SIAM J. OPTIM.
Vol. 11, No. 4, pp. 937–961

SECOND-ORDER ALGORITHMS FOR GENERALIZED FINITE AND SEMI-INFINITE MIN-MAX PROBLEMS*

ELIJAH POLAK†, LIQUN QI‡, AND DEFENG SUN§

Abstract. We present two second-order algorithms, one for solving a class of finite generalized min-max problems and one for solving semi-infinite generalized min-max problems. Our algorithms make use of optimality functions based on second-order approximations to the cost function and of corresponding search direction functions. Under reasonable assumptions we prove that both of these algorithms converge Q-superlinearly, with rate at least 3/2.

This paper is a continuation of [E. Polak, L. Qi, and D. Sun, *Comput. Optim. Appl.*, 13 (1999), pp. 137–161].

Key words. generalized min-max problems, consistent approximations, optimality functions, second-order methods, superlinear convergence

AMS subject classifications. 65K05, 90C34, 90C47

PII. S1052623499358951

1. Introduction.

As is also the case with ordinary min-max problems, generalized min-max problems can be either finite or semi-infinite. Both are of the form

$$\mathbf{P} \qquad \min_{x\in\Re^n} f^0(x), \tag{1.1}$$

where

$$f^0(x) = F(\psi(x)) \ , \tag{1.2}$$

with $F : \Re^m \to \Re$ is a smooth function and $\psi : \Re^n \to \Re^m$ is a nonsmooth, vector-valued function. In the case of generalized finite min-max problems, the components of $\psi(\cdot)$ are of the form[1]

$$\psi^j(x) = \max_{k\in\mathbf{q_j}} \ f^{j,k}(x) \ , \tag{1.3}$$

where the functions $f^{j,k} : \Re^n \to \Re$, $j \in \mathbf{m}$ and $k \in \mathbf{q_j}$, are continuously differentiable and the sets $\mathbf{q_j} := \{1, 2, ..., q_j\}$ are of finite cardinality.[2]

In semi-infinite generalized min-max problems the components of $\psi(\cdot)$ are of the form

$$\psi^j(x) = \max_{y_j\in Y_j} \phi^j(x, y_j) \ , \tag{1.4}$$

*Received by the editors July 12, 1999; accepted for publication (in revised form) September 18, 2000; published electronically March 28, 2001. The research of this paper was supported by the National Science Foundation under grants NSF-INT-9725220 and ECS-9900985 and the Australian Research Council.

http://www.siam.org/journals/siopt/11-4/35895.html

†Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley, CA 94720 (polak@optimum.eecs.berkeley.edu).

‡Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong (maqilq@polyu.edu.hk), and School of Mathematics, The University of New South Wales, Sydney, NSW 2052, Australia (L.Qi@unsw.edu.au).

§School of Mathematics, The University of New South Wales, Sydney, NSW 2052, Australia (sun@maths.unsw.edu.au).

[1] We denote components of a vector by superscripts and elements of a sequence or a set by subscripts.

[2] Given any positive integer q, we use the notation $\mathbf{q} := \{1, 2, \ldots, q\}$.

where the functions $\phi^j : \Re^n \times \Re^{m_j} \to \Re$, $j \in \mathbf{m}$, and $Y_j \subset \Re^{m_j}$, $j \in \mathbf{m}$.

Finite generalized min-max problems are obviously a special case of semi-infinite generalized min-max problems, since when the sets

$$Y_j = \{y_{j,k}\}_{k \in \mathbf{q_j}} \ , \tag{1.5}$$

we can define the functions $f^{j,k}(x)$ by

$$f^{j,k}(x) := \phi^j(x, y_{j,k}) \ . \tag{1.6}$$

The best known generalized min-max problem occurs when an optimization problem with a max function cost and equality and inequality constraints is set up for solution using exact penalty functions, which results in an unconstrained optimization problem with $f^0(x)$ in (1.1) of the form

$$f^0(x) = \max_{i \in \mathbf{p}} c^i(x) + \pi_e \sum_{j=1}^{q} |g^j(x)| + \pi_i \sum_{k=1}^{r} \max\{0, f^k(x)\} \ , \tag{1.7}$$

where π_e and π_i are two positive penalty parameters.

Another simple example occurs in a least squares problem involving max functions, in which case

$$f^0(x) = \sum_{j=1}^{q} \psi^j(x)^2 \ , \tag{1.8}$$

where each $\psi^j(x)$ is as in (1.3).

As a last example, in trying to approximate a structural optimization problem, the aim of which was to minimize the sum of the probability of failure[3] plus the cost of the steel in the structure, using linearizations of a state-limit function, we obtained a cost function of the form

$$f^0(x) = F(-a/(\psi(x) + b)), \tag{1.9}$$

where $F'(y) > 0$, $a > 0$,

$$\psi(x) = \max_{u \in B_\rho} g(x, u) \ , \tag{1.10}$$

B_ρ is a ball of radius ρ, centered at the origin in the space of the random variables u, and $g(x, u)$ is a smooth state-limit function which defined the boundary between outcomes that result in structural failure from those that do not [4].

Functions of the form $f^0(x) = F(\psi(x))$, with $\psi(\cdot)$ as in (1.4), are the best known examples of quasi-differentiable functions and are treated in depth in [3]. Hence generalized min-max problems can be solved using algorithms developed for quasi-differentiable functions; see, e.g., [3, 6, 7, 8]. Under the additional assumption that $\partial F(y)/\partial y^j > 0$ for all $y \in \Re^m$ and $j = 1, \ldots, m$, finite generalized min-max problems

[3]The probability of failure was given by $\int_{g(x,u) \geq 0} \phi(u) du$, with $\phi(\cdot)$ the normal probability density function.

can be solved using transformations[4] into a smooth, constrained nonlinear programming problem (see, e.g., [1, 5, 9]). Direct methods that depend on the assumption that $\partial F(y)/\partial y^j > 0$ for all $y \in \Re^m$ and $j = 1, \ldots, m$ can be found, for example, in [6, 8] and in [17].

We will consider semi-infinite generalized min-max problems under the following hypotheses.

Assumption 1.1. We will assume that

(a) the functions $F(\cdot)$ and $\phi^j(\cdot, y)$, $j \in \mathbf{m}$, $y \in Y_j$, are at least once continuously differentiable;

(b) there exists a positive number $c_F > 0$ such that $\partial F(y)/\partial y^j \geq c_F$ for all $y \in \Re^m$ and $j \in \mathbf{m}$;

(c) the sets Y_j are either compact sets of infinite cardinality, or sets of finite cardinality, of the form given in (1.5).

Parts (a) and (b) of Assumption 1.1 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 all parts of Assumption 1.1 hold, the function $f^0(\cdot)$ has a subgradient. In [17], this fact was used in defining an optimality function and an associated descent direction for the problem $\mathbf{P}$ and in extending the Pshenichnyi–Pironneau–Polak (PPP) Algorithm 4.1 in [13] (see also [18, 10, 11]) to finite generalized min-max problems and the Polak–He PPP Rate-Preserving Algorithm 3.4.9 in [13] (see also [14]) to semi-infinite generalized min-max problems.

In this paper we make use of the following observations, described in section 3.3 of [13] and also used in [15] and [16], for constructing Q-superlinearly converging algorithms for solving finite and semi-infinite min-max problems, of the form (1.1) and (1.2).

First, suppose that the sets Y_j, $j \in \mathbf{m}$, are as in (1.5), i.e., they are of finite cardinality; that the cost function $f^0(\cdot)$ is strongly convex at the minimizer $\hat{x}$, i.e., there exist $\alpha < \infty$ such that

$$f^0(x_i) - f^0(\hat{x}) \geq \alpha \|x_i - \hat{x}\|^2 \ ; \tag{1.11}$$

and that we have a local model $\hat{f}^0(x_i, x - x_i)$ for the cost function at x_i, with the property that for some $\kappa < \infty$,

$$|f^0(x) - \hat{f}^0(x_i, x - x_i)| \leq \kappa \|x - x_i\|^3 \ . \tag{1.12}$$

Then, a local algorithm of the form

$$x_{i+1} \in \arg \min_{x \in \Re^n} \hat{f}^0(x_i, x - x_i) \tag{1.13}$$

converges superlinearly, and, in particular, there exists a $\kappa' < \infty$ such that

$$\|x_{i+1} - \hat{x}\| \leq \kappa' \|x_i - \hat{x}\|^{3/2} \ . \tag{1.14}$$

[4]These transformations result in a smooth problem with more variables than in the nonsmooth problem. There is a fair bit of anecdotal evidence that they can induce considerable ill-conditioning in the smooth problem because they introduce arbitrary scaling. In particular, all methods based on the smooth transformations require the linear independent constraint qualification (LICQ) to be satisfied, which is unlikely to be true for the problem considered here, and some of these methods also require the strict complementarity condition to hold. Instead of using smooth transformations, we directly exploit the problem structure to avoid assuming either LICQ or the strict complementarity condition. However, we do need to solve a slightly more complicated subproblem at each iteration than methods based on smooth transformations.

Next, consider a problem $\mathbf{P}$, with a unique solution $\hat{x}$, and a sequence of approximating problems $\mathbf{P}_N$, with unique solutions $\hat{x}_N$, such that $\hat{x}_N \to \hat{x}$, as $N \to \infty$. Suppose we have an algorithm for solving the problems $\mathbf{P}_N$ such that the iterates that it constructs satisfy the relation

$$\|x_{i+1} - \hat{x}_N\| \leq \gamma \|x_i - \hat{x}_N\|^\tau \tag{1.15}$$

for some $\gamma < \infty$ and $\tau > 1$. If we choose N at each iteration so that

$$\|\hat{x}_N - \hat{x}\| \leq \gamma' \|x_i - \hat{x}_N\|^\sigma \tag{1.16}$$

for some $\gamma' < \infty$ and any $1 < \tau < \sigma$, then there exists a γ'' such that

$$\|x_{i+1} - \hat{x}\| \leq \gamma'' \|x_i - \hat{x}\|^\tau. \tag{1.17}$$

Note that our convergence analysis is heavily dependent on Assumption 2.4, to be introduced in section 2, and hence our results are valid only for convex problems.

In section 2, we present a continuous optimality function and its associated search direction function which, together with a backstepping rule, constitute the backbone of our algorithms. In section 3, we extend the Polak–Mayne–Higgins Newton's method [15], for solving finite min-max problems, to generalized finite min-max problems. We prove the Q-superlinear convergence of this extension in section 4. In section 5, we make use of the theory of consistent approximations developed in [13] and the algorithm presented in section 3 to develop an algorithm for solving generalized semi-infinite min-max problems and prove its convergence and Q-superlinear convergence. Section 6 is devoted to some numerical results to demonstrate the behavior of the proposed algorithms. We sum up in the concluding section 7.

2. Optimality conditions. We will now present optimality conditions for the semi-infinite generalized min-max problem, defined in (1.1), (1.2), (1.4), both in "classical" form and in terms of an optimality function which leads to a superlinearly converging second-order algorithm.

LEMMA 2.1 (see [17]). *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*

$$f^0(x) = F(\psi(x)) \ . \tag{2.1}$$

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

$$df^0(x; h) = \langle \nabla F(\psi(x)), d\psi(x; h) \rangle \ . \tag{2.2}$$

Suppose that Assumption 1.1 is satisfied. Then it follows from Lemma 2.1 that the directional derivative of $f^0(\cdot)$, at a point $x \in \Re^n$ in the direction h, is given by

$$\left\{ \begin{aligned} df^0(x; h) &= \sum_{j \in \mathbf{m}} \frac{\partial F}{\partial y^j}(\psi(x)) d\psi^j(x; h) \\ &= \sum_{j \in \mathbf{m}} \frac{\partial F}{\partial y^j}(\psi(x)) \max_{y_j \in \hat{Y}_j(x)} \langle \nabla_x \phi^j(x, y_j), h \rangle, \end{aligned} \right. \tag{2.3}$$

where

$$\hat{Y}_j(x) := \{y_j \in Y_j \mid \phi^j(x, y_j) = \psi^j(x)\}. \tag{2.4}$$

When all the sets Y_j are as in (1.5), (2.3) assumes the form

$$df^0(x;h) = \sum_{j\in\mathbf{m}} \frac{\partial F}{\partial y^j}(\psi(x)) \max_{k\in\hat{\mathbf{q}}_\mathbf{j}(\mathbf{x})} \langle \nabla f^{j,k}(x), h\rangle \ , \tag{2.5}$$

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

$$f^{j,k}(x) := \phi^j(x, y_{j,k}),\ \ k \in \mathbf{q_j}, \tag{2.6}$$

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

$$\hat{\mathbf{q}}_\mathbf{j}(x) := \{k \in \mathbf{q_j} \mid f^{j,k}(x) = \psi^j(x)\} \ . \tag{2.7}$$

Hence the following result is obvious.

THEOREM 2.2. *Suppose that $\hat{x}$ is a local minimizer for the problem* (1.1), (1.2), (1.4). *Then for all $h \in \Re^n$,*

$$\begin{cases} df^0(\hat{x};h) & = \displaystyle\sum_{j\in\mathbf{m}} \frac{\partial F}{\partial y^j}(\psi(\hat{x})) d\psi^j(\hat{x};h) \\ & = \displaystyle\sum_{j\in\mathbf{m}} \frac{\partial F}{\partial y^j}(\psi(\hat{x})) \max_{y_j\in\hat{Y}_j(\hat{x})} \langle \nabla_x \phi^j(\hat{x}, y_j), h\rangle \geq 0 \ . \end{cases} \tag{2.8}$$

Furthermore, (2.8) *holds if and only if $0 \in \partial f^0(\hat{x})$, where the subgradient $\partial f^0(\hat{x})$ is given by*

$$\partial f^0(\hat{x}) = \sum_{j\in\mathbf{m}} \left\{ \mathrm{conv}_{y_j\in\hat{Y}_j(\hat{x})} \left\{ \frac{\partial F}{\partial y^j}(\psi(\hat{x})) \nabla_x \phi^j(\hat{x}, y_j) \right\} \right\} \ . \tag{2.9}$$

Since (2.8) is a necessary condition of optimality, any point $\hat{x} \in \Re^n$ that satisfies (2.8) will be called *stationary.*

When all the sets Y_j are of the form (1.5), the expressions (2.8) and (2.9) assume the following form:

$$df^0(\hat{x};h) = \sum_{j\in\mathbf{m}} \frac{\partial F}{\partial y^j}(\psi(\hat{x})) \max_{k\in\hat{\mathbf{q}}_\mathbf{j}(\hat{x})} \langle \nabla f^{j,k}(\hat{x}), h\rangle \geq 0 \quad \forall\ h \in \Re^n, \tag{2.10}$$

$$\partial f^0(\hat{x}) = \sum_{j\in\mathbf{m}} \mathrm{conv}_{k\in\hat{\mathbf{q}}_\mathbf{j}(\hat{x})} \left\{ \frac{\partial F}{\partial y^j}(\psi(\hat{x})) \nabla f^{j,k}(\hat{x}) \right\} \ . \tag{2.11}$$

DEFINITION 2.3. *We will say that $\theta : \Re^n \to \Re$ is an* optimality function *for problem* (1.1), (1.2), (1.4) *if*

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

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

(c) *for any $\hat{x} \in \Re^n$,* (2.8) *holds if and only if $\theta(\hat{x}) = 0$.*

Assumption 2.4. 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 (1.1), (1.2), (1.4), are twice Lipschitz continuously differentiable on bounded sets,

(b) the functions $\phi^j(\cdot, y_j)$, $\nabla_x \phi^j(\cdot, y_j)$, and $\nabla_x^2 \phi^j(\cdot, y_j)$ are locally Lipschitz continuous, $j \in \mathbf{m}$, $y_j \in Y_j$, and

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

$$c\|h\|^2 \leq \langle h, \nabla_x^2 \phi^j(x, y_j) h \rangle \leq C\|h\|^2 \tag{2.12}$$

and

$$0 \leq \langle w, \nabla^2 F(\psi(x)) w \rangle \leq C\|w\|^2. \tag{2.13}$$

For the sake of convenience, for any $x, h \in \Re^n$ and $w \in \Re^m$, we define

$$u(x, h, w) := \langle \nabla F(\psi(x)), \hat{\psi}(x, h) - \psi(x) + w \rangle \tag{2.14}$$

and

$$v(x, h, w) := \tfrac{1}{2} \langle \hat{\psi}(x, h) - \psi(x) + w, \nabla^2 F(\psi(x))(\hat{\psi}(x, h) - \psi(x) + w) \rangle, \tag{2.15}$$

where $\hat{\psi}(x, h) = (\hat{\psi}^1(x, h), \ldots, \hat{\psi}^m(x, h))$, and

$$\hat{\psi}^j(x, h) := \max_{y_j \in Y_j} \{\phi^j(x, y_j) + \langle \nabla_x \phi^j(x, y_j), h \rangle + \tfrac{1}{2} \langle h, \nabla_x^2 \phi^j(x, y_j) h \rangle\}. \tag{2.16}$$

The reason for the introduction of the artificial variable w is as follows. The function

$$\tilde{f}^0(x, h) := F(\psi(x)) + u(x, h, 0) + v(x, h, 0) \tag{2.17}$$

is a perfectly good second-order approximation to $F(\psi(x+h))$, but unfortunately, it is not always convex and hence leads to problems in developing an algorithm for solving semi-infinite generalized min-max problems. By introducing the artificial variable w, we can define the function

$$\hat{f}^0(x, h) := \min_{w \in \Re^m_+} \{F(\psi(x)) + u(x, h, w) + v(x, h, w)\} \tag{2.18}$$

which, as we will later see, is a convex second-order approximation to $F(\psi(x+h))$ and hence much more useful in algorithm construction.

We define the function $\theta : \Re^n \to \Re$ and the associated search direction function $H : \Re^n \to \Re^n$ by

$$\theta(x) := \min_{h \in \Re^n} \{ \min_{w \in \Re^m_+} [u(x, h, w) + v(x, h, w)]\} \tag{2.19}$$

and

$$H(x) := \arg \min_{h \in \Re^n} \{ \min_{w \in \Re^m_+} [u(x, h, w) + v(x, h, w)]\} \; . \tag{2.20}$$

Note that

$$\theta(x) = \min_{h \in \Re^n} \{\hat{f}^0(x, h) - f^0(x)\}. \tag{2.21}$$

We will shortly see that the function $\theta(\cdot)$ is an optimality function for the problem (1.1), (1.2), (1.4). For any $y, \delta y \in \Re^m$, let

$$\hat{F}(y, \delta y) := \min_{w \in \Re^m_+} \{F(y) + \langle \nabla F(y), \delta y + w\rangle + \tfrac{1}{2}\langle \delta y + w, \nabla^2 F(y)(\delta y + w)\rangle\}. \tag{2.22}$$

LEMMA 2.5. *Suppose that Assumptions* 1.1 *and* 2.4 *are satisfied. For any* $y, \delta y \in \Re^m$, *let* $\Omega^*(y, \delta y) \subset \Re^m_+$ *be the solution set of* (2.22). *Then* $\Omega^*(y, \delta y)$ *is nonempty and compact and for any* $w^* \in \Omega^*(y, \delta y)$, *we have*

$$\nabla F(y) + \nabla^2 F(y)\delta y + \nabla^2 F(y) w^* \geq 0. \tag{2.23}$$

Proof. Since $\nabla F(y) > 0$ and $\nabla^2 F(y)$ is positive semidefinite, for any $w \in \Re^m_+$ and $\|w\| \to \infty$ we have

$$F(y) + \langle \nabla F(y), \delta y + w\rangle + \tfrac{1}{2}\langle \delta y + w, \nabla^2 F(y)(\delta y + w)\rangle \to +\infty. \tag{2.24}$$

Thus, $\Omega^*(y, \delta y)$ is nonempty and compact.

Suppose that $w^* \in \Omega^*(y, \delta y)$. Then w^* satisfies the following first-order optimality conditions which follow directly from (and are equivalent to) the KKT conditions:

$$\begin{cases} \nabla F(y) + \nabla^2 F(y)(\delta y + w^*) - \lambda^* = 0, \\ \\ w^* \geq 0,\ \lambda^* \geq 0,\ \langle w^*, \lambda^*\rangle = 0, \end{cases} \tag{2.25}$$

i.e.,

$$\begin{cases} \nabla F(y) + \nabla^2 F(y)(\delta y + w^*) \geq 0, \\ \\ w^* \geq 0, \\ \\ \langle w^*, \nabla F(y) + \nabla^2 F(y)(\delta y + w^*)\rangle = 0. \end{cases} \tag{2.26}$$

Clearly, (2.26) implies that for any $w^* \in \Omega^*(y, \delta y)$, we have

$$\nabla F(y) + \nabla^2 F(y)\delta y + \nabla^2 F(y) w^* \geq 0. \qquad \square \tag{2.27}$$

LEMMA 2.6. *Suppose that Assumptions* 1.1 *and* 2.4 *are satisfied. Then for any* $z \in \Re^n$ *there exists an* $\varepsilon > 0$ *such that for all* $h \in \Re^n$ *with* $\|h\| \leq \varepsilon$ *and for all* $x \in \Re^n$ *with* $\|x - z\| \leq \varepsilon$ *we have*

$$\hat{f}^0(x, h) = F(\psi(x)) + u(x, h, 0) + v(x, h, 0), \tag{2.28}$$

i.e.,

$$\hat{f}^0(x, h) = \tilde{f}^0(x, h), \tag{2.29}$$

where $\tilde{f}^0(\cdot,\cdot)$ *is defined by* (2.17).

Proof. Since $F(\psi(x)) + u(x, h, \cdot) + v(x, h, \cdot)$ is a convex quadratic function, any $w \in \Re^m$ satisfying the first-order conditions

$$\begin{cases} w \geq 0, \\ \\ \nabla F(\psi(x)) + \nabla^2 F(\psi(x))(\hat{\psi}(x, h) - \psi(x) + w) \geq 0, \\ \\ \langle w, \nabla F(\psi(x)) + \nabla^2 F(\psi(x))(\hat{\psi}(x, h) - \psi(x) + w)\rangle = 0 \end{cases} \tag{2.30}$$

is a solution of (2.18). Then, because $\partial F(y)/\partial y_j \geq c_F$, for every $j \in \mathbf{m}$ and $y \in \Re^m$, and $\hat{\psi}(\cdot,\cdot)$ is uniformly continuous on any compact set and $\hat{\psi}(x,0) = \psi(x)$, we see that for any $z \in \Re^n$ there exists an $\varepsilon > 0$ such that for all $h \in \Re^n$ with $\|h\| \leq \varepsilon$ and for all $x \in \Re^n$ with $\|x - z\| \leq \varepsilon$, $w = 0$ satisfies (2.30). This implies that for all those h and x, we have

$$\hat{f}^0(x,h) = F(\psi(x)) + u(x,h,0) + v(x,h,0) = \tilde{f}^0(x,h). \tag{2.31}$$

Hence our proof is complete. $\square$

The above lemma shows that $\hat{f}^0(x,h)$ is identical to $\tilde{f}^0(x,h)$ for all h sufficiently small. This fact will be used in proving our superlinear convergence results.

In general, $\tilde{f}^0(x,h)$ is not convex in h. We will now show that $\hat{f}^0(x,h)$ is convex in h.

LEMMA 2.7. *Suppose that Assumptions* 1.1 *and* 2.4 *are satisfied. Then for any fixed* $x \in \Re^n$, $\hat{f}^0(x,\cdot)$ *is a convex function. Moreover,* $\hat{f}^0(\cdot,\cdot)$ *is continuous.*

Proof. First we will show that $\hat{f}^0(x,\cdot)$ is a convex function. For any $y \in \Re^m$ and $\delta y \in \Re^m$, we have

$$\hat{F}(y,\delta y) = F(y) + \langle \nabla F(y), \delta y\rangle + \tfrac{1}{2}\langle \delta y, \nabla^2 F(y)\delta y\rangle + S(\delta y), \tag{2.32}$$

where

$$S(\delta y) = \min_{w\in\Re^m_+} \langle \nabla F(y) + \nabla^2 F(y)\delta y, w\rangle + \tfrac{1}{2}\langle w, \nabla^2 F(y) w\rangle \ . \tag{2.33}$$

It is easy to verify that $S(\delta y)$ is a concave function and that its subgradient is given by

$$\partial S(\delta y) = \mathrm{conv}\{\nabla^2 F(y) w^* : w^* \in \Omega^*(y,\delta y)\}, \tag{2.34}$$

where $\Omega^*(y,\delta y) \subset \Re^m_+$ is the solution set of (2.33). It now follows from (2.32) that for any $y \in \Re^m$, $\hat{F}(y,\cdot)$ is locally Lipschitz continuous and that its generalized gradient at δy in the sense of Clarke [2] is given by

$$\partial_{\delta y}\hat{F}(y,\delta y) = \mathrm{conv}\{\nabla F(y) + \nabla^2 F(y)\delta y + \nabla^2 F(y) w^* : w^* \in \Omega^*(y,\delta y)\}. \tag{2.35}$$

Since, by Lemma 2.5, for any $w^* \in \Omega^*(y,\delta y)$,

$$\nabla F(y) + \nabla^2 F(y)\delta y + \nabla^2 F(y) w^* \geq 0, \tag{2.36}$$

we conclude that $s \geq 0$ for any $s \in \partial_{\delta y}\hat{F}(y,\delta y)$. Hence, since $\hat{\psi}^j(x,\cdot)$ is convex for every $j \in \{1,\ldots,m\}$, it follows that $\hat{f}^0(x,h) = \hat{F}(\psi(x), \hat{\psi}(x,h) - \psi(x))$ is convex in $h \in \Re^n$ (because it is the composition of a convex function with positive elements in the generalized gradient and a vector function whose components are convex).

Next, we will prove that $\hat{f}^0(x,h)$ is continuous. First, since $\partial F(y)/\partial y_j \geq c_F > 0$ and $\nabla^2 F(y)$ is positive semidefinite for all $j \in \{1,\ldots,m\}$ and $y \in \Re^m$, it follows from (2.22) that $\Omega^*(y,\delta y)$ is uniformly bounded in a neighborhood of given point $(z,\delta z) \in \Re^m \times \Re^m$. It now follows from Corollary 5.4.2 in [13] that $\hat{F}(\cdot,\cdot)$ is continuous. Hence

$$\hat{F}(y,\delta y) \to \hat{F}(z,\delta z) \quad \text{as} \quad y \to z, \delta y \to \delta z \ , \tag{2.37}$$

which implies that $\hat{f}^0(x,h)$ is continuous on $\Re^n \times \Re^n$ because

$$\hat{f}^0(x,h) = \hat{F}(\psi(x), \hat{\psi}(x,h) - \psi(x)) \tag{2.38}$$

with $y := \psi(x)$ and $\delta y := \hat{\psi}(x,h) - \psi(x)$. $\square$

The following theorem shows that $\theta(\cdot)$ is indeed an optimality function for the problem (1.1), (1.2), (1.4) and that the set-valued function $H(\cdot)$ is a descent direction function for $f^0(\cdot)$.

THEOREM 2.8. *Suppose that Assumptions* 1.1 *and* 2.4 *are satisfied. Consider the functions* $\theta(\cdot)$ *and* $H(\cdot)$ *defined by* (2.19) *and* (2.20), *respectively. Then the following hold:*

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

$$\theta(x) \le 0 \ . \tag{2.39}$$

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

$$df^0(x;h) \le \theta(x) - \gamma\|h\|^2 \ \forall \ h \in H(x), \tag{2.40}$$

where $df^0(x;h)$ *is the directional derivative of* f^0 *at* x *in the direction* h *and* $\gamma = \frac{1}{2}mc_F c$.

(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* $f^0(\cdot)$ *at* x, *defined in* (2.9). *Moreover, for any* $x \in \Re^n$ *such that* $\theta(x) = 0$ *we have* $H(x) = \{0\}$.

(iv) *The set-valued map* $H(\cdot)$ *is* (a) *bounded on bounded sets,* (b) *compact valued, and* (c) *outer-semicontinuous, i.e., for any* $x \in \Re^n$, $H(x)$ *is closed and, for every compact set* S *such that* $H(x) \cap S = \emptyset$, *there exists a* $\rho > 0$ *such that* $H(z) \cap S = \emptyset$ *for all* $z \in B(x,\rho) := \{y \in \Re^n \,|\, \|y - x\| \ \le \rho\}$.

(v) *The function* $\theta(\cdot)$ *is continuous.*

Proof. (i) Since $h = 0$ is admissible in (2.19) that $\theta(x) \le 0$ for all $x \in \Re^n$.

(ii) Since $\hat{Y}_j(x) \subset Y_j$, it follows directly from the definition of $\theta(x)$ in (2.19) that for any $h \in H(x)$,

$$\left\{ \begin{array}{rl} \theta(x) & \ge \min\limits_{w \in \Re^m_+} \langle \nabla F(\psi(x)), \hat{\psi}(x,h) - \psi(x) + w \rangle \\[2ex] & = \langle \nabla F(\psi(x)), \hat{\psi}(x,h) - \psi(x) \rangle \\[2ex] & \ge \displaystyle\sum_{j \in \mathbf{m}} \frac{\partial F}{\partial y^j}(\psi(x)) \max_{y_j \in \hat{Y}_j(x)} \ \{\phi^j(x,y_j) - \psi^j(x) \\[2ex] & \qquad\qquad + \langle \nabla_x \phi^j(x,y_j), h \rangle + \frac{1}{2}c\|h\|^2\} \\[2ex] & \ge df^0(x,h) \ + \ \frac{1}{2}mc_F c\|h\|^2. \end{array} \right. \tag{2.41}$$

Thus we have shown that (2.40) holds.

(iii) For any $x \in \Re^n$, let

$$\eta(x) := \min_{h \in \Re^n} \min_{w \in \Re^m_+} u(x,h,w) = \min_{h \in \Re^n} u(x,h,0) \ . \tag{2.42}$$

We will first prove that

$$\theta(x) = 0 \Longleftrightarrow \eta(x) = 0. \tag{2.43}$$

It is easy to see that $\eta(x) = 0 \Rightarrow \theta(x) = 0$ because $\theta(x) \geq \eta(x)$ and $\theta(x) \leq 0$. Hence we only need to show that $\theta(x) = 0 \Rightarrow \eta(x) = 0$.

Suppose that $\theta(x) = 0$ but $\eta(x) < 0$. Then, there exists an $h' \in \Re^n$ such that $\eta(x) = u(x, h', 0) < 0$.

For any $j \in \{1, \ldots, m\}$, we have

$$(2.44)\quad \begin{cases} \hat{\psi}^j(x,h) - \psi^j(x) \\ = \max_{y_j \in Y_j} \{\phi^j(x,y_j) + \langle \nabla_x \phi^j(x,y_j), h\rangle + \frac{1}{2}\langle h, \nabla_{xx}\phi^j(x,y_j)h\rangle\} - \psi^j(x) \\ \leq \max_{y_j \in Y_j} \{\langle \nabla_x \phi^j(x,y_j), h\rangle + \frac{1}{2}\langle h, \nabla_{xx}\phi^j(x,y_j)h\rangle\} \end{cases}$$

and

$$(2.45)\quad \begin{cases} \hat{\psi}^j(x,h) - \psi^j(x) \\ \geq \max_{y_j \in \hat{Y}_j(x)} \{\phi^j(x,y_j) + \langle \nabla_x \phi^j(x,y_j), h\rangle + \frac{1}{2}\langle h, \nabla_{xx}\phi^j(x,y_j)h\rangle\} - \psi^j(x) \\ = \max_{y_j \in \hat{Y}_j(x)} \{\langle \nabla_x \phi^j(x,y_j), h\rangle + \frac{1}{2}\langle h, \nabla_{xx}\phi^j(x,y_j)h\rangle\} \\ \geq \min_{y_j \in Y_j} \{\langle \nabla_x \phi^j(x,y_j), h\rangle + \frac{1}{2}\langle h, \nabla_{xx}\phi^j(x,y_j)h\rangle\}\ . \end{cases}$$

Thus, there exists a constant C_0 such that

$$(2.46)\qquad \|\hat{\psi}(x,h) - \psi(x)\| \leq C_0 \max\{\|h\|, \|h\|^2\} \quad \forall\ h \in \Re^n,$$

which further implies that there exists a constant C_1 such that

$$(2.47)\qquad 0 \leq v(x,h,0) \leq C_1 \max\{\|h\|^2, \|h\|^4\} \quad \forall\ h \in \Re^n.$$

Since $u(x,\cdot,0)$ is a convex function and $u(x,0,0) = 0$, for $\lambda > 0$ sufficiently small we have

$$(2.48)\quad \begin{cases} u(x,\lambda h',0) + v(x,\lambda h',0) & \leq & \lambda u(x,h',0) + \lambda^2 C_1 \|h'\|^2 \\ & = & \lambda \eta(x) + \lambda^2 C_1 \|h'\|^2 \\ & < & 0, \end{cases}$$

which contradicts that $\theta(x) = 0$. Hence $\theta(x) = 0 \Rightarrow \eta(x) = 0$.

Next, with $\partial f^0(x)$ the subgradient of $f^0(\cdot)$ at x, defined in (2.9), by emulating the proof of Lemma 2.5.5 in [13], we can prove that for any $x \in \Re^n$, $0 \in \partial f^0(x)$ if and only if $\eta(x) = 0$, and therefore if and only if $\theta(x) = 0$.

Finally we will show that for any $x \in \Re^n$ such that $\theta(x) = 0$ we have $H(x) = \{0\}$. For the sake of contradiction, suppose that there exists an $x \in \Re^n$ such that $\theta(x) = 0$ but $H(x) \neq \{0\}$. Then there exist $0 \neq h \in \Re^n$ and $w \in \Re^m_+$ such that

$$(2.49)\qquad u(x,h,w) + v(x,h,w) = \theta(x) = 0,$$

which, together with the fact that $v(x,h,w) \geq 0$ implies that $u(x,h,w) \leq 0$. Hence we conclude that both $u(x,h,w) = 0$ and $w = 0$ because otherwise $\eta(x) \leq u(x,h,0) \leq u(x,h,w) < 0$, which contradicts (2.43). However, $\eta(x) = u(x,h,0) = 0$ implies that $h = 0$ because $u(x,h,0)$ is strongly convex in h and $u(x,0,0) = 0$.

(iv) According to our definition, for each $h \in \Re^n$ there exists a $w(h) \in \Re^m_+$ such that

$$\hat{f}^0(x,h) = F(\psi(x)) + u(x,h,w(h)) + v(x,h,w(h)), \tag{2.50}$$

which, together with the fact that $\nabla F(y) > 0$, $y \in \Re^m$ and $v(x,h,w(h)) \geq 0$, implies that

$$\hat{f}^0(x,h) \geq F(\psi(x)) + u(x,h,w(h)) \geq F(\psi(x)) + u(x,h,0). \tag{2.51}$$

Since for each $j \in \{1,\dots,m\}$ and $h \in \Re^n$,

$$\hat{\psi}^j(x,h) \geq \max_{y_j \in Y_j} \{\phi^j(x,y_j) + \langle \nabla_x \phi^j(x,y_j), h\rangle + \tfrac{1}{2}c\langle h,h\rangle\}\ , \tag{2.52}$$

it follows from (2.51) that for all s in any bounded neighborhood of x,

$$\hat{f}^0(s,h) \to \infty \quad \text{as} \quad \|h\| \to \infty. \tag{2.53}$$

Consequently, for any $x \in \Re^n$, $H(x)$ is nonempty and bounded and $H(\cdot)$ is bounded on bounded sets. Since $\hat{f}^0(x,h)$ is continuous (Lemma 2.7), it follows that $H(x)$ is closed. Next we will prove that for every $x \in \Re^n$ and every compact set S such that $H(x) \cap S = \emptyset$, there exists a $\rho > 0$ such that $H(z) \cap S = \emptyset$ for all $z \in B(x,\rho)$. Suppose not; then there exists an $x \in \Re^n$ and a compact set S such that $H(x) \cap S = \emptyset$ and a sequence $\{x_i\}$ converging to x such that $H(x_i) \cap S \neq \emptyset$. Hence there exists a sequence $\{h_i\}$ such that $h_i \in H(x_i) \cap S$. Since S is a compact set, without loss of generality, we can assume that

$$h_i \to \bar{h} \in S. \tag{2.54}$$

By definition of $H(x_i)$,

$$\hat{f}^0(x_i,h_i) \leq \hat{f}^0(x_i,h) \quad \forall\ h \in \Re^n. \tag{2.55}$$

Since $\hat{f}^0(\cdot,\cdot)$ is continuous, it follows from (2.55) that

$$\hat{f}^0(x,\bar{h}) \leq \hat{f}^0(x,h) \quad \forall\ h \in \Re^n, \tag{2.56}$$

which implies that $\bar{h} \in H(x)$. This contradicts that $H(x) \cap S = \emptyset$. Thus, we have shown that $H(\cdot)$ is outer-semicontinuous.

(v) Finally, it follows from Corollary 5.4.2 in Polak [13] that θ is continuous. □

By introducing an additional variable, we can rewrite the expression for $\theta(x)$, defined in (2.19), as follows:

$$\begin{cases} \theta(x) = \min\limits_{(p,h)} & \{\langle \nabla F(\psi(x)), p\rangle + \tfrac{1}{2}\langle p, \nabla^2 F(\psi(x))p\rangle\} \\ \qquad\qquad \text{s.t.} & p - \hat{\psi}(x,h) + \psi(x) \geq 0. \end{cases} \tag{2.57}$$

The constraints in (2.57) involve maximum functions, and hence (2.57) appears to be a nonsmooth problem. However, (2.57) can be reformulated as a smooth problem with quadratic cost and quadratic constraints, as follows:

$$\left\{\begin{array}{rl} \theta(x) = \min\limits_{(p,h)} & \{\langle \nabla F(\psi(x)), p\rangle + \frac{1}{2}\langle p, \nabla^2 F(\psi(x))p\rangle\} \\ \text{s.t.} & p^j - \phi^j(x, y_j) - \langle \nabla_x \phi^j(x, y_j), h\rangle - \frac{1}{2}\langle h, \nabla_x^2 \phi^j(x, y_j)h\rangle + \psi^j(x) \geq 0, \\ & j \in \mathbf{m}, y_j \in Y_j. \end{array}\right. \tag{2.58}$$

Under the assumptions in this paper, (2.58) is convex, and hence can be solved using the smoothing Newton method in [19] (see [19, 21] for the details of the implementation of the smoothing Newton method as well as section 6 for numerical results). Alternatively, one can use primal-dual interior point methods, described in [20] and references therein.

THEOREM 2.9. *Suppose that Assumptions* 1.1 *and* 2.4 *are satisfied and the sets* Y_j *are as in* (1.5). *For any* $x \in \Re^n$, *let* $\Gamma(x)$ *be the solution set of* (2.57), *i.e., any* $(p, h) \in \Gamma(x)$ *solves* (2.57). *Then*

(i) *problem* (2.57) *is a convex quadratic problem with convex quadratic constraints;*

(ii) *for* $x \in \Re^n$, $\Gamma(x)$ *is nonempty and compact and* $\Gamma(\cdot)$ *is outer-semicontinuous and bounded on bounded sets;*

(iii) *if* $z \in \Re^n$ *is such that* $\theta(z) = 0$, *then* $\Gamma(z) = \{(0, 0)\}$ *and there exist a neighborhood* $N(z)$ *of* z *and an* $\varepsilon > 0$ *such that for any* $(p, h) \in \Gamma(x)$, $x \in N(z)$, *we have*

$$\theta(x) \leq -\varepsilon\|h\|^2. \tag{2.59}$$

Proof. (i) Under the conditions of Assumptions 1.1 and 2.4, $\nabla^2 F(\psi(x))$ is positive semidefinite and for each $j \in \{1, 2, \ldots, m\}$, $\hat{\psi}^j(x, \cdot)$ is strongly convex. Hence (2.57) is a convex quadratic problem with convex quadratic constraints.

(ii) Since for all z in a bounded neighborhood $N(x)$ of x and $j \in \{1, 2, \ldots, m\}$,

$$\hat{\psi}^j(z, h) - \psi^j(z) \to +\infty \quad \text{as} \quad \|h\| \to \infty, \tag{2.60}$$

it follows that for all $z \in N(x)$ and $(p, h) \in \Re^m \times \Re^n$ satisfying

$$p \geq \hat{\psi}(z, h) - \psi(z), \tag{2.61}$$

we have

$$\langle \nabla F(\psi(z)), p\rangle + \tfrac{1}{2}\langle p, \nabla^2 F(\psi(z))p\rangle \geq c_F \sum_{j \in \mathbf{m}} p^j \to \infty \text{ as } \|(p, h)\| \to \infty. \tag{2.62}$$

Hence, for all $x \in \Re^n$, $\Gamma(x)$ is nonempty and compact, and $\Gamma(\cdot)$ is bounded on bounded sets.

The outer-semicontinuity of $\Gamma(\cdot)$ follows from the fact that $\theta(\cdot)$ is continuous and the constraint set in (2.57) is outer-semicontinuous.

(iii) Since $z \in \Re^n$ is such that $\theta(z) = 0$, $(0, 0) \in \Gamma(z)$. For any $x \in \Re^n$, the KKT conditions for (2.57) are

$$\left\{\begin{array}{l} \nabla F(\psi(x)) + \nabla^2 F(\psi(x))p = \lambda, \\ 0 \in \sum\limits_{j \in \mathbf{m}} \lambda^j \partial_h \hat{\psi}^j(x, h), \\ \lambda \geq 0, \ p - \hat{\psi}(x, h) + \psi(x) \geq 0, \ \lambda^T(p - \hat{\psi}(x, h) + \psi(x)) = 0, \end{array}\right. \tag{2.63}$$

where $\partial_h \hat{\psi}^j(x,h)$ is the subgradient of $\hat{\psi}^j(x,h)$ with respect to h.

Suppose that $(p,h) \in \Gamma(z)$. By (iii) of Theorem 2.8, we have $h = 0$. Hence it follows from (2.63) and the fact that $\hat{\psi}(z,0) = \psi(z)$ that

$$\langle p, \nabla F(\psi(z))\rangle + \langle p, \nabla^2 F(\psi(z))p\rangle = 0, \tag{2.64}$$

which implies that $p = 0$ because $p \geq 0$, $\nabla F(\psi(z)) > 0$ and $\nabla^2 F(\psi(z))$ is positive semidefinite. Thus, we have proved that $\Gamma(z) = \{(0,0)\}$. Hence, since $\Gamma(\cdot)$ is outer-semicontinuous, it follows that if $x \to z$ and $(p,h) \in \Gamma(x)$, then

$$(p,h) \to (0,0)\ . \tag{2.65}$$

It now follows from (2.63), (2.65), and the fact that for any $y \in \Re^m$, $\partial F(y)/\partial y^j \geq c_F > 0$ for $j \in \{1,2,\ldots,m\}$ that there exists a neighborhood $N(z)$ of z such that for all $x \in N(z)$, the multiplier λ in the KKT (2.63) must have all components positive and hence for all $x \in N(z)$, the KKT conditions for (2.57) become

$$\begin{cases} \nabla F(\psi(x)) + \nabla^2 F(\psi(x))p = \lambda, \\ 0 \in \displaystyle\sum_{j\in\mathbf{m}} \lambda^j \partial_h \hat{\psi}^j(x,h), \\ \lambda > 0,\ p - \hat{\psi}(x,h) + \psi(x) = 0. \end{cases} \tag{2.66}$$

Thus, for any $x \in N(z)$ and $j \in \{1,2,\ldots,m\}$, there exist nonnegative numbers $\mu^{j,k} \in [0,1]$ satisfying $\sum_{k\in\mathbf{q_j}} \mu^{j,k} = 1$ such that for any $(p,h) \in \Gamma(x)$

$$\sum_{j\in\mathbf{m}} \lambda^j \sum_{k\in\mathbf{q_j}} \mu^{j,k}(\nabla f^{j,k}(x) + \nabla^2 f^{j,k}(x)h) = 0, \tag{2.67}$$

where

$$\lambda = \nabla F(\psi(x)) + \nabla^2 F(\psi(x))p > 0, \tag{2.68}$$

and for any $k \in \mathbf{q_j}$ such that

$$\hat{\psi}^j(x,h) > f^{j,k}(x) + \langle \nabla f^{j,k}(x), h\rangle + \tfrac{1}{2}\langle h, \nabla^2 f^{j,k}(x)h\rangle, \tag{2.69}$$

we have

$$\mu^{j,k} = 0. \tag{2.70}$$

We conclude from (2.66), (2.67), and (2.68) that for all $x \in N(z)$ and $(p,h) \in \Gamma(x)$,

$$
\left\{
\begin{aligned}
\theta(x) &= \langle \nabla F(\psi(x)), p\rangle + \tfrac{1}{2}\langle p, \nabla^2 F(\psi(x))p\rangle \\
&= \langle \lambda, p\rangle - \tfrac{1}{2}\langle p, \nabla^2 F(\psi(x))p\rangle \\
&\leq \langle \lambda, p\rangle \\
&= \langle \lambda, \hat{\psi}(x,h) - \psi(x)\rangle \\
&= \sum_{j\in\mathbf{m}} \lambda^j \sum_{k\in\mathbf{q_j}} \mu^{j,k}[(f^{j,k}(x) - \psi^j(x)) + \langle \nabla f^{j,k}(x), h\rangle + \tfrac{1}{2}\langle h, \nabla^2 f^{j,k}(x)h\rangle] \\
&= \sum_{j\in\mathbf{m}} \lambda^j \sum_{k\in\mathbf{q_j}} \mu^{j,k}[(f^{j,k}(x) - \psi^j(x)) - \tfrac{1}{2}\langle h, \nabla^2 f^{j,k}(x)h\rangle] \\
&\leq \sum_{j\in\mathbf{m}} \lambda^j \sum_{k\in\mathbf{q_j}} \mu^{j,k}[(f^{j,k}(x) - \psi^j(x)) - \tfrac{1}{2}c\langle h, h\rangle] \\
&\leq -\tfrac{1}{2}c\|h\|^2 \sum_{j\in\mathbf{m}} \lambda^j,
\end{aligned}
\right.
\tag{2.71}
$$

where the last inequality follows from the fact that $f^{j,k}(x) \leq \psi^j(x)$ for all $k \in \mathbf{q_j}$ and $j \in \mathbf{m}$. By shrinking $N(z)$ if necessary, we conclude from (2.68), (2.71), and Assumptions 1.1 and 2.4 that there exists a positive number $\varepsilon > 0$ such that for all $x \in N(z)$ and $(p,h) \in \Gamma(x)$, $\theta(x) \leq -\varepsilon\|h\|^2$. $\square$

3. An algorithm for solving generalized finite min-max problems. An algorithm for solving generalized finite min-max problems is obviously of interest in its own right. However, we will also need it as a subroutine for our algorithms for solving generalized semi-infinite min-max problems. Hence, for the time being, we will assume that the sets Y_j are of the form (1.5) and that the functions $f^{j,k}(\cdot)$ are as in (2.6). As a result, our generalized finite min-max problem assumes the form (1.1), (1.2), (1.4), with

$$
\left\{
\begin{aligned}
&\min_{x\in\Re^n}\ f^0(x) \\
&f^0(x) = F(\psi(x)), \\
&\psi(x) = (\psi^1(x), \ldots, \psi^m(x)), \\
&\psi^j(x) = \max_{k\in\mathbf{q_j}}\ f^{j,k}(x), \quad j \in \mathbf{m}\ ,
\end{aligned}
\right.
\tag{3.1}
$$

where, in view of Assumption 1.1, the functions $F(\cdot)$ and $f^{j,k}(\cdot)$, $j \in \mathbf{m}$, $k \in \mathbf{q_j}$ are all continuously differentiable, where $f^{j,k}(\cdot)$ are defined by (2.6).

We are now ready to state an algorithm for solving generalized finite min-max problems. This algorithm is a generalization of the Polak–Mayne–Higgins Newton's algorithm for solving finite min-max problems [15].

ALGORITHM 3.1 (solves problem (3.1)).
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 a search direction* $h_i \in H(x_i)$ *according to the formulae* (2.19) *and* (2.20).
Step 2. *If* $\theta_i = 0$*, stop. Else, compute the step-size*

$$\lambda_i = \lambda(x_i) := \max_{k\in\mathcal{N}}\{\beta^k |\ f^0(x_i + \beta^k h_i) - f^0(x_i) - \beta^k\alpha\theta_i \le 0\}, \tag{3.2}$$

where $\mathcal{N} := \{0, 1, 2, \dots\}$.
Step 3. *Set*

$$x_{i+1} = x_i + \lambda_i h_i, \tag{3.3}$$

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

LEMMA 3.2 (see [17]). *Suppose that Assumption* 1.1 *holds. Then for any* $y, y' \in \Re^m$ *such that* $y' \ge y$*,*

$$F(y') - F(y) \ge c_F \sum_{j\in\mathbf{m}} (y'_j - y_j). \tag{3.4}$$

LEMMA 3.3 (see [17]). *Suppose that Assumptions* 1.1 *and* 2.4 *are satisfied. Then there exists a constant* $\tau > 0$ *such that for all* $x, x' \in \Re^n$ *and* $\lambda \in [0, 1]$,

$$f^0(\lambda x + (1-\lambda)x') \le \lambda f^0(x) + (1-\lambda)f^0(x') - \tfrac{1}{2}\tau\lambda(1-\lambda)\|x - x'\|^2. \tag{3.5}$$

THEOREM 3.4. *Suppose that Assumptions* 1.1 *and* 2.4 *are satisfied and that all the* Y_j*,* $j \in \mathbf{m}$*, are of the form* (1.5)*, so that problem* (1.1)*,* (1.2)*,* (1.4) *reduces to problem* (3.1)*. If* $\{x_i\}_{i=0}^\infty$ *is an infinite sequence generated by Algorithm* 3.1 *and* x^* *is the unique solution of* (3.1)*, then* $\{x_i\}_{i=0}^\infty$ *converges to* x^*.

Proof. Suppose that $\{x_i\}_{i=0}^\infty$ is an infinite sequence generated by Algorithm 3.1. Since $f(\cdot)$ is strongly convex by Lemma 3.3, the sequence $\{x_i\}_{i=0}^\infty$ is bounded. Suppose that $\hat{x}$ is an accumulation point of this sequence. Since the cost function $f^0(\cdot)$ is continuous, $f^0(\hat{x})$ is an accumulation point of the cost sequence. Hence, since, by construction, the cost sequence $\{f^0(x_i)\}_{i=0}^\infty$ is monotone decreasing, it follows that $f^0(x_i) \to f^0(\hat{x})$, as $i \to \infty$.

Now, for the sake of contradiction, suppose that $\theta(\hat{x}) < 0$. Since for any $x \in \Re^n$, $H(x)$ is compact, and $H(\cdot)$ is bounded on bounded sets and is outer-semicontinuous ((iv) of Theorem 2.8), it follows from Theorem 5.3.7 (b) in Polak [13] that there exists a subsequence $\{j_i\}_{i=0}^\infty$ of the integers such that $x_{j_i} \to \hat{x}$ and $h_{j_i} \to \hat{h} \in H(\hat{x})$, as $i \to \infty$. It follows from (ii) and (iii) of Theorem 2.8 that $\hat{h} \neq 0$ and

$$df^0(\hat{x}; \hat{h}) \le \theta(\hat{x}) - \gamma\|\hat{h}\|^2. \tag{3.6}$$

Let $\varepsilon > 0$ be such that $0 < \alpha - \varepsilon < 1$. Then it follows from the definition of the directional derivative of $f^0(\cdot)$ that there exists a $k_\varepsilon \in \mathcal{N}$ such that

$$\begin{cases} f^0(\hat{x} + \beta^{k_\varepsilon}\hat{h}) - f^0(\hat{x}) & \le \beta^{k_\varepsilon}(\alpha - \varepsilon) df^0(\hat{x}; \hat{h}) \\ & \le \beta^{k_\varepsilon}(\alpha - \varepsilon)[\theta(\hat{x}) - \gamma\|\hat{h}\|^2]. \end{cases} \tag{3.7}$$

Hence,

$$f^0(\hat{x} + \beta^{k_\varepsilon}\hat{h}) - f^0(\hat{x}) - \beta^{k_\varepsilon}\alpha\theta(\hat{x}) \le -\beta^{k_\varepsilon}[\varepsilon\theta(\hat{x}) + (\alpha - \varepsilon)\gamma\|\hat{h}\|^2]. \tag{3.8}$$

Now,

$$\varepsilon\theta(\hat{x}) + (\alpha - \varepsilon)\gamma\|\hat{h}\|^2 > 0 \tag{3.9}$$

for all $\varepsilon > 0$ such that

$$\varepsilon < \varepsilon' := \frac{\alpha\gamma\|\hat{h}\|^2}{-\theta(\hat{x}) + \gamma\|\hat{h}\|^2}. \tag{3.10}$$

Let $\hat{\varepsilon} := \frac{1}{2}\varepsilon'$. Then, since $f^0(\cdot)$ and $\theta(\cdot)$ are continuous and $h_{j_i} \to \hat{h}$, as $i \to \infty$, there exists a $\rho > 0$ such that for all $x_{j_i} \in B(\hat{x};\rho)$,

$$f^0(x_{j_i} + \beta^{k_{\hat{\varepsilon}}}h_{j_i}) - f^0(x_{j_i}) - \beta^{k_{\hat{\varepsilon}}}\alpha\theta(x_{j_i}) < 0, \tag{3.11}$$

which shows that for all $x_{j_i} \in B(\hat{x};\rho)$, $\lambda(x_{j_i}) \ge \beta^{k_{\hat{\varepsilon}}}$. Next, since $\theta(\cdot)$ is continuous, there exists $\hat{\rho} \in (0,\rho)$ such that for all $x_{j_i} \in B(\hat{x};\hat{\rho})$, $\theta(x_{j_i}) \le \frac{1}{2}\theta(\hat{x})$. It therefore follows from the step-size rule (3.2) that for all $x_{j_i} \in B(\hat{x};\hat{\rho})$,

$$f^0(x_{j_i+1}) - f^0(x_{j_i}) \le \beta^{k_{\hat{\varepsilon}}}\alpha\theta(x_{j_i}) \le \tfrac{1}{2}\beta^{k_{\hat{\varepsilon}}}\alpha\theta(\hat{x}) \ . \tag{3.12}$$

Since $\{f^0(x_i)\}_{i=0}^{\infty}$ is monotone decreasing, (3.12) implies that $f^0(x_i) \to -\infty$, as $i \to \infty$, contradicting the fact that $f^0(x_i) \to f^0(\hat{x})$, as $i \to \infty$. Hence we conclude that $\theta(\hat{x}) = 0$, and therefore that $\hat{x} = x^*$. Since by Lemma 3.3, $f^0(\cdot)$ is strongly convex, the whole sequence $\{x_i\}$ converges to x^*. □

4. Rate of convergence of Algorithm 3.1. We will now show that (1.11)–(1.13) hold for Algorithm 3.1.

PROPOSITION 4.1. *Suppose that Assumptions* 1.1 *and* 2.4 *are satisfied and that* $\hat{x}$ *is the unique minimizer of* $f^0(\cdot)$. *Then for all* $x \in \Re^n$,

$$f^0(x) - f^0(\hat{x}) \ge \tfrac{1}{2}cc_F m\|x - \hat{x}\|^2. \tag{4.1}$$

Proof. By Lemma 3.3, $f^0(\cdot)$ is a strongly convex function. Hence, for any $x \in \Re^n$ we have

$$\begin{cases} F(\psi(x)) - F(\psi(\hat{x})) \\ \quad \ge \displaystyle\sum_{j\in\mathbf{m}} \frac{\partial F}{\partial y^j}(\psi(\hat{x}))(\psi^j(x) - \psi^j(\hat{x})) \\ \quad \ge \displaystyle\sum_{j\in\mathbf{m}} \frac{\partial F}{\partial y^j}(\psi(\hat{x})) \max_{k\in\mathbf{q_j}(x)}\{f^{j,k}(\hat{x}) - \psi^j(\hat{x}) \\ \qquad\qquad\qquad + \langle\nabla f^{j,k}(\hat{x}), x - \hat{x}\rangle + \frac{c}{2}\|x - \hat{x}\|^2\} \\ \quad \ge \displaystyle\sum_{j\in\mathbf{m}} \frac{\partial F}{\partial y^j}(\psi(\hat{x})) \max_{k\in\hat{\mathbf{q}}_\mathbf{j}(x)}\{\langle\nabla f^{j,k}(\hat{x}), x - \hat{x}\rangle + \frac{c}{2}\|x - \hat{x}\|^2\}, \end{cases} \tag{4.2}$$

where $\hat{\mathbf{q}}_{\mathbf{j}}(x)$ is defined by (2.7). It now follows from (2.5) and (4.2) that

$$F(\psi(x)) - F(\psi(\hat{x})) \geq df^0(\hat{x}, x - \hat{x}) + \frac{mc_F c}{2}\|x - \hat{x}\|^2, \tag{4.3}$$

Since $df^0(\hat{x}, x - \hat{x}) \geq 0$, (4.1) follows. ◻

PROPOSITION 4.2. *Suppose that Assumptions* 1.1 *and* 2.4 *are satisfied. Then for any compact convex set S there exists a $\kappa > 0$ such that for any $x, z \in S$,*

$$|f^0(x) - \tilde{f}^0(z, x - z)| \leq \kappa\|x - z\|^3, \tag{4.4}$$

where $\tilde{f}^0(z, x - z)$ was defined in (2.17).

Proof. First, it follows from Polak [13, Lemma 2.5.4] or [15] that there exists a constant $L_1 < \infty$ such that for any $x, z \in \Re^n$,

$$|\psi^j(x) - \hat{\psi}^j(z, x - z)| \leq \frac{L_1}{6}\|x - z\|^3, j \in \mathbf{m}. \tag{4.5}$$

Let $S \subset \Re^n$ be a compact set, and let $L_2(\geq C) < \infty$ be a constant associated with S, such that for any $z \in S$,

$$\|\nabla F(\psi(z))\| \leq L_2. \tag{4.6}$$

Then for all $x, z \in S$, by the mean-value theorem, it holds that

$$\left\{\begin{aligned}
& f^0(x) = F(\psi(x)) \\
& \quad = F(\psi(z)) + \langle \nabla F(\psi(z)), \psi(x) - \psi(z)\rangle \\
& \quad + \tfrac{1}{2}\langle \psi(x) - \psi(z), \nabla^2 F(\psi(z))(\psi(x) - \psi(z))\rangle \\
& \quad + \int_0^1 t \int_0^1 \langle \psi(x) - \psi(z), [\nabla^2 F(\psi(z) + st(\psi(x) - \psi(z))) \\
& \qquad\qquad - \nabla^2 F(\psi(z))](\psi(x) - \psi(z))\rangle ds dt \\
& \quad \leq F(\psi(z)) + \langle \nabla F(\psi(z)), \psi(x) - \psi(z)\rangle \\
& \quad + \tfrac{1}{2}\langle \psi(x) - \psi(z), \nabla^2 F(\psi(z))(\psi(x) - \psi(z))\rangle + \tfrac{L_2}{6}\|\psi(x) - \psi(z)\|^3 \\
& \quad = \tilde{f}^0(z, x - z) + \langle \nabla F(\psi(z)), \psi(x) - \hat{\psi}(z, x - z)\rangle \\
& \quad + \tfrac{1}{2}\langle \psi(x) - \psi(z), \nabla^2 F(\psi(z))(\psi(x) - \psi(z))\rangle \\
& \quad - \tfrac{1}{2}\langle \hat{\psi}(z, x - z) - \psi(z), \nabla^2 F(\psi(z))(\hat{\psi}(z, x - z) - \psi(z))\rangle \\
& \quad + \tfrac{L_2}{6}\|\psi(x) - \psi(z)\|^3 .
\end{aligned}\right. \tag{4.7}$$

Thus, according to (4.5) and (4.7), we have

$$\left\{\begin{aligned}
f^0(x) \quad & \leq \tilde{f}^0(z, x - z) + L_3\|x - z\|^3 + \tfrac{L_2}{6}\|\psi(x) - \psi(z)\|^3 \\
& \quad + \tfrac{1}{2}\langle \psi(x) - \psi(z), \nabla^2 F(\psi(z))(\psi(x) - \hat{\psi}(z, x - z))\rangle \\
& \quad + \tfrac{1}{2}\langle \psi(x) - \hat{\psi}(z, x - z), \nabla^2 F(\psi(z))(\hat{\psi}(z, x - z) - \psi(z))\rangle,
\end{aligned}\right. \tag{4.8}$$

where $L_3 := mL_2L_1/6$. For $x, z \in S$ and $j \in \mathbf{m}$, by the definition of $\hat{\psi}^j(\cdot,\cdot)$ (see (2.16)), it holds that

$$\begin{aligned}
&\hat{\psi}^j(z, x-z) - \psi^j(z)\\
&\quad = \max_{k\in\mathbf{q_j}}\{f^{j,k}(z) + \langle \nabla f^{j,k}(z), x-z\rangle + \tfrac{1}{2}\langle x-z, \nabla^2 f^{j,k}(z)(x-z)\rangle\} - \psi^j(z)\\
&\quad \le \max_{k\in\mathbf{q_j}} f^{j,k}(z) + \max_{k\in\mathbf{q_j}}\{\langle \nabla f^{j,k}(z), x-z\rangle + \tfrac{1}{2}\langle x-z, \nabla^2 f^{j,k}(z)(x-z)\rangle\} - \psi^j(z)\\
&\quad = \max_{k\in\mathbf{q_j}}\{\langle \nabla f^{j,k}(z), x-z\rangle + \tfrac{1}{2}\langle x-z, \nabla^2 f^{j,k}(z)(x-z)\rangle\}
\end{aligned}$$

and, on the other hand,

$$\begin{aligned}
&\hat{\psi}^j(z, x-z) - \psi^j(z)\\
&\quad = \max_{k\in\mathbf{q_j}}\{f^{j,k}(z) + \langle \nabla f^{j,k}(z), x-z\rangle + \tfrac{1}{2}\langle x-z, \nabla^2 f^{j,k}(z)(x-z)\rangle\} - \psi^j(z)\\
&\quad \ge \max_{k\in\hat{\mathbf{q}}_{\mathbf{j}}(z)} f^{j,k}(z) + \langle \nabla f^{j,k}(z), x-z\rangle + \tfrac{1}{2}\langle x-z, \nabla^2 f^{j,k}(z)(x-z)\rangle\} - \psi^j(z)\\
&\quad = \max_{k\in\hat{\mathbf{q}}_{\mathbf{j}}(z)}\{\langle \nabla f^{j,k}(z), x-z\rangle + \tfrac{1}{2}\langle x-z, \nabla^2 f^{j,k}(z)(x-z)\rangle\}\ ,
\end{aligned}$$

where the definition of $\hat{\mathbf{q}}_{\mathbf{j}}(z)$ can be found in (2.7). Thus, since S is compact, there exists a positive number L_4 such that for all $x, z \in S$,

$$\|\hat{\psi}(z, x-z) - \psi(z)\| \le L_4\|x-z\|. \tag{4.9}$$

By the Lipschitzian property of ψ and (4.5) it follows that there exists a positive number $L_5(\ge L_4)$ such that for all $x, z \in S$,

$$\|\psi(x) - \psi(z)\| \le L_5\|x-z\| \tag{4.10}$$

and

$$\|\psi(x) - \hat{\psi}(z, x-z)\| \le L_5\|x-z\|^2. \tag{4.11}$$

Hence for all $x, z \in S$,

$$f^0(x) - \tilde{f}^0(z, x-z) \le \kappa\|x-z\|^3 \tag{4.12}$$

with

$$\kappa := L_3 + \frac{L_2L_5}{6} + L_2L_5^2. \tag{4.13}$$

The other half of the inequality of (4.4) follows similarly (with κ as defined in (4.13)). $\square$

THEOREM 4.3. *Suppose that Assumptions* 1.1 *and* 2.4 *are satisfied, that all the* Y_j, $j \in \mathbf{m}$ *are of the form* (1.5), *so that problem* (1.1), (1.2), (1.4) *reduces to problem* (3.1). *If* $\{x_i\}_{i=0}^{\infty}$ *is a sequence constructed by Algorithm* 3.1, *in solving problem* (3.1), *then,* $\{x_i\}_{i=0}^{\infty}$ *converges superlinearly with Q-order at least* 3/2.

Proof. First we will prove that after a finite number of iterations, the step-size λ_i stabilizes to 1, so that eventually $x_{i+1} = x_i + h_i$ holds for the sequence $\{x_i\}_{i=0}^{\infty}$. We will then complete our proof by making use of results in [13, Corollary 2.5.8].

It follows from Theorem 3.4 that the sequence $\{x_i\}_{i=0}^{\infty}$ converges to the unique minimizer $\hat{x}$ of $f^0(\cdot)$. Hence we conclude from Theorem 2.8 that

$$h_i \to 0 \quad \text{as } i \to \infty. \tag{4.14}$$

In view of this, we conclude from Lemma 2.6 that there exist a positive number $\varepsilon > 0$ and a nonnegative integer i_0 such that for all $i \geq i_0$,

$$\hat{f}^0(x_i, h_i) = u(x_i, h_i, 0) + v(x_i, h_i, 0) = \tilde{f}^0(x_i, h_i) = \min_{h \in \Re^n, \|h\| \leq \varepsilon} \tilde{f}^0(x_i, h). \tag{4.15}$$

Suppose that i_0 is sufficiently large to ensure that for all $i \geq i_0$,

$$\|h_i\| \leq \varepsilon, \quad \|x_i - \hat{x}\| \leq \varepsilon\ . \tag{4.16}$$

Then, making use of (4.1), we find that, for $i = i_0, i_0 + 1, i_0 + 2, \ldots$,

$$\left\{ \begin{aligned} &\tfrac{1}{2} c c_F m \|x_i + h_i - \hat{x}\|^2 \\ &\quad \leq f^0(x_i + h_i) - f^0(\hat{x}) \\ &\quad = f^0(x_i + h_i) - \tilde{f}^0(x_i, h_i) + \tilde{f}^0(x_i, h_i) - f^0(\hat{x}) \\ &\quad \leq f^0(x_i + h_i) - \tilde{f}^0(x_i, h_i) + \tilde{f}^0(x_i, \hat{x} - x_i) - f^0(\hat{x}), \end{aligned} \right. \tag{4.17}$$

because $\tilde{f}^0(x_i, h_i) \leq \tilde{f}^0(x_i, \hat{x} - x_i)$, by (4.15). It now follows from Proposition 4.2 that there exists a $\kappa > 0$ such that for all $i \geq i_0$,

$$\left\{ \begin{aligned} &\tfrac{1}{2} c c_F m \|x_i + h_i - \hat{x}\|^2 \\ &\quad \leq \kappa(\|x_i + h_i - x_i\|^3 + \|x_i - \hat{x}\|^3) \\ &\quad \leq \kappa[(\|x_i + h_i - \hat{x}\| + \|x_i - \hat{x}\|)^3 + \|x_i - \hat{x}\|^3]. \end{aligned} \right. \tag{4.18}$$

Now, by Theorem 2.9, there exist a positive integer $i_1 \geq i_0$ and an $\varepsilon_1 > 0$ such that for all $i \geq i_1$,

$$\theta(x_i) \leq -\varepsilon_1 \|h_i\|^2. \tag{4.19}$$

Next, Proposition 4.2 and (4.15) imply that for all $i \geq i_1$,

$$\left\{ \begin{aligned} \theta(x_i) &= \hat{f}^0(x_i, h_i) - f^0(x) \\ &= \tilde{f}^0(x_i, h_i) - f^0(x) \\ &= \tilde{f}^0(x_i, h_i) - f^0(x_i + h_i) + f^0(x_i + h_i) - f^0(x_i) \\ &\geq -\kappa \|h_i\|^3 + f^0(x_i + h_i) - f^0(x_i). \end{aligned} \right. \tag{4.20}$$

Hence, from (4.20) and (4.19), we have

$$\left\{\begin{array}{lll} f^0(x_i+h_i)-f^0(x_i)-\alpha\theta(x_i) & \leq & (1-\alpha)\theta(x_i)+\kappa\|h_i\|^3 \\ & \leq & -(1-\alpha)\varepsilon_1\|h_i\|^2+\kappa\|h_i\|^3. \end{array}\right. \tag{4.21}$$

It now follows from (4.21) and the fact that $h_i \to 0$ as $i \to \infty$ that for all i sufficiently large,

$$x_{i+1} = x_i + h_i. \tag{4.22}$$

We therefore conclude from [13, Corollary 2.5.8] or [15], (4.18), and (4.19) that $\{x_i\}_{i=0}^{\infty}$ converges to $\hat{x}$ superlinearly with Q-order at least 3/2. □

5. 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 (1.1), (1.2), (1.4). 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 [13]) and use them in conjunction with a master algorithm that calls Algorithm 3.1 as a subroutine. We will see that under a reasonable assumption, the resulting algorithm retains the rate of convergence of Algorithm 3.1.

5.1. Consistent approximations. Let N_0 be a strictly positive integer, and, for $N \in \mathcal{N}_0 := \{N_0, N_0+1, N_0+2, \ldots\}$, 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 $\mathbf{P_N}$, $N \in \mathcal{N}_0$, as follows:

$$\mathbf{P_N} \qquad \min_{x \in \Re^n} f_N^0(x), \tag{5.1}$$

where

$$f_N^0(x) := F(\psi_N(x)), \tag{5.2}$$

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

$$\psi_N^j(x) = \max_{y_j \in Y_{j,N}} \phi^j(x, y_j) \ . \tag{5.3}$$

It should be clear that the approximating problems $\mathbf{P_N}$ are of the form (3.1) and that one can define optimality functions $\theta_N(\cdot)$ for them of the form (2.19). We will refer to the original problem (1.1), (1.2), (1.4) as $\mathbf{P}$.

DEFINITION 5.1 (see [13]). *We will say that the pairs $(\mathbf{P_N}, \theta_\mathbf{N})$ in the sequence $\{(\mathbf{P_N}, \theta_\mathbf{N})\}_{\mathbf{N}\in\mathcal{N}_0}$ are consistent approximations to 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 [13]) and for any infinite sequence $\{x_N\}_{N\in K}$, $K \subset \mathcal{N}_0$, such that $x_N \to^K x$, $\overline{\lim}_{N\in K}\theta_N(x_N) \leq \theta(x)$.*

Assumption 5.2. We will assume as follows:

(a) For every $N \in \mathcal{N}_0$, the problem (5.1) has a solution.

(b) There exists a strictly positive valued, strictly monotone decreasing function $\Delta : \mathcal{N} \to \Re$, such that $\Delta(N) \to 0$, as $N \to \infty$, and a $L < \infty$, 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

$$\|y - y'\| \leq L\Delta(N). \tag{5.4}$$

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

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

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

For any $x, h \in \Re^n$ and $w \in \Re^m$, we define

$$u_N(x,h,w) := \langle \nabla F(\psi_N(x)), \hat{\psi}_N(x,h) - \psi_N(x) + w\rangle \tag{5.5}$$

and

$$\begin{cases} v_N(x,h,w) \\ \quad = \frac{1}{2}\langle \hat{\psi}_N(x,h) - \psi_N(x) + w, \nabla^2 F(\psi_N(x))(\hat{\psi}_N(x,h) - \psi_N(x) + w)\rangle, \end{cases} \tag{5.6}$$

where

$$\hat{\psi}_N(x,h) = (\hat{\psi}_N^1(x,h), \ldots, \hat{\psi}_N^m(x,h)) \tag{5.7}$$

and

$$\hat{\psi}_N^j(x,h) = \max_{y_j\in Y_{j,N}}\{\phi^j(x,y_j) + \langle \nabla_x\phi^j(x,y_j), h\rangle + \tfrac{1}{2}\langle h, \nabla_x^2\phi^j(x,y_j)h\rangle\}\ . \tag{5.8}$$

We infer from (2.19) that the optimality functions $\theta_N(\cdot)$, for the problems $\mathbf{P_N}$ have the following form:

$$\theta_N(x) := \min_{h\in\Re^n}\{\min_{w\in\Re^m_+}(u_N(x,h,w) + v_N(x,h,w))\}. \tag{5.9}$$

Since the cardinality of the sets $Y_{j,N}$ is finite, it is obvious that the $\theta_N(x)$ can be evaluated.

As was also done in the Polak–Mayne–Higgins rate-preserving method [16] (see also [17]), we use an alternative optimality function for the problems $\mathbf{P_N}$ for precision adjustment in our algorithm. This optimality function is defined by

$$\bar{\theta}_N(x) := \min_{h\in\Re^n}\bar{f}_N^0(x,h) - \sum_{j\in\mathbf{m}}\frac{\partial F}{\partial y^j}(\psi_N(x))\psi_N^j(x), \tag{5.10}$$

where

$$\begin{cases} \bar{f}_N^0(x,h) \\ \quad = \displaystyle\sum_{j\in\mathbf{m}}\frac{\partial F}{\partial y^j}(\psi_N(x))\max_{y_j\in Y_{j,N}}[\phi^j(x,y_j) + \langle\nabla_x\phi^j(x,y_j),h\rangle + \tfrac{1}{2}\delta\|h\|^2], \end{cases} \tag{5.11}$$

with $\delta > 0$, a constant.

Similarly (as in [17]), we define an alternative optimality function for the problem $\mathbf{P}$ by

$$\bar{\theta}(x) := \min_{h\in\Re^n}\bar{f}^0(x,h) - \sum_{j\in\mathbf{m}}\frac{\partial F}{\partial y^j}(\psi(x))\psi^j(x), \tag{5.12}$$

where

$$\bar{f}^0(x,h) = \sum_{j\in\mathbf{m}} \frac{\partial F}{\partial y^j}(\psi(x)) \max_{y_j\in Y_j}[\phi^j(x,y_j) + \langle \nabla_x \phi^j(x,y_j), h\rangle + \tfrac{1}{2}\delta\|h\|^2], \tag{5.13}$$

with $\delta > 0$ the same constant as in (5.11).

PROPOSITION 5.3 (see [17]). *Suppose that Assumptions* 1.1 *and* 5.2 *are satisfied and that for all* $N \in \mathcal{N}_0$, $f_N^0(\cdot)$ *is defined by* (5.2) *and* $\bar{\theta}_N(\cdot)$ *by* (5.10). *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* $x \in S, N \in \mathcal{N}_0$,

$$|f_N^0(x) - f^0(x)| \le C_S\Delta(N), \tag{5.14}$$

and

$$|\bar{\theta}_N(x) - \bar{\theta}(x)| \le C_S\Delta(N). \tag{5.15}$$

5.2. The superlinear rate-preserving algorithm.

ALGORITHM 5.4 (solves problem (1.1), (1.2), (1.4)).

Parameters. $\alpha, \beta \in (0,1)$, $\delta > 0$, $D > 0$, $\sigma \ge 3$.

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* $\bar{\theta}_N(x_i)$ *according to* (5.10) *and* (5.11).

Step 2. *If*

$$D\Delta(N) \le |\bar{\theta}_N(x_i)|^\sigma, \tag{5.16}$$

go to Step 3. *Else, replace* N *by* $N+1$, *and go to Step* 1.

Step 3. *Compute the second* optimality function value $\theta_N(x_i)$ *according to* (5.9), *i.e.*,

$$\theta_N(x_i) = \min_{h\in\Re^n}\{\min_{w\in\Re^m_+}(u_N(x_i,h,w) + v_N(x_i,h,w))\} \tag{5.17}$$

and the corresponding search direction h_i *according to*

$$h_i \in \arg\min_{h\in\Re^n}\{\min_{w\in\Re^m_+}(u_N(x_i,h,w) + v_N(x_i,h,w))\}. \tag{5.18}$$

Step 4. *Compute the step-size*

$$\lambda_i = \max_{k\in\mathcal{N}}\{\beta^k |\ f_N^0(x_i + \beta^k h_i) - f_N^0(x_i) - \beta^k\alpha\theta_N(x_i) \le 0\}, \tag{5.19}$$

and go to Step 5.

Step 5. *Set*

$$x_{i+1} = x_i + \lambda_i h_i. \tag{5.20}$$

Set $N_i = N$, *replace* i *by* $i+1$, *and go to Step* 1.

Remark.

(a) It follows from Proposition 5.3 that $\bar{\theta}_N(x_i) \to \bar{\theta}(x_i)$, as $N \to \infty$. Hence, whenever $\bar{\theta}(x_i) \ne 0$, the loop consisting of Step 1 and Step 2 of Algorithm 5.4 yields a finite discretization parameter N_i. For simplicity, we will assume that Algorithm 5.4 does not produce an iterate x_i such that $\bar{\theta}(x_i) = 0$.

(b) Note that the work needed to compute x_i by Algorithm 5.4 increases with the iteration number i.

The purpose of the following results is to show that the relations (1.15)–(1.17) hold for Algorithm (5.4)

LEMMA 5.5. *Suppose that Assumptions* 1.1*,* 2.4*, and* 5.2 *are satisfied and that Algorithm* 5.4 *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 $\{x_i\}_{i=0}^{\infty}$ has an accumulation point $\hat{x}$ and 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 $i_0 \in \mathcal{N}$, 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 3.1 applied to problem (5.1) with $N = N^*$. Furthermore, it follows from (5.16) that there exists an $\varepsilon > 0$, such that $\bar{\theta}_i = \bar{\theta}_{N^*}(x_i) \leq -\varepsilon$ for all $i \geq i_0$. However, it follows from Theorem 3.4 that $\theta_{N^*}(\hat{x}) = 0$. Thus, by (iii) of Theorem 2.8, $0 \in \partial f^0_{N^*}(x_i)$. By [17, Theorem 2], $0 \in \partial f^0_{N^*}(x_i)$ implies $\bar{\theta}_{N^*}(x_i) = 0$. Then, from the continuity of $\bar{\theta}_{N^*}(\cdot)$ [17, Theorem 2], it holds that $\bar{\theta}_{N^*}(x_i) \to \bar{\theta}_{N^*}(\hat{x}) = 0$ as $i \to \infty, i \in K$, where the infinite subsequence $\{x_i\}_{i \in K}$, $K \subset \mathcal{N}$, converges to $\hat{x}$, which contradicts the previous finding, and hence completes our proof. ∎

THEOREM 5.6. *Suppose that Assumptions* 1.1*,* 2.4*, and* 5.2 *are satisfied and that Algorithm* 5.4 *has constructed a bounded sequence* $\{x_i\}_{i=0}^{\infty}$*. Then every accumulation point* $\hat{x}$ *of* $\{x_i\}_{i=0}^{\infty}$ *satisfies* $\bar{\theta}(\hat{x}) = 0$*.*

Proof. By applying Theorem 3.3.23 of [13] or theorems in section 5 of [12] and Lemma 5.5 to Algorithm 5.4, we obtain the desired result. ∎

THEOREM 5.7. *Suppose that Assumptions* 1.1*,* 2.4*, and* 5.2 *are satisfied and that Algorithm* 5.4 *has constructed a bounded sequence* $\{x_i\}_{i=0}^{\infty}$*. Then* $\{x_i\}$ *converges to the unique minimizer* $\hat{x}$ *of* $f^0(\cdot)$ *with Q-order* 3/2*.*

Proof. First, by Theorem 5.6 and the fact that $f^0(\cdot)$ has a unique minimizer $\hat{x}$, the whole sequence $\{x_i\}$ converges to $\hat{x}$. Hence one can deduce from Theorem 4.3 and the proof of [13, Theorem 3.4.20], that $\{x_i\}$ converges to $\hat{x}$ with Q-order 3/2. Since the derivation is straightforward, we omit the details here. ∎

6. Some numerical results.

We now present some numerical results that illustrate the behavior of the algorithm proposed in section 5 for generalized semi-infinite programming problems. The algorithm was implemented in Matlab. Throughout the computational experiments, the parameters used in the algorithm were $\alpha = 0.05, \beta = 0.5, \delta = 1.0, D = 10^{-10}$, and $\sigma = 3.1$. For both examples, we used the starting point $(1, 1)$. The iteration of the algorithm is stopped at x_i if for some N the meshsize $\Delta(N) < 0.005$ and $|\theta_N(x_i)| \leq 10^{-8}$. A Matlab code developed in [21], which was based on a smoothing Newton method [19] for variational inequalities, was used to solve our search direction finding subproblem (2.57).

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

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

Table 6.1
Numerical results for Example 1.

Iteration i	0	1	2	3	4
$\|\|x_i - \hat{x}\|\|$	1.6×10^0	5.7×10^{-1}	5.1×10^{-2}	2.9×10^{-4}	0.0
Discretization level	1	1	1	1	9

Table 6.2
Numerical results for Example 2.

Iteration i	0	1	2	3	4
$\|\|x_i - \hat{x}\|\|$	1.7×10^0	6.4×10^{-1}	5.0×10^{-2}	1.9×10^{-4}	0.0
Discretization level	1	1	1	1	9

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 $Y_1 = [0,1]$ and $Y_2 = [-1,0]$.

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

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

The numerical results are summarized in Table 6.1 and Table 6.2. In these two tables the first row represents the iteration number, the second row is the residue $||x_i - \hat{x}||$ (we used the last iterate as a substitute for $\hat{x}$) and the third row shows the discretization level (the meshsize at the present level is decreased to half of the previous one) refined by the master algorithm at the i-th step. It is clear from the numerical results that the rate of convergence is superlinear.

7. Conclusion. We have presented two superlinearly converging algorithms, one for solving finite generalized min-max problems of the form (1.1), (1.2), (1.3) and one for solving generalized semi-infinite min-max problems of the form (1.1), (1.2), (1.4). These algorithms were obtained by making use of the concepts underlying the construction of the Polak–Mayne–Higgins Newton's method [15] and the Polak–Mayne–Higgins rate-preserving method [16], respectively. The construction of the algorithms depends on the cost function having a subgradient and their rate of convergence depends on convexity and second order smoothness, and hence Assumption 2.4 is essential.

Our numerical results are consistent with our theoretical prediction that the algorithms converge Q-superlinearly.

Acknowledgments. The authors wish to thank Prof. R. T. Rockafellar for suggesting the function $\hat{f}^0(x,h)$ as a way to get around the possible nonconvexity of the function $\tilde{f}^0(x,h)$ in h, as well as for the formula (2.57) which shows that our optimality function is defined by a quadratically constrained quadratic programming problem. They also thank two anonymous referees for their helpful comments and suggestions.

REFERENCES

[1] D. P. Bertsekas, *Nondifferentiable optimization via approximation*, Math. Programming Study, 3 (1975), pp. 1–25.
[2] F. H. Clarke, *Optimization and Nonsmooth Analysis*, John Wiley, New York, 1983.
[3] V. F. Dem'yanov and A. M. Rubinov, *Quasidifferential Calculus*, Optimization Software Inc., New York, 1986.
[4] A. der Kiureghian, *private communication*, Department of Civil Engineering, University of California at Berkeley, 1998.
[5] G. Di Pillo, L. Grippo, and S. Lucidi, *A smooth transformation of the generalized minimax problem*, J. Optim. Theory Appl., 95 (1997), pp. 1–24.
[6] K. C. Kiwiel, *A quadratic approximation method for minimizing a class of quasidifferentiable functions*, Numer. Math., 45 (1984), pp. 411–430.
[7] K. C. Kiwiel, *A linearization method for minimizing certain quasidifferentiable functions*, Math. Programming Study, 29 (1986), pp. 85–94.
[8] K. C. Kiwiel, *Descent methods for quasidifferentiable minimization*, Appl. Math. Optim., 18 (1988), pp. 163–180.
[9] G. Papavassilopoulos, *Algorithms for a class of nondifferentiable problems*, J. Optim. Theory Appl., 34 (1981), pp. 41–82.
[10] O. Pironneau and E. Polak, *On the rate of convergence of certain methods of centers*, Math. Programming, 2 (1972), pp. 230–258.
[11] E. Polak, *Basics of minimax algorithms*, in Nonsmooth Optimization and Related Topics, F. H. Clarke, V. F. Dem'yanov, and F. Giannessi, eds., Plenum Press, New York, 1989, pp. 343–367.
[12] E. Polak, *On the use of consistent approximations in the solution of semi-infinite optimization and optimal control problems*, Math. Programming, 62 (1993), pp. 385–414.
[13] E. Polak, *Optimization: Algorithms and Consistent Approximations*, Springer, New York, 1997.
[14] E. Polak and L. He, *Rate-preserving discretization strategies for semi-infinite programming and optimal control*, SIAM J. Control Optim., 30 (1992), pp. 548–572.
[15] E. Polak, D. Q. Mayne, and J. Higgins, *A superlinearly convergent algorithm for minmax problems*, J. Optim. Theory Appl., 89 (1991), pp. 407–439.
[16] E. Polak, D. Q. Mayne and J. Higgins, *On the extension of Newton's method to semi-infinite for minimax problems*, SIAM J. Control Optim., 30 (1992), pp. 367–389.
[17] E. Polak, L. Qi, and D. Sun, *First-Order algorithms for generalized finite and semi-infinite min-max problems*, Comput. Optim. Appl., 13 (1999), pp. 137–161.
[18] B. N. Pshenichnyi and Yu. M. Danilin, *Numerical Methods in Extremal Problems*, Nauka, Moscow, 1975 (in Russian).
[19] L. Qi, D. Sun, and G. Zhou, *A new look at smoothing Newton methods for nonlinear complementarity problems and box constrained variational inequalities*, Math. Programming, 87 (2000), pp. 1–35.
[20] S. J. Wright, *Primal-Dual Interior-Point Methods*, SIAM, Philadelphia, PA, 1996.
[21] G. Zhou, D. Sun, and L. Qi, *Numerical experiments for a class of squared smoothing Newton methods for box constrained variational inequality problems*, in Reformulation: Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods, M. Fukushima and L. Qi, eds., Kluwer, 1999, pp. 421–441.