ON THE EXTENSION OF NEWTON'S METHOD TO SEMI-INFINITE MINIMAX PROBLEMS

by

E. Polak, D. Q. Mayne, and J. E. Higgins

Memorandum No. UCB/ERL M89/92

1 August 1989
(Revised October 9, 1990)

ON THE EXTENSION OF NEWTON'S METHOD TO SEMI-INFINITE MINIMAX PROBLEMS

by

E. Polak, D. Q. Mayne, and J. E. Higgins

Memorandum No. UCB/ERL M89/92

1 August 1989
(Revised October 9, 1990)

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

ON THE EXTENSION OF NEWTON'S METHOD TO SEMI-INFINITE MINIMAX PROBLEMS

by

E. Polak, D. Q. Mayne, and J. E. Higgins

Memorandum No. UCB/ERL M89/92

1 August 1989
(Revised October 9, 1990)

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

ON THE EXTENSION OF NEWTON'S METHOD TO SEMI-INFINITE MINIMAX PROBLEMS[†]

E. Polak*, D. Q. Mayne** and J. E. Higgins*

ABSTRACT

This paper introduces two new techniques for the analysis and construction of semi-infinite optimization algorithms. The first is a very simple technique for establishing the superlinear rate of convergence of semi-infinite optimization algorithms. The second technique enables one to specify discretization rules which preserve the superlinear convergence of *conceptual* superlinearly converging semi-infinite optimization algorithms.

We use natural extensions of Newton's method to semi-infinite optimization, as a vehicle for presenting our techniques. In particular, we show that both local and global versions of the conceptual extension of Newton's method converge Q-superlinearly, with rate at least 3/2, and that their implementations, based on our discretization rules, retain this rate of convergence.

KEY WORDS

Newton's method, minimax problems, non-differentiable optimization, superlinear convergence.

[†] The research reported herein was sponsored in part by the National Science Foundation grant ECS-8713334, the Air Force Office of Scientific Research contract AFOSR-86-0116 and the State of California MICRO Program grant 532410-19900.

*Department of Electrical Engineering and Computer Sciences and the Electronics Research Laboratory, University of California, Berkeley, CA 94720, U.S.A.

**Department of Electrical Engineering, Imperial College of Science and Technology, London, SW7 2BT, Great Britain.

1. INTRODUCTION

This is a dual purpose paper. The first purpose of this paper is to introduce a novel technique for establishing the superlinear of convergence of a class of semi-infinite optimization algorithms; the second is to demonstrate the degree to which various discretization effects, associated with semi-infinite optimization problems can be taken into account. In particular, this paper introduces discretization rules which preserve the superlinear convergence of *conceptual* superlinearly converging semi-infinite optimization algorithms.

In his pioneering paper [27], dealing with perturbed Kuhn-Tucker points, Robinson showed that by applying the Implicit Function Theorem to the first order optimality conditions of a finitely constrained optimization problem and then relating the result to the search direction finding problem of a particular algorithm, one can sometimes establish the superlinear convergence of this algorithm. The same technique can also be used for establishing the superlinear convergence of finite minimax algorithms, see e.g., [25].

Unfortunately, Robinson's technique cannot be used in conjunction with semi-infinite optimization algorithms because the assumptions of the Implicit Function Theorem cannot be met in the semi-infinite case. The technique in this paper is based on function approximations and is therefore not restricted by the by the linear independence requirements associated with Implicit Function Theorem based techniques.

To illustrate both our new technique for establishing the superlinear convergence of a semi-infinite optimization algorithm and the manner in which discretization effects can be taken into account, we chose an extension of Newton's method for the solution of semi-infinite optimization problems. Our choice was motivated partly by the fact that Newton's method is the simplest method method in the class that can be considered, and partly because Newton's method is one of the best understood, most studied, variously modified, adapted, and approximated algorithms in the literature (see, e.g., [5, 18, 8, 15, 16, 19, 24, 27, 28]).

In the area of nonlinear programming, it was at first used as a *local* method for unconstrained optimization of twice locally Lipschitz continuously differentiable, strongly convex functions on $\mathbb{R}^n$. Then, it was shown by Goldstein [7] that, for such functions, the local Newton method can be globally stabilized, i.e., it can be made globally convergent by the addition of the Armijo-Goldstein step size rule [1, 7]. Since this step size rule returns a step size of of unity near a solution (see [7]), the Goldstein version of the globally stabilized Newton method converges quadratically. Finally, referring to [20], we see that it is possible to construct globally stabilized versions of Newton's method which converge quadratically in minimizing twice locally Lipschitz continuously differentiable, but not necessarily convex, functions on $\mathbb{R}^n$, whose local minimizers satisfy second order sufficiency conditions[1].

The extension of Newton's method (largely in the form of sequential quadratic programming) to semi-infinite optimization problems appears to have been confined to constrained problems which can be converted to ordinary nonlinear programming problems by means of the Implicit Function Theorem (see, e.g. [10, 23, 4]). For example, a problem of the form

$$\min \{ f(x) \mid \phi(x, t) \le 0, \ \forall t \in [0, 1] \} , \tag{1.1a}$$

can be converted to the standard nonlinear programming form

$$\min \{ f(x) \mid \phi(x, t^j(x)) \le 0, j = 1, 2, \ldots, q \} , \tag{1.1b}$$

when it is known that for all x near a local solution x^*, $\phi(x, \cdot)$ has exactly q local maximizers, and that $\phi_{tt}(x, t^j(x)) < 0$ for each j. It should be noted that some of these extensions are *conceptual algorithms* because in their analysis it was not taken into account the fact that the local maximizers $t^j(x)$ cannot be computed exactly.

One can convert an unconstrained minimax problem of the form

$$\min_{x \in \mathbb{R}^n} \max_{t \in [0, t]} \phi(x, t) \tag{1.2a}$$

into a constrained problem of the form

$$\min \{ w \mid \phi(x, t) \le w, \ \forall t \in [0, t] \} , \tag{1.2b}$$

and, assuming that the required assumptions are satisfied, apply one of the above mentioned algorithms (i.e., [10, 23, 4]). Such an approach suffers from both aesthetic and practical drawbacks. First, it is displeasing to convert an unconstrained optimization problem into a constrained one. Second, to avoid the Maratos effect [14], one must use a curvilinear step size rule or other modifications, such as the use of the modified Lagrangian merit function of Shittkowski and Powell, which are more complex than the simple Armijo-Goldstein rule mentioned earlier. Third, unlike Newton's method, the methods in [10, 23, 4] do not exhibit invariance under linear transformations. Last, but not least, we have observed that constrained semi-infinite optimization algorithms (such as Algorithm 5.7 in [21]) do not perform on (1.2b) as well as semi-infinite minimax algorithms (such as the version of Algorithm 5.2, based on (5.52) in [21]) do on (1.2a).

In this paper, we present natural extensions of both the local version of Newton's method and of the Goldstein globally stabilized version of Newton's method, for the solution of a class of convex semi-infinite minimax problems. The notable aspects of our work are *(i)* we do not impose the above

[1] Such globally converging methods are obtained by using the Goldstein method if certain conditions are satisfied, and reverting to the Armijo Gradient Method [1] otherwise (see, e.g., [17] for an example).

mentioned restrictive assumption that all the local maximizers are strict and that their number is finite, *(ii)* we take into account the most obvious approximations required to produce implementable algorithms, and *(iii)* we use a new and very simple technique for establishing superlinear convergence of our extensions of Newton's method. Since our technique is not based on the Implicit Function Theorem (as in [27, 25]), it does not require the imposition of a strict complementarity condition[2]. In Section 2 we show that a *conceptual* local Newton's method for semi-infinite minimax problems converges superlinearly with Q-rate 3/2, under assumptions analogous to those needed in the minimization of twice locally Lipschitz continuously differentiable, strongly convex functions on $\mathbb{R}^n$. In Section 3 we present a *conceptual* globally stabilized Newton method and show that it retains the Q-rate of 3/2. In Section 4 we present two *implementable* versions of the local Newton method for semi-infinite minimax problems and show that they converge locally with Q-rate 3/2; a superlinearly converging (with Q-rate 3/2) *implementable* version of our globally stabilized Newton method is presented in Section 5. We present numerical results in Section 6 and our concluding comments and final observations in Section 7.

2. THE LOCAL NEWTON METHOD

We will consider the problem

$$\mathbf{P}: \quad \min_{x \in \mathbb{R}^n} \psi(x), \tag{2.1a}$$

where

$$\psi(x) = \max_{j \in \mathbf{q}} \max_{t \in [0,1]} \phi^j(x, t), \tag{2.1b}$$

where $\mathbf{q} \triangleq \{1, 2, \ldots, q\}$.

In keeping with standard assumptions for Newton's method (see [7]), we make the following hypotheses:

Assumption 2.1:

(i) For all $j \in \mathbf{q}$, the functions $\phi^j : \mathbb{R}^n \times \mathbb{R}$ are twice Lipschitz continuously differentiable in the first argument (uniformly in the second).

(ii) For all $j \in \mathbf{q}$, $\phi^j(\cdot, \cdot)$, $\nabla_x \phi(\cdot, \cdot)$ and $\phi^j_{xx}(\cdot, \cdot)$ are all continuous.

(iii) There exist constants $0 < m \leq M$ such that for all $x \in \mathbb{R}^n$,

[2] Referring to problem (1.2a), we note that an assumption of strict complementary slackness is highly restrictive for semi-infinite minimax problems, because it implies that, at a solution x^*, the active gradients $\nabla_x \phi(x^*, t)$ are affinely independent, which in turn, implies that $\phi(x^*, \cdot)$ has at most $n + 1$ maximizers. However, $\phi(x^*, \cdot)$ may well have a *continuum* of maximizers.

$$m\|h\|^2 \le \langle h, \phi^j_{xx}(x,t)h\rangle \le M\|h\|^2, \quad \forall t \in [0,1], \quad \forall j \in \mathbf{q}. \tag{2.2}$$

Proposition 2.1: Suppose that Assumption 2.1 holds and that x^* is the minimizer of $\psi(\cdot)$. Then for all $x \in \mathbb{R}^n$,

$$\psi(x) - \psi(x^*) \ge \frac{m}{2}\|x - x^*\|^2. \tag{2.3a}$$

Proof: For any $x \in \mathbb{R}^n$, and for any $j \in \mathbf{q}$, let

$$\mathbf{q}^*(x) \triangleq \{ j \in \mathbf{q} \mid \psi(x) = \max_{t \in [0,1]} \phi^j(x,t) \} \tag{2.3b}$$

$$T^{*j}(x) \triangleq \{ t \in [0,1] \mid \phi^j(x,t) = \psi(x) \}. \tag{2.3c}$$

Then, making use of the second order expansion formula [6, p.185] and of (2.2), we obtain that

$$\begin{aligned}\psi(x) - \psi(x^*) &\ge \max_{j \in \mathbf{q}} \max_{t \in [0,1]} \phi^j(x^*,t) - \psi(x^*) + \langle \nabla_x \phi^j(x^*,t), x - x^* \rangle + \frac{m}{2}\|x - x^*\|^2 \\ &\ge \max_{j \in \mathbf{q}^*(x^*)} \max_{t \in T^{*j}(x^*)} \phi^j(x^*,t) - \psi(x^*) + \langle \nabla_x \phi^j(x^*,t), x - x^* \rangle + \frac{m}{2}\|x - x^*\|^2 \\ &= d\psi(x^*, x - x^*) + \frac{m}{2}\|x - x^*\|^2, \end{aligned} \tag{2.3d}$$

where $d\psi(x^*, x - x^*)$ denotes the directional derivative of $\psi(\cdot)$ at x^*, in the direction $(x - x^*)$. Since x^* is the minimizer of $\psi(\cdot)$, $d\psi(x^*, x - x^*) \ge 0$, and hence (2.3a) follows. □

By analogy with Newton's method for differentiable functions, we define a quadratic approximation $\hat{\psi}(\cdot \mid y)$ to $\psi(\cdot)$, around the point y, by

$$\hat{\psi}(x \mid y) \triangleq \max_{j \in \mathbf{q}} \max_{t \in [0,1]} \phi^j(y,t) + \langle \nabla_x \phi^j(y,t), x - y \rangle + \tfrac{1}{2} \langle (x-y), \phi^j_{xx}(y,t)(x-y) \rangle. \tag{2.4a}$$

Algorithm 2.1 (Local Newton Method).

Data: $x_0 \in \mathbb{R}^n$.

Step 0: Set $i = 0$.

Step 1: Compute

$$x_{i+1} = \arg\min_{x \in \mathbb{R}^n} \hat{\psi}(x \mid x_i). \tag{2.4b}$$

Step 2: Replace i by $i+1$ and go to Step 1. □

Proceeding as in the proof of Proposition 2.1, it can be shown that $\hat{\psi}(x \mid x_i) - \hat{\psi}(x_{i+1} \mid x_i) \ge \frac{1}{2} m \| x - x_{i+1} \|^2$. Hence we conclude that x_{i+1} is uniquely defined by (2.4b).

To establish the local convergence and rate of convergence of the above algorithm, we shall need the following lemmas.

Lemma 2.1: Suppose that Assumption 2.1 holds. Then there exists a $\hat{K} < \infty$ such that for any $x, y \in \mathbb{R}^n$,

$$| \psi(x) - \hat{\psi}(x \mid y) | \le \hat{K} \| x - y \|^3 . \tag{2.5}$$

Proof: Let $L < \infty$ be a common Lipschitz constant for the Hessians $\phi^j_{xx}(\cdot, \cdot)$. Then, making use of second order expansions, we obtain that

$$\psi(x) = \max_{j \in \mathbf{q}} \max_{t \in [0,1]} \phi^j(y, t) + \langle \nabla_x \phi^j(y, t), x - y \rangle + \tfrac{1}{2} \langle (x - y), \phi^j_{xx}(y, t)(x - y) \rangle$$

$$+ \int_{s \in [0,1]} (1 - s) \langle (x - y), [\phi^j_{xx}(y + s(x - y), t) - \phi^j_{xx}(y, t)](x - y) \rangle \, ds$$

$$\le \psi(x \mid y) + \frac{L}{6} \| x - y \|^3 . \tag{2.6}$$

The other half of the inequality in (2.5) follows similarly (with $\hat{K} = L/6$). □

Lemma 2.2: Suppose that Assumption 2.1 is satisfied. Let $h : \mathbb{R}^n \to \mathbb{R}^n$ and $\theta : \mathbb{R}^n \to \mathbb{R}$ be defined by

$$h(x) = \arg \min_{h \in \mathbb{R}^n} \hat{\psi}(x + h \mid x) , \tag{2.7a}$$

$$\theta(x) = \min_{h \in \mathbb{R}^n} \hat{\psi}(x + h \mid x) - \psi(x) . \tag{2.7b}$$

Then *(a)* both $h(\cdot)$ and $\theta(\cdot)$ are continuous. *(b)* For all $x \in \mathbb{R}^n$, $d\psi(x, h(x)) \le \theta(x)$. *(c)* If x^* is a solution of (2.1a), then both $h(x^*) = 0$ and $\theta(x^*) = 0$. *(d)* For all $x \ne x^*$, $\theta(x) < 0$.

Proof: *(a)* Continuity of $\theta(\cdot)$ and $h(\cdot)$ follows from the Maximum Theorem in [2], strengthened by Assumption 2.1 *(iii)*. The continuity of $h(\cdot)$ again follows from the Maximum Theorem in [2], which states that it is an upper-semicontinuous set valued map, and the fact that $h(x)$ is always a singleton. *(b)* Clearly, with $q^*(x)$, $T^{*j}(x)$ defined as in (2.3b, c), we must have that

$$\theta(x) \ge \max_{j \in q^*(x)} \max_{t \in T^{*j}(x)} \phi^j(x, t) - \psi(x) + \langle \nabla_x \phi^j(x, t), h(x) \rangle = d\psi(x, h(x)) . \tag{2.7c}$$

Hence $d\psi(x, h(x)) \le \theta(x)$. *(c)* Since $0 \le d\psi(x^*, h(x^*)) \le \theta(x^*) \le 0$, must hold, it follows that both $h(x^*) = 0$ and $\theta(x^*) = 0$. *(d)* In view of Assumption 2.1 *(iii)*, x^* is the only point satisfying the first order necessary and sufficient condition $0 \in \partial\psi(x^*)$, where $\partial\psi(x)$ denotes the Clarke generalized gradient of $\psi(\cdot)$ at x (for a definition, see page 27 in [3]). Hence this part is a simple generalization of Proposition 5.5 in [21], from which we see that $\theta(x) \le 0$ for all $x \in \mathbb{R}^n$ and that $\theta(x) = 0$ if and only if $0 \in \partial\psi(x)$. □

Lemma 2.3: Suppose that $K \in (0, \infty)$, and that $t, s \ge 0$ are such that

$$t^2 \le K[(s+t)^3 + s^3], \tag{2.8a}$$

$$0 \le t \le \frac{1}{9K}, \quad 0 \le s \le \frac{1}{9K}. \tag{2.8b}$$

Then $t \le s$ and

$$t \le 3\sqrt{K}\, s^{3/2}. \tag{2.8c}$$

Furthermore, if $s \le \gamma/9K$, with $\gamma \in (0, 1)$, then

$$t \le \sqrt{\gamma}\, s. \tag{2.8d}$$

Proof: Let $\lambda \triangleq 1/9K$. Then, from (2.8a, b),

$$t^2 \le K[2s^3 + 3s^2t + 3st^2 + t^3]$$

$$\le K[2\lambda s^2 + 3\lambda s^2 + 3\lambda t^2 + \lambda t^2]. \tag{2.9a}$$

Hence,

$$(1 - 4\lambda K)t^2 \le 5K\lambda s^2. \tag{2.9b}$$

Since $(1 - 4\lambda K) = 5K\lambda = 5/9$, it follows that $t \le s$. Hence, replacing t by s in (2.8a), we obtain (2.8c).

Now, if $s \le \gamma/9K$, then $\sqrt{s} \le \sqrt{\gamma}/3\sqrt{K}$. Substituting for $\sqrt{s}$ in (2.8c) we obtain (2.8d), which completes our proof. □

Corollary 2.1: Suppose that $K \in (0, \infty)$, $\gamma \in (0, 1)$, and that $\{\alpha_i\}_{i=0}^{\infty}$ is a real sequence such that

$$\alpha_{i+1}^2 \le K[(\alpha_i + \alpha_{i+1})^3 + \alpha_i^3] \tag{2.10a}$$

$$0 \le \alpha_i \le \frac{\gamma}{9K}, \quad \forall i \in \mathbb{N}. \tag{2.10b}$$

Then $\alpha_i \to 0$ as $i \to \infty$ superlinearly, with Q-rate 3/2.

Proof: It follows from Lemma 2.3 that $\alpha_{i+1} \le \sqrt{\gamma}\,\alpha_i$, for all $i \in \mathbb{N}$. Hence $\alpha_i \to 0$ as $i \to \infty$. The 3/2 Q-rate follows from (2.8c). □

We are finally ready to establish the convergence properties of Algorithm 2.1.

Theorem 2.1: There exists a $\rho > 0$ such that if $\| x_0 - x^* \| \le \rho$, where x^* is the solution of (2.1a), and $\{ x_i \}_{i=0}^{\infty}$ is a sequence constructed by Algorithm 2.1, then, $x_i \to x^*$, as $i \to \infty$, Q-superlinearly, with rate at least 3/2.

Proof: Let $\alpha = m/2$, then, making use of (2.3a) and (2.4b), we obtain, for $i = 0, 1, 2, \ldots$, that

$$\alpha \| x_{i+1} - x^* \|^2 \le \psi(x_{i+1}) - \psi(x^*)$$

$$\le \psi(x_{i+1}) - \hat{\psi}(x_{i+1} \mid x_i) + \hat{\psi}(x_{i+1} \mid x_i) - \psi(x^*)$$

$$\le \psi(x_{i+1}) - \hat{\psi}(x_{i+1} \mid x_i) + \hat{\psi}(x^* \mid x_i) - \psi(x^*), \tag{2.11a}$$

because $\hat{\psi}(x_{i+1} \mid x_i) \le \hat{\psi}(x^* \mid x_i)$, by construction of x_{i+1}. It now follows from (2.11a) and Lemma 2.1 that

$$\| x_{i+1} - x^* \|^2 \le K [\| x_{i+1} - x_i \|^3 + \| x_i - x^* \|^3]$$

$$\le K [\| (x_{i+1} - x^*) - (x_i - x^*) \|^3 + \| x_i - x^* \|^3]$$

$$\le K [(\| x_{i+1} - x^* \| + \| x_i - x^* \|)^3 + \| x_i - x^* \|^3], \tag{2.11b}$$

where $K = \hat{K}/\alpha$ and $\hat{K}$ is as in Lemma 2.1.

Next, since by Lemma 2.2, $h(\cdot)$ is continuous and $h(x^*) = 0$, it follows that given $\gamma^* \in (0, 1)$, there exists a $\bar{\rho} > 0$ such that if $\| x_i - x^* \| \le \bar{\rho}$, then $\| h(x_i) \| = \| x_{i+1} - x_i \| \le \gamma^*/18K$. Let $\rho^* = \min \{ \bar{\rho}, \gamma^*/18K \}$. Then, if $\| x_i - x^* \| \le \rho^*$, we must have that

$$\| x_{i+1} - x^* \| \le \| x_{i+1} - x_i \| + \| x_i - x^* \| \le \frac{\gamma^*}{18K} + \rho^* \le \frac{\gamma^*}{9K}. \tag{2.12}$$

Letting $t \triangleq \| x_{i+1} - x^* \|$ and $s \triangleq \| x_i - x^* \|$, and making use of Lemma 2.3 (see (2.8b)), we obtain that

$$\| x_{i+1} - x^* \| \le \sqrt{\gamma^*} \, \| x_i - x^* \|. \tag{2.13}$$

Hence, if $\| x_0 - x^* \| \le \rho^*$, then $\| x_i - x^* \| \le \rho^*$ for all $i = 1, 2, 3, \ldots$, and therefore, by (2.13), $\| x_i - x^* \| \to 0$ as $i \to \infty$. It now follows from (2.11b) and Corollary 2.1 (via (2.8c)) that

$$\| x_{i+1} - x^* \| \le 3\sqrt{K} \, \| x_i - x^* \|^{3/2}, \quad \forall i \in \mathbb{N}, \tag{2.14}$$

which completes our proof. □

3. THE GLOBAL NEWTON METHOD

We will now present an extension of the globally stabilized Newton method, proposed by Goldstein in [7] (see also [22, p. 33]). Stabilization is achieved by adding an Armijo type step size rule to the Local Newton Method. The rate of convergence of the Local Newton Method is preserved, because, as we will show, near the solution of (2.1a), under Assumption 2.1, the step size becomes unity, i.e., the Global Newton Method reverts to the Local Newton Method.

Algorithm 3.1 (Global Newton Method).

Data: $x_0 \in \mathbb{R}^n$, $\alpha, \beta \in (0,1)$, $S \triangleq \{1, \beta, \beta^2, \ldots\}$.

Step 0: Set i = 0.

Step 1: Compute $\theta(x_i)$, and $h_i = h(x_i)$, according to (2.7b), (2.7a).

Step 2: Compute the step size $\lambda_i \triangleq \max\{\lambda \in S \mid \psi(x_i + \lambda h_i) - \psi(x_i) \le \lambda\alpha\theta(x_i)\}$.

Step 3: Set $x_{i+1} = x_i + \lambda_i h_i$. Replace i by $i+1$ and go to Step 1. □

First we show that Algorithm 3.1 is globally convergent.

Theorem 3.1: Suppose that Assumption 2.1 holds and that x^* is the solution of (2.1a). Then any sequence $\{x_i\}_{i=0}^{\infty}$, constructed by Algorithm 3.1, converges to x^*.

Proof: First, because of Assumption 2.1 *(iii)*, the level sets of $\psi(\cdot)$ are bounded and, by construction in Step 2, $\psi(x_{i+1}) < \psi(x_i)$. Hence any sequence $\{x_i\}_{i=0}^{\infty}$, constructed by Algorithm 3.1, must have accumulation points. For the sake of contradiction, suppose that the sequence $\{x_i\}_{i=0}^{\infty}$ does not converge to x^*. Then it must have an accumulation point $x^{**} \ne x^*$. By Lemma 2.2, we then have that $\theta(x^{**}) < 0$ and $h(x^{**}) \ne 0$. Since by Lemma 2.2, the directional derivative $d\psi(x^{**}, h(x^{**})) \le \theta(x^{**}) < 0$, it follows that there is a $s^{**} \in S$ such that

$$\psi(x^{**} + s^{**}h(x^{**})) - \psi(x^{**}) < s^{**}\alpha\theta(x^{**}) . \tag{3.1a}$$

Hence, making use of the continuity of $\theta(\cdot)$ and $h(\cdot)$, for all x_i sufficiently near x^{**}, the stepsize $\lambda_i \ge s^{**}$ and $\theta(x_i) < \theta(x^{**})/2$. Therefore, for all such x_i,

$$\psi(x_i + \lambda_i h(x_i)) - \psi(x_i) \le \lambda_i \alpha\theta(x_i) \le s^{**}\alpha\theta(x^{**}) . \tag{3.1b}$$

Since the sequence $\{\psi(x_i)\}_{i=0}^{\infty}$ is monotone decreasing, (3.1b) implies that $\psi(x_i) \to -\infty$ as $i \to \infty$, which is a contradiction. Hence the theorem must be true. □

Next we establish superlinear convergence.

Theorem 3.2: Suppose that Assumption 2.1 holds and that x^* is the solution of problem (2.1a). Then any sequence $\{x_i\}_{i=0}^{\infty}$, constructed by Algorithm 3.1, converges to x^*, superlinearly, with Q-rate at least 3/2.

Proof: Since $\{x_i\}_{i=0}^{\infty}$ converges to x^* by Theorem 3.1, we only need to show that there exists an i_0 such that $\lambda_i = 1$ for all $i \ge i_0$, so that Algorithm 3.1 reduces to Algorithm 2.1 and invoke Theorem 2.1.

Now, it follows from (2.5) that

$$\theta(x_i) = \hat{\psi}(x_i + h(x_i) \mid x_i) - \psi(x_i + h(x_i)) + \psi(x_i + h(x_i)) - \psi(x_i)$$

$$\ge \psi(x_i + h(x_i)) - \psi(x_i) - \hat{K}\|h(x_i)\|^3 . \tag{3.2a}$$

Hence

$$\psi(x_i + h(x_i)) - \psi(x_i) \le \alpha\theta(x_i) + [(1-\alpha)\theta(x_i) + \hat{K}\|h(x_i)\|^3] . \tag{3.2b}$$

Next we establish a relationship between $\theta(x)$ and $\|h(x)\|$. Since $x + h(x)$ is the minimizer of $\hat{\psi}(\cdot \mid x)$, it follows that it satisfies the first order condition:

$$0 \in \partial\hat{\psi}(x + h(x) \mid x) . \tag{3.3a}$$

For any integer $p \ge 1$, let $\Sigma_p \triangleq \{\mu \in \mathbb{R}^p \mid \sum_{j=1}^{p} \mu^j = 1,\ \mu^j \ge 0,\ \forall j \in \mathbf{q}\}$. Then it follows from (3.3a), the definition of the generalized gradient $\partial\hat{\psi}(x + h(x) \mid x)$ (see [3]), and the Caratheodory Theorem [29], that there exists a multiplier $\mu \in \Sigma_q$, multipliers $\nu_j \in \Sigma_{n+1}$, and $t_j^k \in [0,1]$, with $j \in \mathbf{q}$ and $k \in \mathbf{n+1}$, such that

$$0 = \sum_{j=1}^{q} \mu^j \sum_{k=1}^{n+1} \nu_j^k [\nabla_x \phi^j(x, t_j^k) + \phi_{xx}^j(x, t_j^k) h(x)] , \tag{3.3b}$$

which implies that[3]

$$h(x) = -\left[\sum_{j=1}^{q} \mu^j \sum_{k=1}^{n+1} \nu_j^k \phi_{xx}^j(x, t_j^k)\right]^{-1} \sum_{j=1}^{q} \mu^j \sum_{k=1}^{n+1} \nu_j^k \nabla_x \phi^j(x, t_j^k) . \tag{3.3c}$$

Furthermore the following complementary slackness condition (see (5.12a, b) in [21]) is satisfied:

$$\theta(x) = \sum_{j=1}^{q} \mu^j \sum_{k=1}^{n+1} \nu_j^k \left\{ [\phi^j(x, t_j^k) - \psi(x)] + \langle \nabla_x \phi^j(x, t_j^k), h(x) \rangle + \frac{1}{2} \langle h(x), \phi_{xx}^j(x, t_j^k) h(x) \rangle \right\} \tag{3.3d}$$

Substituting for $h(x)$ from (3.3c) into (3.3d), we obtain, in view of Assumption 2.1 *(iii)*, that

[3] Since the $\mu^j \ge 0$ and the $\nu_j^k \ge 0$ in (3.3c), if follows from (2.2) that the matrix $\left[\sum_{j=1}^{q} \mu^j \sum_{k=1}^{n+1} \nu_j^k \phi_{xx}^j(x, t_j^k)\right]$ is invertible.

$$\theta(x) = \sum_{j=1}^{q} \mu^j \sum_{k=1}^{n+1} \nu_j^k [\phi^j(x, t_j^k) - \psi(x)] - \frac{1}{2} \langle h(x), \left[\sum_{j=1}^{q} \mu^j \sum_{k=1}^{n+1} \nu_j^k \phi_{xx}^j(x, t_j^k) \right]^{-1} h(x) \rangle$$

$$\leq - \frac{1}{2M} \| h(x) \|^2 , \tag{3.4}$$

with the last line following from the fact that $\phi^j(x, t_j^k) - \psi(x) \leq 0$ for all t_j^k.

Substituting for $\theta(x)$ from (3.4) into (3.2b), we obtain that

$$\psi(x_i + h(x_i)) - \psi(x_i) \leq \alpha\theta(x_i) - [(1-\alpha)/2M - \hat{K} \| h(x_i) \|] \| h(x_i) \|^2 . \tag{3.5a}$$

Since $h(x_i) \to 0$ as $i \to \infty$, it follows that there exists an i_0 such that for all $i \geq i_0$,

$$\psi(x_i + h(x_i)) - \psi(x_i) \leq \alpha\theta(x_i) , \tag{3.5b}$$

i.e., that $\lambda_i = 1$. This completes our proof. □

4. IMPLEMENTATIONS OF THE LOCAL ALGORITHM

Note that numerical evaluations of $\psi(x)$ and $\theta(x)$, and hence of $h(x)$, are only aproximate: for $\psi(x)$ because intervals must be discretized, and for $\theta(x)$ and $h(x)$, because they are defined by a convex optimization problem which can only be solved approximately. Hence both the local and the global Newton methods that we have presented (Algorithms 2.1 and 3.1, respectively) must be viewed as *conceptual*. This brings us to the question as to whether it is possible to construct *implementable* algorithms, using some form of discretization of the interval $[0, 1]$, appearing in (2.1b), as well as some truncation rule for the algorithm used in computing approximations to $\theta(x)$, which retain the basic properties of Algorithms 2.1 and 3.1.

We need to strengthen Assumption 2.1, by adding the following hypothesis:

Assumption 4.1: There exists a Lipschitz constant $L < \infty$[4], such that for all $x \in \mathbb{R}^n$,

$$| \phi^j(x, t) - \phi^j(x, t') | \leq L | t - t' | , \quad \forall t, t' \in [0, 1] , \quad \forall j \in \mathbf{q} . \tag{4.1a}$$

$$\| \nabla_x \phi^j(x, t) - \nabla_x \phi^j(x, t') \| \leq L | t - t' | , \quad \forall t, t' \in [0, 1] , \quad \forall j \in \mathbf{q} . \tag{4.1b}$$

$$\| \phi_{xx}^j(x, t) - \phi_{xx}^j(x, t') \| \leq L | t - t' | , \quad \forall t, t' \in [0, 1] , \quad \forall j \in \mathbf{q} . \tag{4.1c}$$

□

[4] At the expense of some complication, it is possible to carry out the following analysis using local, rather than global Lipschitz constants.

We begin with the following observations. For any integer $N > 0$, let[5]

$$T_N \triangleq \{ t \mid t = \frac{k}{N} ,\ k = 0 , 1 , 2 , \ldots , N \} , \tag{4.2a}$$

$$\psi_N(x) \triangleq \max_{j \in \mathbf{q}} \max_{t \in T_N} \phi^j(x , t) , \tag{4.2b}$$

$$\hat{\psi}_N(x \mid y) \triangleq \max_{j \in \mathbf{q}} \max_{t \in T_N} \phi^j(y , t) + \langle \nabla_x \phi^j(y , t) , x - y \rangle + ½ \langle (x - y) , \phi^j_{xx}(y)(x - y) \rangle , \tag{4.2c}$$

$$h_N(x) \triangleq \arg \min_{h \in \mathbb{R}^n} \hat{\psi}_N(x + h \mid x) , \tag{4.2d}$$

$$\theta_N(x) \triangleq \min_{h \in \mathbb{R}^n} \hat{\psi}_N(x + h \mid x) - \psi_N(x) . \tag{4.2e}$$

The relationships between the quantities associated with the original problem P in (2.1a) and the approximating problems

$$P_N : \quad \min_{x \in \mathbb{R}^n} \psi_N(x) , \tag{4.3}$$

are as follows:

Proposition 4.1: Suppose that Assumptions 2.1 and 4.1 hold. Let x^* denote the solution of (2.1), and, for any positive integer N, let $x_N{}^*$ denote the solution of the discretized problem P_N. Then

$$| \psi(x^*) - \psi_N(x_N{}^*) | \leq \frac{L}{2N} , \tag{4.4a}$$

$$\| x^* - x_N{}^* \|^2 \leq \frac{2L}{mN} , \tag{4.4b}$$

and, for every bounded set $B \subset \mathbb{R}^n$, there exists a $L' < \infty$ such that

$$| \theta(x) - \theta_N(x) | \leq \frac{L'}{2N} , \quad \forall x \in B , \tag{4.4c}$$

$$\| h(x) - h_N(x) \|^2 \leq \frac{L'}{mN} , \quad \forall x \in B . \tag{4.4d}$$

Proof: First, let $x \in \mathbb{R}^n$ be arbitrary. Then, because $T_N \subset [0 , 1]$,

[5] Note that there is nothing special about the discretization (4.2a). Any family of discrete sets $T_N \subset T$, where $T \triangleq [0 , 1]$ can be used provided that *(i)* $d(T_N , T) \to 0$ as $N \to \infty$, and *(ii)* for any sequence of integers $\{ N_i \}_{i=0}^{\infty}$, such that $N_{i+1} \geq 2N_i$, $\sum d(T_N i , T) < \infty$, where $d(\cdot , \cdot)$ denotes the Hausdorff distance.

$$\frac{-L}{2N} < 0 \le \psi(x) - \psi_N(x) \,. \tag{4.5a}$$

Next, let $\hat{t} \in [0\,,1]$ and $\hat{j} \in \mathbf{q}$ be such that $\psi(x) = \phi^{\hat{j}}(x\,,\hat{t})$. Then there exist points $t' \in T_N$, such that $|\,t' - \hat{t}\,| \le 1/2N$ and hence, making use of (4.1),

$$\psi_N(x) \ge \phi^{\hat{j}}(x\,,t') \ge \phi^{\hat{j}}(x\,,\hat{t}) - \frac{L}{2N} = \psi(x) - \frac{L}{2N} \,. \tag{4.5b}$$

Thus we have shown that

$$|\,\psi(x) - \psi_N(x)\,| \le \frac{L}{2N} \,. \tag{4.5c}$$

As a result, we have

$$\psi(x^*) \le \psi(x_N^*) \le \psi_N(x_N^*) + \frac{L}{2N} \,, \tag{4.5d}$$

and

$$\psi_N(x_N^*) \le \psi_N(x^*) \le \psi(x^*) \,, \tag{4.5e}$$

which gives us (4.4a).

Next, making use of (2.3a) and (4.5c), we obtain

$$\frac{m}{2}\|x_N^* - x^*\|^2 \le \psi(x_N^*) - \psi(x^*) \le \psi_N(x^*_N) + \frac{L}{2N} - \psi(x^*) \le \frac{L}{2N} < \frac{L}{N} \,, \tag{4.6a}$$

which establishes (4.4b).

Now suppose $B \subset \mathbb{R}^n$ is bounded, and let $x \in B$. Let the functions $\eta_x^j : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$, $j \in \mathbf{q}$, be defined by

$$\eta_x^j(h\,,t) \triangleq \phi^j(x\,,t) + \langle \nabla_x \phi^j(x\,,t)\,,h\,\rangle + \frac{1}{2}\,\langle h\,,\phi_{xx}^j(x\,,t)h\,\rangle \,. \tag{4.6b}$$

Let $G \triangleq \max_{x \in B} \max_{j \in \mathbf{q}} \max_{t \in [0,1]} \|\nabla_x \phi^j(x\,,t)\|$. Then it follows from the inequality

$$\max_{j \in \mathbf{q}} \max_{t \in [0,1]} \eta_x^j(h\,,t) \ge \max_{j \in \mathbf{q}} \max_{t \in [0,1]} \phi^j(x\,,t) - G\|h\| + \frac{m}{2}\|h\|^2$$

$$= \psi(x) - G\|h\| + \frac{m}{2}\|h\|^2 \,, \tag{4.6c}$$

that if $\|h\| > 2G/m$, then the left hand side of (4.6c) is greater than $\psi(x)$. Since $\theta(x) \le 0$, we must have $\|h(x)\| \le 2G/m$. A similar analysis shows that $\|h_N(x)\| \le 2G/m$ also. Now suppose that $\|h\| \le 2G/m$, and let $L_B \triangleq L(1 + 2G/m + 2G^2/m^2)$. Then

$$|\eta_x^j(h, t) - \eta_x^j(h, t')| \leq L_B |t - t'| \ , \ \forall t, t' \in [0, 1], \ \forall j \in \mathbf{q} . \tag{4.6d}$$

It now follows from (2.7b), (4.2e) and (4.5c), and an argument similar to that used to establish (4.4a), (4.4b) that (4.4c), (4.4d) hold, with $L' \triangleq L_B + L$. □

Comment 4.1: Note that it follows from duality theory [13] that because (4.2e) is a convex problem, the dual of (4.2e) is given by

$$\theta_N(x) = \max J_d(\mu) , \tag{4.7a}$$

where

$$J_{d,N}(\mu) \triangleq - \sum_{\substack{j \in \mathbf{q} \\ t \in T_N}} \mu^{j,t} [\psi_N(x) - \phi^j(x, t)] + \frac{1}{2} \| \sum_{\substack{j \in \mathbf{q} \\ t \in T_N}} \mu^{j,t} \nabla_x \phi^j(x, t) \|^2 \mid \sum_{\substack{j \in \mathbf{q} \\ t \in T_N}} \mu^{j,t} = 1 , \ \mu^{j,t} \geq 0 \} . \tag{4.7b}$$

Hence, given any set of admissible multipliers $\mu^{j,t}$, we have that $\theta_N(x) \geq J_{d,N}(\mu)$.

Now an algorithm such as the barrier function method in [26], applied to (4.2e), produces not only approximations $\hat{h}$ to $h_N(x)$, but also associated multipliers $\mu^{j,t}$, while an algorithm such as the Levitin-Polyak method [12], applied to (4.7a) produces multipliers $\mu^{j,t}$ which, via (3.3c) can be used to obtain an approximation $\hat{h}$ to $h_N(x)$. In either event, we have that

$$J_{p,N}(\hat{h}) \triangleq \hat{\psi}_N(x + \hat{h}) - \psi_N(x) \geq \theta_N(x) \geq J_{d,N}(\mu) . \tag{4.7c}$$

Therefore, given any $\varepsilon > 0$, to determine when such an algorithm has constructed an approximation $h_{N,\varepsilon}(x)$ such that

$$0 \leq \hat{\psi}_N(x + h_{N,\varepsilon}(x) \mid x)) - \psi_N(x) - \theta_N(x) \leq \varepsilon , \tag{4.7d}$$

we need only to check whether $J_{p,N}(h_{N,\varepsilon}(x) - J_{d,N}(\mu) \leq \varepsilon$. Hence we see that the construction of such $h_{N,\varepsilon}(x)$ is a finite process. Furthermore, it follows from Proposition 2.1, applied to the function $h \mapsto \hat{\psi}_N(x + h \mid x)$ and (4.7d) that

$$\varepsilon \geq \hat{\psi}_N(x + h_{N,\varepsilon}(x) \mid x) - \hat{\psi}_N(x + h_N(x) \mid x) \geq \frac{m}{2} \| h_{N,\varepsilon}(x) - h_N(x) \|^2 . \tag{4.7e}$$

□

We can now follow one of two alternatives. The first is to decide on an acceptable level of error and then to use (4.4a) or (4.4b) to determine the required level of discretization, i.e., the parameter N. In that case, one proposes to solve P_N and one only needs to invent a scheme for truncating the computation of $h_N(x)$. Such a scheme is incorporated in the following implementation of the local Newton Method for solving problems P_N. The second alternative involves increasing the discretization mesh progressively, rather than using a fixed discretization. This second alternative will be discussed subsequently.

Algorithm 4.1 (Implementable Finite Minimax Local Newton Method for P_N).

Data: $x_0 \in \mathbb{R}^n, \hat{\varepsilon} \in (0, 1)$.

Step 0: Set $i = 0$.

Step 1: Set $\varepsilon = \hat{\varepsilon}$.

Step 2: Compute a vector $h_{N,\varepsilon}(x_i) \in \mathbb{R}^n$ such that[6]

$$0 \le \hat{\psi}_N(x_i + h_{N,\varepsilon}(x_i) \mid x_i) - \psi_N(x_i) - \theta_N(x_i) \le \varepsilon . \tag{4.8a}$$

Step 3: If

$$\hat{\psi}_N(x_i + h_{N,\varepsilon}(x_i) \mid x_i) - \psi_N(x_i) \le -2\varepsilon , \tag{4.8b}$$

and

$$\varepsilon \le \| h_{N,\varepsilon}(x_i) \|^3 , \tag{4.8c}$$

set $x_{i+1} = x_i + h_{N,\varepsilon}(x_i)$, $(\varepsilon_i = \varepsilon)$[7] and go to step 4. Else replace ε by $\varepsilon/2$ and go to Step 2.

Step 4: Replace i by $i+1$ and go to Step 1. □

Comment 4.2: The structure of the tests (4.8a-c) is dictated by the proofs to follow, which establish the 3/2 rate of convergence of the algorithm. Note that (4.8b) ensures that

$$\varepsilon_i \le \frac{-\theta_N(x_i)}{2} . \tag{4.8d}$$

Hence, since $\theta_N(x_i) \to 0$ as $x_i \to x_N^*$, it follows that Algorithm 4.1 computes approximates $h_N(x_i)$ more and more accurately as the solution of P_N is approached. Also, if Algorithm 4.1 is initialized with $x_0 = x^*$, it cycles indefinitely up in the loop defined by Step 2 and Step 3, reducing ε to zero. □

Theorem 4.1: There exists a $\rho > 0$ such that if $\| x_0 - x_N^* \| \le \rho$, where x_N^* is the solution of (4.3), and $\{ x_i \}_{i=0}^{\infty}$ is a sequence constructed by Algorithm 4.1, then, $x_i \to x_N^*$, as $i \to \infty$, Q-superlinearly, with rate at least 3/2.

Proof: First we note that (2.5) holds with $\psi(\cdot)$, $\hat{\psi}(\cdot \mid \cdot)$ replaced by $\psi_N(\cdot)$, $\hat{\psi}_N(\cdot \mid \cdot)$, respectively, that we may assume that $K \ge 1$ in (2.5), and that Theorem 2.1 equally applies to the obvious simplification of the local Newton Method for problem P_N.

---

[6] Note that $\theta_N(x_i)$ is *not* evaluated. See the paragraph preceding (4.7d).

[7] Note that the computation of ε_i need not always begin with $\hat{\varepsilon}$. Rather, it is more efficient to start with ε_{i-1}.

Next, for any i, let $x'_{i+1} \triangleq x_i + h_N(x_i)$. Then, by (2.11b),

$$\|x'_{i+1} - x_N^*\|^2 \le K[\|x'_{i+1} - x_i\|^3 + \|x_i - x_N^*\|^3] . \tag{4.9a}$$

Since by (4.7d) and (4.8c), we have that

$$\|x_{i+1} - x'_{i+1}\|^2 = \|h_{N,\varepsilon_i}(x_i) - h_N(x_i)\|^2$$

$$\le \frac{2\varepsilon_i}{m} \le \frac{2}{m}\|h_{N,\varepsilon_i}(x_i)\|^3 = \frac{2}{m}\|x_{i+1} - x_i\|^3 , \tag{4.9b}$$

we obtain, using (4.9a) and the fact that $\|x + y\|^p \le 2^{p-1}[\|x\|^p + \|y\|^p]$, $p = 2, 3$, that, with $K \ge 1$,

$$\|x_{i+1} - x_N^*\|^2 \le 2[\|x'_{i+1} - x_N^*\|^2 + \|x_{i+1} - x'_{i+1}\|^2]$$

$$\le 2K[\|x'_{i+1} - x_i\|^3 + \|x_i - x_N^*\|^3 + \|x_{i+1} - x'_{i+1}\|^2]$$

$$\le 8K[\|x_{i+1} - x_i\|^3 + \|x_{i+1} - x'_{i+1}\|^3 + \|x_i - x_N^*\|^3 + \|x_{i+1} - x'_{i+1}\|^2] . \tag{4.9c}$$

Assuming that $x_i - x_N^*$ is sufficiently small, we must have, in view of the fact that by Lemma 2.1 $\theta_N(x_i) \to 0$ as $x_i \to x_N^*$ and (4.8d), that $2\varepsilon_i/m < 1$ and hence, by (4.9b), that $\|x_{i+1} - x'_{i+1}\| < 1$. Therefore, making use of (4.9b), (4.9c) leads to the conclusion that there exists a $K' \in [16K, \infty)$, depending on m, such that (4.9c) reduces to

$$\|x_{i+1} - x_N^*\|^2 \le 16K[\|x_{i+1} - x_i\|^3 + \|x_i - x_N^*\|^3 + \|x_{i+1} - x'_{i+1}\|^2]$$

$$\le K'[\|x_{i+1} - x_i\|^3 + \|x_i - x_N^*\|^3] . \tag{4.10a}$$

The proof can now be completed by using arguments similar to those following (2.11b) in the proof of Theorem 2.1. This requires that we show that given any $\delta > 0$, there exists a $\rho > 0$ such that if $\|x_i - x_N^*\| \le \rho$, then $\|x_{i+1} - x_i\| \le \delta$. Making use of the triangle inequality, (4.7d) and (4.8d), we obtain that

$$\|x_{i+1} - x_i\| = \|h_{N,\varepsilon_i}(x_i)\|$$

$$\le \|h_{N,\varepsilon_i}(x_i) - h_N(x_i)\| + \|h_N(x_i)\|$$

$$\le \sqrt{\frac{-2\theta_N(x_i)}{m}} + \|h_N(x_i)\| . \tag{4.10b}$$

The desired continuity result now follows from Lemma 2.2, and one can now proceed as in the proof of Theorem 2.1, following (2.11b), to complete this proof. □

There is evidence in the literature (see, e.g., [11, 9]) that one can reduce computing times considerably by increasing the discretization mesh size progressively, rather than using the finest mesh from the very start. This idea is incorporated in the following implementation of the Local Newton Method which adjust both the precision with which successive iterates are computed as well as the mesh size.

Algorithm 4.2 (Implementable Finite Minimax Local Newton Method for P).

Data: $x_0 \in \mathbb{R}^n, \hat{\varepsilon} \in (0, 1), K \ll 1, N_0 \in \mathbb{N}$.

Step 0: Set $i = 0$.

Step 1: Set $\varepsilon = \hat{\varepsilon}, N = N_i$[8].

Step 2: Compute a vector $h_{N,\varepsilon}(x_i) \in \mathbb{R}^n$ such that (see Comment 4.1)

$$0 \le \hat{\psi}_N(x_i + h_{N,\varepsilon}(x_i) \mid x_i) - \psi_N(x_i) - \theta_N(x_i) \le \varepsilon . \tag{4.11a}$$

Step 3: If

$$\hat{\psi}_N(x_i + h_{N,\varepsilon}(x_i) \mid x_i) - \psi_N(x_i) \le -2\varepsilon , \tag{4.11b}$$

and

$$\varepsilon \le \| h_{N,\varepsilon}(x_i) \|^3 , \tag{4.11c}$$

set $x'_{i+1} = x_i + h_{N,\varepsilon}(x_i)$ and go to step 4. Else replace ε by $\varepsilon/2$ and go to Step 2.

Step 4: If

$$\frac{K}{N} \le \| x'_{i+1} - x_i \|^3 , \tag{4.11d}$$

set $x_{i+1} = x'_{i+1}$, ($\varepsilon_i = \varepsilon, N_i = N$) and go to Step 5. Else replace N by $2N$ and go to Step 1.

Step 5: Replace i by $i+1$ and go to Step 1. □

Comment 4.3: The function of the coefficient K in (4.11d) is to limit the growth of the discretization parameter N. Thus, suppose that we are willing to accept a solution corresponding to N^* discretization points, and that our stopping criterion is $\| h_{N^*}(x_i) \| \le \omega$, with $\omega \ll 1$. Then we would set $K \le N^* \omega$. □

Note that Algorithm 4.2 solves problem $P_{N_{i-1}}$ until the test (4.11d) fails. In view of Theorem 4.1, this will happen after a finite number of iterations provided that $\| x_0 - x_{N_i}^* \| \le \rho$, where $\rho > 0$ is as in Theorem 4.1. Now, suppose that N_0 is such that $\sqrt{2L/mN_0} \le \rho/2$. Then, from (4.4b) we have that if $\| x_0 - x^* \| \le \rho/2$, with x^* the solution of P, then, $\| x_0 - x_{N_0}^* \| \le \rho$.

[8] Although it is reasonable to key ε_i, which controls the precision with which $\theta_{N_i}(x_i)$ is approximated, to the actual value of $\theta_{N_i}(x_i)$, so that ε_i may or may not decrease monotonically, it makes better sense to increase the discretization parameter N_i monotonically.

Theorem 4.2: There exists a $\rho > 0$ and a integer $N_0 < \infty$, such that if $\|x_0 - x^*\| \le \rho$, where x^* is the solution of (2.1a), and $\{x_i\}_{i=0}^{\infty}$ is a sequence constructed by Algorithm 4.2, then, $x_i \to x^*$, as $i \to \infty$, Q-superlinearly, with rate at least 3/2.

Proof: First, assuming that $\|x_i - x_{N_i}{}^*\| \le 1$, it follows from (4.10) that for some $K' < \infty$, independent of N_i,

$$\|x_{i+1} - x_{N_i}{}^*\|^2 \le K' [\|x_{i+1} - x_i\|^3 + \|x_i - x_{N_i}{}^*\|^3] . \tag{4.12a}$$

Hence, assuming, without loss of generality, that $K' \ge 1$ and that N_0 is sufficiently large to ensure that for all $N_i \ge N_0$, $\|x_{N_i}{}^* - x^*\| \le 1$, we get

$$\begin{aligned}\|x_{i+1} - x^*\|^2 &\le 2[\|x_{i+1} - x_{N_i}{}^*\|^2 + \|x_{N_i}{}^* - x^*\|^2] \\ &\le 2K' [\|x_{i+1} - x_i\|^3 + \|x_i - x_{N_i}{}^*\|^3 + \|x_{N_i}{}^* - x^*\|^2] \\ &\le 8K' [\|x_{i+1} - x_i\|^3 + \|x_i - x^*\|^3 + \|x_{N_i}{}^* - x^*\|^3 + \|x_{N_i}{}^* - x^*\|^2] \\ &\le 16K' [\|x_{i+1} - x_i\|^3 + \|x_i - x^*\|^3 + \|x_{N_i}{}^* - x^*\|^2] .\end{aligned} \tag{4.12b}$$

Now, it follows from (4.4b) and (4.11d) that

$$\|x_{N_i}{}^* - x^*\|^2 \le \frac{2L}{mN_i} \le \frac{2L}{m}\|x_{i+1} - x_i\|^3 . \tag{4.12c}$$

Substituting into (4.12b), we obtain that there exists a $K'' < \infty$, independent of N_i, such that

$$\|x_{i+1} - x^*\|^2 \le K'' [\|x_{i+1} - x_i\|^3 + \|x_i - x^*\|^3 . \tag{4.12d}$$

To continue, let $B \triangleq \{ x \in \mathbb{R}^n \mid \|x - x^*\| \le 1 \}$. It follows from Proposition 4.1 that there exists a $L' < \infty$ such that (4.4c), (4.4d) hold. As in the proof of Theorem 4.1, we will show that if N_0 is sufficiently large, then given any $\delta > 0$ there exists a $\rho > 0$ (where $\rho \le 1$ without loss of generality) such that for all $N_i \ge N_0$ and $\|x_i - x^*\| \le \rho$, $\|x_{i+1} - x_i\| \le \delta$. Using the triangle inequality, (4.7d) and (4.4d), we obtain

$$\begin{aligned}\|x_{i+1} - x_i\| &= \|h_{N_i, \varepsilon_i}(x_i)\| \\ &\le \|h_{N_i, \varepsilon_i}(x_i) - h_{N_i}(x_i)\| + \| h_{N_i}(x_i) - h(x_i)\| + \|h(x_i)\| \\ &\le \sqrt{\frac{2\varepsilon_i}{m}} + \sqrt{\frac{4L'}{mN_i}} + \|h(x_i)\| .\end{aligned} \tag{4.13a}$$

Furthermore, analogously to (4.8d), we have that

$$\|x_{i+1}-x_i\| \le \sqrt{\frac{|\theta_{N_i}(x_i)|}{m}}+\sqrt{\frac{4L'}{mN_i}}+\|h(x_i)\|\,. \tag{4.13b}$$

Applying the triangle inequality once more and utilizing (4.4c), we obtain

$$\|x_{i+1}-x_i\| \le \sqrt{\frac{[\,|\theta(x_i)|+|\theta(x_i)-\theta_{N_i}(x_i)|\,]}{m}}+\sqrt{\frac{4L'}{mN_i}}+\|h(x_i)\|$$

$$\le \sqrt{\frac{[\,|\theta(x_i)|+\frac{L'}{N_i}\,]}{m}}+\sqrt{\frac{4L'}{mN_i}}+\|h(x_i)\|\,. \tag{4.13c}$$

It now follows from the continuity of $\theta(\cdot)$ and $h(\cdot)$ and the fact that $\theta(x^*)=0$, $h(x^*)=0$, that if N_0 is chosen sufficiently large and ρ sufficiently small, then the desired continuity result holds. One can now proceed as in the proof of Theorem 2.1, following (2.11b), to complete the proof. □

5. IMPLEMENTATION OF THE GLOBAL ALGORITHM

To produce an implementation of Algorithm 3.1 (the Global Newton Method), we propose to use two mechanisms for controlling the precision of the approximations used. The first one will be taken from Algorithm 4.2, and will ensure superlinear rate of convergence, while the second one, which we will allow to dominate the first one, will be an extension of the mechanism described in Appendix A of [22]. For our case, this extension can be described abstractly as follows. Suppose that for every integer $N \ge N_0 > 0$, $A_N : \mathbb{R}^n \to 2^{\mathbb{R}^n \times \mathbb{N}}$ is a (possibly) set-valued iteration map. The reason for introducing a second integer N' is that given an integer N, the algorithm may have to increase it to a new value $N' \ge N$ before it can satisfy all the internal tests. Now consider the following algorithm model form solving the problem P in (2.1a).

Algorithm Model 5.1

Data: $x_0 \in \mathbb{R}^n$, $N_0 \in \mathbb{N}$.

Step 0: Set $i = 0$.

Step 1: Set $N = N_i$.

Step 2: Compute a pair

$$(y\,, N') \in A_N(x_i)\,. \tag{5.1a}$$

Step 3: If

$$\psi_{N'}(y)-\psi_{N'}(x_i) \le -\frac{1}{N'}\,, \tag{5.1b}$$

go to Step 4; else replace N by $2N'$ and go to Step 2.

Step 4: Set $x_{i+1} = y$, $N_{i+1} = N$.

Step 5: Replace i by $i+1$ and go to Step 1. □

Our proof of convergence requires the following technical result:

Lemma 5.1: Suppose that the sequences of real numbers $\{ \beta_i \}_{i=0}^{\infty}$ and $\{ \eta_i \}_{i=0}^{\infty}$ satisfy the following conditions: *(i)* $\eta_i \geq 0$ for all $i \in \mathbb{N}$, *(ii)* $\sum_{i=0}^{\infty} \eta_i < \infty$, and, *(iii)* $\beta_{i+1} \leq \beta_i + \eta_i$, for all $i \in \mathbb{N}$. Then either the sequence $\{ \beta_i \}_{i=0}^{\infty}$ converges, or $\beta_i \to -\infty$ as $i \to \infty$.

Proof: It is clear from the assumptions that the following holds:

$$\beta_n - \beta_0 = \sum_{i=0}^{n-1} (\beta_{i+1} - \beta_i) \leq \sum_{i=0}^{\infty} \eta_i \,. \tag{5.2a}$$

Hence, β_i is bounded from above, and therefore $\hat{\beta} \triangleq \overline{\lim}_{i \to \infty} \beta_i < \infty$ Obviously, if $\hat{\beta} = -\infty$, then $\beta_i \to -\infty$ as $i \to \infty$.

Now suppose that $\hat{\beta} > -\infty$. To prove convergence of the sequence $\{ \beta_i \}_{i=0}^{\infty}$, we will show by contradiction that $\underline{\lim}_{i \to \infty} \beta_i \geq \hat{\beta}$. Thus, let $\varepsilon > 0$ be arbitrary, and suppose that there is no i_0 such that $\beta_i > \hat{\beta} - \varepsilon$ for all $i > i_0$. Clearly, there exists an i_1 such that $\sum_{k=i}^{\infty} \eta_k < \varepsilon / 2$ for all $i \geq i_1$. It follows from our hypothesis that there exists an $i_2 \geq i_1$, such that $\beta_{i_2} \leq \hat{\beta} - \varepsilon$. It follows from (5.2a) that for $i > i_2$,

$$\beta_i - \beta_{i_2} = \sum_{k=i_2}^{i-1} (\beta_{k+1} - \beta_k) \leq \sum_{k=i_2}^{\infty} \eta_k \leq \frac{\varepsilon}{2} \,. \tag{5.2b}$$

Hence $\beta_i \leq \hat{\beta} - \frac{\varepsilon}{2}$ for all i sufficiently large, which contradicts the definition of $\hat{\beta}$. It follows that $\lim_{i \to \infty} \beta_i = \hat{\beta}$. □

Theorem 5.1: Suppose that Assumptions 2.1 and 4.1 hold, so that (4.4a) is valid, and that for every $x \in \mathbb{R}^n$ such that $0 \notin \partial\psi(x)$ there exist a $\rho_x > 0$, a $\delta_x > 0$ and an integer $N_x > 0$ such that

$$\psi_{N'}(y') - \psi_{N'}(x') \leq -\delta_x \tag{5.3}$$

for all $N \geq N_x$, and all x', $y' \in \mathbb{R}^n$ such that $\|x' - x\| \leq \rho_x$, $(y', N') \in A_N(x')$.

Under these assumptions, if $\{ x_i \}_{i=0}^{\infty}$ $\{ N_i \}_{i=0}^{\infty}$ are a pair of sequences constructed by Algorithm Model 5.1, then $x_i \to x^*$ and $N_i \to \infty$, as $i \to \infty$, where x^* is the solution of (2.1a).

Proof: First we use Lemma 5.1 to show that the sequence $\{ \psi_{N_i}(x_i) \}_{i=0}^{\infty}$ converges. Let $I \triangleq \{ i \in \mathbb{N} \mid N_{i+1} \neq N_i \}$, and let the sequence $\{ \eta_i \}_{i=0}^{\infty}$ be defined by

$$\eta_i \triangleq \begin{cases} L/N_i & i \in I \\ 0 & \text{otherwise} \end{cases}, \tag{5.4a}$$

where L is the Lipschitz constant in Assumption 2.1*(i)*. Now suppose that $i \in I$, and let $i_+ \triangleq \min\{ j \in I \mid j > i \}$. Since by construction, we have that $N_{i_+} \geq 2N_i$, it follows that

$$\sum_{i=0}^{\infty} \eta_i = \sum_{i \in I} \eta_i \leq \sum_{k=0}^{\infty} \frac{L}{2^k N_0} < \infty . \tag{5.4b}$$

Hence the sequence $\{ \eta_i \}_{i=0}^{\infty}$ is summable. From (4.5c), we have that for all $i \in I$,

$$| \psi_{N_{i+1}}(x_i) - \psi_{N_i}(x_i) | \leq | \psi_{N_{i+1}}(x_i) - \psi(x_i) | + | \psi(x_i) - \psi_{N_i}(x_i) | \leq \frac{L}{2N_{i+1}} + \frac{L}{2N_i} \leq \frac{L}{N_i} . \tag{5.4c}$$

Clearly, for all $i \in \mathbb{N} \setminus I$, i.e. such that $N_{i+1} = N_i$, $| \psi_{N_{i+1}}(x_i) - \psi_{N_i}(x_i) | = 0$. Hence for all $i \in \mathbb{N}$,

$$| \psi_{N_{i+1}}(x_i) - \psi_{N_i}(x_i) | \leq | \psi_{N_{i+1}}(x_i) - \psi(x_i) | + | \psi(x_i) - \psi_{N_i}(x_i) | \leq \frac{L}{2N_{i+1}} + \frac{L}{2N_i} \leq \frac{L}{N_i} ,$$

Consequently, using (5.1b) we obtain

$$\psi_{N_{i+1}}(x_{i+1}) - \psi_{N_i}(x_i) \leq \psi_{N_{i+1}}(x_{i+1}) - \psi_{N_{i+1}}(x_i) + \psi_{N_{i+1}}(x_i) - \psi_{N_i}(x_i)$$

$$\leq -\frac{1}{N_{i+1}} + \eta_i \leq \eta_i . \tag{5.4d}$$

Furthermore, it follows from Proposition 2.1, (4.5c) and the fact that $N_{i+1} \geq N_i$ that

$$\psi_{N_i}(x) \geq \psi(x^*) + \frac{m}{2} \| x - x^* \|^2 - \frac{L}{2N_0} . \tag{5.4e}$$

Hence the sequence $\{ \psi_{N_i}(x_i) \}_{i=0}^{\infty}$ satisfies the hypotheses of Lemma 5.1, and in addition, it is bounded below. We therefore conclude that it converges.

Next, we show that $N_i \to \infty$. If this is not true, then since $N_{i+1} \geq N_i$ we must have $N_i = N^*$ for i sufficiently large. For such i, (5.1b) implies that

$$\psi_{N^*}(x_{i+1}) - \psi_{N^*}(x_i) \leq -\frac{1}{N^*} , \tag{5.4f}$$

which contradicts the fact that the sequence $\{ \psi_{N_i}(x_i) \}_{i=0}^{\infty}$ converges.

As a consequence of (5.4e), the sequence $\{ x_i \}_{i=0}^{\infty}$ is bounded, and hence it must have accumulation points. For the sake of contradiction, suppose that the sequence $\{ x_i \}_{i=0}^{\infty}$ does not converge to x^*. Then it must have an accumulation point $x^{**} \neq x^*$. Let $K \subset \mathbb{N}$ be the set of indices of the subsequence converging to x^{**}.

Since $x^{**} \neq x^*$, we have $0 \notin \partial\psi(x^{**})$, from which it follows, by assumption, that there exists a $\delta > 0$ such that for $i \in K$ sufficiently large

$$\psi_{N_{i+1}}(x_{i+1}) - \psi_{N_{i+1}}(x_i) \leq -\delta \,. \tag{5.4g}$$

Referring to (5.4d), we see that for $i \in K$ sufficiently large,

$$\psi_{N_{i+1}}(x_{i+1}) - \psi_{N_i}(x_i) \leq \psi_{N_{i+1}}(x_{i+1}) - \psi_{N_{i+1}}(x_i) + \psi_{N_{i+1}}(x_i) - \psi_{N_i}(x_i)$$

$$\leq -\delta + \eta_i \,. \tag{5.4h}$$

However, since $\eta_i \to 0$, (5.4h) contradicts the fact that the sequence $\{ \psi_{N_i}(x_i) \}_{i=0}^{\infty}$ converges. Hence $0 \in \partial\psi(x^*)$. □

The above Algorithm Model and our desire to retain the superlinear rate of convergence of the implementation of the local Newton method, Algorithm 4.2, leads us to the following algorithm.

Algorithm 5.2 (Implementable Global Newton Method for P).

Data: $x_0 \in \mathbb{R}^n, N_0 \in \mathbb{N}, \alpha, \beta \in (0, 1), \varepsilon_0 > 0, K << 1, S \triangleq \{1, \beta, \beta^2, \dots\}$.

Step 0: Set $i = 0$.

Step 1: Set $N = N_i$.

Step 2: Set $\varepsilon = \varepsilon_0$.

Step 3: Compute a vector $h_{N,\varepsilon}(x_i) \in \mathbb{R}^n$ such that (see Comment 4.1)

$$0 \leq \hat{\psi}_N(x_i + h_{N,\varepsilon}(x_i) \mid x_i) - \psi_N(x_i) - \theta_N(x_i) \leq \varepsilon \,. \tag{5.3a}$$

Step 4: If

$$\hat{\psi}_N(x_i + h_{N,\varepsilon}(x_i)) - \psi_N(x_i) \leq -2\varepsilon \,, \tag{5.3b}$$

and

$$\varepsilon \leq \| h_{N,\varepsilon}(x_i) \|^3 \,, \tag{5.3c}$$

go to step 5. Else replace ε by $\varepsilon/2$ and go to Step 3.

Step 5: If

$$\frac{K}{N} \leq \| h_{N,\varepsilon}(x_i) \|^3 \,, \tag{5.3d}$$

set $h_i = h_{N,\varepsilon}(x_i)$, and go to Step 6. Else replace N by $2N$ and go to Step 2.

Step 6: Compute the step size

$$\lambda_i \triangleq \max \{ \lambda \in S \ / \ \psi_N(x_i + \lambda h_i) - \psi_N(x_i) \le \lambda\alpha[\hat{\psi}_N(x_i + h_i \mid x_i) - \psi_N(x_i)] \} . \tag{5.3e}$$

Step 7: If

$$\psi_N(x_i + \lambda h_i) - \psi_N(x_i) \le -\frac{1}{N} , \tag{5.3f}$$

set $x_{i+1} = x_i + \lambda_i h_i$, $N_{i+1} = N$ and go to Step 8; else replace N by $2N$ and go to step 2.

Step 8: Replace i by $i+1$ and go to Step 1. □

Theorem 5.1 can now be used to show that sequences constructed by Algorithm 5.2 converge to the solution x^*, while Theorem 4.2 leads to the conclusion that these sequences converge superlinearly.

Theorem 5.2: Suppose that Assumptions 2.1 and 4.1 hold and that x^* is the solution of problem (2.1a). Then any sequence $\{ x_i \}_{i=0}^{\infty}$, constructed by Algorithm 5.2, converges to x^*, superlinearly, with Q-rate at least 3/2.

Proof: The proof consists of two parts. The first part shows that the Algorithm map $A_N(\cdot)$ satisfies the hypotheses of Theorem 5.1, from which we conclude that $x_i \to x^*$ as $i \to \infty$. The second part shows that for i sufficiently large, the step size λ_i is one. In this case Algorithm 5.2 reduces to the local algorithm, Algorithm 4.2, and we may apply Theorem 4.2 (along with the fact that $N_i \to \infty$) to conclude that the interates converge superlinearly.

(a) To show that $A_N(\cdot)$ satisfies the hypotheses of Theorem 5.1, suppose that $x \in \mathbb{R}^n$ is such that $0 \notin \partial\psi(x)$. Then Lemma 2.2 *(d)* implies that $\theta(x) < 0$. By continuity, there exist $\rho > 0$, $\delta > 0$ such that for all $x' \in B(x, \rho) \triangleq \{ x' \in \mathbb{R}^n \mid \| x' - x \| \le \rho \}$, $\theta(x') \le -\delta < 0$. Furthermore, since $B(x, \rho)$ is bounded, (4.4c) implies that there exists a $\bar{N} \in \mathbb{N}$ such that for all $N \ge \bar{N}$ and $x' \in B(x, \rho)$ $\theta_N(x') \le -\delta/2$.

Suppose that $x' \in B(x, \rho)$ and that the algorithm map $A_N(\cdot)$ produces a pair $(y', N') \in A_N(x')$ (with $N' \ge N$), and an ε satisfying the tests in Steps 3 and 4. Then from (5.3b) we have

$$\theta_{N'}(x') \le \hat{\psi}_{N'}(x' + h_{N', \varepsilon}(x') \mid x') - \psi_{N'}(x') \le -2\varepsilon , \tag{5.4}$$

which yields $-\theta_{N'}(x')/2 \ge \varepsilon$. Furthermore, using (5.3a) we obtain that

$$\begin{aligned} \hat{\psi}_{N'}(x' + h_{N', \varepsilon}(x') \mid x') - \psi_{N'}(x') &\le \varepsilon + \theta_{N'}(x') \\ &\le \frac{\theta_{N'}(x')}{2} . \end{aligned} \tag{5.5}$$

It follows from the convexity of the function $\lambda \mapsto \hat{\psi}_{N'}(x' + \lambda h_{N', \varepsilon}(x') \mid x') - \psi_{N'}(x')$ that for all $\lambda \in [0, 1]$,

$$\hat{\psi}_{N'}(x' + \lambda h_{N',\varepsilon}(x') \mid x') - \psi_{N'}(x') \le \lambda[\hat{\psi}_N(x' + h_{N,\varepsilon}(x') \mid x') - \psi_N(x')]\,. \tag{5.6}$$

By Lemma 2.1, applied to the functions $\psi_{N'}$, $\hat{\psi}_{N'}$, we have the estimate

$$|\,\hat{\psi}_{N'}(x' + \lambda h_{N',\varepsilon}(x') \mid x') - \psi_{N'}(x' + \lambda h_{N',\varepsilon}(x'))\,| \le \hat{K}\lambda^3 \| h_{N',\varepsilon}(x')\|^3\,. \tag{5.7}$$

Combining (5.6) and (5.7) yields

$$\psi_{N'}(x' + \lambda h_{N',\varepsilon}(x')) - \psi_{N'}(x') \le \hat{K}\lambda^3 \| h_{N',\varepsilon}(x')\|^3 + \lambda[\hat{\psi}_{N'}(x' + h_{N',\varepsilon}(x') \mid x') - \psi_{N'}(x')]\,. \tag{5.8}$$

Using (5.5) and (5.8), we obtain that for all $\lambda \in [0\,,1]$,

$$\begin{aligned}
&\psi_{N'}(x' + \lambda h_{N',\varepsilon}(x')) - \psi_{N'}(x') - \alpha\lambda[\hat{\psi}_{N'}(x' + h_{N',\varepsilon}(x') \mid x') - \psi_{N'}(x')] \\
&\qquad \le \hat{K}\lambda^3 \| h_{N',\varepsilon}(x')\|^3 + (1-\alpha)\lambda[\hat{\psi}_{N'}(x' + h_{N',\varepsilon}(x') \mid x') - \psi_{N'}(x')] \\
&\qquad = \lambda\Big[(1-\alpha)[\hat{\psi}_{N'}(x' + h_{N',\varepsilon}(x') \mid x') - \psi_{N'}(x')] + \hat{K}\lambda^2 \| h_{N',\varepsilon}(x')\|^3\Big] \\
&\qquad \le \lambda\Big[(1-\alpha)\frac{\theta_{N'}(x')}{2} + \hat{K}\lambda^2 \| h_{N',\varepsilon}(x')\|^3\Big] \\
&\qquad \le \lambda\Big[-(1-\alpha)\frac{\delta}{4} + \hat{K}\lambda^2 \| h_{N',\varepsilon}(x')\|^3\Big]\,.
\end{aligned} \tag{5.9}$$

Combining the fact that $h(\cdot)$ is continuous and $B(x\,,\rho)$ is bounded, with (4.4d) and (4.7d), we conclude that there exists a constant $\Delta < \infty$ such that for all $x' \in B(x\,,\rho)$ and all $N' \ge N$, $\| h_{N',\varepsilon}(x')\| \le \Delta$. Thus (5.9) yields

$$\psi_N(x' + \lambda h_{N,\varepsilon}(x')) - \psi_N(x') - \alpha\lambda[\hat{\psi}_N(x' + h_{N,\varepsilon}(x') \mid x') - \psi_N(x')] \le \lambda\Big[-(1-\alpha)\frac{\delta}{4} + \hat{K}\lambda^2\Delta^3\Big]\,, \tag{5.10}$$

from which we conclude that there exists a $0 < \lambda_0 < 1$ such that for all $x' \in B(x\,,\rho)$, the step size λ' produced by Step 6 satisfies $\lambda' \ge \lambda_0$. Consequently, using (5.3e), (5.5) and the fact that $\theta_N(x') \le -\delta/2$, we obtain that for all $x' \in B(x\,,\rho)$ and all $N' \ge N \ge \bar{N}$,

$$\begin{aligned}
\psi_{N'}(x' + \lambda' h_{N',\varepsilon}(x')) - \psi_{N'}(x') &\le \alpha\lambda_0[\hat{\psi}_{N'}(x' + h_{N',\varepsilon}(x') \mid x') - \psi_{N'}(x')] \\
&\le -\frac{\alpha\lambda_0\delta}{4}\,.
\end{aligned} \tag{5.11}$$

If we let $N_x = \bar{N}$, $\rho_x = \rho$ and $\delta_x = \alpha\lambda_0\delta/4$, we see that the map $A_N(\cdot)$ satisfies the conditions of Theorem 5.1. Hence $x_i \to x^*$, and as a consequence of Step 7, we see that $N_i \to \infty$.

(b) To complete the proof, we must show that for i sufficiently large the step size λ_i is one. If we set $\lambda = 1$ in the second last line of (5.9) we obtain

$$\psi_{N_{i+1}}(x_i+h_i) - \psi_{N_{i+1}}(x_i) - \alpha[\hat{\psi}_{N_{i+1}}(x_i+h_i \mid x_i) - \psi_{N_{i+1}}(x_i)] \le \left[(1-\alpha)\frac{\theta_{N_{i+1}}(x_i)}{2} + \hat{K}\|h_i\|^3 \right] , \tag{5.12}$$

where h_i is as defined in Step 5 of Algorithm 5.2. Using a result similar to (3.4) we obtain that $\theta_{N_{i+1}}(x_i) \le -(1/2M)\|h_{N_{i+1}}(x_i)\|^2$. Using (4.7d), (5.3c) and the fact that $\|x\|^2 \ge \frac{1}{2}\|y\|^2 - \|x-y\|^2$, we obtain that

$$\begin{aligned} \|h_{N_{i+1}}(x_i)\|^2 &\ge \frac{1}{2}\|h_i\|^2 - \|h_i - h_{N_{i+1}}(x_i)\|^2 \\ &\ge \frac{1}{2}\|h_i\|^2 - \frac{2\varepsilon_i}{m} \\ &\ge \frac{1}{2}\|h_i\|^2 - \frac{2}{m}\|h_i\|^3 . \end{aligned} \tag{5.13}$$

Hence we obtain the bound

$$\theta_{N_{i+1}}(x_i) \le \frac{-1}{4M}\|h_i\|^2 + \frac{1}{mM}\|h_i\|^3 . \tag{5.14}$$

Substituting this bound into (5.12) we obtain

$$\begin{aligned} &\psi_{N_{i+1}}(x_i+h_i) - \psi_{N_{i+1}}(x_i) - \alpha[\hat{\psi}_{N_{i+1}}(x_i+h_i \mid x_i) - \psi_{N_{i+1}}(x_i)] \\ &\qquad \le \frac{(1-\alpha)}{2}\left[\frac{-1}{4M}\|h_i\|^2 + \frac{1}{mM}\|h_i\|^3 \right] + \hat{K}\|h_i\|^3 . \end{aligned} \tag{5.15}$$

From (4.13b) we note that $h_i \to 0$ as $i \to \infty$, and hence the right hand side of (5.15) is negative for i sufficiently large. Hence $\lambda_i = 1$ for i sufficiently large. This completes the proof. □

6. A NUMERICAL EXAMPLE

We will present the solution of a semi-infinite minimax problem which was constructed by converting an optimal control problem with control and state space constraints, by means of an exact penalty function, into an unconstrained minimax problem.

The original optimal control problem is as follows:

$$\min_{x \in \mathbb{R}^{21}} \{ \tfrac{1}{2}\left[\|z(x,1)\|^2 + 10^{-6}\|x\|^2 \right] \mid z^2(x,t) - 0.15 \le 0,\ \forall t \in [0,20],\ x_j^2 - 1 \le 0,\ \forall j \in \mathbf{p} \} , \tag{6.1}$$

where $\mathbf{p} \triangleq \{0, \ldots, 20\}$, and $z : \mathbb{R}^{21} \times [0,20] \to \mathbb{R}^2$ solves the differential equation

$$\frac{d}{dt} z(x\,,t) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} z(z\,,t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(x\,,t)\,, \tag{6.2}$$

where $x(z\,,0) = (-2.5\,,0)^T$, and the control $u : \mathbb{R}^{21}\times[0\,,20] \to \mathbb{R}$ is defined by the 21 dimensional parameter vector x, through linear interpolation, as follows: for any $j = 0\,,1\,,\ldots,19$, and $t = \lambda j + (1-\lambda)(j+1)$, $u(x\,,j) = \lambda x_j + (1-\lambda)x_{j+1}$. Note that the dynamics in (6.1) are so simple that we can integrate them exactly, for the resulting piecewise linear control.

To convert the above optimal control problem into unconstrained form, we use a parameter of 100 in the exact penalty function, and obtain a problem of the form (2.1a), (2.1b), with $q = 23$, and the functions $\phi^j(\cdot\,,\cdot)$, $j \in \mathbf{q}$ defined by

$$\phi^1(x\,,t) \triangleq ½\Big[\,\|z(x\,,1)\|^2 + 10^{-6}\|x\|^2\Big]\,, \tag{6.3a}$$

$$\phi^2(x\,,t) \triangleq \phi^1(x\,,t) + 100(x_2(x\,,t/20) - 0.15)\,, \tag{6.3b}$$

$$\phi^j(x\,,t) \triangleq \phi^1(x\,,t) + 100(x_{j-2}^2 - 1)\,,\ \forall j \in \{3\,,\ldots,23\}\,, \tag{6.3c}$$

with $t \in [0\,,1]$. The algorithm was started from the initial point $x_0 = (1\,,-1\,,1\,,-1\,,1\,,-1\,,1\,,-1\,,1\,,-1\,,1\,,-1\,,1\,,-1\,,1\,,-1\,,1\,,-1\,,1\,,-1\,,1)$.

Figures 1 and 2 present our computational results. Figure 1 presents plots of the control sequence x^i versus i, while Figure 2 presents phase-plane plots, i.e., plots of $x^1(t)$ versus $x^2(t)$, traditional in the control literature. We see from these plots that the problem (6.1) is solved in 2 iterations, at the end of which the original semi-infinite constraint, $z^2(u\,,t) - 0.15 \le 0$ for all $t \in [0\,,20]$ has been satisfied. The initial value of the cost was $\psi_5(x_0) = 3.12501$, the final value of the cost was $\psi_{28}(2) = 2.09003\times10^{-7}$; the final value of the optimality function was $\theta_{28}(x_2) = -2.01827\times10^{-7}$. To limit the growth of the discretization index, we set $K = 10^{-20}$ in (5.3d) so as to be able to satisfy (5.3d) with $N = 100$ when $\|h_{N\,,\varepsilon}\| \ge 10^{-10}$. We set $N_0 = 5$, and the algorithm set N_1, $N_2 = 28$.

To reduce the looping in the computation of the discretization parameter N, instead of simply doubling it whenever the test (5.3d) failed to be satisfied, we set the next value to the integer part of $\max\{\,6.5\times10^{-19}/\|h_{N\,,\varepsilon}\|^3\,,\,1.1N\,\}$.

We computed the search direction, $h_{N\,,\varepsilon}(x)$, satisfying (5.3a), by applying to the dual of problem (4.2e) the rapidly converging Levitin-Polyak constrained Newton method [12].

To illustrate how the inequality (5.3a) is solved, consider the primal problem (4.2e). When expanded, it assumes the form

$$\theta \triangleq \min_{h \in \mathbb{R}^n} \max_{j \in \mathbf{q}} f^j + \langle g_j\,,h\rangle + ½\langle h\,,H^j h\rangle\,, \tag{6.4a}$$

The dual of this problem is

$$\theta = -\min_{\mu \in \Sigma} J_d(\mu) \ , \tag{6.4b}$$

where $J_d(\mu) \triangleq -\sum_{j=1}^{q} \mu^j f^j + ½ \langle \sum_{j=1}^{q} \mu^j g_j \, , (\sum_{j=1}^{q} \mu^j H^j)^{-1} \sum_{j=1}^{q} \mu^j g_j \rangle$, and Σ is the unit simplex in $\mathbb{R}^n$. The formula for the second derivative matrix of the dual cost function, required by the Levitin-Polyak method, is given in an appendix in [25].

When applied to the dual problem, the Levitin-Polyak method computes a sequence $\{ \mu_i \}_{i=0}^{\infty}$ which converges quadratically to $\hat{h}$, a solution of the dual problem (6.4b). Furthermore, if we define $h : \Sigma \to \mathbb{R}^n$ by

$$h(\mu) \triangleq (\sum_{j=1}^{q} \mu^j H^j)^{-1} \sum_{j=1}^{q} \mu^j g_j \ , \tag{6.5}$$

it is straightforward to show that the corresponding sequence $\{ h(\mu_i) \}_{i=0}^{\infty}$ converges to $\hat{h}$, the unique solution of the primal problem (6.4a). Noting that the iterations of the Levitin-Polyak algorithm generate both upper and lower bounds on θ, as given by

$$J_p(\mu_i) \geq \theta \geq -J_d(\mu_i) \, , \tag{6.6a}$$

where

$$J_p(\mu_i) \triangleq \max_{j \in \mathbf{q}} f^j + \langle g_j \, , h(\mu_i) \rangle + ½ \langle h(\mu_i) \, , H^j h(\mu_i) \rangle \, , \tag{6.6b}$$

and making use of the fact that $J_p(\mu_i) - J_d(\mu_i) \to 0$, we see that a point, $h(\mu_i)$ satisfying (5.3a) can be computed in a finite number of iterations of the Levitin-Polyak method.

7. CONCLUSION

We have used a new and very simple proof technique to show that natural *conceptual* extensions of Newton's method converge superlinearly on a class of semi-infinite minimax problems. This technique has also enabled us to construct rate preserving *implementations* of these extensions. Our implementations are interesting for two reasons: first, they account for all the significant approximations involved, and second, they do not require the knowledge of the Lipschitz constants or eigenvalue bounds associated with the problem functions and their first and second order derivatives.

Apart from the intrinsic interest that a theoretical extension of Newton's method to semi-infinite optimization possesses, our numerical results show that it is a viable procedure for the solution of such classical problems as state and control constrained optimal control problems with linear dynamics.

7. ACKNOWLEDGEMENT

We wish to thank the referees for their most helpful comments. **9. REFERENCES**

[1] L. ARMIJO, *Minimization of functions having Lipschitz continuous first partial derivatives*, Pacific Journal of Mathematics Vo. 16, pp 1-3, 1966.

[2] C. BERGE, *Topological Spaces*, Macmillan, New York, (1963).

[3] F. H. CLARKE, *Optimization and Nonsmooth Analysis*, Wiley-Interscience, New York, (1983).

[4] I. D. COOPE and G. A. WATSON, *A projected Lagrangian algorithm for semi-infinite programming*, Mathematical Programming, Vol. 32, pp. 337-356 (1985).

[5] J. E. DENNIS, Jr., R. B. SCHNABEL, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations* Prentice-Hall, Englewood Cliffs, N.J., 1983.

[6] J. DIEUDONNE, *Foundations of Modern Analysis*, Academic Press, New York, 1960.

[7] A. A. GOLDSTEIN, *Constructive Real Analysis*, Harper and Row, New York, 1967.

[8] S. P. HAN, *Variable metric methods for minimizing a class of nondifferentiable functions*, Mathematical Programming, Vol. 20, No. 1, pp. 1-13, 1981.

[9] L. HE and E. POLAK, *An Optimal Diagonalization Strategy for the Solution of a Class of Optimal design Problems*, University of California, Berkeley, Electronics Research Laboratory Memo No. UCB/ERL M88/41, 6 June 1988. IEEE, Trans. on Automatic Contr., in press.

[10] R. HETTICH and W. Van HOSTEDE, *On quadratically convergent methods for semi-infinite programming*, in: R. Hettich, ed., *Semi-infinite Programming, Lecture Notes in Control and Information Science 15*, Springer-Verlag, New York, pp. 97-111, (1979).

[11] R. KLESSIG and E. POLAK, *An adaptive algorithm for unconstrained optimization with applications to optimal control*, SIAM J. Control, Vol. 11, No. 1, pp. 80-94, 1973.

[12] E. S. LEVITIN and B. T. POLLYAK, "Constrained Minimization Methods", *U.S.S.R. Computational Mathematics and Mathematical Physics*, Vol. 6, No. 5, pp. 1-50, 1966.

[13] O. L. MANGASARIAN, *Nonlinear Programming*, MacGraw-Hill, New York, 1969.

[14] D. Q. MAYNE, *On the use of exact penalty functions to determine the step length of optimization algorithms*, Lecture Notes in Mathematics, 773, Numerical Analysis, pp. 98-109, Springer Verlag, 1980.

[15] D. Q. MAYNE and E. POLAK, *A quadratically convergent algorithm for solving infinite dimensional inequalities*, Applied Mathematics and Optimization, Vol. 9, pp. 25-40, 1982.

[16] D. Q. MAYNE and E. POLAK, *A superlinearly convergent algorithm for constrained optimization problems*, Mathematical Programming Study Vol. 16 pp. 45-61 (1982).

[17] H. MUKAI and E. POLAK, *A Second Order Algorithm for Unconstrained Optimization*, JOTA, Vol. 26, No. 4, 1978.

[18] J. M. ORTEGA and W. C. RHEINBOLDT, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.

[19] M. L. OVERTON, *On minimizing the maximum eigenvalue of a symmetric matrix*, in *Linear Algebra in Signals, Systems and Control*, (B. N. Datta, Ed.), SIAM, Philadelphia, 1988.

[20] E. POLAK, R. W. H. SARGENT and D. J. SEBASTIAN, *On the convergence of sequential minimization algorithms*, JOTA, Vol. 14, No. 4, pp. 439-442, 1974.

[21] E. POLAK, *On the mathematical foundations of nondifferentiable optimization in engineering design*, SIAM Review Vol. 29(1) pp. 21-89 (March 1987).

[22] E. POLAK, *Computational Methods in Optimization: A Unified Approach*, Academic Press, 1971.

[23] E. POLAK and A. L. TITS, *A recursive quadratic programming algorithm for semi-infinite optimization problems*, Appl. Math. Optim, Vol. 8, pp. 325-349, 1982.

[24] E. POLAK *A modified secant method for unconstrained optimization*, Mathematical Programming, Vol. 6, No. 3, pp. 264-280, 1974.

[25] E. POLAK, D. Q. MAYNE and J. HIGGINS, *A superlinearly convergent algorithm for min-max problems*, University of California, Berkeley, Electronics Research Laboratory Memo No. M86/103, Nov. 15, 1986. JOTA, in press.

[26] E. POLAK, J. HIGGINS and D. Q. MAYNE, *A Barrier Function Method for Minimax Problems*, University of California, Electronics Research Laboratory, Memo UCB/ERL M88/64, 20 October 1988. Mathematical Programming, in press.

[27] S. M. ROBINSON, *Perturbed Kuhn-Tucker points and rates of convergence for a class of nonlinear-programming algorithms*, Mathematical Programming Vol. 7 pp. 1-16 (1974).

[28] S. M. ROBINSON, *Extension of Newton's method to mixed systems of nonlinear equations and inequalities*, Numerische Mathematik, Vol. 19, pp 341-347, 1972.

[29] R. T. ROCKAFELLAR, *Convex Analysis*, Princeton University Press, Princeton (1970).

Fig. 1. Plots of Control Sequence at Iterations 0, 1, 2.

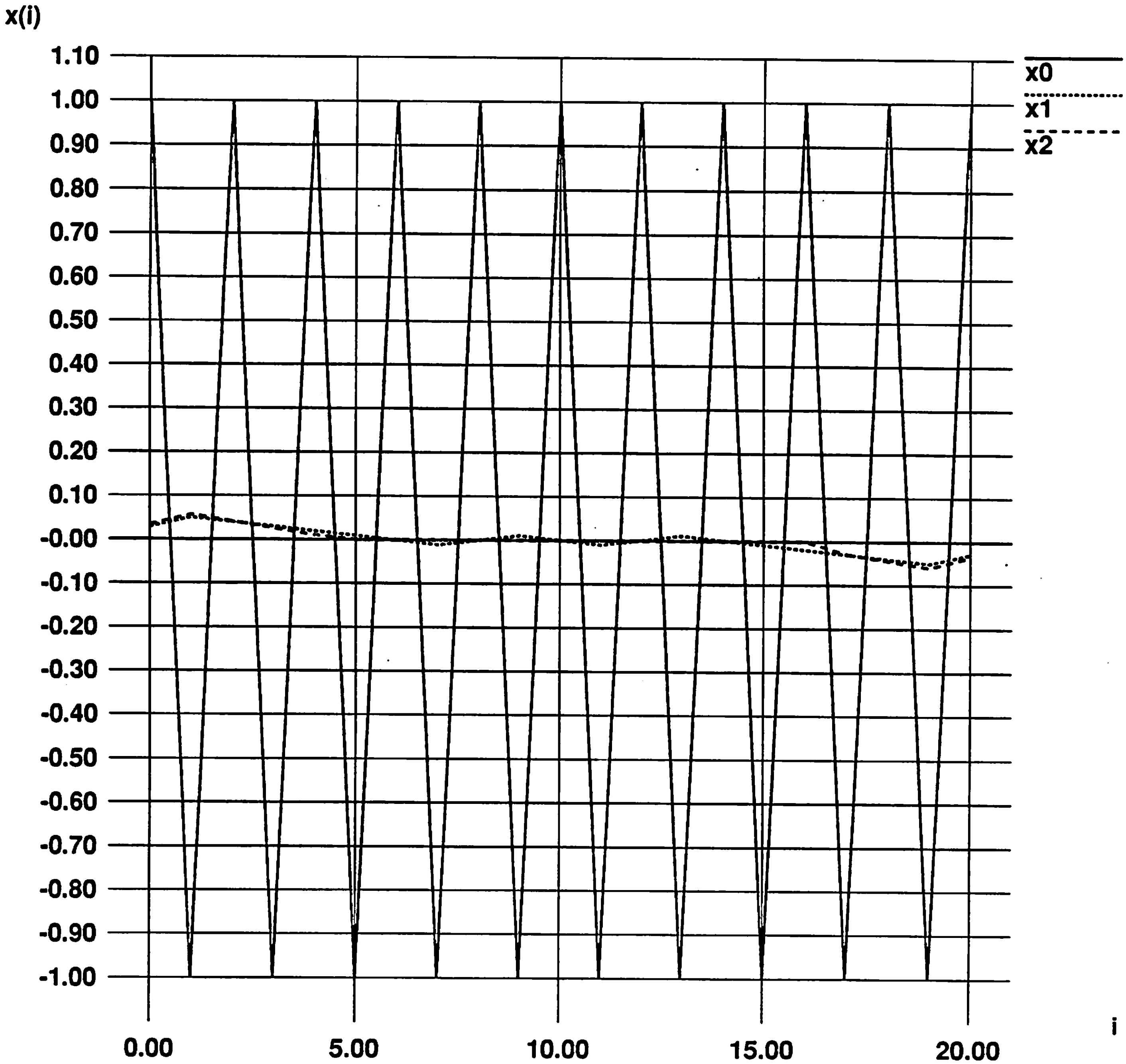

Fig. 2. Phase-Plane Plots at Iterations 0, 1, 2.