Computational Optimization and Applications, 13, 137–161 (1999)

First-Order Algorithms for Generalized Semi-Infinite Min-Max Problems *

ELIJAH POLAK polak@optimum.eecs.berkeley.edu
Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley, CA 94720, USA

LIQUN QI l.qi@unsw.edu.au

DEFENG SUN sun@maths.unsw.edu.au
School of Mathematics, the University of New South Wales, Sydney 2052, Australia

Received June 20, 1997; Accepted July 27, 1998

Abstract. We present a first-order algorithm for solving semi-infinite generalized min-max problems which consist of minimizing a function $f^0(x) = F(\psi^1(x), ..., \psi^m(x))$, where F is a smooth function and each ψ^i is the maximum of an infinite number of smooth functions.

In Section 3.3 of [17] Polak finds a methodology for solving infinite dimensional problems by expanding them into an infinite sequence of consistent finite dimensional approximating problems, and then using a master algorithm that selects an appropriate subsequence of these problems and applies a number of iterations of a finite dimensional optimization algorithm to each of these problems, sequentially. Our algorithm was constructed within this framework; it calls an algorithm by Kiwiel as a subroutine. The number of iterations of the Kiwiel algorithm to be applied to the approximating problems is determined by a test that ensures that the overall scheme retains the rate of convergence of the Kiwiel algorithm.

Under reasonable assumptions we show that all the accumulation points of sequences constructed by our algorithm are stationary, and, under an additional strong convexity assumption, that the Kiwiel algorithm converges at least linearly, and that our algorithm also converges at least linearly, with the same rate constant bounds as Kiwiel's.

Keywords: generalized min-max problems, consistent approximations, optimality functions, first-order methods, linear convergence

Dedication: Olvi, please accept this modest tribute in celebration of your 65th birthday. I wish you another 65 years of good health, happiness, and important contributions to nonlinear programming. - Elijah Polak

1. Introduction

Generalized min-max problems have the form

P

$$\min_{x \in \Re^n} f^0(x), \tag{1}$$

where

$$f^0(x) := F(\psi(x)), \tag{2}$$

* The research of the first author was supported by the National Science Foundation grant NSF-INT-9725220, and the research of the second author and the third author was supported by the Australian Research Council.

where $F : \Re^m \to \Re$, $\psi : \Re^n \to \Re^m$, with $\psi(x) = (\psi^1(x), ..., \psi^m(x))^1$, and, for $j = 1, ..., m$, $\psi^j : \Re^n \to \Re$ are of the form

$$\max_{y_j \in Y_j} \phi^j(x, y_j), \tag{3}$$

and $\phi^j : \Re^n \times \Re^{m_j} \to \Re$. It is usual to assume that the functions $F(\cdot)$, $\phi(\cdot, \cdot)$ are smooth and that the sets $Y_j \subset \Re^{m_j}$ are compact. When the cardinality of the sets Y_j is finite, $\mathbf{P}$ is a *finite* generalized min-max problem, while when the cardinality of at least one of the sets Y_j is infinite, $\mathbf{P}$ is a *semi-infinite* generalized min-max problem.

Generalized min-max problems are among the simplest of nonsmooth problems beyond ordinary min-max problems that have manifestly exploitable structure. Quite early, they were recognized to be in the class of quasi-differentiable problems, see, e.g., [4] and hence, at least in principle, finite-dimensional generalized min-max problems are solvable by algorithms developed for minimizing quasi-differentiable functions, see, e.g., [2, 3, 7, 8, 9, 10, 11, 18]. Under the additional assumption that $\partial F(y)/\partial y^j \geq 0$ for all $y \in \Re^m$ and $j = 1, ..., m$, finite generalized min-max problems can also be solved by means of nonlinear programming algorithms, using a transformation that converts a generalized min-max problem into a smooth constrained nonlinear programming problem[2] (see e.g., [1, 5, 12]). Other methods that depend on the assumption that $\partial F(y)/\partial y^j \geq 0$ for all $y \in \Re^m$ and $j = 1, ..., m$, can be found, for example, in [7]. To date, only finite-dimensional generalized min-max problems have been considered.

Generalized min-max problems are not only of theoretical interest. For example, when $F(y) = \sum_{j=0}^{m} y^j$, or $F(y) = \sum_{j=0}^{m} (y^j)^2$, or $F(y) = \sum_{j=0}^{m} \log y^j$, minimizing $f^0(\cdot)$ corresponds to minimizing the average value, mean square value, or nonlinearly weighted mean value of the $\psi^j(x)$, respectively. Less obvious cases are bound to emerge in the engineering and economics literatures as efficient algorithms for the solution of generalized min-max problems become available.

It is a well known fact that the less problem structure is taken into account, the less efficient are the resulting nonsmooth optimization algorithms. Thus, there are both linearly and superlinearly converging algorithms for the solution of min-max problems of the form (1) (see. e.g., Chapter 4 in [17]), which make strong use of the structure of these problems. In this paper we make full use of the structure of the nonsmooth problem and of the theory of consistent approximations in [17] to obtain a linearly converging algorithm.

We will consider problem (1) under the following hypotheses.

Assumption 1. *We will assume that*

(a) *The functions $F(\cdot)$ and $\phi^j(\cdot, \cdot)$, $j \in \mathbf{m}$, are at least once continuously differentiable.*

(b) *There exists a $c_F > 0$ such that for all $y \in \Re^m$ and $j \in \mathbf{m}$, $\frac{\partial F}{\partial y^j}(y) \geq c_F$.*

(c) *The sets Y_j are either convex and compact, or of finite cardinality, of the form*

$$Y_j = \{y_{j,1}, \ldots, y_{j,q_j}\}\ . \tag{4}$$

Parts *(a)* and *(b)* of Assumption 1 ensure that when both the $F(\cdot)$ and the $\psi^j(\cdot)$ are convex, the function $f^0(\cdot)$ is also convex. In addition, as we will see, when parts *(a)* and *(b)* of Assumption 1 hold, the function $f^0(\cdot)$ has a subgradient.

In Section 2 we derive optimality conditions in optimality function form that will be needed in the construction of our generalized semi-infinite optimization algorithm. In Section 3 we show that a particular case of the Kiwiel extension [7, 10] of the Pshenichnyi-Pironneau-Polak Minimax Algorithm (see [19, 13, 14] and Algorithm 4.1 in [17]) to generalized finite min-max problems converges linearly[3]. In Section 4 we present our algorithm for solving generalized semi-infinite min-max problems, which can be seen to be an extension of the Polak-He PPP Rate-Preserving Algorithm 3.4.9 in [17] (see also [15]). The algorithm consists of a master algorithm that constructs consistent finite dimensional approximations by discretizing the intervals in the max functions in (1), calls the Kiwiel algorithm as a subroutine for solving the resulting finite dimensional problems, and determines how long the Kiwiel algorithm should be used at a given level of discretization. We show that our algorithm converges linearly. In Section 5 we present a couple of numerical results that illustrate the behavior of our algorithm, and our conclusions are in Section 6.

2. Optimality Conditions

In this section we will present optimality conditions for problem (1) both in "classical" form and in terms of an optimality function which leads to a linearly converging first-order algorithm. First, we need the following straightforward result.

LEMMA 1 *Suppose that $F : \Re^m \to \Re$ is continuously differentiable and that $\psi : \Re^n \to \Re^m$ is a locally Lipschitz continuous function that has directional derivatives at every $x \in \Re^n$. Let $f^0 : \Re^n \to \Re$ be defined by*

$$f^0(x) = F(\psi(x)) \ . \tag{5}$$

Then, given any $x \in \Re^n$, and direction vector $h \in \Re^n$, the function $f^0(\cdot)$ has a directional derivative $df^0(x;h)$ which is given by

$$df^0(x;h) = \langle \nabla F(\psi(x)), d\psi(x;h)\rangle \ . \tag{6}$$

Suppose that Assumption 1 is satisfied. Then it follows from Lemma 1 that the directional derivative of $f^0(\cdot)$, at a point $x \in \Re^n$ in the direction h, is given by

$$\begin{aligned} df^0(x;h) &= \sum_{j\in\mathbf{m}} \frac{\partial F}{\partial y^j}(\psi(x)) d\psi^j(x;h) \\ &= \sum_{j\in\mathbf{m}} \frac{\partial F}{\partial y^j}(\psi(x)) \max_{y_j\in\hat{Y}_j(x)} \langle \nabla_x \phi^j(x,y_j), h\rangle, \end{aligned} \tag{7}$$

where

$$\hat{Y}_j(x) := \{y_j \in Y_j \mid \phi^j(x,y_j) = \psi^j(x)\}. \tag{8}$$

When all the sets Y_j are as in (4), (7) assumes the form

$$df^0(x;h) = \sum_{j\in\mathbf{m}} \frac{\partial F}{\partial y^j}(\psi(x)) \max_{k\in\hat{\mathbf{q}}_\mathbf{j}(\mathbf{x})} \langle \nabla f^{j,k}(x), h\rangle, \tag{9}$$

where the functions $f^{j,k}(\cdot)$ are defined by

$$f^{j,k}(x) := \phi^j(x, y_{j,k}),\ k \in \mathbf{q}_\mathbf{j}, \tag{10}$$

and the sets $\hat{\mathbf{q}}_\mathbf{j}(x)$ by

$$\hat{\mathbf{q}}_\mathbf{j}(x) := \{k \in \mathbf{q}_\mathbf{j} \mid f^{j,k}(x) = \psi^j(x)\} \ . \tag{11}$$

Hence the following result is obvious.

THEOREM 1 *Suppose that $\hat{x}$ is a local minimizer for the problem (1). Then for all $h \in \Re^n$,*

$$\begin{aligned} df^0(\hat{x};h) &= \sum_{j\in\mathbf{m}} \frac{\partial F}{\partial y^j}(\psi(\hat{x})) d\psi^j(\hat{x};h) \\ &= \sum_{j\in\mathbf{m}} \frac{\partial F}{\partial y^j}(\psi(\hat{x})) \max_{y_j\in\hat{Y}_j(\hat{x})} \langle \nabla_x \phi^j(\hat{x}, y_j), h\rangle \geq 0 \ . \end{aligned} \tag{12}$$

Furthermore, (12) holds if and only if $0 \in \partial f^0(\hat{x})$, where the subgradient $\partial f^0(\hat{x})$ is given by

$$\partial f^0(\hat{x}) = \sum_{j\in\mathbf{m}} \left\{ \mathrm{conv}_{y_j\in\hat{Y}_j(\hat{x})} \left\{ \frac{\partial F}{\partial y^j}(\psi(\hat{x})) \nabla_x \phi^j(\hat{x}, y_j) \right\} \right\} \ . \tag{13}$$

Since (12) is a necessary condition of optimality, any point $\hat{x} \in \Re^n$ that satisfies (12) will be called *stationary*.

When all the sets Y_j are of the form (4), the expressions (12) and (13) assume the following form:

$$df^0(\hat{x};h) = \sum_{j\in\mathbf{m}} \frac{\partial F}{\partial y^j}(\psi(\hat{x})) \max_{k\in\hat{\mathbf{q}}_\mathbf{j}(\hat{x})} \langle \nabla f^{j,k}(\hat{x}), h\rangle \geq 0 \ , \quad \forall h \in \Re^n, \tag{14}$$

$$\partial f^0(\hat{x}) = \sum_{j\in\mathbf{m}} \mathrm{conv}_{k\in\hat{\mathbf{q}}_\mathbf{j}(\hat{x})} \left\{ \frac{\partial F}{\partial y^j}(\psi(\hat{x})) \nabla f^{j,k}(\hat{x}) \right\} \ . \tag{15}$$

Definition 1. *We will say that $\theta : \Re^n \to \Re$ is an* optimality function *for problem (1) if*

(a) $\theta(\cdot)$ is upper-semicontinuous,

(b) $\theta(x) \leq 0$ for all $x \in \Re^n$, and

(c) for any $\hat{x} \in \Re^n$, (12) holds if and only if $\theta(\hat{x}) = 0$.

Let $\delta > 0$ be arbitrary. Then we define the function $\theta : \Re^n \to \Re$ and the associated search direction function $h : \Re^n \to \Re^n$ by[4]

$$\theta(x) = \min_{h \in \Re^n} \sum_{j \in \mathbf{m}} \frac{\partial F}{\partial y^j}(\psi(x)) \max_{y_j \in Y_j} \{\phi^j(x, y_j) - \psi^j(x) + \langle \nabla_x \phi^j(x, y_j), h\rangle + \tfrac{1}{2}\delta \|h\|^2\} \tag{16}$$

and

$$h(x) = \arg \min_{h \in \Re^n} \sum_{j \in \mathbf{m}} \frac{\partial F}{\partial y^j}(\psi(x)) \max_{y_j \in Y_j} \{\phi^j(x, y_j) - \psi^j(x) + \langle \nabla_x \phi^j(x, y_j), h\rangle + \tfrac{1}{2}\delta \|h\|^2\}. \tag{17}$$

For any $j \in \mathbf{m}$, and any $x \in \Re^n$, let $\bar{G}\psi^j(x) \subset \Re^{n+1}$ be a set with elements $\bar{\xi}_j = (\xi_j^0, \xi_j)$, $\xi_j^0 \in \Re$, $\xi_j \in \Re^n$, defined by

$$\bar{G}\psi^j(x) := \text{conv}_{y_j \in Y_j} \left\{ \begin{bmatrix} \psi^j(x) - \phi^j(x, y_j) \\ \nabla_x \phi^j(x, y_j) \end{bmatrix} \right\}, \quad j \in \mathbf{m}\,, \tag{18}$$

and let

$$a(x) := (\frac{\partial F}{\partial y^1}(\psi(x)), ..., \frac{\partial F}{\partial y^m}(\psi(x)))\,. \tag{19}$$

Then we conclude from (16) that an alternative expression for $\theta(x)$ is as follows

$$\theta(x) = \min_{h \in \Re^n} \sum_{j \in \mathbf{m}} a^j(x) \max_{\bar{\xi}_j \in \bar{G}\psi^j(x)} \{-\xi_j^0 + \langle \xi_j, h\rangle + \tfrac{1}{2}\delta \|h\|^2\}. \tag{20}$$

Using the fact that

$$\max_{j \in \mathbf{p}} u^j + \max_{k \in \mathbf{q}} v^k = \max_{j \in \mathbf{p}} \max_{k \in \mathbf{q}} (u^j + v^k), \tag{21}$$

we conclude that

$$\theta(x) = \min_{h \in \Re^n} \max_{\bar{\xi} \in \bar{G}f^0(x)} \{-\xi^0 + \langle \xi, h\rangle + \tfrac{1}{2}\gamma(x)\|h\|^2\}\,, \tag{22}$$

where $\bar{G}f^0(x) \subset \Re^{n+1}$ is defined by

$$\bar{G}f^0(x) := \sum_{j \in \mathbf{m}} a^j(x) \bar{G}\psi^j(x), \tag{23}$$

and

$$\gamma(x) := \delta \sum_{j \in \mathrm{m}} a^j(x). \tag{24}$$

Since the set $\bar{G}f^0(x)$ is convex and compact, and the function $-\xi_j^0 + \langle \xi, h\rangle + \frac{1}{2}\gamma(x)\|h\|^2$ is convex in h, concave in $\bar{\xi}$, and $-\xi_j^0 + \langle \xi, h\rangle + \frac{1}{2}\gamma(x)\|h\|^2 \to \infty$ as $\|h\| \to \infty$, we can make use of a corollary to the von Neumann Theorem (see Corollary 5.5.6 in [17]) and interchange the max and the min in (22), to obtain

$$\theta(x) = \max_{\bar{\xi}\in\bar{G}f^0(x)} \min_{h\in\Re^n} \{-\xi^0 + \langle \xi, h\rangle + \tfrac{1}{2}\gamma(x)\|h\|^2\}. \tag{25}$$

Now, solving the inner minimization problem in (25) for h, we find that

$$h = -\frac{1}{\gamma(x)}\xi \tag{26}$$

and hence (25) simplifies out to

$$\theta(x) = -\min_{\bar{\xi}\in\bar{G}f^0(x)} \{\xi^0 + \frac{1}{2\gamma(x)}\|\xi\|^2\}\ . \tag{27}$$

Now, let

$$\bar{\xi}^*(x) = (\xi^{0*}(x), \xi^*(x)) = -\arg\min_{\bar{\xi}\in\bar{G}f^0(x)} \{\xi^0 + \frac{1}{2\gamma(x)}\|\xi\|^2\}\ . \tag{28}$$

Then it follows from Corollary 5.5.6 in [17] and (26) that an alternative expression for $h(x)$ is given by

$$h(x) = -\frac{1}{\gamma(x)}\xi^*(x)\ . \tag{29}$$

When all the sets Y_j are of the form (4), (27) reduces to the following quadratic programming problem:

$$\theta(x) = -\min_{\mu_j \in \sum_{j\in\mathbf{m}} q_j} \left\{ \sum_{j\in\mathbf{m}} \sum_{k\in\mathbf{q_j}} a^j(x)\mu_j^k(\psi^j(x) - f^{j,k}(x)) + \frac{1}{2}\left(\sum_{j\in\mathbf{m}} a^j(x)\delta\right)^{-1} \|\sum_{j\in\mathbf{m}}\sum_{k\in\mathbf{q_j}} a^j(x)\mu_j^k \nabla f^{j,k}(x)\|^2 \right\}, \tag{30}$$

where, for any positive integer q, Σ_q denotes the unit simplex in $\Re^q$, which is defined by

$$\Sigma_q = \{\mu \in \Re^q \mid \mu^j \geq 0,\ j \in \mathbf{q},\ \sum_{j\in\mathbf{q}} \mu^j = 1\}. \tag{31}$$

Let $\mu(x) = (\mu_1(x), ..., \mu_m(x))$, with $\mu_j(x) \in \Re^{m_j}$, denote any solution of (30). Then we see that

$$h(x) = -\sum_{j\in\mathbf{m}}\sum_{k\in\mathbf{q_j}} a^j(x)\mu_j^k(x)\nabla f^{j,k}(x)\ . \tag{32}$$

The following theorem shows that $\theta(\cdot)$ is indeed an optimality function for the problem (1) and that the search direction function $h(\cdot)$ is a descent direction function for $f^0(\cdot)$.

THEOREM 2 *Suppose that Assumption 1 is satisfied. Consider the functions $\theta(\cdot)$ and $h(\cdot)$ defined by (16) and (17). Then*

(i) For all $x \in \Re^n$,

$$\theta(x) \leq 0 \,. \tag{33}$$

(ii) For all $x \in \Re^n$,

$$df^0(x; h(x)) \leq \theta(x) - \tfrac{1}{2} \sum_{j \in \mathbf{m}} a_j(x)\delta \|h(x)\|^2, \tag{34}$$

where $df^0(x; h(x))$ is the directional derivative of f^0 at x in the direction $h(x)$.

(iii) For any $x \in \Re^n$, $0 \in \partial f^0(x)$ if and only if $\theta(x) = 0$, where $\partial f^0(x)$ is the subgradient of $f^0(x)$ at x, defined in (13).

(iv) The function $\theta(\cdot)$ is continuous.

(v) The function $h(\cdot)$ is point-valued and continuous.

The proof of this theorem can be obtained by straightforward extension of the proof of Theorem 3.1.6 in [17] and is therefore omitted.

3. An Algorithm for Solving Generalized Finite Min-Max Problems

We will need a special case of the generalized finite min-max algorithm described in [7, 10] as a subroutine, for solving problems of the form (1) when the sets Y_j are of the form (4). Defining the functions $f^{j,k}(\cdot)$ are as in (10), these generalized finite min-max problems are of the form (1), with

$$\left.\begin{aligned} &\min_{x \in \Re^n} \; f^0(x) \\ &f^0(x) = F(\psi(x)), \\ &\psi(x) = (\psi^1(x), ..., \psi^m(x)), \\ &\psi^j(x) = \max_{k \in \mathbf{q_j}} \; f^{j,k}(x), \;\; j \in \mathbf{m} \,, \end{aligned}\right\} \tag{35}$$

where, in view of Assumption 1, the functions $F(\cdot)$ and $f^{j,k}(\cdot)$, $j \in \mathbf{m}$, $k \in \mathbf{q_j}$ are all continuously differentiable.

The special case of the Kiwiel algorithm [7, 10] that we need is a straightforward extension of the Pshenichnyi-Pironneau-Polak Minimax Algorithm 2.4.1 in [17] and has the following form:

Algorithm 1 *(Solves Problem (35))*

Parameters. $\alpha \in (0,1)$, $\beta \in (0,1)$, *and* $\delta > 0$.

Data. $x_0 \in \Re^n$.

Step 0. *Set* $i = 0$.

Step 1. *Compute the optimality function value* $\theta_i := \theta(x_i)$ *and the search direction* $h_i := h(x_i)$ *according to the formulae (30) and (32).*

Step 2. *If* $\theta_i = 0$*, stop. Else, compute the step-size*

$$\lambda_i = \lambda(x_i) := \max_{k \in N} \{\beta^k \,|\, f^0(x_i + \beta^k h_i) - f^0(x_i) - \beta^k \alpha \theta_i \le 0\}, \tag{36}$$

where $N := \{0, 1, 2, ..., \}$.

Step 3. *Set*

$$x_{i+1} = x_i + \lambda_i h_i, \tag{37}$$

replace i *by* $i + 1$*, and go to Step 1.*

THEOREM 3 *[7] Suppose that Assumption 1 is satisfied and that all the* Y_j*,* $j \in \mathbf{m}$ *are of the form (4), so that problem (1)-(2) reduces to problem (35). If* $\{x_i\}_{i=0}^{\infty}$ *is an infinite sequence generated by Algorithm* 1 *and* $\hat{x}$ *is an accumulation point of* $\{x_i\}_{i=0}^{\infty}$*, then* $\theta(\hat{x}) = 0$.

We will show that Algorithm 1 converges linearly under the following additional assumption.

Assumption 2. *We will assume that*

(a) *the functions* $\phi^j(\cdot, y_j)$*,* $j \in \mathbf{m}$*,* $y_j \in Y_j$*, and* $F(\cdot)$*, in (1)-(2), are twice continuously differentiable, and*

(b) *there exist constants* $0 < c \le C < \infty$*, such that for all* $j \in \mathbf{m}$*,* $y_j \in Y_j$*,* $x \in \Re^n$*, and* $h \in \Re^n$*,*

$$c\|y\|^2 \le \langle h, \phi^j_{xx}(x, y_j)h \rangle \le C\|h\|^2 \tag{38}$$

and

$$0 \le \langle h, F_{xx}(x)h \rangle \le C\|h\|^2. \tag{39}$$

The following result is obvious.

LEMMA 2 *Suppose that Assumption 1 holds. Then for any* $y, y' \in \Re^m$ *such that* $y' \geq y$,

$$F(y') - F(y) \geq c_F \sum_{j \in \mathbf{m}} (y'_j - y_j). \tag{40}$$

LEMMA 3 *Suppose that Assumptions 1 and 2 are satisfied. Then there exists a constant* $\tau > 0$ *such that for all* $x, x' \in \Re^n$ *and* $\lambda \in [0, 1]$,

$$f^0(\lambda x + (1-\lambda)x') \leq \lambda f^0(x) + (1-\lambda) f^0(x') - \tfrac{1}{2}\tau\lambda(1-\lambda)\|x - x'\|^2. \tag{41}$$

Proof: It follows from Assumption 2 that for any $x, x' \in \Re^n$ and $\lambda \in [0, 1]$,

$$f^{j,k}(\lambda x + (1-\lambda)x') \leq \lambda f^{j,k}(x) + (1-\lambda) f^{j,k}(x') - \tfrac{1}{2}c\lambda(1-\lambda)\|x - x'\|^2. \tag{42}$$

Hence, for all $j \in \mathbf{m}$,

$$\psi^j(\lambda x + (1-\lambda)x') \leq \lambda \psi^j(x) + (1-\lambda)\psi^j(x') - \tfrac{1}{2}c\lambda(1-\lambda)\|x - x'\|^2. \tag{43}$$

It follows from (40) and (43) that

$$F(\lambda\psi(x) + (1-\lambda)\psi(x')) \geq F(\psi(\lambda x + (1-\lambda)x')) + \tfrac{1}{2}c_F cm\lambda(1-\lambda)\|x - x'\|^2. \tag{44}$$

Since F is convex, we conclude from (44) that

$$\begin{aligned} &\lambda F(\psi(x)) + (1-\lambda)F(\psi(x')) \\ &\quad \geq F(\psi(\lambda x + (1-\lambda)x')) + \tfrac{1}{2}c_F cm\lambda(1-\lambda)\|x - x'\|^2. \end{aligned} \tag{45}$$

By letting $\tau = c_F cm$, we conclude that

$$\lambda f^0(x) + (1-\lambda) f^0(x') \geq f^0(\lambda x + (1-\lambda)x') + \tfrac{1}{2}\tau\lambda(1-\lambda)\|x - x'\|^2. \tag{46}$$

This proves (41). ∎

PROPOSITION 1 *Suppose that Assumptions 1 and 2 are satisfied, that all the* Y_j, $j \in \mathbf{m}$ *are of the form (4), so that problem (1)-(2) reduces to problem (35). Then any sequence* $\{x_i\}_{i=0}^{\infty}$ *constructed by Algorithm* 1, *in solving problem (35), converges to the unique minimizer* $\hat{x}$ *of* $f^0(\cdot)$.

Proof: First, by Lemma 3, the function $f^0(\cdot)$ is strongly convex, and hence it has a unique minimizer $\hat{x}$. This minimizer is also the unique stationary point of $f^0(\cdot)$. Also, since the function $f^0(\cdot)$ is strongly convex, its level sets are compact. Consequently, since the cost sequence $\{f^0(x_i)\}_{i=0}^{\infty}$ is monotone decreasing, the sequence $\{x_i\}_{i=0}^{\infty}$ must have accumulation points. It now follows from Theorem 3 that each of these accumulation points must be stationary. Since $f^0(\cdot)$ has a unique stationary point $\hat{x}$, it follows that the sequence $\{x_i\}_{i=0}^{\infty}$ converges to $\hat{x}$. ∎

THEOREM 4 *Suppose that Assumptions 1 and 2 are satisfied, that all the Y_j, $j \in \mathbf{m}$ are of the form (4), so that problem (1)-(2) reduces to problem (35), and that $\delta \in [c, C]$. If $\{x_i\}_{i=0}^{\infty}$ is a sequence constructed by Algorithm 1, in solving problem (35), then*

$$\frac{f^0(x_{i+1}) - f^0(\hat{x})}{f^0(x_i) - f^0(\hat{x})} \leq 1 - \alpha\beta \frac{c}{C^*} , \quad \forall i \in N , \tag{47}$$

where

$$C^* = C \sup_{i \in N} \max_{s \in [0,1]} \{1 + \max_{j \in \mathbf{m}} \max_{k \in \mathbf{q_j}} \|\nabla f^{j,k}(x_i + sh_i)\|^2 / \frac{\partial F}{\partial y^j}(\psi(x_i))\} . \tag{48}$$

Proof: Since by Lemma 2,

$$F(y') - F(y) \geq 0, \tag{49}$$

whenever $y' \geq y$, it follows that for any $x', x \in \Re^n$,

$$\begin{aligned} f^0(x') - f^0(x) &= F(\psi(x')) - F(\psi(x)) \\ &\geq F(\tilde{\psi}(x, x'-x)) - F(\psi(x)), \end{aligned} \tag{50}$$

where $\tilde{\psi}(x, x'-x) = (\tilde{\psi}^1(x, x'-x), ..., \tilde{\psi}^m(x, x'-x))$, with

$$\tilde{\psi}^j(x, x'-x) = \max_{k \in \mathbf{q_j}} f^{j,k}(x) + \langle \nabla f^{j,k}(x), x'-x\rangle + \tfrac{1}{2}c\|x'-x\|^2, \quad j \in \mathbf{m}. \tag{51}$$

Next, since $F(\cdot)$ is convex, for any $y, y' \in \Re^m$,

$$F(y') - F(y) \geq \langle \nabla F(y), y'-y\rangle, \tag{52}$$

which, together with (50), implies that for any $x, x' \in \Re^n$,

$$\begin{aligned} f^0(x') - f^0(x) &\geq \sum_{j \in \mathbf{m}} \frac{\partial F}{\partial y^j}(\psi(x))[\max_{k \in \mathbf{q_j}} \{f^{j,k}(x) - \psi^j(x) \\ &\quad + \langle \nabla f^{j,k}(x), x'-x\rangle + \tfrac{1}{2}c\|x'-x\|^2\}] \\ &\geq \frac{\delta}{c} \sum_{j \in \mathbf{m}} \frac{\partial F}{\partial y^j}(\psi(x))[\max_{k \in \mathbf{q}_j} \{f^{j,k}(x) - \psi^j(x) \\ &\quad + \langle \nabla f^{j,k}(x), \tfrac{c}{\delta}(x'-x)\rangle + \tfrac{1}{2}\delta\|\tfrac{c}{\delta}(x'-x)\|^2\}]. \end{aligned} \tag{53}$$

It follows directly from (53) that

$$f^0(x') - f^0(x) \geq \frac{\delta}{c}\theta(x). \tag{54}$$

Since $\hat{x}$ is a global minimizer of $f^0(\cdot)$, we deduce from (54) that for all $x \in \Re^n$,

$$f^0(\hat{x}) - f^0(x) \geq \frac{\delta}{c}\theta(x). \tag{55}$$

To complete our proof of linear convergence, we need one more inequality. By the Mean Value Theorem, for every $j \in \mathbf{m}$, $x \in \Re^n$, and $\lambda > 0$,

$$\psi^j(x + \lambda h(x)) = \max_{k \in \mathbf{q_j}} \{f^{j,k}(x) + \lambda \langle \nabla f^{j,k}(\xi_{j,k}), h(x)\rangle\}, \quad \xi_{j,k} \in [x, x + \lambda h(x)]. \tag{56}$$

Hence, for every $j \in \mathbf{m}$, $x \in \Re^n$, and $\lambda > 0$

$$\lambda \min_{k \in \mathbf{q_j}} \langle \nabla f^{j,k}(\xi_{j,k}), h(x)\rangle \leq \psi^j(x + \lambda h(x)) - \psi^j(x) \leq \lambda \max_{k \in \mathbf{q_j}} \langle \nabla f^{j,k}(\xi_{j,k}), h(x)\rangle, \tag{57}$$

which leads to the conclusion that

$$|\psi^j(x + \lambda h(x)) - \psi^j(x)| \leq \lambda \max_{k \in \mathbf{q_j}} |\langle \nabla f^{j,k}(\xi_{j,k}), h(x)\rangle|. \tag{58}$$

Because of Assumption 2, for any $\lambda > 0$ and $j \in \mathbf{m}$, and $x \in \Re^n$,

$$\begin{aligned} \psi^j(x + \lambda h(x)) - \psi^j(x) \leq \max_{k \in \mathbf{q_j}} \{ & f^{j,k}(x) - \psi^j(x) + \lambda \langle \nabla f^{j,k}(x), h(x)\rangle \\ & + \tfrac{1}{2}\lambda^2 C \|h(x)\|^2\}, \end{aligned} \tag{59}$$

which, together with Assumptions 1 and 2, and (58), implies that for any $\lambda \in (0, 1]$,

$$\begin{aligned} & F(\psi(x + \lambda h(x))) - F(\psi(x)) \\ & \quad \leq \langle \nabla F(\psi(x)), \psi(x + \lambda h(x)) - \psi(x)\rangle + \tfrac{1}{2} C \|\psi(x + \lambda h(x)) - \psi(x)\|^2 \\ & \quad \leq \lambda \left\{ \sum_{j \in \mathbf{m}} \frac{\partial F}{\partial y^j}(\psi(x)) [\max_{k \in \mathbf{q_j}} \{f^{j,k}(x) - \psi^j(x) + \langle \nabla f^{j,k}(x), h(x)\rangle \right. \\ & \qquad \left. + \frac{\lambda C}{2\delta} \delta \|h(x)\|^2\}] + \frac{\lambda C}{2\delta} \max_{k \in \mathbf{q_j}} \|\nabla f^{j,k}(\xi^{j,k})\|^2 \delta \|h(x)\|^2 \right\}. \end{aligned} \tag{60}$$

Consequently, if

$$\frac{\lambda C}{\delta} [1 + \max_{k \in \mathbf{q_j}} \|\nabla f^{j,k}(\xi^{j,k})\|^2 / \frac{\partial F}{\partial y^j}(\psi(x))] \leq 1, \tag{61}$$

then we conclude from (60) that

$$f^0(x + \lambda h(x)) - f^0(x) \leq \lambda \theta(x). \tag{62}$$

Now suppose that $\{x_i\}_{i=0}^{\infty}$ is a sequence constructed by Algorithm 1 from an initial point $x_0 \in L^* := \{x \in \Re^n |\ f(x) \leq f(x^0)\}$. Then, because by construction the sequence

$\{f^0(x_i)\}_{i=0}^{\infty}$ is monotone decreasing, it follows that the whole sequence $\{x_i\}_{i=0}^{\infty}$ is in L^*. Hence it follows from (48) and (62) that for every $i \in N$,

$$\lambda_i \geq \frac{\delta\beta}{C^*}. \tag{63}$$

It follows from (36) and (55) that for every $i \in \mathcal{N}$

$$f^0(x_{i+1}) - f^0(x_i) \leq \lambda_i \alpha c \delta^{-1}(f^0(\hat{x}) - f^0(x_i)). \tag{64}$$

So,

$$\frac{f^0(x_{i+1}) - f^0(\hat{x})}{f^0(x_i) - f^0(\hat{x})} \leq 1 - \lambda_i \alpha c \delta^{-1}. \tag{65}$$

Hence from (63) and (65) we deduce that

$$\frac{f^0(x_{i+1}) - f^0(\hat{x})}{f^0(x_i) - f^0(\hat{x})} \leq 1 - \alpha\beta\frac{c}{C^*}. \tag{66}$$

This completes our proof. ■

The following consequence of Theorem 4 should be obvious:

COROLLARY 1 *Suppose Assumptions 1 and 2 are satisfied. Let $\hat{x}$ be the unique minimizer of $f^0(\cdot)$, and let*

$$\hat{C} = C\{1 + \max_{j \in \mathbf{m}} \max_{k \in \mathbf{q_j}} \|\nabla f^{j,k}(\hat{x})\|^2 / \frac{\partial F}{\partial y^j}(\psi(\hat{x}))\}\ . \tag{67}$$

If $\delta \in [c, \hat{C}]$, then, for any sequence $\{x_i\}_{i=0}^{\infty}$ constructed by Algorithm 1

$$\overline{\lim} \frac{f^0(x_{i+1}) - f^0(\hat{x})}{f^0(x_i) - f^0(\hat{x})} \leq 1 - \alpha\beta\frac{c}{\hat{C}}\ . \tag{68}$$

4. An Algorithm for Solving Generalized Semi-Infinite Min-Max Problems

We are now ready to tackle the generalized semi-infinite min-max problems defined in (1)-(2). Such problems can be solved only by discretization techniques. We will use discretizations that result in consistent approximations (as defined in Section 3.3 of [17]) and use them in conjunction with a master algorithm that calls Algorithm 1 as a subroutine. We will see that under a reasonable assumption, the resulting algorithm retains the rate of convergence of Algorithm 1.

4.1. Consistent Approximations

Let N_0 be a strictly positive integer, and, for $N \in N_0 := \{N_0, N_0+1, N_0+2, ...\}$, let $Y_{j,N}$ be finite cardinality subsets of Y_j, $j \in \mathbf{m}$, such that $Y_{j,N} \subset Y_{j,N+1}$ for all N and the closure of the set $\lim Y_{j,N}$ is equal to Y_j, $j \in \mathbf{m}$. Then we define the family of approximating problems $\mathbf{P_N}$, $N \in N_0$, as follows:

$\mathbf{P_N}$

$$\min_{x \in \Re^n} f_N^0(x), \tag{69}$$

where

$$f_N^0(x) := F(\psi_N(x)) \tag{70}$$

where $\psi_N(x) = (\psi_N^1(x), ..., \psi_N^m(x))$, and for $j \in \mathbf{m}$,

$$\psi_N^j(x) = \max_{y_j \in Y_{j,N}} \phi^j(x, y_j)\ . \tag{71}$$

It should be clear that the approximating problems $\mathbf{P_N}$ are of the form (35) and that one can define optimality functions $\theta_N(\cdot)$ for them of the form (4). We will refer to the original problem (1) as $\mathbf{P}$.

Definition 2. [17] We will say that the pairs $(\mathbf{P_N}, \theta_\mathbf{N})$, in the sequence $\{(\mathbf{P_N}, \theta_\mathbf{N})\}_{\mathbf{N} \in \mathbf{N_0}}$ are consistent approximations for the pair $(\mathbf{P}, \theta)$, if the problems $\mathbf{P_N}$ epi-converge to $\mathbf{P}$, (i.e., the epigraphs of the $f_N^0(\cdot)$ converge to the epigraph of $f^0(\cdot)$ in the sense defined in Definition 5.3.6 in [17]), and for any infinite sequence $\{x_N\}_{N \in K}$, $K \subset N_0$, such that $x_N \to x$, $\overline{\lim}\theta_N(x_N) \leq \theta(x)$.

Assumption 3. We will assume as follows:

(a) *For every $N \in N_0$, the problem (69) has a solution.*

(b) *There exists a strictly positive valued, strictly monotone decreasing function $\Delta : N \to \Re$, such that $\Delta(N) \to 0$, as $N \to \infty$, and a $K < \infty$, such that for every $N \geq N_0$, $j \in \mathbf{m}$, and $y \in Y_j$, there exists a $y' \in Y_{j,N}$ such that*

$$\|y - y'\| \leq K\Delta(N). \tag{72}$$

For example, if for all $j \in \mathbf{m}$, Y_j is the unit cube in $\Re^{m_j}$, i.e., $Y_j = I^{m_j}$, with $I := [0, 1]$, then we can define $Y_{j,N} = I_N^{m_j}$, where

$$I_N = \{0, 1/a(N), 2/a(N), ..., (a(N) - 1)/a(N), 1\},$$

with $a(N) := 2^{N-N_0}$. In this case it is easy to see that $\Delta(N) = 1/a(N)$ and $K = \frac{1}{2}\max_{j \in \mathbf{m}} m_j{}^{(1/m_j)}$. Similar constructions can be obtained for other polyhedral sets.

We infer from (16) that the optimality functions $\theta_N(\cdot)$, for the problems $\mathbf{P_N}$ have the following form:

$$\theta_N(x) := \min_{h\in\Re^n} \tilde{f}^0_N(x, x+h) - \sum_{j\in\mathbf{m}} \frac{\partial F}{\partial y^j}(\psi_N(x))\psi^j_N(x) \tag{73}$$

where

$$\tilde{f}^0_N(x, x+h) \; = \sum_{j\in\mathbf{m}} \frac{\partial F}{\partial y^j}(\psi_N(x)) \max_{y_j\in Y_{j,N}} [\phi^j(x, y_j) + \langle \nabla_x\phi^j(x, y_j), h\rangle + \tfrac{1}{2}\delta\|h\|^2]. \tag{74}$$

Since the cardinality of the sets $Y_{j,N}$ is finite, it is obvious that the $\theta_N(x)$ can be evaluated, using the dual form (30), by means of a quadratic programming code.

LEMMA 4 *Suppose that Assumptions 1 and 3 are satisfied, and that for all $N \in N_0$, $f^0_N(\cdot)$ is defined by (70) and $\theta_N(\cdot)$ by (73). Let $S \subset \Re^n$ be a bounded subset and let $L < \infty$ be a Lipschitz constant valid for the functions $\phi^j(\cdot,\cdot)$ and $\nabla_x\phi^j(\cdot,\cdot)$ on $S \times Y_j$, $j \in \mathbf{q}$. Then there exists a constant $C_S < \infty$ such that for all $x \in S, N \in N_0$,*

$$|f^0_N(x) - f^0(x)| \le C_S\Delta(N), \tag{75}$$

and

$$|\theta_N(x) - \theta(x)| \le C_S\Delta(N), \tag{76}$$

provided that both $\theta_N(x)$, $N \in N_0$, and $\theta(x)$ are defined using the same parameter $\delta > 0$.

Proof: First, because $Y_{j,N} \subset Y_j$, $j \in \mathbf{m}$, we always have that

$$\psi^j_N(x) \le \psi^j(x). \tag{77}$$

It now follows from Lemma 2 that

$$f^0_N(x) \le f^0(x). \tag{78}$$

Next, for any $x \in S$, and each $j \in \mathbf{m}$, there must exist a $y_{j,x} \in Y_j$ such that

$$\psi^j(x) = \phi^j(x, y_{j,x})\ . \tag{79}$$

By Assumption 3, for each $j \in \mathbf{m}$, there exits $y'_{j,x} \in Y_{j,N}$ such that $\|y'_{j,x} - y_{j,x}\| \le K\Delta(N)$. Hence, for each $j \in \mathbf{m}$, we have that

$$\psi^j_N(x) \ge \phi^j(x, y'_{j,x}) \ge \phi^j(x, y_{j,x}) - LK\Delta(N) = \psi^j(x) - LK\Delta(N)\ . \tag{80}$$

It therefore follows from Lemma 2 that

$$f^0_N(x) = F(\psi_N(x)) \ge F(\psi^1(x) - LK\Delta(N), ..., \psi^m(x) - LK\Delta(N)). \tag{81}$$

Since $\psi(\cdot)$ is continuous and $\nabla F(\cdot)$ is Lipschitz continuous, there exists a constant $0 < L_1 < \infty$ such that for any $x \in S$, and each $j \in \mathbf{q}$,

$$|\psi^j(x)| \le L_1, \quad \left|\frac{\partial F}{\partial y^j}(\psi_N(x))\right| \le L_1, \quad \left|\frac{\partial F}{\partial y^j}(\psi(x))\right| \le L_1 \tag{82}$$

and

$$\left|\frac{\partial F}{\partial y^j}(\psi_N(x)) - \frac{\partial F}{\partial y^j}(\psi(x))\right| \le L_1|\psi_N(x) - \psi(x)|. \tag{83}$$

Making use of the Mean Value Theorem, (81), and (82), we find that

$$\begin{aligned} f_N^0(x) &= F(\psi_N(x)) \\ &\ge F(\psi(x)) + \sum_{j\in\mathrm{m}} \frac{\partial F}{\partial y^j}(\xi)(-LK\Delta(N)) \\ &\ge f^0(x) - mL_1LK\Delta(N), \end{aligned} \tag{84}$$

where $\xi \in \Re^m$ is such that for each $j \in \mathbf{m}$,

$$\xi^j \in [\psi^j(x) - LK\Delta(N), \psi^j(x)].$$

It now follows from (78) and (84) that (75) holds with $C_S = mL_1LK$.

Next, we turn to (76). For any $x, h \in \Re^n$, let

$$\tilde{f}^0(x, x+h) := \sum_{j\in\mathbf{m}} \frac{\partial F}{\partial y^j}(\psi(x)) \max_{y_j\in Y_j}[\phi^j(x, y_j) + \langle \nabla_x\phi^j(x, y_j), h\rangle + \tfrac{1}{2}\delta\|h\|^2], \tag{85}$$

and for any $N \in N_0$, $x, h \in \Re^n$, let $\tilde{f}_N^0(x, x+h)$ be defined by (74). Then, by inspection, $\tilde{f}_N^0(x, x+h) \le \tilde{f}^0(x, x+h)$ always holds. Therefore, given any $h' \in \Re^n$,

$$\min_{h\in\Re^n} \tilde{f}_N^0(x, x+h) \le \tilde{f}_N^0(x, x+h') \le \tilde{f}^0(x, x+h'), \tag{86}$$

and hence

$$\min_{h\in\Re^n} \tilde{f}_N^0(x, x+h) \le \min_{h'\in\Re^n} \tilde{f}^0(x, x+h'). \tag{87}$$

Since

$$\theta(x) = \min_{h\in\Re^n} \tilde{f}^0(x, x+h) - \sum_{j\in\mathbf{m}} \frac{\partial F}{\partial y^j}(\psi(x))\psi^j(x),$$

and

$$\theta_N(x) = \min_{h\in\Re^n} \tilde{f}_N^0(x, x+h) - \sum_{j\in\mathbf{m}} \frac{\partial F}{\partial y^j}(\psi_N(x))\psi_N^j(x),$$

we conclude, making use of (87), (80), (82), and (83), that

$$\begin{aligned}\theta_N(x) &\le \theta(x) + \sum_{j\in\mathbf{m}} \frac{\partial F}{\partial y^j}(\psi(x))\psi^j(x) - \sum_{j\in\mathbf{m}} \frac{\partial F}{\partial y^j}(\psi_N(x))\psi_N^j(x) \\ &= \theta(x) + \sum_{j\in\mathbf{m}} \frac{\partial F}{\partial y^j}(\psi_N(x))[\psi^j(x) - \psi_N^j(x)] \\ &\quad + \sum_{j\in\mathbf{m}} \left[\frac{\partial F}{\partial y^j}(\psi(x)) - \frac{\partial F}{\partial y^j}(\psi_N(x))\right]\psi^j(x) \\ &\le \theta(x) + C\Delta(N)\,, \end{aligned} \tag{88}$$

with $C = m(L_1 + L_1^2)LK$.

Next, referring to Theorem 2, (23) and (29), we see that for any $x \in \Re^n$, and any $N \in N$, $h_N(x)$, the minimizer of (73), satisfies

$$h_N(x) \in -\gamma_N(x)^{-1} \sum_{j\in\mathbf{m}} a_N^j(x)\mathrm{conv}_{y_j\in Y_{j,N}}\{\nabla_x\phi^j(x, y_{j,N})\}\,, \tag{89}$$

where

$$a_N(x) := (\frac{\partial F}{\partial y^1}(\psi_N(x)), ..., \frac{\partial F}{\partial y^m}(\psi_N(x))), \tag{90}$$

and

$$\gamma_N(x) := \delta \sum_{j\in\mathbf{m}} a_N^j(x). \tag{91}$$

Hence, since the gradients $\nabla_x\phi^j(\cdot,\cdot)$ are continuous, and since the set S and the sets Y_j are bounded, there exists a $\kappa < \infty$ such that $\|h_N(x)\| \le \kappa$ for all $x \in S$, and $N \in N$. For any $x \in \Re^n$ and each $j \in \mathbf{m}$, let $y_{j,x} \in Y_{j,N}$ be such that

$$\begin{aligned}\tilde{f}^0(x, x + h_N(x)) &= \sum_{j\in\mathbf{m}} \frac{\partial F}{\partial y^j}(\psi(x))[\phi^j(x, y_{j,x}) \\ &\quad + \langle \nabla_x\phi^j(x, y_{j,x}), h_N(x)\rangle + \tfrac{1}{2}\delta\|h_N(x)\|^2]. \end{aligned} \tag{92}$$

By Assumption 3, for each $j \in \mathbf{m}$, there exists a $y'_{j,x} \in Y_j$ such that $\|y'_{j,x} - y_{j,x}\| \le K\Delta(N)$. Hence, because (82) holds, and because

$$\|f^{j,j_x}(x, y_{j,j_x}) - f^{j,j_x}(x, y'_{j,j_x})\| \le LK\Delta(N), \tag{93}$$

and

$$\|\langle \nabla_x f^{j,j_x}(x, y_{j,j_x}), h_N(x)\rangle - \langle \nabla_x f^{j,j_x}(x, y'_{j,j_x}), h_N(x)\rangle\| \le LK\Delta(N)\kappa, \tag{94}$$

we conclude that

$$\tilde{f}_N^0(x, x + h_N(x)) \geq \sum_{j \in \mathbf{m}} \frac{\partial F}{\partial y^j}(\psi(x))[\phi^j(x, y'_{j,x}) + \langle \nabla_x \phi^j(x, y'_{j,x}), h_N(x) \rangle$$

$$+\tfrac{1}{2}\delta \|h_N(x)\|^2]$$

$$\geq \tilde{f}^0(x, x + h_N(x)) - (mL_1LK + \kappa m L_1 LK)\Delta(N). \tag{95}$$

Hence, since $\theta_N(x) = \tilde{f}_N^0(x, x + h_N(x)) - \sum_{j \in \mathbf{m}} \frac{\partial F}{\partial y^j}(\psi_N(x))\psi_N^j(x)$, $\frac{\partial F}{\partial y^j}(\psi_N(x)) > 0$, and $\psi^j(x) \geq \psi_N^j(x)$, $j \in \mathbf{m}$,

$$\theta_N(x) \geq \theta(x) + \sum_{j \in \mathbf{m}} \frac{\partial F}{\partial y^j}(\psi(x))\psi^j(x) - \sum_{j \in \mathbf{m}} \frac{\partial F}{\partial y^j}(\psi_N(x))\psi_N^j(x)$$

$$-(mL_1LK + \kappa m L_1 LK)\Delta(N)$$

$$= \theta(x) + \sum_{j \in \mathbf{m}} \frac{\partial F}{\partial y^j}(\psi_N(x))[\psi^j(x) - \psi_N^j(x)]$$

$$+ \sum_{j \in \mathbf{m}} \left[\frac{\partial F}{\partial y^j}(\psi(x)) - \frac{\partial F}{\partial y^j}(\psi_N(x))\right]\psi^j(x)$$

$$-(mL_1LK + \kappa m L_1 LK)\Delta(N)$$

$$\geq \theta(x) + \sum_{j \in \mathbf{m}} \left[\frac{\partial F}{\partial y^j}(\psi(x)) - \frac{\partial F}{\partial y^j}(\psi_N(x))\right]\psi^j(x)$$

$$-(mL_1LK + \kappa m L_1 LK)\Delta(N). \tag{96}$$

Finally, from (87) and (96), we deduce that

$$\theta_N(x) \geq \theta(x) - [m(L_1 + L_1^2)LK + \kappa m L_1 LK]\Delta(N). \tag{97}$$

It follows by inspection from (88) and (97) that (76) holds with $C_S = m(L_1 + L_1^2)LK + \kappa m L_1 LK$. Obviously, this value of C_S is also good for (75), and therefore our proof is completed. ■

COROLLARY 2 *The pairs* $(\mathbf{P_N}, \theta_\mathbf{N})$*, defined by (69)-(71) and (73) respectively, are consistent approximations for the pair* $(\mathbf{P}, \theta)$*, defined by (1)-(2), (16), respectively.*

Proof: Let $x \in \Re^n$ be arbitrary, and let $\{x_N\}_{N \in N_0}$ be a sequence in $\Re^n$ that converges to x. Then, since $\{x_N\}_{N \in N_0}$ is bounded, it follows from Lemma 4 that there exists a constant $C^* < \infty$ such that

$$\lim |f_N^0(x_N) - f^0(x)| \leq \lim\{|f_N^0(x_N) - f^0(x_N)| + |f^0(x_N) - f^0(x)|\}$$
$$\leq \lim\{C^*\Delta(N) + |f_N^0(x_N) - f^0(x_N)|\} = 0\,. \quad (98)$$

Referring to Theorem 3.3.2 in [17], we see that (98) implies epi-convergence.

Similar reasoning leads to the conclusion that $\theta_N(x_N) \to \theta(x)$, as $N \to \infty$, which completes our proof. ■

Note that it follows directly from (75) that if $\hat{x}$ is a minimizer of $f^0(\cdot)$ and $\hat{x}_N$ is a minimizer of $f_N^0(\cdot)$, then $|f_N^0(\hat{x}_N) - f^0(\hat{x})| \leq C_S\Delta(N)$.

4.2. Rate Preserving Algorithm

Algorithm 2 *(Solves Problem (1))*

Parameters. $\alpha, \beta \in (0,1)$, $\delta > 0$, $D > 0$, $\sigma > 1$.

Data. $x_0 \in \Re^n$, $N_0 \in \mathcal{N}$.

Step 0. *Set* $i = 0$, $N = N_0$.

Step 1. *Compute the* optimality function value $\theta_N(x_i)$ *according to (73), i.e.,*

$$\theta_N(x_i) = \min_{h\in\Re^n} \sum_{\in \mathbf{m}} \frac{\partial F}{\partial y^j}(\psi_N(x_i)) \max_{y_j\in Y_{j,N}} [\phi^j(x_i,y_j)$$
$$+\langle \nabla_x \phi^j(x_i,y_j), h\rangle + \tfrac{1}{2}\delta\|h\|^2] - \sum_{j\in\mathbf{m}} \frac{\partial F}{\partial y^j}(\psi_N(x_i))\psi_N^j(x_i) \quad (99)$$

and the corresponding search direction $h_N(x_i)$ *according to*

$$h_N(x_i) = \arg\min_{h\in\Re^n} \sum_{j\in\mathbf{m}} \frac{\partial F}{\partial y^j}(\psi_N(x_i)) \max_{k\in\mathbf{q_j}} \max_{y_j\in Y_{j,N}} [\phi^j(x_i,y_j)$$
$$+\langle \nabla_x \phi^j(x_i,y_j), h\rangle + \tfrac{1}{2}\delta\|h\|^2]. \quad (100)$$

Step 2. *If*

$$D\Delta(N) \leq |\theta_N(x_i)|^\sigma, \quad (101)$$

set $\theta_i = \theta_N(x_i)$, $h_i = h_N(x_i)$, $N_i = N$, *and go to Step 3. Else replace* N *by* $N+1$ *and go to Step 1.*

Step 3. *Compute the step-size*

$$\lambda_i = \max_{k \in \mathcal{N}} \{\beta^k \mid f^0_{N_i}(x_i + \beta^k h_i) - f^0_{N_i}(x_i) - \beta^k \alpha \theta_i \leq 0\}, \tag{102}$$

and go to Step 4.

Step 4. *Set*

$$x_{i+1} = x_i + \lambda_i h_i, \tag{103}$$

replace i by $i+1$ and go to Step 1.

Remark.

(a) It follows from Lemma 4 that $\theta_N(x_i) \to \theta(x_i)$, as $N \to \infty$. Hence, whenever $\theta(x_i) \neq 0$, the loop consisting Step 1 and Step 2 of Algorithm 2 yields a finite discretization parameter N_i. For simplicity, we will assume that Algorithm 2 does not produce an iterate x_i such that $\theta(x_i) = 0$.

(b) Note that the work needed to compute x_i by Algorithm 2 increases with the iteration number i.

LEMMA 5 *Suppose that Assumption 3 is satisfied, and that Algorithm 2 has constructed a sequence $\{x_i\}_{i=0}^{\infty}$ together with the corresponding sequence of discretization parameters $\{N_i\}_{i=0}^{\infty}$. If the sequence $\{x_i\}_{i=0}^{\infty}$ has at least one accumulation point, then $N_i \to \infty$, as $i \to \infty$.*

Proof: For the sake of contradiction, suppose that the sequence $\{N_i\}_{i=0}^{\infty}$ is bounded. Then, because $\{N_i\}_{i=0}^{\infty}$ is a monotonically increasing sequence of integers, there exists an $i_0 \in N$, such that for all $i \geq i_0$, $N_i = N_{i_0} =: N^*$. Hence for $i \geq i_0$, the construction of the sequence $\{x_i\}_{i=0}^{\infty}$ is carried out by Algorithm 1 applied to problem (69) with $N = N^*$. Furthermore, it follows from (101) that there exists an $\varepsilon > 0$, such that $\theta_i = \theta_{N^*}(x_i) \leq -\varepsilon$ for all $i \geq i_0$. Since by assumption the sequence $\{x_i\}_{i=0}^{\infty}$ has at least one accumulation point $\hat{x}$, with $x_i \to^K \hat{x}$, as $i \to \infty$, for some infinite sequence $K \subset N$. It now follows from Theorem 3 that $\theta_{N^*}(\hat{x}) = 0$ and from the continuity of $\theta_{N^*}(\cdot)$, that $\theta_{N^*}(x_i) \to^K \theta_{N^*}(\hat{x}) = 0$, as $i \to \infty$, which contradicts the previous finding, and completes our proof. ■

By applying Theorem 3.3.23 of [17] or Theorems in Section 5 of [16] to Algorithm 2, we obtain the following global convergence theorem:

THEOREM 5 *Suppose that Assumptions 1 and 3 are satisfied, and that Algorithm 2 has constructed a bounded sequence $\{x_i\}_{i=0}^{\infty}$. Then every accumulation point $\hat{x}$ of $\{x_i\}_{i=0}^{\infty}$ satisfies $\theta(\hat{x}) = 0$.*

THEOREM 6 *Suppose that Assumptions 1, 2, and 3 are satisfied, and that $\delta \in [c, C]$. Then*

(i) *any bounded sequence $\{x_i\}_{i=0}^{\infty}$, generated by Algorithm 2, converges to the unique solution $\hat{x}$ of (1.3), and, in addition,*

(ii) *we have that*

$$\overline{\lim}\ \frac{f^0(x_{i+1}) - f^0(\hat{x})}{f^0(x_i) - f^0(\hat{x})} \leq 1 - \alpha\beta\frac{c}{C^*}\ , \tag{104}$$

where

$$C^* = C \sup_{i \in N} \left[1 + \max_{j \in \mathbf{m}} \max_{\lambda \in [0,1]} \max_{y_j \in Y_j} \|\nabla_x \phi^j(\hat{x}, y_j))\|^2 / \frac{\partial F}{\partial y^j}(\psi_{N_i}(\hat{x})) \right]\ . \tag{105}$$

Proof: *(i)* First we note that it follows from Assumption 2 that the function $f^0(\cdot)$ has a unique stationary point $\hat{x}$. Next, since the sequence $\{x_i\}_{i=0}^{\infty}$ is bounded it must have accumulation points, and finally, by Theorem 5, these accumulation points must be stationary. Hence the sequence $\{x_i\}_{i=0}^{\infty}$ converges to $\hat{x}$.

(ii) We deduce from the optimality of $\hat{x}_N$ and Lemma 4 that for some $\kappa < \infty$,

$$f_N^0(\hat{x}_N) \leq f_N^0(\hat{x}) \leq f^0(\hat{x}) + \kappa\Delta(N). \tag{106}$$

Similarly, we see that

$$f^0(\hat{x}) \leq f^0(\hat{x}_N) \leq f_N^0(\hat{x}_N) + \kappa\Delta(N). \tag{107}$$

Hence

$$|f^0(\hat{x}) - f_N^0(\hat{x}_N)| \leq \kappa\Delta(N). \tag{108}$$

Next, because $\delta \in [c, C]$, an examination of the proof of Theorem 4 shows that (47) must hold for all $i \in N$, on the sequence $\{x_i\}_{i=0}^{\infty}$ constructed by Algorithm 2, i.e., for all $i \in N$,

$$f_{N_i}^0(x_{i+1}) - f_{N_i}^0(\hat{x}_{N_i}) \leq [1 - \alpha\beta\frac{c}{C_i}](f_{N_i}^0(x_i) - f_{N_i}^0(\hat{x}_{N_i}))\ , \tag{109}$$

where

$$C_i := C\left[1 + \max_{j \in \mathbf{m}} \max_{\lambda \in [0,1]} \max_{y_j \in Y_j} \|\nabla_x \phi^j(x_i, y_j)(x_i + \lambda h_{N_i}(x_i))\|^2 / \frac{\partial F}{\partial y^j}(\psi_{N_i}(x_i))\right] \tag{110}$$

is finite. It therefore follows from Lemma 4 and (108) that

$$f^0(x_{i+1}) - f^0(\hat{x}) \leq [1 - \alpha\beta\frac{c}{C_i}]\left(f^0(x_i) - f^0(\hat{x})\right) + 4\kappa\Delta(N_i). \tag{111}$$

Next, we conclude from the proof of Theorem 4 that, for any $\lambda \in (0, 1]$,

$$\begin{aligned} &F(\psi_{N_i}(x_i + \lambda h_{N_i}(x_i))) - F(\psi_{N_i}(x_i)) \\ &\leq \lambda \left\{ \sum_{j \in \mathbf{m}} \frac{\partial F}{\partial y^j}(\psi_{N_i}(x_i)) \max_{y_j \in Y_{j,N_i}} [\phi^j(x_i, y_j) \right. \\ &\left. -\psi_{N_i}^j(x_i) + \langle \nabla_x \phi^j(x, y_j), h\rangle + \frac{\lambda C_i}{2\delta}\delta\|h_{N_i}(x_i)\|^2] \right\}\ . \end{aligned} \tag{112}$$

Hence, if $\lambda C/\delta \leq 1$, then $\lambda C_i/\delta \leq 1$ and

$$F(\psi_{N_i}(x_i + \lambda h_{N_i}(x_i))) - F(\psi_{N_i}(x_i)) \leq \lambda \theta_{N_i}(x_i). \tag{113}$$

We therefore conclude from (113) and (102) of Algorithm 2 that

$$\lambda_i \geq \frac{\delta\beta}{C_i}. \tag{114}$$

It now follows from (114) and (102) that

$$f^0_{N_i}(\hat{x}_{N_i}) - f^0_{N_i}(x_i) \leq f^0_{N_i}(x_{i+1}) - f^0_{N_i}(x_i) \leq \frac{\delta\alpha\beta}{C_i}\theta_{N_i}(x_i)\,. \tag{115}$$

Next, making use of Lemma 4, (108), (115), the fact that $\theta_i = \theta_{N_i}(x_i)$ by definition, and that $-\theta_i \geq [D\Delta(N_i)]^{1/\sigma}$ by construction in (101), we conclude that

$$\begin{aligned}
\frac{C_i}{\delta\beta}[f^0(x_i) - f^0(\hat{x})] &\geq \frac{C_i}{\delta\alpha\beta}[f^0_{N_i}(x_i) - f^0_{N_i}(\hat{x}_{N_i})] - \frac{2C_i\kappa}{\delta\beta}\Delta(N_i) \\
&\geq -\theta_i - \frac{2C_i\kappa}{\delta\beta}\Delta(N_i) \\
&\geq [D\Delta(N_i)]^{1/\sigma} - \frac{2C^*\kappa}{\delta\beta}\Delta(N_i) \\
&= [D\Delta(N_i)]^{1/\sigma}[1 - 2C_i\kappa/(\delta\beta D^{1/\sigma})\Delta(N_i)^{(\sigma-1)/\sigma}]\,. \qquad (116)
\end{aligned}$$

Since $\sigma > 1$ and since by Lemma 5, $N_i \to \infty$, as $i \to \infty$, we conclude that there exists an i_0 such that for all $i \geq i_0$,

$$\frac{C_i}{\delta\alpha\beta}[f^0(x_i) - f^0(\hat{x})] \geq \tfrac{1}{2}[D\Delta(N_i)]^{1/\sigma}, \tag{117}$$

and hence that

$$\frac{C_i}{\delta\alpha\beta}[D\Delta(N_i)]^{(\sigma-1)/\sigma}[f^0(x_i) - f^0(\hat{x})] \geq \tfrac{1}{2}D\Delta(N_i)\,. \tag{118}$$

It now follows from (111) and (118) that for all $i \geq i_0$,

$$f^0(x_{i+1}) - f^0(x_i) \leq \left[1 - \frac{\alpha\beta c}{C_i} + \frac{8\kappa C_i \Delta(N_i)^{(\sigma-1)/\sigma}}{\delta\beta D}\right][f^0(x_i) - f^0(\hat{x})]\,. \tag{119}$$

Since $C_i \to \hat{C}$, as $i \to \infty$, (104) follows from (119) and the fact that, by Lemma 5, $N_i \to \infty$, as $i \to \infty$. This completes our proof. ■

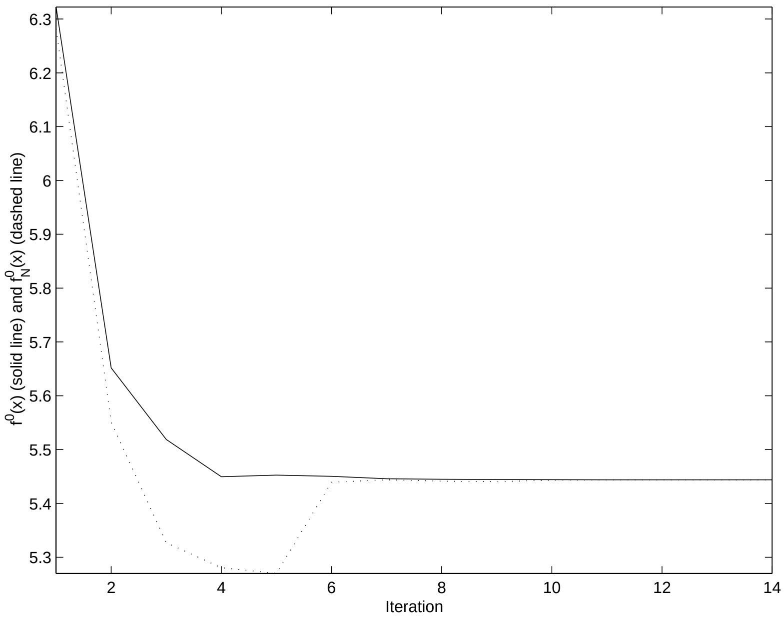

Figure 1. Numerical results for Example-1

5. Some Numerical Results

We now present some numerical results that illustrate the behavior of the algorithm proposed in Section 4 for generalized semi-infinite programming problems. The algorithm was implemented in Matlab and run on a DEC Alpha Server 8200. Throughout the computational experiments, the parameters used in the algorithm were $\alpha = 0.5, \beta = 0.85, \delta = 1, D = 0.01$, and $\sigma = 1.1$.

Example-1. In this case, $f^0(x) = F(\psi^1(x), \psi^2(x))$, with $x = (x^1, x^2) \in \Re^2$, $F(z) = z^1 + z^2$, with $z = (z^1, z^2) \in \Re^2$, and

$$\psi^1(x) = \max_{t \in Y_1} \{t^2 - (tx^1 + e^t x^2) + (x^1 + x^2)^2 + (x^1)^2 + (x^2)^2\}$$

and

$$\psi^2(x) = \max_{t \in Y_2} \{(t-1)^2 + 0.5(x^1 + x^2)^2 - 2t(x^1 + x^2) + 0.5[(x^1)^2 + (x^2)^2]\},$$

where $Y_1 = [0, 2]$ and $Y_2 = [-1, 1]$. We used the starting point $x_0 = [1, 1]$.

Figure 1 displays both the exact value of $f^0(x_i)$ and the current approximating value $f^0_{N_i}(x_i)$, constructed by the algorithm at iteration i. Note the expected sudden increase in the approximating value $f^0_{N_i}(x_i)$ when the discretization level is refined by the master algorithm. The observed rate of convergence is linear.

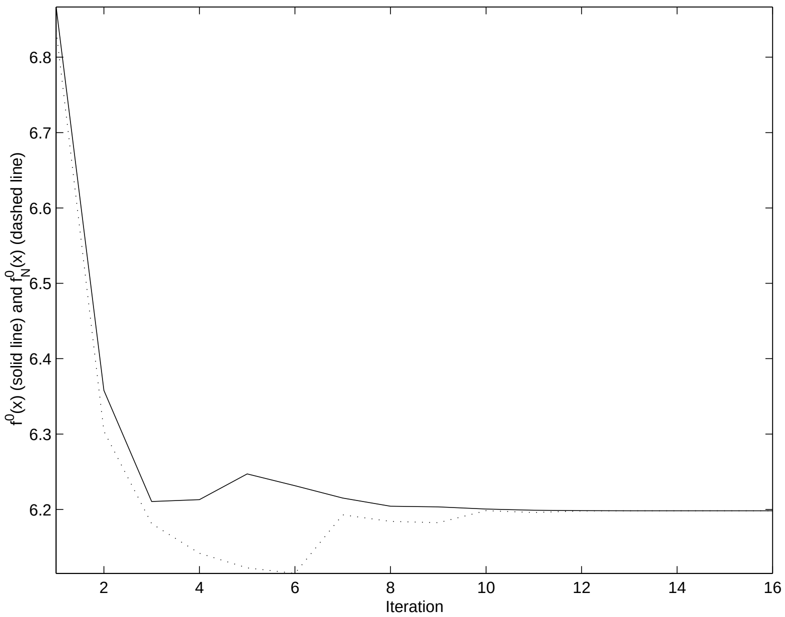

Figure 2. Numerical results for Example-2

Example-2. In this case, the functions $f^0(\cdot)$, $\psi^1(\cdot)$, and $\psi^2(\cdot)$ are also defined as in Example-1, but $F(\cdot)$ is defined by

$$F(z) = 0.5(z^1 + \sqrt{(z^1)^2 + 4}) + \ln(1 + e^{z^2}), \quad z = (z^1, z^2) \in \Re^2.$$

Figure 2 displays both the exact value of $f^0(x_i)$ and the current approximating value $f^0_{N_i}(x_i)$, constructed by the algorithm at iteration i. Again note the expected sudden increase in the approximating value $f^0_{N_i}(x_i)$ when the discretization level is refined by the master algorithm. The observed rate of convergence is linear.

6. Conclusion

We have presented an algorithm for solving semi-infinite generalized min-max problems of the form (1) and (2), which we have obtained by making use of the Kiwiel algorithm in [7] and of the concepts underlying the construction the Polak-He PPP Rate-Preserving Algorithm in [15] (see also Algorithm 3.4.9 in [17], respectively. The construction of the algorithm depends on the cost function having a subgradient and hence Assumption 2 is essential.

Our numerical results are consistent with our theoretical predictions and show that the algorithm was efficient in solving our two generalized semi-infinite min-max test problems.

Acknowledgments

The authors would like to thank Ismail Khalil for his help in preparing the computational results, as well as the two anonymous referees for their helpful comments.

Notes

1. We denote the components of a vector by superscripts, and elements of a set by subscripts.
2. These transformations result in a smooth problem with more variables than in the nonsmooth problem. Although not documented in the literature, it is widely observed that these transformations induce considerable ill-conditioning in the resulting smooth problem because they introduce arbitrary scaling, and hence lead to less efficient solution of the original nonsmooth problems than using algorithms that exploit problem structure.
3. We were unable to show that the general case of the Kiwiel algorithm in [7, 10] converges linearly.
4. In [7], Kiwiel defines a family of algorithms with the term $\sum_{j\in\mathbf{m}} \frac{\partial F}{\partial y^j}(\psi(x))\frac{1}{2}\delta\|h\|^2$ replaced by $\frac{1}{2}\langle h, Bh\rangle$, with B any symmetric positive definite matrix. However, our rate of convergence analysis does not appear to carry over to the entire family of algorithms even when Assumptions 1 and 2 are satisfied.

References

1. D. P. Bertsekas, "Nondifferentiable optimization via approximation," Mathematical Programming Study, vol. 3, pp. 1–25, 1975.
2. V. F. Dem'yanov, "Quasidifferentiable functions: necessary conditions and descent directions", Mathematical Programming Study, vol. 29, pp. 20–43, 1986.
3. V. F. Dem'yanov, S. Gamidov, and I. T. Sivelina, "An algorithm for minimizing a certain class of quasidifferentiable functions", Mathematical Programming Study, vol. 29, pp. 74–84, 1986.
4. V. F. Dem'yanov and A. M. Rubinov, Quasidifferential Calculus, Optimization Software Inc.: New York, 1986.
5. G. Di Pillo, L. Grippo, and S. Lucidi, "A smooth transformation of the generalized minimax problem", Journal of Optimization Theory and Applications vol. 95, pp. 1–24, 1997.
6. M. Fukushima, "A descent algorithm for nonsmooth convex optimization," Mathematical Programming, vol. 30, pp. 163–175, 1984.
7. K. C. Kiwiel, "A quadratic approximation method for minimizing a class of quasidifferentiable functions," Numerische Mathematik, vol. 45, 411–430, 1984.
8. K.C. Kiwiel, "A linearization method for minimizing certain quasidifferentiable functions," Mathematical Programming Study, vol. 29, pp. 85–94, 1986.
9. K. C. Kiwiel, "Randomized search directions in descent methods for minimizing certain quasidifferentiable functions," Optimization, vol. 17, pp. 475–485, 1986.
10. K. C. Kiwiel, "Descent methods for quasidifferentiable minimization," Applied Mathematics and Optimization, vol. 18, pp. 163-180, 1988.
11. L. Kuntz and A. Pielczyk, "The method of common descent for a certain class of quasidifferentiable functions," Optimization, vol. 22, pp. 669–679, 1991.
12. G. Papavassilopoulos, "Algorithms for a class of nondifferentiable problems," Journal of Optimization Theory and Applications, vol. 34 pp. 41–82, 1981.
13. O. Pironneau and E. Polak, "On the rate of convergence of certain methods of centers," Mathematical Programming, vol. 2, pp. 230–258, 1972.
14. E. Polak, "Basics of minimax algorithms," in Nonsmooth Optimization and Related Topics, Plenum Press: New York, pp. 343–367, 1989.
15. E. Polak and L. He, "Rate-preserving discretization strategies for semi-infinite programming and optimal control," SIAM Journal on Control and Optimization, vol. 30, pp. 548-572, 1992.
16. E. Polak, "On the use of consistent approximations in the solution of semi-infinite optimization and optimal control problems," Mathematical Programming, vol. 62, pp. 385–414, 1993.
17. E. Polak, Optimization: Algorithms and Consistent Approximations, Springer: New York, 1997.

18. L.N. Polyakova, “On the minimization of a quasidifferentiable function subject to equality-type quasidifferentiable constraints, ” Mathematical Programming Study, vol. 29, pp. 44–55, 1986.
19. B. N. Pshenichnyi and Yu. M. Danilin, Numerical Methods in Extremal Problems, Chislennye Metody v Ekstremal’nykh Zadachakh, Nauka, Moscow, 1975.