J. Nonlinear Var. Anal. 9 (2025), No. 4, pp. 539-560
Available online at http://jnva.biemdas.com
https://doi.org/10.23952/jnva.9.2025.4.05

LOG-EXPONENTIAL APPROXIMATION IN SEMI-INFINITE PROGRAMMING: A VARIATIONAL APPROACH

RAFAEL CORREA[1], MARCO LÓPEZ[2,*], PEDRO PÉREZ-AROS[1]

[1]*Departamento de Ingeniería Matemática and Centro de Modelamiento Matemático (CNRS UMI 2807), Universidad de Chile, Santiago, Chile*
[2]*Department of Mathematics, University of Alicante, Alicante, Spain*

Abstract. In this paper, we propose an approximation technique for a function which is the supremum of a general family of functions by means of a function, of LogExp-type, involving a finite number of the data functions. A study of variational properties of such an approximating function is carried out in the paper. In particular, the epigraphic convergence of these approximating functions to the supremum function is proven. Moreover, refined calculus specifying the relations among the subdifferentials (regular and general) of the approximating and the data functions are provided in the first part of the paper. In the second part, we propose to approximate a general semi-infinite programming problem by a simple problem with a single constraint. Applying the results of the first part we establish, under standard hypotheses, the convergence of optimal values and solutions of these problems to the corresponding ones of the original problem. The Lagrangian duality of this problem is also studied but restricted to the semi-infinite convex programming problem, and optimal dual solutions are built by means of a sequential procedure.

Keywords. Approximation technique; Duality; Epigraphical convergence; Supremum function; Semi-infinite programming.

1. INTRODUCTION

Semi-infinite programming is a branch of optimization where the objective is to minimize (or maximize) a function subject to an infinite number of constraints. The term semi-infinite comes from the fact that there is a finite number of variables but the number of inequeality constraints is infinite. Semi-infinite programming arises in various applications, including engineering design, economics, robust optimization and control theory, where it is necessary to ensure that certain conditions hold over a continuum of scenarios or parameter values. The complexity of semi-infinite programming lies in handling this infinite constraint set efficiently while finding optimal solutions. A standard formulation of the semi-infinite programming problems is given by:

$$\min \psi(x) \text{ subject to } f_t(x) \leq 0, \text{ for all } t \in T,$$

*Corresponding author.
E-mail address: rcorrea@dim.uchile.cl (R. Correa), marco.antonio@ua.es (M. López), pperez@dim.uchile.cl (P. Pérez-Aros).
Received 27 January 2025; Accepted 17 March 2025; Published online 15 May 2025.

where $\psi, f_t : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, with $t \in T$, are general functions, frequently assumed to be lower-semicontinuous, ψ is the objective function, and $f_t(x) \leq 0$, $t \in T$, represents an infinite system of constraints, indexed by a parameter t in T.

To the best of our knowledge, the roots of semi-infinite programming can be traced back to Chebyshev approximation, with significant contributions from Haar on linear semi-infinite systems [23] and Fritz John's optimality conditions. However, it was not until the influential works of Charnes, Cooper, and Kortanek that the term semi-infinite programming was formally coined in their research papers [14, 15]. Nowdays, semi-infinite programming has been an important part of mathematical optimization with several applications in different topics of applied mathematics. We refer to [25, 29] and the references therein for more details about the subject.

A proficient approach to understand the mathematical properties of semi-infinite programming problems from the perspective of nonsmooth analysis involves considering the pointwise supremum function of the family f_t for $t \in T$; that is, the function $f := \sup_{t \in T} f_t$. In this way, we succeed to replace infinitely many constraints by a single one $f(x) \leq 0$. Then, by examining properties of this function, such as its subdifferential, one moves into the classical framework of nonsmooth constrained programming. For that reason, several authors have studied the supremum function, demonstrating its central role in mathematical optimization. Some classical contributions to this topic are attributed to Danskin [21], Hiriart-Urruty & Phelps [26], Ioffe & Tikhomirov [27], Levin [28], Pshenichnyi [42], Rockafellar [43], Valadier [49], Volle [50], and more contemporary research can be found in [19, 24, 32, 34, 35, 37, 38] and the references therein. A detailed account of results concerning the subdifferential of the supremum function of an arbitrary family of convex functions can be found in [18]. However, while this approach is successful in understanding the theoretical properties of semi-infinite programming, it has the disadvantage that the supremum function is well-known for being nonsmooth. As a consequence of the considerations above, the reformulation of the initial problem using the supremum function loses its theoretical potential for solving semi-infinite programming problems from a numerical point of view, and accounts for the interest of alternative techniques as the smoothing and approximation methods. Smoothing and approximation techniques have been proposed by many authors, mainly in the setting of convex-minimax optimization; e.g., by Auslender *et al.* [5], Bertsekas [7], Ben-Tal and Teboulle [6], Nesterov [33], Polak *et al.* [41], Sheu & Wu [46], and by Sheu & Lin [45] for continuous min-max problems, related to the global approach of Fang & Wu [22] who used an integral analog.

In this paper, we investigate an approximation φ_k of the supremum function f, which is a log-exponential function involving a finite number of the original functions f_{t_i}, $i = 1, \ldots, k$. Our primary focus is on understanding the approximation properties of this class of functions. Given that the supremum function is not necessarily smooth, traditional analysis via classical uniform or pointwise convergence of values is insufficient for mathematical programming purposes. This limitation motivates us to explore the approximation within the framework of variational analysis, allowing us to assess not only the convergence of values in a general setting but also the convergence of (sub-)gradients. This theoretical framework provides a basis for applying our results and the tools of variation analysis to the approximation of general semi-infinite programming problems using the sequence of functions φ_k. Our approach is aligned with the methodology of epigrafical convergence and consistent approximations (see, e.g., [5, 12, 39, 41, 44] and the references therein).

The contents summary of the paper is the following. After this introduction, Section 2 fixes the notation and provides some preliminaries on generalized differentiation and convergence of sets and mappings. Section 3 presents an exhaustive analysis of the properties of such an approximating function from the perspective of variational analysis. The fundamental result is Theorem 3.1 in which the epigraphic convergence of these approximating functions to the supremum function is proven, that is, $\varphi_k \xrightarrow{e} f$, which is key for the methodology of consistent approximation. A bit more is given in Corollary 3.1 which establishes continuous convergence but under the requirement of uniform continuity of the original functions with respect to the parameter t. Proposition 3.2 provides refined calculus rules that specify the relations between the subdifferential (regular and general) of φ_k and that of the functions f_{t_i}, $i = 1,\ldots,k$. Finally, in this section, Theorem 3.2, under the additional hypothesis of the continuity of the function $(t,x) \mapsto (f_t(x), \nabla f_t(x))$ on $T \times \mathbb{B}_\varepsilon(\bar{x})$, where $\mathbb{B}_\varepsilon(\bar{x})$ is the closed ball centered at $\bar{x}$ and of radius ε, proves the Painlevé-Kuratowski convergence of the sequence of gradients $\nabla f_{t_1}(\bar{x})$, $\nabla f_{t_2}(\bar{x}),\ldots,\nabla f_{t_k}(\bar{x}),\ldots$ to the subdifferential $\partial f(\bar{x})$.

In section 4, we study a general semi-infinite programming problem $(P): \quad \min \psi(x)$ subject to $f_t(x) \le 0$, $t \in T$, and we propose to replace it by the simple smooth problem with a single constraint $(P_k): \quad \min \psi(x)$ subject to $\varphi_k(x) \le 0$. Applying the variational results of the previous section, Theorem 4.1 establishes that, under standard hypotheses, the optimal values of the problems (P_k) converge to the optimal value of (P), and that the accumulation points of any sequence of optimal solutions of the problems (P_k) are optimal solutions of problem (P). The Lagrangian duality of this problem is studied in the second part of this section, but restricted to the semi-infinite convex programming problem. Theorem 4.3 establishes a procedure that conceptually converges to an optimal solution of the dual problem (D). Finally, the paper ends with some concluding remarks.

2. Notation and Preliminaries

Throughout this work, we represent by (T,d) a metric space. We adopt the standard notation of convex and variational analysis as detailed in the monographs [30, 31, 47]. We denote by $\mathbb{R}^n$ the n-dimensional Euclidean space with the Euclidean norm $\|\cdot\|$ and the inner product $\langle \cdot,\cdot \rangle$. The closed ball centered at $\bar{x}$ with radius ε is denoted by $\mathbb{B}_\varepsilon(\bar{x})$. Additionally, we follow the conventions $\exp(+\infty) = +\infty$ and $\ln(+\infty) = +\infty$.

2.1. Generalized differentiation.

Given a function $\psi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ we denote its domain as $\operatorname{dom}\psi := \{x \in \mathbb{R}^n : \psi(x) \in \mathbb{R}\}$. For a point $\bar{x} \in \operatorname{dom}\psi$, we say that the vector x^* is a *(regular) subgradient* of ψ at $\bar{x}$, and write $x^* \in \hat{\partial}\psi(\bar{x})$, if

$$\psi(x) \ge \psi(\bar{x}) + \langle x^*, x - \bar{x} \rangle + o(\|x - \bar{x}\|), \text{ for all } x \in \mathbb{R}^n;$$

equivalently, if

$$\liminf_{\substack{x \to \bar{x} \\ x \ne \bar{x}}} \left(\frac{\psi(x) - \psi(\bar{x}) - \langle x^*, x - \bar{x} \rangle}{\|x - \bar{x}\|} \right) \ge 0.$$

Moreover, x^* is a *(general) subgradient* of ψ at $\bar{x}$, written $x^* \in \partial\psi(\bar{x})$, if there exist sequences $x_k \to \bar{x}$ and $x_k^* \to x^*$ with $\psi(x_k) \to \psi(\bar{x})$ and $x_k^* \in \hat{\partial}\psi(x_k)$. We set $\hat{\partial}\psi(\bar{x}) := \emptyset$ and $\partial\psi(\bar{x}) := \emptyset$ for any $\bar{x} \notin \operatorname{dom}\psi$. The sets $\hat{\partial}\psi(\bar{x})$ and $\partial\psi(\bar{x})$ are closed, $\hat{\partial}\psi(\bar{x})$ is convex and $\hat{\partial}\psi(\bar{x}) \subset \partial\psi(\bar{x})$.

Finally, x^* is called a *horizon subgradient* of ψ at $\bar{x}$, written $x^* \in \partial^\infty\psi(\bar{x})$ if in the last definition we require $\lambda_k x_k^* \to x^*$ for some sequence $\lambda_k \searrow 0$.

Furthermore, following [30, Definition 1.91], we say that ψ is *lower regular* at $\bar{x}$ provided that $\hat{\partial}\psi(\bar{x}) = \partial\psi(\bar{x})$. It is well-known that any of the above subdifferentials coincide with the classical subdifferential of convex functions. In other words, if ψ is convex, then we have

$$x^* \in \hat{\partial}\psi(\bar{x}) = \partial\psi(\bar{x}) \Longleftrightarrow \psi(x) \geq \psi(\bar{x}) + \langle x^*, x - \bar{x}\rangle\,, \text{ for all } x \in \mathbb{R}^n.$$

Here, we recall the following *approximate* (or *decoupled*) *chain rule* for composition of smooth functions and extended real valued functions. This result can be found in [11, Theorem 3.5.7].

Theorem 2.1. *Let $\psi : \mathbb{R}^k \to \mathbb{R}$ be a C^1-function strictly increasing for each of its variables, and $f_i : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ with $i = 1, \ldots, k$. Consider the composite function $\Phi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, given by*

$$\Phi(x) := \begin{cases} \psi(f_1(x), f_2(x), \ldots, f_k(x)) & \text{if } x \in \bigcap_{i=1}^k \operatorname{dom} f_i, \\ +\infty & \text{if } x \notin \bigcap_{i=1}^k \operatorname{dom} f_i. \end{cases}$$

Then, for every $x^* \in \hat{\partial}\Phi(\bar{x})$ and every $\varepsilon > 0$, there exists $x_i \in \mathbb{B}_\varepsilon(\bar{x})$ with $|f_i(x_i) - f_i(\bar{x})| \leq \varepsilon$ for $i = 1, \ldots, k$ and

$$x^* \in \sum_{i=1}^k \hat{\partial}(\mu_i f_i)(x_i) + \mathbb{B}_\varepsilon(0),$$

where $(\mu_1, \mu_2, ..., \mu_k) = \nabla\psi(f_1(\bar{x}), f_2(\bar{x}), ..., f_k(\bar{x}))$.

2.2. **Convergence of sets and set-valued mappings.** Given a set $S \subset \mathbb{R}^n$ and a point $x \in \mathbb{R}^n$, we denote by $d(x,S) := \inf\{\|x - y\| : y \in S\}$ the (Euclidean) distance from x to S.

For a sequence of sets $\{S_k\}_{k\in\mathbb{N}}$ in $\mathbb{R}^n$, we define the *upper limit* and the *lower limit* of the sequence S_k by

$$\operatorname{Ls} S_k := \{x \in \mathbb{R}^n : \liminf d(x, S_k) = 0\},$$
$$\operatorname{Li} S_k := \{x \in \mathbb{R}^n : \limsup d(x, S_k) = 0\},$$

respectively. A subset S is the *limit* of the sequence S_k in the sense of *Painlevé-Kuratowski* (or *P.K.*), denoted $S_k \overset{P.K.}{\to} S$, provided that

$$\operatorname{Ls} S_k = S = \operatorname{Li} S_k.$$

In such a case, we simply denote $S = \operatorname{Lim} S_k$. Let us recall the following compactness result about the *Painlevé-Kuratowski* convergence (see, e.g., [3, Theorem 1.17] or [44, Theorem 4.18]).

Theorem 2.2 (Zarankiewicz)**.** *Every sequence of closed subsets $\{S_k\}_{k\in\mathbb{N}}$ on $\mathbb{R}^n$ contains a subsequence which has a (possibly empty) limit in the Painlevé-Kuratowski sense.*

Remark 2.1. When the limit of $\{S_k\}_{k\in\mathbb{N}}$ is the empty set it is said that the sequence *escapes to the horizon* (e.g., the Corollary 4.11 in [44, Theorem 4.18]).

Given a family of set-valued mappings $F_k : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, we define the set valued-mappings $\operatorname{Ls} F_k, \operatorname{Li} F_k : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ given by

$$\operatorname{gph}(\operatorname{Ls} F_k) := \operatorname{Ls}(\operatorname{gph} F_k), \text{ and } \operatorname{gph}(\operatorname{Li} F_k) := \operatorname{Li}(\operatorname{gph} F_k). \tag{2.1}$$

Moreover, we say that F_k *converges graphically* to F, denoted $F_k \xrightarrow{g} F$, provided that $\operatorname{gph} F_k$ converges to $\operatorname{gph} F$. The last is equivalent to the fact that

$$\operatorname{Ls} F_k(x) \subseteq F(x) \subseteq \operatorname{Li} F_k(x), \text{ for all } x \in \mathbb{R}^n.$$

Now, we recall the notion of epigraphical convergence of functions.

Definition 2.1. Let $f_k : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, $k = 1, 2, \ldots$, be a sequence of functions. This sequence f_k is said to be *epiconvergent* to f, denoted $f_k \xrightarrow{e} f$, if, for every $x \in \mathbb{R}^n$, the following statements hold:

a) $\liminf\limits_{k\to\infty} f_k(x_k) \geq f(x)$ for all sequence $x_k \to x$.
b) $\limsup\limits_{k\to\infty} f_k(x_k) \leq f(x)$ for some sequence $x_k \to x$.

Moreover a sequence of functions f_k *converge continuously* to a function f provided that $f_k(x_k) \to f(x)$ for every sequence $x_k \to x$. We refer to [2, 44] for more details about the theory of epigraphical convergence.

2.3. **The supremum function.** Given a family of functions $f_t : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ with $t \in T$, we consider the *supremum function* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ given by

$$f(x) := \sup_{t\in T} f_t(x). \tag{2.2}$$

Let us establish the following lemma, which provides a sufficient condition to ensure that the supremum function in (2.2) can be determined by considering only a dense subset of the index set T.

Lemma 2.1. *Let $D \subseteq T$ be a dense subset of T. Consider $\bar{x} \in \mathbb{R}^n$ such that the function $T \ni t \mapsto f_t(\bar{x})$ is lsc. Then*

$$f(\bar{x}) = \sup_{t\in D} f_t(\bar{x}). \tag{2.3}$$

Proof. Since the inequality $\geq$ holds by definition of f, let us focus on the opposite inequality. Consider an arbitrary point $\bar{t} \in T$. By density of D, one sees that there exists a sequence $t_j \to \bar{t}$. Using the lower semicontinuity of $t \mapsto f_t(\bar{x})$, we see that

$$\sup_{s\in D} f_s(\bar{x}) \geq \liminf_{j\to\infty} f_{t_j}(\bar{x}) \geq f_{\bar{t}}(\bar{x}).$$

Since, $\bar{t} \in T$ was chosen arbitrary, we have that the inequality $\leq$ holds in (2.3). □

The next result provides a formula for the subdifferential of the supremum function (see, e.g., [44, Theorem 10.31]).

Proposition 2.1. *Let us suppose that T is compact and that the functions $f_t : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, $t \in T$, are differentiable on the ball $\mathbb{B}_\varepsilon(\bar{x})$, for some $\varepsilon > 0$, and the function $(t,x) \mapsto (f_t(x), \nabla f_t(x))$ is continuous on $T \times \mathbb{B}_\varepsilon(\bar{x})$. Then $\partial f(\bar{x}) = \operatorname{co}\{\nabla f_t(\bar{x}) : t \in T(\bar{x})\}$, where $\operatorname{co} A$ denotes de convex hull of the set A, and $T(\bar{x}) := \{t \in T : f_t(\bar{x}) = f(\bar{x})\}$.*

3. Log-exponential approximation of supremum function

In this section, we define a sequence of log-exponential approximating functions of the supremum function f given in (2.2). We investigate the variational approximation properties of this sequence of functions, focusing on the convergence of values and subgradients.

Consider a countably infinite set $\{t_i\}_{i\in\mathbb{N}} \subset T$ and $\rho_k > 0$ with $k \in \mathbb{N}$. Associated with each $k \in \mathbb{N}$, we introduce the log-exponential approximating function $\varphi_k : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, given by

$$\varphi_k(x) := \frac{1}{\rho_k} \ln\left(\frac{1}{k}\sum_{i=1}^{k} \exp\left[\rho_k f_{t_i}(x)\right]\right). \tag{3.1}$$

Remark 3.1. We emphasize that the construction in (3.1) and our analysis will make sense in the context of supremum functions considered in semi-infinite programming, that is, when set T contains infinitely many points. When T is a finite set, say $T = \{1,\dots,p\}$, the function $\varphi_k(x)$, and $k \geq p$, reduces to the classical log-exponential approximation of the maximum function $\max_{i=1,\dots,p} f_i$, given by

$$\varphi_k(x) = \frac{1}{\rho_k} \ln\left(\frac{1}{p}\sum_{i=1}^{p} \exp\left[\rho_k f_i(x)\right]\right), \quad \text{with } \rho_k \to +\infty.$$

This framework has been extensively studied in the literature (see, e.g., [8, 39, 41, 44] and references therein).

In contrast, the main challenge in our approach arises from the fact that both the number of points $\{t_1,\dots,t_k\}$ and the regularization parameter ρ_k vary simultaneously as $k \to \infty$.

Remark 3.2. An alternative to the smoothing and approximation technique proposed here consists of using integral methods. They have been applied by many researchers, such as Auslender [4], Fang-Wu [22], Polak-Higgins-Mayne [40], etc., and they are based on *integral smoothing functions* of the following type

$$\phi_k(x) := \frac{1}{\rho_k} \log\left(\int_T \exp\left[\rho_k f(t,x)\right] dt\right).$$

A comparative analysis of these two methodologies was performed in [5].

The following result provides a series of simple properties of the function φ_k, which can be proved by using standard arguments. We only give a detailed prove of statment *(ii)* for completness.

Proposition 3.1 (Basic properties of φ_k)**.** *Let us consider the family of functions $f_t : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, $t \in T$, and the sequence $\varphi_k : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ defined in (3.1). Then, the following properties hold:*

(i) φ_k is lower semicontinous provided that each f_{t_i}, $i = 1,\dots,k$, is lower semicontinuous.
(ii) φ_k is convex provided that each f_{t_i}, $i = 1,\dots,k$, is convex.
(iii) For any $h : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and all $k \in \mathbb{N}$, we have that

$$\varphi_k(x) + h(x) = \frac{1}{\rho_k} \ln\left(\frac{1}{k}\sum_{i=1}^{k} \exp\left[\rho_k\Big(f_{t_i}(x) + h(x)\Big)\right]\right), \textit{ for all } x \in \mathbb{R}^n. \tag{3.2}$$

Consequently, $\varphi + h$ is convex provided that each $f_{t_i} + h$, $i = 1,\dots,k$, is convex.

Proof. *(i)* The lower semicontinuity of φ_k follows from the continuity and nondecreasing character of ln and exp together with the lower semicontinuity of f_{t_i}.

(ii) Let us prove that the convexity of f_{t_i}, for all $i = 1,\dots,k$, implies that φ_k is convex. Indeed, for $x, y \in \mathbb{R}^n$ and $\lambda \in]0,1[$, we have

$$\begin{aligned}\varphi_k(\lambda x+(1-\lambda)y) &= \frac{1}{\rho_k}\ln\left(\frac{1}{k}\sum_{i=1}^{k}\exp\left[\rho_k\left(f_{t_i}(\lambda x+(1-\lambda)y)\right]\right)\right.\\ &\le \frac{1}{\rho_k}\ln\left(\frac{1}{k}\sum_{i=1}^{k}\exp\left[\rho_k\left(\lambda f_{t_i}(x)+(1-\lambda)f_{t_i}(y)\right]\right)\right.\\ &= \frac{1}{\rho_k}\ln\left(\frac{1}{k}\sum_{i=1}^{k}\exp\left[\rho_k f_{t_i}(x)\right]^{\lambda}\exp\left[\rho_k f_{t_i}(y)\right]^{1-\lambda}\right).\end{aligned}$$

Hence, by the Hölder inequality given by

$$\left(\sum_{i=1}^{k}|x_i|^{\alpha}|y_i|^{\beta}\right) \le \left(\sum_{i=1}^{k}|x_i|\right)^{\alpha}\left(\sum_{i=1}^{k}|y_i|\right)^{\beta} \text{ with } \alpha+\beta=1,$$

we have that

$$\sum_{i=1}^{k}\exp\left[\rho_k f_{t_i}(x)\right]^{\lambda}\exp\left[\rho_k f_{t_i}(y)\right]^{1-\lambda} \le \left(\sum_{i=1}^{k}\exp\left[\rho_k f_{t_i}(x)\right]\right)^{\lambda}\left(\sum_{i=1}^{k}\exp\left[\rho_k f_{t_i}(y)\right]\right)^{1-\lambda}.$$

Therefore,

$$\begin{aligned}\varphi_k(\lambda x+(1-\lambda)y) &\le \frac{1}{\rho_k}\ln\left(\left(\frac{1}{k}\sum_{i=1}^{k}\exp\left[\rho_k f_{t_i}(x)\right]\right)^{\lambda}\left(\frac{1}{k}\sum_{i=1}^{k}\exp\left[\rho_k f_{t_i}(y)\right]\right)^{1-\lambda}\right)\\ &\le \lambda\varphi_k(x)+(1-\lambda)\varphi_k(y),\end{aligned}$$

which shows the convexity of φ_k.

(iii) It is obvious that (3.2) holds, and the convexity of $x \mapsto \varphi_k(x)+h(x)$ comes from the previous part, provided that $x \mapsto f_{t_i}(x)+h(x)$ is convex, for $i = 1,\dots,k$.

□

The next crucial result formally establishes the epigraphical convergence of the sequence φ_k to the supremum function f. It is important to notice that in this result the index set T is an arbitrary metric space, and the functions f_t, with $t \in T$, could be extended real valued functions.

Theorem 3.1. *Consider a family of functions $f_t : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ with $t \in T$ and the sequence $\varphi_k : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ defined in (3.1), with $\rho_k > 0$ such that $\rho_k^{-1}\ln(k) \to 0$. In addition, suppose that $\{t_i\}_{i\in\mathbb{N}}$ is dense, and $(t,x) \mapsto f_t(x)$ is lower semicontinuous. Then,*

(i) For all $x \in \mathbb{R}^n$, $\varphi_k(x) \to f(x)$.

(ii) For any $x \in \mathbb{R}^n$, and all $x_k \to x$, $f(x) \le \liminf_{k\to\infty}\varphi_k(x_k)$.

Particularly, $\varphi_k \xrightarrow{e} f$.

Proof. Let us consider the auxiliary function

$$m_k(x) := \max\{f_{t_i}(x) : i = 1,\dots,k\}. \tag{3.3}$$

Then, it is easy to check that (see, e.g., [39, Eq. 4.9b])

$$m_k(x) - \frac{\ln(k)}{\rho_k} \leq \varphi_k(x) \leq m_k(x), \text{ for all } x \in \mathbb{R}^n. \tag{3.4}$$

It follows from (2.3) that $m_k(x) \nearrow \sup_{k\in\mathbb{N}} m_k(x) = f(x)$ as $k \nearrow \infty$ for all $x \in \mathbb{R}^n$. Using [44, Proposition 7.4 (d)] we have that $m_k \xrightarrow{e} \sup_{k\in\mathbb{N}} m_k = f$. Finally, using (3.4), we see that φ_k satisfies properties (i). Moreover, if we consider any sequence $x_k \to x$, we obtain by (3.4) that

$$\liminf_{k\to\infty} \varphi_k(x_k) \geq \liminf_{k\to\infty} \left(m_k(x_k) - \frac{\ln(k)}{\rho_k} \right) = \liminf_{k\to\infty} m_k(x_k) \geq f(x),$$

which shows that (ii) holds, and that completes the proof. □

Remark 3.3. It it worth mentioning that Theorem 3.1 implies that, for any increasing function $j : \mathbb{N} \to \mathbb{N}$, we have that $\varphi_{j(k)} \xrightarrow{e} f$. This can be follow from the fact that the epigraphical convergence is preserved by considering subsequences, due to the fact that the space of lower semicontinuous functions with the topology of the epigraphic convergence is metrizable (see, e.g., [44, Theorem 7.58]).

Now, we establish the continuous convergence of the sequence φ_k to the supremum function f (see, e.g., [44] for more details and consequences of this notation).

Corollary 3.1. *In addition to the assumptions of Theorem 3.1 suppose that the family f_t, $t \in T$, is continuous at $\bar{x}$ uniformly with respect to $t \in T$, that is, $\lim_{x\to\bar{x}} \sup_{t\in T} |f_t(x) - f_t(\bar{x})| = 0$. Then, for every sequence $x_k \to \bar{x}$ one has that $\lim_{k\to\infty} \varphi_k(x_k) = f(\bar{x})$. Consequently, if the family f_t, $t \in T$, is continuous at any $x \in \mathbb{R}^n$ uniformly with respect to $t \in T$, we have that φ_k converges continuously to f.*

Proof. Now, choose $\hat{t}_k \in \{t_1, \ldots, t_k\}$ such that $m_k(x_k) = f_{\hat{t}_k}(x_k)$, where m_k was defined in (3.3). Then, using (3.4) we get that

$$\varphi_k(x_k) \leq m_k(x_k) = f_{\hat{t}_k}(x_k) \leq f_{\hat{t}_k}(\bar{x}) + \left| f_{\hat{t}_k}(x_k) - f_{\hat{t}_k}(\bar{x}) \right| \leq f(\bar{x}) + \sup_{t\in T} |f_t(x_k) - f_t(\bar{x})|. \tag{3.5}$$

Therefore our continuity assumption and (3.5) imply that $\limsup_{k\to\infty} \varphi_k(x_k) \leq f(\bar{x})$, which together with Theorem 3.1 item (ii) implies the result. □

In the following corollary, we quantify the uniform difference between φ_k and the supremum f under certain Hölder-type assumption of the family $\{f_t : t \in T\}$ with respect to $t \in T$.

Corollary 3.2. *Under the assumptions of Theorem 3.1, suppose that there exist a set $K \subseteq \mathbb{R}^n$, $r > 0$ and $\alpha \in (0,1]$ such that*

$$|f(t_1,x) - f(t_2,x)| \leq rd(t_1,t_2)^\alpha, \textit{ for all } t_1,t_2 \in T \textit{ and all } x \in K. \tag{3.6}$$

Then,

$$\sup_{x\in K} |\varphi_k(x) - f(x)| \leq r \sup_{t\in T} \min_{i=1,\ldots,k} d(t,t_i)^\alpha + \rho_k^{-1} \ln(k). \tag{3.7}$$

Proof. Let us consider $k \in \mathbb{N}$, $t \in T$ and $x \in K$. Using (3.4) and for all $i = 1, \ldots, k$ one has

$$\begin{aligned} f_t(x) &\leq f_{t_i}(x) + |f_t(x) - f_{t_i}(x)| \leq m_k(x) + rd(t_i,t)^\alpha \\ &\leq \varphi_k(x) + rd(t_i,t)^\alpha + \rho_k^{-1} \ln(k). \end{aligned}$$

Therefore, the above inequality implies that

$$f_t(x) - \varphi_k(x) \le r \min_{i=1,\dots,k} d(t,t_i)^\alpha + \rho_k^{-1} \ln(k).$$

Taking supremum on $t \in T$ in the above inequality, we see that

$$0 \le f(x) - \varphi_k(x) \le r \sup_{t\in T} \min_{i=1,\dots,k} d(t,t_i)^\alpha + \rho_k^{-1} \ln(k),$$

and (3.7) holds, which ends the proof. □

3.1. **Subgradients of the approximating function.** We begin this section by presenting the results that enable us to understand the first-order variations of the function φ_k in terms of the data function f_{t_i}, with $i = 1,\dots,k$.

Proposition 3.2 (Refined calculus rules for φ_k). *Consider the family of functions $f_t : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$, $t \in T$, and the sequence $\varphi_k : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ defined in (3.1). Then*

(i) For all $\bar{x} \in \mathbb{R}^n$, the following inclusion holds

$$\sum_{i=1}^{k} \alpha_i(\bar{x}) \hat{\partial} f_{t_i}(\bar{x}) \subseteq \hat{\partial} \varphi_k(\bar{x}), \tag{3.8}$$

where

$$\alpha_i(\bar{x}) := \frac{\exp\left[\rho_k f_{t_i}(\bar{x})\right]}{\sum_{j=1}^{k} \exp\left[\rho_k f_{t_j}(\bar{x})\right]}, \tag{3.9}$$

and with the convention $\alpha\emptyset = \emptyset$ when $\alpha > 0$.

(ii) For any $\bar{x} \in \mathbb{R}^n$, $v \in \hat{\partial}\varphi_k(\bar{x})$ and all $\varepsilon > 0$, there exist $x_i \in \mathbb{B}_\varepsilon(\bar{x})$ with $|f_{t_i}(x_i) - f_{t_i}(\bar{x})| \le \varepsilon$, for $i = 1,\dots,k$, such that $v \in \sum_{i=1}^{k} \alpha_i(\bar{x}) \hat{\partial} f_{t_i}(x_i) + \mathbb{B}_\varepsilon(0)$.

(iii) It holds that

$$\partial \varphi_k(\bar{x}) \subseteq \sum_{i=1}^{k} \alpha_i(\bar{x}) \hat{\partial} f_{t_i}(\bar{x}) \tag{3.10}$$

provided that the following qualification condition holds

$$v_i \in \partial f_{t_i}^\infty(\bar{x}), \text{ and } \sum_{i=1}^{k} v_i = 0 \Rightarrow v_i = 0. \tag{3.11}$$

In addition, if the functions f_{t_i} with $i = 1,\dots,k$ are lower-regular at $\bar{x}$, then φ_k is also lower regular at $\bar{x}$ and (3.10) holds as an equality.

(iv) φ_k is differentiable at $\bar{x} \in \mathbb{R}^n$ provided that each f_{t_i}, $i = 1,\dots,k$, is differentiable at $\bar{x} \in \mathbb{R}^n$. In this case

$$\nabla\varphi_k(\bar{x}) = \sum_{i=1}^{k} \alpha_i(\bar{x}) \nabla f_{t_i}(\bar{x}), \tag{3.12}$$

(v) If $\{f_t\}_{t\in T}$ are C^p around $\bar{x}$, then φ_k is also $\mathscr{C}^p$ around $\bar{x}$.

Proof. Notice that $(\alpha_i(\bar{x}))_{i=1}^k$ corresponds to the gradient of the function $\psi : \mathbb{R}^k \to \mathbb{R}$ given by

$$\psi(y) = \frac{1}{\rho_k} \ln\left(\frac{1}{k}\sum_{i=1}^{k} \exp(\rho_k y_i)\right), \tag{3.13}$$

at the point $\bar{y} := (f_{t_i}(\bar{x}))_{i=1}^k$. The function ψ is convex due to the statement ii) in Proposition 3.1. Therefore,

$$\sum_{i=1}^k \alpha_i(\bar{x})(y_i - \bar{y}_i) = \langle (\alpha_i(\bar{x}))_{i=1}^k, y - \bar{y}\rangle \leq \psi(y) - \psi(\bar{y}), \text{ for all } y \in \mathbb{R}^k. \tag{3.14}$$

Now, let us consider $w^* \in \sum_{i=1}^k \alpha_i(\bar{x})\hat{\partial} f_{t_i}(\bar{x})$. Then, there exist $w_i^* \in \hat{\partial} f_{t_i}(\bar{x})$, $i = 1,...,k$, such that $w^* = \sum_{i=1}^k \alpha_i(\bar{x}) w_i^*$ and

$$\begin{aligned} \langle w^*, x - \bar{x}\rangle &= \sum_{i=1}^k \alpha_i(\bar{x})\langle w_i^*, x - \bar{x}\rangle \\ &\leq \sum_{i=1}^k \alpha_i(\bar{x})\left(f_{t_i}(x) - f_{t_i}(\bar{x})\right) + o(\|x - \bar{x}\|). \end{aligned}$$

The, using (3.14) with $\bar{y} = (f_{t_i}(\bar{x}))_{i=1}^k$, $y = (f_{t_i}(x))_{i=1}^k$ and $x \in \operatorname{dom}\varphi_k$, we obtain

$$\langle w^*, x - \bar{x}\rangle \leq \varphi_k(x) - \varphi_k(\bar{x}) + o(\|x - \bar{x}\|.$$

That concludes the proof of formula (3.8).

Now, in order to prove item ii), let us consider $x^* \in \hat{\partial}\varphi_k(\bar{x})$. Then, applying Theorem 2.1 to function φ_k, we see that there exist $x_i \in \mathbb{B}_\varepsilon(\bar{x})$ with $|f_{t_i}(x_i) - f_{t_i}(\bar{x})| \leq \varepsilon$, for $i = 1,\dots,k$, such that

$$x^* \in \sum_{i=1}^k \hat{\partial}\left(\alpha_i(\bar{x}) f_{t_i}\right)(x_i) + \mathbb{B}_\varepsilon(0).$$

Now, notice that $\alpha_i(\bar{x}) > 0$. Then

$$\hat{\partial}\left(\alpha_i(\bar{x}) f_{t_i}\right)(x_i) = \alpha_i(\bar{x})\hat{\partial} f_{t_i}(x_i),$$

which concludes the proof of this item.

To prove statement (iii), let us consider $x^* \in \partial\varphi_k(\bar{x})$. In view of the definition, one sees that there exist sequences $x_j \to \bar{x}$ and $x_j^* \in \hat{\partial}\varphi_k(x_j)$ such that $\varphi_k(x_j) \to \varphi_k(\bar{x})$ and $x_j^* \to x^*$. Let us show that

$$\lim_{j\to\infty} f_{t_i}(x_j) = f_{t_i}(\bar{x}) \text{ for each } i = 1,\dots,k. \tag{3.15}$$

Suppose by contradiction that (3.15) does not hold. Then, we can take $i_0 \in \{1,\dots,k\}$ and a subsequence $x_{\nu(j)}$ for an increasing function $\nu : \mathbb{N} \to \mathbb{N}$ such that

$$\lim_{j\to\infty} f_{t_{i_0}}(x_{\nu(j)}) > f_{t_{i_0}}(\bar{x}). \tag{3.16}$$

On the one hand, by the lower semicontinuity of f_{t_i}, we have that

$$f_{t_i}(\bar{x}) \leq \liminf_{j\to\infty} f_{t_i}(x_{\nu(j)}) \text{ for all } i = 1,\dots,k. \tag{3.17}$$

On the other hand, since the function ψ, defined in (3.13), is strictly increasing with respect to each variable, we have that, for all $j \in \mathbb{N}$,

$$\psi((\inf_{p\geq j} f_{t_i}(x_{\nu(p)}))_{i=1}^k) \leq \psi((f_{t_i}(x_{\nu(p)}))_{i=1}^k), \text{ for all } p \geq j.$$

This implies that

$$\psi((\inf_{p\geq j} f_{t_i}(x_{\nu(p)}))_{i=1}^k) \leq \liminf_{j\to\infty} \psi((f_{t_i}(x_{\nu(j)}))_{i=1}^k) = \liminf_{j\to\infty} \varphi_k(x_{\nu(j)}) = \varphi_k(\bar{x}). \tag{3.18}$$

Hence, from (3.17) and the fact that ψ is continuous and strictly increasing, we see that

$$\varphi_k(\bar{x}) = \psi((f_{t_i}(\bar{x}))_{i=1}^k) \leq \psi((\liminf_{j\to\infty} f_{t_i}(x_{\nu(j)}))_{i=1}^k) = \sup_{j\in\mathbb{N}} \psi((\inf_{p\geq j} f_{t_i}(x_{\nu(p)}))_{i=1}^k). \tag{3.19}$$

Therefore, using (3.18) and (3.19), we obtain

$$\psi((\liminf_{p\to\infty} f_{t_i}(x_{\nu(p)}))_{i=1}^k) = \psi((f_{t_i}(\bar{x}))_{i=1}^k).$$

Consequently, due to the fact that the function is strictly increasing in each variable, we see that

$$f_{t_i}(\bar{x}) = \liminf_{j\to\infty} f_{t_i}(x_{\nu(j)}) \text{ for all } i = 1,\dots,k,$$

which contradicts (3.16).

Now, applying item (ii) to x_j and $x_j^* \in \hat{\partial}\varphi_k(x_j)$ with $\varepsilon_j \to 0$ yields the existence of $x_{i,j} \in \mathbb{B}_{\varepsilon_j}(x_j)$ with $|f_{t_i}(x_{i,j}) - f_{t_i}(x_j)| \leq \varepsilon_j$ and

$$x_j^* \in \sum_{i=1}^k \alpha_i(x_j) w_{i,j}^* + \mathbb{B}_{\varepsilon_j}(0), \tag{3.20}$$

where $w_{i,j}^* \in \hat{\partial} f_{t_i}(x_{i,j})$. Then, using (3.15), it is easy to see that $x_{i,j} \to \bar{x}$ and $|f_{t_i}(x_{i,j}) - f_{t_i}(\bar{x})| \to 0$ for all $i = 1,\dots,k$.

Finally, if $((w_{i,j}^*))_j$, $j = 1,...,k$, are bounded (by passing to a subsequence if it is necessary), we can assume that $w_{i,j}^* \to w_i^*$. It follows from the definition of the basic subdifferential that $w_i^* \in \partial f_{t_i}(\bar{x})$, and, taking the limits in (22), $x^* \in \sum_{i=1}^k \alpha_i(\bar{x})\partial f_{t_i}(\bar{x})$. On the other hand, if there exists some $i_0 \in \{1,\dots,k\}$ such that $\lambda_{i_0,j} := \|w_{i_0,j}^*\|$ is not bounded, then, by passing to a subsequence, we can assume that $\lambda_j := \max_{i\in\{1,...,k\}}\{\lambda_{i,j}\} \to \infty$. Then, (3.20) implies that

$$\sum_{i=1}^k \frac{\alpha_i(x_j)}{\lambda_j} w_{i,j}^* \to 0.$$

Moreover, it is not difficult to see that $\frac{\alpha_i(x_j)}{\lambda_j} w_{i,j}^*$ are bounded, and by passing to a subsequence if it is necessary, we see that $\frac{\alpha_i(x_j)}{\lambda_j} w_{i,j}^* \to v_i^* \in \partial^\infty f_{t_i}(\bar{x})$, with not all $v_i = 0$ and

$$\sum_{i=1}^k v_i^* = 0.$$

Indeed, there must exist $i_1 \in \{1,2,...,k\}$ such that, passing to a subsequence if needed, $\left\|w_{i_1,j}^*\right\| = \lambda_j$, $j = 1,2,...$ Then, without loss of generality, we can assume the existence of a vector u of norm one, such that $\lim_{j\to\infty} \frac{w_{i_1,j}^*}{\lambda_j} = u$ and

$$\lim_{j\to\infty} \frac{\alpha_{i_1}(x_j)}{\lambda_j} w_{i,j}^* = \alpha_{i_1}(\bar{x})u = v_{i_1}^*,$$

and $v_{i_1}^* \neq 0$ because $\alpha_{i_1}(\bar{x}) > 0$. The above contradicts (3.11), and this concludes the proof of (3.10).

Now, in (iv), if the functions f_t, $t \in T$, are differentiable, then φ_k is also differentiable because it is a composition of differentiable functions. Consequently, the regular subdifferential simplifies to the classical gradient, and formula (3.12) follows from (3.8) as an equality in this case. Furthermore, in (v), the last argumentation applies if f_t, with $t \in T$, are of class C^p. $\square$

The following example illustrates the fact that φ_k has better calculus rules that the classical max-function.

Example 3.1. Let us consider the functions $f_1, f_2 : \mathbb{R} \to \mathbb{R}$ given by

$$f_1(x) = \begin{cases} 0 & \text{if } x \le 0, \\ -\sqrt{x} & \text{if } x > 0, \end{cases}$$

and $f_2(x) = 0$, for $x \in \mathbb{R}$. Then, following [30, Theorem 3.46], we see that

$$\{0\} = \partial m(0) \subseteq \{\lambda_1 \circ \partial f_1(0) + \lambda_2 \circ \partial f_2(0) : \lambda_1 + \lambda_2 = 1 \text{ and } \lambda_1, \lambda_2 \ge 0\} =]-\infty, 0],$$

where $m(x) := \max\{f_1(x), f_2(x)\}$ and

$$\lambda_i \circ \partial f_i(0) = \begin{cases} \lambda_1 \partial f_i(0) & \text{if } \lambda_i > 0, \\ \partial^\infty f_i(0) & \text{if } \lambda_i = 0. \end{cases}$$

Finally, it is easy to see that (3.10) provides an equality in this example for the associated function φ_2, given by (3.1).

Now, we investigate how the subgradients of φ_k approximate the subgradients of the supremum function f. To understand this from a variational perspective, we analyze the convergence in terms of Painlevé-Kuratowski convergence.

Theorem 3.2. *Under the assumptions of Theorem 3.1, suppose that there exists $\varepsilon > 0$ such that f_t is C^1 on $\mathbb{B}_\varepsilon(\bar{x})$ for all $t \in T$. Then, for any increasing function $j : \mathbb{N} \to \mathbb{N}$*

$$\partial f(\bar{x}) \subseteq \mathrm{Ls}\, \nabla \varphi_{j(k)}(\bar{x}), \tag{3.21}$$

where the set-valued mapping $\mathrm{Ls}\, \nabla \varphi_{j(k)}$ is interpreted in the sense of (2.1). In addition, if T is compact and the function $(t,x) \mapsto (f_t(x), \nabla f_t(x))$ is continuous on $T \times \mathbb{B}_\varepsilon(\bar{x})$, then

$$\mathrm{Ls}\, \nabla \varphi_k(\bar{x}) = \mathrm{Li}\, \nabla \varphi_k(\bar{x}) = \partial f(\bar{x}). \tag{3.22}$$

Proof. Let us notice that (3.21) follows from the epigraphical convergence of the sequence $\varphi_{j(k)}$ and [48, Theorem 5.3] (see also [12, Lemma 3.4]). Now, let us assume that the function $(t,x) \mapsto (f_t(x), \nabla f_t(x))$ is continuous on $T \times \mathbb{B}_\varepsilon(\bar{x})$. We divide the rest of the proof into some claims.

Claim 1: For any increasing function $j : \mathbb{N} \to \mathbb{N}$, the equality in (3.21) holds.

Since for any increasing function $j : \mathbb{N} \to \mathbb{N}$ the inclusion $\mathrm{Ls}\, \nabla \varphi_{j(k)}(\bar{x}) \subset \mathrm{Ls}\, \nabla \varphi_k(\bar{x})$ holds, it is enough to prove that

$$\mathrm{Ls}\, \nabla \varphi_k(\bar{x}) \subset \partial f(\bar{x}). \tag{3.23}$$

To prove the above inclusion, let us take x^* in the left-hand side set in (3.23) and any $\eta \in (0, \varepsilon)$. We claim that

$$x^* \in S_\eta := \mathrm{co}\left\{ \nabla f_t(x) : t \in T_\eta(\bar{x}) \text{ and } x \in \mathbb{B}_\eta(\bar{x}) \right\},$$

where $T_\eta(\bar{x}) := \{t \in T : f_t(\bar{x}) \ge f(\bar{x}) - \eta\}$. Indeed, let us consider $x_j \to \bar{x}$ and $k_j \to \infty$ such that $x^* = \lim_{j\to\infty} \nabla \varphi_{k_j}(x_j)$. We can assume, without loss of generality, that x_j belongs to the ball $\mathbb{B}_\eta(\bar{x})$. Now, taking into account (3.12), we have that

$$\nabla \varphi_{k_j}(x_j) = \sum_{i=1}^{k_j} \lambda_i^j \nabla f_{t_i}(x_j),$$

with $\lambda_i^j = \alpha_i(x_j)$, defined in (3.9), which can be also written

$$\lambda_i^j := \frac{\exp(\rho_{k_j}(f_i(x_j) - m_{k_j}(x_j)))}{\sum_{l=1}^{k_j} \exp\left(\rho_{k_j}\left[f_{t_l}(x_j) - m_{k_j}(x_j)\right]\right)},$$

where m_k was defined in (3.3). Now, let us consider

$$y_j := \sum_{\substack{i=1 \\ t_i \notin T_\eta(\bar{x})}}^{k_j} \lambda_i^j \nabla f_{t_i}(x_j), \; \nu_j := \sum_{\substack{i=1 \\ t_i \notin T_\eta(\bar{x})}}^{k_j} \lambda_i^j \text{ and } z_j := \sum_{\substack{i=1 \\ t_i \in T_\eta(\bar{x})}}^{k_j} \lambda_i^j \nabla f_{t_i}(x_j).$$

Let us show now that $\nu_j \to 0$ as $j \to \infty$. Indeed, on the one hand, due to the continuity of $(t,x) \mapsto f_t(x)$ and the compactness of T, we conclude the existence of $j_1 \in \mathbb{N}$ such that

$$\left|f_t(x_j) - f_t(\bar{x})\right| \leq \eta/4 \text{ for all } j \geq j_1 \text{ and } t \in T.$$

On the other hand, using Corollary 3.1 and (3.4), we can consider $j_0 \geq j_1$ such that

$$\left|m_{k_j}(x_j) - f(\bar{x})\right| \leq \eta/4, \text{ for all } j \geq j_0.$$

It follows that for every $t_i \notin T_\eta(\bar{x})$ and $j \geq j_0$

$$f_{t_i}(x_j) - m_{k_j}(x_j) \leq f_{t_i}(\bar{x}) + |f_{t_i}(x_j) - f_{t_i}(\bar{x})| - f(\bar{x}) + |m_{k_j}(x_j) - f(\bar{x})| \leq -\eta/2.$$

Hence, noticing that $f_{t_l}(x_{k_j}) = m_{k_j}(x_{k_j})$ for some $l \in \{1, \ldots, k_j\}$, we have

$$\sum_{l=1}^{k_j} \exp\left(\rho_{k_j}\left[f_{t_l}(x_j) - m_{k_j}(x_j)\right]\right) \geq 1,$$

which implies that, for all $j \geq j_0$ and $i = 1, \ldots, k_j$ with $t_i \notin T_\eta(\bar{x})$,

$$\lambda_i^j \leq \exp(\rho_{k_j}(f_{t_i}(x_j) - m_{k_j}(x_j))) \leq \exp(-\rho_{k_j}\eta/2).$$

Therefore, this yields

$$\nu_j \leq k_j \exp(-\rho_{k_j}\eta/2) = \exp\left(-\rho_{k_j}\left(\eta/2 + \frac{\ln(k_j)}{\rho_{k_j}}\right)\right) \to 0.$$

(Recall that $\rho_k^{-1}\ln(k) \to 0$, entailing $\rho_k \to \infty$.). Hence,

$$\|y_j\| \leq \nu_j \sup_{(t,x) \in T \times \mathbb{B}_\eta(\bar{x})} \|\nabla f_t(x)\| \to 0.$$

Now, the vector $w_j := \left(1 - \nu_j\right)^{-1} z_j$ belongs to S_η (for large enough $j \in \mathbb{N}$), which shows that x^* belongs to the closure of S_η, but S_η is closed due to the continuity and compactness assumption of the data. Finally, some standard arguments give rise to

$$\text{co}\{\nabla f_t(\bar{x}) : t \in T(\bar{x})\} = \bigcap_{\eta > 0} S_\eta,$$

so $x^* \in \text{co}\{\nabla f_t(\bar{x}) : t \in T(\bar{x})\}$, and Proposition 2.1 implies that $x^* \in \partial f(\bar{x})$, and we are done with the claim.

Claim 2: Equation (3.22) holds.

Let us consider $F_k : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ given by $F_k(x) := \nabla \varphi_k(x)$ and $F(x) := \partial f(x)$. Now, we suppose, by contradiction, that (3.22) does not hold, that is, there exists $v \in \operatorname{Ls} F_k(\bar{x})$ such that $v \notin \operatorname{Li} F_k(\bar{x})$. It follows that there exists an increasing function $j : \mathbb{N} \to \mathbb{N}$ such that

$$\lim d((\bar{x}, v), \operatorname{gph} F_{j(k)}) > 0. \tag{3.24}$$

Now, taking into account that $\operatorname{gph} \nabla \varphi_k$ is closed and using Lemma 2.2 applied to $F_{j(k)}$, we conclude that there exist a set S and an increasing function $\gamma : \mathbb{N} \to \mathbb{N}$ such that $\operatorname{gph} F_{\nu(k)}$ converges to S, where $\nu := \gamma \circ j$. Now, using Claim 1, we see that

$$\operatorname{Li} F_{\nu(k)}(\bar{x}) = S(\bar{x}) = \operatorname{Ls} F_{\nu(k)}(\bar{x}) = \partial f(\bar{x}).$$

Particularly, the above equality implies that the point $v \in \operatorname{Li} F_{\nu(k)}(\bar{x})$, which means that

$$\liminf_{k\to\infty} d((x, v), \operatorname{gph} F_{\nu(k)}) = 0,$$

but this is a contradiction with respect to (3.24), which ends the proof. □

Remark 3.4. Notice that equality (3.22) implies the notion of the gradient consistency introduced in [16]. This concept plays a pivotal role in investigating first-order information arising from abstract functions applied to nonsmooth data. Moreover, it was widely employed by various authors in diverse contexts (see, e.g., [10, 13, 36]).

Now, our intention is to establish a similar convergence result of subgradients that the one given in Theorem 3.2, but replacing the smoothness assumptions over the family of functions f_t, with $t \in T$, with some degree of convexity of this family. Following [17], let us introduce the following notation, for a given function $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ the set

$$\Delta_g := \{(x, g(x), x^*) : x^* \in \partial g(x)\}. \tag{3.25}$$

Theorem 3.3. *Consider a family of functions $f_t : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ with $t \in T$ and the sequence $\varphi_k : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ defined in (3.1), with $\rho_k > 0$ such that $\rho_k^{-1} \ln(k) \to 0$. Suppose that there exist a continuously differentiable function $h : \mathbb{R}^n \to \mathbb{R}$ such that $f_t + h$ is proper lower semicontinuous convex function for all $t \in T$. Then $\partial \varphi_k \xrightarrow{g} \partial f$ and $\Delta_{\varphi_k} \xrightarrow{P.K.} \Delta_f$.*

Proof. We notice that the functions $\hat{\varphi}_k := \varphi_k + h$, $k \in \mathbb{N}$, and $\hat{f} = f + h$ are convex by Proposition 3.1. Moreover, due to Theorem 3.1 and the continuity of h, we have that $\hat{\varphi}_k \xrightarrow{e} \hat{f}$. Then, by Attouch's Theorem (see [44, Theorem 12.35] and [17, Theorem 2.5]) we have that

$$\partial \hat{\varphi}_k \xrightarrow{g} \partial \hat{f} \text{ and } \Delta_{\hat{\varphi}_k} \xrightarrow{P.K.} \Delta_{\hat{f}}. \tag{3.26}$$

The rest of the proof is divided into two claims.

Claim 1: The functions φ_k and f are lower-regular at any point, and for all $x \in \mathbb{R}^n$

$$\partial(\varphi_k + h)(x) = \partial \varphi_k(x) + \nabla h(x)$$
$$\partial(f + h)(x) = \partial f(x) + \nabla h(x).$$

Indeed, since $\varphi_k + h$ is convex, we have that it is lower regular

$$\hat{\partial}(\varphi_k + h)(x) = \partial(\varphi_k + h)(x), \text{ for all } x \in \mathbb{R}^n. \tag{3.27}$$

Moreover, by the smoothness of h and the sum rules (see, e.g., [30, Proposition 1.107] or [31, Proposition 1.30]), we see that, for all $x \in \mathbb{R}^n$,

$$\hat{\partial}\varphi_k(x) + \nabla h(x) = \hat{\partial}(\varphi_k + h)(x) \text{ and } \partial\varphi_k(x) + \nabla h(x) = \partial(\varphi_k + h)(x). \tag{3.28}$$

Hence, (3.27) and (3.28) imply that $\partial\varphi_k(x) = \hat{\partial}_k\varphi(x)$ for all $x \in \mathbb{R}^n$. This concludes the proof of the claim for φ_k. The proof for the function f is similar, so we omit it.

Claim 2: $\partial\varphi_k \xrightarrow{g} \partial f$ and $\Delta_{\varphi_k} \xrightarrow{P.K.} \Delta_f$.

Indeed, first let $(x_k, x_k^*) \in \operatorname{gph}\partial\varphi_{j(k)}$ and $(u_k, \varphi_{j(k)}(u_k), u_k^*) \in \Delta_{\varphi_{j(k)}}$ for some increasing function $j : \mathbb{N} \to \mathbb{N}$ such that $(x_k, x_k^*) \to (x, x^*)$ and $(u_k, \varphi_{j(k)}(u_k), u_k^*) \to (u, \alpha, u^*)$. Then

$$(x_k, x_k^* + \nabla h(x_k)) \in \operatorname{gph}\partial\hat{\varphi}_{j(k)} \text{ and } (x_k, \varphi_{j(k)}(u_k) + h(x_k), x_k^* + \nabla h(x_k)) \in \Delta_{\hat{\varphi}_k}.$$

Hence, using (3.26) and Claim 1, we see that $x^* \in \partial f(x)$, $u^* \in \partial f(u)$ and $\alpha = f(x)$. This shows that

$$\operatorname{Ls}(\operatorname{gph}\partial\varphi_k) \subseteq \operatorname{gph}\partial f \text{ and } \operatorname{Ls}\Delta_{\varphi_k} \subseteq \Delta_f.$$

Now, we consider a point $(x, x^*) \in \operatorname{gph}\partial f$ and $(u, f(u), u^*) \in \Delta_f$. Then, $(x, x^* + \nabla h(x)) \in \operatorname{gph}\partial\hat{f}$ and $(u, f(u) + h(u), u^* + \nabla h(u)) \in \Delta_{\hat{f}}$. Imitating the arguments presented above, it follows from (3.26) and Claim 1 that there are sequences $x_k^* \in \partial\varphi_k(x_k)$ and $u_k^* \in \partial\varphi_k(u_k)$ such that $(x_k, x_k^*) \to (x, x^*)$ and $(u_k, \varphi_k(u_k), u_k^*) \to (u, f(u), u^*)$. This proof that

$$\operatorname{gph}\partial f \subset \operatorname{Li}(\operatorname{gph}\partial\varphi_k) \text{ and } \Delta_f \subseteq \operatorname{Li}\Delta_{\varphi_k},$$

and that concludes the proof. □

Remark 3.5. It is important to note that the class of mappings g for which there exists a function h such that $g + h$ is convex is quite large. This class particularly includes weakly convex functions, which were extensively studied from both theoretical and, more recent, numerical perspectives (see, e.g., [1, 20] and the references therein). This ensures that such functions exhibit similar properties to convex functions.

4. Application to Semi-infinite Programming

In this section, we consider the following optimization problem

$$\min \psi(x) \text{ subject to } f_t(x) \le 0, \text{ for all } t \in T, \tag{P}$$

where ψ, $f_t : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, $t \in T$, are lower semicontinuous functions. The purpose of this section is to provide asymptotic relations between optimization problem (P) and the following approximation

$$\min \psi(x) \text{ subject to } \varphi_k(x) \le 0. \tag{P$_k$}$$

We denote by v and v_k the optimal value of problems (P) and (P$_k$), respectively. Moreover, we say that $\bar{x}$ is a *stationary point* of problem (P) if there exists $\lambda \ge 0$ such that $\lambda f(\bar{x}) = 0$ and

$$0 \in \partial\psi(\bar{x}) + \lambda\partial f(\bar{x}).$$

Similarly, x_k is a stationary point of (P$_k$) if there is $\lambda_k \ge 0$ such that $\lambda_k\varphi(x_k) = 0$ and

$$0 \in \partial\psi(x_k) + \lambda_k\partial\varphi_k(x_k).$$

The next result strongly relies on the epigraphical convergence established in Theorem 3.1. We provide the proof for the sake of completeness.

Theorem 4.1. *Under the assumptions of Theorem 3.1, consider the sequence of functions $\varphi_k : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ defined in (3.1), with $\rho_k > 0$ such that $\rho_k^{-1}\ln(k) \to 0$. In addition, suppose that the objective function of (P) is coercive and $(\operatorname{dom}\psi) \cap \{x : f(x) \le 0\} \ne \emptyset$. Then,*

(i) v_k converges to v.

(ii) *Let* $x_k \in \operatorname{argmin}(\mathrm{P}_k)$ *for* $k \in \mathbb{N}$. *Then* (x_k) *is bounded and every accumulation point of* (x_k) *is a solution of problem* (P).

Proof. Notice that optimization problem (P) has an optimal solution since the coercivity of ψ implies $\{x \in \mathbb{R}^n : \psi(x) \le M, \text{ and } f(x) \le 0\}$ is compact and nonempty for M large enough.

(i) We start by showing that $\liminf_{k\to\infty} v_k \ge v$. Indeed, suppose by contradiction that there exist $\alpha \in \mathbb{R}$, and a sequence $x_{k_j} \in \mathbb{R}^n$ such that $\varphi_{k_j}(x_{k_j}) \le 0$ and $\psi(x_k) < \alpha < v$. Since ψ is coercive, we can assume that x_{k_j} converges to some $\bar{x}$. Now, using Theorem 3.1, we obtain that $f(\bar{x}) \le 0$ and $\psi(\bar{x}) \le \alpha < v$, which is a contradiction.

Now, we focus on $\limsup_{k\to\infty} v_k \le v$. Let us consider a point $x_0 \in \operatorname{argmin}(P)$. Then, using (3.4), we see that x_0 is a feasible point of optimization problem (P_k). Therefore, $v_k \le \psi(x_0) = v$, and it implies that $\limsup_{k\to\infty} v_k \le v$.

(ii) First, let us consider a point $x_0 \in \operatorname{dom}\psi \cap \{x : f(x) \le 0\}$. Then, $x_0 \in \operatorname{dom}\psi \cap \{x : \varphi_k(x) \le 0\}$, and by optimality of x_k, we have that $x_k \in \{x : \psi(x) \le \psi(x_0)\}$, where the last set is bounded due to the coercitivity of ψ, and this yields the conclusion that (x_k) is bounded. Now, let us suppose that $x_{k_j} \to \bar{x}$. First, let us notice that by Theorem 3.1 $f(\bar{x}) \le \liminf_{k_j\to\infty} \varphi_{k_j}(x_{k_j}) \le 0$, which implies that $\bar{x}$ is a feasible point of problem (P). Second, we consider $k \in \mathbb{N}$ and $x \in \mathbb{R}^n$ be a feasible point of problem (P), that is, $f(x) \le 0$. Using (3.4), we obtain that $\varphi_k(x) \le m_k(x) \le f(x) \le 0$. Hence, the point x is also a feasible point of problem (P_k). It follows from the optimality of x_k that $\psi(x_k) \le \psi(x)$, which implies by the lower semicontinuity of ψ that $\psi(\bar{x}) \le \psi(x)$, which shows the optimality of $\bar{x}$.

□

Now, we are in a position to understand the asymptotic behavior of the accumulation points of the stationary points of problem (P_k). Naturally, these results rely on the properties of the limit of subgradients of the function φ_k in relation to the supremum function, which were explored in Section 3. The first result relies in the setting of Theorem 3.2.

Corollary 4.1. *Under the assumptions of Theorem 4.1, let* x_k *be a stationary point of problem* (P_k) *for* $k \in \mathbb{N}$, *and* $\bar{x}$ *an accumulation point of* x_k. *Suppose that* T *is compact and there exists* $\varepsilon > 0$ *such that the functions* $\{f_t : t \in T\}$ *are differentiable on* $\mathbb{B}_\varepsilon(\bar{x})$ *and* $(t,x) \mapsto (\psi(x), f_t(x), \nabla f_t(x))$ *is continuous on* $T \times \mathbb{B}_\varepsilon(\bar{x})$. *Then the point* $\bar{x}$ *is a stationary point of problem* (P) *provided that the following condition holds*

$$(-\partial^\infty \psi(\bar{x})) \cap \operatorname{co}\{\nabla f_t(\bar{x}) : t \in T(\bar{x})\} = \emptyset. \tag{4.1}$$

Proof. Let x_k be a stationary point of problem (P_k) for $k \in \mathbb{N}$, and $\bar{x}$ an accumulation point of $\{x_k\}$, let us say $x_{k_j} \to \bar{x}$. It follows from the definition that there exist λ_{k_j} such that $\lambda_{k_j}\varphi_{k_j}(x_{k_j}) = 0$ and

$$0 \in \partial\psi(x_k) + \lambda_{k_j}\nabla\varphi_{k_j}(x_{k_j}).$$

Under our assumptions, it is not difficult to show that $\{\nabla\varphi_{k_j}(x_{k_j})\}$ is bounded, so we can assume (without relabeling) that $\varphi_{k_j}(x_{k_j}) \to w^*$.

On the one hand, let us suppose that λ_{k_j} is bounded, so without relabeling we can assume that $\lambda_{k_j} \to \bar{\lambda}$. It follows from Theorem 3.2 that $w^* \in \partial f(\bar{x})$. Moreover, since $\psi(x_k) \to \psi(\bar{x})$, we have that $0 \in \partial\psi(\bar{x}) + \bar{\lambda}\partial f(\bar{x})$. Moreover, by compactness of T and continuity of $(t,x) \to f_t(x)$, we see that the assumption of Corollary 3.1 holds. Then $\lim_{j\to\infty}\varphi_{k_j}(x_{k_j}) = f(\bar{x})$. Therefore $\bar{\lambda} f(\bar{x}) = 0$, so, in this case, the conclusion of the corollary holds.

On the other hand, let us suppose that (λ_{k_j}) is unbounded. Taking a subsequence if it is necessary, we can suppose that $\lambda_{k_j} \to +\infty$. Then, it follows that there exists $w^*_{k_j} \in \partial\psi(x_{k_j})$ such that $\nabla\varphi_{k_j}(x_{k_j}) = -\lambda_{k_j}^{-1} w^*_{k_j}$. Now, using Proposition 2.1 and Theorem 3.2, we have that $w^* \in \partial f(\bar{x}) = \mathrm{co}\{\nabla f_t(\bar{x}) : t \in T(\bar{x})\}$. It implies that

$$v^* \in (-\partial^\infty\psi(\bar{x})) \cap \mathrm{co}\{\nabla f_t(\bar{x}) : t \in T(\bar{x})\},$$

but, by definition of the singular subdifferential, we also have that $-w^* \in \partial^\infty\psi(\bar{x})$, which contradicts (4.1), and that ends the proof of the corollary. □

Now, we estate a similar result to the previous corollary, under the assumptions of Theorem 3.3. The proof uses similar arguments, so we omit the details.

Corollary 4.2. *Under the assumptions of Theorem 4.1, let x_k be a stationary point of problem (P_k) for $k \in \mathbb{N}$, and $\bar{x}$ an accumulation point of x_k. Suppose that ψ is Lipschitz continuous around $\bar{x}$ and there exist a continuously differentiable function $h : \mathbb{R}^n \to \mathbb{R}$ such that $f_t + h$ is proper lower semicontinuous convex function for all $t \in T$. Then the point $\bar{x}$ is a stationary point of problem* (P).

Remark 4.1 (min-max problems)**.** Notice that Theorem 4.1 can be applied to the following min-max problem

$$\min_{x\in C} f(x), \tag{4.2}$$

where f is the supremum function defined in (2.2), and its corresponding approximation

$$\min_{x\in C} \varphi_k(x). \tag{4.3}$$

In order to see the connection, it is enough to notice that problem (4.2) can be equivalently reformulated as

$$\min \psi(x,\beta) \text{ subject to } \hat{f}(x,\beta) \le 0, \tag{4.4}$$

where $\psi(x,\beta) := \beta + \delta_C(x)$ and $\hat{f}(x,\beta) = f(x) - \beta$. In that case, we can compute the regularization (3.1) associated to the function $\hat{f}$, let us say, $\hat{\varphi}_k$ corresponds to $\hat{\varphi}_k(x,\beta) = \varphi_k(x) - \beta$, with φ_k the regularization (3.1) associated to the function f. Therefore, the regularized problem proposed in (P_k) associated to (4.4) is nothing more than

$$\min \psi(x,\beta) \text{ subject to } \varphi_k(x) \le \beta,$$

which is equivalent to problem (4.3).

4.1. **Duality results.** In this section, we use the notation $f(t,x) \equiv f_t(x)$ and $\nabla_x f(t,x)) \equiv \nabla f_t(x))$. We assume that ψ is C^1 on the whole space $\mathbb{R}^n$, and we suppose that the functions ψ and $f_t, t \in T$, are convex on $\mathbb{R}^n$, and $(t,x) \mapsto (f(t,x), \nabla_x f(t,x))$ is continuous on $T \times \mathbb{R}^n$. Here $C(T)$ is the Banach space of real-valued continuous functions on T, equipped with the maximum norm $\|h\| = \max\{|h(t)| : t \in T\}$, and $C_+(T)$ is the cone of nonnegative valued functions in $C(T)$. By $M(T)$ we represent the topological dual of $C(T)$; i.e., the space of finite signed Borel measures on T, embedded with the total variation norm. We have

$$\langle h,\sigma\rangle := \int_T h(t)\sigma(dt)\ \forall\sigma\in M(T),\ \forall h\in\mathscr{C}(T).$$

Since T is a metric space, $C(T)$ is separable and every finite signed Borel measure on T is regular (see, e.g., [9, Example 2.37].)

By $M_+(T)$, we represent the positive cone of $M(T)$, i.e., the subset of $M(T)$ composed by the finite Borel measures on T. For $\sigma\in M_+(T)$, we have $||\sigma|| = \int_T \sigma(dt)$. The key in this section is the notion of *Lagrangian function* associated with (P), i.e.,

$$L(x,\sigma) := \psi(x) + \langle f(\cdot,x),\sigma\rangle = \psi(x) + \int_T f(t,x)\sigma(dt),$$

with $x\in\mathbb{R}^n$ and $\sigma\in M_+(T)$.

Associated with our primal problem (P), we define the usual *Lagrangian dual* problem

$$\sup\nolimits_{\sigma\in M_+(T)} \Psi(\sigma), \tag{(D)}$$

where

$$\Psi(\sigma) := \inf\{L(x,\sigma) :\ x\in\mathbb{R}^n\}.$$

If $v(\mathrm{D})$ represents the optimal value of (D), the so-called *weak duality* inequality $v(\mathrm{P})\geq v(\mathrm{D})$ always holds. The optimal set of (D) is denoted by S_D.

The following theorem gathers the most relevant properties of the dual pair. The proof comes straightforwardly from Theorems 5.97 and 5.98, Corollary 5.109, and 5.278 in [9].

Theorem 4.2. *Assume that ψ is level bounded on $F := \{x : f(x)\leq 0\}$, where $f=\sup_{t\in T} f_t$, and the Slater condition holds; i.e., there exists x^0 such that $f(x^0)<0$. Then, the following statements hold:*

(i) The strong duality $v(\mathrm{D}) = v(\mathrm{P})$ is satisfied, and the dual optimal set S_D is nonempty and bounded for the total variation norm.

(ii) If $\overline{\sigma}\in M_+(T)$ and $\overline{x}\in \operatorname{argmin}\Psi(\overline{\sigma})$ are such that $f(t,\overline{x})\leq 0$ for all $t\in T$ and $\langle f(\cdot,\overline{x}),\overline{\sigma}\rangle = 0$, then $\overline{x}$ and $\overline{\sigma}$ are optimal for (P) and (D), respectively. Moreover, under the current assumptions,

$$\overline{x}\in \operatorname{argmin}\Psi(\overline{\sigma}) \Longleftrightarrow \nabla_x L(\overline{x},\overline{\sigma}) = \nabla\psi(\overline{x}) + \int_T \nabla_x f(t,\overline{x})\overline{\sigma}(dt) = 0.$$

Suppose that x_k is an optimal solution of (P_k). Since this problem is convex, assuming that Slater holds, the KKT-conditions are necessary and there will exist $\lambda_k\geq 0$, $k=1,2,\ldots$, satisfying

$$0 = \nabla\psi(x_k) + \lambda_k\nabla\varphi_k(x_k) \text{ and } \lambda_k\varphi_k(x_k) = 0,\ k=1,2,\ldots.$$

Now we introduce the sequence of discrete measures $\{\sigma_k\}$ associated with the sequence $\{x_k\}$ by

$$\sigma_k := \lambda_k\sum_{i=1}^{k}\alpha_i(x_k)\delta_i,\ \text{where } \alpha_i(x_k) := \frac{\exp\left[\rho_k f_{t_i}(x_k)\right]}{\sum_{j=1}^{k}\exp\left[\rho_k f_{t_j}(x)\right]} \tag{4.5}$$

and δ_i is the Dirac distribution concentrated at point t_i. Observe that $||\sigma_k|| = \int_T \sigma_k(dt) = \lambda_k$, $k = 1, 2, \ldots$, and thanks to (3.12),

$$\langle \nabla_x f(., x_k), \sigma_k \rangle = \int_T \nabla_x f(t, x_k)\sigma_k(dt) \tag{4.6}$$

$$= \lambda_k \sum_{i=1}^{k} \alpha_i(x_k)\nabla f_{t_i}(x_k) = \lambda_k \nabla \varphi_k(x_k). \tag{4.7}$$

Next we proceed by applying Theorem 4.2 in order to prove the following dual convergence theorem, in which we establish the weak*-convergence of a sequence $\{\sigma_k\} \in M(T)$, to some element $\overline{\sigma} \in M(T)$; i.e., $\lim_{k\to\infty} \langle h, \sigma^k \rangle = \langle h, \overline{\sigma} \rangle$, for all $h \in C(T)$.

Theorem 4.3. *Assume that ψ is level bounded on $F := \{x : f(x) \leq 0\}$, that we are under the same assumptions that in the third statement of Theorem 4.1 and, additionally, the Slater condition holds and $\rho_k^{-1} \ln(k) \to 0$. Then, the following statements hold:*

(i) The sequence $\{\sigma_k\}$ given in (4.5) is strongly bounded.

(ii) There exists at least a weak$^*-$limit point of this sequence, and each weak$^*-$limit point of this sequence belongs to S_D.

Remark 4.2 (Previous to the proof)**.** From the assumption that ψ is C^1 on the whole space $\mathbb{R}^n$, we have that $\partial^\infty \psi(\bar{x}) = \{0_n\}$ and (4.1) becomes $0 \notin \operatorname{co}\{\nabla f_t(\bar{x}) : t \in T(\bar{x})\}$, which is implied by Slater.

Proof. *(i)* In the proof of the statement 3 in Theorem 4.1, we have established that $\{\sigma_k\}$ is strongly bounded as $\|\sigma_k\| = \lambda_k$ and there is no sequence $\{\lambda_{k_j}\}$ converging to $+\infty$. Therefore, without loss of generality, we assume that $\lambda_k \to \overline{\lambda} \in [0, +\infty[$.

(ii) Observe that $\{\sigma_k\}$ is bounded. Since $\mathbb{B}^* = \{\sigma \in M(T) : \|\sigma\| \leq 1\}$ is weak$^*-$ sequentially compact, we see that there exist at least a weak*-limit point. Let $\overline{\sigma}$ be an arbitrary weak*-limit point of this sequence. Given an accumulation point $\bar{x}$ of $\{x_k\}$, that is optimal for problem (P), and there will exist a subsequence $\{\sigma_k\}_{k\in K}$ such that

$$\lim_{k\to\infty, k\in K} x_k = \bar{x} \text{ and } w^* - \lim_{k\to\infty, k\in K} \sigma_k = \overline{\sigma} \in M_+(T).$$

Since

$$\langle f(\cdot, x_k), \sigma_k \rangle = \lambda_k \sum_{i=1}^{k} \alpha_i(x_k) f(t_i, x_k) \leq \lambda_k m_k(x_k)$$

$$\leq \lambda_k \left(\varphi_k(x_k) + \frac{\ln(k)}{\rho_k} \right) \leq \frac{\lambda_k \ln(k)}{\rho_k},$$

and $\langle f(\cdot, x_k), \sigma_k \rangle \to \langle f(\cdot, \bar{x}), \overline{\sigma} \rangle \leq 0$. We can say more. Indeed, following the same argument that in the proof of Claim 1 in Theorem 3.2, we conclude that the support of $\overline{\sigma}$ is included in $T(\bar{x})$ and, consequently, $\langle f(\cdot, \bar{x}, \overline{\sigma} \rangle = 0$.

Finally (see the proof of statement 3 in Theorem 4.1), thanks to (4.6) and (4.7), one has

$$0 = \nabla \psi(x_k) + \lambda_k \nabla \varphi_k(x_k) = \nabla \psi(x_k) + \langle \nabla_x f(\cdot, x_k), \sigma_k \rangle.$$

Taking limits yields $0 = \nabla \psi(\bar{x}) + \langle \nabla_x f(\cdot, \bar{x}), \overline{\sigma} \rangle$. Then applying Theorem 4.2, it follows that $\overline{\sigma}$ is optimal for dual problem (D).

□

5. Concluding Remarks

In this paper, we investigate an approximation technique for a supremum function $f := \sup_{t\in T} f_t$, where $\{f_t, t \in T\}$ is a general family of functions defined on the Euclidean space and valued in the extended real line, whereas T is a metric space. The main conclusions of the paper are the following:

- We start by observing that if the functions $T \ni t \mapsto f_t(x)$, $x \in \mathbb{R}^n$, are lsc, the supremum function can be determined by considering only a countable dense subset $\{t_i\}_{i\in\mathbb{N}}$ of the T.
- The approximating function is log-exponential. More precisely, if $\rho_k > 0$ with $k \in \mathbb{N}$, we consider the function $\varphi_k : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, given by
$$\varphi_k(x) := \frac{1}{\rho_k} \ln\left(\frac{1}{k}\sum_{i=1}^{k} \exp\left[\rho_k f_{t_i}(x)\right]\right).$$
- The functions φ_k, $k = 1,2,...$, epigraphically converge to f.
- Under the additional requirement of uniform continuity of the original functions with respect to t, the functions φ_k, $k = 1,2,...$, are continuously convergent to f.
- Refined calculus rules are established specifying the relations among the subdifferentials (regular and general) of φ_k and that of the functions f_{t_i}, $i = 1,...,k$.
- Under mild assumptions, the Painlevé-Kuratowski convergence of the (sub-)gradients of φ_k to the subdifferential $\partial f(\bar{x})$ is also proven.
- Given a general semi-infinite programming problem $(\mathrm{P}) : \min \psi(x)$ subject to $f_t(x) \le 0$, $t \in T$, we propose to replace it by the simple smooth problem, with a single constraint, $(\mathrm{P}_k) : \quad \min \psi(x)$ subject to $\varphi_k(x) \le 0$.
- Under standard hypotheses, we prove that the optimal values of the problems (P_k) converge to the optimal value of (P), and that the accumulation points of any sequence of optimal solutions of the problems (P_k) are optimal solutions of the problem (P).
- The Lagrangian duality of this problem is also studied, but restricted to the semi-infinite convex programming problem. We establish a procedure to construct an optimal dual solution.

Acknowledgments
The first author was partially supported by Centro de Modelamiento Matemático (CMM), ACE210010 and FB210005, BASAL funds for center of excellence and ANID-Chile grant Fondecyt Regular 1240335. The second author has been partially supported by Grant PID2022-136399NB-C21 founded byb MICIU/AEI/10.13039/501100011033. The third author was partially supported by Centro de Modelamiento Matemático (CMM), ACE210010 and FB210005, BASAL funds for center of excellence and ANID-Chile grant: MATH-AMSUD 23-MATH-09 and MATH-AMSUD 23-MATH-17, ECOS-ANID ECOS320027, Fondecyt Regular 1220886, Fondecyt Regular 1240335, Fondecyt Regular 1240120 and Exploración 13220097.

References

[1] F. Atenas, C. Sagastizábal, P. J. S. Silva, M. Solodov, A unified analysis of descent sequences in weakly convex optimization, including convergence rates for bundle methods, SIAM J. Optim. 33 (2023), 89-115.

[2] H. Attouch, Variational convergence for functions and operators, Applicable Mathematics Series, Pitman (Advanced Publishing Program), Boston, 1984.

[3] J.P. Aubin, H. Frankowska, Set-Valued Analysis, Modern Birkhäuser Classics, Birkhäuser, Boston, 2009.

[4] A Auslender, Méthodes et théorèmes de dualité, Revue française d'informatique et de recherche opérationnelle. Série rouge, 4 (1970), 9-45.

[5] A. Auslender, M. A. Goberna, M. A. López, Penalty and smoothing methods for convex semi-infinite programming, Math. Oper. Res. 34 (2009), 303-319.

[6] A. Ben-Tal, M. Teboulle, A smoothing technique for nondifferentiable optimization problems, Optimization, Proc. 5th French-German Conf., Varetz/Fr 1988, Lect. Notes Math. 1405, pp. 1-11, 1989.

[7] D. P. Bertsekas, Approximation procedures based on the method of multipliers, J. Optim. Theory Appl. 23 (1977), 487-510.

[8] D. P. Bertsekas, Constrained optimization and Lagrange multiplier methods, Computer Science and Applied Mathematics. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London, 1982.

[9] J. F. Bonnans and A. Shapiro, Perturbation analysis of optimization problems, Springer Science & Business Media, 2013.

[10] P. Borges, C. Sagastizábal, and M. Solodov, A regularized smoothing method for fully parameterized convex problems with applications to convex and nonconvex two-stage stochastic programming, Math. Program. 189 (2021), 117-149.

[11] J. Borwein and Q. Zhu, Techniques of Variational Analysis, CMS Books in Mathematics, Springer New York, 2006.

[12] J. V. Burke and T. Hoheisel, Epi-convergent smoothing with applications to convex composite functions, SIAM J. Optim. 23 (2013), 1457-1479.

[13] J. V. Burke, T. Hoheisel, and C. Kanzow, Gradient consistency for integral-convolution smoothing functions, Set-Valued Var. Anal. 21 (2013), 359-376.

[14] A. Charnes, W. W. Cooper, and K. Kortanek, Duality in semi-infinite programs and some works of Haar and Carathéodory, Manage. Sci. 9 (1963), 209-228.

[15] A. Charnes, W. W. Cooper, and K. O. Kortanek, On the theory of semi-infinite programming and a generalization of the Kuhn-Tucker saddle point theorem for arbitrary convex functions, Nav. Res. Logist. Q. 16 (1969), 41-51.

[16] X. Chen, Smoothing methods for nonsmooth, nonconvex minimization, Math. Program. 134 (2012), 71-99.

[17] C. Combari and L. Thibault, On the graph convergence of subdifferentials of convex functions, Proc. Amer. Math. Soc. 126 (1998), 2231-2240.

[18] R. Correa, A. Hantoute, and M. A. López, Fundamentals of convex analysis and optimization. A supremum function approach, Springer Ser. Oper. Res. Financ. Eng. Springer, Cham, 2023.

[19] R. Correa, M. A. López, and P. Pérez-Aros, Necessary and sufficient optimality conditions in DC semi-infinite programming, SIAM J. Optim. 31 (2021), 837-865.

[20] R. Correa, M. A. López, and P. Pérez-Aros, Optimality conditions in dc-constrained mathematical programming problems, J. Optim. Theory Appl. 198 (2023), 1191-1225.

[21] J. M. Danskin, The theory of max-min and its application to weapons allocation problems, volume 5, Springer Science & Business Media, 2012.

[22] S.-C. Fang and S.-Y. Wu, Solving min-max problems and linear semi-infinite programs, Comput. Math. Appl. 32 (1996), 87-93.

[23] A. Haar, Über lineare Ungleichungen. Acta Litt. Sci. Szeged, 2 (1924), 1-14.

[24] 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), 863-882.

[25] R. Hettich and K. O. Kortanek, Semi-infinite programming: Theory, methods, and applications, SIAM Rev. 35 (1993), 380-429.

[26] J.-B. Hiriart-Urruty and R. R. Phelps, Subdifferential calculus using ε-subdifferentials, J. Funct. Anal. 118 (1993), 154-166.

[27] A. D. Ioffe and V. M. Tihomirov, Theory of extremal problems. Translated from the Russian by K. Makowski, volume 6 of Stud. Math. Appl. Elsevier, Amsterdam, 1979.

[28] V. L. Levin, Application of e. helly's theorem to convex programming, problems of best approximation and related questions, Mathematics of the USSR-Sbornik, 8 (1969), 235.

[29] M. A. López and G. Still, Semi-infinite programming, Eur. J. Oper. Res. 180 (2007), 491-518.

[30] B. S. Mordukhovich, Variational analysis and generalized differentiation. I, volume 330 of Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Springer-Verlag, Berlin, 2006.

[31] B. S. Mordukhovich, Variational Analysis and Applications, Springer, Cham, 2018.

[32] 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), 406-431.

[33] Yu. Nesterov, Smooth minimization of non-smooth functions, Math. Program. 103 (2005), 127-152.

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

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

[36] P. Pérez-Aros, Ergodic approach to robust optimization and infinite programming problems, Set-Valued Var. Anal. 29 (2021), 409-423.

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

[38] P. Pérez-Aros and E. Vilches, Moreau envelope of supremum functions with applications to infinite and stochastic programming, SIAM J. Optim. 31 (2021), 1635-1657.

[39] E. Polak, Smoothing techniques for the solution of finite and semi-infinite min-max-min problems, In: High performance algorithms and software for nonlinear optimization (Erice, 2001), volume 82 of Appl. Optim., pp. 343–362. Kluwer Acad. Publ., Norwell, MA, 2003.

[40] E. Polak, J.E. Higgins, and D. Mayne, A barrier function method for minimax problems, Math. Program. 54 (1992), 155-176.

[41] E. Polak, J. O. Royset, and R. S. Womersley, Algorithms with adaptive smoothing for finite minimax problems, J. Optim. Theory Appl. 119 (2003), 459-484.

[42] B. N. Pshenichnyj, Necessary conditions for an extremum, volume 4 of Pure Appl. Math., Marcel Dekker. Marcel Dekker, Inc., New York, NY, 1971.

[43] R. T. Rockafellar, Directionally Lipschitzian functions and subdifferential calculus. Proc. Lond. Math. Soc. 3 (1979), 331-355.

[44] R. T. Rockafellar and R. J.-B. Wets, Variational analysis, volume 317 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, 1998.

[45] R. L. Sheu and J. Y. Lin, Solving continuous min-max problems by an iterative entropic regularization method, J. Optim. Theory Appl. 121 (2004), 597-612.

[46] R. L. Sheu and S. Y. Wu, Combined entropic regularization and path-following method for solving finite convex min-max problems subject to infinitely many linear constraints, J. Optim. Theory Appl. 101 (1999), 167-190.

[47] L. Thibault, Unilateral variational analysis in Banach spaces. Part I: General theory. Part II: Special classes of functions and sets, World Scientific, Singapore, 2023.

[48] L. Thibault and T. Zakaryan, Convergence of subdifferentials and normal cones in locally uniformly convex Banach space, Nonlinear Anal. 98 (2014), 110-134.

[49] M. Valadier, Sous-différentiels d'une borne supérieure et d'une somme continue de fonctions convexes, C. R. Acad. Sci., Paris, Sér. A, 268 (1969), 39-42.

[50] M. Volle, On the subdifferential of an upper envelope of convex functions, Acta Math. Vietnam. 19 (1994), 137-148.