JOTA manuscript No.
(will be inserted by the editor)

Optimality Functions and Lopsided Convergence

Johannes O. Royset · Roger J-B Wets

Communicated by Igor Konnov

Received: date / Accepted: date

Abstract Optimality functions pioneered by E. Polak characterize stationary points, quantify the degree with which a point fails to be stationary, and play central roles in algorithm development. For optimization problems requiring approximations, optimality functions can be used to ensure consistency in approximations, with the consequence that optimal and stationary points of the approximate problems indeed are approximately optimal and stationary for an original problem. In this paper, we review the framework and illustrate its application to nonlinear programming and other areas. Moreover, we introduce lopsided convergence of bifunctions on metric spaces and show that

Johannes O. Royset, Corresponding author

Naval Postgraduate School, Monterey, California

joroyset@nps.edu

Roger J-B Wets

University of California, Davis, California

rjbwets@ucdavis.edu

this notion of convergence is instrumental in establishing consistency of approximations. Lopsided convergence also leads to further characterizations of stationary points under perturbations and approximations.

Keywords epi-convergence, lopsided convergence, consistent approximations, optimality functions, optimality conditions

Mathematics Subject Classification (2000) 90C46 · 49J53

Dedication. We dedicate this paper to our long-time friend, colleague, collaborator, and advisor Elijah (Lucien) Polak in honor of his 85th birthday. We wish him fair weather and following snow conditions.

1 Introduction

It is well-known that optimality conditions are central to both theoretical and computational advances in optimization. They were developed over centuries starting with the pioneering works of Bishop N. Oresme (14th century) and P. de Fermat (17th century), and brought to their modern form by Karush, John, Kuhn, Tucker, Polak, Mangasarian, Fromowitz, and many others. In this paper, we discuss quantification of first-order necessary optimality conditions in terms of *optimality functions* as developed by E. Polak and co-authors; see [16] for numerous examples in nonlinear programming, semi-infinite optimization, and optimal control as well as [19, 18, 8, 13] for recent applications in stochastic and semi-infinite programming, nonsmooth optimization, and control of uncertain systems.

It is apparent that how "far" a set of equalities, inequalities, and inclusions are from being satisfied can be quantified in numerous ways; see for example [11] for a survey as well as early work in [22]. The framework of optimality functions, as laid out in [16, Section 3.3] and references therein, stipulates axiomatic requirements that such quantifications should satisfy to facilitate the study and computation of approximate stationary points. Specifically, for an optimization problem that can only be "solved" through the solution of an approximating problem, one seeks to determine whether a near-stationary point of the approximating problem is an approximate stationary point of the original problem. The requirements on optimality functions exactly ensure this property. Moreover, there is ample empirical indications and some theoretical evidence (see for example [18,21,20,12]) that computational benefits accrue from approximately solving a sequence of approximating problems with increasing fidelity, each warm-started with the previously obtained point. Optimality functions are tools to carry out such a scheme and give rise to adaptive rules for determining the timing of switches to higher-fidelity approximations. Consequently, the framework of optimality functions provides a pathway to constructing implementable algorithms consisting only of a finite number of arithmetic operations and function evaluations[1].

We here use the terminology "optimality functions," but this does not exclude the consideration of many familiar "residual functions" and "gap func-

[1] The distinction between implementable and conceptual algorithms appears to be due to E. Polak [15,14].

tions;" see [11] for examples. In fact, after minor adjustments many of these functions can be viewed as optimality functions. Our examples are simply illustrations.

In this paper, we review the notion of optimality functions and illustrate the vast number of possibilities through several examples. In an application to nonlinear programming, we establish the convergence of a primal interior point method in the absence of constraint qualifications and convexity assumptions. We show that lopsided convergence of bifunctions [2, 9, 10] is a useful tool for analyzing optimality functions and the associated stationary points. In particular, we prove that lopsided convergence of certain bifunctions, defining optimality functions of approximating problems, to a bifunction associated with an optimality function of the original problem, guarantees the axiomatic requirements on optimality functions. Using lopsided convergence, we provide results on existence of stationary points as well as characterizations of stationary points under perturbations and approximations. In the process, we extend the primary definitions and results on lopsided convergence in [9, 10] from finite dimensions to metric spaces.

The paper is organized as follows. Section 2 defines optimality functions and gives several examples. Section 3 introduces approximating optimization problems, epi-convergence, and consistent approximations as defined by corresponding optimality functions, and demonstrates the implication for algorithmic development. Section 4 develops lopsided convergence for metric spaces.

The paper ends by utilizing lopsided convergence in the context of optimality functions.

2 Optimality Functions: Definitions and Examples

We consider optimization problems defined on a metric space $(\mathcal{X}, d_{\mathcal{X}})$, where $C \subset \mathcal{X}$ is a nonempty feasible set and $f : C \to I\!R$ an objective function, i.e., problems of the form

$$\text{minimize } f(x) \text{ subject to } x \in C \subset \mathcal{X}.$$

The function f might be defined and finite-valued outside C, but that will be immaterial to the following treatment. Thus, the notation $f : C \to I\!R$ specifies the components f and C of optimization problems of this form, without implying that f is necessarily finite *only* on C.

We denote by $\inf_C f \in [-\infty, \infty[$ and $\operatorname{argmin}_C f \subset C$ the corresponding optimal value and set of optimal points, respectively, the latter possibly being empty. For $\varepsilon \geq 0$, the set of ε-optimal solutions is denoted by

$$\varepsilon\text{-}\operatorname{argmin}_C f := \{x \in C \ : \ f(x) \leq \inf_C f + \varepsilon\}.$$

As usually, we say that $x^* \in \mathcal{X}$ is locally optimal (for $f : C \to I\!R$) if and only if there exists a $\delta > 0$ such that $f(x^*) \leq f(x)$ for all $x \in C$ with $d_{\mathcal{X}}(x, x^*) \leq \delta$.

Throughout the paper, we have that C is a nonempty subset of $\mathcal{X}$ and $I\!R_- := [-\infty, 0]$. We characterize stationary points in terms of optimality functions as defined next.

Definition 2.1 (optimality function) An upper semicontinuous function $\theta : X \to I\!R_-$ is an optimality function for $f : C \to I\!R$ if and only if $C \subset X \subset \mathcal{X}$ and

$$x \in C \text{ locally optimal for } f : C \to I\!R \implies \theta(x) = 0.$$

The corresponding sets of stationary points and quasi-stationary points are $\mathcal{S}_{C,\theta} := \{x \in C \ : \ \theta(x) = 0\}$ and $\mathcal{Q}_\theta := \{x \in X \ : \ \theta(x) = 0\}$, respectively.

A series of examples help illustrate the concept; see also §5 and [16,19,18, 8,13]. For related "residual functions" see for example [11].

Example 2.1 (constrained optimization over convex set) Consider the case $\mathcal{X} = I\!R^n$, $C \subset \mathcal{X}$ closed and convex, and $f : C \to I\!R$ continuously differentiable. Then, the function

$$\theta(x) = \min_{y \in C} \left\{ \langle \nabla f(x), y - x \rangle + \frac{1}{2}\|y - x\|^2 \right\}, x \in X = C,$$

satisfies the requirements of Definition 2.1 and is therefore an optimality function for $f : C \to I\!R$. If $C = I\!R^n$, then the expression simplifies to

$$\theta(x) = -\frac{1}{2}\|\nabla f(x)\|^2, \tag{1}$$

which, of course, corresponds to the classical stationarity condition $\nabla f(x) = 0$.

Example 2.2 (nonlinear programming) Consider the case $\mathcal{X} = I\!R^n$, constraint set $C = \{x \in I\!R^n \ : \ f_j(x) \leq 0, j = 1, ..., q\}$, and f, f_1, ..., f_q real-valued and continuously differentiable on $I\!R^n$. Let $\psi(x) = \max_{j=1,...,q} f_j(x)$ and constraint violation $\psi_+(x) = \max\{0, \psi(x)\}$. Then, the function $\theta : I\!\!R^n \to I\!\!R_-$ given by

$$\theta(x) = \min_{y \in \boldsymbol{R}^n} \max \Big\{ -\psi_+(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2}\|y - x\|^2, \qquad (2)$$
$$\max_{j=1,\dots,q} \{ f_j(x) - \psi(x)_+ + \langle \nabla f_j(x), y - x \rangle \} + \frac{1}{2}\|y - x\|^2 \Big\}$$

satisfies the requirements of Definition 2.1 and is therefore an optimality function for $f : C \to I\!\!R$. The condition $\theta(x) = 0$ is equivalent to the Fritz-John conditions in the sense that when $x \in C$,

$$\theta(x) = 0 \iff \text{ there exist } \mu_0, \mu_1, \dots, \mu_q \geq 0, \text{ with } \sum_{j=0}^{q} \mu_j = 1,$$
$$\text{such that } \mu_0 \nabla f(x) + \sum_{j=1}^{q} \mu_j \nabla f_j(x) = 0, \; \sum_{j=1}^{q} \mu_j f_j(x) = 0.$$

However, since θ is defined beyond C, it might also be associated with quasi-stationary points outside C. We refer to [16, Theorem 2.2.8] for proofs and further discussion.

Example 2.3 (minimax problem) Consider the case $\mathcal{X} = C = I\!\!R^n$ and objective $f(x) = \max_{z \in Z} \varphi(x, z)$, $x \in I\!\!R^n$, where $\varphi : I\!\!R^n \times I\!\!R^p \to I\!\!R$ is continuous, the gradient $\nabla_x \varphi : I\!\!R^n \times I\!\!R^p \to I\!\!R^n$ with respect to the first argument exists and is continuous in both arguments, and Z is a compact subset of $I\!\!R^p$. Then, the function $\theta : I\!\!R^n \to I\!\!R_-$ given by

$$\theta(x) = \min_{y \in \boldsymbol{R}^n} \max_{z \in Z} \Big\{ \varphi(x, z) - f(x) + \langle \nabla_x \varphi(x, z), y - x \rangle + \frac{1}{2}\|y - x\|^2 \Big\} \qquad (3)$$

satisfies the requirements of Definition 2.1 and is therefore an optimality function for $f : I\!\!R^n \to I\!\!R$. Moreover, $\theta(x) = 0$ if and only if $0 \in \partial f(x)$ (the subdifferential of f); see [16, Theorem 3.1.6] for details.

We note that the upper semicontinuity of optimality functions ensures the computationally significant property that if a sequence $x^\nu \to x$ and $\theta(x^\nu) \nearrow 0$, for example with $\{x^\nu\}$ obtained as approximate solutions of a corresponding optimization problem with gradually smaller tolerance, then $\theta(x) = 0$ and $x \in \mathcal{Q}_\theta$, i.e., x is quasi-stationary. Although not discussed further here, the optimality functions in Examples 1-3, and others, are also instrumental in constructing descent directions for the respective optimization problems; see [16] for details.

3 Approximations and Implementable Algorithms

Problems involving functions defined in terms of integrals or optimization problems (as the maximization in Example 2.3), functions defined on infinite-dimensional spaces, and/or feasible sets defined by an infinite number of constraints almost always require approximations. For example, one might resort to an approximating space $\mathcal{X}^\nu \subset \mathcal{X}$ with points characterized by a finite number of parameters. Here, the superscript ν indicates that we might consider a family of such approximating spaces, $\nu \in I\!N := \{1, 2, ...,\}$, with usually $\cup_{\nu \in \mathbf{N}} \mathcal{X}^\nu$ dense in $\mathcal{X}$. A feasible set $C^\nu \subset \mathcal{X}^\nu$ may be an approximation of C or simply $C^\nu = C \cap \mathcal{X}^\nu$; see §5 for a concrete illustration in the area of optimal control. A function $f^\nu : C^\nu \to I\!R$ could be a tractable approximation of $f : C \to I\!R$. An example helps illustrate the situation.

Example 3.1 (minimax problem) Continuing from Example 2.3, suppose that $f^\nu(x) = \max_{z \in Z^\nu} \varphi(x, z)$, $x \in I\!R^n$, with $Z^\nu \subset Z$ consists of a finite number of points. Clearly, f^ν is a (lower bounding) approximation of $f = \max_{z\in Z} \varphi(\cdot, z)$ as defined above. The function $f^\nu : I\!R^n \to I\!R$ can be associated with the optimality function $\theta^\nu : I\!R^n \to I\!R_-$ given by

$$\theta^\nu(x) = \min_{y\in \boldsymbol{R}^n} \max_{z\in Z^\nu} \left\{ \varphi(x,z) - f^\nu(x) + \langle \nabla_x \varphi(x,z), y - x\rangle + \frac{1}{2}\|y - x\|^2 \right\},$$

which, as formalized in §5, approximates the optimality function θ in (3). We note that θ^ν can be evaluated in finite time by solving a convex quadratic program with linear constraints; see [16, Theorem 2.1.6].

We next examine approximating functions $f^\nu : C^\nu \to I\!R$ and review the notion of epi-convergence, which provides a path to establishing that optimal points of the corresponding approximating problems indeed approximate optimal points of an original problem. To establish the analogous results for stationary points, we turn to optimality functions and slightly extend the approach in [16, Section 3.3] by considering arbitrary metric spaces and other minor generalizations. The section ends with a result that facilitates the development of implementable algorithms for the minimization of $f : C \to I\!R$, which is then illustrated with the construction of an interior point method. Throughout the paper, we have that C^ν is a nonempty subset of $\mathcal{X}$.

3.1 Epi-Convergence

We recall that epi-convergence is the key property when examining approximations of optimization problems; see [1, 4, 17] for comprehensive treatments.

Definition 3.1 (epi-convergence) The functions $\{f^\nu : C^\nu \to I\!R\}_{\nu \in \mathbb{N}}$ epi-converge to $f : C \to I\!R$ if and only if

(i) for every sequence $x^\nu \to x \in \mathcal{X}$, with $x^\nu \in C^\nu$, we have $\liminf f^\nu(x^\nu) \geq f(x)$ if $x \in C$ and $f^\nu(x^\nu) \to \infty$ otherwise;

(ii) for every $x \in C$, there exists a sequence $\{x^\nu\}_{\nu \in \mathbb{N}}$, with $x^\nu \in C^\nu$, such that $x^\nu \to x$ and $\limsup f^\nu(x^\nu) \leq f(x)$.

A main consequence of epi-convergence is the following well-known result.

Theorem 3.1 (convergence of minimizers) *Suppose that* $\{f^\nu : C^\nu \to I\!R\}_{\nu \in \mathbb{N}}$ *epi-converges to* $f : C \to I\!R$. *Then,* $\limsup\,(\inf_{C^\nu} f^\nu) \leq \inf_C f$.

Moreover, if $x^k \in \operatorname{argmin}_{C^{\nu_k}} f^{\nu_k}$ *and* $x^k \to x$ *for some increasing subsequence* $\{\nu_1, \nu_2, ...\} \subset I\!N$, *then* $x \in \operatorname{argmin}_C f$ *and* $\lim_{k\to\infty} \inf_{C^{\nu_k}} f^{\nu_k} = \inf_C f$.

Proof. The second part is essentially in [3, Theorem 2.5], except for the finite-valued setting. The first and second parts are in [9, Theorem 2.6] for the $I\!R^n$ case. The proof carries over essentially verbatim. □

A strengthening of epi-convergence ensures the convergence of infima.

Definition 3.2 (tight epi-convergence) The functions $\{f^\nu : C^\nu \to I\!R\}_{\nu \in \mathbb{N}}$ epi-converge tightly to $f : C \to I\!R$ if and only if f^ν epi-converge to f and for all $\varepsilon > 0$, there exist a compact set $B_\varepsilon \subset \mathcal{X}$ and an integer ν_ε such that

$$\inf_{C^\nu \cap B_\varepsilon} f^\nu \leq \inf_{C^\nu} f^\nu + \varepsilon \quad \text{for all } \nu \geq \nu_\varepsilon.$$

Theorem 3.2 (convergence of infima) *Suppose that* $\{f^\nu : C^\nu \to I\!R\}_{\nu \in \mathbb{N}}$ *epi-converges to* $f : C \to I\!R$ *and* $\inf_C f$ *is finite. Then, they epi-converge tightly*

(i) *if and only if* $\inf_{C^\nu} f^\nu \to \inf_C f$.

(ii) *if and only if there exists a sequence* $\varepsilon^\nu \searrow 0$ *such that* ε^ν-$\operatorname{argmin}_{C^\nu} f^\nu$ *set-converges*[2] *to* $\operatorname{argmin}_C f$.

Proof. Again, the proof in [9, Theorem 2.8] can be immediately translated to the present setting. □

3.2 Consistent Approximations

The convergence of optimal points is fundamental, but an analogous result for stationary points is also important, especially for nonconvex problems. Optimality functions play a central role in the development of such results. Combining epi-convergence with a limiting property for optimality functions lead to consistent approximations in the sense of E. Polak as defined next. We note that our definition is an extension from that in [16, Section 3.3] as we consider arbitrary metric spaces and not only normed linear spaces.

Definition 3.3 (consistent approximations) The function and optimality function pairs $\{(f^\nu : C^\nu \to I\!R,\ \ \theta^\nu : X^\nu \to I\!R_-)\}_{\nu\in I\!N}$ are weakly consistent approximations of a pair $(f : C \to I\!R,\ \ \theta : X \to I\!R_-)$ if and only if

(i) $\{f^\nu : C^\nu \to I\!R\}_{\nu\in I\!N}$ epi-converge to $f : C \to I\!R$ and

[2] We recall that the outer limit of a sequence of sets $\{A^\nu\}_{\nu\in I\!N}$, denoted by $\limsup A^\nu$, is the collection of points y to which a subsequence of $\{y^\nu\}_{\nu\in I\!N}$, with $y^\nu \in A^\nu$, converges. The inner limit, denoted by $\liminf A^\nu$, is the points to which a sequence of $\{y^\nu\}_{\nu\in I\!N}$, with $y^\nu \in A^\nu$, converges. If both limits exist and are identical, we say that the set is the Painlevé-Kuratowski limit of $\{A^\nu\}_{\nu\in I\!N}$ and that A^ν set-converges to this set; see [6,17].

(ii) for every $x^\nu \to x \in \mathcal{X}$, with $x^\nu \in X^\nu$, $\limsup \theta^\nu(x^\nu) \leq \theta(x)$ if $x \in X$, and $\theta^\nu(x^\nu) \to -\infty$ otherwise.

If in addition $\theta^\nu(x) < 0$ for all $x \in X^\nu \backslash C^\nu$ and ν, then the pairs are consistent approximations of $(f : C \to I\!R,\ \ \theta : X \to I\!R_-)$.

We recall that the epigraph of $f : C \to I\!R$ is defined by

$$\operatorname{epi} f := \{(x, x_0) \in \mathcal{X} \times I\!R \ : \ x \in C, f(x) \leq x_0\}.$$

Since epi-convergence is equivalent to the set-convergence[3] of the corresponding epigraphs, we have that Definition 3.3(i) amounts to $\operatorname{epi} f^\nu$ set-converges to $\operatorname{epi} f$. Similarly, the hypograph of $f : C \to I\!R$ is defined by

$$\operatorname{hypo} f := \{(x, x_0) \in \mathcal{X} \times I\!R \ : \ x \in C, f(x) \geq x_0\}.$$

In view of the definition of set-convergence, we therefore have that Definition 3.3(ii) amounts to $\limsup \operatorname{hypo} \theta^\nu \subset \operatorname{hypo} \theta$.

The additional condition in Definition 3.3 removing "weakly" can be viewed as a constraint qualification as it eliminates the possibility of quasi-stationary points that are not stationary point for $f^\nu : C^\nu \to I\!R$, which might occur if the domain of θ^ν is not restricted to C^ν or other conditions are included.

The main consequence of consistency is given next.

Theorem 3.3 (convergence of stationary points) *Suppose that the pairs* $\{(f^\nu : C^\nu \to I\!R,\ \ \theta^\nu : X^\nu \to I\!R_-)\}_{\nu \in N}$ *are weakly consistent approximations*

[3] Here, we consider set-convergence of subsets of $\mathcal{X} \times I\!R$, which is equipped with the metric $\rho((x, x_0), (x', x_0')) = \max\{d_{\mathcal{X}}(x, x'), |x_0 - x_0'|\}$ for $x, x' \in \mathcal{X}$ and $x_0, x_0' \in I\!R$.

of $(f : C \to I\!R,\ \ \theta : X \to I\!R_-)$ *and* $\{x^\nu\}_{\nu\in N}$, $x^\nu \in X^\nu$, *is a sequence satisfying*

$$\theta^\nu(x^\nu) \geq -\varepsilon^\nu \text{ for all } \nu, \text{ with } \varepsilon^\nu \geq 0 \text{ and } \varepsilon^\nu \to 0.$$

Then, every cluster point x *of* $\{x^\nu\}_{\nu\in N}$ *satisfies* $x \in \mathcal{Q}_\theta$, *i.e.,* $\theta(x) = 0$.

If in addition the pairs are consistent approximations, $\varepsilon^\nu = 0$ *for sufficiently large* ν, *and* $\{f^\nu(x^\nu)\}_{\nu\in N}$ *is bounded from above, then* $x \in \mathcal{S}_{C,\theta}$.

Proof. Suppose that $x^\nu \to x$. Since $-\varepsilon^\nu \leq \theta^\nu(x^\nu)$ for all ν, $x \in X$. Moreover, $0 \leq \operatorname{limsup} \theta^\nu(x^\nu) \leq \theta(x) \leq 0$ and the first conclusion follows. By the definition of consistent approximations, $\theta^\nu(x^\nu) = 0$ for sufficiently large ν and therefore $x^\nu \in C^\nu$ for such ν. The epi-convergence of $f^\nu : C^\nu \to I\!R$ to $f : C \to I\!R$ implies that $\operatorname{liminf} f^\nu(x^\nu) \geq f(x)$ if $x \in C$ and $f^\nu(x^\nu) \to \infty$ if $x \notin C$. The latter possibility is ruled out by assumption and therefore $x \in C$. □

3.3 Algorithms

Theorem 3.3 provides a direct path to the construction of an implementable algorithm for minimizing $f : C \to I\!R$. Specifically, construct a family of approximations $\{f^\nu : C^\nu \to I\!R\}$ and a corresponding collection of optimality functions $\{\theta^\nu : X^\nu \to I\!R_-\}$, and then implement the following algorithm.

Algorithm.

1. Select $\{\varepsilon^\nu\}_{\nu\in N}$, with $\varepsilon^\nu \geq 0$ and $\varepsilon^\nu \to 0$. Initiate the iteration counter by setting $\nu = 1$.

2. Obtain an approximate (quasi-)stationary point x^ν for $f^\nu : C^\nu \to I\!R$ that satisfies $\theta^\nu(x^\nu) \geq -\varepsilon^\nu$.
3. Replace ν by $\nu + 1$ and go to Step 2.

If the pairs $\{(f^\nu : C^\nu \to I\!R,\ \ \theta^\nu : X^\nu \to I\!R_-)\}_{\nu\in \mathbb{N}}$ are weakly consistent approximations of $(f : C \to I\!R,\ \ \theta : X \to I\!R_-)$, then every cluster point of the constructed sequence $\{x^\nu\}$ will be quasi-stationary for $f : C \to I\!R$ by Theorem 3.3. The algorithm is fully implementable under the practically reasonable assumption that one can obtain an approximate quasi-stationary point of $f^\nu : C^\nu \to I\!R$ in finite time.

Example 3.2 (nonlinear programming) Continuing from Example 2.2, consider the standard logarithmic barrier approximation

$$f^\nu(x) = f(x) - t^\nu \sum_{j=1}^{q} \log[-f_j(x)],\ \ x \in C^\nu = \{x \in I\!R^n\ :\ f_j(x) < 0, j = 1, ..., q\},$$

where $t^\nu \searrow 0$. We first establish epi-convergence of $f^\nu : C^\nu \to I\!R$ to $f : C \to I\!R$. Suppose that $x^\nu \to x$, with $x^\nu \in C^\nu$. Since $C^\nu \subset C$ and C is closed, $x \in C$. Let $\varepsilon > 0$. There exists a ν_ε such that $-t^\nu \log[-f_j(x^\nu)] > -\varepsilon/q$ for all j with $\log[-f_j(x^\nu)] \geq 0$ and $\nu \geq \nu_\varepsilon$. Hence, $f^\nu(x^\nu) \geq f(x^\nu) - \varepsilon$ for all $\nu \geq \nu_\varepsilon$. In view of the continuity of f and the fact that ε is arbitrary, we conclude that Definition 3.1(i) is satisfied. Next, let $x \in C$. There exists a sequence $\{x^\nu\}_{\nu\in \mathbb{N}}$ such that $x^\nu \in C^\nu$ tends to x sufficiently slowly such that $t^\nu \sum_{j=1}^{q} \log[-f_j(x^\nu)] \to 0$. Consequently, $f^\nu(x^\nu) \to f(x)$, which satisfies Definition 3.1(ii). Therefore, $f^\nu : C^\nu \to I\!R$ epi-converge to $f : C \to I\!R$. We next analyze optimality functions. Using a minmax theorem, one can show that (2) is equivalently expressed as

$$\theta(x) = -\min_{\mu\in M}\left\{\mu_0\psi_+(x) + \sum_{j=1}^{q}\mu_j[\psi_+(x) - f_j(x)] \right. \tag{4}$$
$$\left. + \frac{1}{2}\left\|\mu_0\nabla f(x) + \sum_{j=1}^{q}\mu_j\nabla f_j(x)\right\|^2\right\},\quad x\in X = I\!R^n$$

where $M = \{(\mu_0,\mu_1,...,\mu_q) \ : \ \mu_j \geq 0, j = 0,1,...,q, \sum_{j=0}^{q}\mu_j = 1\}$; see [16, Theorem 2.2.8]. By (1) and direct differentiation of f^ν, we obtain an approximating optimality function

$$\theta^\nu(x) = -\frac{1}{2}\left\|\nabla f(x) + \sum_{j=1}^{q} m_j^\nu(x)\nabla f_j(x)\right\|^2,\quad x\in C^\nu,$$

where $m_j^\nu(x) = -t^\nu/f_j(x)$. Suppose that $x^\nu \to x \in I\!R^n$, with $x^\nu \in C^\nu$. Since $x^\nu \in C^\nu \subset C$ and C is closed, $x\in C$. Let

$$c^\nu = 1 + \sum_{j=1}^{q} m_j^\nu(x^\nu),\quad \mu_0^\nu = \frac{1}{c^\nu},\quad\text{and}\quad \mu_j^\nu = \frac{m_j^\nu(x^\nu)}{c^\nu}, j = 1,...,q.$$

Consequently, $\mu^\nu = (\mu_0^\nu,\mu_1^\nu,...,\mu_q^\nu)\in M$ for all ν. Since M is compact, $\{\mu^\nu\}$ has at least one convergent subsequence. Suppose that $\mu^\nu \to^N \mu^\infty$, with N an infinite subsequence of $I\!N$. If j is such that $f_j(x) < 0$, then $\mu_j^\nu \to^N 0$ and consequently $\mu_j^\infty = 0$ necessarily. In view of the continuity of the gradients,

$$\frac{\theta^\nu(x^\nu)}{(c^\nu)^2} = -\frac{1}{2}\left\|\frac{1}{c^\nu}\nabla f(x^\nu) + \sum_{j=1}^{q}\frac{m_j^\nu(x^\nu)}{c^\nu}\nabla f_j(x^\nu)\right\|^2$$
$$\to^N -\frac{1}{2}\left\|\mu_0^\infty\nabla f(x) + \sum_{j=1}^{q}\mu_j^\infty\nabla f_j(x)\right\|^2.$$

Since $x\in C$, $\psi_+(x) = 0$. Therefore we also have that

$$\frac{\theta^\nu(x^\nu)}{(c^\nu)^2} \to^N -\mu_0^\infty\psi_+(x) - \sum_{j=1}^{q}\mu_j^\infty[\psi_+(x) - f_j(x)]$$
$$-\frac{1}{2}\left\|\mu_0^\infty\nabla f(x) + \sum_{j=1}^{q}\mu_j^\infty\nabla f_j(x)\right\|^2 \leq \theta(x),$$

where the inequality follows from the fact that $\mu^\infty \in M$ furnishes a possibly suboptimal solution in (4). Because $\theta^\nu(x^\nu) \leq 0$ and $(c^\nu)^2 \geq 1$, the inequality remains valid when we drop the denominator on the left-hand side. Hence, we have shown that $\limsup \theta^\nu(x^\nu) \leq \theta(x)$. Thus $\{f^\nu : C^\nu \to I\!R,\ \theta^\nu : C^\nu \to I\!R_-)\}$ is consistent. Consequently, the above algorithm, which can then be viewed as a primal interior point method, generates cluster points that are stationary for $f : C \to I\!R$ in the sense of Fritz-John. We observe that this is achieved without any constraint qualifications and convexity assumptions. In this case, Step 2 of the algorithm can be achieved by any of the standard unconstrained optimization methods in finite time.

The key technical challenge associated with the above scheme is to establish (weak) consistency. In the next section, we provide tools for this purpose that rely on lopsided convergence.

4 Lopsided Convergence

In view of the definition of optimality functions, it is apparent that

$$\text{if } \mathcal{Q}_\theta \neq \emptyset, \text{ then } \mathcal{Q}_\theta = \operatorname{argmax}_X \theta.$$

Moreover, Examples 1-3 indicate that many optimality functions take the form

$$\theta(x) = \inf_{y \in Y} F(x, y), \ \text{ with } Y \subset \mathcal{Y} \tag{5}$$

for some metric space $(\mathcal{Y}, d_{\mathcal{Y}})$ and function F. In fact, in our examples, $\mathcal{Y} = I\!R^n$ and F involves gradients and other quantities; §5 provides an example in infinite dimensions. From these observations it is apparent that the consideration of maxinf-problems of the form

$$\max_{x\in X}\inf_{y\in Y} F(x,y)$$

for bifunction $F : X \times Y \to I\!R$ will provide direct insight about stationary and quasi-stationary points of optimization problems. We therefore set out to describe the fundamental tool for examining the convergence of such maxinf-problems, which is lopsided convergence first defined in [2]. (The stronger notion of epi/hypo-convergence [3,5] appears less suitable as it is directed towards saddle points; see the discussion in [10].) In the process, we extend some of the results in [9, 10] to general metric spaces.

Suppose that $(\mathcal{X}, d_{\mathcal{X}})$ and $(\mathcal{Y}, d_{\mathcal{Y}})$ are metric spaces, $X \subset \mathcal{X}$ and $Y \subset \mathcal{Y}$ are nonempty, and $F : X \times Y \to I\!R$ is a bifunction. We say that x^* is a maxinf-point of F if and only if

$$x^* \in \operatorname*{argmax}_{x\in X}\left\{\inf_{y\in Y} F(x,y)\right\}.$$

The study of such functions is facilitated by the notion of lopsided convergence.

Definition 4.1 (lopsided convergence) The sequence of bifunctions $\{F^\nu : X^\nu \times Y^\nu \to I\!R\}_{\nu\in \mathbb{N}}$ lop-converges to $F : X \times Y \to I\!R$ if and only if

(i) for all $y \in Y$ and $x^\nu \to x \in \mathcal{X}$, with $x^\nu \in X^\nu$, there exists $y^\nu \to y$, with $y^\nu \in Y^\nu$, such that $\limsup F^\nu(x^\nu, y^\nu) \le F(x,y)$ if $x \in X$ and $F^\nu(x^\nu, y^\nu) \to -\infty$ otherwise.

(ii) for all $x \in X$, there exists $x^\nu \to x$, with $x^\nu \in X^\nu$, such that for all $y^\nu \to y \in \mathcal{Y}$, with $y^\nu \in Y^\nu$, $\liminf F^\nu(x^\nu, y^\nu) \ge F(x,y)$ if $y \in Y$ and $F^\nu(x^\nu, y^\nu) \to \infty$ otherwise.

We assume throughout that the sets $X^\nu \subset \mathcal{X}$ and $Y^\nu \subset \mathcal{Y}$ are nonempty. We start with a preliminary result.

Proposition 4.1 (epi-convergence of slices) *Suppose* $\{F^\nu : X^\nu \times Y^\nu \to I\!R\}_{\nu \in N}$ *lop-converges to* $F : X \times Y \to I\!R$*. Then, for all* $x \in X$*, there exists* $x^\nu \to x$*, with* $x^\nu \in X^\nu$ *such that the functions* $F^\nu(x^\nu, \cdot) : Y^\nu \to I\!R$ *epi-converge to* $F(x, \cdot) : Y \to I\!R$*.*

Proof. We follow the same arguments as in [9, Proposition 3.2], where $\mathcal{X} = I\!R^n$ is considered. From Definition 4.1(ii) there exists $x^\nu \to x$, with $x^\nu \in X^\nu$, such that the functions $\{F^\nu(x^\nu, \cdot)\}_{\nu \in N}$ and $F(x, \cdot)$ satisfy Definition 3.1(i). From Definition 4.1(i), for any $y \in Y$ and $x^\nu \to x$, with $x^\nu \in X^\nu$, one can find $y^\nu \to y$, with $y^\nu \in Y^\nu$, such that Definition 3.1(ii) is also satisfied. □

We recall that the inf-projections of the bifunctions $F^\nu : X^\nu \times Y^\nu \to I\!R$ and $F : X \times Y \to I\!R$ are defined as the functions

$$h(x) := \inf_{y \in Y} F(x, y), \text{ for } x \in X, \text{ and } h^\nu(x) := \inf_{y \in Y^\nu} F^\nu(x, y), \text{ for } x \in X^\nu.$$

In addition to their overall interest, inf-projections of bifunctions are central to the study of optimality functions as clearly highlighted by (5). We start by recording a well-known condition for upper semicontinuity of inf-projections. We include the proof as it is short.

Proposition 4.2 (upper semicontinuity of inf-projection) *For a bifunction* $F : X \times Y \to I\!R$ *that has* $F(\cdot, y)$ *upper semicontinuous on* X *for all* $y \in Y$*, the corresponding inf-projection* $h(x) = \inf_Y F(x, \cdot), x \in X$*, is upper semicontinuous.*

Proof. Let $\{x^\nu\}_{\nu\in\mathbb{N}}$ be a sequence in X converging to $x \in X$. If $h(x)$ is finite, then for every $\varepsilon > 0$ there is a $y_\varepsilon \in Y$ such that $h(x) \geq F(x, y_\varepsilon) - \varepsilon$. Thus, $\limsup F(x^\nu, y_\varepsilon) \leq F(x, y_\varepsilon) \leq h(x) + \varepsilon$ and $\limsup h(x^\nu) \leq h(x) + \varepsilon$. If $h(x) = -\infty$, then for every $M < \infty$ there is a $y_M \in Y$ such that $F(x, y_M) < -M$. Since $\limsup F(x^\nu, y_M) \leq F(x, y_M) < -M$, we have $\limsup h(x^\nu) < -M$. Since ε and M are arbitrary, the conclusion follows. □

Applications of this proposition to Examples 1-3 establish the upper semicontinuity of the corresponding optimality functions.

Theorem 4.1 (containment of inf-projections) *Suppose that the bifunctions* $\{F^\nu : X^\nu \times Y^\nu \to \mathbb{R}\}_{\nu\in\mathbb{N}}$ *lop-converge to* $F : X \times Y \to \mathbb{R}$ *and* $-\infty < \inf_Y F(x, \cdot)$ *for some* $x \in X$. *Then, the inf-projections* $h^\nu : X^\nu \to [-\infty, \infty[$ *and* $h : X \to [-\infty, \infty[$ *satisfy* $\limsup \operatorname{hypo} h^\nu \subset \operatorname{hypo} h$.

Proof. Suppose that $(x, x_0) \in \limsup \operatorname{hypo} h^\nu$. Then there exists a sequence $\{(x^\nu, x_0^\nu)\}_{\nu\in N}$, with N an infinite subsequence of $\mathbb{N}$, $x^\nu \in X^\nu$, $h^\nu(x^\nu) \geq x_0^\nu$, $x^\nu \to^N x$, and $x_0^\nu \to^N x_0$. If $x \notin X$, then take $y \in Y$ and construct a sequence $y^\nu \to y$, with $y^\nu \in Y^\nu$, such that $F^\nu(x^\nu, y^\nu) \to^N -\infty$, which exists by Definition 4.1(i). However,

$$x_0^\nu \leq h^\nu(x^\nu) \leq F^\nu(x^\nu, y^\nu),\ \nu \in N,$$

imply a contradiction since $x_0^\nu \to^N x_0 \in \mathbb{R}$ and $F^\nu(x^\nu, y^\nu) \to^N -\infty$. Thus, $x \in X$. If $h(x) = -\infty$, then there exists $y \in Y$ such that $F(x, y) \leq x_0 - 1$. Definition 4.1(i) ensures that there exists a sequence $y^\nu \to y$, with $y^\nu \in Y^\nu$, such that $\limsup F^\nu(x^\nu, y^\nu) \leq F(x, y)$. Consequently, $x_0 = \limsup_{\nu\in N} x_0^\nu \leq$

$\mathrm{limsup}_{\nu\in N} h^\nu(x^\nu) \le \mathrm{limsup}_{\nu\in N} F^\nu(x^\nu, y^\nu) \le F(x,y) \le x_0 - 1$, which is a contradiction. Hence, it suffices to consider the case with $h(x)$ finite. Given any $\varepsilon > 0$ arbitrarily small, pick $y_\varepsilon \in Y$ such that $F(x, y_\varepsilon) - \varepsilon \le h(x)$. Then Definition 4.1(i) again yields $y^\nu \to y_\varepsilon$, with $y^\nu \in Y^\nu$, such that

$$\mathrm{limsup}_{\nu\in N} h^\nu(x^\nu) \le \mathrm{limsup}_{\nu\in N} F^\nu(x^\nu, y^\nu) \le F(x, y_\varepsilon) \le h(x) + \varepsilon,$$

implying $\mathrm{limsup}_{\nu\in N} h^\nu(x^\nu) \le h(x)$. Thus, the conclusion follows from $x_0 = \mathrm{limsup}_{\nu\in N} x_0^\nu \le \mathrm{limsup}_{\nu\in N} h^\nu(x^\nu) \le h(x)$. □

Additional results can be obtained under a strengthening of lopsided convergence analogous to tight epi-convergence.

Definition 4.2 (ancillary-tight lop-convergence) The lop-convergence of bifunctions $\{F^\nu : X^\nu \times Y^\nu \to I\!R\}_{\nu\in I\!N}$ to $F : X \times Y \to I\!R$ is ancillary-tight if and only if Definition 4.1 holds and for any $\varepsilon > 0$ one can find a compact set $B_\varepsilon \subset \mathcal{Y}$ and an integer ν_ε, depending possibly on the sequence $x^\nu \to x$ selected in Definition 4.1(ii), such that

$$\inf_{y\in Y^\nu\cap B_\varepsilon} F^\nu(x^\nu, y) \le \inf_{y\in Y^\nu} F^\nu(x^\nu, y) + \varepsilon \text{ for all } \nu \ge \nu_\varepsilon.$$

Under ancillary-tight lop-convergence, we can strengthen the conclusion of Theorem 4.1 as follows.

Theorem 4.2 (hypo-convergence of inf-projections) *Suppose that the bifunctions $\{F^\nu : X^\nu \times Y^\nu \to I\!R\}_{\nu\in I\!N}$ lop-converge ancillary-tightly to a bifunction $F : X \times Y \to I\!R$ and $-\infty < \inf_Y F(x, \cdot)$ for some $x \in X$. Then, the corresponding inf-projections $h^\nu : X^\nu \to [-\infty, \infty[$ hypo-converge to the inf-projection $h : X \to [-\infty, \infty[$, i.e.,* $\mathrm{hypo}\, h^\nu$ *set-converge to* $\mathrm{hypo}\, h$.

Proof. We follow nearly the same argument as in the proof of [9, Theorem 3.4]. Let $x \in X$ be such that $h(x)$ is finite. Now, choose $x^\nu \to x$, with $x^\nu \in X^\nu$, such that $F^\nu(x^\nu, \cdot)$ epi-converge to $F(x, \cdot)$, cf. Proposition 4.1. In fact, they epi-converge tightly as an immediate consequence of ancillary-tightness. Thus,

$$h^\nu(x^\nu) = \inf_{y \in Y^\nu} F^\nu(x^\nu, y^\nu) \to \inf_{y \in Y} F(x, y) = h(x),$$

via Theorem 3.2. In view of Theorem 4.1, the conclusion then follows. □

We recall that hypo h^ν set-converges to hypo h if and only if epi $-h^\nu$ set-converges to epi $-h$. Thus, Theorem 4.2 implies that Theorem 3.1 holds[4] with $f^\nu = -h^\nu$ and $f = -h$. This observation leads to the following corollary.

Corollary 4.1 *Suppose that the bifunctions* $\{F^\nu : X^\nu \times Y^\nu \to \mathbb{R}\}_{\nu \in \mathbb{N}}$ *lop-converge ancillary-tightly to* $F : X \times Y \to \mathbb{R}$ *and* $-\infty < \inf_Y F(x, \cdot)$ *for some* $x \in X$*. Then, the corresponding inf-projections* $h^\nu : X^\nu \to [-\infty, \infty[$ *and* $h : X \to [-\infty, \infty[$ *satisfy the following:*

(i) $\liminf(\sup_{X^\nu} h^\nu) \geq \sup_X h$.

(ii) If $x^k \in \operatorname{argmax}_{X^{\nu_k}} h^{\nu_k}$ *and* $x^k \to x$ *for some increasing subsequence* $\{\nu_1, \nu_2, ...\} \subset \mathbb{N}$*, then* $x \in \operatorname{argmax}_X h$ *and* $\lim_{k \to \infty} \sup_{X^{\nu_k}} h^{\nu_k} = \sup_X h$.

Further strengthening of the notion is also beneficial.

Definition 4.3 (tight lop-convergence) The lop-convergence of bifunctions $\{F^\nu : X^\nu \times Y^\nu \to \mathbb{R}\}_{\nu \in \mathbb{N}}$ to $F : X \times Y \to \mathbb{R}$ is tight if and only if Definition 4.2 holds and for any $\varepsilon > 0$ one can find a compact set $A_\varepsilon \subset \mathcal{X}$ and an

[4] We note that Theorem 3.1 is stated for finite-valued functions and h^ν and h might be extended-real valued. However, the conclusions hold under this slight extension.

integer ν_ε such that

$$\sup_{x\in X^\nu\cap A_\varepsilon}\inf_{y\in Y^\nu} F^\nu(x,y) \geq \sup_{x\in X^\nu}\inf_{y\in Y^\nu} F^\nu(x,y) - \varepsilon \text{ for all } \nu \geq \nu_\varepsilon.$$

Under tight lop-convergence, we can strengthen Theorem 4.2 as follows.

Theorem 4.3 (approximating maxinf-points) *Suppose that the bifunctions* $\{F^\nu : X^\nu \times Y^\nu \to I\!R\}_{\nu\in \mathbb{N}}$ *lop-converge tightly to* $F : X \times Y \to I\!R$ *and* $\sup_X \inf_Y F$ *is finite. Then,*

$$\sup_{x\in X^\nu}\inf_{y\in Y^\nu} F^\nu(x,y) \to \sup_{x\in X}\inf_{y\in Y} F(x,y).$$

Moreover, for every $x^* \in \operatorname{argmax}_{x\in X}\inf_{y\in Y} F(x,y)$*, there exist an infinite subsequence* N *of* $I\!N$*,* $\{\varepsilon^\nu\}_{\nu\in N}$*, with* $\varepsilon^\nu \searrow 0$*, and* $\{x^\nu\}_{\nu\in N}$*, with*

$$x^\nu \in \varepsilon^\nu\text{-}\operatorname{argmax}_{x\in X^\nu}\inf_{y\in Y^\nu} F^\nu(x,y),$$

such that $x^\nu \to^N x$*. Conversely, if such sequences exist, then we have that* $\sup_{x\in X^\nu}\inf_{y\in Y^\nu} F^\nu(x,y) \to^N \inf_{y\in Y} F(x^*,\cdot)$*.*

Proof. The result is a direct consequence of Theorem 4.2 in conjunction with tight lop-convergence and Theorem 3.2 reoriented to maximization. □

It is well-known that the supremum over a compact set of an upper semicontinuous function is attained. Consequently, in view of Proposition 4.2, if $F(\cdot,y)$ is upper semicontinuous on X for all $y \in Y$ and X is compact, then there exists a maxinf-point of F. We next state a result that relaxes the compactness requirement.

Theorem 4.4 (existence of maxinf-point) *Suppose* $\{F^\nu : X^\nu \times Y^\nu \to I\!R\}_{\nu\in \mathbb{N}}$ *lop-converges ancillary-tightly to* $F : X \times Y \to I\!R$*,* X^ν *is compact,* $F^\nu(\cdot, y)$ *is upper semicontinuous for all* $y \in Y^\nu$*, and* $-\infty < \inf_Y F(x, \cdot)$ *for some* $x \in X$*.*

Then, for all ν *there exists a maxinf-point* x^ν *of* $F^\nu : X^\nu \times Y^\nu \to I\!R$ *and every cluster point of* $\{x^\nu\}_{\nu\in \mathbb{N}}$ *is a maxinf-point of* $F : X \times Y \to I\!R$*.*

Proof. The discussion prior to the theorem ensures the existence of maxinf-points of $F^\nu : X^\nu \times Y^\nu \to I\!R$ for every ν. The result is then a consequence of Corollary 4.1. □

In view of Theorem 4.4, we see that the existence of a maxinf-point of $F : X \times Y \to I\!R$ is established through constructing $F^\nu : X^\nu \times Y^\nu \to I\!R$, with X^ν compact, that lop-converge ancillary-tightly to $F : X \times Y \to I\!R$ and that have a sequence of maxinf-points with a cluster point. The theorem does not guarantee the existence of such a cluster point, an additional condition needs to be brought in. Obviously, the simplest such condition is the containment of $\{X^\nu\}_{\nu\in \mathbb{N}}$ in a compact set. Still, the compactness of X is not required.

5 Applications and Further Examples

We now return to the context of optimality functions of the form (5) and start with the requirement for consistency in Definition 3.3(ii).

Proposition 5.1 (sufficient condition for consistency, optimality function part) *Suppose that* $\{F^\nu : X^\nu \times Y^\nu \to I\!R\}_{\nu\in \mathbb{N}}$ *lop-converges to* $F : X \times Y \to I\!R$ *and that the bifunctions define* $\theta^\nu = \inf_{y\in Y^\nu} F^\nu(\cdot, y)$ *and* $\theta = \inf_{y\in Y} F(\cdot, y)$*, with*

$-\infty < \theta(x)$ for some $x \in X$. Then,

$$\text{for every } x^\nu \to x \in \mathcal{X}, \text{ with } x^\nu \in X^\nu, \ \limsup \theta^\nu(x^\nu) \le \theta(x) \text{ if } x \in X, \text{ and}$$
$$\theta^\nu(x^\nu) \to -\infty \text{ otherwise.}$$

Proof. The result is a direct consequence of Theorem 4.1. □

In view of this result, it is clear that (weak) consistency will be ensured by epi-convergence of the approximating objective functions and feasible sets as well as lopsided convergence of the approximating bifunctions defining the corresponding optimality functions. We illustrate Proposition 5.1 by continuing from Example 2.3.

Example 5.1 (minimax problem) Continuing from Example 2.3, suppose that for every $z \in Z$ there exists a sequence $z^\nu \in Z^\nu$ such that $z^\nu \to z$. Let for $x, y \in \mathbb{R}^n$,

$$F^\nu(x,y) = \max_{z \in Z^\nu} \left\{ \varphi(x,z) - f^\nu(x) + \langle \nabla_x \varphi(x,z), y - x \rangle + \frac{1}{2}\|y - x\|^2 \right\}$$

and F be defined similarly with the superscripts removed. We next show lopsided convergence of F^ν to F. First consider Definition 4.1(i). Let $y \in \mathbb{R}^n$ and $x^\nu \to x \in \mathbb{R}^n$. Set $y^\nu = y$ for all ν. Clearly, $\limsup F^\nu(x^\nu, y^\nu) \le \limsup F(x^\nu, y) = F(x,y)$ by the continuity of F and part (i) holds. Second, we consider part (ii). Let $x \in \mathbb{R}^n$ and $y^\nu \to y \in \mathbb{R}^n$. Set $x^\nu = x$ for all ν. Let

$$z_x \in \operatorname{argmax}_{z \in Z} \left\{ \varphi(x,z) - f(x) + \langle \nabla_x \varphi(x,z), y - x \rangle + \frac{1}{2}\|y - x\|^2 \right\}.$$

Let $\varepsilon > 0$. By assumption on Z^ν and the continuity of $\varphi(x, \cdot)$ and $\nabla_x\varphi(x, \cdot)$, there exist $z^\nu \in Z^\nu$ and ν_0 such that $\varphi(x, z^\nu) - \varphi(x, z_x) > -\varepsilon$ and

$$\|\nabla_x\varphi(x, z^\nu) - \nabla_x\varphi(x, z_x)\| < \min\left\{\varepsilon, \frac{\varepsilon}{\|y - x\|}\right\}$$

for all $\nu \geq \nu_0$. Consequently, $\nu \geq \nu_0$,

$$\begin{aligned}
F^\nu(x^\nu, y^\nu) &= F^\nu(x, y^\nu) \\
&= \max_{z \in Z^\nu}\left\{\varphi(x, z) - f^\nu(x) + \langle\nabla_x\varphi(x, z), y^\nu - x\rangle + \frac{1}{2}\|y^\nu - x\|^2\right\} \\
&\geq \varphi(x, z^\nu) - f(x) + \langle\nabla_x\varphi(x, z^\nu), y^\nu - x\rangle + \frac{1}{2}\|y^\nu - x\|^2 \\
&= \varphi(x, z_x) - f(x) + \langle\nabla_x\varphi(x, z_x), y - x\rangle + \frac{1}{2}\|y - x\|^2 + \varphi(x, z^\nu) \\
&\quad - \varphi(x, z_x) + \langle\nabla_x\varphi(x, z^\nu) - \nabla_x\varphi(x, z_x), y - x\rangle \\
&\quad + \langle\nabla_x\varphi(x, z^\nu), y^\nu - y\rangle + \frac{1}{2}\|y^\nu - x\|^2 - \frac{1}{2}\|y - x\|^2 \\
&> \varphi(x, z_x) - f(x) + \langle\nabla_x\varphi(x, z_x), y - x\rangle + \frac{1}{2}\|y - x\|^2 \\
&\quad - \varepsilon - \varepsilon + \langle\nabla_x\varphi(x, z^\nu), y^\nu - y\rangle + \frac{1}{2}\|y^\nu - x\|^2 - \frac{1}{2}\|y - x\|^2 \\
&= F(x, y) - 2\varepsilon + \langle\nabla_x\varphi(x, z^\nu), y^\nu - y\rangle + \frac{1}{2}\|y^\nu - x\|^2 - \frac{1}{2}\|y - x\|^2
\end{aligned}$$

Since $y^\nu \to y$, $\{z^\nu\}$ is bounded and $\nabla_x\varphi$ is continuous, $\liminf F^\nu(x^\nu, y^\nu) \geq F(x, y) - 2\varepsilon$. Since ε was arbitrary, part (ii) of Definition 4.1 holds and F^ν therefore lop-converge to F. In view of Proposition 5.1 and the fact that epi-convergence is also easily established, we have that the pairs $\{(f^\nu : \mathbb{R}^n \to \mathbb{R}, \theta^\nu : \mathbb{R}^n \to \mathbb{R}_-)\}$ are consistent approximations of the pair $\{(f : \mathbb{R}^n \to \mathbb{R}, \theta : \mathbb{R}^n \to \mathbb{R}_-)\}$ in this case. The above algorithm therefore is implementable for the solution of the semi-infinite minimax problem $\min_{x \in \mathbb{R}^n} \max_{z \in Z} \varphi(x, z)$.

Under slightly stronger assumptions, the approximating bifunctions do not need to be associated with an optimality function to achieve convergence to quasi-stationary points.

Theorem 5.1 (convergence to quasi-stationary points) *Suppose that the bifunctions* $\{F^\nu : X^\nu \times Y^\nu \to \mathbb{R}\}_{\nu \in N}$ *lop-converge ancillary-tightly to the bifunction* $F : X \times Y \to \mathbb{R}$ *and* $\theta : X \to \mathbb{R}_-$, *with* $\theta = \inf_{y \in Y} F(\cdot, y)$, *and* $\theta(x) > -\infty$ *for some* $x \in X$. *Then,* θ *is upper semicontinuous.*

Moreover, if $x^\nu \in \operatorname{argmax}_{x \in X^\nu} \inf_{y \in Y^\nu} F^\nu(x, y)$ *for all* ν *and* $\mathcal{Q}_\theta \neq \emptyset$, *then every cluster point* x *of* $\{x^\nu\}_{\nu \in N}$ *is quasi-stationary, i.e.,* $x \in \mathcal{Q}_\theta$.

Proof. In view of Theorem 4.2, the inf-projections of $F^\nu : X^\nu \times X^\nu \to \mathbb{R}$ hypo-converge to θ. This implies that θ is upper semicontinuous since set limits (of hypo-graphs) are necessarily closed. By Corollary 4.1, $x \in \operatorname{argmax}_X \theta$. Since $\mathcal{Q}_\theta \neq \emptyset$, $\mathcal{Q}_\theta = \operatorname{argmax}_X \theta$ and the conclusion follows. □

Further characterization of (quasi-)stationary points is available under tight lopsided convergence.

Theorem 5.2 (characterization of quasi-stationary points) *Suppose that the bifunctions* $\{F^\nu : X^\nu \times Y^\nu \to \mathbb{R}\}_{\nu \in N}$ *lop-converge tightly to the bifunction* $F : X \times Y \to \mathbb{R}$ *and* $\theta = \inf_{y \in Y} F(\cdot, y)$, *with* $\mathcal{Q}_\theta \neq \emptyset$. *For every* $x \in \mathcal{Q}_\theta$ *there exist an infinite subsequence* N *of* $\mathbb{N}$, $\{\varepsilon^\nu\}_{\nu \in N}$, *with* $\varepsilon^\nu \searrow 0$, *and* $\{x^\nu\}_{\nu \in N}$, *with* $x^\nu \in \varepsilon^\nu\text{-}\operatorname{argmax}_{x \in X^\nu} \inf_{y \in Y^\nu} F^\nu(x, y)$, *such that* $x^\nu \to^N x$.

Proof. The result is a direct consequence of Theorem 4.3. □

The next result establishes a pathway to show the existence of a quasi-stationary point, i.e., $\mathcal{Q}_\theta \neq \emptyset$. We note that the scope is reduced to linear spaces and $X = Y$ to facilitate the application of a Ky Fan Inequality.

Theorem 5.3 (existence of quasi-stationary point) *Let $\mathcal{X}$ be a linear space and $\theta : X \to \mathbb{R}_-$, with $X \subset \mathcal{X}$, be defined by $\theta = \inf_{y\in X} F(\cdot, y)$ for a bifunction $F : X \times X \to \mathbb{R}$ and $\theta(x) > -\infty$ for some $x \in X$. Suppose that there exist bifunctions $\{F^\nu : X^\nu \times X^\nu \to \mathbb{R}\}_{\nu\in \mathbb{N}}$ that lop-converge to $F : X \times X \to \mathbb{R}$, with (i) $X^\nu \subset \mathcal{X}$ convex and compact, (ii) $F^\nu(\cdot, y)$ upper semicontinuous for all $y \in X^\nu$, (iii) $F^\nu(x, \cdot)$ convex for all $x \in X^\nu$, and (iv) $F^\nu(y, y) \geq 0$ for all $y \in X^\nu$.*

Then, for all ν there exists $x^\nu \in X^\nu$ with $\inf_{X^\nu} F^\nu(x^\nu, \cdot) \geq 0$. Moreover, every cluster point $\bar{x}$ of $\{x^\nu\}_{\nu\in \mathbb{N}}$ satisfies $\theta(\bar{x}) = 0$, i.e., $\mathcal{Q}_\theta \neq \emptyset$.

Proof. We invoke the Ky Fan Inequality as applied to $F^\nu : X^\nu \times X^\nu \to \mathbb{R}$, which establishes that there exists $x^\nu \in X^\nu$ such that $\inf_{X^\nu} F^\nu(x^\nu, \cdot) \geq 0$; see [7]. Let $\bar{x}$ be a cluster point of $\{x^\nu\}_{\nu\in \mathbb{N}}$. Then, in view of Proposition 5.1 we find that $\bar{x} \in X$ and $\limsup \inf_{X^\nu} F^\nu(x^\nu, \cdot) \leq \theta(\bar{x})$. Since the left-hand side is nonnegative and $\theta(x) \leq 0$ for all $x \in X$, the conclusion follows. □

We stress that the theorem does not guarantee the existence of a cluster point of $\{x^\nu\}_{\nu\in \mathbb{N}}$. Of course, the containment of $\{x^\nu\}_{\nu\in \mathbb{N}}$ in a compact set would suffice, but other application dependent conditions might also be used for the purpose. Thus, the theorem provides a way of establishing existence of a quasi-stationary point without insisting on the compactness of X.

We end the paper with an example from the area of optimal control and adjust the notation accordingly.

Example 5.2 (optimal control) We here follow the set-up in Section 5.6 and Chapter 4 of [16], which contain further details. For $g : I\!R^n \times I\!R^m \to I\!R^n$, we consider the dynamical system

$$\dot{x}(t) = g(x(t), u(t)), \text{ for } t \in [0,1], \text{ with } x(0) = \xi \in I\!R^n,$$

where the control $u \in \mathcal{L}_\infty^m := \{u : [0,1] \to I\!R^m \ : \ \text{measurable, ess. bounded}\}$. Since such controls are contained in the space of square-integrable functions from $[0,1]$ to $I\!R^m$, the usual L_2-norm applies; see [16, p.709] for a motivation for this "hybrid" set-up. Let $\mathcal{H} := I\!R^n \times \mathcal{L}_\infty^m$. For initial condition and control pairs $\eta = (\xi, u) \in \mathcal{H}$ and $\bar{\eta} = (\bar{\xi}, \bar{u}) \in \mathcal{H}$, we equip $\mathcal{H}$ with the inner product and norm

$$\langle \eta, \bar{\eta} \rangle_{\mathcal{H}} := \langle \xi, \bar{\xi} \rangle + \int_0^1 \langle u(t), \bar{u}(t) \rangle dt \ \text{ and } \ \|\eta\|_{\mathcal{H}}^2 := \langle \eta, \eta \rangle_{\mathcal{H}}.$$

We consider control constraints of the form $u(t) \in C$, for almost every $t \in [0,1]$ for some given convex and compact set $C \subset I\!R^m$. By imposing the constraints for almost every t instead of every t, we deviate slightly from [16] and follow [13]. We therefore also define the feasible set

$$U = \mathcal{L}_\infty^m \cap \{u \ : \ u(t) \in C, \text{ for almost every } t \in [0,1]\} \text{ and } H = I\!R^n \times U.$$

Under standard assumptions, a solution of the differential equation, for a given $\eta \in H$, denoted by x_η is unique, Lipschitz continuous, and Gâteaux differentiable in η. Consequently, for a given $\varphi : I\!R^n \times I\!R^n \to I\!R$, Lipschitz continuously differentiable on bounded sets, the function $f : H \to I\!R$ defined by

$$f(\eta) = \varphi(\xi, x_\eta(1)), \text{ for } \eta = (\xi, u) \in H,$$

has a Gâteaux differential of the form $\langle \nabla f(\eta), \bar{\eta} - \eta \rangle_{\mathcal{H}}$ for some Lipschitz continuous gradient $\nabla f(\eta)$ given in [16, Corollary 5.6.9]. The optimal control problem

$$\text{minimize } f(\eta) \text{ subject to } \eta \in H,$$

analogous to Example 2.1, has an optimality function

$$\theta(\eta) = \min_{\bar{\eta} \in H} F(\eta, \bar{\eta}), \text{ for } \eta \in H,$$

where

$$F(\eta, \bar{\eta}) = \langle \nabla f(\eta), \bar{\eta} - \eta \rangle_{\mathcal{H}} + \frac{1}{2} \|\bar{\eta} - \eta\|_{\mathcal{H}}^2, \text{ for } \eta, \bar{\eta} \in H.$$

We next consider approximations. Let $U^\nu \subset U$, $\nu \in I\!N$, consist of the piecewise constant functions that are constant on each interval $[(k-1)/\nu, k/\nu)$, $k = 1, ..., \nu$. Set $H^\nu = I\!R^n \times U^\nu$. Moreover, let x_η^ν be the (unique) solution of the forward Euler approximation of the differential equation, using time-step $1/\nu$, given input $\eta = (\xi, u) \in \mathcal{H}$. An approximate problem then takes the form

$$\text{minimize } f^\nu(\eta) \text{ subject to } \eta \in H^\nu,$$

where

$$f^\nu(\eta) = \varphi(\xi, x_\eta^\nu(1)).$$

One can show that

$$\theta^\nu(\eta) = \min_{\bar{\eta} \in H^\nu} F^\nu(\eta, \bar{\eta}), \text{ for } \eta \in H^\nu,$$

where

$$F^\nu(\eta,\bar\eta) = \langle \nabla f^\nu(\eta), \bar\eta - \eta\rangle_{\mathcal{H}} + \frac{1}{2}\|\bar\eta - \eta\|^2_{\mathcal{H}}, \text{ for } \eta, \bar\eta \in H^\nu,$$

is an optimality function of $f^\nu : H^\nu \to I\!R$, where the Lipschitz continuous gradient $\nabla f^\nu(\eta)$ is given in [16, Theorem 5.6.19].

By [16, Theorem 4.3.2], for every bounded set $S \subset H$, there exists a $C_S < \infty$ such that $|f(\eta) - f^\nu(\eta)| \le C_S/\nu$ and $\|\nabla f(\eta) - \nabla f^\nu(\eta)\|_{\mathcal{H}} \le C_S/\nu$ for all $\eta \in S$. Moreover, $\cup_{\nu\in I\!N} H^\nu$ is dense in H. Consequently, it is easily established that $f^\nu : H^\nu \to I\!R$ epi-converge to $f : H \to I\!R$. We next consider the optimality functions. Let $\bar\eta \in H$ and $\eta^\nu \to \eta \in \mathcal{H}$, with $\eta^\nu \in H^\nu$. Necessarily, $\eta \in H$. Due to the density result, there exists $\bar\eta^\nu \to \bar\eta$, with $\bar\eta^\nu \in H^\nu$. Hence,

$$\begin{aligned}|F^\nu(\eta^\nu,\bar\eta^\nu) - F(\eta,\bar\eta)| &\le \|\nabla f^\nu(\eta^\nu) - \nabla f(\eta)\|_{\mathcal{H}}\|\bar\eta^\nu - \eta^\nu\|_{\mathcal{H}} \\ &+ \|\nabla f(\eta)\|_{\mathcal{H}}\|\bar\eta^\nu - \eta^\nu - \bar\eta + \eta\|_{\mathcal{H}} + \frac{1}{2}\|\bar\eta^\nu - \eta^\nu\|^2_{\mathcal{H}} - \frac{1}{2}\|\bar\eta^\nu - \eta^\nu\|^2_{\mathcal{H}} \to 0\end{aligned}$$

and we have shown Definition 4.1(i). Using similar arguments, we also establish part (ii) and the lopsided convergence of F^ν to F. Consequently, $\{(f^\nu : H^\nu \to I\!R,\ \theta^\nu : H^\nu \to I\!R_-)\}_{\nu\in I\!N}$ are consistent approximations of $(f : H \to I\!R,\ \theta : H \to I\!R_-)$. Since the minimization of $f^\nu : H^\nu \to I\!R$ is equivalent to an optimization problem on a Euclidean space, the above algorithm is implementable for the infinite-dimensional problem $f : H \to I\!R$.

6 Conclusions

We have shown that the lopsided convergence of a certain class of bifunctions provides a general pathway for constructing implementable algorithms for optimization problems requiring approximations. The bifunctions, with their inf-projections called optimality functions, quantifies near-stationarity and therefore convergence of the algorithm to stationary points can be guaranteed. A series of examples from constrained optimization, nonlinear programming, minimax problems, and optimal control illustrate the framework.

Acknowledgements This material is based upon work supported in part by the U. S. Army Research Laboratory and the U. S. Army Research Office under grant numbers 00101-80683, W911NF-10-1-0246 and W911NF-12-1-0273.

References

1. Attouch, H.: Variational Convergence for Functions and Operators. Applicable Mathematics Sciences. Pitman (1984)
2. Attouch, H., Wets, R.J.B.: Convergence des points min/sup et de points fixes. Comptes Rendus de l'Académie des Sciences de Paris **296**, 657–660 (1983)
3. Attouch, H., Wets, R.J.B.: A convergence theory for saddle functions. Transactions of the American Mathematical Society **280**(1), 1–41 (1983)
4. Aubin, J.P., Frankowska, H.: Set-Valued Analysis. Birkhäuser (1990)
5. Bagh, A.: Epi/hypo-convergence: The slice topology and saddle points approximation. J. Applied Analysis **4**, 13–39 (1996)
6. Beer, G.: Topologies on Closed and Closed Convex Sets, *Mathematics and its Applications*, vol. 268. Kluwer (1992)
7. Fan, K.: A minimax inequality and applications. In: O. Shisha (ed.) Inequalities - III, pp. 103–113. Academic Press (1972)
8. Foraker, J., Royset, J., Kaminer, I.: Search-trajectory optimization: Part 1, formulation and theory. J. Optimization Theory and Application (2015)
9. Jofre, A., Wets, R.J.: Variational convergence of bivariate functions: Lopsided convergence. Mathematical Programming B **116**, 275–295 (2009)

10. Jofre, A., Wets, R.J.: Variational convergence of bifunctions: Motivating applications. SIAM J. Optimization **24**(4), 1952–1979 (2014)
11. Pang, J.S.: Error bounds in mathematical programming. Mathematical Programming **79**, 299–332 (1997)
12. Pasupathy, R.: On choosing parameters in retrospective-approximation algorithms for stochastic root finding and simulation optimization. Operations Research **58**(4), 889–901 (2010)
13. Phelps, C., Royset, J., Gong, Q.: Optimal control of uncertain systems using sample average approximations. SIAM J. Control and Optimization (2015)
14. Polak, E.: Computational Methods in Optimization. A Unified Approach. Academic Press (1971)
15. Polak, E.: On the use of models in the synthesis of optimization algorithms. In: H. Kuhn, G. Szego (eds.) Differential Games and Related Topics, pp. 263–279. North Holland, Amsterdam (1971)
16. Polak, E.: Optimization. Algorithms and Consistent Approximations, *Applied Mathematical Sciences*, vol. 124. Springer (1997)
17. Rockafellar, R., Wets, R.: Variational Analysis, *Grundlehren der Mathematischen Wissenschaft*, vol. 317, 3rd printing-2009 edn. Springer (1998)
18. Royset, J., Pee, E.: Rate of convergence analysis of discretization and smoothing algorithms for semi-infinite minimax problems. Journal of Optimization Theory and Applications **155**(3), 855–882 (2012)
19. Royset, J.O.: Optimality functions in stochastic programming. Mathematical Programming **135**(1), 293–321 (2012)
20. Royset, J.O.: On sample size control in sample average approximations for solving smooth stochastic programs. Computational Optimization and Applications **55**(2), 265–309 (2013)
21. Royset, J.O., Szechtman, R.: Optimal budget allocation for sample average approximation. Operations Research **61**, 762–776 (2013)
22. Zangwill, W.: Nonlinear Programming: A Unified Approach. Prentice-Hall (1969)