SIAM J. OPTIM.
Vol. 31, No. 1, pp. 837–865

NECESSARY AND SUFFICIENT OPTIMALITY CONDITIONS IN DC SEMI-INFINITE PROGRAMMING*

RAFAEL CORREA†, M. A. LÓPEZ‡, AND PEDRO PÉREZ-AROS§

Abstract. This paper deals with particular families of DC optimization problems involving suprema of convex functions. We show that the specific structure of this type of function allows us to cover a variety of problems in nonconvex programming. Necessary and sufficient optimality conditions for these families of DC optimization problems are established, where some of these structural features are conveniently exploited. More precisely, we derive necessary and sufficient conditions for (global and local) optimality in DC semi-infinite programming and DC cone-constrained optimization, under natural constraint qualifications. Finally, a penalty approach to DC abstract programming problems is developed in the last section.

Key words. DC functions, supremum function, semi-infinite programming, cone-constraint programming

AMS subject classifications. 90C30, 90C34, 90C26

DOI. 10.1137/19M1303320

1. Introduction.

In [34, Chapter 3] the author emphasizes the universality of DC functions as "virtually all the most frequently encountered functions in practice are DC." This claim is supported by the most relevant properties of these functions, particularly by their stability relative to operations frequently used in optimization. The class of DC functions is considered in [1] a remarkable subclass of locally Lipschitz functions, and it is the smallest vector space containing all continuous convex functions on a given set. Moreover, every continuous function f defined on a compact convex set K in a normed space can be approached by a sequence of DC functions which converges to f uniformly on K. Actually, in $\mathbb{R}^n$ this property follows from the fact that polynomials are DC, since every function $f \in \mathcal{C}^2(\mathbb{R}^n)$ is DC on any compact convex set.

In mathematical optimization, several models can be formulated in terms of the maximum/supremum of a finite/infinite family of data functions. Moreover, many convex functions, such as the Fenchel conjugate, the sum, the composition with affine applications, etc., can be expressed as the supremum of affine or convex functions. Even more, any formula for the subdifferential of the supremum function can be seen as a useful tool in deriving KKT-type optimality conditions for a convex optimization problem as any set of convex constraints, even an infinite set, can be replaced by a unique convex constraint by using the supremum function. For that reason, several

*Received by the editors December 2, 2019; accepted for publication (in revised form) December 6, 2020; published electronically March 11, 2021.
https://doi.org/10.1137/19M1303320
Funding: The first author was partially supported by ANID Chile under grant Fondecyt Regular 1190110. The second author is supported by Research Project PGC2018-097960-B-C21 from MICINN Spain, and Australian ARC–Discovery Projects DP 180100602. The third author was partially supported by ANID Chile under grants Fondecyt regular 1190110, Fondecyt regular 1200283, and Programa Regional Mathamsud 20-Math-08 CODE: MATH190003.
†Universidad de O'Higgins, Rancagua, Chile and DIM-CMM of Universidad de Chile, Santiago, Chile (rafael.correa@uoh.cl).
‡Department of Mathematics, University of Alicante, Alicante 03080, Spain, and CIAO, Federation University, Australia (marco.antonio@ua.es).
§Instituto de Ciencias de la Ingeniería, Universidad de O'Higgins, Rancagua 2820000, Chile (pedro.perez@uoh.cl).

authors have focused their research on computing gradients, subdifferentials, and calculus rules via the supremum function, and remarkable contributions to this topic can be found in the literature, starting in the decades of the 1960s and 1970s of the past century. In [18], Hiriart-Urruty claimed: "One of the most specific constructions in convex or nonsmooth analysis is certainly taking the supremum of a (possibly infinite) collection of functions. In the years 1965–1970, various calculus rules concerning the subdifferential of sup-functions started to emerge; working in that direction and using various assumptions, several authors contributed to this calculus rule: B.N. Pshenichnyi, A.D. Ioffe, V.L. Levin, R.T. Rockafellar, A. Sotskov, etc.; however, the most elaborated results of that time were due to M. Valadier (1969) [35]." Recent contributions in this field are [4, 5, 6, 7, 8, 15, 16, 26, 28, 30, 31] and references therein. On the other hand, for the class of DC problems there also exist necessary and sufficient conditions for optimality, many times in terms of (exact/approximate) subdifferentials of the involved DC decomposition (see, e.g., [9, 10, 12, 14, 17, 20, 21, 22] and references therein). Unfortunately, the class of DC functions is not stable in general for supremum functions (see Example 2.1). Nevertheless, in this work we restrict our study to particular families of DC optimization problems involving suprema of convex functions. In this framework, we combine formulae for the (approximate) subdifferential of supremum functions and optimality conditions for DC programming to get new optimality conditions for several classes of optimization problems which are relevant in mathematical programming.

The rest of the paper is organized as follows: In section 2, we give the main definitions and notation used in this work; in section 3 we introduce the class of DC functions on which the paper is focused, and some of their properties are established; in section 4, we recall and develop formulae for the subdifferential of DC functions and supremum functions in several specific frameworks; in section 5 we apply our formulae to provide necessary and sufficient conditions for global and local optimality of problems related to (nonconvex) semi-infinite programming; in section 6 we show that general problems in cone-constrained optimization can be translated into our setting, and with the help of our formulae, we derive necessary and sufficient conditions for (global and local) optimality; in section 7 we focus on general nonconvex optimization problems with DC objective functions and also provide necessary and sufficient conditions for optimality; and in section 8 we develop optimality conditions for perturbations of general nonconvex optimization problems. Finally, the work ends with some concluding remarks.

2. Notation. Throughout the paper, unless we stipulate something else, we consider $\mathbb{R}^n$ equipped with the Euclidean norm $\|\cdot\|$. Given a set $A \subset \mathbb{R}^n$, we denote by $cl(A)$, $int(A)$, $conv(A)$, $cone(A)$ the *closure*, the *interior*, the *convex hull*, and the the *convex cone* generated by A. By 0_n we represent the zero vector in $\mathbb{R}^n$. For two sets $A, B \subset \mathbb{R}^n$ we define the following operations,

$$A + B := \{a + b,\ a \in A \text{ and } b \in B\}$$

and

$$A \ominus B := \{x \in \mathbb{R}^n : x + B \subseteq A\}$$

with the following standard conventions:

$$A + \emptyset = \emptyset + B = \emptyset \text{ and } conv(\emptyset) = \emptyset. \tag{2.1}$$

Given a family of sets $\{A_i\}_{i\in I}$, in order to avoid possible confusion it is convenient to adopt the following notation:

$$\bigcup[A_i : i \in I] := \bigcup_{i\in I} A_i.$$

If A is convex and $\varepsilon \geq 0$, we define the *ε-normal set* to A at x as

$$\mathrm{N}_A^{\varepsilon}(x) := \{x^* \in X^* : \ \langle x^*, y - x\rangle \leq \varepsilon \ \forall y \in A\}$$

if $x \in A$, and $\mathrm{N}_A^{\varepsilon}(x) = \emptyset$ if $x \notin A$. If $\varepsilon = 0$, $\mathrm{N}_A^0(x) \equiv \mathrm{N}_A(x)$ is the so-called *normal cone* to A at x.

If K is a cone, the *polar cone of* K is

$$K^{\circ} := \{x^* \in X^* : \ \langle x^*, y\rangle \leq 0 \ \forall y \in K\} \equiv \mathrm{N}_K(0_n).$$

Given an arbitrary set T, $\mathcal{P}_f(T)$ denotes the family of finite subsets of T.

If $Z \subset \mathbb{R}$, by $Z^{(T)}$ we represent the family of functions $\lambda : T \to Z$ (i.e., $\lambda \in Z^T$) such that $\lambda(t) \equiv \lambda_t = 0$ for all $t \in T$ except perhaps for finitely many $t \in T$. The *support* of λ is defined as $\mathrm{supp}\lambda := \{t \in T : \lambda_t \neq 0\}$. The *$\varepsilon$-generalized simplex* on T is the set

$$\Delta^{\varepsilon}(T) := \left\{ \lambda \in [0,1]^{(T)} : \sum_{t\in T} \lambda_t = \sum_{t\in \ \mathrm{supp}\lambda} \lambda_t = \varepsilon \right\}.$$

In particular, the *generalized simplex* on T is $\Delta^1(T)$, which is denoted simply by $\Delta(T)$. Furthermore, we also denote $\Delta_n^{\varepsilon} := \Delta^{\varepsilon}(\{1,\ldots,n\})$ and $\Delta_n := \Delta(\{1,\ldots,n\})$. Finally, for a family of sets $\{C_t\}_{t\in T}$ and $\lambda \in \mathbb{R}^{(T)}$ we define

$$\sum_{t\in T} \lambda_t C_t := \sum_{t\in \mathrm{supp}\,\lambda} \lambda_t C_t.$$

Given a function $f : \mathbb{R}^n \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$, the (*effective*) *domain* and the *epigraph* of f are

$$\mathrm{dom}\, f := \{x \in \mathbb{R}^n : \ f(x) < +\infty\} \text{ and } \mathrm{epi}\, f := \{(x,\alpha) \in \mathbb{R}^n \times \mathbb{R} : \ f(x) \leq \alpha\},$$

respectively. We say that f is *proper* if $\mathrm{dom}\, f \neq \emptyset$.

For $\varepsilon \geq 0$, the *ε-subdifferential* (or *approximate subdifferential*) of f at a point $x \in \mathbb{R}^n$, where f is finite, is the set

$$\partial_{\varepsilon} f(x) := \{x^* \in \mathbb{R}^n : \ \langle x^*, y - x\rangle \leq f(y) - f(x) + \varepsilon \ \forall y \in \mathbb{R}^n\};$$

if $f(x)$ is not finite, we set $\partial_{\varepsilon} f(x) = \emptyset$. The special case $\varepsilon = 0$ yields the classical (*Moreau-Rockafellar*) *convex subdifferential*, denoted by $\partial f(x)$.

The *Fréchet subdifferential* of f at $x \in dom f$ is the set

$$\hat{\partial} f(x) := \left\{ x^* \in \mathbb{R}^n : \ \liminf_{h\to 0_n} \frac{f(x+h) - f(x) - \langle x^*, h\rangle}{\|h\|} \geq 0 \right\},$$

and its elements, usually called (*Fréchet*) *subgradients* (also *regular subgradients* [33]), are affine functions "supporting" f from below. The set $\hat{\partial} f(x)$ is closed and convex and generalizes simultaneously the notions of Fréchet derivative and subdifferential of a convex function.

The following *sum rule* is applied in the paper (see, e.g., [25, Proposition 1.107]): If the function g is Fréchet differentiable at x, and f is finite at this point, then

$$\hat{\partial}(f+g)(x) = \hat{\partial} f(x) + \nabla g(x). \tag{2.2}$$

Moreover, for a mapping $F : \mathbb{R}^n \to \mathbb{R}^m$ which is *calm* at x, i.e., such that

$$\|F(u) - F(x)\| \leq \ell \|u - x\|,$$

for some $\ell > 0$ and u in a certain neighborhood of x, the *Fréchet co-derivative* of F at x is given by the set-valued map $\hat{D}F(x) : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ defined as

$$\hat{D}F(x)(y^*) := \hat{\partial}\left(\langle y^*, F(\cdot)\rangle\right)(x),$$

where $\langle y^*, F(\cdot)\rangle$ is the scalar function defined by $\langle y^*, F(\cdot)\rangle(u) := \langle y^*, F(u)\rangle$.

Given the set $A \subset \mathbb{R}^p$, the *characteristic function* and the *indicator function* are, respectively, defined as follows:

$$\mathbf{1}_A(x) := \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A, \end{cases} \qquad \delta_A(x) := \begin{cases} 0 & \text{if } x \in A, \\ +\infty & \text{if } x \notin A. \end{cases}$$

Since some of the result concerning a representation of difference of a convex function is local, we introduce a precise notation for the space of functions that we are dealing with.

Consider a convex set $U \subseteq \mathbb{R}^n$. First, we denote by $\Gamma_0(U)$ the family of all the *lower semicontinuous proper convex functions* $f : U \to \overline{\mathbb{R}}$, and by $\mathcal{DC}(U)$ the family of functions f which are a *difference of two convex functions* in $\Gamma_0(U)$, i.e., such that there exist two functions $g, h \in \Gamma_0(U)$ such that $f(x) = g(x) - h(x)$ for all $x \in U$, with the conventions $+\infty - (+\infty) = +\infty$. When there is no ambiguity in U, we simply say that f is a DC function or that f belongs to $\mathcal{DC}$.

Given a family of functions $\{f_t,\ t \in T\} \subseteq \overline{\mathbb{R}}^T$, in this paper we are especially interested in the supremum function

$$f := \sup_{t \in T} f_t.$$

Given a point $\bar{x} \in \operatorname{dom} f$ and $\varepsilon \geq 0$, the following set of indices is a key tool in our approach:

$$\begin{aligned} T_\varepsilon^f(\bar{x}) &:= \{t \in T : \ f_t(\bar{x}) \geq f(\bar{x}) - \varepsilon\}, \\ T^f(\bar{x}) &:= \{t \in T : \ f_t(\bar{x}) = f(\bar{x})\}. \end{aligned}$$

We simply use $T_\varepsilon(\bar{x})$ and $T(\bar{x})$ when there is no ambiguity in f.

The following example shows that the class of DC functions is, in general, not stable under the supremum operation and pointwise convergence.

Example 2.1. Let $f(x) = 1 - \sqrt{|(1/2) - x|}$ with $x \in [0,1]$. Since the function f is not locally Lipschitz it cannot be DC over $[0,1]$ (see [1, p. 974] for more details). Consider a sequence of polynomials $(p_n)_{n \in \mathbb{N}}$ which converges uniformly to f on $[0,1]$; we can assume that

$$\sum_{k \in \mathbb{N}} \|p_k - p_{k+1}\|_\infty < +\infty.$$

Now, we define the sequence of functions

$$f_n(x) := p_n(x) - \sum_{k=n}^{\infty} \|p_k - p_{k+1}\|_\infty,\ n \in \mathbb{N}.$$

We have $f_n \leq f_{n+1}$ as

$$f_n(x) - f_{n+1}(x) = p_n(x) - p_{n+1}(x) - \|p_n - p_{n+1}\|_\infty \leq 0,$$

yielding

$$\sup_{n\in\mathbb{N}} f_n(x) = \lim_{n\to\infty} f_n(x) = f(x),$$

and this leads us to the desired conclusion.

In contrast to Example 2.1, the following proposition shows a criterion to ensure that the supremum of DC functions is still a DC function.

Consider families of functions $g_t, h_t \in \Gamma_0(\mathbb{R}^n)$ with $t \in T$. Given $F \in \mathcal{P}_f(T)$ we denote by

$$g_F(x) := \max_{t\in F}\left(g_t(x) + \sum_{s\in F\setminus\{t\}} h_s(x)\right), \qquad h_F(x) := \sum_{s\in F} h_s(x),$$

and the supremum functions

$$g(x) := \sup_{F\in\mathcal{P}_f(T)} g_F(x), \quad h(x) := \sup_{F\in\mathcal{P}_f(T)} h_F(x). \tag{2.3}$$

Proposition 2.2. *Consider a family of functions $f_t = g_t - h_t$, where $g_t, h_t \in \Gamma_0(\mathbb{R}^n)$, $t \in T$, which are all nonnegative. Then, $f := \sup_{t\in T} f_t$ is a DC function over* $\mathrm{dom} h$. *Furthermore, $f(x) = g(x) - h(x)$ for all $x \in \mathrm{dom} h$, where g and h are defined in* (2.3).

Proof. Let us notice first that, for every $F_1 \subseteq F_2$,

$$\sum_{s\in F_1} h_s(x) \leq \sum_{s\in F_2} h_s(x). \tag{2.4}$$

Given $F \in \mathcal{P}_f(T)$ and $x \in \mathbb{R}^n$, we denote by $t(F,x)$ an index in F such that

$$g_F(x) = g_{t(F,x)}(x) + \sum_{s\in F\setminus\{t(F,x)\}} h_s(x).$$

For any $t \in T$ and $F \in \mathcal{P}_f(T)$ we have

$$f_t = g_t - h_t = g_t + \sum_{s\in F\setminus\{t\}} h_s - \sum_{s\in F\cup\{t\}} h_s \leq g_{F\cup\{t\}} - h_{F\cup\{t\}}.$$

Now, fix $x \in \mathrm{dom} h$ and $\varepsilon > 0$. On the one hand, there exists $F \in \mathcal{P}_f(T)$ such that $h(x) \leq \sum_{s\in F} h_s(x) + \varepsilon \leq \sum_{s\in F\cup\{t\}} h_s(x)$, where the last inequality is justified by (2.4), so for any $t \in T$,

$$f_t(x) \leq g_{F\cup\{t\}} - h_{F\cup\{t\}} \leq g(x) - h(x) + \varepsilon.$$

Since $\varepsilon > 0$ and $t \in T$ were arbitrarily chosen we get $f(x) \leq g(x) - h(x)$. To prove the opposite inequality, we assume that $g(x) \in \mathbb{R}$ (otherwise the equality holds trivially), so there exist $F \in \mathcal{P}_f(T)$ such that $g(x) \leq g_F(x) + \varepsilon$. Hence,

$$\begin{aligned} g(x) - h(x) &\leq g_F(x) - h_F(x) + \varepsilon \\ &= g_{t(F,x)}(x) + \sum_{s \in F \setminus \{t(F,x)\}} h_s(x) + \varepsilon - h_{F \cup \{t(F,x)\}\}}(x) \\ &= g_{t(F,x)}(x) - h_{t(F,x)}(x) + \varepsilon \leq f(x) + \varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ was arbitrarily chosen we get $g(x) - h(x) \leq f(x)$, and we are done. ☐

3. The class $\Gamma_h(\mathbb{R}^n)$.

We introduce now a class of DC functions, which play a crucial role in the paper. If U is a convex set in $\mathbb{R}^n$ and $h \in \Gamma_0(U)$, we define the family of functions

$$\Gamma_h(U) := \{f : U \to \overline{\mathbb{R}} : \ f + h \in \Gamma_0(U)\}.$$

For example, $x^3 \in \Gamma_{x^2}([-1/3, +\infty[)$. Obviously, $\Gamma_h(U) \subset \mathcal{DC}(U)$, and we can also write $\Gamma_h(U) = \Gamma_0(U) - h$. We say that h is a *control function* for the functions in $\Gamma_h(U)$.

The following result is a simple criterion to guarantee that a function f belongs to $\Gamma_h(\mathbb{R}^n)$. For that purpose, we need a concept of second-order derivative. Given an open set U and a function $f : U \to \mathbb{R}$ which is $\mathcal{C}^{1+}$ at $\bar{x}$ (i.e., f is differentiable at $\bar{x}$ with locally Lipschitz continuous gradient), we define the *generalized Hessian* of f at $\bar{x}$ by

$$\overline{H}_f(\bar{x}) := \{A \in \mathbb{R}^{n \times n} : \ \exists x_n \to \bar{x} \text{ such that } \nabla^2 f(x_n) \to A\}.$$

It has been proved that the generalized Hessian is a nonempty and compact set of symmetric matrices (see, e.g., [33, Theorem 13.52] and also [2, 3]). Furthermore, it can be used to provide the following characterization of convexity for not necessary $\mathcal{C}^2$-functions.

PROPOSITION 3.1 ([19, Example 2.2]). *A $\mathcal{C}^{1+}$-function f defined on the convex open set U is convex if and only if for all $x \in U$ and all $A \in \overline{H}_f(x)$ one has $A \succcurlyeq 0$ (i.e., $\langle Au, u\rangle \geq 0$ for all $u \in \mathbb{R}^n$).*

Using the above result we can establish the following characterization of DC functions.

PROPOSITION 3.2. *Let U be a convex open set in $\mathbb{R}^n$. Consider a $\mathcal{C}^{1+}$-function $f : U \to \mathbb{R}$ and a $\mathcal{C}^2$-convex function $h : U \to \mathbb{R}$. Then, if*

$$A \succcurlyeq -\nabla^2 h(x) \ \forall x \in U \ \text{ and } \forall A \in \overline{H}_f(x),$$

we have $f \in \Gamma_h(U)$.

Proof. We need to show that the function $g := f + h$ is convex. Since the function g is $\mathcal{C}^{1+}$ and h is $\mathcal{C}^2$ it is not difficult to show that

$$\overline{H}_g(x) = \overline{H}_f(x) + \nabla^2 h(x) \ \forall x \in U.$$

Consequently, every $B \in \overline{H}_g(x)$ is a positive semidefinite matrix, and this implies the convexity of g on U in virtue of Proposition 3.1. ☐

One important class of DC functions is the class $\Gamma_{\rho J}(\mathbb{R}^n)$, where J is the so-called *duality function*, that is, $J(x) = \frac{1}{2}\|x\|^2$. For the sake of brevity, we represent this class by $\Gamma_\rho(\mathbb{R}^n)$. Let us recall here that

$$\partial_\varepsilon J(x) = \mathbb{B}(x, \sqrt{2\varepsilon}). \tag{3.1}$$

Thanks to Proposition 3.2 it is not difficult to prove that any C^2-function belongs to $\Gamma_\rho(K)$ over every bounded set K (for some $\rho \geq 0$). More generally, Proposition 3.2 also can be used to show that the class of lower-C^2 functions satisfies the same property. Formally, the class of lower-C^2 functions is defined as follows.

DEFINITION 3.3. *A function $f : U \to \mathbb{R}$, where $U \subset \mathbb{R}^n$ is an open set, is said to be* lower-$\mathcal{C}^2$ *on U if, for every $\bar{x} \in U$, there exists a neighborhood $\mathcal{V}$ of $\bar{x}$ where f admits a representation*

$$f(x) = \sup_{t \in T} f_t(x),$$

where T is a compact space and $f_t(x)$, $\nabla f_t(x)$, and $\nabla^2 f_t(x)$ depend continuously jointly on $(t, x) \in T \times \mathcal{V}$.

Any finite convex function is lower-$\mathcal{C}^2$ (see, e.g., [33, Theorem 10.33]). It has been shown that if the function f is lower-$\mathcal{C}^2$ on a convex open set U, then f belongs to $\Gamma_\rho(K)$ over each convex compact set $K \subseteq U$ (see, e.g., [36, Proposition 3.3]).

4. Mixing supremum and DC formulae. This section starts by recalling a well-known formula for the ε-subdifferential of DC functions and continues by establishing some results for ε-subdifferential of the supremum of a family of convex functions. These formulae will be our working horse to establish different optimality conditions in the next sections.

PROPOSITION 4.1 ([24, Theorem 1]). *Let $g, h \in \Gamma_0(\mathbb{R}^n)$ such that both are finite at x. Then, for every $\varepsilon \geq 0$*

$$\partial_\varepsilon (g - h)(x) = \bigcap_{\eta \geq 0} \left(\partial_{\eta+\varepsilon} g(x) \ominus \partial_\eta h(x)\right).$$

Next we provide characterizations of the subdifferential and the ε-subdifferential of the supremum function of a family of convex functions $\{g_t,\ t \in T\}$. The first result, which is a Valadier-type formula (see [35, Theorem 2]), follows directly from [6, Corollary 3.13], and the second one, given in Theorem 4.3, is a corollary of [28, Theorem 5.4].

PROPOSITION 4.2 ([6, Corollary 3.13]). *Let $\{g_t,\ t \in T\} \subseteq \Gamma_0(\mathbb{R}^n)$, and suppose that $g := \sup_{t \in T} g_t$ is finite and continuous at $\bar{x}$. Assume additionally that for some $\gamma_0 > 0$,*

(i) *the set $T_{\gamma_0}(\bar{x})$ is compact,*

(ii) *for every $z \in \operatorname{dom} g$, the function $t \mapsto g_t(z)$ is upper semicontinuous on $T_{\gamma_0}(\bar{x})$.*

Then

$$\partial g(\bar{x}) = \operatorname{co}\left(\bigcup_{t \in T(\bar{x})} \partial g_t(\bar{x})\right).$$

In the proof of the next theorem we need the following result established in [28, Theorem 5.1] (see [29] for a corrected proof).

PROPOSITION 4.3. *Consider a family of functions* $\{g_t,\, t \in T\} \subset \Gamma_0(\mathbb{R}^n)$, *together with the supremum function* $g := \sup_{t\in T} g_t$, *and suppose that* $\operatorname{dom} g$ *has a nonempty interior. Then for every* $\varepsilon \geq 0$ *one has, for all* $\bar{x} \in \mathbb{R}^n$,

$$\partial_\varepsilon g(\bar{x}) = \bigcup_{(\varepsilon_1,\varepsilon_2)\in\Delta_2^\varepsilon} \left(S(\bar{x},\varepsilon_1) + \mathrm{N}^{\varepsilon_2}_{\operatorname{dom} g}(\bar{x}) \right), \tag{4.1}$$

where

$$S(\bar{x},\varepsilon_1) := \bigcap_{\gamma>0} \operatorname{cl}\left(\bigcup \left[\sum_{t\in\operatorname{supp}\lambda} \lambda_t \partial_{\nu_t+\gamma} g_t(\bar{x}) : \begin{array}{l} \lambda \in \Delta(T),\ \rho \in \Delta^{\varepsilon_1}(T),\ \nu \in \mathbb{R}_+^{(T)}, \\ \textit{and } g_t(\bar{x}) + \rho_t/\lambda_t + \gamma \geq g(\bar{x}) + \nu_t \end{array} \right] \right).$$

Let us notice that if $\varepsilon_1 \leq \varepsilon_1'$ and $\varepsilon_2 \leq \varepsilon_2'$ we have $S(\bar{x},\varepsilon_1) \subseteq S(\bar{x},\varepsilon_1')$ and $\mathrm{N}^{\varepsilon_2}_{\operatorname{dom} g}(\bar{x}) \subseteq \mathrm{N}^{\varepsilon_2'}_{\operatorname{dom} g}(\bar{x})$. Consequently, (4.1) also holds with the union over all $\varepsilon_1, \varepsilon_2 \geq 0$ and $\varepsilon_1 + \varepsilon_2 \leq \varepsilon$. Moreover, if g is finite on $\mathbb{R}^n$, and so continuous on the whole space, $\mathrm{N}^{\varepsilon_2}_{\operatorname{dom} g}(\bar{x}) = \{0_n\}$ for all $\varepsilon_2 \geq 0$, and (4.1) collapses to

$$\partial_\varepsilon g(\bar{x}) = S(\bar{x},\varepsilon). \tag{4.2}$$

THEOREM 4.4. *Consider a family of functions* $\{g_t, t \in T\} \subset \Gamma_0(\mathbb{R}^n)$ *and suppose that* $g := \sup_{t\in T} g_t$ *is finite on* $\mathbb{R}^n$. *Let* $\bar{x} \in \mathbb{R}^n$ *be such that for all* $\gamma > 0$,

(i) *the set* $T_\gamma(\bar{x})$ *is compact,*

(ii) *for every* $z \in \mathbb{R}^n$ *the function* $t \mapsto g_t(z)$ *is upper-semicontinuous on* $T_\gamma(\bar{x})$.

Then, for all $\varepsilon > 0$,

$$\partial_\varepsilon g(\bar{x}) = \bigcup \left[\sum_{t\in T} \lambda_t \partial_{\eta_t} g_t(\bar{x}) : \ (\lambda,\eta) \in \Lambda^\varepsilon((g_t)_{t\in T}, \bar{x}) \right], \tag{4.3}$$

where

$$\Lambda^\varepsilon((g_t)_{t\in T}, z) := \left\{ (\lambda,\eta) \in \Delta(T) \times \mathbb{R}_+^{(T)} : \sum_{t\in\operatorname{supp}\lambda} \lambda_t \left(g(z) - g_t(z) + \eta_t \right) \leq \varepsilon \right\}. \tag{4.4}$$

Proof. First, let us prove that the right-hand side of (4.3) is contained in $\partial_\varepsilon g(\bar{x})$. Indeed, consider $(\lambda,\eta) \in \Lambda^\varepsilon((g_t)_{t\in T}, \bar{x})$ and $u^* = \sum_{t\in\operatorname{supp}\lambda} \lambda_t u_t^*$ with $u_t^* \in \partial_{\eta_t} g_t(\bar{x})$. Then, for every $y \in \mathbb{R}^n$ we have that

$$\begin{aligned} \langle u^*, y - \bar{x} \rangle &= \sum_{t\in\operatorname{supp}\lambda} \lambda_t \langle u_t^*, y - \bar{x} \rangle \leq \sum_{t\in\ \operatorname{supp}\lambda} \lambda_t \left(g_t(y) - g_t(\bar{x}) + \eta_t \right) \\ &\leq g(y) - g(\bar{x}) + \sum_{t\in\operatorname{supp}\lambda} \lambda_t \left(g(\bar{x}) - g_t(\bar{x}) + \eta_t \right) \\ &\leq g(y) - g(\bar{x}) + \varepsilon, \end{aligned}$$

and $u^* \in \partial_\varepsilon g(\bar{x})$.

Second, consider $u^* \in \partial_\varepsilon g(\bar{x})$. Since $\operatorname{dom} g = \mathbb{R}^n$, (4.2) holds, i.e.,

$$\partial_\varepsilon g(\bar{x}) = S(\bar{x},\varepsilon).$$

Now, let us consider a sequence of positive scalars $(\gamma_k)_{k\in\mathbb{N}}$ converging to 0. We divide the proof into three claims.

Claim 1. We can take a sequence of multipliers $(\lambda_{1,k},\ldots,\lambda_{n+1,k}) \in \Delta_{n+1}$, together with numbers $(\rho_{1,k},\ldots,\rho_{n+1,k}) \in \Delta^{\varepsilon}_{n+1}$, $(\nu_{1,k},\ldots,\nu_{n+1,k}) \in \mathbb{R}^{n+1}_{+}$, subgradients $u^*_{i,k} \in \partial_{\nu_{i,k}+\gamma_k} g_{t_{i,k}}(\bar{x})$, and points $t_{i,k} \in T$, $i = 1,\ldots,n+1$, $k \in \mathbb{N}$, such that

$$u^* = \lim_{k\to+\infty} \sum_{i=1}^{n+1} \lambda_{i,k} u^*_{i,k}, \tag{4.5}$$

and, for $i = 1,\ldots,n+1$ and $k \in \mathbb{N}$,

$$g(\bar{x}) + \nu_{i,k} \le g_{t_{i,k}}(\bar{x}) + \rho_{i,k}/\lambda_{i,k} + \gamma_k. \tag{4.6}$$

Indeed, to prove this claim we notice that by definition of $S(\bar{x},\varepsilon)$ for every $k \in \mathbb{N}$ there is $y^*_k \in A_{\gamma_k}$ such that $\|x^* - y^*_k\| \le k^{-1}$, where

$$A_\gamma := \bigcup \left[\sum_{t\in\operatorname{supp}\lambda} \lambda_t \partial_{\nu_t+\gamma} g_t(\bar{x}) : \begin{array}{c} \lambda \in \Delta(T),\ \rho \in \Delta^{\varepsilon_1}(T),\ \nu \in \mathbb{R}^{(T)}_{+} \\ \text{and } g_t(\bar{x}) + \rho_t/\lambda_t + \gamma \ge g(\bar{x}) + \nu_t \end{array} \right].$$

Due to the definition of A_γ, we can write y^*_k as

$$y^*_k = \sum_{t\in\operatorname{supp}\lambda_{t,k}} \lambda_{t,k} u^*_{t,k}$$

for some multipliers $\lambda_k \in \Delta(T)$ and $u^*_{t,k} \in \partial_{\nu_{t,k}+\gamma_k} g_t(\bar{x})$ with $\rho_k \in \Delta^{\varepsilon_1}(T)$, $\nu_k \in \mathbb{R}^{(T)}_{+}$, and $g_t(\bar{x})+\rho_{t,k}/\lambda_{t,k}+\gamma_k \ge g(\bar{x})+\nu_{t,k}$. Now, by Carathéodory's theorem, the cardinal number of $\operatorname{supp}\lambda_k$ is bounded by $n+1$. Therefore, relabelling the sequences we get the proof of the claim.

Claim 2. There exist $(\lambda_1,\ldots,\lambda_{n+1}) \in \Delta_{n+1}$, $(\rho_1,\ldots,\rho_{n+1}) \in \Delta^{\varepsilon}_{n+1}$, $(u^*_1,\ldots,u^*_{n+1}) \in (\mathbb{R}^n)^{n+1}$ and $(t_1,\ldots,t_{n+1}) \in T^{n+1}$ such that for all $i = 1,\ldots,n+1$

$$\lambda_{i,k} \to \lambda_i,\ \rho_{i,k} \to \rho_i,\ \text{and } u^*_{i,k} \to u^*_i \text{ as } k \to \infty,$$

and for all $i \in \mathcal{I}$

$$t_{i,k} \to t_i \text{ as } k \to \infty,$$

where

$$\mathcal{I} := \{i \in \{1,\ldots,n+1\} :\ \lambda_i \ne 0\}.$$

Let us prove Claim 2. Since Δ_{n+1} and $\Delta^{\varepsilon}_{n+1}$ are compact sets we can assume that $\lambda_{i,k} \to \lambda_i$ and $\rho_{i,k} \to \rho_i$ as $k \to \infty$ with $(\lambda_1,\ldots,\lambda_{n+1}) \in \Delta_{n+1}$, $(\rho_1,\ldots,\rho_{n+1}) \in \Delta^{\varepsilon}_{n+1}$. Let us consider the constant

$$r := \sup\left\{\frac{\rho_{i,k}}{\lambda_{i,k}} :\ i \in \mathcal{I},\ k \in \mathbb{N}\right\} + \sup\{\gamma_k :\ k \in \mathbb{N}\} < +\infty.$$

Hence, by (4.6) we have that

$$t_{i,k} \in T_r(\bar{x})\ \forall k \in \mathbb{N}\ \forall i \in \mathcal{I}.$$

Thus, the compactness of $T_r(\bar{x})$ allows us to assume without loss of generality (w.l.o.g.) that $t_{i,k} \to t_k$ for all $i \in \mathcal{I}$ (via, perhaps, a subnet). For $i \in \{1,\ldots,n+1\}\setminus\mathcal{I}$ we can take t_i as any element in T.

Now we proceed by proving that the subgradients $u_{k,i}^*$ are bounded uniformly. Indeed, let M be such that

$$g(\bar{x}+h) \leq M \ \forall h \in \mathbb{B}.$$

Then, for all $i = 1, \ldots, n+1$, $k \in \mathbb{N}$, and all $h \in \mathbb{B}$,

$$\langle u_{i,k}^*, h \rangle \leq g_{t_{i,k}}(\bar{x}+h) - g_{t_{i,k}}(\bar{x}) + \nu_{i,k} + \gamma_k \leq M - g_{t_{i,k}}(\bar{x}) + \nu_{i,k} + \gamma_k.$$

Then, using (4.6) we get that

$$\|u_{i,k}^*\| \leq M - g(\bar{x}) + r + \sup\{\gamma_k : \ k \in \mathbb{N}\} \ \forall k \in \mathbb{N}, \forall i = 1, \ldots, n+1. \tag{4.7}$$

Hence, by (4.7) we can assume w.l.o.g. that $u_{i,k}^* \to u_i^*$ for $i = 1, \ldots, n+1$. This concludes the proof of Claim 2.

It is important to notice that there could exist $i \neq j$ such that the limiting points of the sequences (or of some subnets) $(t_{i,k})$ coincide, i.e., $t_i = t_j$ for $i \neq j$. Consequently, we define

$$\mathcal{I}_t := \{i \in \{1, \ldots, n+1\} : \ t_i = t \text{ and } \lambda_i \neq 0\}.$$

Using this set we denote

$$\lambda_t := \begin{cases} \sum\limits_{i \in \mathcal{I}_t} \lambda_i & \text{if } \mathcal{I}_t \neq \emptyset, \\ 0 & \text{if } \mathcal{I}_t = \emptyset, \end{cases} \qquad \rho_t := \begin{cases} \frac{1}{\lambda_t} \sum\limits_{i \in \mathcal{I}_t} \rho_i & \text{if } \mathcal{I}_t \neq \emptyset, \\ 0 & \text{if } \mathcal{I}_t = \emptyset, \end{cases}$$

$$\eta_t := \begin{cases} g_t(\bar{x}) - g(\bar{x}) + \rho_t & \text{if } \mathcal{I}_t \neq \emptyset, \\ 0 & \text{if } \mathcal{I}_t = \emptyset, \end{cases} \qquad u_t^* := \begin{cases} \frac{1}{\lambda_t} \sum\limits_{i \in \mathcal{I}_t} \lambda_i u_i^* & \text{if } \mathcal{I}_t \neq \emptyset, \\ 0_n & \text{if } \mathcal{I}_t = \emptyset. \end{cases}$$

Then, let us conclude the proof of the theorem proving the following claim.

Claim 3. For these λ and η, we have $(\lambda, \eta) \in \Lambda^{\varepsilon}((g_t)_{t \in T}, \bar{x})$, and

$$u^* = \sum_{t \in \operatorname{supp}\lambda} \lambda_t u_t^* \text{ and } u_t^* \in \partial_{\eta_t} g_t(\bar{x}) \ \forall t \in \operatorname{supp}\lambda.$$

To prove Claim 3, let us show first that $(\lambda, \eta) \in \Lambda^{\varepsilon}((g_t)_{t \in T}, \bar{x})$. Indeed, by (4.6) we have, for all t with $\mathcal{I}_t \neq \emptyset$ and all $i \in I_t$,

$$\begin{aligned} 0 &\leq \limsup_{k \to \infty} \left(g_{t_{i,k}}(\bar{x}) - g(\bar{x}) + \rho_{i,k}/\lambda_{i,k} + \gamma_k \right) \\ &\leq g_t(\bar{x}) - g(\bar{x}) + \rho_i/\lambda_i, \end{aligned}$$

where the last inequality is a consequence of the upper semicontinuity of the function $t \mapsto g_t(z)$. Thus,

$$\begin{aligned} \eta_t = g_t(\bar{x}) - g(\bar{x}) + \rho_t &= \frac{1}{\lambda_t} \left(\lambda_t g_t(\bar{x}) - \lambda_t g(\bar{x}) + \sum_{i \in \mathcal{I}_t} \rho_i \right) \\ &= \frac{1}{\lambda_t} \left(\sum_{i \in \mathcal{I}_t} \lambda_i \left(g_t(\bar{x}) - g(\bar{x}) + \rho_i/\lambda_i \right) \right) \geq 0, \end{aligned}$$

which means that $\eta \in \mathbb{R}_+^{(T)}$. Furthermore,

$$\sum_{t \in T} \lambda_t = \sum_{i=1}^{n+1} \lambda_i = 1,$$

which implies that $\lambda \in \Delta(T)$.

Now, using again the definition of λ and η, we have that

$$\sum_{t\in\operatorname{supp}\lambda}\lambda_t\left(g(\bar{x})-g_t(\bar{x})+\eta_t\right)=\sum_{t\in\operatorname{supp}\lambda}\lambda_t\rho_t=\sum_{t\in\operatorname{supp}\lambda}\sum_{i\in\mathcal{I}_t}\rho_i$$
$$=\sum_{i\mid\lambda_i\neq0}\rho_i\leq\sum_{i=1}^{n+1}\rho_i=\varepsilon.$$

In this way, we concluded that $(\lambda,\eta)\in\Lambda^{\varepsilon}((g_t)_{t\in T},\bar{x})$.

Second, using (4.5) we have

$$u^*=\sum_{i=1}^{n+1}\lambda_iu_i^*=\sum_{t\in T}\sum_{i\in\mathcal{I}_t}\lambda_iu_i^*=\sum_{t\in\operatorname{supp}\lambda}\lambda_t\left(\frac{1}{\lambda_t}\sum_{i\in\mathcal{I}_t}\lambda_iu_i^*\right)$$
$$=\sum_{t\in\operatorname{supp}\lambda}\lambda_tu_t^*.$$

Finally, let us check that $u_t^*\in\partial_{\eta_t}g_t(\bar{x})$ for all $t\in\operatorname{supp}\lambda$. Indeed, for a fixed $t\in\operatorname{supp}\lambda_t$, and all $i\in\mathcal{I}_t$, $k\in\mathbb{N}$,

$$\langle u_{i,k}^*,y-\bar{x}\rangle\leq g_{t_{k,i}}(y)-g_{t_{k,i}}(\bar{x})+\nu_{i,k}+\gamma_k\quad\forall y\in\mathbb{R}^n.$$

Whence, for all $y\in\mathbb{R}^n$,

$$\left\langle\sum_{i\in\mathcal{I}_t}\lambda_{i,k}u_{i,k}^*,y-\bar{x}\right\rangle\leq\sum_{i\in\mathcal{I}_t}\lambda_{i,k}g_{t_{k,i}}(y)+\sum_{i\in\mathcal{I}_t}\lambda_{i,k}\left(-g_{t_{k,i}}(\bar{x})+\nu_{i,k}+\gamma_k\right).\tag{4.8}$$

Now, the upper semicontinuity of the function $t\mapsto g_t(z)$ entails (recall that $t_{i,k}\to t$ for all $i\in\mathcal{I}_t$)

$$\limsup_{k\to\infty}\sum_{i\in\mathcal{I}_t}\lambda_{i,k}g_{t_{k,i}}(y)\leq\sum_{i\in\mathcal{I}_t}\lambda_i\limsup_{k\to\infty}g_{t_{k,i}}(y)\leq\sum_{i\in\mathcal{I}_t}\lambda_ig_t(y)=\lambda_tg_t(y).\tag{4.9}$$

Moreover, by (4.6) we have

$$\sum_{i\in\mathcal{I}_t}\lambda_{i,k}\left(-g_{t_{k,i}}(\bar{x})+\nu_{i,k}\right)\leq\sum_{i\in\mathcal{I}_t}\lambda_{i,k}\left(-g(\bar{x})+\rho_{i,k}/\lambda_{i,k}+\gamma_k\right).$$

Then, taking the limit as $k\to\infty$ we get that

$$\begin{aligned}\limsup_{k\to+\infty}\left(\sum_{i\in\mathcal{I}_t}\lambda_{i,k}\left(-g_{t_{k,i}}(\bar{x})+\nu_{i,k}\right)\right)&\leq\sum_{i\in\mathcal{I}_t}\lambda_i\left(-g(\bar{x})+\rho_i/\lambda_i\right)\\&=-\lambda_tg_t(\bar{x})+\lambda_t\left(g_t(\bar{x})-g(\bar{x})+\rho_t\right)\\&=-\lambda_tg_t(\bar{x})+\lambda_t\eta_t.\end{aligned}\tag{4.10}$$

Therefore, taking the limits in (4.8), and using inequalities (4.9) and (4.10), we obtain

$$\lambda_t\langle u_t^*,y-\bar{x}\rangle\leq\lambda_t\left(g_t(y)-g_t(\bar{x})+\eta_t\right)\quad\forall y\in\mathbb{R}^n,$$

and, dividing the last inequality by λ_t, we are done. ∎

Remark 4.5. Observe that while Proposition 4.2 provides an expression for the $\partial g(\bar{x})$ under the compacity/upper semicontinuity assumptions (i) and (ii), Proposition 4.3 and Theorem 4.4 provide formulas for $\partial_\varepsilon g(\bar{x})$, which is a rather more complicated issue as g is a supremum function. Formula (4.1) given in Proposition 4.3 only requires that $\operatorname{int}(\operatorname{dom} g) \neq \emptyset$, whereas Theorem 4.4 assumes that $\operatorname{dom} g = \mathbb{R}^n$, and the compactness of all the sets $T_\gamma(\bar{x})$, $\gamma > 0$ to avoid the use of ε-normal sets and compute the estimation with exact multipliers $(\lambda, \eta) \in \Lambda^\varepsilon((g_t)_{t\in T}, \bar{x})$. Finally, condition (i) in Theorem 4.4 is actually implied by the compactness of the whole index set T together with assumption (ii).

Next we establish some corollaries under additional assumptions on the data functions.

COROLLARY 4.6. *In the setting of Theorem* 4.4, *assume that the functions g_t are affine, i.e., $g_t(x) = \langle c_t, x\rangle + \alpha_t$ with $c_t \in \mathbb{R}^n$ and $\alpha_t \in \mathbb{R}$. Then, we have*

$$\partial_\varepsilon g(\bar{x}) = \bigcup \left[\sum_{t\in T} \lambda_t c_t : \quad \begin{array}{l} (\lambda_t) \in \Delta(T) \text{ and} \\ \sum\limits_{t\in T} \lambda_t \left(g(\bar{x}) - \langle c_t, \bar{x}\rangle - \alpha_t \right) \leq \varepsilon \end{array} \right].$$

COROLLARY 4.7. *In the setting of Theorem* 4.4, *assume that T is a convex set of a linear space and that the function $t \mapsto g_t(z)$ is concave for all $z \in \mathbb{R}^n$. Then, we have*

$$\partial_\varepsilon g(\bar{x}) = \bigcup \left[\partial_\eta g_t(\bar{x}) : \begin{array}{c} \eta \geq 0 \text{ and } t \in T \text{ such that} \\ g(\bar{x}) - g_t(\bar{x}) \leq \varepsilon - \eta \end{array} \right]. \tag{4.11}$$

Proof. First consider u^* in the right-hand side of (4.11), i.e., then there exist $t \in T$ and $\eta \geq 0$ such that $u^* \in \partial_\eta g_t(\bar{x})$ and $g(\bar{x}) - g_t(\bar{x}) \leq \varepsilon - \eta$. Then, for all $y \in \mathbb{R}^n$,

$$\begin{aligned} \langle u^*, y - x\rangle &\leq g_t(y) - g_t(\bar{x}) + \eta \\ &\leq g(y) - g(\bar{x}) + \varepsilon, \end{aligned}$$

which implies that $u^* \in \partial_\varepsilon g(\bar{x})$.

Conversely, consider $u^* \in \partial_\varepsilon g(\bar{x})$. Then, by Theorem 4.4, there exists $(\lambda, \eta) \in \Delta(T) \times \mathbb{R}_+^{(T)}$ such that $u^* = \sum_{t\in T} \lambda_t u_t^*$ for some $u_t^* \in \partial_{\eta_t} g_t(\bar{x})$ and

$$\sum_{t\in \operatorname{supp}\lambda} \lambda_t \left(g(\bar{x}) - g_t(\bar{x}) + \eta_t \right) \leq \varepsilon.$$

Let us take

$$t^* := \sum_{t\in \operatorname{supp}\lambda} \lambda_t t \in T$$

and

$$\eta^* := \sum_{t\in \operatorname{supp}\lambda} \lambda_t \left(\eta_t + g_{t^*}(\bar{x}) - g_t(\bar{x}) \right).$$

By the concavity of $t \mapsto g_t(\bar{x})$ we have that $\sum_{t\in \operatorname{supp}\lambda} \lambda_t g_t(\bar{x}) \leq g_{t^*}(\bar{x})$, and consequently $\eta^* \geq 0$. Furthermore, by the definition of η^* and the fact that $(\lambda, \eta) \in$

$\Lambda^{\varepsilon}((g_t)_{t\in T}, \bar{x})$, we obtain

$$\begin{aligned} g(\bar{x}) - g_{t^*}(\bar{x}) &= g(\bar{x}) - \sum_{t\in \mathrm{supp}\lambda} \lambda_t g_t(\bar{x}) + \sum_{t\in \mathrm{supp}\lambda} \lambda_t g_t(\bar{x}) - g_{t^*}(\bar{x}) \\ &= \sum_{t\in \mathrm{supp}\lambda} \lambda_t \left(g(\bar{x}) - g_t(\bar{x}) + \eta_t\right) - \sum_{t\in \mathrm{supp}\lambda} \lambda_t \left(\eta_t + g_{t^*}(\bar{x}) - g_t(\bar{x})\right) \\ &\leq \varepsilon - \eta^*. \end{aligned}$$

Finally, let us check that $u^* \in \partial_{\eta^*} g_t(\bar{x})$. Indeed, if we consider $y \in \mathbb{R}^n$, by the concavity of $t \mapsto g_t(y)$ we have $\sum_{t\in \mathrm{supp}\lambda} \lambda_t g_t(y) \leq g_{t^*}(y)$, and

$$\begin{aligned} \langle u^*, y - x\rangle &= \sum_{t\in \mathrm{supp}\lambda} \lambda_t \langle u_t^*, y - x\rangle \leq \sum_{t\in \mathrm{supp}\lambda} \lambda_t \left(g_t(y) - g_t(\bar{x}) + \eta_t\right) \\ &\leq g_{t^*}(y) - g_{t^*}(\bar{x}) + \sum_{t\in \mathrm{supp}\lambda} \lambda_t \left(\eta_t + g_{t^*}(\bar{x}) - g_t(\bar{x})\right) \\ &\leq g_{t^*}(y) - g_{t^*}(\bar{x}) + \eta^*. \end{aligned}$$

□

Following the same arguments as in the above result we can establish the following result for the exact subdifferential, in the framework of Proposition 4.2.

COROLLARY 4.8. *In the setting of Proposition* 4.2, *assume that T is a convex set of a linear space and the function $t \mapsto g_t(z)$ is concave for all $z \in \operatorname{dom} g$. Then, we have*

$$\partial g(\bar{x}) := \bigcup_{t\in T(\bar{x})} \partial g_t(\bar{x}).$$

5. Semi-infinite programming.

In this section we consider the following class of *semi-infinite programming* problems with DC data functions:

$$\begin{aligned} &\min \psi(x) \\ &\text{s.t. } \varphi_t(x) \leq 0 \ \forall t \in T. \end{aligned} \tag{5.1}$$

The multipliers associated with ψ are distinguished by using the symbol "ˆ" over the multiplier, that is, $\hat{\lambda}, \hat{\eta}$, etc. In what follows we appeal to the set.

$$T_\varepsilon(\bar{x}) := \{t \in T : \ \varphi_t(\bar{x}) \geq \sup_{s\in T} \varphi_s(\bar{x}) - \varepsilon\}.$$

If $\sup_{s\in T} \varphi_s(\bar{x}) = 0$, the indices in $T_\varepsilon(\bar{x})$ are called ε-*active* at $\bar{x}$.

We divide this section into two parts devoted to studying global minima and local minima of the (5.1) separately.

5.1. Global minima.

In this section we provide necessary and sufficient optimality conditions for nonconvex optimization problems with DC data functions, where we benefit from results established in the previous section applied to a supremum function (4.4) involving all the data functions. Let us establish the main assumptions over the data functions.

Assumption 5.1. Our data functions $\varphi_t, t \in T$, and ψ belong to $\Gamma_h(\mathbb{R}^n)$ for some convex function $h : \mathbb{R}^n \to \mathbb{R}$, and the following hold:

- All of them are finite-valued, that is, $\psi, \varphi_t : \mathbb{R}^n \to \mathbb{R}$, $t \in T$.
- For all $z \in \mathbb{R}^n$ the function $t \mapsto \varphi_t(z)$ is upper-semicontinuous.

- The supremum function $\sup_{t\in T}\varphi_t$ is finite-valued.

THEOREM 5.2. *Let $\bar{x}$ be a feasible point of* (5.1). *Suppose that Assumption* 5.1 *holds and additionally that, for all $\varepsilon \geq 0$, the set $T_\varepsilon(\bar{x})$ is compact. Then, if $\bar{x}$ is a minimum of* (5.1), *for every $\varepsilon \geq 0$ and all $u \in \partial_\varepsilon h(\bar{x})$ there are $(\lambda_t), (\eta_t) \in \mathbb{R}_+^{(T)}$, $\hat{\lambda}, \hat{\eta} \in \mathbb{R}_+$, $\sum_{t\in T}\lambda_t + \hat{\lambda} = 1$, such that*

$$u \in \hat{\lambda}\partial_{\hat{\eta}}(\psi + h)(\bar{x}) + \sum_{t\in T}\lambda_t\partial_{\eta_t}(\varphi_t + h)(\bar{x}) \tag{5.2}$$

with

$$\sum_{t\in T}\lambda_t\left(\eta_t - \varphi_t(\bar{x})\right) + \hat{\lambda}\hat{\eta} \leq \varepsilon. \tag{5.3}$$

The converse is true provided that $\bar{x}$ is a feasible point of (5.1) *and* (5.2) *always holds with multiplier $\hat{\lambda} \neq 0$.*

Consider a point $\hat{t} \notin T$ and define $\hat{T} := T \cup \{\hat{t}\}$ together with the functions $\{g_t,\ t \in \widehat{T}\}$, given by

$$g_t := \begin{cases} \varphi_t + h & \text{if } t \in T, \\ \psi + h & \text{if } t = \hat{t}. \end{cases}$$

Consider now the supremum function

$$g := \sup_{t\in\hat{T}} g_t,$$

and the following unconstrained DC optimization problem:

$$\min_{x\in\mathbb{R}^n}\ (g - h)\,. \tag{5.4}$$

Before presenting the proof of Theorem 5.2 we establish the relationship between problems (5.1) and (5.4). From now on, we shall assume w.l.o.g. that $\psi(\bar{x}) = 0$.

LEMMA 5.3. *If $\bar{x}$ is a minimum of* (5.1), *then $\bar{x}$ is a minimum of* (5.4).

Proof. The key is the relation

$$g - h = \sup\left\{\varphi_t,\ t \in T;\ \psi\right\}. \tag{5.5}$$

Suppose that $\bar{x}$ is a minimum of (5.1) and, reasoning by contradiction, let $x \in \mathbb{R}^n$ be such that

$$(g - h)(x) < (g - h)(\bar{x}) = \psi(\bar{x}) = 0.$$

Then, by (5.5),

$$\varphi_t(x) < 0\ \forall t \in T,$$

and x is feasible for (5.1). Moreover,

$$\begin{aligned}(g - h)(\bar{x}) = \psi(\bar{x}) \leq \psi(x) &= \psi(x) + h(x) - h(x) \\ &\leq g(x) - h(x),\end{aligned}$$

which constitutes a contradiction, and $\bar{x}$ is also a minimum of (5.4). ∎

Proof of Theorem 5.2. By Lemma 5.3 we have that $\bar{x}$ is a minimum of (5.4). Now, by the definition of subdifferential, we have that $\bar{x}$ is a minimum of (5.4) if and only if

$$0_n \in \partial\,(g-h)\,(\bar{x}),$$

and, by Proposition 4.1, this happens if and only if

$$\partial_\varepsilon h(\bar{x}) \subseteq \partial_\varepsilon g(\bar{x}) \;\forall \varepsilon \geq 0. \tag{5.6}$$

Now, Theorem 4.4 allows us to express the ε -subdifferential of g at $\bar{x}$, as $\widehat{t}$ is an isolated point of any $\hat{T}_\varepsilon(\bar{x})$. In fact, inclusion (5.6) is equivalent to having that, for all $\varepsilon \geq 0$ and every $u \in \partial_\varepsilon h(\bar{x})$, there exist $((\lambda_t)_{t\in T}, \hat{\lambda}, (\eta_t)_{t\in T}, \hat{\eta}) \in \Lambda^\varepsilon((g_t)_{t\in\hat{T}}, \bar{x})$ such that

$$u \in \hat{\lambda}\partial_{\hat{\eta}}(\psi+h)(\bar{x}) + \sum_{t\in T} \lambda_t \partial_{\eta_t}(\varphi_t+h)(\bar{x}).$$

Finally, let us check that the elements of $\Lambda^\varepsilon((g_t)_{t\in\hat{T}}, \bar{x})$ satisfy (5.3). Indeed, $((\lambda_t)_{t\in T}, \hat{\lambda}, (\eta_t)_{t\in T}, \hat{\eta}) \in \Lambda^\varepsilon((g_t)_{t\in\hat{T}}, \bar{x})$ if and only if $(\lambda_t), (\eta_t) \in \mathbb{R}_+^{(T)}$, $\hat{\lambda}, \hat{\eta} \in \mathbb{R}_+$, $\sum_{t\in T} \lambda_t + \hat{\lambda}_t = 1$, and

$$\sum_{t\in T} \lambda_t \left(g(\bar{x}) - g_t(\bar{x}) + \eta_t\right) + \hat{\lambda}\left(g(\bar{x}) - g_{\hat{t}}(\bar{x}) + \hat{\eta}\right) \leq \varepsilon. \tag{5.7}$$

Then, using in the last inequality that $g(\bar{x}) - h(\bar{x}) = \psi(\bar{x}) = 0$, we see that (5.7) gives rise to (5.3).

Now, to prove the converse, consider $y \in \mathbb{R}^n$, a feasible point of the optimization problem (5.1), that is, $\varphi_t(y) \leq 0$ for all $t \in T$. On the one hand, by the continuity of h, there exists $x^* \in \partial h(y)$, so

$$\begin{aligned} h(y) - h(\bar{x}) &= \langle x^*, y\rangle - h^*(x^*) - h(\bar{x}) \\ &= \langle x^*, y - x\rangle - (h^*(x^*) + h(\bar{x}) - \langle x^*, \bar{x}\rangle) \\ &= \langle x^*, y - \bar{x}\rangle - \varepsilon, \text{ with } \varepsilon := h^*(x^*) + h(\bar{x}) - \langle x^*, \bar{x}\rangle \geq 0. \end{aligned}$$

By the definition of the ε-subdifferential, we have that $x^* \in \partial_\varepsilon h(\bar{x})$. Therefore, we conclude that

$$h(y) - h(\bar{x}) + \varepsilon = \langle x^*, y - \bar{x}\rangle \text{ for some } x^* \in \partial_\varepsilon h(\bar{x}). \tag{5.8}$$

On the other hand, by (5.2), for some suitable numbers $(\lambda_t), (\eta_t) \in \mathbb{R}_+^{(T)}$, $\hat{\lambda}, \hat{\eta} \in \mathbb{R}_+$, we have

$$\begin{aligned} \langle x^*, y - \bar{x}\rangle &\leq \hat{\lambda}\left(\psi(y) + h(y) - \psi(\bar{x}) - h(\bar{x}) + \hat{\eta}\right) \\ &\quad + \sum_{t\in\sup\lambda_t} \lambda_t \left(\varphi_t(y) + h(y) - \varphi_t(\bar{x}) - h(\bar{x}) + \eta_t\right) \\ &\leq \hat{\lambda}\left(\psi(y) - \psi(\bar{x})\right) + \hat{\lambda}\left(h(y) - h(\bar{x})\right) + \hat{\lambda}\hat{\eta} \\ &\quad + \sum_{t\in\sup\lambda_t} \lambda_t \left(-\varphi_t(\bar{x}) + \eta_t\right) + \sum_{t\in\sup\lambda_t} \lambda_t \left(h(y) - h(\bar{x})\right) \\ &= \hat{\lambda}\left(\psi(y) - \psi(\bar{x})\right) + \left(h(y) - h(\bar{x})\right) + \sum_{t\in T} \lambda_t \left(\eta_t - \varphi_t(\bar{x})\right) + \hat{\lambda}\hat{\eta} \\ &\leq \hat{\lambda}\left(\psi(y) - \psi(\bar{x})\right) + \left(h(y) - h(\bar{x})\right) + \varepsilon. \end{aligned}$$

So, we conclude that

$$\langle x^*, y - \bar{x}\rangle \leq \hat{\lambda}\left(\psi(y) - \psi(\bar{x})\right) + \left(h(y) - h(\bar{x})\right) + \varepsilon. \tag{5.9}$$

Hence, combining (5.8) and (5.9), we yield $0 \leq \hat{\lambda}\left(\psi(y) - \psi(\bar{x})\right)$. Finally, using our extra assumption we have that necessarily $\hat{\lambda} \neq \emptyset$, which concludes the proof. ∎

Remark 5.4. We can provide a characterization of optimal solutions to problems with an abstract constraint

$$\begin{aligned} &\min \psi(x)\\ \text{s.t. }\ &\varphi_t(x) \leq 0 \ \forall t \in T,\\ &x \in C \end{aligned}$$

with C being a closed convex set, by using the characterization given in Proposition 4.3.

Since one of the most important classes of DC functions corresponds to the class $\Gamma_\rho(\mathbb{R}^n)$, in the following corollary we deal with the particular case when the data belongs to this family of functions.

COROLLARY 5.5. *Assume that the data functions ψ, $\varphi_t, t \in T$, belong to $\Gamma_\rho(\mathbb{R}^n)$ for some $\rho > 0$. In the setting of Theorem* 5.2, *if $\bar{x}$ is a minimum of* (5.1), *then for every $\varepsilon \geq 0$ and every $u \in \mathbb{B}(\bar{x}, \sqrt{\frac{2\varepsilon}{\rho}})$ there are $(\lambda_t), (\eta_t) \in \mathbb{R}_+^{(T)}$, $\hat{\lambda}, \hat{\eta} \in \mathbb{R}_+$, $\sum_{t\in T}\lambda_t + \hat{\lambda} = 1$, satisfying* (5.3) *and such that*

$$\rho u \in \hat{\lambda}\partial_{\hat{\eta}}(\psi + \rho J)(\bar{x}) + \sum_{t\in T}\lambda_t \partial_{\eta_t}(\varphi_t + \rho J)(\bar{x}). \tag{5.10}$$

The converse is true provided that $\bar{x}$ is a feasible point of (5.1) *and* (5.10) *always holds with multiplier $\hat{\lambda} \neq 0$.*

Proof. The proof follows from Theorem 5.2 by taking into account that $\partial_\varepsilon J(\bar{x}) = \mathbb{B}(\bar{x}, \sqrt{2\varepsilon})$ for all $\varepsilon \geq 0$. ∎

Remark 5.6. It is worth mentioning that the assumptions of the above results are weaker than the classical assumptions for nonconvex semi-infinite programming where, in a majority of cases, a certain assumption of continuity of the function $(t, x) \mapsto \nabla_x \varphi_t(x)$ is made (see, e.g., [23]). Other weaker conditions are also assumed in diverse works, but they are also stronger that our assumptions. Let us provide a very simple example. Considering a function $c : [0, 1] \to \mathbb{R}$, which is not continuous, but upper-semicontinuous, for instance, $c(t) := \mathbf{1}_{[t_i,1]}$, where t_i is some point in $]0, 1[$. Consequently, the function $t \mapsto c(t)z^2$ is just upper-semicontinuous for every $z \in \mathbb{R}$.

The following corollary is the application of the above result and Proposition 3.2, which constitutes a criterion to determine when the function belongs to $\Gamma_h(\mathbb{R}^n)$.

COROLLARY 5.7. *Under the Assumption* 5.1, *assume that ψ, $\varphi_t, t \in T$, belong to $\mathcal{C}^2(\mathbb{R}^n)$, and suppose that there exists a $\mathcal{C}^2(\mathbb{R}^n)$-convex function h such that, for all $x \in \mathbb{R}^n$,*

$$\begin{aligned} \nabla^2\varphi_t(x) &\succcurlyeq -\nabla^2 h(x) \ \forall t \in T,\\ \nabla^2\psi(x) &\succcurlyeq -\nabla^2 h(x). \end{aligned} \tag{5.11}$$

Then, in the setting of Theorem 5.2*, if* $\bar{x}$ *is a minimum of* (5.1)*, then for every* $\varepsilon \geq 0$ *and every* $u \in \partial_\varepsilon h(x)$ *there are* $(\lambda_t), (\eta_t) \in \mathbb{R}_+^{(T)}$, $\hat{\lambda}, \hat{\eta} \in \mathbb{R}_+$, $\sum_{t\in T} \lambda_t + \hat{\lambda}_t = 1$, *satisfying* (5.3)*and such that*

$$u \in \hat{\lambda}\partial_{\hat{\eta}} (\psi + h) (\bar{x}) + \sum_{t\in T} \lambda_t \partial_{\eta_t} (\varphi_t + h) (\bar{x}). \tag{5.12}$$

The converse is true provided that $\bar{x}$ *is a feasible point of* (5.1) *and* (5.12) *always holds with multiplier* $\hat{\lambda} \neq 0$.

Proof. By (5.11) and Proposition 3.2 we have that the functions ψ, $\varphi_t, t \in T$, belong to $\Gamma_h(\mathbb{R})$. Thus, Theorem 5.2 yields the result. ∎

Example 5.8. Consider a family of $n \times n$-symmetric matrices $A(t)$, n-vectors $b(t)$ and d, and numbers $c(t)$, with $t \in T$, and define the following optimization problem:

$$\begin{aligned} &\min \langle d, x\rangle \\ &\text{s.t. } (1/2) \langle A(t)z, z\rangle + \langle b(t), z\rangle + c(t) \leq 0 \ \forall t \in T. \end{aligned} \tag{5.13}$$

An easy criterion to verify (5.11) is that the infimum of the smallest eigenvalues of the matrices $A(t)$ be finite. This condition, formulated as an optimization problem, is given by

$$\inf_{(z,t)\in\mathbb{S}^{n-1}\times T} z^\top A(t)z > -\infty,$$

where

$$\mathbb{S}^{n-1} := \left\{ z \in \mathbb{R}^n : \sum_{i=1}^n z_i^2 = 1 \right\}.$$

Consequently, under the assumptions of Theorem 5.2, in particular, that the function

$$t \mapsto \varphi_t(z) := (1/2) \langle A(t)z, z\rangle + \langle b(t), z\rangle + c(t)$$

is upper-semicontinous for all $z \in \mathbb{R}^n$, and $\sup_{t\in T} \varphi_t$ is finite-valued, a necessary and sufficient condition for optimality of (5.13) is given by Corollary 5.7. Indeed, consider

$$\rho := \max \left\{ 0, -\inf_{(z,t)\in\mathbb{S}^{n-1}\times T} z^\top A(t)z \right\},$$

as well as the convex function $h(x) = \frac{\rho}{2}\|x\|^2$, where $\|\cdot\|$ is the Euclidean norm. Then, we have that for all $t \in T$

$$\nabla^2\varphi_t(x) = A(t) \succcurlyeq -\nabla^2 h(x) = -\rho I_d,$$

where I_d represents the identity matrix.

5.2. Local minima. In this section we study local minima of (5.1), with functions ψ, $\varphi_t, t \in T$, defined on $\mathbb{R}^n$ and values in $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$. Let us establish our main assumptions in this subsection, which are similar to Assumptions 5.1, but of local nature.

Assumption 5.9 (at $\bar{x}$). There exists a convex closed neighborhood U of $\bar{x}$ such that ψ, $\varphi_t, t \in T$, belong to $\Gamma_h(U)$ for some function $h \in \Gamma_0(\mathbb{R}^n)$ which is Fréchet differentiable at $\bar{x}$. Additionally the following hold:

- The data functions ψ, $\varphi_t, t \in T$, are finite-valued on U.
- For some $\varepsilon_0 > 0$, the set $T_{\varepsilon_0}(\bar{x})$ is compact.
- For all $z \in U$ the function $t \mapsto \varphi_t(z)$ is upper-semicontinuous on $T_{\varepsilon_0}(\bar{x})$.
- For all $z \in U$, we have $\sup_{t\in T} \varphi_t(z) < +\infty$.

THEOREM 5.10. *Under the Assumption* 5.9, *let $\bar{x}$ be a local minimum of* (5.1). *Then,*

$$0_n \in \operatorname{co}\left(\hat{\partial}\psi(\bar{x}) \cup \bigcup \left[\hat{\partial}\varphi_t(\bar{x}) : \ t \in T(\bar{x})\right]\right). \tag{5.14}$$

In addition, if $0_n \notin \operatorname{co}(\bigcup[\hat{\partial}\varphi_t(\bar{x}) : \ t \in T(\bar{x})])$, then

$$0_n \in \hat{\partial}\psi(\bar{x}) + \operatorname{cone}\left(\bigcup \left[\hat{\partial}\varphi_t(\bar{x}) : \ t \in T(\bar{x})\right]\right). \tag{5.15}$$

Proof. Without loss of generality we can assume that $\psi(\bar{x}) = 0$, that U is small enough to guarantee that h is finite on U, and that $\bar{x}$ is a minimum of (5.1) over U. As in the proof of Theorem 5.2, we consider a point $\hat{t} \notin T$ and define $\hat{T} = T \cup \{\hat{t}\}$ and the functions $\{g_t\}_{t\in\hat{T}}$

$$g_t := \begin{cases} \varphi_t + h + \delta_U & \text{if } t \in T, \\ \psi + h & \text{if } t = \hat{t}, \end{cases}$$

as well as the supremum function

$$g := \sup_{t\in\hat{T}} g_t.$$

Now, by Lemma 5.3 we know that $\bar{x}$ is a minimum of the optimization problem

$$\min_{x\in\mathbb{R}^n} \ (g - h)(x),$$

and, consequently, $0_n \in \partial (f - g)(\bar{x})$. Then, using Proposition 4.1 we have that

$$\nabla h(\bar{x}) \in \partial g(\bar{x}). \tag{5.16}$$

Now, applying Proposition 4.2 to the function g ($\hat{t}$ is isolated of $T_{\varepsilon_0}(\bar{x})$) we get

$$\partial g(\bar{x}) = \operatorname{co}\Bigg(\bigcup_{t\in\widehat{T}(\bar{x})} \partial g_t(\bar{x})\Bigg), \tag{5.17}$$

where

$$\begin{aligned} \widehat{T}(\bar{x}) &= \left\{t \in \hat{T} : \ g_t(\bar{x}) = g(\bar{x})\right\} \\ &= \{t \in T : \ g_t(\bar{x}) = g(\bar{x}) = h(\bar{x})\} \cup \left\{\widehat{t}\right\}, \end{aligned}$$

since we are assuming that $\psi(\bar{x}) = 0$ and $\bar{x}$ is feasible for (5.1). Moreover, since h is Fréchet differentiable at $\bar{x}$ and due to the sum rule (2.2), we have that

$$\begin{aligned} \partial (\varphi_t + h)(\bar{x}) &= \hat{\partial}\varphi_t(x) + \nabla h(\bar{x}), \\ \partial (\psi + h)(\bar{x}) &= \hat{\partial}\psi(x) + \nabla h(\bar{x}). \end{aligned} \tag{5.18}$$

Thus, combining (5.17) and (5.18) we obtain

$$\begin{aligned}\partial g(\bar{x}) &= \operatorname{co}\left(\partial(\psi+h)(\bar{x}) \cup \bigcup_{t\in T(\bar{x})} \partial\left(\varphi_t + h + \delta_U\right)(\bar{x})\right)\\ &= \operatorname{co}\left(\hat{\partial}\psi(\bar{x}) \cup \bigcup\left[\hat{\partial}\varphi_t(\bar{x}) : \ t\in T(\bar{x})\right]\right) + \nabla h(\bar{x}),\end{aligned}$$

where

$$T(\bar{x}) = \{t \in T : \ g_t(\bar{x}) = g(\bar{x})\}.$$

Therefore, using the above equality and (5.16) we get that (5.14) holds. Finally, by (5.14) there exits $\hat{\lambda} \in \mathbb{R}$ and $(\lambda_t) \in \mathbb{R}_+^{(T)}$ such that

$$0_n \in \hat{\lambda}\hat{\partial}\psi(\bar{x}) + \sum_{t\in T} \lambda_t \hat{\partial}\psi_t(\bar{x}), \tag{5.19}$$

and if $0_n \notin \operatorname{co}(\bigcup[\hat{\partial}\varphi_t(\bar{x}) : \ t\in T(\bar{x})])$, we have that necessarily $\hat{\lambda} \neq 0$. Thus, dividing (5.19) by $\hat{\lambda}$ we conclude that (5.15) is fulfilled. ∎

The following corollary represents a tighter necessary optimality condition when the data functions are smooth.

COROLLARY 5.11. *In the setting of Theorem* 5.10 *assume that the data* $\psi,\ \varphi_t, t \in T$, *are differentiable at* $\bar{x}$. *Then,*

$$0_n \in \operatorname{co}\left(\{\nabla\psi(\bar{x})\} \cup \{\nabla\varphi_t(\bar{x}) : \ t \in T(\bar{x})\}\right).$$

In addition, if $0_n \notin \operatorname{co}\{\nabla\varphi_t(\bar{x}) : \ t\in T(\bar{x})\}$, *then*

$$-\nabla\psi(\bar{x}) \in \operatorname{cone}\{\nabla\varphi_t(\bar{x}) : \ t\in T(\bar{x})\}.$$

6. DC cone-constraint optimization. This section is devoted to establishing necessary and sufficient conditions for cone-constraint optimization problems. More precisely we consider the following optimization problem:

$$\min \psi(x) \text{ s.t. } F(x) \in K, \tag{6.1}$$

where $\psi : \mathbb{R}^n \to \mathbb{R}$, $F : \mathbb{R}^n \to \mathbb{R}^m$ is a vector-valued mapping, and $K \subset \mathbb{R}^m$ is a closed convex cone with nonempty interior. We also consider the following basis of the polar cone K°:

$$K_1^\circ := \left\{y^* \in K^\circ : \sum_{i=1}^m |y_i^*| \le 1\right\} \equiv K^\circ \cap \mathbb{B}_1,$$

where $\mathbb{B}_1$ is the closed unit ball for the norm $\|\cdot\|_1$.

DEFINITION 6.1. *The mapping* $F : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ *is called a* DC vector valued mapping *on the open set* U *if there exists convex function* $h : U \to \mathbb{R}$, *called a* control function, *such that for all* $y^* \in \mathbb{R}^m$ *with* $\sum_{i=1}^m |y_i^*| = 1$ *the function*

$$x \mapsto \langle y^*, F(x)\rangle + h(x)$$

is convex on U.

Remark 6.2. In the original definition of vector valued DC function, given in [37], the authors consider a vector valued function $F : U \subseteq X \to Y$, where X and Y are Banach spaces, and U is an open set. In this work the mapping F is said to be a *delta-convex function* if and only if there exists a convex function $h : U \to \mathbb{R}$ such that for all $y^* \in Y^*$ with $\|y^*\| = 1$ the function $x \mapsto \langle y^*, F(x)\rangle + h(x)$ is convex. It is not difficult to prove that, in the Euclidean space, both definitions are equivalent. Indeed, consider another norm $\|\cdot\|$, and let us suppose that there exists $C > 0$ such that $\|x^*\|_1 \leq C\|x^*\|$ for all $x^* \in \mathbb{R}^n$. Then, for every $\|x^*\| = 1$ we have that

$$\langle x^*, F(x)\rangle + Ch(x) = \|x^*\|_1 \left(\left\langle \frac{x^*}{\|x^*\|_1}, F(x) \right\rangle + h(x) \right) + (C - \|x^*\|_1)h(x),$$

which is a convex function; therefore F is a DC function with control function Ch.

The next theorem establishes necessary and sufficient conditions of optimality of problem (6.1).

THEOREM 6.3. *Assume that F is a DC vector valued function on $\mathbb{R}^n$ with control function h and that $\psi \in \Gamma_h(\mathbb{R}^n)$. Then, if $\bar{x}$ is a minimum of* (6.1)*, then we have for every $\varepsilon \geq 0$*

$$\partial_\varepsilon h(\bar{x}) \subseteq \bigcup \left[\partial_\eta \left(\langle y^*, F\rangle + \lambda\psi + h \right)(\bar{x}) : \begin{array}{c} \exists (y^*, \lambda) \in C, \exists \eta \geq 0 \text{ s.t.} \\ \eta \leq \langle y^*, F\rangle(\bar{x}) + \varepsilon \end{array} \right], \tag{6.2}$$

where $\langle y^*, F\rangle(x) := \langle y^*, F(x)\rangle$ and

$$C = \operatorname{co}\left((K_1^\circ \times \{0\}) \cup \{(0_m, 1)\} \right).$$

The converse is true when $\bar{x}$ is a feasible point of (6.1) *and inclusion in* (6.2) *holds with the set in the right-hand side with multiplier $\lambda \neq 0$.*

In the proof of Theorem 6.3 we also assume that $\psi(\bar{x}) = 0$.

Now, for every $(y^*, \lambda) \in C$ we define the function $g_{(y^*,\lambda)} : \mathbb{R}^n \to \mathbb{R}$ defined by

$$g_{(y^*,\lambda)}(x) := \langle y^*, F(x)\rangle + \lambda\psi(x) + h(x).$$

Here, it is important to point out that, for each $(y^*, \lambda) \in C$, the function $g_{(y^*,\lambda)}$ is convex, because it can be written as the sum of convex functions. Actually, if $y^* = 0_m$, we have

$$g_{(0_m,\lambda)}(x) = (1 - \lambda)h(x) + \lambda \left(\psi(x) + h(x) \right), \tag{6.3}$$

and $g_{(0_m,\lambda)}$ is sum of two convex functions. Alternatively, $y^* \neq 0_m$, and then $\lambda < 1$. Considering $z^* = (1 - \lambda)^{-1} y^*$ and defining $\mu := \sum_{i=1}^m |z_i^*| \in (0, 1]$, we have that $x \mapsto \langle \mu^{-1} z^*, F\rangle(x) + h(x)$ is convex, and consequently

$$\begin{aligned} g_{(y^*,\lambda)}(x) &= \langle y^*, F(x)\rangle + \lambda\psi(x) + h(x) \\ &= (1 - \lambda)\mu \left(\langle \mu^{-1} z^*, F(x)\rangle + h(x) \right) \\ &\quad + (1 - \lambda)(1 - \mu)h(x) + \lambda \left(\psi(x) + h(x) \right), \end{aligned} \tag{6.4}$$

which shows that $g_{(y^*,\lambda)}$ is again sum of convex functions. Therefore, $g_{(y^*,\lambda)}$ is a convex function for all $(y^*, \lambda) \in C$.

Our approach in this section also relies on a supremum function; this is

$$g := \sup_{(y^*,\lambda) \in C} g_{(y^*,\lambda)}.$$

Furthermore, the function $(y^*, \lambda) \mapsto g_{(y^*,\lambda)}(z)$ is affine for all $z \in \mathbb{R}^n$.

Now, let us consider once again the unconstrained optimization problem

$$\min_{x\in\mathbb{R}^n} \ (g(x) - h(x)), \tag{6.5}$$

where

$$g(x) - h(x) = \sup_{(y^*,\lambda)\in C} \{\langle y^*, F(x)\rangle + \lambda\psi(x)\}.$$

We notice that, due to the fact that $(0_m, 0) \in C$,

$$g(x) - h(x) \geq 0 \ \forall x \in \mathbb{R}^n. \tag{6.6}$$

Next we formally establish the relation between problems (6.1) and (6.5). The proof of this lemma follows similar arguments to the proof of Lemma 5.3, so we omit the proof.

Lemma 6.4. *If $\bar{x}$ is a minimum of* (6.1)*, then $\bar{x}$ is a minimum of* (6.5)*.*

Proof. Suppose that $\bar{x}$ is a minimum of (6.1), and remember that we are assuming that $\psi(\bar{x}) = 0$. Since $g(\bar{x})-h(\bar{x}) = 0$, and by (6.6), it turns out that $\bar{x}$ is a a minimum of (6.5). ◻

Proof of Theorem 6.3. By Lemma 6.4 we have that $\bar{x}$ is a minimum of (6.5), and this happens if and only if $0_n \in \partial(g-h)(\bar{x})$. Thus, by Proposition 4.1 we have that $\bar{x}$ is a minimum of (6.5) if and only if

$$\partial_\varepsilon h(\bar{x}) \subseteq \partial_\varepsilon g(\bar{x}) \ \forall \varepsilon \geq 0. \tag{6.7}$$

Finally, using Corollary 4.7 we can express the subdifferential of g at $\bar{x}$, that is,

$$\partial_\varepsilon g(\bar{x}) = \bigcup \left[\partial_\eta g_{(y^*,\lambda)}(\bar{x}) : \begin{array}{c} \exists \eta \geq 0,\ (y^*,\lambda) \in C \text{ such that} \\ g(\bar{x}) - g_{(y^*,\lambda)}(\bar{x}) \leq \varepsilon - \eta \end{array} \right]. \tag{6.8}$$

From the fact that $g(\bar{x}) - h(\bar{x}) = \psi(\bar{x}) = 0$, we get that

$$\begin{aligned} & g(\bar{x}) - g_{(y^*,\lambda)}(\bar{x}) \leq \varepsilon - \eta \\ \Leftrightarrow \quad & h(\bar{x}) - \langle y^*, F(\bar{x})\rangle - h(\bar{x}) \leq \varepsilon - \eta \\ \Leftrightarrow \quad & -\langle y^* F(\bar{x})\rangle \leq \varepsilon - \eta, \\ \Leftrightarrow \quad & \eta \leq \langle y^*, F(\bar{x})\rangle + \varepsilon. \end{aligned} \tag{6.9}$$

Thus, using (6.7), (6.8), and (6.9), we get the result.

The converse follows similar arguments to the one given in Theorem 5.2, so we omit the proof. ◻

Now, we focus on necessary optimality conditions for a local optimal solution of problem (6.1). We refer to [27] and the references therein for similar results.

Theorem 6.5. *Assume that F is a DC mapping on an open convex neighborhood U of $\bar{x}$, with control function h which is differentiable at $\bar{x}$. Assume also that $\psi \in \Gamma_h(U)$. Then, if $\bar{x}$ is a local minimum of* (6.1)*, then there exists $u^* \in K^\circ$ such that either*

$$0 = \langle u^*, F(\bar{x})\rangle \ \text{and} \ 0_n \in \hat{D}F(\bar{x})(u^*) \tag{6.10}$$

or

$$0 = \langle u^*, F(\bar{x})\rangle \ \text{and} \ 0_n \in \hat{D}F(\bar{x})(u^*) + \hat{\partial}\psi(\bar{x}). \tag{6.11}$$

Proof. We proceed similarly to the proof of Theorem 6.3. Thus, using the same notation as in such a theorem, we have that $0_n \in \partial\,(g-h)\,(\bar{x})$. Thus, applying Proposition 4.1, if $\bar{x}$ is a local minimum of (6.1), then

$$\nabla h(\bar{x}) \in \partial g(\bar{x}).$$

Now, we use Corollary 4.8 to get the existence of $(y^*, \lambda) \in C$, where

$$C = \operatorname{co}\left(K_1^\circ \times \{0\} \cup \{(0_m, 1)\}\right),$$

such that

$$g(\bar{x}) - h(\bar{x}) = 0 = \langle y^*, F(\bar{x})\rangle$$

and

$$\nabla h(\bar{x}) \in \partial\left(\langle y^*, F\rangle + \lambda\psi + h\right)(\bar{x}). \tag{6.12}$$

Now, let us compute the right-hand side of the above inclusion. We analyze first the case $y^* = (1-\lambda)z^*$, $z^* \in \mathbb{B}_1 \setminus \{0_m\}$, and $\lambda \in [0, 1[$.

If we define $\mu := \sum_{i=1}^m |z_i^*|$, we have that $\langle y^*, F\rangle + \lambda\psi + h$ can be expressed as in (6.4), and the sum rule for the convex subdifferential (see, e.g., [32]) together with the sum rule for the Fréchet subdifferential (see (2.2)) yields

$$\begin{aligned}
\partial\left(\langle y^*, F\rangle + \lambda\psi + h\right)(\bar{x}) &= (1-\lambda)\mu\partial\left(\langle \mu^{-1}z^*, F\rangle + h\right)(\bar{x}) + (1-\lambda)(1-\mu)\nabla h(\bar{x}) \\
&\quad + \lambda\partial\left(\psi + h\right)(\bar{x}) \\
&= (1-\lambda)\mu\widehat{\partial}\left(\langle \mu^{-1}z^*, F\rangle + h\right)(\bar{x}) + (1-\lambda)(1-\mu)\nabla h(\bar{x}) \\
&\quad + \lambda\hat{\partial}\left(\psi + h\right)(\bar{x}) \\
&= \widehat{\partial}\left(\langle (1-\lambda)z^*, F\rangle\right)(\bar{x}) + \lambda\hat{\partial}\psi(\bar{x}) + \nabla h(\bar{x}) \\
&= \hat{D}F(\bar{x})(y^*) + \lambda\hat{\partial}\psi(\bar{x}) + \nabla h(\bar{x}).
\end{aligned}$$

Thus, replacing it in (6.12) we conclude that there exists $(y^*, \lambda) \in C$ such that the following equations hold:

$$0 = \langle y^*, F(\bar{x})\rangle \text{ and } 0_n \in \hat{D}F(\bar{x})(y^*) + \lambda\hat{\partial}\psi(\bar{x}). \tag{6.13}$$

It is easy to see that the last result is also valid when $y^* = 0_m$, since $\hat{D}F(\bar{x})(0_m) = 0_n$ (see (6.3)). Therefore, on the one hand if $\lambda = 0$, (6.13) implies that (6.10) holds defining $u^* = y^*$. On the other hand if $\lambda > 0$, we can divide (6.13) by λ, and conclude (6.11) by defining $u^* = \lambda^{-1}y^*$. ∎

7. Problems with DC/supremum objective function.

In this section we deal with the following optimization problem:

$$\begin{aligned} &\min\ f(x) \\ &\text{s.t. } x \in K, \end{aligned} \tag{7.1}$$

where

$$f := g - h,\ g = \sup_{t\in T_1} g_t, \quad \text{and } h = \sup_{t\in T_2} h_t,$$

T_1 and T_2 being two index sets and $K \subset \mathbb{R}^n$. In the following, we simply denote by $T_\varepsilon^g(x) := (T_1)_\varepsilon^g$ and $T_\varepsilon^h(x) := (T_1)_\varepsilon^h$. Let us consider a point $\bar{x} \in K$, and assume that the following assumptions hold for the problem (7.1) at $\bar{x}$.

Assumption 7.1 (at $\bar{x}$).

- All the functions $g_t, t \in T_1,\ h_t, t \in T_2$ are finite-valued convex functions defined on $\mathbb{R}^n$, and g and h are both finite everywhere.
- K is a closed convex set.
- For all $\varepsilon \geq 0$ the indices sets $T_\varepsilon^g(\bar{x})$ and $T_\varepsilon^h(\bar{x})$ are compact.
- For all $z \in \mathbb{R}^n$, the functions $T_1 \ni t \to g_t(z)$ and $T_2 \ni t \to h_t(z)$ are upper-semicontinuous.

THEOREM 7.2 (global solution). *Let us assume that Assumption* 7.1 *holds* $\bar{x} \in \mathbb{R}^n$. *Then,* $\bar{x}$ *is a minimum of problem* (7.1) *if and only if for every* $\varepsilon \geq 0$,

$$\bigcup_{(\lambda,\eta)\in\Lambda_h^\varepsilon} \sum_{t\in T_2} \lambda_t \partial_{\eta_t} h_t(\bar{x}) \subseteq \bigcup \left[\sum_{t\in T_1} \lambda_t \partial_{\eta_t} g_t(\bar{x}) + \mathrm{N}_K^{\varepsilon_2}(\bar{x}) : \begin{array}{c} (\varepsilon_1,\varepsilon_2) \in \Delta_2^\varepsilon \\ (\lambda,\eta) \in \Lambda_g^{\varepsilon_1} \end{array} \right], \tag{7.2}$$

where

$$\Lambda_g^\varepsilon = \left\{ (\lambda,\eta) \in \Delta(T_1) \times \mathbb{R}_+^{(T_1)} : \sum_{t\in \mathrm{supp}\lambda} \lambda_t \left(g(\bar{x}) - g_t(\bar{x}) + \eta_t \right) \leq \varepsilon \right\} \text{ and}$$

$$\Lambda_h^\varepsilon = \left\{ (\lambda,\eta) \in \Delta(T_2) \times \mathbb{R}_+^{(T_2)} : \sum_{\mathrm{supp}\lambda} \lambda_t \left(h(\bar{x}) - h_t(\bar{x}) + \eta_t \right) \leq \varepsilon \right\}.$$

Proof. It is obvious that problem (7.1) is equivalent to the optimization problem

$$\min_{x\in\mathbb{R}^n} \left((g + \delta_K)(x) - h(x) \right).$$

Then, by definition of the subdifferential, and using Proposition 4.1, we have that $\bar{x}$ is a minimum of problem (7.1) if and only if, for every $\varepsilon \geq 0$,

$$\partial_\varepsilon h(\bar{x}) \subseteq \partial_\varepsilon \left(g + \delta_K \right)(\bar{x}).$$

Now, according to the well-known formula for the ε-subdifferential of the sum of convex functions (see, e.g., [38, Theorem 2.8.3]) we have that

$$\partial_\varepsilon \left(g + \delta_K \right)(\bar{x}) = \bigcup_{(\varepsilon_1,\varepsilon_2)\in\Delta_2^\varepsilon} \left(\partial_{\varepsilon_1} g(\bar{x}) + \mathrm{N}_K^{\varepsilon_2}(\bar{x}) \right).$$

Now, we apply Theorem 4.4 to the functions g and h to get the right- and left-hand side of (7.2), respectively. This concludes the proof. ∎

Now, we establish the local optimality conditions for problem (7.1), under the following assumptions at $\bar{x}$.

Assumption 7.3 (at $\bar{x}$).

- The functions $g_t,\ t \in T_1$, and $h_t,\ t \in T_2$, belong to $\Gamma_0(\mathbb{R}^n)$, and g and h are both finite and continuous at $\bar{x}$.
- For some $\varepsilon_0 \geq 0$ the sets $T_{\varepsilon_0}^g(\bar{x})$ and $T_{\varepsilon_0}^h(\bar{x})$ are compact.
- For all $y \in \operatorname{dom} g$ and all $y \in \operatorname{dom} h$ the functions $t \to g_t(y)$ and $t \to h_t(z)$ are upper-semicontinuous on $T_{\varepsilon_0}^g(\bar{x})$ and $T_{\varepsilon_0}^h(\bar{x})$, respectively.

The proof of the following result is based on similar arguments to those used in the proof of Theorem 7.2, but appealing to Proposition 4.2 instead of Theorem 4.4. Therefore, we shall omit the proof.

THEOREM 7.4 (local solution). *Let us suppose that Assumption* 7.3 *holds at* $\bar{x}$. *Then, if* $\bar{x}$ *is a local optimal solution of problem* (7.1), *then*

$$\operatorname{co}\left(\bigcup_{t\in T^h(\bar{x})} \partial h_t(\bar{x})\right) \subseteq \operatorname{co}\left(\bigcup_{t\in T^g(\bar{x})} \partial g_t(\bar{x})\right) + \mathrm{N}_K(x).$$

Before presenting the following example let us establish the following formula for the ε-subdifferential of the *Asplund function*. We recall that, given a closed set $C \subset \mathbb{R}^n$, the Asplund function associated with C is defined as

$$A_C(x) = \sup_{c\in C}\left(\langle c, x\rangle - \frac{1}{2}\|c\|^2\right).$$

It is well-known that A_C is the conjugate of the function $\phi_C : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ defined by

$$\phi_C(y) := \frac{1}{2}\|y\|^2 + \delta_C(y)$$

and that

$$A_C(x) = \frac{1}{2}\left(\|x\|^2 - d_C^2(x)\right), \tag{7.3}$$

where d_C is the distance function to C.

PROPOSITION 7.5. *Let* $C \subseteq \mathbb{R}^n$ *be a closed set,* $x \in \mathbb{R}^n$, *and* $\varepsilon \geq 0$. *Then,*

$$\partial_\varepsilon A_C(x) = \left\{ \sum_{c\in \operatorname{supp}\lambda} \lambda_c \cdot c : \begin{array}{c} (\lambda_c) \in \Delta(C), \ \textit{and} \\ \sum\limits_{c\in\operatorname{supp}\lambda} \lambda_c\left(\|x-c\|^2 - d_C^2(x)\right) \leq 2\varepsilon \end{array} \right\}. \tag{7.4}$$

Particularly, for $\varepsilon = 0$ *the above formula reduces to*

$$\partial A_C(x) = \operatorname{co}\left(P_C(x)\right), \tag{7.5}$$

where

$$P_C(x) := \{c \in C : \|x-c\| = d_C(x)\}.$$

Proof. Let us check that the assumptions of Corollary 4.6 hold. Indeed, for every $c \in C$ the function $x \mapsto \langle x, c\rangle - \frac{1}{2}\|c\|^2$ is affine. Moreover, by applying (7.3) we observe that the set of ε-active indices at x is given by

$$C_\varepsilon(x) = \left\{ c \in C : \frac{1}{2}\|c-x\|^2 \leq \varepsilon + \frac{1}{2}d_C^2(x)\right\},$$

which is a compact set. Finally, the function $c \mapsto \langle z, c\rangle - \frac{1}{2}\|c\|^2$ is continuous for all $z \in \mathbb{R}^n$.

Therefore, applying Corollary 4.6 we get that

$$\partial_\varepsilon A_C(x) = \left\{ \sum_{c\in\operatorname{supp}\lambda} \lambda_c c : \begin{array}{c} (\lambda_c) \in \Delta(C), \ \text{and} \\ \sum\limits_{c\in\operatorname{supp}\lambda} \lambda_c\left(A_C(x) - \left(\langle x, c\rangle - \frac{1}{2}\|c\|^2\right)\right) \leq \varepsilon \end{array} \right\}.$$

Applying (7.3) again, we conclude the proof of (7.4). Finally, for $\varepsilon = 0$ the formula (7.4) reduces to (7.5). $\square$

Example 7.6. Given a closed convex set K and a closed set C , we consider the optimization problem

$$\begin{aligned} &\min \frac{1}{2} d_C^2(x), \\ &\text{s.t. } x \in K. \end{aligned} \tag{7.6}$$

Thanks to (7.3), the above minimization problem can be seen as a DC optimization problem using the function $g(x) = \frac{1}{2}\|x\|^2$ and the Asplund function $h(x) = A_C(x)$.

On the one hand, by Theorem 7.2 we have that $\bar{x}$ is a global solution of (7.6) if and only if (7.2) holds. Furthermore, by Corollary 4.6 we have that condition (7.2) of Theorem 7.2 can be written as for all $\varepsilon \geq 0$

$$\partial_\varepsilon A_C(\bar{x}) \subseteq \bigcup_{(\varepsilon_1,\varepsilon_2)\in\Delta_2^\varepsilon} \left\{\mathbb{B}(\bar{x}, \sqrt{2\varepsilon_1}) + \mathrm{N}_K^{\varepsilon_2}(\bar{x})\right\},$$

where in the right-hand side of the above inclusion we have used the well-known formula for the ε-subdifferential of the sum of convex functions (see, e.g., [38, Theorem 2.8.3]).

On the other hand, by Theorem 7.4 we get that if $\bar{x}$ is a local solution of (7.6), then

$$P_C(\bar{x}) \subseteq \bar{x} + \mathrm{N}_K(\bar{x}). \tag{7.7}$$

Let us emphasize that even the above condition is stronger that the classical Fermat rule for critical points for nonconvex sets C. Here, we recall that the Fermat rule for critical points (using the Mordukhovich, or the Clarke subdifferential ∂) should be $0_n \in \partial(\frac{1}{2}d_C^2)(\bar{x}) + \mathrm{N}_K(\bar{x})$, where $\bar{x} - P_C(\bar{x}) \subseteq \partial(\frac{1}{2}d_C^2)(\bar{x})$. Indeed, let us consider $C = C_1 \cup C_2$, where

$$\begin{aligned} C_1 &:= \left\{(x,y) \in [-1/2, 1/2] \times \mathbb{R} :\ y = \sqrt{1-x^2}\right\} \text{ and} \\ C_2 &:= \left\{(x,y) \in \mathbb{R}^2 :\ |x| > 1/2 \text{ and } y = \sqrt{3}/2\right\}. \end{aligned}$$

Moreover, consider $K = \{(x,y) :\ \ y = 0\}$. Then, with these sets the optimization problem (7.6) has a critical point at $(x,y) = (0,0)$. Indeed, $(0,1) \in P_C(0,0)$ and $\mathrm{N}_K(0,0) = \{(0,y) :\ y \in \mathbb{R}\}$. Nevertheless, the inclusion (7.7) is not satisfied at $(0,0)$, but the Fermat rule holds.

8. A penality approach to abstract programming problems. In this section we study a particular class of optimization problems with a DC objective function, whose importance justifies the convenience of writing the associated optimality conditions separately. More precisely, we consider

$$\begin{aligned} &\min g(x) - h(x) \\ &\text{s.t. } x \in C, \end{aligned}$$

where $g, h : \mathbb{R}^n \to \mathbb{R}$ are convex functions and C is closed set in $\mathbb{R}^n$.

In order to solve this problem, at least approximately, we consider a constant $\mu > 0$ and we provide sufficient optimality conditions for the following penalized problem:

$$\min_{x\in\mathbb{R}^n} \left(g(x) - h(x) + \frac{1}{2\mu} d_C^2(x)\right), \tag{8.1}$$

where $d_C^2(x)$ is the square of the distance function to C.

THEOREM 8.1. *The point $\bar{x}_\mu$ is a minimum of* (8.1) *if and only if*

$$\bigcup_{(\varepsilon_1,\varepsilon_2)\in\Delta_2^\varepsilon}\left(\partial_{\varepsilon_1}h(\bar{x}_\mu)+\frac{1}{\mu}C(\mu\varepsilon_2)\right)\subseteq\bigcup_{(\varepsilon_1,\varepsilon_2)\in\Delta_2^\varepsilon}\left(\partial_{\varepsilon_1}g(\bar{x}_\mu)+\frac{1}{\mu}\mathbb{B}(\bar{x}_\mu,\sqrt{2\mu\varepsilon_2})\right),$$

where

$$C(\varepsilon)=\left\{\sum_{c\in\operatorname{supp}\lambda}\lambda_c\cdot c:\begin{array}{c}(\lambda_c)\in\Delta(C),\ \textit{and}\\ \sum_{c\in\operatorname{supp}\lambda}\lambda_c\left(\|x-c\|^2-d_C^2(x)\right)\leq 2\varepsilon\end{array}\right\}.$$

In particular, if $\bar{x}_\mu$ is a minimum of (8.1), *then*

$$\partial h(\bar{x}_\mu)+\frac{1}{\mu}\operatorname{co}\left(P_C(\bar{x}_\lambda)\right)\subseteq\partial g(\bar{x}_\mu)+\frac{1}{\mu}\bar{x}_\mu.$$

Proof. Since

$$\frac{1}{2\mu}d_C^2(x)=\frac{1}{\mu}J(x)-\frac{1}{\mu}A_C(x),$$

where J and A_C are the duality map and the Asplund function, respectively, we can write problem (8.1) as

$$\min_{x\in\mathbb{R}^n}\left(\hat{g}(x)-\hat{h}(x)\right),$$

where

$$\hat{g}(x)=g(x)+\frac{1}{\mu}J(x)\text{ and }\hat{h}(x)=h(x)+\frac{1}{\mu}A_C(x).$$

Thus, by Proposition 4.1 we have that $\bar{x}_\mu$ is a minimum of (8.1) if and only if

$$\partial_\varepsilon\hat{h}(\bar{x})\subseteq\partial_\varepsilon\hat{g}(\bar{x})\ \forall\varepsilon\geq 0. \tag{8.2}$$

Now, by the sum rule for the ε-subdifferential (see, e.g., [38, Theorem 2.8.3]), we can compute the left- and the right-hand side of (8.2); it yields

$$\partial_\varepsilon\hat{h}(\bar{x}_\mu)=\bigcup_{(\varepsilon_1,\varepsilon_2)\in\Delta_2^\varepsilon}\left(\partial_{\varepsilon_1}h(\bar{x}_\mu)+\frac{1}{\mu}\partial_{\varepsilon_2/\mu}A_C(\bar{x}_\mu)\right), \tag{8.3}$$

$$\partial_\varepsilon\hat{g}(\bar{x}_\mu)=\bigcup_{(\varepsilon_1,\varepsilon_2)\in\Delta_2^\varepsilon}\left(\partial_{\varepsilon_1}g(\bar{x}_\mu)+\frac{1}{\mu}\partial_{\varepsilon_2/\mu}J(\bar{x}_\mu)\right), \tag{8.4}$$

where we have used that

$$\partial_{\varepsilon_2}\left(\frac{1}{\mu}A_C\right)(\bar{x}_\mu)=\frac{1}{\mu}\partial_{\varepsilon_2/\mu}A_C(\bar{x}_\mu)\text{ and }\partial_{\varepsilon_2}\left(\frac{1}{\mu}J\right)(\bar{x}_\mu)=\frac{1}{\mu}\partial_{\varepsilon_2/\mu}J(\bar{x}_\mu).$$

Now, we compute using (3.1) and (7.4) that

$$\frac{1}{\mu}\partial_{\varepsilon_2/\mu}A_C(\bar{x}_\mu) = \frac{1}{\mu}C(\mu\varepsilon_2), \text{ and} \frac{1}{\mu}\partial_{\varepsilon_2/\mu}J(\bar{x}_\mu) = \mathbb{B}(\bar{x}, \sqrt{2\mu\varepsilon_2}). \tag{8.5}$$

Finally, mixing (8.3) and (8.5) we conclude the proof. ∎

9. Concluding remarks. In this work, we provided necessary and sufficient optimality conditions for nonconvex optimization problems, which can be formulated as DC optimization problems involving suprema of convex functions. Our approach exploits some structural properties of this type of DC model and uses calculus rules for the ε-subdifferential of the supremum functional to provide the optimality conditions for global and local minima.

In section 4, we derived formulas for the subdifferential and the ε-subdifferential of the supremum function. Particularly, Theorem 4.4 is our working horse in the second part of the paper. In addition, it has its own interest as it characterizes the ε-subdifferential of the supremum functions, in contrast with Proposition 4.2, which only computes the subdifferential. Corollaries 4.7 and 4.8 analyze important particular cases, and they are applied later in the paper.

The main results in section 5 are Theorem 5.5 (global optimality) and Theorem 5.13 (local optimality), which provide necessary and sufficient KKT optimality conditions for semi-infinite optimization problems involving only functions in the family Γ_h. There are in the literature a few works related to our results in section 5, but the optimality conditions for the semi-infinite problems considered in those works are derived under more restrictive assumptions, and none of the data functions is a supremum of convex functions. For instance, in [11, Theorem 1] the objective function is DC but the constraints are convex. Moreover, the constraint qualification used in that paper is a kind of closedness condition leading to KKT-type optimality conditions involving finitely many active constraints. Another paper for parametrized DC problems under very similar assumptions as in [11] is [13].

In section 6, we applied our methodology to cone-constraint optimization problems with DC data function. Here, the main results are Theorems 6.3 and 6.5, again for global and local optimality, respectively. As far as we know, the global optimality conditions in Theorem 6.3 are new, and no similar result can be found in the literature. In contrast Theorem 6.5 can be compared with [27, Theorem 4.1]. Nevertheless, the necessary optimality conditions are expressed in terms of the Mordukhovich subdifferential and co-derivative, which is a larger object than the Fréchet subdifferential and co-derivative used in our approach.

Here, it is worth mentioning that our approach allowed us to get not only necessary conditions for optimality but also sufficient conditions under some qualification conditions of strict positivity of the multiplier.

Finally, in section 7, we focus on abstract optimization problems with an objective function which is the difference of two supremum functions. In Theorem 7.2 we established necessary and sufficient conditions for global optimality without any "extra" qualification condition like the ones used in sections 5 and 6. This framework is remarkably relevant when the distance function, a key tool in optimization and variational analysis, appears. Particularly, in section 8, we showed the importance of the distance function in a penalty approach to optimization problems.

Acknowledgment. The authors are grateful to the anonymous referees for their careful reading, comments, and suggestions, which have significantly improved the quality of this article.

REFERENCES

[1] M. Bacak and J. M. Borwein, *On difference convexity of locally Lipschitz functions*, Optimization, 60 (2011), pp. 961–978.

[2] R. Cominetti and R. Correa, *Sur une dérivée du second ordre en analyse non différentiable*, C. R. Acad. Sci. Paris Sér. I Math., 303 (1986), pp. 861–864.

[3] R. Cominetti and R. Correa, *A generalized second-order derivative in nonsmooth optimization*, SIAM J. Control Optim., 28 (1990), pp. 789–809, https://doi.org/10.1137/0328045.

[4] R. Correa, A. Hantoute, and M. A. López, *Towards supremum-sum subdifferential calculus free of qualification conditions*, SIAM J. Optim., 26 (2016), pp. 2219–2234.

[5] R. Correa, A. Hantoute, and M. A. López, *Weaker conditions for subdifferential calculus of convex functions*, J. Funct. Anal., 271 (2016), pp. 1177–1212.

[6] R. Correa, A. Hantoute, and M. A. López, *Valadier-like formulas for the supremum function* II: *The compactly indexed case*, J. Convex Anal., 26 (2019), pp. 299–324.

[7] R. Correa, A. Hantoute, and M. A. López-Cerdá, *Moreau-Rockafellar-type formulas for the subdifferential of the supremum function*, SIAM J. Optim., 29 (2019), pp. 1106–1130.

[8] N. Dinh, M. A. Goberna, M. A. López, and T. Q. Son, *New Farkas-type constraint qualifications in convex infinite programming*, ESAIM Control Optim. Calc. Var., 13 (2007), pp. 580–597.

[9] N. Dinh, M. A. Goberna, M. A. López, and M. Volle, *Convex inequalities without constraint qualification nor closedness condition, and their applications in optimization*, Set-Valued Var. Anal., 18 (2010), pp. 423–445.

[10] N. Dinh, B. Mordukhovich, and T. T. A. Nghia, *Subdifferentials of value functions and optimality conditions for DC and bilevel infinite and semi-infinite programs*, Math. Program., 123 (2010), pp. 101–138.

[11] N. Dinh, B. S. Mordukhovich, and T. T. A. Nghia, *Qualification and optimality conditions for DC programs with infinite constraints*, Acta Math. Vietnam., 34 (2009), pp. 125–155.

[12] N. Dinh, T. T. A. Nghia, and G. Vallet, *A closedness condition and its applications to DC programs with convex constraints*, Optimization, 59 (2010), pp. 541–560.

[13] D. Fang, W. Liu, and C. Wen, *Generalized subdifferentials of the value function for parametrized DC optimization problems*, J. Nonlinear Convex Anal., 17 (2016), pp. 639–654.

[14] D. Fang and X. Zhao, *Optimality conditions for convex and DC infinite optimization problems*, J. Nonlinear Convex Anal., 17 (2016), pp. 683–700.

[15] A. Hantoute, M. A. López, and C. Zălinescu, *Subdifferential calculus rules in convex analysis: A unifying approach via pointwise supremum functions*, SIAM J. Optim., 19 (2008), pp. 863–882.

[16] A. Hantoute and A. Svensson, *A general representation of δ-normal sets to sublevels of convex functions*, Set-Valued Var. Anal., 25 (2017), pp. 651–678.

[17] J.-B. Hiriart-Urruty, *Generalized differentiability, duality and optimization for problems dealing with differences of convex functions*, in Proceedings of the Symposium on Convexity and Duality in Optimization (Groningen, 1984), Lecture Notes in Econom. and Math. Systems 256, Springer, Berlin, 1985, pp. 37–70.

[18] J.-B. Hiriart-Urruty, *Convex analysis and optimization in the past* 50 *years: Some snapshots*, in Constructive Nonsmooth Analysis and Related Topics, Springer Optim. Appl. 87, Springer, Berlin, 2014, pp. 245–253.

[19] J.-B. Hiriart-Urruty, J.-J. Strodiot, and V. H. Nguyen, *Generalized Hessian matrix and second-order optimality conditions for problems with $C^{1,1}$ data*, Appl. Math. Optim., 11 (1984), pp. 43–56.

[20] R. Horst and N. V. Thoai, *DC programming: Overview*, J. Optim. Theory Appl., 103 (1999), pp. 1–43.

[21] V. Jeyakumar, A. M. Rubinov, B. M. Glover, and Y. Ishizuka, *Inequality systems and global optimization*, J. Math. Anal. Appl., 202 (1996), pp. 900–919.

[22] M. Laghdir, *Optimality conditions in DC-constrained optimization*, Acta Math. Vietnam., 30 (2005), pp. 169–179.

[23] M. López and G. Still, *Semi-infinite programming*, European J. Oper. Res., 180 (2007), pp. 491–518.

[24] J.-E. Martínez-Legaz and A. Seeger, *A formula on the approximate subdifferential of the difference of convex functions*, Bull. Aust. Math. Soc., 45 (1992), pp. 37–41.

[25] B. S. Mordukhovich, *Variational Analysis and Generalized Differentiation.* I: *Basic Theory*, Grundlehren Math. Wiss. 330, Springer, Berlin, 2006.

[26] B. S. Mordukhovich and T. T. A. Nghia, *Subdifferentials of nonconvex supremum functions and their applications to semi-infinite and infinite programs with Lipschitzian data*, SIAM J. Optim., 23 (2013), pp. 406–431.

[27] B. S. Mordukhovich and T. T. A. Nghia, *Nonsmooth cone-constrained optimization with applications to semi-infinite programming*, Math. Oper. Res., 39 (2014), pp. 301–324.

[28] P. Pérez-Aros, *Formulae for the conjugate and the subdifferential of the supremum function*, J. Optim. Theory Appl., 180 (2019), pp. 397–427.

[29] P. Pérez-Aros, *FORUM about formulae for the conjugate and the subdifferential of the supremum function*, J. Optim. Theory Appl., 183 (2019), pp. 1177–1178.

[30] P. Pérez-Aros, *Subdifferential formulae for the supremum of an arbitrary family of functions*, SIAM J. Optim., 29 (2019), pp. 1714–1743.

[31] P. Pérez-Aros, D. Salas, and E. Vilches, *On formulae for the Ioffe geometric subdifferential of supremum function*, J. Convex Anal., 27 (2020), pp. 487–508.

[32] R. T. Rockafellar, *Convex analysis*, Princeton Math. Ser. 28, Princeton University Press, Princeton, NJ, 1970.

[33] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Grundlehren Math. Wiss. 317, Springer, Berlin, 1998.

[34] H. Tuy, *Convex Analysis and Global Optimization*, Nonconvex Optim. Appl. 22, Kluwer Academic, Dordrecht, 1998.

[35] M. Valadier, *Sous-différentiels d'une borne supérieure et d'une somme continue de fonctions convexes*, C. R. Acad. Sci. Paris Sér. A-B, 268 (1969), pp. A39–A42.

[36] W. van Ackooij and W. de Oliveira, *Nonsmooth and nonconvex optimization via approximate difference-of-convex decompositions*, J. Optim. Theory Appl., 182 (2019), pp. 49–80.

[37] L. Veselý and L. Zajíček, *Delta-convex mappings between Banach spaces and applications*, Dissertationes Math. (Rozprawy Mat.), 289 (1989).

[38] C. Zălinescu, *Convex Analysis in General Vector Spaces*, World Scientific, River Edge, NJ, 2002.