m, \\ f_p^{k,\text{lower}}(x^{k,i+1}; x^{k,i}) \geq f_p^k(x^{k,i+1}) - \epsilon_k, & p \in I_2, \\ \|x^{k,i+1} - x^{k,i}\| \leq \delta_k / (\lambda + \ell_k), \end{cases} \quad (29)$$

---

**Algorithm 1** The prox-ADC method for solving (CP<sub>1</sub>)

---

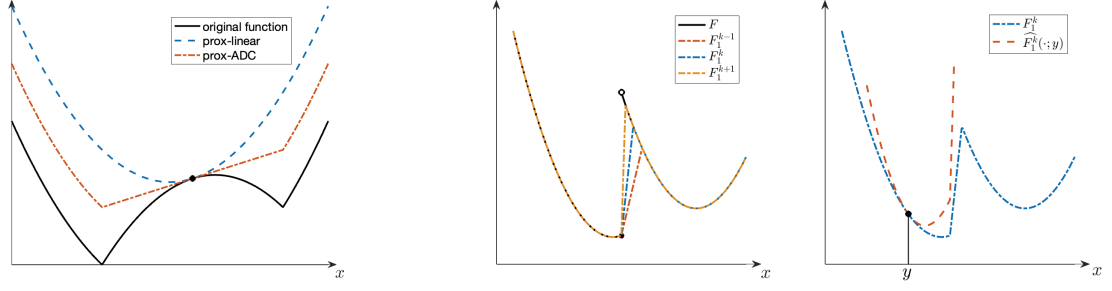
**Input:** Given  $x^0$  and  $\{\widehat{\alpha}_p^k\}_{k \in \mathbb{N}}$  satisfying Assumption 2. Let  $\{\ell_k\}$  be a sequence satisfying Assumption 3. Choose  $\lambda > 0$ , a positive sequence  $(\epsilon_k, \delta_k) \downarrow 0$  such that  $\delta_k / (\lambda + \ell_k) \downarrow 0$ . Set  $k = 0$ .

- 1: **while** a prescribed stopping criterion is not met **do**
- 2:      $x^{k,0} = x^k$
- 3:     **for**  $i = 0, 1, \dots$  **do**
- 4:         Take  $a_p^{k,i} \in \partial g_p^k(x^{k,i})$  for  $p = 1, \dots, m$  and  $b_p^{k,i} \in \partial h_p^k(x^{k,i})$  for  $p = 1, \dots, m_1$
- 5:         Solve the strongly convex subproblem:

$$x^{k,i+1} = \left[ \begin{array}{l} \underset{x \in \mathbb{R}^n}{\text{argmin}} \quad \sum_{p=1}^{m_1} \widehat{F}_p^k(x; x^{k,i}) + \frac{\lambda}{2} \|x - x^{k,i}\|^2 \\ \text{subject to} \quad f_p^{k,\text{upper}}(x; x^{k,i}) \leq 0, \quad p = m_1 + 1, \dots, m \end{array} \right] \quad (30)$$

- 6:         **if** the conditions (29) hold for  $\lambda, \ell_k, \epsilon_k, \delta_k$ , and  $\sum_{k'=k}^{+\infty} \widehat{\alpha}_p^{k'}$  **then**
  - 7:             Break the for-loop
  - 8:         **else**
  - 9:              $i \leftarrow i + 1$
  - 10:         **end if**
  - 11:     **end for**
  - 12:      $x^{k+1} = x^{k,i}$
  - 13:      $k \leftarrow k + 1$
  - 14: **end while**
- 

In contrast to the prox-linear algorithm that is designed to minimize amenable functions and adopts complete linearization of the inner maps, the prox-ADC method retains more curvature information inherent in these maps (see Figure 1). We emphasize that the prox-ADC method differs from [13, Algorithm 7.1.2] that is designed for solving a problem with a convex composite DC objective and DC constraints. Central to the prox-ADC method is the double-loop structure, where, in contrast to [13, Algorithm 7.1.2], the DC sequence  $f_p^k$  is dynamically updated in the outer



(a)  $F_1 = \varphi_1 \circ f_1$  for a convex  $\varphi_1$  and a smooth  $f_1$ . (b)  $F_1 = \varphi_1 \circ f_1$  for a convex nondecreasing  $\varphi_1$  and a lsc  $f_1$ .

Figure 1: Illustrations of the prox-ADC method. (a): a comparison of the prox-ADC and the prox-linear method for minimizing an amenable function. (b): asymptotic approximations of a discontinuous composite function  $F_1 = \varphi_1 \circ f_1$  that are constructed by an epi-convergent sequence  $\{F_1^k = \varphi_1 \circ f_1^k\}$ , and a convex majorization  $\widehat{F}_1^k(\cdot; y)$  for  $F_1^k$ .

loop rather than remaining the same. This adaptation necessitates specialized termination criteria (29) and the incorporation of  $\widehat{\alpha}_p^k$  to maintain feasibility with each update of  $f_p^k$ . In the following, we demonstrate the well-definedness of the prox-ADC method. Specifically, we establish that for each iteration  $k$ , the criteria detailed in (29) are attainable in finitely many steps.

**Theorem 3** (convergence of the inner loop). *Suppose that Assumptions 1-4 hold. Then the following statements hold.*

(a) Problem (30) is feasible for any  $k, i \in \mathbb{N}$ .

(b) The stopping rule of the inner loop is achievable in finitely many steps, i.e., the smallest integer  $i$  satisfying conditions (29), denoted by  $i_k$ , is finite for any  $k \in \mathbb{N}$ .

*Proof.* We prove (a) and (b) by induction. For  $k = 0$ , notice from Assumption 2 and (28a) that  $f_p^{0, \text{upper}}(x^0; x^0) = f_p^0(x^0) + \sum_{k'=0}^{+\infty} \widehat{\alpha}_p^{k'} \leq 0$  for  $p = m_1 + 1, \dots, m$ . Thus, problem (30) is feasible for  $k = i = 0$ . Assume that (30) is feasible for  $k = 0$  and some  $i = \bar{i} (\in \mathbb{N})$ . Consequently,  $x^{0, \bar{i}+1}$  is well-defined and for  $p = m_1 + 1, \dots, m$ ,

$$f_p^{0, \text{upper}}(x^{0, \bar{i}+1}; x^{0, \bar{i}+1}) \stackrel{(28a)}{=} f_p^0(x^{0, \bar{i}+1}) + \sum_{k'=0}^{+\infty} \widehat{\alpha}_p^{k'} \stackrel{(28a)}{\leq} f_p^{0, \text{upper}}(x^{0, \bar{i}+1}; x^{0, \bar{i}}) \leq 0,$$

which yields the feasibility of (30) for  $k = 0, i = \bar{i} + 1$ . Hence, by induction, problem (30) is feasible for  $k = 0$  and any  $i \in \mathbb{N}$ . To proceed, recall the function  $H^k$  defined in Assumption 4. From the update of  $x^{0, i+1}$ , we have

$$H^0(x^{0, i+1}) = \sum_{p=1}^{m_1} F_p^0(x^{0, i+1}) \stackrel{(28c)}{\leq} \sum_{p=1}^{m_1} \widehat{F}_p^0(x^{0, i+1}; x^{0, i}) \leq H^0(x^{0, i}) - \frac{\lambda}{2} \|x^{0, i+1} - x^{0, i}\|^2 \quad \forall i \in \mathbb{N}. \quad (31)$$

The last inequality follows from the definition of  $x^{0, i+1}$  and the second relation in (28c) that  $\widehat{F}_p^0(x^{0, i}; x^{0, i}) = F_p^0(x^{0, i})$  for  $p = 1, \dots, m_1$ . Observe that  $H^0$  is bounded from below by the continuity of  $F_p^0 = \varphi_p \circ f_p^0$  for  $p = 1, \dots, m_1$  (see the discussion following model (CP<sub>1</sub>)) and the level-boundedness of  $H^0$ . Suppose for contradiction that the stopping rule of the inner loop is not achievable in finitely many steps. Then from (31),  $\{H^0(x^{0, i})\}$  converges and  $\sum_{i=0}^{\infty} \|x^{0, i+1} - x^{0, i}\|^2 < +\infty$ .

The latter further yields  $\|x^{0,i+1} - x^{0,i}\| \rightarrow 0$  and thus the last condition in (29) is achievable in finitely many iterations. Next, to derive a contradiction, it suffices to prove that the first two conditions in (29) can also be achieved in finitely many steps. We only show the first one since the other can be done with similar arguments. By the level-boundedness of  $H^0$ , the set  $S^0 \triangleq \{x \mid H^0(x) \leq H^0(x^{0,0})\}$  is compact. Notice that  $x^{0,i} \in S^0$  for all  $i \in \mathbb{N}$  due to (31). For  $p = 1, \dots, m$ , we then have

$$0 \leq f_p^{0,\text{upper}}(x^{0,i+1}; x^{0,i}) - f_p^0(x^{0,i+1}) - \sum_{k'=0}^{+\infty} \widehat{\alpha}_p^{k'} = h_p^0(x^{0,i+1}) - h_p^0(x^{0,i}) - (a_p^{0,i})^\top (x^{0,i+1} - x^{0,i}) \rightarrow 0,$$

because  $h_p^0$  is uniformly continuous on the compact set  $S^0$  and  $\{a_p^{0,i}\}_{i \in \mathbb{N}} \subset \bigcup \{\partial h_p^0(x) \mid x \in S^0\}$  is bounded by [27, Theorem 24.7]. Therefore, for a fixed  $\epsilon_0 > 0$ , there exists some  $i_0$  such that  $f_p^{0,\text{upper}}(x^{0,i_0+1}; x^{0,i_0}) \leq f_p^0(x^{0,i_0+1}) + \sum_{k'=0}^{+\infty} \widehat{\alpha}_p^{k'} + \epsilon_0$  holds for  $p = 1, \dots, m$ . Thus, (a)-(b) hold for  $k = 0$ .

Now assume that (a)-(b) hold for some  $k = \bar{k} (\in \mathbb{N})$  and, hence  $i_{\bar{k}}$  is finite. It then follows from  $x^{\bar{k}+1,0} = x^{\bar{k},i_{\bar{k}}} \in X^{\bar{k}}$  and  $f_p^{\bar{k},\text{upper}}(x^{\bar{k},i_{\bar{k}}}; x^{\bar{k},i_{\bar{k}}}) \leq 0$  that for each  $p = m_1 + 1, \dots, m$ ,

$$\begin{aligned} & f_p^{\bar{k}+1,\text{upper}}(x^{\bar{k}+1,0}; x^{\bar{k}+1,0}) \stackrel{(28a)}{=} f_p^{\bar{k}+1}(x^{\bar{k}+1,0}) + \sum_{k'=\bar{k}+1}^{+\infty} \widehat{\alpha}_p^{k'} \\ & \leq f_p^{\bar{k}}(x^{\bar{k}+1,0}) + \sup_{x \in X^{\bar{k}}} [f_p^{\bar{k}+1}(x) - f_p^{\bar{k}}(x)]_+ + \sum_{k'=\bar{k}+1}^{+\infty} \widehat{\alpha}_p^{k'} \\ & \leq f_p^{\bar{k}}(x^{\bar{k}+1,0}) + \sum_{k'=\bar{k}}^{+\infty} \widehat{\alpha}_p^{k'} \stackrel{(28a)}{=} f_p^{\bar{k},\text{upper}}(x^{\bar{k}+1,0}; x^{\bar{k},i_{\bar{k}}}) \leq 0. \end{aligned}$$

Thus, problem (30) is feasible for  $k = \bar{k} + 1$  and any  $i \in \mathbb{N}$ . Building upon this, we can now clearly see the validity of (b) for  $k = \bar{k} + 1$ , as we have shown similar results earlier in the case of  $k = 0$ . By induction, we complete the proof of (a)-(b).  $\square$

For any  $k \in \mathbb{N}$ , define the set of multipliers for problem (30) as

$$Y^k(x^{k+1}) \triangleq \left\{ \left( \begin{array}{c} y_{1,1}^k \\ y_{1,2}^k \\ \vdots \\ y_{m,1}^k \\ y_{m,2}^k \end{array} \right) \mid \left. \begin{array}{l} 0 \in \sum_{p=1}^m \left[ y_{p,1}^k \partial f_p^{k,\text{upper}}(x^{k,i_k+1}; x^{k,i_k}) + y_{p,2}^k \partial f_p^{k,\text{lower}}(x^{k,i_k+1}; x^{k,i_k}) \right] \\ \quad + \lambda(x^{k,i_k+1} - x^{k,i_k}), \\ y_{p,1}^k \in \partial \varphi_p^\uparrow(f_p^{k,\text{upper}}(x^{k,i_k+1}; x^{k,i_k})), p = 1, \dots, m, \\ y_{p,2}^k \in \partial \varphi_p^\downarrow(f_p^{k,\text{lower}}(x^{k,i_k+1}; x^{k,i_k})), p = 1, \dots, m. \end{array} \right\}.$$

Here  $x^{k,i_k+1}$  is uniquely determined by  $x^{k+1} = x^{k,i_k}$  as the minimizer of a strongly convex problem (30). Notice that  $y_{p,2}^k = 0$  for  $p \in I_1$  since  $\varphi_p$  is nondecreasing and  $\varphi_p^\downarrow = 0$  for  $p \in I_1$ . Let  $\{x^{k+1}\}_{k \in \mathbb{N}}$  be a subsequence that converges to some point  $\bar{x}$ . As we will see in the following lemma, the asymptotic constraint qualification in Assumption 5 implies the non-emptiness and the compactness of  $Y^k(x^{k+1})$  for all sufficiently large  $k \in \mathbb{N}$  and the eventual boundedness of  $\{Y^k(x^{k+1})\}_{k \in \mathbb{N}}$ . These technical results play an important role in the convergence analysis of the prox-ADC method. However, a stronger property of equi-boundedness appears necessary for designing practical termination criteria for the algorithm. We will establish this strengthened property in section 4.3 under non-asymptotic constraint qualifications.

**Lemma 2** (non-emptiness and eventual boundedness of multipliers). *Let  $\bar{x} \in \bigcap_{p=1}^m \text{dom } F_p$  be a feasible point of problem (CP<sub>1</sub>). Suppose that Assumptions 1-5 hold. Consider any sequence  $\{x^k\}$  generated by the prox-ADC method, with a subsequence  $\{x^{k+1}\}_{k \in N}$  converging to  $\bar{x}$ . The following statements hold.*

- (a) *The set of multipliers  $Y^k(x^{k+1})$  is non-empty and compact for all sufficiently large  $k \in N$ .*  
(b) *Additionally, if  $\partial_A^\infty f_p(\bar{x}) = \{0\}$  for  $p \in I_2$  (with the definition of  $I_2$  in (8)), then the sequence  $\{Y^k(x^{k+1})\}_{k \in N}$  is eventually bounded.*

*Proof.* (a) Observe that  $x^{k,i_k+1} \rightarrow_N \bar{x}$  because  $x^{k+1} = x^{k,i_k} \rightarrow_N \bar{x}$  and  $\|x^{k,i_k} - x^{k,i_k+1}\| \leq \delta_k / (\lambda + \ell_k) \downarrow 0$  by conditions (29). The non-emptiness and compactness of  $Y^k(x^{k+1})$  for all sufficiently large  $k \in N$  is a direct consequence of the nonsmooth Lagrange multiplier rule [30, Exercise 10.52] for problem (30) if we can show that, for all sufficiently large  $k \in N$ ,  $y_{m_1+1}^k = \dots = y_m^k = 0$  is the unique solution of the following system

$$0 \in \sum_{p=m_1+1}^m y_p^k \partial f_p^{k,\text{upper}}(x^{k,i_k+1}; x^{k,i_k}), \quad y_p^k \in \mathcal{N}_{(-\infty,0]}(f_p^{k,\text{upper}}(x^{k,i_k+1}; x^{k,i_k})), \quad p = m_1 + 1, \dots, m. \quad (32)$$

Suppose that the above claim does not hold. Then, there exists a subsequence  $N' \subset N$  such that  $y_{m_1+1}^k = \dots = y_m^k = 0$  is not the unique solution of (32) for all  $k \in N'$ . Without loss of generality, suppose  $N' = N$  and take  $\{y_p^k\}_{k \in N}$  for  $p = m_1 + 1, \dots, m$  satisfying (32) and  $\sum_{p=m_1+1}^m |y_p^k| = 1$ . For each  $p$  and  $k \in N$ , define

$$A_p^k \triangleq \left\{ y_p^k v_p^k \mid v_p^k \in \left\{ \partial g_p^k(x^{k,i_k}) - \partial h_p^k(x^{k,i_k}) \right\} \cup \left\{ \partial g_p^k(x^{k,i_k+1}) - \partial h_p^k(x^{k,i_k+1}) \right\} \right\}.$$

Then, for all  $k \in N$ , we have

$$\begin{aligned} & \text{dist} \left( 0, \sum_{p=m_1+1}^m A_p^k \right) \\ & \stackrel{(i)}{\leq} \text{dist} \left( 0, \sum_{p=m_1+1}^m y_p^k [\partial g_p^k(x^{k,i_k+1}) - \partial h_p^k(x^{k,i_k})] \right) + \sum_{p=m_1+1}^m \mathbb{D}(y_p^k [\partial g_p^k(x^{k,i_k+1}) - \partial h_p^k(x^{k,i_k})], A_p^k) \\ & \stackrel{(ii)}{\leq} \text{dist} \left( 0, \sum_{p=m_1+1}^m y_p^k \partial f_p^{k,\text{upper}}(x^{k,i_k+1}; x^{k,i_k}) \right) \\ & \quad + \sum_{p=m_1+1}^m |y_p^k| \cdot \min \left\{ \mathbb{D}(\partial g_p^k(x^{k,i_k+1}), \partial g_p^k(x^{k,i_k})), \mathbb{D}(\partial h_p^k(x^{k,i_k}), \partial h_p^k(x^{k,i_k+1})) \right\} \\ & \stackrel{(iii)}{\leq} 0 + \sum_{p=m_1+1}^m |y_p^k| \cdot \min \left\{ \mathbb{H}(\partial g_p^k(x^{k,i_k+1}), \partial g_p^k(x^{k,i_k})), \mathbb{H}(\partial h_p^k(x^{k,i_k+1}), \partial h_p^k(x^{k,i_k})) \right\} \\ & \stackrel{(iv)}{\leq} \sum_{p=m_1+1}^m |y_p^k| \cdot \ell_k \|x^{k,i_k+1} - x^{k,i_k}\| \stackrel{(v)}{\leq} \delta_k, \end{aligned}$$

where (i) uses the inequalities  $\mathbb{D}(A, C) \leq \mathbb{D}(A, B) + \mathbb{D}(B, C)$  and  $\mathbb{D}(A + B, A' + B') \leq \mathbb{D}(A, A') + \mathbb{D}(B, B')$ ; the first term in (ii) is because of the construction of upper convex majorization (26);



the second term in (ii) is due to  $\mathbb{D}(A, B \cup C) \leq \min\{\mathbb{D}(A, B), \mathbb{D}(A, C)\}$  so that

$$\begin{aligned} & \mathbb{D}(y_p^k [\partial g_p^k(x^{k,i_k+1}) - \partial h_p^k(x^{k,i_k})], A_p^k) \\ &= |y_p^k| \cdot \mathbb{D}\left(\partial g_p^k(x^{k,i_k+1}) - \partial h_p^k(x^{k,i_k}), \{\partial g_p^k(x^{k,i_k}) - \partial h_p^k(x^{k,i_k})\} \cup \{\partial g_p^k(x^{k,i_k+1}) - \partial h_p^k(x^{k,i_k+1})\}\right) \\ &\leq |y_p^k| \cdot \min\left\{\mathbb{D}(\partial g_p^k(x^{k,i_k+1}), \partial g_p^k(x^{k,i_k})), \mathbb{D}(\partial h_p^k(x^{k,i_k}), \partial h_p^k(x^{k,i_k+1}))\right\}. \end{aligned}$$

Inequality (iii) is due to (32) and  $\mathbb{D}(A, B) \leq \mathbb{H}(A, B)$ ; (iv) is by Assumption 3; and (v) is implied by conditions (29) and  $\sum_{p=m_1+1}^m |y_p^k| = 1$ . Equivalently, for all  $k \in N$  and  $p = m_1 + 1, \dots, m$ , there exist  $y_p^k \in \mathcal{N}_{(-\infty, 0]}(f_p^{k, \text{upper}}(x^{k,i_k+1}; x^{k,i_k}))$  with  $\sum_{p=m_1+1}^m |y_p^k| = 1$  and

$$v_p^k \in \left\{ \partial g_p^k(x^{k,i_k}) - \partial h_p^k(x^{k,i_k}) \right\} \cup \left\{ \partial g_p^k(x^{k,i_k+1}) - \partial h_p^k(x^{k,i_k+1}) \right\}$$

such that  $\|\sum_{p=m_1+1}^m y_p^k v_p^k\| \leq \delta_k$ . For  $p = m_1 + 1, \dots, m$ , since the subsequence  $\{f_p^k(x^{k,i_k+1})\}_{k \in N}$  is bounded by Assumption 1(b), we can assume without loss of generality that  $f_p^k(x^{k,i_k+1})$  converges to some  $\bar{z}_p \in T_p(\bar{x})$  as  $k \in N \rightarrow +\infty$ . Furthermore, it can be easily seen from (28a) and (29) that  $f_p^{k, \text{upper}}(x^{k,i_k+1}; x^{k,i_k})$  converges to the same limit point  $\bar{z}_p$  for  $p = m_1 + 1, \dots, m$ . Notice that  $f_p^{k, \text{upper}}(x^{k,i_k+1}; x^{k,i_k}) \leq 0$  for all  $k \in N$  and  $p = m_1 + 1, \dots, m$  from Theorem 3(a) and, thus, each  $\bar{z}_p$  must satisfy  $\bar{z}_p \leq 0$ . Suppose that  $y_p^k \rightarrow_N \bar{y}_p$  for each  $p$ . Then, by the outer semicontinuity of the normal cone [30, Proposition 6.6],

$$\bar{y}_p \in \mathcal{N}_{(-\infty, 0]}(\bar{z}_p) \subset \bigcup \{ \mathcal{N}_{\text{dom } \varphi_p}(t_p) \mid t_p \in T_p(\bar{x}) \}, \quad p = m_1 + 1, \dots, m.$$

Obviously,  $\sum_{p=m_1+1}^m |\bar{y}_p| = 1$ , and the sequence  $\{\bar{y}_p\}_{p=m_1+1}^m$  has at least one nonzero element. Consider two cases.

Case 1. If  $\{v_p^k\}_{k \in N}$  is bounded for  $p = m_1 + 1, \dots, m$ , then there are vectors  $\{\bar{v}_p\}_{p=m_1+1}^m$  with  $\bar{v}_p \in \partial_A f_p(\bar{x})$  such that  $v_p^k \rightarrow_N \bar{v}_p$  and  $0 = \sum_{p=m_1+1}^m \bar{y}_p \bar{v}_p \in \sum_{p=m_1+1}^m \bar{y}_p \partial_A f_p(\bar{x})$ , contradicting Assumption 5 since  $\bar{y}_{m_1+1}, \dots, \bar{y}_m$  are not all zeros.

Case 2. Otherwise, there exists some  $p$  such that  $\{v_p^k\}_{k \in N}$  is unbounded. Define the index sets

$$I_{\text{ub}} \triangleq \left\{ p \in \{m_1 + 1, \dots, m\} \mid \{v_p^k\}_{k \in N} \text{ unbounded} \right\} (\neq \emptyset) \quad \text{and} \quad I_{\text{b}} \triangleq \{m_1 + 1, \dots, m\} \setminus I_{\text{ub}}.$$

Notice that  $\{\sum_{p \in I_{\text{b}}} y_p^k v_p^k\}_{k \in N}$  is bounded. Without loss of generality, assume that this sequence converges to some  $\bar{w}$  and, thus,  $\sum_{p \in I_{\text{ub}}} y_p^k v_p^k \rightarrow_N (-\bar{w})$ .

Step 1: Next we prove by contradiction that, for each  $p \in I_{\text{ub}}$ , the sequence  $\{y_p^k v_p^k\}_{k \in N}$  is bounded. Suppose that the boundedness fails and  $\sum_{p \in I_{\text{ub}}} \|y_p^k v_p^k\| \rightarrow_N +\infty$  by passing to a subsequence. Consider  $\tilde{w}_p^k \triangleq y_p^k v_p^k / \sum_{p \in I_{\text{ub}}} \|y_p^k v_p^k\|$  for  $p \in I_{\text{ub}}$ . Then  $\sum_{p \in I_{\text{ub}}} \tilde{w}_p^k \rightarrow_N 0$ . Since  $\sum_{p \in I_{\text{ub}}} \|\tilde{w}_p^k\| = 1$  for all  $k \in N$ , we can assume that there exist  $p_1 \in I_{\text{ub}}$  and  $\tilde{w}_{p_1} \neq 0$  such that  $\tilde{w}_{p_1}^k \rightarrow_N \tilde{w}_{p_1}$ . It then follows from the construction of  $\tilde{w}_p^k$  that  $\{\tilde{w}_p^k\}_{k \in N}$  has a subsequence converging to some element of  $\pm \partial_A^\infty f_p(\bar{x})$  for each  $p \in I_{\text{ub}}$  and, in particular,  $\tilde{w}_{p_1} \in [\pm \partial_A^\infty f_{p_1}(\bar{x}) \setminus \{0\}]$ . From  $\sum_{p \in I_{\text{ub}}} \tilde{w}_p^k \rightarrow_N 0$ , we obtain

$$0 \in [\pm \partial_A^\infty f_{p_1}(\bar{x}) \setminus \{0\}] + \sum_{p \in I_{\text{ub}} \setminus \{p_1\}} [\pm \partial_A^\infty f_p(\bar{x})],$$

a contradiction to Assumption 5 since the coefficient of the term  $[\pm \partial_A^\infty f_{p_1}(\bar{x}) \setminus \{0\}]$  is nonzero. So far, we have shown the boundedness of  $\{y_p^k v_p^k\}_{k \in N}$  for each  $p \in I_{\text{ub}}$ .

*Step 2:* Now suppose that  $y_p^k v_p^k \rightarrow_N \bar{w}_p$  for each  $p \in I_{\text{ub}}$  with  $\sum_{p \in I_{\text{ub}}} \bar{w}_p = -\bar{w}$ . Thus  $y_p^k \rightarrow_N 0$  and  $\bar{w}_p \in [\pm \partial_A^\infty f_p(\bar{x})]$  for each  $p \in I_{\text{ub}}$ . Since  $\sum_{p=m_1+1}^m |\bar{y}_p| = \sum_{p \in I_b} |\bar{y}_p| = 1$ , there exists  $p_2 \in I_b$  such that  $\bar{y}_{p_2} \neq 0$ . Then  $\sum_{p=m_1+1}^m y_p^k v_p^k \rightarrow_N 0$  implies

$$0 \in \bar{y}_{p_2} \partial_A f_{p_2}(\bar{x}) + \sum_{p \in I_b \setminus \{p_2\}} \bar{y}_p \partial_A f_p(\bar{x}) + \sum_{p \in I_{\text{ub}}} [\pm \partial_A^\infty f_p(\bar{x})],$$

which leads to a contradiction to Assumption 5. Thus,  $Y^k(x^{k+1})$  is non-empty and compact for all sufficiently large  $k \in N$ .

(b) By part (a), assume from now on that  $Y^k(x^{k+1}) \neq \emptyset$  for all  $k \in N$  without loss of generality. We also assume that  $\{f_p^k(x^{k,i_k+1})\}_{k \in N}$  converges to some point  $\bar{z}_p \in T_p(\bar{x})$  for  $p = 1, \dots, m$ . Then, by (28a), (28b) and (29),  $f_p^{k,\text{upper}}(x^{k,i_k+1}; x^{k,i_k}) \rightarrow_N \bar{z}_p$  for  $p = 1, \dots, m$  and  $f_p^{k,\text{lower}}(x^{k,i_k+1}; x^{k,i_k}) \rightarrow_N \bar{z}_p$  for  $p \in I_2$ . For any  $k \in N$  and any  $(y_{1,1}^k, y_{1,2}^k, \dots, y_{m,1}^k, y_{m,2}^k) \in Y^k(x^{k+1})$ , we have  $y_{p,1}^k \in \partial \varphi_p^\uparrow(f_p^{k,\text{upper}}(x^{k,i_k+1}; x^{k,i_k}))$  and  $y_{p,2}^k \in \partial \varphi_p^\downarrow(f_p^{k,\text{lower}}(x^{k,i_k+1}; x^{k,i_k}))$  for  $p = 1, \dots, m$  satisfying

$$0 \in \sum_{p=1}^m \left[ y_{p,1}^k [\partial g_p^k(x^{k,i_k+1}) - \partial h_p^k(x^{k,i_k})] + y_{p,2}^k [\partial g_p^k(x^{k,i_k}) - \partial h_p^k(x^{k,i_k+1})] \right] + \lambda(x^{k,i_k+1} - x^{k,i_k}). \quad (33)$$

Due to Assumption 3 and similar arguments in the proof of part (a), the optimality condition (33) implies that

$$\left\{ \begin{array}{l} \left\| \sum_{p=1}^m (y_{p,1}^k v_{p,1}^k + y_{p,2}^k v_{p,2}^k) \right\| \leq \left[ \lambda + \sum_{p=1}^m (|y_{p,1}^k| + |y_{p,2}^k|) \ell_k \right] \frac{\delta_k}{\lambda + \ell_k} \leq \max \left\{ 1, \sum_{p=1}^m (|y_{p,1}^k| + |y_{p,2}^k|) \right\} \delta_k, \\ \left[ \begin{array}{l} v_{p,1}^k \in \partial g_p^k(x^{k,i_k}) - \partial h_p^k(x^{k,i_k}) \\ v_{p,2}^k \in \partial g_p^k(x^{k,i_k+1}) - \partial h_p^k(x^{k,i_k+1}) \end{array} \right] \text{ or } \left[ \begin{array}{l} v_{p,1}^k \in \partial g_p^k(x^{k,i_k+1}) - \partial h_p^k(x^{k,i_k+1}) \\ v_{p,2}^k \in \partial g_p^k(x^{k,i_k}) - \partial h_p^k(x^{k,i_k}) \end{array} \right], p = 1, \dots, m. \end{array} \right. \quad (34)$$

Note that, for  $p \in I_1$ ,  $\varphi_p$  is nondecreasing, i.e.,  $\varphi_p^\downarrow = 0$ . Then  $y_{p,2}^k = 0$  for all  $k \in N$  and  $p \in I_1$ , and the first inequality of (34) is equivalent to

$$\left\| \sum_{p \in I_1} y_{p,1}^k v_{p,1}^k + \sum_{p \in I_2} (y_{p,1}^k v_{p,1}^k + y_{p,2}^k v_{p,2}^k) \right\| \leq \max \left\{ 1, \sum_{p \in I_1} |y_{p,1}^k| + \sum_{p \in I_2} (|y_{p,1}^k| + |y_{p,2}^k|) \right\} \delta_k. \quad (35)$$

Observe that the sequences  $\{v_{p,1}^k\}_{k \in N}$  and  $\{v_{p,2}^k\}_{k \in N}$  must be bounded for  $p \in I_2$ . Otherwise, we could assume  $\|v_{p,1}^k\| \rightarrow_N +\infty$ . Then every accumulation point of the unit vectors  $\{v_{p,1}^k / \|v_{p,1}^k\|\}_{k \in N}$  would be in the set  $\partial_A^\infty f_p(\bar{x})$ , contradicting our assumption that  $\partial_A^\infty f_p(\bar{x}) = \{0\}$  for each  $p \in I_2$ .

For  $p \in I_2 \subset \{1, \dots, m_1\}$ , given that  $\varphi_p^\uparrow$  is convex, real-valued, and  $f_p^{k,\text{upper}}(x^{k,i_k+1}; x^{k,i_k}) \rightarrow_N \bar{z}_p$ , we can invoke [27, Theorem 24.7] to deduce the boundedness of  $\{y_{p,1}^k\}_{k \in N}$ . A parallel reasoning applies to demonstrate the boundedness of  $\{y_{p,2}^k\}_{k \in N}$ .

For  $p \in I_1$ , we proceed by contradiction to establish the boundedness of  $\{y_{p,1}^k\}_{k \in N}$  based on Assumption 5. Suppose that  $\{\sum_{p \in I_1} |y_{p,1}^k|\}_{k \in N}$  is unbounded and  $\sum_{p \in I_1} |y_{p,1}^k| \rightarrow_N +\infty$  by

passing to a subsequence. Consider the normalized subsequences  $\{\tilde{y}_{p,1}^k \triangleq y_{p,1}^k / \sum_{p' \in I_1} |y_{p',1}^k|\}_{k \in N}$  and  $\{\tilde{y}_{p,2}^k \triangleq y_{p,2}^k / \sum_{p' \in I_1} |y_{p',1}^k|\}_{k \in N}$  for each  $p$ . Consequently,  $\tilde{y}_{p,1}^k \rightarrow_N 0$  and  $\tilde{y}_{p,2}^k \rightarrow_N 0$  for  $p \in I_2$ . By the triangle inequality and (35), we have

$$\begin{aligned} & \left\| \left\| \sum_{p \in I_1} \tilde{y}_{p,1}^k v_{p,1}^k \right\| - \left\| \sum_{p \in I_2} (\tilde{y}_{p,1}^k v_{p,1}^k + \tilde{y}_{p,2}^k v_{p,2}^k) \right\| \right\| \leq \left\| \sum_{p \in I_1} \tilde{y}_{p,1}^k v_{p,1}^k + \sum_{p \in I_2} (\tilde{y}_{p,1}^k v_{p,1}^k + \tilde{y}_{p,2}^k v_{p,2}^k) \right\| \\ & \leq \max \left\{ \frac{1}{\sum_{p \in I_1} |y_{p,1}^k|}, 1 + \frac{\sum_{p \in I_2} (|y_{p,1}^k| + |y_{p,2}^k|)}{\sum_{p \in I_1} |y_{p,1}^k|} \right\} \delta_k \rightarrow_N 0, \end{aligned}$$

which further implies  $\|\sum_{p \in I_1} \tilde{y}_{p,1}^k v_{p,1}^k\| \rightarrow_N 0$  by the boundedness of  $\{v_{p,1}^k\}_{k \in N}$  and  $\{v_{p,2}^k\}_{k \in N}$  for  $p \in I_2$ . Now suppose that  $\tilde{y}_{p,1}^k \rightarrow_N \tilde{y}_{p,1}$  for  $p \in I_1$ . Then from a similar reasoning in (20), for  $p \in I_1$ ,

$$\tilde{y}_{p,1} \in \text{Lim sup}_{k \in N \rightarrow +\infty}^\infty \partial \varphi_p^\uparrow \left( f_p^{k, \text{upper}}(x^{k, i_k+1}, x^{k, i_k}) \right) \subset \partial^\infty \varphi_p^\uparrow(\bar{z}_p) = \mathcal{N}_{\text{dom } \varphi_p^\uparrow}(\bar{z}_p),$$

and obviously  $\sum_{p \in I_1} |\tilde{y}_{p,1}| = 1$ . The remaining argument to derive a contradiction to Assumption 5 follows the same steps as the proof of part (a) for the two cases, with the exception that the index set  $\{m_1 + 1, \dots, m\}$  is replaced by  $I_1$ . Thus, the sequences  $\{y_{p,1}^k\}_{k \in N}$  for  $p \in I_1 \cup I_2$  and  $\{y_{p,2}^k\}_{k \in N}$  for  $p \in I_1$  are bounded. We can conclude that  $\bigcup \{Y^k(x^{k+1}) \mid k \in N, k \geq K\}$  is bounded for sufficiently large integer  $K$ , because otherwise we could extract a subsequence of multipliers from  $Y^k(x^{k+1})$  whose norms diverge to  $+\infty$  as  $k \in N \rightarrow +\infty$ , in contradiction to the result of boundedness that we have shown. Hence, the subsequence  $\{Y^k(x^{k+1})\}_{k \in N}$  is eventually bounded.  $\square$

We make a remark on Lemma 2(b) about the additional assumption. According to the proof of part (b), the assumption  $\partial_A^\infty f_p(\bar{x}) = \{0\}$  for  $p \in I_2$  ensures the boundedness of the set  $\partial_A f_p(\bar{x})$  for  $p \in I_2$ . There are some sufficient conditions for  $\partial_A^\infty f_p(\bar{x}) = \{0\}$  to hold: (i) If  $f_p$  is locally Lipschitz continuous and bounded from below, by Theorem 1(b), we have  $\partial_A^\infty f_p(x) = \{0\}$  at any  $x \in \text{dom } f_p$  for the approximating sequence generated by the Moreau envelope. (ii) If  $f_p$  is icc associated with  $\bar{f}_p$  satisfying all assumptions in Proposition 2, it then follows from Proposition 2(b) that  $\partial_A^\infty f_p(x) = \{0\}$  at any  $x \in \text{int}(\text{dom } f_p)$  for the approximating sequence based on the partial Moreau envelope. It is worth mentioning that the icc function  $f_p$  under condition (ii) is not necessarily locally Lipschitz continuous.

The main convergence result of the prox-ADC method follows.

**Theorem 4.** *Suppose that Assumptions 1-5 hold. Let  $\{x^k\}$  be the sequence generated by the prox-ADC method. Suppose that  $\{x^k\}$  has an accumulation point  $\bar{x}$  and, in addition,  $\partial_A^\infty f_p(\bar{x}) = \{0\}$  for  $p \in I_2$ . Then  $\bar{x}$  is a weakly A-stationary point of  $(\text{CP}_1)$ . Moreover, if for each  $p \in I_2$ , the functions  $g_p^k$  and  $h_p^k$  are  $\ell_k$ -smooth for all  $k \in \mathbb{N}$ , i.e., there exists a sequence  $\{\ell_k\}$  such that for all  $k \in \mathbb{N}$ ,*

$$\max \left\{ \|\nabla g_p^k(x) - \nabla g_p^k(x')\|, \|\nabla h_p^k(x) - \nabla h_p^k(x')\| \right\} \leq \ell_k \|x' - x\| \quad \forall x, x' \in \mathbb{R}^n, p \in I_2, \quad (36)$$

*then  $\bar{x}$  is also an A-stationary point of  $(\text{CP}_1)$ .*

*Proof.* Let  $\{x^{k+1}\}_{k \in \mathbb{N}}$  be a subsequence converging to  $\bar{x}$ . By the stopping conditions (29) and  $x^{k,i_k} \rightarrow_N \bar{x}$ , we also have  $x^{k,i_k+1} \rightarrow_N \bar{x}$ . First, we prove  $\bar{x} \in \bigcap_{p=1}^m \text{dom } F_p$ . From Theorem 3(a), we have  $f_p^k(x^{k,i_k+1}) \leq 0$  for  $p = m_1 + 1, \dots, m$  and all  $k \in \mathbb{N}$ . Due to epi-convergence in Assumption 1(c), it holds that

$$\delta_{(-\infty, 0]}(f_p(\bar{x})) \leq \liminf_{k \in \mathbb{N} \rightarrow +\infty} \delta_{(-\infty, 0]}(f_p^k(x^{k,i_k+1})) = 0, \quad p = m_1 + 1, \dots, m.$$

Thus  $f_p(\bar{x}) \leq 0$  for  $p = m_1 + 1, \dots, m$  and  $\bar{x} \in \bigcap_{p=m_1+1}^m \text{dom } F_p$ . By Assumption 1(a),  $\text{dom } \varphi_p = \mathbb{R}^n$  for all  $p = 1, \dots, m_1$ . This implies  $\bar{x} \in \bigcap_{p=1}^{m_1} \text{dom } F_p$ , and we can conclude that  $\bar{x} \in \bigcap_{p=1}^m \text{dom } F_p$ .

By Lemma 2(a), for all sufficiently large  $k \in N$ , we have

$$0 \in \sum_{p=1}^m \left[ y_{p,1}^k \left( \partial g_p^k(x^{k,i_k+1}) - \partial h_p^k(x^{k,i_k}) \right) + y_{p,2}^k \left( \partial g_p^k(x^{k,i_k}) - \partial h_p^k(x^{k,i_k+1}) \right) \right] + \lambda(x^{k,i_k+1} - x^{k,i_k}), \quad (37)$$

where  $y_{p,1}^k \in \partial \varphi_p^\uparrow(f_p^{k,\text{upper}}(x^{k,i_k+1}; x^{k,i_k}))$  and  $y_{p,2}^k \in \partial \varphi_p^\downarrow(f_p^{k,\text{lower}}(x^{k,i_k+1}; x^{k,i_k}))$  for  $p = 1, \dots, m$ . It follows from Lemma 2(b) that the subsequences  $\{y_{p,1}^k\}_{k \in N}$  and  $\{y_{p,2}^k\}_{k \in N}$  are bounded for  $p = 1, \dots, m$ . Suppose that  $y_{p,1}^k \rightarrow_N \bar{y}_{p,1}$  and  $y_{p,2}^k \rightarrow_N \bar{y}_{p,2}$  for  $p = 1, \dots, m$ . Recall that the subsequence  $\{f_p^k(x^{k,i_k+1})\}_{k \in N}$  is bounded by Assumption 1(b) for  $p = 1, \dots, m$ . Without loss of generality, assume that  $\{f_p^k(x^{k,i_k+1})\}_{k \in N}$  converges to some point  $\bar{z}_p \in T_p(\bar{x})$  for  $p = 1, \dots, m$ . Then, by (28a), (28b) and (29),  $f_p^{k,\text{upper}}(x^{k,i_k+1}; x^{k,i_k}) \rightarrow_N \bar{z}_p$  for  $p = 1, \dots, m$  and  $f_p^{k,\text{lower}}(x^{k,i_k+1}; x^{k,i_k}) \rightarrow_N \bar{z}_p$  for  $p \in I_2$ . From the outer semicontinuity of  $\partial \varphi_p^\uparrow$  and  $\partial \varphi_p^\downarrow$ , we have  $\bar{y}_{p,1} \in \partial \varphi_p^\uparrow(\bar{z}_p)$  for  $p = 1, \dots, m$  and  $\bar{y}_{p,2} \in \partial \varphi_p^\downarrow(\bar{z}_p)$  for  $p \in I_2$ .

To proceed, we prove by contradiction that the sequence  $\{y_{p,1}^k v_{p,1}^k\}_{k \in N}$  is bounded for  $p \in I_1$ . Suppose that  $\sum_{p \in I_1} \|y_{p,1}^k v_{p,1}^k\| \rightarrow_N +\infty$ . For each  $p \in I_2$ , the boundedness of  $\{v_{p,1}^k\}_{k \in N}$  and  $\{v_{p,2}^k\}_{k \in N}$  follows from the assumption  $\partial_A^\infty f_p(\bar{x}) = \{0\}$ ; otherwise, any accumulation point of the unit vectors  $\{v_{p,1}^k / \|v_{p,1}^k\|\}_{k \in N}$  would be in  $\partial_A^\infty f_p(\bar{x})$ , leading to a contradiction. Since  $\{y_{p,1}^k\}_{k \in N}$  and  $\{y_{p,2}^k\}_{k \in N}$  for  $p \in I_2$  are also bounded, we conclude that the subsequence  $\{\sum_{p \in I_2} (y_{p,1}^k v_{p,1}^k + y_{p,2}^k v_{p,2}^k)\}_{k \in N}$  is bounded. Thus, we can assume that

$$\sum_{p \in I_2} (y_{p,1}^k v_{p,1}^k + y_{p,2}^k v_{p,2}^k) \rightarrow_N \bar{w} \left( \in \sum_{p \in I_2} (\bar{y}_{p,1} \partial_A f_p(\bar{x}) + \bar{y}_{p,2} \partial_A f_p(\bar{x})) \right).$$

By (35), it follows that  $\sum_{p \in I_1} y_{p,1}^k v_{p,1}^k \rightarrow_N (-\bar{w})$ . Consider  $\tilde{w}_p^k \triangleq y_{p,1}^k v_{p,1}^k / \sum_{p' \in I_1} \|y_{p',1}^k v_{p',1}^k\|$  for  $p \in I_1$ , and then  $\sum_{p \in I_1} \tilde{w}_p^k \rightarrow_N 0$ . Given  $\sum_{p \in I_1} \|\tilde{w}_p^k\| = 1$  for all  $k \in N$ , there must exist  $p_1 \in I_1$  such that  $\tilde{w}_{p_1}^k \rightarrow_N \tilde{w}_{p_1} \neq 0$ . For each  $p \in I_1$ , it then follows from  $y_{p,1}^k / \sum_{p' \in I_1} \|y_{p',1}^k v_{p',1}^k\| \rightarrow_N 0$  that  $\{\tilde{w}_p^k\}_{k \in N}$  has a subsequence converging to some element in  $\partial_A^\infty f_p(\bar{x})$ . In particular,  $\tilde{w}_{p_1} \in \partial_A^\infty f_{p_1}(\bar{x}) \setminus \{0\}$ . Since  $\sum_{p \in I_1} \tilde{w}_p^k \rightarrow_N 0$ , this implies that

$$0 \in [\partial_A^\infty f_{p_1}(\bar{x}) \setminus \{0\}] + \sum_{p \in I_1 \setminus \{p_1\}} \partial_A^\infty f_p(\bar{x}),$$

which contradicts Assumption 5. Hence,  $\{y_{p,1}^k v_{p,1}^k\}_{k \in N}$  is bounded for  $p \in I_1$ .

We are now ready to prove that  $\bar{x}$  is a weakly A-stationary point. Suppose that  $y_{p,1}^k v_{p,1}^k \rightarrow_N \bar{w}_p$  for  $p \in I_1$  with  $\sum_{p \in I_1} \bar{w}_p = -\bar{w}$ . It remains to show that for each  $p \in I_1$ , there exists  $\bar{y}_{p,1} \in \bigcup \{\partial\varphi_p^\uparrow(t_p) \mid t_p \in T_p(\bar{x})\}$  such that

$$\bar{w}_p \in \{\bar{y}_{p,1} \partial_A f_p(\bar{x})\} \cup [\partial_A^\infty f_p(\bar{x}) \setminus \{0\}],$$

which can be derived similarly as the proof of (19) in Theorem 2. Summarizing these arguments, we conclude that  $\bar{x}$  is a weakly A-stationary point of  $(\text{CP}_1)$ .

Under the additional assumption of the theorem, there exist  $y_{p,1}^k \in \partial\varphi_p^\uparrow(f_p^{k,\text{upper}}(x^{k,i_k+1}, x^{k,i_k}))$ ,  $y_{p,2}^k \in \partial\varphi_p^\downarrow(f_p^{k,\text{lower}}(x^{k,i_k+1}, x^{k,i_k}))$  for  $p = 1, \dots, m$ , and

$$v_{p,1}^k \in \left\{ \partial g_p^k(x^{k,i_k}) - \partial h_p^k(x^{k,i_k}) \right\} \cup \left\{ \partial g_p^k(x^{k,i_k+1}) - \partial h_p^k(x^{k,i_k+1}) \right\} \text{ for } p \in I_1$$

such that

$$\begin{aligned} & \left\| \sum_{p \in I_1} y_{p,1}^k v_{p,1}^k + \sum_{p \in I_2} (y_{p,1}^k + y_{p,2}^k) \left[ \nabla g_p^k(x^{k,i_k}) - \nabla h_p^k(x^{k,i_k}) \right] \right\| \\ & \stackrel{\text{(vi)}}{\leq} \lambda \|x^{k,i_k+1} - x^{k,i_k}\| + \sum_{p \in I_1} |y_{p,1}^k| \cdot \min \left\{ \mathbb{H} \left( \partial g_p^k(x^{k,i_k+1}), \partial g_p^k(x^{k,i_k}) \right), \mathbb{H} \left( \partial h_p^k(x^{k,i_k+1}), \partial h_p^k(x^{k,i_k}) \right) \right\} \\ & \quad + \sum_{p \in I_2} \left( |y_{p,1}^k| \cdot \|\nabla g_p^k(x^{k,i_k}) - \nabla g_p^k(x^{k,i_k+1})\| + |y_{p,2}^k| \cdot \|\nabla h_p^k(x^{k,i_k+1}) - \nabla h_p^k(x^{k,i_k})\| \right) \\ & \stackrel{\text{(vii)}}{\leq} \left[ \lambda + \left( \sum_{p \in I_1} |y_{p,1}^k| + \sum_{p \in I_2} (|y_{p,1}^k| + |y_{p,2}^k|) \right) \ell_k \right] \|x^{k,i_k+1} - x^{k,i_k}\| \\ & \stackrel{\text{(viii)}}{\leq} \max \left\{ 1, \sum_{p \in I_1} |y_{p,1}^k| + \sum_{p \in I_2} (|y_{p,1}^k| + |y_{p,2}^k|) \right\} \delta_k \quad \forall k \in \mathbb{N}, \end{aligned}$$

where (vi) is implied by the optimality condition (37), (vii) employs (36) and Assumption 3, and (viii) follows from conditions (29). This inequality is a tighter version of (35) in the sense that, for each  $p \in I_2$  and  $k \in \mathbb{N}$ ,  $v_{p,1}^k$  and  $v_{p,2}^k$  are elements taken from the single-valued mapping  $\nabla g_p^k(\cdot) - \nabla h_p^k(\cdot)$  evaluated at the same point  $x^{k,i_k}$ . A straightforward adaptation of the preceding argument confirms that  $\bar{x}$  is an A-stationary point of  $(\text{CP}_1)$ .  $\square$

### 4.3 Termination criteria.

The previous subsection demonstrates the asymptotic convergence of the algorithm, showing that any accumulation point of the sequence generated by the prox-ADC method is weakly A-stationary. This subsection is dedicated to the non-asymptotic analysis of verifiable termination criteria for practical implementation.

**Assumption 6 (non-asymptotic constraint qualifications)** Let  $\lambda$  be the parameter in Algorithm 1. For all  $k \in \mathbb{N}$  and any pair  $(x', x'')$  satisfying

$$x'' = \left[ \begin{array}{l} \operatorname{argmin}_{x \in \mathbb{R}^n} \sum_{p=1}^{m_1} \widehat{F}_p^k(x; x') + \frac{\lambda}{2} \|x - x'\|^2 \\ \text{subject to } f_p^{k, \text{upper}}(x; x') \leq 0, p = m_1 + 1, \dots, m \end{array} \right],$$

if there exist  $y_p^k \in \mathcal{N}_{(-\infty, 0]}(f_p^{k, \text{upper}}(x''; x'))$  for  $p = m_1 + 1, \dots, m$  such that

$$0 \in \sum_{p=m_1+1}^m y_p^k \partial f_p^{k, \text{upper}}(x''; x'),$$

then we must have  $y_{m_1+1}^k = \dots = y_m^k = 0$ .

A direct consequence of Assumption 6 and the nonsmooth Lagrange multiplier rule [30, Exercise 10.52] is that the set of multipliers  $Y^k(x^{k+1})$  is non-empty and compact for any fixed  $k \in \mathbb{N}$ . This is in contrast with Lemma 2(a), where the results only hold for sufficiently large  $k \in N$ . We will show below that the result on the eventual boundedness of the subsequence  $\{Y^k(x^{k+1})\}_{k \in N}$  can be strengthened to the equi-boundedness under this assumption.

**Proposition 5** (equi-boundedness of multipliers). *Suppose that Assumptions 1-6 hold. Consider any sequence  $\{x^k\}$  generated by the prox-ADC method. The following statements hold.*

(a) *If there is a subsequence  $\{x^{k+1}\}_{k \in N}$  converging to some  $\bar{x}$  and  $\partial_A^\infty f_p(\bar{x}) = \{0\}$  for  $p \in I_2$ , then the subsequence  $\{Y^k(x^{k+1})\}_{k \in N}$  is equi-bounded.*

(b) *If  $\{x^k\}$  is bounded and  $\partial_A^\infty f_p(x) = \{0\}$  for any  $x \in \bigcap_{p=1}^m \operatorname{dom} F_p$  and  $p \in I_2$ , then the sequence  $\{Y^k(x^{k+1})\}$  is equi-bounded, i.e.,*

$$D \triangleq \sup_{k \in \mathbb{N}} \sup_{y \in Y^k(x^{k+1})} \|y\| < +\infty. \quad (38)$$

*Proof.* (a) We know from Lemma 2(b) that the subsequence  $\{Y^k(x^{k+1})\}_{k \in N}$  is eventually bounded. This implies the existence of an index  $K \in N$  such that  $\bigcup\{Y^k(x^{k+1}) \mid k \in N, k \geq K\}$  is bounded. On the other hand, it follows from Assumption 6 that  $Y^k(x^{k+1})$  is non-empty and compact for any fixed  $k \in N$ . Thus,  $\bigcup\{Y^k(x^{k+1}) \mid k \in N\}$  is bounded, and  $\{Y^k(x^{k+1})\}_{k \in N}$  is equi-bounded.

(b) Suppose for contradiction that  $\{Y^k(x^{k+1})\}$  is not equi-bounded. Then for any nonnegative integer  $j$ , there is an index  $k_j \in \mathbb{N}$  such that  $\|y^{k_j}\| \geq j$  for some multiplier  $y^{k_j} \in Y^{k_j}(x^{k_j+1})$ . Observe that the nonnegative sequence of indices  $\{k_j\}_{j \in \mathbb{N}}$  is either bounded or unbounded. It suffices to consider these two cases separately.

Suppose first that  $\{k_j\}_{j \in \mathbb{N}}$  is bounded. There must be an index  $\bar{k} \in \mathbb{N}$  that appears infinitely many times in  $\{k_j\}_{j \in \mathbb{N}}$ . Consequently, the set  $Y^{\bar{k}}(x^{\bar{k}+1})$  is unbounded, a contradiction to Assumption 6.

Suppose next that  $\{k_j\}_{j \in \mathbb{N}}$  is unbounded. For some index set  $N' \in \mathbb{N}_\infty^\#$ , we have  $k_j \rightarrow +\infty$  as  $j(\in N') \rightarrow +\infty$ . Notice that the subsequence  $\{x^{k_j}\}_{j \in N'}$  is bounded since  $\{x^k\}$  is bounded. By passing to a subsequence if necessary, we assume that  $\{x^{k_j}\}_{j \in N'}$  converges to some  $\tilde{x}$ . Using epi-convergence in Assumption 1(c) and following the same procedure as in the proof of Theorem 4, we can obtain that  $\tilde{x} \in \bigcap_{p=1}^m \operatorname{dom} F_p$ . Then, by the assumption in (b),  $\partial_A^\infty f_p(\tilde{x}) = \{0\}$  for  $p \in I_2$ . Henceforth, there is a subsequence  $\{x^{k_j}\}_{j \in N'} \subset \{x^k\}$  converging to some  $\tilde{x}$  with a corresponding

subsequence of multipliers  $\{y^{k_j} \in Y^{k_j}(x^{k_j+1})\}_{j \in N'}$  such that  $\|y^{k_j}\| \rightarrow +\infty$  as  $j(\in N') \rightarrow +\infty$ , which is a contradiction to the result of part (a).

We have obtained contradictions for the two cases where  $\{k_j\}_{j \in \mathbb{N}}$  is bounded or unbounded. Then we can conclude that  $\{Y^k(x^{k+1})\}$  is equi-bounded and the quantity  $D$  defined in (38) is finite.  $\square$

After obtaining the equi-boundedness of the multipliers, we next introduce a relaxation of the weakly A-stationary point for preparation of the termination criteria. For a proper and convex function  $f$  and any  $\beta > 0$ , we denote  $\partial^\beta f(\bar{x}) \triangleq \bigcup\{\partial f(x) \mid x \in \mathbb{B}(\bar{x}, \beta)\}$ , which is related to the Goldstein's  $\beta$ -subdifferential [18].

**Definition 5.** *Given any  $\bar{\eta} > 0$ ,  $\bar{\beta} > 0$  and  $\bar{k} \in \mathbb{N}$ , we say a point  $x$  is a  $(\bar{\eta}, \bar{\beta}, \bar{k})$ -weakly A-stationary point of problem  $(\text{CP}_0)$  if there exists a nonnegative integer  $k \geq \bar{k}$  such that*

$$\text{dist} \left( 0, \sum_{p=1}^m \bigcup \left\{ y_{p,1} [\partial^{\bar{\beta}} g_p^k(x) - \partial^{\bar{\beta}} h_p^k(x)] + y_{p,2} [\partial^{\bar{\beta}} g_p^k(x) - \partial^{\bar{\beta}} h_p^k(x)] \mid \begin{array}{l} y_{p,1} \in \partial^{\bar{\beta}} \varphi_p^\uparrow(f_p^k(x)), \\ y_{p,2} \in \partial^{\bar{\beta}} \varphi_p^\downarrow(f_p^k(x)) \end{array} \right\} \right) \leq \bar{\eta}.$$

We remark that, if each outer function  $\varphi_p$  is an identity function, i.e.,  $\varphi_p(t) = t$  for any  $t \in \mathbb{R}$ , and each inner function  $f_p$  is DC rather than ADC, the above definition in the context of a DC program is independent of  $k$  and says about nearness to a  $\bar{\eta}$ -critical point [35, definition 2]. For nonsmooth optimization problem, similar definitions based on the idea of small nearby subgradients, together with the termination criteria, have appeared in the literature [18, 9].

The following proposition reveals the relationship between a  $(\bar{\eta}, \bar{\beta}, \bar{k})$ -weakly A-stationary point and a weakly A-stationary point.

**Proposition 6.** *Let  $\bar{x} \in \bigcap_{p=1}^m \text{dom } F_p$  be a feasible point of  $(\text{CP}_0)$ . Suppose that Assumption 1 holds and  $\partial_A^\infty f_p(\bar{x}) = \{0\}$  for each  $p = 1, \dots, m$ . For any nonnegative sequence  $(\eta_k, \beta_k) \downarrow 0$  and some index set  $N \in \mathbb{N}_\infty^\sharp$ , if each  $x^k$  is a  $(\eta_k, \beta_k, k)$ -weakly A-stationary point of  $(\text{CP}_0)$  for  $k \in N$  and  $x^k \rightarrow_N \bar{x}$ , then  $\bar{x}$  is a weakly A-stationary point of  $(\text{CP}_0)$ .*

*Proof.* By Assumption 1(b), the subsequence  $\{f_p^k(x^k)\}_{k \in N}$  is bounded for each  $p$ . Then, there is an index set  $N'(\subset N) \in \mathbb{N}_\infty^\sharp$  such that  $\{f_p^k(x^k)\}_{k \in N'}$  converges to some  $\bar{t}_p \in T_p(\bar{x})$  for each  $p$ . Using the outer semicontinuity of the subdifferential mapping of a convex function, we have

$$\limsup_{k(\in N') \rightarrow +\infty} \partial^{\beta_k} \varphi_p^\uparrow(f_p^k(x^k)) \subset \partial \varphi_p^\uparrow(\bar{t}_p), \quad \limsup_{k(\in N') \rightarrow +\infty} \partial^{\beta_k} \varphi_p^\downarrow(f_p^k(x^k)) \subset \partial \varphi_p^\downarrow(\bar{t}_p)$$

and

$$\limsup_{k(\in N') \rightarrow +\infty} [\partial^{\beta_k} g_p^k(x^k) - \partial^{\beta_k} h_p^k(x^k)] \subset \partial_A f_p(\bar{x}).$$

Thus, by taking an outer limit of the subdifferentials involved in the condition that  $x^k$  is  $(\eta_k, \beta_k, k)$ -weakly A-stationary for all  $k \in N$ , we know that  $\bar{x}$  is a weakly A-stationary point of  $(\text{CP}_0)$ .  $\square$

We conclude this section with our main result on the termination criteria.

**Proposition 7** (termination criteria). *Suppose that Assumptions 1-6 hold. Let  $\{x^k\}$  be the sequence generated by the prox-ADC method. Suppose that  $\{x^k\}$  is bounded and  $\partial_A^\infty f_p(x) = \{0\}$  for any  $x \in \bigcap_{p=1}^m \text{dom } F_p$  and  $p \in I_2$ . For any  $\bar{\eta} > 0$ ,  $\bar{\beta} > 0$  and  $\bar{k} \in \mathbb{N}$ , there exists a nonnegative integer  $k_0 \geq \bar{k}$  such that*

$$\max_{1 \leq p \leq m} \sum_{k'=k_0}^{+\infty} \widehat{\alpha}_p^{k'} + \epsilon_{k_0} \leq \bar{\beta}, \quad \frac{\delta_{k_0}}{\lambda + \ell_{k_0}} \leq \bar{\beta}, \quad \delta_{k_0} \leq \bar{\eta}. \quad (39)$$

Consequently,  $x^{k_0, i_{k_0} + 1}$  is a  $(\bar{\eta} \cdot \max\{1, \sqrt{2mD}\}, \bar{\beta}, \bar{k})$ -weakly A-stationary point of problem (CP<sub>1</sub>), where  $D$  is the constant defined in (38).

*Proof.* The existence of  $k_0 \geq \bar{k}$  satisfying (39) is a direct consequence of  $\sum_{k'=0}^{+\infty} \widehat{\alpha}_p^{k'} < +\infty$ ,  $(\epsilon_k, \delta_k) \downarrow 0$ , and  $\delta_k/(\lambda + \ell_k) \downarrow 0$ . Recalling (34) in the proof of Lemma 2, at the  $k_0$ -th outer iteration, we have

$$\left\{ \left\| \sum_{p=1}^m (y_{p,1}^{k_0} v_{p,1}^{k_0} + y_{p,2}^{k_0} v_{p,2}^{k_0}) \right\| \leq \max \left\{ 1, \sum_{p=1}^m (|y_{p,1}^{k_0}| + |y_{p,2}^{k_0}|) \right\} \delta_{k_0}, \right. \\ \left. \left[ \begin{array}{l} v_{p,1}^{k_0} \in \partial g_p^{k_0}(x^{k_0, i_{k_0}}) - \partial h_p^{k_0}(x^{k_0, i_{k_0}}) \\ v_{p,2}^{k_0} \in \partial g_p^{k_0}(x^{k_0, i_{k_0} + 1}) - \partial h_p^{k_0}(x^{k_0, i_{k_0} + 1}) \end{array} \right] \text{ or } \left[ \begin{array}{l} v_{p,1}^{k_0} \in \partial g_p^{k_0}(x^{k_0, i_{k_0} + 1}) - \partial h_p^{k_0}(x^{k_0, i_{k_0} + 1}) \\ v_{p,2}^{k_0} \in \partial g_p^{k_0}(x^{k_0, i_{k_0}}) - \partial h_p^{k_0}(x^{k_0, i_{k_0}}) \end{array} \right], p = 1, \dots, m, \right.$$

where  $y_{p,1}^{k_0} \in \partial \varphi_p^\uparrow(f_p^{k_0, \text{upper}}(x^{k_0, i_{k_0} + 1}; x^{k_0, i_{k_0}}))$ ,  $y_{p,2}^{k_0} \in \partial \varphi_p^\downarrow(f_p^{k_0, \text{lower}}(x^{k_0, i_{k_0} + 1}; x^{k_0, i_{k_0}}))$  for  $p = 1, \dots, m$ .

Thus, at the point  $x^* = x^{k_0, i_{k_0} + 1}$ , we have

$$\text{dist} \left( 0, \sum_{p=1}^m \bigcup \left\{ y_{p,1} [\partial^\beta g_p^{k_0}(x^*) - \partial^\beta h_p^{k_0}(x^*)] + y_{p,2} [\partial^\beta g_p^{k_0}(x^*) - \partial^\beta h_p^{k_0}(x^*)] \mid \begin{array}{l} y_{p,1} \in \partial^\beta \varphi_p^\uparrow(f_p^{k_0}(x^*)), \\ y_{p,2} \in \partial^\beta \varphi_p^\downarrow(f_p^{k_0}(x^*)) \end{array} \right\} \right) \leq \eta,$$

where the parameters  $\beta$  and  $\eta$  are given by

$$\beta = \max \left\{ \begin{array}{l} \max_{1 \leq p \leq m} \left[ f_p^{k_0, \text{upper}}(x^*; x^{k_0, i_{k_0}}) - f_p^{k_0}(x^*) \right], \\ \max_{1 \leq p \leq m} \left[ f_p^{k_0}(x^*) - f_p^{k_0, \text{lower}}(x^*; x^{k_0, i_{k_0}}) \right], \\ \|x^* - x^{k_0, i_{k_0}}\| \end{array} \right\} \stackrel{(29)}{\leq} \max \left\{ \begin{array}{l} \max_{1 \leq p \leq m} \sum_{k'=k_0}^{+\infty} \widehat{\alpha}_p^{k'} + \epsilon_{k_0}, \\ \delta_{k_0}/(\lambda + \ell_{k_0}) \end{array} \right\} \stackrel{(39)}{\leq} \bar{\beta},$$

$$\eta = \max \left\{ 1, \sum_{p=1}^m (|y_{p,1}^{k_0}| + |y_{p,2}^{k_0}|) \right\} \delta_{k_0} \stackrel{\text{Proposition 5}}{\leq} \max\{1, \sqrt{2mD}\} \cdot \delta_{k_0} \stackrel{(39)}{\leq} \bar{\eta} \cdot \max\{1, \sqrt{2mD}\}.$$

Henceforth, for  $k_0$  satisfying (39),  $x^* = x^{k_0, i_{k_0} + 1}$  is a  $(\bar{\eta} \cdot \max\{1, \sqrt{2mD}\}, \bar{\beta}, \bar{k})$ -weakly A-stationary point of problem (CP<sub>1</sub>).  $\square$

## 5 Numerical examples.

We present some preliminary experiments to illustrate the performance of our algorithm on the inverse optimal value optimization with or without constraints. The first experiment aims to demonstrate the practical performance of the prox-ADC method under the termination criteria



in section 4.3, by varying different approximating sequences and initial points. To demonstrate the computation of ADC constrained problems, especially the choice of the quantity  $\widehat{\alpha}_p^k$  and a feasible initial point in Assumption 2, we further consider the constrained inverse optimal value optimization. These experiments were tested on a MacBook Air laptop with an Apple M1 chip and 16GB of memory using Julia 1.10.2.

### 5.1 Inverse optimal value optimization with simple constraints.

Based on the setting in (2), we aim to find a vector  $x \in [-1, 1]^n$  to minimize the errors between the observed optimal values  $\{\nu_p\}_{p=1}^m$  and true optimal values  $\{f_p(x)\}_{p=1}^m$ :

$$\underset{x \in [-1, 1]^n}{\text{minimize}} \quad F(x) \triangleq \sum_{p=1}^m |\nu_p - f_p(x)|, \quad (40)$$

where each  $f_p$  is the optimal value function as defined in (1). We fix  $n = 10$ ,  $m = 11$ ,  $d = 10$ , and the number of inequality constraints  $\ell = 5$  in the minimization problem (1). Vectors  $b^p$  and  $c^p$ , and matrices  $A^p, B^p, C^p$  are randomly generated with each entry independent and normally distributed with mean  $\mu = 0$  and variance  $\sigma = 1$ . For numerical stability, we then normalize matrices  $C^p$  and  $A^p$  by a factor of  $\sqrt{n}$ . We also generate a positive definite matrix  $Q^p$  and a random solution  $x^* = u/\|u\|$  with  $u \sim \text{Normal}(0, I_n)$ . We set  $\nu_p = f_p(x^*)$  for each  $p$  and, therefore,  $F(x^*) = 0$ .

We adopt the ADC decomposition in (6), denoted by  $f_p^k = (g_p)_{\gamma_k} - (h_p)_{\gamma_k}$  with a sequence  $\{\gamma_k = 1/(k+1)^\rho\}$  for some exponent  $\rho > 0$ . Consequently,  $\ell_k = 1/\gamma_k = (k+1)^\rho$ . We apply the prox-ADC algorithm to solve this example with  $\epsilon_k = \delta_k = 1/(k+1)^\rho$  and  $\lambda = 5$ . In this example, the strongly convex subproblem (30) can be easily reformulated to a problem with linear objective and convex quadratic constraints, which is solved by Gurobi in our experiments.

We first investigate the performance of our algorithm under the termination criteria with different values of parameters. Figure 2 displays the logarithm of the objective values against the number of outer iterations and the total number of inner iterations. We mark three different points on the curve where the termination criteria (39) with  $\bar{\eta} = \bar{\beta} = 10^{-1}, 10^{-2}, 10^{-3}$  and  $\bar{k} = 10, 20, 40$  are satisfied.

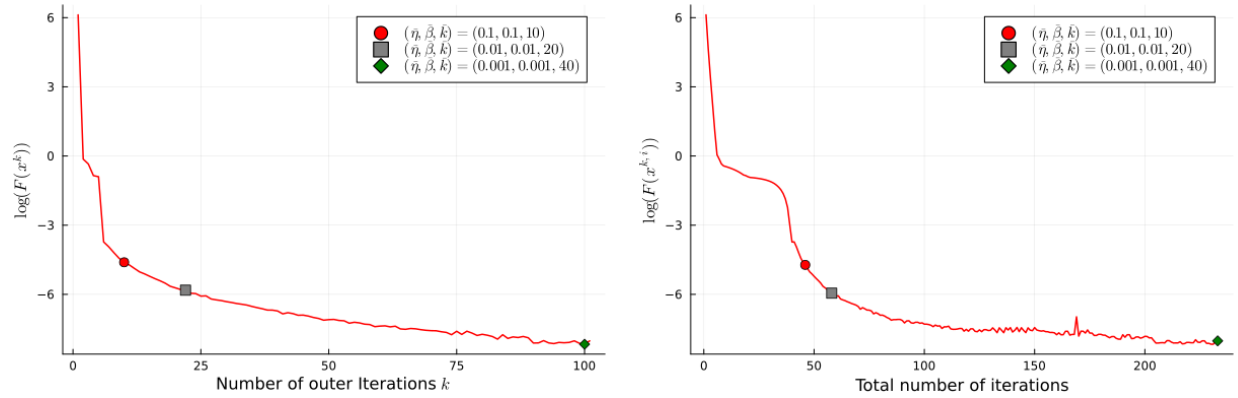


Figure 2: Performance of the prox-ADC method for problem (40), under the termination criteria (39) with  $\bar{\eta} = \bar{\beta} = 10^{-1}, 10^{-2}, 10^{-3}$  and  $\bar{k} = 10, 20, 40$ , for a fixed exponent  $\rho = 1.5$  and a fixed initial point.

We have also experimented with various values of exponent  $\rho$  that determine the convergence rate of the approximating sequence and various initial points. In both cases, we terminate the algorithm under the conditions (39) with  $\bar{\eta} = \bar{\beta} = 10^{-2}$  and  $\bar{k} = 10$ . In Figure 3, we observe that setting different values of  $\rho$  under the same termination criteria leads to candidate solutions with similar objective values, and there are roughly two phases of convergence in terms of the total number of iterations. Initially, the objective value decreases faster for smaller  $\rho$ , corresponding to poorer approximation. When the objective value is sufficiently small ( $10^{-3}$  on this particular instance), larger  $\rho$  results in faster convergence to high accuracy. We remark that for  $\rho = 0.5$ , the algorithm reaches the maximum number of outer iterations and does not output a  $(10^{-2}, 10^{-2}, 10)$ -weakly A-stationary point. Figure 4 demonstrates the influence of using various initial points that are uniformly distributed on  $[-1, 1]^n$ . On this instance, two of the initial points find  $(10^{-2}, 10^{-2}, 10)$ -weakly A-stationary points with large objective values. For these two initial points, we rerun the algorithm with  $\bar{\eta} = \bar{\beta} = 10^{-3}$ , and the algorithm still terminates with large objective values.

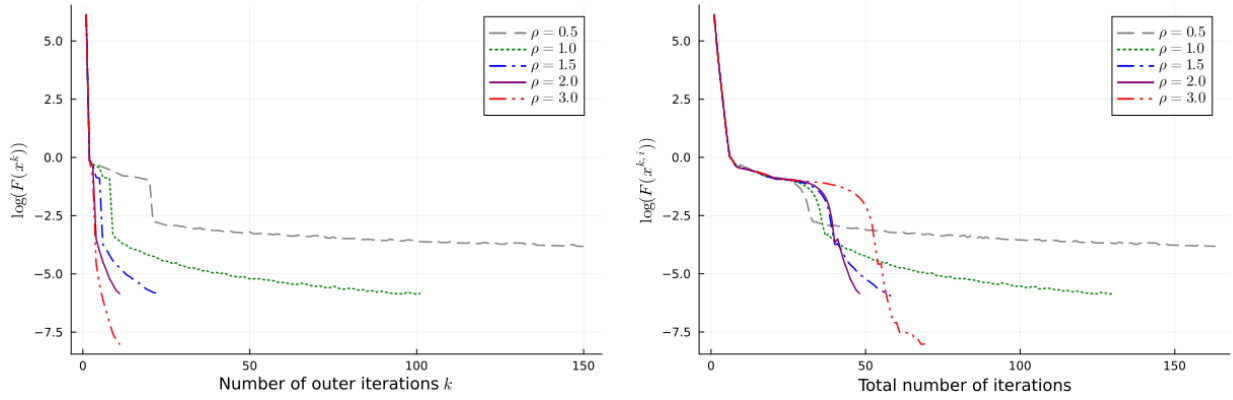


Figure 3: Performance of the prox-ADC method for problem (40) using different values of exponent  $\rho$ , under the termination criteria (39) with  $\bar{\eta} = \bar{\beta} = 10^{-2}$  and  $\bar{k} = 10$ , for a fixed initial point.

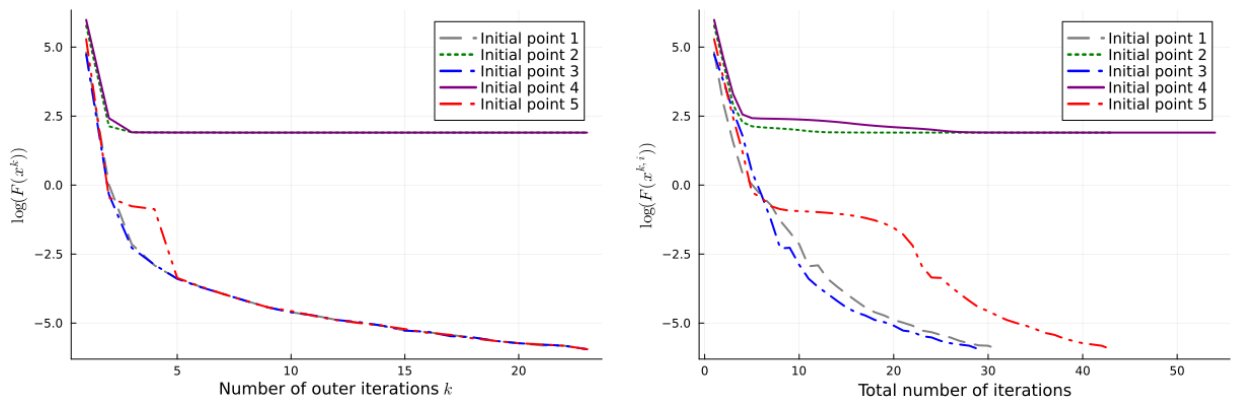


Figure 4: Performance of the prox-ADC method for problem (40) using five initial points uniformly distributed on  $[-1, 1]^n$ , under the termination criteria (39) with  $\bar{\eta} = \bar{\beta} = 10^{-2}$  and  $\bar{k} = 10$ , for a fixed exponent  $\rho = 1.5$ .

## 5.2 Inverse optimal value optimization with ADC constraints.

We consider a variant of the inverse optimal value optimization that is defined as follows:

$$\begin{aligned} & \underset{x \in [-1, 1]^n}{\text{minimize}} && F(x) = \sum_{p=1}^{m_1} |\nu_p - f_p(x)| \\ & \text{subject to} && \frac{\nu_p - f_p(x)}{\max\{1, |\nu_p|\}} \leq \varepsilon, \quad \frac{f_p(x) - \nu_p}{\max\{1, |\nu_p|\}} \leq \varepsilon, \quad p = m_1 + 1, \dots, m. \end{aligned} \quad (41)$$

In this formulation, the observations of the optimal values  $\{\nu_p\}_{p=1}^m$  are divided into two groups indexed by  $\{1, \dots, m_1\}$  and  $\{m_1 + 1, \dots, m\}$ . We aim to minimize the errors for the first group while ensuring the relative errors for the second group do not exceed a specified feasibility tolerance, denoted by  $\varepsilon$ . In our experiment, we fix  $n = 10$ ,  $m = 11$ ,  $m_1 = 8$ ,  $\varepsilon = 10^{-1}$ ,  $d = 10$ , and the number of inequality constraints  $\ell = 5$  in the minimization problem (1). The solution  $x^*$  and the data, including  $\{\nu_p\}_{p=1}^m$ , are randomly generated in the same way as in section 5.1. We can see that  $x^*$  is feasible to (41) and attains the minimal objective value  $F(x^*) = 0$ .

Similar to the first example, we adopt the ADC decomposition in (6), denoted by  $f_p^k = (g_p)_{\gamma_k} - (h_p)_{\gamma_k}$  with a sequence  $\{\gamma_k = 1/(k + \tilde{k})^\rho\}$  for some positive integer  $\tilde{k}$  and  $\rho > 0$ . Due to the feasibility problem in Assumption 2, we introduce the additional parameter  $\tilde{k}$  to control the approximating sequences, which will be explained in details later. We also note that treating  $\frac{\nu_p - f_p(x)}{\max\{1, |\nu_p|\}} \leq \varepsilon$  and  $\frac{f_p(x) - \nu_p}{\max\{1, |\nu_p|\}} \leq \varepsilon$  as two separate constraints for  $p = m_1 + 1, \dots, m$  leads to the failure of the asymptotic constraint qualification in Assumption 5 because the approximate subdifferentials of the ADC functions  $f_p$  and  $-f_p$  are linearly dependent. This issue can be resolved by rewriting the constraints in a composite ADC form  $\frac{|\nu_p - f_p(x)|}{\max\{1, |\nu_p|\}} \leq \varepsilon$  and assuming a corresponding version of Assumption 5. We omit this technical detail since the main focus of this section is to illustrate the practical implementation of our algorithm.

To verify Assumption 2 that states the existence of a strictly feasible point, we first follow the discussion after Assumption 2 to construct the quantity  $\widehat{\alpha}_p^k = (\gamma_k - \gamma_{k+1})L_p^2 / (2 \max\{1, |\nu_p|\})$  where  $L_p$  is the Lipschitz constant of  $\bar{f}_p(\cdot, x)$  for all  $x \in [-1, 1]^n$ . We can derive the Lipschitz constant  $L_p$  by characterizing the subdifferential  $\partial_1 \bar{f}_p(\cdot, x)$  for a fixed  $x$  based on Danskin's Theorem [11, Theorem 2.1] and then upper bounding the norm of this subdifferential over  $x \in [-1, 1]^n$ . The extra denominator  $\max\{1, |\nu_p|\}$  in the expression of  $\widehat{\alpha}_p^k$  is due to the scaling of the constraints in (41). Then, consider the following problem:

$$\underset{x \in [-1, 1]^n}{\text{minimize}} \quad V(x) \triangleq \sum_{p=m_1+1}^m \max \left\{ 0, |\nu_p - f_p^0(x)| - \underbrace{\left( \varepsilon - \sum_{k'=0}^{+\infty} \widehat{\alpha}_p^{k'} \right) \max\{1, |\nu_p|\}}_{\triangleq s_p \text{ (constant)}} \right\}, \quad (\text{Feas})$$

where the objective is the sum of the compositions of univariate convex functions  $\varphi_p(t) = \max\{0, |\nu_p - t| - s_p\}$  and DC functions  $f_p^0$ . Notice that problem (Feas) takes the same form as (23). Thus, we can apply the inner loop of the prox-ADC method to solve it approximately. If solving this problem gives a solution  $x^0$  with  $V(x^0) = 0$ , then

$$\frac{|\nu_p - f_p^0(x)|}{\max\{1, |\nu_p|\}} \leq \varepsilon - \sum_{k'=0}^{+\infty} \widehat{\alpha}_p^{k'} = \varepsilon - \frac{(L_p)^2}{2 \tilde{k}^\rho \cdot \max\{1, |\nu_p|\}}, \quad (42)$$

and  $x^0$  is a strictly feasible point satisfying Assumption 2. We emphasize that using the inner loop of the prox-ADC method for solving problem (Feas) to obtain a strictly feasible point is merely a heuristic. Although this approach works well in our experiments, it is generally not easy to verify Assumption 2. We make a final remark on the role of  $\tilde{k}$ . For small values of  $\rho$  and  $\tilde{k}$ , it is possible that  $\varepsilon < \frac{(L_p)^2}{2\tilde{k}^\rho \cdot \max\{1, |\nu_p|\}}$ , and, from (42), there is no strictly feasible point satisfying Assumption 2 for this fixed approximating sequence. Henceforth, the flexibility of the parameter  $\tilde{k}$  is necessary to ensure the validity of Assumption 2.

We implement the above procedure to find an initial point and then apply the prox-ADC method with  $\epsilon_k = \delta_k = 1/(k+1)^\rho$  and  $\lambda = 5$ . On most of the randomly generated instances, we observe that the point given by solving (Feas) is also feasible to the original problem (41) along the iterations, although this result cannot be implied by Assumption 2. In Figure 5, we again plot the logarithm of the objective values against the total number of iterations, using various combination of  $\tilde{k}$  and  $\rho$  and various initial points. It is worth mentioning that for this constrained problem, the random initial point is not directly utilized in the prox-ADC method. Instead, it is first used in problem (Feas) to generate a strictly feasible point satisfying Assumption 2, and the candidate solution for (Feas) then becomes the initial point of the prox-ADC method for solving (41).

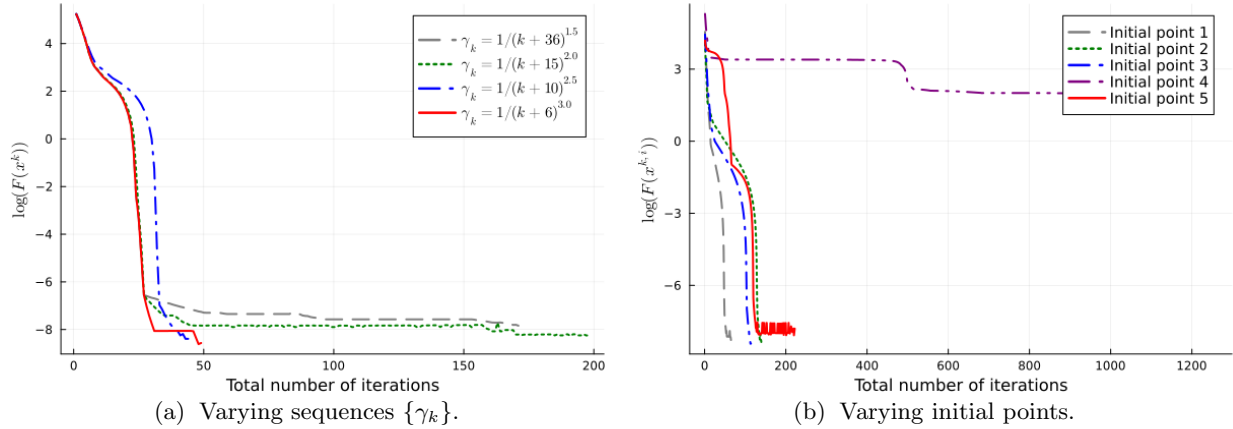


Figure 5: Performance of the prox-ADC method for problem (41) under the termination criteria (39) with  $\bar{\eta} = \bar{\beta} = 2 \times 10^{-2}$  and  $\tilde{k} = 5$ . (a): using different sequences  $\{\gamma_k\}$  for a fixed initial point. (b): using five initial points uniformly distributed on  $[-1, 1]^n$  for a fixed sequence  $\{\gamma_k = 1/(k+10)^{2.5}\}$ .

## Acknowledgments.

The authors are partially supported by the National Science Foundation under grants CCF-2416172 and DMS-2416250, and the National Institutes of Health under grant 1R01CA287413-01. The authors are grateful to the associate editor and two reviewers for their careful reading and constructive comments that have substantially improved the paper.

## References

- [1] Carlo Acerbi. Spectral measures of risk: A coherent representation of subjective risk aversion. *Journal of Banking and Finance*, 26(7):1505–1518, 2002.
- [2] Shabbir Ahmed and Yongpei Guan. The inverse optimal value problem. *Mathematical Programming*, 102(1):91–110, 2005.
- [3] Amir Beck and Marc Teboulle. Smoothing and first order methods: A unified framework. *SIAM Journal on Optimization*, 22(2):557–580, 2012.
- [4] James V Burke. Descent methods for composite nondifferentiable optimization problems. *Mathematical Programming*, 33(3):260–279, 1985.
- [5] James V Burke and Michael C Ferris. A Gauss-Newton method for convex composite optimization. *Mathematical Programming*, 71(2):179–194, 1995.
- [6] James V Burke and Tim Hoheisel. Epi-convergent smoothing with applications to convex composite functions. *SIAM Journal on Optimization*, 23(3):1457–1479, 2013.
- [7] James V Burke and Tim Hoheisel. Epi-convergence properties of smoothing by infimal convolution. *Set-Valued and Variational Analysis*, 25(1):1–23, 2017.
- [8] James V Burke, Tim Hoheisel, and Christian Kanzow. Gradient consistency for integral-convolution smoothing functions. *Set-Valued and Variational Analysis*, 21(2):359–376, 2013.
- [9] James V Burke, Adrian S Lewis, and Michael L Overton. A robust gradient sampling algorithm for nonsmooth, nonconvex optimization. *SIAM Journal on Optimization*, 15(3):751–779, 2005.
- [10] Xiaojun Chen. Smoothing methods for nonsmooth, nonconvex minimization. *Mathematical Programming*, 134(1):71–99, 2012.
- [11] Frank H Clarke. Generalized gradients and applications. *Transactions of the American Mathematical Society*, 205:247–262, 1975.
- [12] Frank H Clarke. *Optimization and Nonsmooth Analysis*. Society for Industrial and Applied Mathematics, Philadelphia, 1990.
- [13] Ying Cui and Jong-Shi Pang. *Modern Nonconvex Nondifferentiable Optimization*. Society for Industrial and Applied Mathematics, Philadelphia, 2021.
- [14] Dmitriy Drusvyatskiy and Courtney Paquette. Efficiency of minimizing compositions of convex functions and smooth maps. *Mathematical Programming*, 178(1–2):503–558, 2019.
- [15] Yuri M Ermoliev, Vladimir I Norkin, and Roger JB Wets. The minimization of semicontinuous functions: Mollifier subgradients. *SIAM Journal on Control and Optimization*, 33(1):149–167, 1995.
- [16] Roger Fletcher. A model algorithm for composite nondifferentiable optimization problems. *Mathematical Programming Studies*, 17:67–76, 1982.

- [17] Gerald B Folland. *Real Analysis: Modern Techniques and Their Applications*, volume 40. John Wiley & Sons, New York, 1999.
- [18] Allen A Goldstein. Optimization of Lipschitz continuous functions. *Mathematical Programming*, 13(1):14–22, 1977.
- [19] Adrian S Lewis and Stephen J Wright. A proximal method for composite minimization. *Mathematical Programming*, 158(1-2):501–546, 2016.
- [20] Hanyang Li and Ying Cui. A decomposition algorithm for two-stage stochastic programs with nonconvex recourse. *SIAM Journal on Optimization*, 34(1):306–335, 2024.
- [21] Junyi Liu, Ying Cui, Jong-Shi Pang, and Suvrajeet Sen. Two-stage stochastic programming with linearly bi-parameterized quadratic recourse. *SIAM Journal on Optimization*, 30(3):2530–2558, 2020.
- [22] Boris S Mordukhovich. Generalized differential calculus for nonsmooth and set-valued mappings. *Journal of Mathematical Analysis and Applications*, 183(1):250–288, 1994.
- [23] Matthew Norton, Valentyn Khokhlov, and Stan Uryasev. Calculating CVaR and bPOE for common probability distributions with application to portfolio optimization and density estimation. *Annals of Operations Research*, 299(1):1281–1315, 2021.
- [24] Giuseppe Paleologo and Samer Takriti. Bandwidth trading: A new market looking for help from the OR community. *AIRO News*, 6(3):1–4, 2001.
- [25] RA Poliquin and R Tyrrell Rockafellar. Amenable functions in optimization. *Nonsmooth Optimization: Methods and Applications (Erice, 1991)*, pages 338–353, 1992.
- [26] RA Poliquin and R Tyrrell Rockafellar. A calculus of epi-derivatives applicable to optimization. *Canadian Journal of Mathematics*, 45(4):879–896, 1993.
- [27] R Tyrrell Rockafellar. *Convex Analysis*, volume 18. Princeton University Press, Princeton, NJ, 1970.
- [28] R Tyrrell Rockafellar. Lagrange multipliers and optimality. *SIAM Review*, 35(2):183–238, 1993.
- [29] R Tyrrell Rockafellar and Stanislav Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2(3):21–42, 2000.
- [30] R Tyrrell Rockafellar and Roger JB Wets. *Variational Analysis*, volume 317. Springer Science & Business Media, New York, 2009.
- [31] Johannes O Royset. Approximations of semicontinuous functions with applications to stochastic optimization and statistical estimation. *Mathematical Programming*, 184(1):289–318, 2020.
- [32] Johannes O Royset. Consistent approximations in composite optimization. *Mathematical Programming*, 201(1–2):339–372, 2022.

- [33] Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory (Third Edition)*. Society for Industrial and Applied Mathematics, Philadelphia, 2021.
- [34] Wim van Ackooij. A discussion of probability functions and constraints from a variational perspective. *Set-Valued and Variational Analysis*, 28(4):585–609, 2020.
- [35] Yao Yao, Qihang Lin, and Tianbao Yang. Large-scale optimization of partial AUC in a range of false positive rates. *Advances in Neural Information Processing Systems*, 35:31239–31253, 2022.

## Appendix A. Proofs of Proposition 2 and Proposition 3.

*Proof of Proposition 2.* (a) We first generalize the convergence result of the classical Moreau envelopes when  $\gamma_k \downarrow 0$  (see, e.g., [30, Theorem 1.25]) to the partial Moreau envelopes. Fixing any  $\gamma_0 > 0$ , we consider the function  $\psi(z, x, \gamma) \triangleq \bar{f}(z, x) + \delta_{\text{dom } f}(x) + \psi_0(z, x, \gamma)$  with

$$\psi_0(z, x, \gamma) \triangleq \begin{cases} \|z - x\|^2 / (2\gamma) & \text{if } \gamma \in (0, \gamma_0], \\ 0 & \text{if } \gamma = 0, z = x, \\ \infty & \text{otherwise.} \end{cases}$$

Notice that  $f^k(x) = g_{\gamma_k}(x) - h_{\gamma_k}(x) + \delta_{\text{dom } f}(x) = \inf_z \psi(z, x, \gamma_k)$ . It is easy to verify that  $\psi$  is proper and lsc based on our assumptions. Under the assumption that  $\bar{f}$  is bounded from below on  $\text{dom } f \times \text{dom } f$ , we can also show by contradiction that  $\psi(z, x, \gamma)$  is level-bounded in  $z$  locally uniformly in  $(x, \gamma)$ . Consequently, it follows from [30, Theorem 1.17] that  $f^k(x) = \inf_z \psi(z, x, \gamma_k) \uparrow f(x)$  for any fixed  $x$  and each  $f^k$  is lsc.

Hence,  $f^k \xrightarrow{c} f$  is a direct consequence of [30, Proposition 7.4(d)] by  $f^k(x) \uparrow f(x)$  for all  $x$  and the lower semicontinuity of  $f^k$ . If  $\text{dom } f = \mathbb{R}^n$ , then  $f$  is continuous, and thus  $f^k \xrightarrow{c} f$  by [30, Proposition 7.4(c-d)]. We then complete the proof of (a).

(b) For any  $\bar{x} \in \text{int}(\text{dom } f)$ ,

$$\begin{aligned} \partial_A f(\bar{x}) &= \bigcup_{x^k \rightarrow \bar{x}} \text{Lim sup}_{k \rightarrow +\infty} \left\{ \partial g_{\gamma_k}(x^k) - \partial h_{\gamma_k}(x^k) \right\} \\ &\stackrel{(i')}{=} \bigcup_{x^k \rightarrow \bar{x}} \text{Lim sup}_{k \rightarrow +\infty} \left\{ \frac{x^k}{\gamma_k} - \partial_2(-\bar{f})(z^k, x^k) - \frac{z^k}{\gamma_k} \mid z^k = \arg \min_{z \in \mathbb{R}^n} \left[ \bar{f}(z, x^k) + \frac{1}{2\gamma_k} \|z - x^k\|^2 \right] \right\} \\ &\stackrel{(ii')}{\subset} \bigcup_{(x^k, z^k) \rightarrow (\bar{x}, \bar{x})} \text{Lim sup}_{k \rightarrow +\infty} \left[ \partial_1 \bar{f}(z^k, x^k) - \partial_2(-\bar{f})(z^k, x^k) \right] \\ &\stackrel{(iii')}{=} \partial_1 \bar{f}(\bar{x}, \bar{x}) - \partial_2(-\bar{f})(\bar{x}, \bar{x}), \end{aligned}$$

where (i') follows from the convexity of  $(-\bar{f})(z, \cdot)$  for any  $z \in \text{dom } f$  and Danskin's Theorem [11, Theorem 2.1]; (ii') is due to the optimality condition for  $z^k$ , and  $z^k \rightarrow \bar{x}$  is obtained by similar arguments in the proof of Theorem 1(b) due to our assumption that  $\bar{f}$  is bounded from below on  $\text{dom } f \times \text{dom } f$ ; and (iii') uses the outer semicontinuity of  $\partial_1 \bar{f}$  and  $\partial_2(-\bar{f})$  at  $(\bar{x}, \bar{x})$  [20, Lemma 5]. Therefore, for any  $\bar{x} \in \text{int}(\text{dom } f)$ ,  $\partial f(\bar{x}) \subset \partial_A f(\bar{x}) \subset \partial_1 \bar{f}(\bar{x}, \bar{x}) - \partial_2(-\bar{f})(\bar{x}, \bar{x})$ . Moreover, due to the local boundedness of the mappings  $\partial_1 \bar{f}$  and  $\partial_2(-\bar{f})$  at  $(\bar{x}, \bar{x})$  [20, Lemma 5], it follows from [30, Example 4.22] that  $\partial_A^\infty f(\bar{x}) = \{0\}$ .  $\square$

*Proof of Proposition 3.* (a) Note that for any  $x \in \mathbb{R}^n$ ,  $\text{CVaR}_\alpha[c(x, Z)]$  is well-defined and takes finite value due to  $\mathbb{E}[|c(x, Z)|] < +\infty$ . Since  $c(x, Z)$  follows a continuous distribution for any  $x \in \mathbb{R}^n$ , we know from [29, Theorem 1] and [1] that CVaR has the following equivalent representations:

$$\text{CVaR}_\alpha[c(x, Z)] = \inf_{t \in \mathbb{R}} \left\{ t + \frac{1}{1-\alpha} \mathbb{E}[\max\{c(x, Z) - t, 0\}] \right\} = \frac{1}{1-\alpha} \int_\alpha^1 \text{VaR}_t[c(x, Z)] dt.$$

Moreover,  $\text{CVaR}_\alpha[c(\cdot, Z)]$  is convex by the convexity of  $c(\cdot, z)$  for any fixed  $z \in \mathbb{R}^m$  (cf. [29, Theorem 2]). Therefore, both  $g^k$  and  $h^k$  defined in (7) are convex. By the definitions of  $g^k$  and  $h^k$ , we have

$$\begin{aligned} g^k(x) - h^k(x) &= [k(1-\alpha) + 1] \text{CVaR}_{\alpha-1/k}[c(x, Z)] - k(1-\alpha) \text{CVaR}_\alpha[c(x, Z)] \\ &= \frac{k(1-\alpha) + 1}{1 - (\alpha - 1/k)} \int_{\alpha-1/k}^1 \text{VaR}_t[c(x, Z)] dt - \frac{k(1-\alpha)}{1-\alpha} \int_\alpha^1 \text{VaR}_t[c(x, Z)] dt \\ &= k \int_{\alpha-1/k}^1 \text{VaR}_t[c(x, Z)] dt - k \int_\alpha^1 \text{VaR}_t[c(x, Z)] dt \\ &= k \int_{\alpha-1/k}^\alpha \text{VaR}_t[c(x, Z)] dt. \end{aligned}$$

Note that  $\text{VaR}_t[c(x, Z)]$  is nondecreasing as a function of  $t$  for any fixed  $x \in \mathbb{R}^n$ . Namely,

$$\int_{\alpha-1/k}^\alpha \text{VaR}_{\alpha-1/k}[c(x, Z)] dt \leq \int_{\alpha-1/k}^\alpha \text{VaR}_t[c(x, Z)] dt \leq \int_{\alpha-1/k}^\alpha \text{VaR}_\alpha[c(x, Z)] dt.$$

Thus,  $\text{VaR}_{\alpha-1/k}[c(x, Z)] \leq g^k(x) - h^k(x) \leq \text{VaR}_\alpha[c(x, Z)]$  for any  $x \in \mathbb{R}^n$  and  $k > 1/\alpha$ . Since  $\text{VaR}_t[c(x, Z)]$  as a function of  $t$  on  $(0, 1)$  is left-continuous, it follows that  $[g^k(x) - h^k(x)] \uparrow \text{VaR}_\alpha[c(x, Z)]$  for all  $x$ . Observe that

$$\{x \mid \text{VaR}_\alpha[c(x, Z)] \leq r\} = \{x \mid \mathbb{P}(c(x, Z) \leq r) \geq \alpha\}.$$

Based on our assumptions and [34, Proposition 2.2], for any  $r \in \mathbb{R}$ , the probability function  $x \mapsto -\mathbb{P}(c(x, Z) \leq r)$  is lsc, which implies the closedness of the level set  $\{x \mid \mathbb{P}(c(x, Z) \leq r) \geq \alpha\}$  for any  $(r, \alpha) \in \mathbb{R} \times (0, 1)$ . Hence,  $\text{VaR}_\alpha[c(\cdot, Z)]$  is lsc for any given  $\alpha \in (0, 1)$  and is continuous if  $c(\cdot, \cdot)$  is further assumed to be continuous. Then (a) is a direct consequence of [30, Proposition 7.4(c-d)] by the monotonicity  $[g^k(x) - h^k(x)] \uparrow \text{VaR}_\alpha[c(x, Z)]$  and the continuity of  $\text{VaR}_\alpha[c(\cdot, Z)]$ .

(b) We use  $\mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})$  to denote the space of all random variables  $X : \Omega \rightarrow \mathbb{R}$  with  $\mathbb{E}[|X(\omega)|] < +\infty$ . According to [33, Example 6.19], the function  $\text{CVaR}_\alpha : \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}$  is subdifferentiable (see [33, (9.281)] for the definition). Consider any fixed  $x \in \mathbb{R}^n$ . Given that  $c(x, Z)$  is a continuous random variable in  $\mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})$ , it follows from [33, (6.81)] that the subdifferential of  $\text{CVaR}_\alpha[\cdot]$  at  $c(x, Z)$  is:

$$\partial(\text{CVaR}_\alpha[\cdot])[c(x, Z)] = \left\{ \phi \in \mathcal{L}_\infty(\Omega, \mathcal{F}, \mathbb{P}) \left| \begin{array}{ll} \phi(\omega) = (1-\alpha)^{-1} & \text{if } c(x, Z(\omega)) > \text{VaR}_\alpha[c(x, Z)] \\ \phi(\omega) \in [0, (1-\alpha)^{-1}] & \text{if } c(x, Z(\omega)) = \text{VaR}_\alpha[c(x, Z)] \\ \phi(\omega) = 0 & \text{if } c(x, Z(\omega)) < \text{VaR}_\alpha[c(x, Z)] \end{array} \right. \right\}. \quad (43)$$

We would like to mention that the event  $\{\omega \in \Omega \mid c(x, Z(\omega)) = \text{VaR}_\alpha[c(x, Z)]\}$  has zero probability and, thus,  $\mathbb{E}[\phi] = (1-\alpha)^{-1} \cdot (1-\alpha) = 1$  for every random variable  $\phi \in \partial(\text{CVaR}_\alpha[\cdot])[c(x, Z)]$ .



Let  $\mathbb{P}_Z$  denote the probability measure associated with  $Z$ . By using [33, Theorem 6.14], we obtain the subdifferential of the convex function  $\text{CVaR}_\alpha[c(\cdot, Z)]$  at  $x$ :

$$\begin{aligned}
\partial(\text{CVaR}_\alpha[c(\cdot, Z)])(x) &= \text{cl} \left( \bigcup_{\phi \in \partial(\text{CVaR}_\alpha[\cdot])[c(x, Z)]} \int \partial_1 c(x, Z(\omega)) \phi(\omega) d\mathbb{P}_Z(\omega) \right) \\
&\stackrel{(\text{iv}')}{=} \text{cl} \left( \int \partial_1 c(x, Z(\omega)) \bar{\phi}(\omega) d\mathbb{P}_Z(\omega) \right) \quad \forall \bar{\phi} \in \partial(\text{CVaR}_\alpha[\cdot])[c(x, Z)] \\
&\stackrel{(\text{v}')}{=} \int \partial_1 c(x, Z(\omega)) \bar{\phi}(\omega) d\mathbb{P}_Z(\omega) \quad \forall \bar{\phi} \in \partial(\text{CVaR}_\alpha[\cdot])[c(x, Z)].
\end{aligned} \tag{44}$$

To see (iv'), it suffices to show that, for arbitrary two elements  $\phi_1$  and  $\phi_2$  in the set  $\partial(\text{CVaR}_\alpha[\cdot])[c(x, Z)]$ , we have

$$\int \partial_1 c(x, Z(\omega)) \phi_1(\omega) d\mathbb{P}_Z(\omega) = \int \partial_1 c(x, Z(\omega)) \phi_2(\omega) d\mathbb{P}_Z(\omega). \tag{45}$$

To this end, we take any measurable selection  $a(x, Z(\omega)) \in \partial_1 c(x, Z(\omega))$ . By the assumption that  $|c(x, z) - c(x', z)| \leq \kappa(z)\|x - x'\|$  for all  $x, x' \in \mathbb{R}^n$  and  $z \in \mathbb{R}^m$ , it holds that  $\|a(x, Z)\| \leq \kappa(Z)$  since subgradients of a convex function are uniformly bounded in norm by the Lipschitz constant. Consequently, both  $a(x, Z(\omega)) \phi_1(\omega)$  and  $a(x, Z(\omega)) \phi_2(\omega)$  are integrable as  $|\phi_1(\omega)| \leq (1 - \alpha)^{-1}$  and  $|\phi_2(\omega)| \leq (1 - \alpha)^{-1}$  for any  $\omega$  by (43) and  $\mathbb{E}[|a(x, Z)|] \leq \mathbb{E}[\kappa(Z)] < +\infty$  by our assumption. Observing that  $a(x, Z(\omega)) \phi_1(\omega) = a(x, Z(\omega)) \phi_2(\omega)$  almost surely, we can conclude from [17, Proposition 2.23] that  $\int a(x, Z(\omega)) \phi_1(\omega) d\mathbb{P}_Z(\omega) = \int a(x, Z(\omega)) \phi_2(\omega) d\mathbb{P}_Z(\omega)$ . This completes the proof of (45).

Next, we will explain why the closure can be removed in (44). By the convexity of  $c(\cdot, z)$  for any fixed  $z \in \mathbb{R}^m$  and the existence of a measurable function  $\kappa$ , it follows from [12, Theorem 2.7.2] that

$$\int \partial_1 c(x, Z(\omega)) \bar{\phi}(\omega) d\mathbb{P}_Z(\omega) = \partial \left( \int c(\cdot, Z(\omega)) \bar{\phi}(\omega) d\mathbb{P}_Z(\omega) \right) (x),$$

where the right-hand-side is the subdifferential of a convex function and, thus, is a closed set. Then, we can omit the closure to obtain the equation (v') in (44).

Now we use the expression of  $\partial(\text{CVaR}_\alpha[c(\cdot, Z)])(x)$  to characterize  $\partial_A \text{VaR}_\alpha[c(\cdot, Z)](\bar{x})$ . For any  $k > 1/\alpha$ , taking any  $\phi_3 \in \partial(\text{CVaR}_{\alpha-1/k}[\cdot])[c(x, Z)]$  and  $\phi_4 \in \partial(\text{CVaR}_\alpha[\cdot])[c(x, Z)]$ , we have

$$\begin{aligned}
\partial g^k(x) - \partial h^k(x) &= [k(1 - \alpha) + 1] \partial \text{CVaR}_{\alpha-1/k}[c(\cdot, Z)](x) - k(1 - \alpha) \partial \text{CVaR}_\alpha[c(\cdot, Z)](x) \\
&\stackrel{(44)}{=} \int \partial_1 c(x, Z(\omega)) \cdot ([k(1 - \alpha) + 1] \phi_3(\omega) - k(1 - \alpha) \phi_4(\omega)) d\mathbb{P}_Z(\omega) \\
&\stackrel{(43)}{=} \int \partial_1 c(x, Z(\omega)) \cdot \phi(\omega) d\mathbb{P}_Z(\omega),
\end{aligned}$$

with

$$\phi(\omega) \triangleq \begin{cases} 0 & \text{if } c(x, Z(\omega)) > \text{VaR}_\alpha[c(x, Z)] \text{ or } c(x, Z(\omega)) < \text{VaR}_{\alpha-1/k}[c(x, Z)] \\ [0, k] & \text{if } c(x, Z(\omega)) = \text{VaR}_\alpha[c(x, Z)] \text{ or } c(x, Z(\omega)) = \text{VaR}_{\alpha-1/k}[c(x, Z)] \\ k & \text{if } \text{VaR}_{\alpha-1/k}[c(x, Z)] < c(x, Z(\omega)) < \text{VaR}_\alpha[c(x, Z)] \end{cases}.$$

Since the event  $\{\omega \in \Omega \mid c(x, Z(\omega)) = \text{VaR}_\alpha[c(x, Z)] \text{ or } \text{VaR}_{\alpha-1/k}[c(x, Z)]\}$  has zero probability, we have

$$\begin{aligned} \partial g^k(x) - \partial h^k(x) &= \int \partial_1 c(x, Z(\omega)) k \mathbf{1}\{\text{VaR}_{\alpha-1/k}[c(x, Z)] < c(x, Z) < \text{VaR}_\alpha[c(x, Z)]\} d\mathbb{P}_Z(\omega) \\ &= \int \partial_1 c(x, Z(\omega)) \cdot \frac{\mathbf{1}\{\text{VaR}_{\alpha-1/k}[c(x, Z)] < c(x, Z) < \text{VaR}_\alpha[c(x, Z)]\}}{\mathbb{P}(\text{VaR}_{\alpha-1/k}[c(x, Z)] < c(x, Z) < \text{VaR}_\alpha[c(x, Z)])} d\mathbb{P}_Z(\omega) \\ &= \mathbb{E}[\partial_1 c(x, Z) \mid \text{VaR}_{\alpha-1/k}[c(x, Z)] < c(x, Z) < \text{VaR}_\alpha[c(x, Z)]] . \end{aligned}$$

By the definition of the approximate subdifferential, the proof is then completed.  $\square$

## Appendix B. Proof of Proposition 4.

We start with the chain rules for  $\partial(\varphi \circ f)$  and  $\partial^\infty(\varphi \circ f)$ , where the inner function  $f$  is merely lsc. These results are extensions of the nonlinear rescaling [30, Proposition 10.19(b)] to the case where  $\varphi$  may lack the strictly increasing property at a given point. One can also derive the same results through a general chain rule of the coderivative for composite set-valued mappings [22, Theorem 5.1]. However, to avoid the complicated computations accompanied by the introduction of the coderivative, we give an alternative proof below that is more straightforward. To prepare for the chain rules, we need a technical lemma about the proximal normal cone.

**Lemma 3.** *Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be a lsc function. For  $\bar{\alpha} > f(\bar{x})$ , it holds that*

$$\mathcal{N}_{\text{epi } f}^p(\bar{x}, \bar{\alpha}) \subset \mathcal{N}_{\text{epi } f}^p(\bar{x}, f(\bar{x})).$$

*Proof of Lemma 3.* For  $\bar{\alpha} > f(\bar{x})$ . By [30, Example 6.16], we have

$$\mathcal{N}_{\text{epi } f}^p(\bar{x}, \bar{\alpha}) = \left\{ \lambda [(x, \alpha) - (\bar{x}, \bar{\alpha})] \mid (x, \alpha) \in \mathbb{R}^{n+1} \text{ such that } (\bar{x}, \bar{\alpha}) \in \Pi_{\text{epi } f}(x, \alpha), \lambda \geq 0 \right\},$$

where  $\Pi_{\text{epi } f} : \mathbb{R}^{n+1} \rightarrow \text{epi } f$  is the projection operator. For any  $(x, \alpha) \in \mathbb{R}^{n+1}$  with  $(\bar{x}, \bar{\alpha}) \in \Pi_{\text{epi } f}(x, \alpha)$ , we have

$$(\bar{x}, \bar{\alpha}) \in \underset{(u, t) \in \text{epi } f}{\text{argmin}} \|(u, t) - (x, \alpha)\|^2,$$

which can be equivalently written as

$$(\bar{x}, f(\bar{x})) \in \underset{(u, t + \bar{\alpha} - f(\bar{x})) \in \text{epi } f}{\text{argmin}} \|(u, t) - (x, \alpha - \bar{\alpha} + f(\bar{x}))\|^2.$$

Then restrict the feasible region of the above problem to a subset  $\{(u, t) \mid (u, t) \in \text{epi } f\}$ . Since  $(\bar{x}, f(\bar{x}))$  is still a feasible point in this subset, we have  $(\bar{x}, f(\bar{x})) \in \Pi_{\text{epi } f}(x, \alpha - \bar{\alpha} + f(\bar{x}))$ . Hence,  $(\bar{x}, \bar{\alpha}) \in \Pi_{\text{epi } f}(x, \alpha)$  implies  $(\bar{x}, f(\bar{x})) \in \Pi_{\text{epi } f}(x, \alpha - \bar{\alpha} + f(\bar{x}))$ . Using this result and the expression of  $\mathcal{N}_{\text{epi } f}^p(\bar{x}, \bar{\alpha})$ , we conclude that  $\mathcal{N}_{\text{epi } f}^p(\bar{x}, \bar{\alpha}) \subset \mathcal{N}_{\text{epi } f}^p(\bar{x}, f(\bar{x}))$  for  $\bar{\alpha} > f(\bar{x})$ .  $\square$

We present the chain rules with a self-contained proof in the following lemma.

**Lemma 4** (chain rules for the limiting subdifferential). *Let  $\varphi : \mathbb{R} \rightarrow \overline{\mathbb{R}}$  be proper, lsc, convex, and nondecreasing with  $\sup \varphi = +\infty$ , and  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be lsc. Consider  $\bar{x} \in \text{dom}(\varphi \circ f)$ . If the only scalar  $y \in \text{Lim sup}_{x \rightarrow (\varphi \circ f)\bar{x}} \mathcal{N}_{\text{dom } \varphi}(f(x))$  with  $0 \in y \cdot \text{Lim sup}_{x \rightarrow \bar{x}} \partial f(x)$  is  $y = 0$ , then*

$$\begin{aligned} \partial(\varphi \circ f)(\bar{x}) &\subset \bigcup \left\{ y \cdot \text{Lim sup}_{x \rightarrow \bar{x}} \partial f(x) \mid y \in \text{Lim sup}_{x \rightarrow (\varphi \circ f)\bar{x}} \partial \varphi(f(x)) \right\} \cup \left[ \left( \text{Lim sup}_{x \rightarrow \bar{x}}^\infty \partial f(x) \right) \setminus \{0\} \right], \\ \partial^\infty(\varphi \circ f)(\bar{x}) &\subset \bigcup \left\{ y \cdot \text{Lim sup}_{x \rightarrow \bar{x}} \partial f(x) \mid y \in \text{Lim sup}_{x \rightarrow (\varphi \circ f)\bar{x}} \mathcal{N}_{\text{dom } \varphi}(f(x)) \right\} \cup \left[ \left( \text{Lim sup}_{x \rightarrow \bar{x}}^\infty \partial f(x) \right) \setminus \{0\} \right]. \end{aligned}$$

*Proof of Lemma 4.* The basic idea is to rewrite  $\varphi \circ f$  as a parametric minimization problem and apply [30, Theorem 10.13]. Note that  $\varphi(f(x)) = \inf_\alpha [g(x, \alpha) \triangleq \delta_{\text{epi } f}(x, \alpha) + \varphi(\alpha)]$  for  $x \in \text{dom}(\varphi \circ f)$ . Define the corresponding set of optimal solutions as  $M(x)$  for any  $x \in \text{dom}(\varphi \circ f)$ . Then, we have  $f(\bar{x}) \in M(\bar{x})$  and  $\varphi(\alpha) = \varphi(f(\bar{x}))$  for any  $\alpha \in M(\bar{x})$ . By our assumptions, it is obvious that  $\text{dom } \varphi \in \{(-\infty, b), (-\infty, b]\}$  for some  $b \in \mathbb{R} \cup \{+\infty\}$ . Based on our assumption that  $\sup \varphi = +\infty$  and  $f$  is lsc, it is easy to verify that  $g$  is proper, lsc, and level-bounded in  $\alpha$  locally uniformly in  $x$ . Then we apply [30, Theorem 10.13] to obtain

$$\partial(\varphi \circ f)(\bar{x}) \subset \{v \mid (v, 0) \in \partial g(\bar{x}, \bar{\alpha}), \bar{\alpha} \in M(\bar{x})\}, \quad \partial^\infty(\varphi \circ f)(\bar{x}) \subset \{v \mid (v, 0) \in \partial^\infty g(\bar{x}, \bar{\alpha}), \bar{\alpha} \in M(\bar{x})\}. \quad (46)$$

Step 1: We will show that for any  $\bar{\alpha} \in M(\bar{x})$ ,

$$\mathcal{N}_{\text{epi } f}(\bar{x}, \bar{\alpha}) \cap (\{0\} \times [-\mathcal{N}_{\text{dom } \varphi}(\bar{\alpha})]) = \{0\}. \quad (47)$$

We divide the proof of (47) into two cases.

Case 1. If  $M(\bar{x})$  is a singleton  $\{f(\bar{x})\}$ , we can characterize  $\mathcal{N}_{\text{epi } f}(\bar{x}, f(\bar{x}))$  by using the result in [30, Theorem 8.9]. Since  $\partial f(\bar{x}) \subset \text{Lim sup}_{x \rightarrow \bar{x}} \partial f(x)$  and  $\mathcal{N}_{\text{dom } \varphi}(f(\bar{x})) \subset \text{Lim sup}_{x \rightarrow (\varphi \circ f)\bar{x}} \mathcal{N}_{\text{dom } \varphi}(f(x))$ , it follows from our assumption that either  $0 \notin \partial f(\bar{x})$  or  $\mathcal{N}_{\text{dom } \varphi}(f(\bar{x})) = \{0\}$ . Hence, based on the characterization of  $\mathcal{N}_{\text{epi } f}(\bar{x}, f(\bar{x}))$ , (47) is satisfied.

Case 2. Otherwise, there exists  $\bar{\alpha}_{\max} \in (f(\bar{x}), +\infty)$  such that  $M(\bar{x}) = [f(\bar{x}), \bar{\alpha}_{\max}]$  since  $\varphi$  is lsc, nondecreasing and  $\sup \varphi = +\infty$ . Thus, from (46),

$$\begin{aligned} \partial(\varphi \circ f)(\bar{x}) &\subset [\{v \mid (v, 0) \in \partial g(\bar{x}, f(\bar{x}))\} \cup \{v \mid (v, 0) \in \partial g(\bar{x}, \bar{\alpha}), f(\bar{x}) < \bar{\alpha} \leq \bar{\alpha}_{\max}\}], \\ \partial^\infty(\varphi \circ f)(\bar{x}) &\subset [\{v \mid (v, 0) \in \partial^\infty g(\bar{x}, f(\bar{x}))\} \cup \{v \mid (v, 0) \in \partial^\infty g(\bar{x}, \bar{\alpha}), f(\bar{x}) < \bar{\alpha} \leq \bar{\alpha}_{\max}\}]. \end{aligned} \quad (48)$$

Let  $M_1(\bar{x}) \triangleq \{\bar{\alpha} \in (f(\bar{x}), \bar{\alpha}_{\max}] \mid \exists x^k \rightarrow \bar{x} \text{ with } f(x^k) \rightarrow \bar{\alpha}\}$  and  $M_2(\bar{x}) \triangleq M(\bar{x}) \setminus M_1(\bar{x})$ . In the following, we characterize  $\mathcal{N}_{\text{epi } f}(\bar{x}, \bar{\alpha})$  and verify (47) separately for  $\bar{\alpha} \in M_1(\bar{x})$  and  $\bar{\alpha} \in M_2(\bar{x})$ .

Case 2.1. For any  $\bar{\alpha} \in M_1(\bar{x})$ , we first prove the inclusion:

$$\mathcal{N}_{\text{epi } f}(\bar{x}, \bar{\alpha}) \subset \left[ \left\{ \lambda(v, -1) \mid v \in \text{Lim sup}_{x \rightarrow \bar{x}} \partial f(x), \lambda > 0 \right\} \cup \left\{ (v, 0) \mid v \in \text{Lim sup}_{x \rightarrow \bar{x}}^\infty \partial f(x) \right\} \right]. \quad (49)$$

Observe that for any  $\bar{\alpha} \in M_1(\bar{x})$ , it holds that

$$\mathcal{N}_{\text{epi } f}(\bar{x}, \bar{\alpha}) \subset \text{Lim sup}_{(x, \alpha) \in (\text{epi } f) \rightarrow (\bar{x}, \bar{\alpha})} \mathcal{N}_{\text{epi } f}^p(x, \alpha) \subset \text{Lim sup}_{x \rightarrow \bar{x}} \mathcal{N}_{\text{epi } f}^p(x, f(x)) \subset \text{Lim sup}_{x \rightarrow \bar{x}} \mathcal{N}_{\text{epi } f}(x, f(x)), \quad (50)$$

where the first inclusion is because any normal vector is a limit of proximal normals at nearby points [30, Exercise 6.18]; the second one uses Lemma 3; the last inclusion follows from the fact that the proximal normal cone is a subset of the limiting normal cone [30, Example 6.16]. Based on the the result of [30, Theorem 8.9] that

$$\mathcal{N}_{\text{epi } f}(x, f(x)) = \{\lambda(v, -1) \mid v \in \partial f(x), \lambda > 0\} \cup \{(v, 0) \mid v \in \partial^\infty f(x)\},$$

we conclude that  $\mathcal{N}_{\text{epi } f}(\bar{x}, \bar{\alpha}) \subset \mathbb{R}^n \times \mathbb{R}_-$  for any  $\bar{\alpha} \in M_1(\bar{x})$ . For any  $(v, -1) \in \mathcal{N}_{\text{epi } f}(\bar{x}, \bar{\alpha})$  with  $\bar{\alpha} \in M_1(\bar{x})$ , there exist  $x^k \rightarrow \bar{x}$ ,  $v^k \rightarrow v$  with  $v^k \in \partial f(x^k)$ . Then  $v \in \text{Lim sup}_{x \rightarrow \bar{x}} \partial f(x)$ .

To prove (49), it remains to show that  $v \in \text{Lim sup}_{x \rightarrow \bar{x}} \partial^\infty f(x)$  whenever  $(v, 0) \in \mathcal{N}_{\text{epi } f}(\bar{x}, \bar{\alpha})$ . It follows from (50) that  $(v, 0)$  is a limit of proximal normals of  $\text{epi } f$  at  $(x^k, f(x^k))$  for some sequence  $x^k \rightarrow \bar{x}$ . (i) First consider the case  $(v^k, 0) \rightarrow (v, 0)$  with  $(v^k, 0) \in \mathcal{N}_{\text{epi } f}^p(x^k, f(x^k))$ . Following the argument in the proof of [30, Theorem 8.9], we can derive  $v^k \in \partial^\infty f(x^k)$ . Therefore,

$$v \in \text{Lim sup}_{k \rightarrow +\infty} \partial^\infty f(x^k) \subset \text{Lim sup}_{k \rightarrow +\infty} \left( \bigcup_{x^k, i \rightarrow_f x^k} \text{Lim sup}_{i \rightarrow +\infty} \partial^\infty f(x^{k,i}) \right) \subset \bigcup_{x^j \rightarrow \bar{x}} \text{Lim sup}_{j \rightarrow +\infty} \partial^\infty f(x^j),$$

where the first inclusion is due to the definition of the horizon subdifferential, and the last inclusion follows from a standard diagonal extraction procedure. (ii) In the other case, we have  $\lambda_k(v^k, -1) \rightarrow (v, 0)$  with  $\lambda_k \downarrow 0$  and  $v^k \in \partial f(x^k)$  for all  $k \in \mathbb{N}$ . It is easy to see  $v \in \text{Lim sup}_{x \rightarrow \bar{x}} \partial^\infty f(x)$ . So far, we obtain inclusion (49). Since  $\bar{\alpha} \in M_1(\bar{x})$ , we have  $\mathcal{N}_{\text{dom } \varphi}(\bar{\alpha}) \subset \text{Lim sup}_{x \rightarrow (\varphi \circ f)\bar{x}} \mathcal{N}_{\text{dom } \varphi}(f(x))$ , and our assumption implies that  $\lambda = 0$  is the unique solution satisfying  $0 \in \lambda \cdot \text{Lim sup}_{x \rightarrow \bar{x}} \partial f(x)$  with  $\lambda \in \mathcal{N}_{\text{dom } \varphi}(\bar{\alpha})$ . Combining this with (49), we immediately obtain (47).

Case 2.2. For any  $\bar{\alpha} \in M_2(\bar{x})$ , consider any sequence  $\{(x^k, \alpha^k)\} \subset \text{epi } f$  converging to  $(\bar{x}, \bar{\alpha})$ . Then  $\alpha^k > f(x^k)$  for all sufficiently large  $k$  since  $\bar{\alpha} \notin M_1(\bar{x})$ . It is easy to see that  $\mathcal{N}_{\text{epi } f}^p(x^k, \alpha^k) \subset \mathbb{R}^n \times \{0\}$ , which gives us  $\mathcal{N}_{\text{epi } f}(x^k, \alpha^k) \subset \mathbb{R}^n \times \{0\}$  due to [30, Exercise 6.18]. By following a similar pattern as the final part of Case 2.1, it is not difficult to obtain, for any  $\bar{\alpha} \in M_2(\bar{x})$ ,

$$\mathcal{N}_{\text{epi } f}(\bar{x}, \bar{\alpha}) \subset \left\{ (v, 0) \mid v \in \text{Lim sup}_{x \rightarrow \bar{x}} \partial^\infty f(x) \right\}. \quad (51)$$

In this case, (47) holds trivially. Hence, we have verified (47) for cases 1 and 2.

Step 2: Based on (47) in step 1, we can now apply the sum rule [30, corollary 10.9] for  $\partial g(\bar{x}, \bar{\alpha})$  to obtain

$$\partial g(\bar{x}, \bar{\alpha}) \subset \mathcal{N}_{\text{epi } f}(\bar{x}, \bar{\alpha}) + \{0\} \times \partial \varphi(\bar{\alpha}), \quad \partial^\infty g(\bar{x}, \bar{\alpha}) \subset \mathcal{N}_{\text{epi } f}(\bar{x}, \bar{\alpha}) + \{0\} \times \mathcal{N}_{\text{dom } \varphi}(\bar{\alpha}). \quad (52)$$

Case 1. For  $M(\bar{x}) = \{f(\bar{x})\}$ , by combining (52) with (46), we can derive the stated results for  $\partial(\varphi \circ f)(\bar{x})$  and  $\partial^\infty(\varphi \circ f)(\bar{x})$  based on the observations that  $\partial \varphi(f(\bar{x})) \subset \text{Lim sup}_{x \rightarrow (\varphi \circ f)\bar{x}} \varphi(f(x))$  and  $\partial^\infty f(\bar{x}) \subset \text{Lim sup}_{x \rightarrow \bar{x}} \partial^\infty f(x)$ .

Case 2. Otherwise, by (52), we have

$$\begin{aligned} & \{v \mid (v, 0) \in \partial g(\bar{x}, \bar{\alpha}), f(\bar{x}) < \bar{\alpha} \leq \bar{\alpha}_{\max}\} \\ (49)(51) \quad & \subset \bigcup \left\{ y \cdot \text{Lim sup}_{x \rightarrow \bar{x}} \partial f(x) \mid y \in \partial \varphi(\bar{\alpha}), \bar{\alpha} \in M_1(\bar{x}) \right\} \cup \left[ \bigcup \left\{ \text{Lim sup}_{x \rightarrow \bar{x}} \partial^\infty f(x) \mid 0 \in \partial \varphi(\bar{\alpha}), f(\bar{x}) < \bar{\alpha} \leq \bar{\alpha}_{\max} \right\} \right] \\ & \subset \bigcup \left\{ y \cdot \text{Lim sup}_{x \rightarrow \bar{x}} \partial f(x) \mid y \in \text{Lim sup}_{x \rightarrow (\varphi \circ f)\bar{x}} \partial \varphi(f(x)) \right\} \cup \left[ \left( \text{Lim sup}_{x \rightarrow \bar{x}} \partial^\infty f(x) \right) \setminus \{0\} \right], \end{aligned}$$

where the last inclusion is because 0 will be included in the first set if  $0 \in \partial\varphi(\bar{\alpha})$  for some  $\bar{\alpha} \in (f(\bar{x}), \bar{\alpha}_{\max}]$  and the second set will be empty otherwise. Similarly,

$$\begin{aligned} & \{v \mid (v, 0) \in \partial g^\infty(\bar{x}, \bar{\alpha}), f(\bar{x}) < \bar{\alpha} \leq \bar{\alpha}_{\max}\} \\ \subset & \bigcup \left\{ y \cdot \operatorname{Lim sup}_{x \rightarrow \bar{x}} \partial f(x) \mid y \in \operatorname{Lim sup}_{x \rightarrow (\varphi \circ f) \bar{x}} \mathcal{N}_{\operatorname{dom} \varphi}(f(x)) \right\} \cup \left[ \left( \operatorname{Lim sup}_{x \rightarrow \bar{x}}^\infty \partial f(x) \right) \setminus \{0\} \right]. \end{aligned}$$

We then complete the proof by using the inclusions in (48).  $\square$

Equipped with the chain rules, we are now ready to prove Proposition 4.

*Proof of Proposition 4.* Let  $\bar{x}$  be any feasible point, i.e.,  $\bar{x} \in \bigcap_{p=1}^m \operatorname{dom} F_p$ . Suppose for contradiction that (15) does not hold at  $\bar{x}$ . Thus, there exist  $p_1 \in \{1, \dots, m\}$ ,  $\{x^k\} \in S_{p_1}(\bar{x})$  and an index set  $N \in \mathbb{N}_\infty^\#$  such that  $0 \in \partial_C f_{p_1}^k(x^k)$  and  $\mathcal{N}_{\operatorname{dom} \varphi_{p_1}}(f_{p_1}^k(x^k)) \neq \{0\}$  for all  $k \in N$ . Take an arbitrary nonzero scalar  $y^k \in \mathcal{N}_{\operatorname{dom} \varphi_{p_1}}(f_{p_1}^k(x^k))$  for all  $k \in N$ . Let  $\tilde{y}$  be any accumulation point of the unit scalars  $\{y^k/|y^k|\}_{k \in N}$ . Then, we have  $(0 \neq) \tilde{y} \in \bigcup \{\mathcal{N}_{\operatorname{dom} \varphi_{p_1}}(t_{p_1}) \mid t_{p_1} \in T_{p_1}(\bar{x})\}$  and  $0 \in \operatorname{con} \partial_A f_{p_1}(\bar{x})$ , contradicting Assumption 5. This proves condition (15).

For any fixed  $p = 1, \dots, m$ , let  $y_{p'} = 0$  for any  $p' \in \{1, \dots, m\} \setminus \{p\}$  in Assumption 5. Then the only scalar  $y_p \in \bigcup \{\mathcal{N}_{\operatorname{dom} \varphi_p}(t_p) \mid t_p \in T_p(\bar{x})\}$  with  $0 \in y_p \operatorname{con} \partial_A f_p(\bar{x})$  is  $y_p = 0$ , which completes the proof of (16).

To derive the constraint qualification (17), we consider two cases.

Case 1. For  $p \in I_2$ , we have  $\mathcal{N}_{\operatorname{dom} \varphi_p}(f_p(\bar{x})) \subset \bigcup \{\mathcal{N}_{\operatorname{dom} \varphi_p}(t_p) \mid t_p \in T_p(\bar{x})\}$  due to  $f_p^k \xrightarrow{e} f_p$  and  $\partial(y f_p)(\bar{x}) \subset y \partial_C f_p(\bar{x}) \subset y \cdot \operatorname{con} \partial_A f_p(\bar{x})$  for any  $y$  by Theorem 1(a). Together with Assumption 5, we deduce that the only scalar  $y \in \mathcal{N}_{\operatorname{dom} \varphi_p}(f_p(\bar{x}))$  with  $0 \in \partial(y f_p)(\bar{x})$  is  $y = 0$ . From this condition and the local Lipschitz continuity of  $f_p$  for  $p \in I_2$ , we can apply the chain rule [30, Theorem 10.49] to get

$$\partial^\infty(\varphi_p \circ f_p)(\bar{x}) \subset \bigcup \{y \cdot \operatorname{con} \partial_A f_p(\bar{x}) \mid y \in \mathcal{N}_{\operatorname{dom} \varphi_p}(t_p), t_p \in T_p(\bar{x})\}. \quad (53)$$

Case 2. For  $p \in I_1$ , to utilize the chain rules (Lemma 4) for  $\partial^\infty(\varphi_p \circ f_p)$ , we must first confirm the validity of the condition:

$$\left[ 0 \in y \cdot \operatorname{Lim sup}_{x \rightarrow \bar{x}} \partial f_p(x), \quad y \in \operatorname{Lim sup}_{x \rightarrow F_p \bar{x}} \mathcal{N}_{\operatorname{dom} \varphi_p}(f_p(x)) \right] \implies y = 0. \quad (54)$$

Indeed, it suffices to consider the case of  $\operatorname{dom} \varphi_p^\uparrow = (-\infty, r_p)$  or  $(-\infty, r_p]$  for some  $r_p \in \mathbb{R}$ , because the statement holds trivially when  $\varphi_p^\uparrow$  is real-valued. For any element  $\bar{y} \in \operatorname{Lim sup}_{x \rightarrow F_p \bar{x}} \mathcal{N}_{\operatorname{dom} \varphi_p}(f_p(x))$ , there exist  $(x^k, y^k) \rightarrow (\bar{x}, \bar{y})$  with  $y^k \in \mathcal{N}_{\operatorname{dom} \varphi_p}(f_p(x^k))$  and  $F_p(x^k) \rightarrow F_p(\bar{x})$ . Since  $\bar{x} \in \operatorname{dom} F_p$ , we must have  $x^k \in \operatorname{dom} F_p$  for all sufficiently large  $k$ , i.e.,  $f_p(x^k) \in \operatorname{dom} \varphi_p^\uparrow$ , and  $\{f_p(x^k)\}_{k \in \mathbb{N}}$  is bounded from above due to  $\operatorname{dom} \varphi_p^\uparrow = (-\infty, r_p)$  or  $(-\infty, r_p]$ . The sequence  $\{f_p(x^k)\}_{k \in \mathbb{N}}$  is also bounded from below since  $f_p$  is lsc as a consequence of  $f_p^k \xrightarrow{e} f_p$ . Then, we can assume that the bounded sequence  $\{f_p(x^k)\}_{k \in \mathbb{N}}$  converges to some  $\bar{z}_p$ . Note that  $\bar{z}_p \in \operatorname{dom} \varphi_p$  due to  $F_p(\bar{x}) = \liminf_{k \rightarrow +\infty} \varphi_p(f_p(x^k)) \geq \varphi_p(\bar{z}_p)$ . Thus, by the outer semicontinuity,  $y^k \rightarrow \bar{y} \in \mathcal{N}_{\operatorname{dom} \varphi_p}(\bar{z}_p)$ . By  $f_p^k \xrightarrow{e} f_p$ , each  $f_p(x^k)$  can be expressed as the limit of a sequence  $\{f_p^i(x^{k,i})\}_{i \in \mathbb{N}}$  with  $x^{k,i} \rightarrow x^k$  for

any fixed  $k \in \mathbb{N}$ . Using a standard diagonal extraction procedure, one can extract a subsequence  $f_p^{i_k}(x^{k,i_k}) \rightarrow \bar{z}_p$  with  $x^{k,i_k} \rightarrow \bar{x}$ . Hence,  $\bar{z}_p \in T_p(\bar{x})$  and

$$\text{Lim sup}_{x \rightarrow_{F_p} \bar{x}} \mathcal{N}_{\text{dom } \varphi_p}(f_p(x)) \subset \bigcup \{ \mathcal{N}_{\text{dom } \varphi_p}(t_p) \mid t_p \in T_p(\bar{x}) \}. \quad (55)$$

Using the subdifferentials relationships in Theorem 1 and the outer semicontinuity of  $\partial_A f_p$ , we have

$$\text{Lim sup}_{x \rightarrow \bar{x}} \partial f_p(x) \subset \text{Lim sup}_{x \rightarrow \bar{x}} \partial_A f_p(x) = \partial_A f_p(\bar{x}). \quad (56)$$

By (55), (56) and Assumption 5, we immediately get (54). Thus, we can apply the chain rule in Lemma 4, and use (55), (56) again to obtain

$$\begin{aligned} \partial^\infty(\varphi_p \circ f_p)(\bar{x}) &\subset \bigcup \{ y \partial_A f_p(x) \mid y \in \mathcal{N}_{\text{dom } \varphi_p}(t_p), t_p \in T_p(\bar{x}) \} \cup \left[ \text{Lim sup}_{x \rightarrow \bar{x}}^\infty \partial f_p(x) \setminus \{0\} \right] \\ &\subset \bigcup \{ y \partial_A f_p(x) \mid y \in \mathcal{N}_{\text{dom } \varphi_p}(t_p), t_p \in T_p(\bar{x}) \} \cup [\partial_A^\infty f_p(\bar{x}) \setminus \{0\}]. \end{aligned} \quad (57)$$

For the last inclusion, we use  $\text{Lim sup}_{x \rightarrow \bar{x}}^\infty \partial f_p(x) \subset \text{Lim sup}_{x \rightarrow \bar{x}}^\infty \partial_A f_p(x) \subset \partial_A^\infty f_p(\bar{x})$  by Theorem 1(a) and using a standard diagonal extraction procedure. Combining inclusions (53), (57) for two cases with Assumption 5, we derive (17) and complete the proof.  $\square$