

Optimality Functions and Lopsided Convergence*

Johannes O. Royset

Roger J-B Wets

Operations Research Department
Naval Postgraduate School
joroyset@nps.edu

Department of Mathematics
University of California, Davis
rjbwets@ucdavis.edu

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 apply it to nonlinear programming. This results in a convergence result for a primal interior point method without constraint qualifications or convexity assumptions. Moreover, we introduce lopsided convergence of bifunctions on metric spaces and show that 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.

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.

Keywords: epi-convergence, lopsided convergence, consistent approximations, optimality functions, optimality conditions, interior point method

AMS Classification:

Date: March 16, 2015

*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.

Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE 16 MAR 2015		2. REPORT TYPE		3. DATES COVERED 00-00-2015 to 00-00-2015	
4. TITLE AND SUBTITLE Optimality Functions and Lopsided Convergence				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School (NPS), Operations Research Department, Monterey, CA, 93943				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES					
14. 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 apply it to nonlinear programming. This results in a convergence result for a primal interior point method without constraint qualifications or convexity assumptions. Moreover we introduce lopsided convergence of bifunctions on metric spaces and show that 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.					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT Same as Report (SAR)	18. NUMBER OF PAGES 17	19a. NAME OF RESPONSIBLE PERSON
a REPORT unclassified	b ABSTRACT unclassified	c THIS PAGE unclassified			

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 [12] for numerous examples in nonlinear programming, semiinfinite optimization, and optimal control as well as [14, 17, 5, 9] for recent applications in stochastic and semiinfinite 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. The framework of optimality functions, as laid out in [12, 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 [17, 16, 15, 8]) 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[†].

In this paper, we review the notion of optimality functions and illustrate the vast number of possibilities through several examples. In a novel 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 [6, 7] 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. Moreover, we provide characterizations of stationary points under perturbations and approximations using lopsided convergence. In the process, we extend the primary definitions and results on lopsided convergence in [6, 7] from finite dimensions to any metric space.

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

[†]The distinction between implementable and conceptual algorithms appears to be due to E. Polak [11, 10].

2 Optimality Functions: Definitions and Examples

We consider optimization problems defined on a metric space $(\mathcal{X}, \rho_{\mathcal{X}})$, where $C \subset \mathcal{X}$ is a nonempty feasible set and $f : C \rightarrow \mathbb{R}$ 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. Therefore, the notation $f : C \rightarrow \mathbb{R}$ 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

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

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

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

2.1 Definition (optimality function) *An upper semicontinuous function $\theta : X \rightarrow \mathbb{R}_-$ is an optimality function for $f : C \rightarrow \mathbb{R}$ if $C \subset X \subset \mathcal{X}$ and*

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

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

A series of examples help illustrate the concept; see also §5 and [12, 14, 17, 5, 9].

Example 1: Constrained Optimization over Convex Set. Consider the case $\mathcal{X} = \mathbb{R}^n$, $C \subset \mathcal{X}$ closed and convex, and $f : C \rightarrow \mathbb{R}$ continuously differentiable. Then, the function

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

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

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

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

Example 2: Nonlinear Programming. Consider the case $\mathcal{X} = \mathbb{R}^n$, $C = \{x \in \mathbb{R}^n \mid f_j(x) \leq 0, j = 1, \dots, q\}$, and $f, f_1, \dots, f_q$ real-valued and continuously differentiable on $\mathbb{R}^n$. Let $\psi(x) = \max_{j=1, \dots, q} f_j(x)$

and $\psi_+(x) = \max\{0, \psi(x)\}$. Then, the function

$$\theta(x) = \min_{y \in \mathbb{R}^n} \max \left\{ -\psi_+(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2} \|y - x\|^2, \right. \\ \left. \max_{j=1, \dots, q} \{f_j(x) - \psi(x)_+ + \langle \nabla f_j(x), y - x \rangle\} + \frac{1}{2} \|y - x\|^2 \right\}, \quad x \in X = \mathbb{R}^n, \quad (2)$$

satisfies the requirements of Definition 2.1 and is therefore an optimality function for $f : C \rightarrow \mathbb{R}$. 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 [12, Theorem 2.2.8] for proofs and further discussion.

Example 3: Minimax Problem. Consider the case $\mathcal{X} = C = \mathbb{R}^n$ and $f(x) = \max_{z \in Z} \varphi(x, z)$, $x \in \mathbb{R}^n$, where $\varphi : \mathbb{R}^n \times \mathbb{R}^p \rightarrow \mathbb{R}$ is continuous, the gradient $\nabla_x \varphi : \mathbb{R}^n \times \mathbb{R}^p \rightarrow \mathbb{R}^n$ with respect to the first argument exists and is continuous in both arguments, and Z is a compact subset of $\mathbb{R}^p$. Then, the function

$$\theta(x) = \min_{y \in \mathbb{R}^n} \max_{z \in Z} \left\{ \varphi(x, z) - f(x) + \langle \nabla_x \varphi(x, z), y - x \rangle + \frac{1}{2} \|y - x\|^2 \right\}, \quad x \in X = \mathbb{R}^n, \quad (3)$$

satisfies the requirements of Definition 2.1 and is therefore an optimality function for $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Moreover, $\theta(x) = 0$ if and only if $0 \in \partial f(x)$ (the subdifferential of f); see [12, Theorem 3.1.6] for details.

We note that the upper semicontinuity of optimality functions ensures the computationally significant property that if a sequence $x^\nu \rightarrow x$ 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 [12] for details.

3 Approximations and Implementable Algorithms

Problems involving functions defined in terms of integrals or optimization problems (as the maximization in Example 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, $\nu \in \mathbb{N} = \{1, 2, \dots\}$, with usually $\cup_{\nu \in \mathbb{N}} \mathcal{X}^\nu$ 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 $f^\nu : C^\nu \rightarrow \mathbb{R}$ could be a tractable approximation of $f : C \rightarrow \mathbb{R}$. An example helps illustrate the situation.

Example 3: Minimax Problem (cont.). Suppose that $f^\nu(x) = \max_{z \in Z^\nu} \varphi(x, z)$, $x \in \mathbb{R}^n$, with $Z^\nu \subset Z$ consisting 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 $f^\nu : \mathbb{R}^n \rightarrow \mathbb{R}$ can be associated with the optimality function

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

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 [12, Theorem 2.1.6].

We next examine approximating functions $f^\nu : C^\nu \rightarrow \mathbb{R}$ 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 [12, 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 $f : C \rightarrow \mathbb{R}$, 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, as defined next, is the key property when examining approximations of optimization problems; see [1, 3, 13] for more comprehensive treatments.

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

- (i) *for every sequence $x^\nu \rightarrow x \in \mathcal{X}$, with $x^\nu \in C^\nu$, we have that $\liminf f^\nu(x^\nu) \geq f(x)$ if $x \in C$ and $f^\nu(x^\nu) \rightarrow \infty$ 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 $x^\nu \rightarrow x$ and $\limsup f^\nu(x^\nu) \leq f(x)$.*

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

3.2 Theorem (convergence of minimizers) *Suppose that the functions $\{f^\nu : C^\nu \rightarrow \mathbb{R}\}_{\nu \in \mathbb{N}}$ epi-converge to $f : C \rightarrow \mathbb{R}$. Then,*

$$\limsup (\inf_{C^\nu} f^\nu) \leq \inf_C f.$$

Moreover, if $x^k \in \arg\max_{C^{\nu_k}} f^{\nu_k}$ and $x^k \rightarrow x$ for some increasing subsequence $\{\nu_1, \nu_2, \dots\} \subset \mathbb{N}$, then $x \in \arg\max_C f$ and $\lim_{k \rightarrow \infty} \inf_{C^{\nu_k}} f^{\nu_k} = \inf_C f$.

Proof. The second part is essentially in [2, Theorem 2.5], but not for the finite-valued setting. The first and second parts are in [6, Theorem 2.6] for the $\mathbb{R}^n$ case. The proof carries over essentially verbatim. $\square$

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

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

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

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

- (i) *if and only if $\inf_{C^\nu} f^\nu \rightarrow \inf_C f$.*
- (ii) *if and only if there exists a sequence $\varepsilon^\nu \searrow 0$ such that ε^ν -argmin f^ν set-converges[‡] to argmin f .*

Proof. Again, the proof in [6, Theorem 2.8] can be immediately translated to the present setting. $\square$

3.2 Consistent Approximations

The convergence of optimal points as stipulated above 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 of consistent approximations is an extension from that in [12, Section 3.3] as we consider arbitrary metric spaces and not only normed linear spaces.

3.5 Definition (consistent approximations) *The pairs $\{(f^\nu : C^\nu \rightarrow \mathbb{R}, \theta^\nu : X^\nu \rightarrow \mathbb{R}_-)\}_{\nu \in N}$ of functions and corresponding optimality functions are weakly consistent approximations of the function and optimality-function pair $(f : C \rightarrow \mathbb{R}, \theta : X \rightarrow \mathbb{R}_-)$ if*

- (i) *$\{f^\nu : C^\nu \rightarrow \mathbb{R}\}_{\nu \in N}$ epi-converge to $f : C \rightarrow \mathbb{R}$ and*
- (ii) *for every $x^\nu \rightarrow x \in \mathcal{X}$, with $x^\nu \in X^\nu$, $\limsup \theta^\nu(x^\nu) \leq \theta(x)$ if $x \in X$, and $\theta^\nu(x^\nu) \rightarrow -\infty$ otherwise.*

If in addition $\theta^\nu(x) < 0$ for all $x \in X^\nu \setminus C^\nu$ and ν , then $\{(f^\nu : C^\nu \rightarrow \mathbb{R}, \theta^\nu : X^\nu \rightarrow \mathbb{R}_-)\}_{\nu \in N}$ are consistent approximations of $(f : C \rightarrow \mathbb{R}, \theta : X \rightarrow \mathbb{R}_-)$.

We recall that the epigraph of $f : C \rightarrow \mathbb{R}$ is defined by

$$\text{epi } f = \{(x, x_0) \in \mathcal{X} \times \mathbb{R} \mid x \in C, f(x) \leq x_0\}.$$

[‡]We recall that the outer limit of a sequence of sets $\{A^\nu\}_{\nu \in N}$, denoted by $\limsup A^\nu$, is the collection of points y to which a subsequence of $\{y^\nu\}_{\nu \in N}$, with $y^\nu \in A^\nu$, converges. The inner limit, denoted by $\liminf A^\nu$, is the points to which a sequence of $\{y^\nu\}_{\nu \in N}$, 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 $\{A^\nu\}_{\nu \in N}$ and that A^ν set-converges to this set; see [4, 13].

Since epi-convergence is equivalent to the set-convergence[§] of the corresponding epigraphs, we have that Definition 3.5(i) amounts to

$$\text{epi } f^\nu \text{ set-converges to } f.$$

Similarly, the hypograph of $f : C \rightarrow \mathbb{R}$ is defined by

$$\text{hypo } f = \{(x, x_0) \in \mathcal{X} \times \mathbb{R} \mid x \in C, f(x) \geq x_0\}.$$

In view of the definition of set-convergence, we therefore have that Definition 3.5(ii) amounts to

$$\limsup \text{hypo } \theta^\nu \subset \text{hypo } \theta.$$

The additional condition in Definition 3.5 removing “weakly” can be viewed as a constraint qualification as it eliminates the possibility of quasi-stationary points that are not stationary point for $f^\nu : C^\nu \rightarrow \mathbb{R}$, 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.

3.6 Theorem (convergence of stationary points) *Suppose that the pairs $\{(f^\nu : C^\nu \rightarrow \mathbb{R}, \theta^\nu : X^\nu \rightarrow \mathbb{R}_-)\}_{\nu \in N}$ are weakly consistent approximations of $(f : C \rightarrow \mathbb{R}, \theta : X \rightarrow \mathbb{R}_-)$ and $\{x^\nu\}_{\nu \in N}$, $x^\nu \in X^\nu$, is a sequence satisfying*

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

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 $x \in C$, i.e., $x \in \mathcal{S}_{C,\theta}$.

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

3.3 Algorithms

Theorem 3.6 provides a direct path to the construction of an implementable algorithm for minimizing $f : C \rightarrow \mathbb{R}$. Specifically, construct a family of approximations $\{f^\nu : C^\nu \rightarrow \mathbb{R}\}$ and corresponding optimality functions $\{\theta^\nu : X^\nu \rightarrow \mathbb{R}_-\}$, and then implement the following algorithm.

Algorithm.

1. Set $\varepsilon^\nu \geq 0$, with $\varepsilon^\nu \rightarrow 0$, and $\nu = 1$.
2. Obtain an approximate (quasi-)stationary point x^ν for $f^\nu : C^\nu \rightarrow \mathbb{R}$ that satisfies $\theta^\nu(x^\nu) \geq -\varepsilon^\nu$.

[§]Here, we consider set-convergence of subsets of $\mathcal{X} \times \mathbb{R}$, which is equipped with the metric $\rho((x, x_0), (x', x'_0)) = \max\{\rho_{\mathcal{X}}(x, x'), |x_0 - x'_0|\}$ for $x, x' \in \mathcal{X}$ and $x_0, x'_0 \in \mathbb{R}$.

3. Replace ν by $\nu + 1$ and go to Step 2.

If the pairs $\{(f^\nu : C^\nu \rightarrow \mathbb{R}, \theta^\nu : X^\nu \rightarrow \mathbb{R}_-)\}_{\nu \in \mathbb{N}}$ are weakly consistent approximations of $(f : C \rightarrow \mathbb{R}, \theta : X \rightarrow \mathbb{R}_-)$, then every cluster point of the constructed sequence $\{x^\nu\}$ will be quasi-stationary for $f : C \rightarrow \mathbb{R}$ by Theorem 3.6. The algorithm is fully implementable under the practically reasonable assumption that one can obtain an approximate quasi-stationary point of $f^\nu : C^\nu \rightarrow \mathbb{R}$ in finite time.

Example 2: Nonlinear Programming (cont.). Consider the standard logarithmic barrier approximation

$$f^\nu(x) = f(x) - t^\nu \sum_{j=1}^q \log[-f_j(x)], \quad x \in C^\nu = \{x \in \mathbb{R}^n \mid f_j(x) < 0, j = 1, \dots, q\},$$

where $t^\nu \searrow 0$. We first establish epi-convergence of $f^\nu : C^\nu \rightarrow \mathbb{R}$ to $f : C \rightarrow \mathbb{R}$. Suppose that $x^\nu \rightarrow x$, 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 $\log[-f_j(x)] \geq 0$ and $\nu \geq \nu_\varepsilon$. Hence,

$$f^\nu(x^\nu) \geq f(x^\nu) - \varepsilon \text{ for all } \nu \geq \nu_\varepsilon.$$

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 $t^\nu \sum_{j=1}^q \log[-f_j(x^\nu)] \rightarrow 0$. Consequently, $f^\nu(x^\nu) \rightarrow f(x)$, which satisfies Definition 3.1(ii). Thus, $f^\nu : C^\nu \rightarrow \mathbb{R}$ epi-converge to $f : C \rightarrow \mathbb{R}$. We next analyze optimality functions. Using a minmax theorem, one can show that (2) is equivalently expressed as

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

where $M = \{(\mu_0, \mu_1, \dots, \mu_q) \mid \mu_j \geq 0, j = 0, 1, \dots, q, \sum_{j=0}^q \mu_j = 1\}$; see [12, 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

$$m_j^\nu(x) = \frac{-t^\nu}{f_j(x)}.$$

Suppose that $x^\nu \rightarrow x \in \mathbb{R}^n$, with $x^\nu \in C^\nu$. Since $x^\nu \in C^\nu \subset C$ and C is closed, $x \in C$. Let

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

Consequently, $\mu^\nu = (\mu_0^\nu, \mu_1^\nu, \dots, \mu_q^\nu) \in M$ for all ν . Since M is compact, $\{\mu^\nu\}$ has at least one convergent subsequence. Suppose that $\mu^\nu \rightarrow^N \mu^\infty$, with N an infinite subsequence of $\mathbb{N}$. If j is such that $f_j(x) < 0$,

then $\mu_j^\nu \rightarrow^N 0$ and consequently $\mu_j^\infty = 0$ necessarily. In view of the continuity of the gradients, we then have that

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

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

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

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 $\limsup \theta^\nu(x^\nu) \leq \theta(x)$. This establishes the consistency of $\{f^\nu : C^\nu \rightarrow \mathbb{R}, \theta^\nu : C^\nu \rightarrow \mathbb{R}_-\}$. Consequently, the above algorithm, which can then be viewed as a primal interior point method, generates cluster points that are stationary for $f : C \rightarrow \mathbb{R}$ 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 associate 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

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

for some metric space $(\mathcal{Y}, \rho_Y)$ and function F . In fact, in our examples, $\mathcal{Y} = \mathbb{R}^n$ 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 $F : X \times Y \rightarrow \mathbb{R}$ will provide direct insight about (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. In the process, we extend some of the results in [6, 7] to general metric spaces.

Suppose that $(\mathcal{X}, \rho_{\mathcal{X}})$ and $(\mathcal{Y}, \rho_{\mathcal{Y}})$ are metric spaces, $X \subset \mathcal{X}$ and $Y \subset \mathcal{Y}$ are nonempty, and $F : X \times Y \rightarrow \mathbb{R}$ is a bifunction. We say that x^* is a maxinf-point of F 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 as defined next.

4.1 Definition (lopsided convergence) *The bifunctions $\{F^\nu : X^\nu \times Y^\nu \rightarrow \mathbb{R}\}_{\nu \in N}$ lop-converge to $F : X \times Y \rightarrow \mathbb{R}$ if*

- (i) *for all $y \in Y$ and $x^\nu \rightarrow x \in X$, with $x^\nu \in X^\nu$, there exists $y^\nu \rightarrow y$, with $y^\nu \in Y^\nu$, such that $\limsup F^\nu(x^\nu, y^\nu) \leq F(x, y)$ if $x \in X$ and $F^\nu(x^\nu, y^\nu) \rightarrow -\infty$ otherwise.*
- (ii) *for all $x \in X$, there exists $x^\nu \rightarrow x$, with $x^\nu \in X^\nu$, such that for all $y^\nu \rightarrow y \in Y$, with $y^\nu \in Y^\nu$, $\liminf F^\nu(x^\nu, y^\nu) \geq F(x, y)$ if $y \in Y$ and $F^\nu(x^\nu, y^\nu) \rightarrow \infty$ 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.

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

Proof. We follow the same arguments as in [6, Proposition 3.2], where $\mathcal{X} = \mathbb{R}^n$ is considered. From Definition 4.1(ii) there exists $x^\nu \rightarrow x$, 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 $x^\nu \rightarrow x$, with $x^\nu \in X^\nu$, one can find $y^\nu \rightarrow y$, with $y^\nu \in Y^\nu$, such that Definition 3.1(ii) is also satisfied. $\square$

We recall that the inf-projections of the bifunctions $F^\nu : X^\nu \times Y^\nu \rightarrow \mathbb{R}$ and $F : X \times Y \rightarrow \mathbb{R}$ are defined as the functions

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

In addition to their overall interest, inf-projections of bifunctions are central to the study of optimality functions as clearly highlighted by (5).

4.3 Theorem (containment of inf-projections) *Suppose that the bifunctions $\{F^\nu : X^\nu \times Y^\nu \rightarrow \mathbb{R}\}_{\nu \in N}$ lop-converge to $F : X \times Y \rightarrow \mathbb{R}$ and $-\infty < \inf_Y F(x, \cdot)$ for some $x \in X$. Then, the inf-projections $h^\nu : X^\nu \rightarrow [-\infty, \infty)$ and $h : X \rightarrow [-\infty, \infty)$ 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 $\mathbb{N}$, $x^\nu \in X^\nu$, $h^\nu(x^\nu) \geq x_0^\nu$, $x^\nu \rightarrow^N x$, and $x_0^\nu \rightarrow^N x_0$. If $x \notin X$, then take $y \in Y$ and construct a sequence $y^\nu \rightarrow y$, with $y^\nu \in Y^\nu$, such that $F^\nu(x^\nu, y^\nu) \rightarrow^N -\infty$, which exists by Definition 4.1(i). However,

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

imply to a contradiction since $x_0^\nu \rightarrow^N x_0 \in \mathbb{R}$. 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 $y^\nu \rightarrow y$, with $y^\nu \in Y^\nu$, such that $\limsup F^\nu(x^\nu, y^\nu) \leq F(x, y)$. Consequently,

$$x_0 = \limsup_{\nu \in N} x_0^\nu \leq \limsup_{\nu \in N} h^\nu(x^\nu) \leq \limsup_{\nu \in N} F^\nu(x^\nu, y^\nu) \leq F(x, y) \leq x_0 - 1,$$

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 $F(x, y_\varepsilon) - \varepsilon \leq h(x)$. Then Definition 4.1(i) again yields $y^\nu \rightarrow y_\varepsilon$, with $y^\nu \in Y^\nu$, such that

$$\limsup_{\nu \in N} h^\nu(x^\nu) \leq \limsup_{\nu \in N} F^\nu(x^\nu, y^\nu) \leq F(x, y_\varepsilon) \leq h(x) + \varepsilon,$$

implying $\limsup_{\nu \in N} h^\nu(x^\nu) \leq h(x)$. Since

$$x_0 = \limsup_{\nu \in N} x_0^\nu \leq \limsup_{\nu \in N} h^\nu(x^\nu) \leq h(x),$$

the conclusion follows. $\square$

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

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

$$\inf_{y \in Y^\nu \cap B_\varepsilon} F^\nu(x^\nu, y) \leq \inf_{y \in Y^\nu} F^\nu(x^\nu, y) + \varepsilon \text{ for sufficiently large } \nu.$$

Under ancillary-tight lop-convergence, we can strengthen the conclusion of Theorem 4.3 as follows.

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

Proof. Since it is a very short proof, we include it for completeness sake. It is verbatim the same as that of [6, Theorem 3.4]. Let $x \in X$ be such that $h(x)$ is finite. Now, choose $x^\nu \rightarrow x$, with $x^\nu \in X^\nu$, such that $F^\nu(x^\nu, \cdot)$ epi-converge to $F(x, \cdot)$, cf. Proposition 4.2. In fact, they epi-converge tightly as an immediate consequence of ancillary-tightness. Thus,

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

via Theorem 3.4. $\square$

Further strengthening of the notion is also beneficial.

4.6 Definition (tight lop-convergence) *The lop-convergence of bifunctions $\{F^\nu : X^\nu \times Y^\nu \rightarrow \mathbb{R}\}_{\nu \in N}$ to $F : X \times Y \rightarrow \mathbb{R}$ is tight if Definition 4.4 holds and for any $\varepsilon > 0$ one can find a compact set $A_\varepsilon \subset \mathcal{X}$ such that*

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

Under tight lop-convergence, we can strengthen the conclusion of Theorem 4.5 as follows.

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

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

Moreover, for every $x^* \in \operatorname{argmax}_{x \in X} \inf_{y \in Y} F(x, y)$, there exist an infinite subsequence N of $\mathbb{N}$, $\{\varepsilon^\nu\}_{\nu \in N}$, with $\varepsilon^\nu \searrow 0$, and $\{x^\nu\}_{\nu \in N}$, with $x^\nu \in \varepsilon^\nu$ - $\operatorname{argmax}_{x \in X^\nu} \inf_{y \in Y^\nu} F^\nu(x, y)$, such that $x^\nu \rightarrow^N x$. Conversely, if such sequences exists, then $\sup_{x \in X^\nu} \inf_{y \in Y^\nu} F^\nu(x, y) \rightarrow^N \inf_{y \in Y} F(x^*, \cdot)$

Proof. We refer to the arguments of the proof of [6, Theorem 3.7]. $\square$

5 Applications and Further Examples

We now return to the context of optimality functions of the form (5). We start with providing a sufficient condition for the required upper semicontinuity of an optimality function; see Definition 2.1. We state the result for general inf-projections.

5.1 Theorem (upper semicontinuity of inf-projection) *If the bifunction $F : X \times Y \rightarrow \mathbb{R}$ is lower semicontinuous, $F(\cdot, y)$ is upper semicontinuous for all $y \in Y$, and Y is closed, then the corresponding inf-projection $h(x) = \inf_Y F(x, \cdot)$, $x \in X$, is upper semicontinuous.*

Proof. Let $x^\nu \rightarrow x \in X$, with $x^\nu \in X$. We show that the functions $F(x^\nu, \cdot)$ epi-converge to $F(x, \cdot)$, which are all defined on Y . Suppose that $y^\nu \rightarrow y$, with $y^\nu \in Y$. By the closedness of Y , $y \in Y$. Moreover, $\liminf F(x^\nu, y^\nu) \geq F(x, y)$ by the lower semicontinuity of F at (x, y) . Consequently, part (i) of Definition 3.1 is satisfied. Next, let $y^\nu = y \in Y$. Then, $\limsup F(x^\nu, y^\nu) = \limsup F(x^\nu, y) \leq F(x, y)$ by the upper semicontinuity of $F(\cdot, y)$ and part (ii) of Definition 3.1 is satisfied. Since $F(x^\nu, \cdot)$ epi-converge to $F(x, \cdot)$, Theorem 3.2 demonstrates that $\limsup h(x^\nu) \leq h(x)$. $\square$

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

We next turn to the requirement for consistency in Definition 3.5(ii).

5.2 Theorem (sufficient condition for consistency, optimality function part) *Suppose that the bifunctions $\{F^\nu : X^\nu \times Y^\nu \rightarrow \mathbb{R}\}_{\nu \in N}$ lop-converge to $F : X \times Y \rightarrow \mathbb{R}$ and that the bifunctions define the optimality functions $\theta^\nu = \inf_{y \in Y^\nu} F^\nu(\cdot, y)$ and $\theta = \inf_{y \in Y} F(\cdot, y)$, with $-\infty < \theta(x)$ for some $x \in X$. Then,*

for every $x^\nu \rightarrow x \in X$, with $x^\nu \in X^\nu$, $\limsup \theta^\nu(x^\nu) \leq \theta(x)$ if $x \in X$, and $\theta^\nu(x^\nu) \rightarrow -\infty$ otherwise.

Proof. The result is a direct consequence of Theorem 4.3. $\square$

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 Theorem 5.2 in the context of Example 3.

Example 3: Minimax Problem (cont.). Suppose that for every $z \in Z$, there exists a sequence $z^\nu \in Z^\nu$ such that $z^\nu \rightarrow z$. Let

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

and F be defined similarly with the superscripts removed. We next show lopsided convergence of F^ν to F . First consider part (i) of Definition 4.1. Let $y \in \mathbb{R}^n$ and $x^\nu \rightarrow x \in \mathbb{R}^n$. Set $y^\nu = y$ for all ν . Clearly, $\limsup F^\nu(x^\nu, y^\nu) \leq \limsup F(x^\nu, y) = F(x, y)$ by the continuity of F and part (i) holds. Second, we consider part (ii). Let $x \in \mathbb{R}^n$ and $y^\nu \rightarrow y \in \mathbb{R}^n$. 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 exists $z^\nu \in Z^\nu$ and ν_0 such that

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

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

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

Since $y^\nu \rightarrow y$, $\{z^\nu\}$ is bounded, and $\nabla_x \varphi$ is continuous, it follows that $\liminf F^\nu(x^\nu, y^\nu) \geq F(x, y) - 2\varepsilon$. Since ε was arbitrary, part (ii) of Definition 4.1 holds and F^ν therefore lop-converge to F . In view of Theorem 5.2 and the fact that epi-convergence is also easily established, we have that $\{(f^\nu : \mathbb{R}^n \rightarrow \mathbb{R}, \theta^\nu : \mathbb{R}^n \rightarrow \mathbb{R}_-)\}$ are consistent approximations of $\{(f : \mathbb{R}^n \rightarrow \mathbb{R}, \theta : \mathbb{R}^n \rightarrow \mathbb{R}_-)\}$ in this case. The above algorithm therefore is implementable for the solution of the semiinfinite minimax problem $\min_{x \in \mathbb{R}^n} \max_{z \in Z} \varphi(x, z)$.

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

5.3 Theorem (convergence to quasi-stationary points) *Suppose that the bifunctions $\{F^\nu : X^\nu \times Y^\nu \rightarrow \mathbb{R}\}_{\nu \in N}$ lop-converge ancillary-tightly to $F : X \times Y \rightarrow \mathbb{R}$ and $\theta = \inf_{y \in Y} F(\cdot, y)$ is an optimality function for $f : C \rightarrow \mathbb{R}$ with $\mathcal{Q}_\theta \neq \emptyset$. If $x^\nu \in \operatorname{argmax}_{x \in X^\nu} \inf_{y \in Y^\nu} F^\nu(x, y)$ for all ν , then every cluster point x of $\{x^\nu\}_{\nu \in N}$ is quasi-stationary for $f : C \rightarrow \mathbb{R}$, i.e., $x \in \mathcal{Q}_\theta$.*

Proof. By Theorem 4.5, $h^\nu = \inf_{y \in Y^\nu} F^\nu(\cdot, y)$ hypo-converge to $h = \inf_{y \in Y} F(\cdot, y)$. By translating Theorem 3.2 from the stated minimization framework to a maximization framework we obtain that $x \in \operatorname{argmax}_X h$. Since $\mathcal{Q}_\theta \neq \emptyset$, $\mathcal{Q}_\theta = \operatorname{argmax}_X h$ and the conclusion follows. $\square$

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

5.4 Theorem (characterization of quasi-stationary points) *Suppose that the bifunctions $\{F^\nu : X^\nu \times Y^\nu \rightarrow \mathbb{R}\}_{\nu \in N}$ lop-converge tightly to $F : X \times Y \rightarrow \mathbb{R}$ and $\theta = \inf_{y \in Y} F(\cdot, y)$ is an optimality function for $f : C \rightarrow \mathbb{R}$ 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 $x^\nu \in \varepsilon^\nu$ - $\operatorname{argmax}_{x \in X^\nu} \inf_{y \in Y^\nu} F^\nu(x, y)$, such that $x^\nu \xrightarrow{N} x$.*

Proof. The result is a direct consequence of Theorem 4.7. $\square$

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

Example 4: Optimal Control. We here follow the set-up in Section 5.6 and Chapter 4 of [12], which contain further details. For $g : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$, we consider the dynamical system

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

where the control $u \in \mathcal{L}_\infty^m = \{u : [0, 1] \rightarrow \mathbb{R}^m \mid \text{measurable, essentially bounded}\}$. Since such controls are contained in the space of square-integrable functions from $[0, 1]$ to $\mathbb{R}^m$, the usual L_2 -norm applies; see [12, p.709] for a motivation for this “hybrid” set-up. Let $\mathcal{H} = \mathbb{R}^n \times \mathcal{L}_\infty^m$. 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

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

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 $C \subset \mathbb{R}^m$. By imposing the constraints for almost every t instead of every t , we deviate slightly from [12] and follow [9]. We therefore also define the feasible set

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

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

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

has a Gateaux differential of the form $\langle \nabla f(\eta), \bar{\eta} - \eta \rangle_{\mathcal{H}}$ for some Lipschitz continuous gradient $\nabla f(\eta)$ given in [12, Corollary 5.6.9]. The optimal control problem

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

analogous to Example 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 $U^\nu \subset U$, $\nu \in \mathbb{N}$, consist of the piecewise constant functions that are constant on each of the intervals $[(k-1)/\nu, k/\nu)$, $k = 1, \dots, \nu$. Set $H^\nu = \mathbb{R}^n \times U^\nu$. 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

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

is an optimality function of $f^\nu : H^\nu \rightarrow \mathbb{R}$, where the Lipschitz continuous gradient $\nabla f^\nu(\eta)$ is given in [12, Theorem 5.6.19].

By [12, Theorem 4.3.2], for every bounded set $S \subset H$, there exists a $C_S < \infty$ such that $|f(\eta) - f^\nu(\eta)| \leq C_S/\nu$ and $\|\nabla f(\eta) - \nabla f^\nu(\eta)\|_{\mathcal{H}} \leq C_S/\nu$ for all $\eta \in S$. Moreover, $\cup_{\nu \in \mathbb{N}} H^\nu$ is dense in H . Consequently, it is easily established that $f^\nu : H^\nu \rightarrow \mathbb{R}$ epi-converge to $f : H \rightarrow \mathbb{R}$. We next consider the optimality functions. Let $\bar{\eta} \in H$ and $\eta^\nu \rightarrow \eta \in \mathcal{H}$, with $\eta^\nu \in H^\nu$. Necessarily, $\eta \in H$. Due to the density result, there exists $\bar{\eta}^\nu \rightarrow \bar{\eta}$, with $\bar{\eta}^\nu \in H^\nu$. Hence,

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

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, $\{(f^\nu : H^\nu \rightarrow \mathbb{R}, \theta^\nu : H^\nu \rightarrow \mathbb{R}_-)\}_{\nu \in \mathbb{N}}$ are consistent approximations of $(f : H \rightarrow \mathbb{R}, \theta : H \rightarrow \mathbb{R}_-)$. Since the minimization of $f^\nu : H^\nu \rightarrow \mathbb{R}$ is equivalent to an optimization problem on a Euclidean space, the above algorithm is implementable for the infinite-dimensional problem $f : H \rightarrow \mathbb{R}$.

References

- [1] H. Attouch. *Variational Convergence for Functions and Operators*. Applicable Mathematics Sciences. Pitman, 1984.
- [2] H. Attouch and R. Wets. A convergence theory for saddle functions. *Transactions of the American Mathematical Society*, 280(1):1–41, 1983.
- [3] J.-P. Aubin and H. Frankowska. *Set-Valued Analysis*. Birkhäuser, 1990.
- [4] G. Beer. *Topologies on Closed and Closed Convex Sets*, volume 268 of *Mathematics and its Applications*. Kluwer, 1992.
- [5] J.C. Foraker, J.O. Royset, and I. Kaminer. Search-trajectory optimization: Part 1, formulation and theory. in review.
- [6] A. Jofre and R. J-B Wets. Variational convergence of bivariate functions: Lopsided convergence. *Mathematical Programming B*, 116:275–295, 2009.
- [7] A. Jofre and R. J-B Wets. Variational convergence of bifunctions: Motivating applications. *SIAM J. Optimization*, 24(4):1952–1979, 2014.
- [8] R. Pasupathy. On choosing parameters in retrospective-approximation algorithms for stochastic root finding and simulation optimization. *Operations Research*, 58(4):889–901, 2010.
- [9] C. Phelps, J.O. Royset, and Q. Gong. Optimal control of uncertain systems using sample average approximations. in review.
- [10] E. Polak. *Computational Methods in Optimization. A Unified Approach*. Academic Press, 1971.
- [11] E. Polak. On the use of models in the synthesis of optimization algorithms. In H. Kuhn and G. Szego, editors, *Differential Games and Related Topics*, pages 263–279. North Holland, Amsterdam, 1971.
- [12] E. Polak. *Optimization. Algorithms and Consistent Approximations*, volume 124 of *Applied Mathematical Sciences*. Springer, 1997.
- [13] R.T. Rockafellar and R. Wets. *Variational Analysis*, volume 317 of *Grundlehren der Mathematischen Wissenschaft*. Springer, 3rd printing-2009 edition, 1998.
- [14] J. O. Royset. Optimality functions in stochastic programming. *Mathematical Programming*, 135(1):293–321, 2012.
- [15] J. O. Royset. On sample size control in sample average approximations for solving smooth stochastic programs. *Computational Optimization and Applications*, 55(2):265–309, 2013.
- [16] J. O. Royset and R. Szechtman. Optimal budget allocation for sample average approximation. *Operations Research*, 61:762–776, 2013.

- [17] J.O. Royset and E.Y. Pee. 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.