ON THE RATE OF CONVERGENCE OF TWO MINIMAX ALGORITHMS

by

E. J. Wiest and E. Polak

Memorandum No. UCB/ERL M89/111

15 August 1989

ON THE RATE OF CONVERGENCE OF TWO MINIMAX ALGORITHMS

by

E. J. Wiest and E. Polak

Memorandum No. UCB/ERL M89/111

15 August 1989

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

ON THE RATE OF CONVERGENCE OF TWO MINIMAX ALGORITHMS

by

E. J. Wiest and E. Polak

Memorandum No. UCB/ERL M89/111

15 August 1989

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

ON THE RATE OF CONVERGENCE OF TWO MINIMAX ALGORITHMS[1]

E. J. Wiest[2] and E. Polak[3]

Abstract. We show that the sequences of function values constructed by two versions of a minimax algorithm converge linearly to the minimum values. Both versions use the Pironneau-Polak-Pshenichnyi search direction subprocedure; the first uses an exact line search to determine step size, while the second one uses an Armijo-type step size rule. The proofs depend on a second-order sufficiency condition, but not on strict complementary slackness. Minimax problems in which each function appearing in the max is a composition of a twice continuously differentiable function with a linear function typically do not satisfy second-order sufficiency conditions. Nevertheless, we show that, on such minimax problems, the two algorithms do converge linearly when the outer functions are convex and strict complementary slackness holds at solutions.

Key Words. Minimax, nonsmooth optimization, composite nondifferentiable optimization, linear convergence.

---

[1] 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.

[2]Graduate student, Department of Electrical Engineering and Computer Sciences and the Electronics Research Laboratory, University of California, Berkeley, CA 94720, U.S.A.

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

1. Introduction

Most minimax algorithms have been shown to converge locally or globally under various conditions. However, the literature dealing with their *rate* of convergence is rather fragmentary (see, e.g., Refs. 1-9). In this paper, we establish the rate of convergence of two versions of a minimax algorithm which was first proposed by Pironneau and Polak (Ref. 10) as a subprocedure in an implementation of the Huard method of centers[1] (Ref. 11), and, later, independently, by Pshenichnyi (Ref. 1), who calls it the method of linearizations. (See Ref. 13 for extensions to semi-infinite minimax problems.)

We are concerned with algorithms for solving minimax problems of the form

$$\min_{x \in \mathbb{R}^n} \max_{j \in \underline{p}} f^j(x), \tag{1a}$$

where $\underline{p} \triangleq \{ 1,2,\ldots,p \}$ and each $f^j:\mathbb{R}^n \to \mathbb{R}$ is continuously differentiable.

We will briefly review the literature dealing with the rate of convergence of first-order minimax algorithms for solving problem (1a). For this, we need to define the function $\psi:\mathbb{R}^n \to \mathbb{R}$ by

$$\psi(x) = \max_{j \in \underline{p}} f^j(x) \,. \tag{1b}$$

First, since problem (1a) can be transcribed into the equivalent constrained form

$$\min\{ w \mid f^j(x) - w \leq 0, j \in \underline{p} \} \,, \tag{2}$$

it can also be solved by first-order nonlinear programming algorithms.[2] For example, it can be solved by the Pironneau-Polak method of centers (Ref. 10) which, as shown by Chaney (Ref. 14), converges linearly on (2) whenever a strengthened second-order sufficiency condition is satisfied.

Subgradient and bundle methods designed for the more general problem of minimizing locally Lipschitz functions can be used for minimizing the function $\psi(\cdot)$. In Polyak (Ref. 2) and Shor (Ref. 3), we find proofs that several subgradient methods converge linearly when $\psi(\cdot)$ is strongly convex. In Ref. 4, Kiwiel has proved that a bundle algorithm for constrained minimax optimization converges

[1] One of the definitions of the "center" of a set described by inequalities, given by Huard (Ref. 11), is in terms of a minimax subproblem of the form (1a). Consequently, every implementation of the Huard method of centers (e.g. - Refs. 10, 12) incorporates a minimax algorithm as a subprocedure. This fact was not widely recognized, and some of these imbedded minimax algorithms were later rediscovered independently.

[2] The transcription of (1a) into (2) is not recommended, because nonlinear programming algorithms converge more slowly on (2) than minimax algorithms designed specifically for (1a).

linearly.

Next, there are several algorithms which were designed specifically for solving minimax problems of the form (1a). One of the oldest is that of Demyanov (Refs. 15, 16), which computes δ-approximations to the minimum value of $\psi(\cdot)$ with $\delta > 0$. It computes search directions by solving a linear program defined by the linearizations of the δ-active functions $f^j(\cdot)$ (c.f., the Zoutendijk method of feasible directions (Ref. 17)), and employs an Armijo-like step size rule. It was shown by Pevny (Ref. 5) that, when $\psi(\cdot)$ is strongly convex, the Demyanov algorithm converges linearly in function value.[3] Madsen et al (Ref. 6) propose a trust region algorithm for the linearly constrained minimax problem in which a linear program is solved at each iteration. When the solution $\hat{x}$ of (1a) is a "vertex" solution (also called a *Chebyshev point* or a *Haar point*), the algorithm in Ref. 6 converges quadratically. However, when $\hat{x}$ is not a vertex solution, the rate of convergence of this algorithm is unknown.

The minimax algorithms which we will discuss in this paper belong to a family conforming to the following algorithm model, which uses the Pironneau-Polak-Pshenichnyi (PPP) search direction subprocedure (Refs. 1, 10):

PPP Algorithm Model

Step 1: Given x_i, compute the search direction,

$$h_i \triangleq \arg \min_{h \in \mathbb{R}^n} \max_{j \in I_i} f^j(x_i) + \langle \nabla f^j(x_i), h \rangle + \tfrac{1}{2}\gamma \| h \|^2 . \tag{3}$$

Step 2: Compute the step size λ_i.

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

Algorithms in this family are specified by the quantities $I_i \subset \underline{p}$, $\gamma > 0$, and a rule for computing the step size λ_i. Thus, in Ref. 1, we find a minimax algorithm in the PPP family with $I_i \triangleq \{ j \in \underline{p} \mid f^j(x_i) \geq \psi(x_i) - \delta \}$ (with $\delta > 0$), $\gamma = 1$, and the *constant* step size $\lambda_i \equiv \lambda$. It is shown in Ref. 1 that the resulting algorithm converges linearly, provided the initial point is sufficiently close to $\hat{x}$. The proof assumes that λ is sufficiently small, and that strict complementary slackness, affine indepen-

[3]Pevnyi also shows that, if the functions fail to be strongly convex but are convex with bounded level sets, convergence to a δ-optimal value is arithmetic.

dence of the gradients of the active functions and second-order sufficiency conditions hold at $\hat{x}$. It is also shown in Ref. 1 that, if $\lambda = 1$ and $\hat{x}$ is a "vertex" solution, then the *local* algorithm converges quadratically.

In Ref. 1, we also find a PPP minimax algorithm which uses the step size rule

$$\lambda_i = \max_{k \in \mathbb{N}} \{ 2^{-k} \mid \psi(x_i + 2^{-k}h_i) - \psi(x_i) \le 2^{-k}\alpha\|h_i\|^2 \}, \quad \alpha \in (½,1) , \tag{4}$$

where $\mathbb{N}$ is the set of all nonnegative integers. It was shown in Ref. 7 that, if (1a) has a "vertex" solution $\hat{x}$, then the step size in the above algorithm eventually becomes unity. It therefore follows from Ref. 1, that if a sequence $\{x_i\}_{i=0}^{\infty}$, constructed by the PPP algorithm using (4), converges to a "vertex" solution $\hat{x}$, then it converges quadratically.

In Ref. 8, $I_i = \underline{p}$, $\gamma > 0$ for all $j \in \underline{p}$ and an Armijo step size rule (Ref. 19) similar to (4) is used, while, in Ref. 9, $I_i = \underline{p}$, $\gamma = 1$ for all $j \in \underline{p}$ and an exact minimizing line search is used to determine step size. It was shown in these papers that both of these PPP methods converge linearly under the assumption that the functions $f^j(\cdot)$ are strongly convex.[4]

In Sections 3 and 4 of this paper, we show that the PPP algorithms, considered in Refs. 8 and 9, converge linearly under a slightly strengthened form of the standard second-order sufficiency condition. This condition is considerably weaker than the strong convexity assumption used in Refs. 8, 9. Furthermore, unlike in Ref.1, we assume neither strict complementary slackness nor affine independence of the gradients of the active functions.

In Section 5, we consider the *composite* minimax problem,

$$\min_{x \in \mathbb{R}^n} \max_{j \in \underline{p}} g^j(A_j x) , \tag{5}$$

in which each continuously differentiable function $g^j : \mathbb{R}^{l_j} \to \mathbb{R}$ is composed with a different linear function $A_j : \mathbb{R}^n \to \mathbb{R}^{l_j}$. Minimax problems of this form arise in the design of feedback compensators and of discrete time optimal controls. We show that, despite the fact that the solution set is generally

[4]To obtain linear convergence in the case that the solution set is an affine set (Ref. 8), strict complementary slackness was assumed in addition.

nonunique,[5] the PPP algorithm described in Ref. 8 converges linearly, under somewhat more stringent assumptions than for the general case.

2. The PPP Minimax Algorithm with Exact Line Search

In this section, we will consider the algorithm which results when the step size λ_i in the Algorithm Model is computed by exact minimization along the search direction. To simplify notation, we define

$$\phi^j(h \mid x) \triangleq f^j(x) + \langle \nabla f^j(x), h \rangle + \tfrac{1}{2}\gamma \|h\|^2 . \tag{6}$$

Algorithm 2.1 (PPP-ELS): (see Algorithm 5.2 and Corollary 5.1 in Ref. 13)

Data: $x_0 \in \mathbb{R}^n$; $\gamma > 0$.

Step 0: Set $i = 0$.

Step 1: Compute the search direction[6],

$$h(x_i) \triangleq \arg \min_{h \in \mathbb{R}^n} \max_{j \in \underline{p}} \phi^j(h \mid x_i) - \psi(x_i). \tag{7}$$

Step 2: Compute the minimizing step size, $\lambda_i = \arg \min_{\lambda \in \mathbb{R}} \psi(x_i + \lambda h(x_i))$.

Step 3: Set $x_{i+1} = x_i + \lambda_i h(x_i)$, replace i by $i + 1$ and go to Step 1. ■

Let the standard unit simplex be denoted by $\Sigma_p \triangleq \{ \mu \in \mathbb{R}^p \mid \mu^j \geq 0 , \sum_{j \in \underline{p}} \mu^j = 1 \}$. Then the search direction finding problem (7) can be transformed as follows:

$$\theta(x) \triangleq \min_{h \in R^n} \left[\max_{j \in \underline{p}} \phi^j(h \mid x) - \psi(x) \right]$$
$$= \min_{h \in R^n} \left[\max_{\mu \in \Sigma_p} \sum_{j \in \underline{p}} \mu^j \phi(h \mid x) - \psi(x) \right] . \tag{8}$$

Next, by an extension to von Neumann's Minimax Theorem (Ref. 13), the max and min in (8) can be interchanged, and hence we obtain that

[5] In fact, the solution set must contain a translation of the intersection of the null spaces of the matrices A_j.

[6] For the convenience of the proofs to follow, we subtract the term $\psi(x_i)$ from the minimand in (4), so as to make the value $\theta(x_i)$ of the search direction finding problem less than or equal to zero. This has no effect on the resulting search direction.

$$\theta(x) = \max_{\mu \in \sum_p} \min_{h \in \mathbb{R}^n} \sum_{j \in \underline{p}} \mu^j \phi(h \mid x) - \psi(x)$$

$$= \max_{\mu \in \sum_p} \min_{h \in \mathbb{R}^n} \sum_{j \in \underline{p}} \mu^j \left(f^j(x) + \langle \nabla f^j(x), h \rangle - \psi(x) \right) + \tfrac{1}{2}\gamma |h|^2 . \tag{9}$$

The solution μ of (9) is not always unique, and hence we define the solution set

$$U(x) \triangleq \arg \max_{\mu \in \sum_p} \left[\min_{h \in \mathbb{R}^n} \sum_{j \in \underline{p}} \mu^j \left(f^j(x) + \langle \nabla f^j(x), h \rangle - \psi(x) \right) + \tfrac{1}{2}\gamma |h|^2 \right] . \tag{10a}$$

By solving the inner minimization problem in (10a), we see that $U(x)$ is the solution set to a positive semi-definite quadratic program,

$$U(x) = \arg \max_{\mu \in \sum_p} \sum_{j \in \underline{p}} \mu^j \left(f^j(x) - \psi(x) \right) - \frac{1}{2\gamma} \Big| \sum_{j \in \underline{p}} \mu^j \nabla f^j(x) \Big|^2 . \tag{10b}$$

Several methods exist for solving such problems (see, for example, Refs. 21-25).

As a consequence of the extended von Neumann Minimax Theorem, for any $\bar{\mu} \in U(x)$,

$$\sum_{j \in \underline{p}} \bar{\mu}^j \phi^j(h(x) \mid x) \le \max_{\mu \in \Sigma_p} \sum_{j \in \underline{p}} \mu^j \phi^j(h(x) \mid x)$$

$$= \min_{h \in \mathbb{R}^n} \max_{\mu \in \Sigma_p} \sum_{j \in \underline{p}} \mu^j \phi^j(h \mid x)$$

$$= \max_{\mu \in \Sigma_p} \min_{h \in \mathbb{R}^n} \sum_{j \in \underline{p}} \mu^j \phi^j(h \mid x)$$

$$= \min_{h \in \mathbb{R}^n} \sum_{j \in \underline{p}} \bar{\mu}^j \phi^j(h \mid x) . \tag{11}$$

Hence, any multiplier vector $\bar{\mu} \in U(x)$ yields the solution,

$$h(x) = \arg \min_{h \in \mathbb{R}^n} \sum_{j \in \underline{p}} \bar{\mu}^j \phi^j(h \mid x) \tag{12}$$

to the search direction finding problem (7) (for $x = x_i$), which is unique since the function $\max_{j \in \underline{p}} \phi^j(\cdot \mid x)$ is strictly convex.

Next we recall the following necessary optimality condition for problem (1) (see Ref. 13).

Theorem 2.1: (Ref. 13) *If $\hat{x} \in \mathbb{R}^n$ is a solution to problem (1), then there exists a vector of multipliers $\hat{\mu} \in \Sigma_p$ such that*

$$\sum_{j \in \underline{p}} \hat{\mu}^j \nabla f^j(\hat{x}) = 0 , \tag{13a}$$

$$\sum_{j \in \underline{p}} \hat{\mu}^j [f^j(\hat{x}) - \psi(\hat{x})] = 0 . \tag{13b}$$

■

When the functions $f^j(\cdot)$ are convex, equations (13a, 13b) are also a sufficient condition for optimality. We denote the minimum value for problem (1) by $\hat{\psi} \triangleq \min_{x \in \mathbb{R}^n} \psi(x)$ and the set of minimizers by $\hat{G} \triangleq \arg\min_{x \in \mathbb{R}^n} \psi(x)$. For any $\hat{x} \in \hat{G}$, the set of multiplier vectors $\hat{\mu} \in \Sigma_p$ which satisfy equations (13a, 13b) together with $\hat{x}$ is exactly $U(\hat{x})$.

Theorem 2.2: (Ref. 13) *Suppose that the functions $f^j(\cdot)$ in problem (1) have continuous derivatives. If $\bar{x}$ is an accumulation point of a sequence $\{ x_i \}_{i=0}^{\infty}$ constructed by Algorithm 2.1, then $\bar{x}$ satisfies the optimality condition (13a, 13b).* ■

3. Rate of Convergence of the PPP-ELS Algorithm

We now proceed to prove that the sequence of values, $\{ \psi(x_i) \}_{i=0}^{\infty}$, constructed by Algorithm 2.1 converges linearly to the minimum value under weaker assumptions than those used in Refs. 8 and 9. Our proof draws on ideas which appeared in the proofs of linear convergence of the Pironneau-Polak algorithm for inequality-constrained minimization in Refs. 10 and 14. We make the following assumptions.

Hypothesis 3.1: *We will assume that*

(i) *the functions $f^j(\cdot)$ are twice continuously differentiable,*

(ii) *there exists $T \in \mathbb{R}$ such that the set $S \triangleq \{ x \in \mathbb{R}^n \mid \psi(x) \leq T \}$ is bounded and such that there is a single point $\hat{x} \in S$ which satisfies the necessary conditions* (13a, 13b),

(iii) *for some $M' < \infty$*, all $x \in \mathbb{R}^n$ *and all $j \in \underline{p}$*, $\| F^j(x) \|_2 < M'$. ■

For any stationary point $\hat{x}$, we define the set of indices of the functions active at $\hat{x}$ by

$$J(\hat{x}) \triangleq \{ j \in \underline{p} \mid \exists \mu \in U(\hat{x}) : \mu^j > 0 \} . \tag{15}$$

Hypothesis 3.2: *Let $\hat{x}$ be as defined in Hypothesis 3.1, let B denote the subspace spanned by the vectors $\{ \nabla f^j(\hat{x}) \}_{j \in J(\hat{x})}$, and let $B^\perp$ denote the orthogonal complement of B. We will assume that there exists $m' > 0$ such that, for all $\hat{\mu} \in U(\hat{x})$,*

$$m' \|h\|^2 < \langle h, \left[\sum_{j \in \underline{p}} \hat{\mu}^j F^j(\hat{x}) \right] h \rangle \quad \forall\, h \in B^\perp . \tag{16}$$

■

Remark: Hypothesis 3.2 and equations (13a, 13b) together constitute a variation on the standard second-order sufficiency conditions for $\hat{x}$ to be a local minimizer of $\psi(\cdot)$ (Ref. 18). Note that, while (16) must hold for *all* multiplier vectors in $U(\hat{x})$, the subspace $B^\perp$ over which the inequality must hold may be quite small, because all of the multiplier vectors in $U(\hat{x})$ are used to determine the set $J(\hat{x})$. ■

The proof of linear convergence requires several technical lemmas involving the following quantities. With $\hat{x}$ as in Hypothesis 3.1 (ii) and B as in Hypothesis 3.2, let $P : \mathbb{R}^n \to \mathbb{R}^n$ denote the projection operator with range equal to B, and let $P^\perp$ be the projection operator with range equal to $B^\perp$. Let

$$m \triangleq \min\{ m', \gamma \} . \tag{17a}$$

For any $y \in \mathbb{R}^n$ and $\mu \in \Sigma_p$, we define

$$R(y,\mu) \triangleq \tfrac{1}{2} m\, I - \int_0^1 (1 - s) \sum_{j \in \underline{p}} \mu^j F^j(\hat{x} + (1 - s) y) ds . \tag{17b}$$

The function $R(\cdot,\cdot)$ is continuous, and, by Hypothesis 3.2, for any $\hat{\mu} \in U(\hat{x})$, $R(0,\hat{\mu})$ is negative definite on the subspace $B^\perp$.

We will use the notation $z_i \to Z$ to indicate the convergence of the sequence $\{ z_i \}_{i=0}^\infty \subset \mathbb{R}^n$ to the set $Z \subset \mathbb{R}^n$, *i.e.*, the fact that $\lim_{i \to \infty} \min_{y \in Z} \|z_i - y\| = 0$. The following two results are established in the Appendix.

Lemma 3.1: *If Hypotheses 3.1 and 3.2 hold, and $\hat{x}$ is defined as in Hypothesis 3.1, then there exists $K > 0$ such that*

$$\limsup_{\substack{y \to 0 \\ \mu \to U(\hat{x})}} \frac{\langle y, R(y,\mu) y \rangle}{\|y\| \, \|Py\|} < K . \tag{18}$$

■

Lemma 3.2: *If Hypotheses 3.1 and 3.2 hold, and* $\hat{x}$ is defined as in Hypothesis 3.1, then

$$\lim_{x \to \hat{x}} \frac{|x - \hat{x}|\, |P(x - \hat{x})|}{\psi(x) - \hat{\psi}} = 0 . \tag{19}$$

■

We now relate the potential decrease in the function $\psi(\cdot)$ at x to the decrease predicted by $\theta(x)$.

Lemma 3.3: *If Hypotheses 3.1 and 3.2 hold, then*

$$\limsup_{x \to \hat{x}} \frac{\theta(x)}{\psi(x) - \hat{\psi}} \leq -\frac{m}{\gamma} . \tag{20}$$

Proof: Referring to (9) and (10), we see that For any $\bar{\mu} \in U(x)$,

$$\theta(x) = \min_{h \in \mathbb{R}^n} \sum_{j \in \underline{p}} \bar{\mu}^j \phi^j(h \mid x) - \psi(x) . \tag{21}$$

Let $s(x) \triangleq m / \gamma$. By the definition of m above, $s(x) < 1$. Substituting $h = s(x)(\hat{x} - x)$ in (21) and using the definition of $\phi^j(\cdot \mid \cdot)$ in (2), we obtain that

$$\begin{aligned}
\theta(x) &\leq \sum_{j \in \underline{p}} \bar{\mu}^j \phi^j(s(x)(\hat{x} - x) \mid x) - \psi(x) \\
&= \sum_{j \in \underline{p}} \bar{\mu}^j \left[f^j(x) - \psi(x) + \langle \nabla f^j(x), s(x)(\hat{x} - x)\rangle + \tfrac{1}{2}\gamma s(x)^2 |\hat{x} - x|^2 \right] \\
&\leq s(x) \left\{ \sum_{j \in \underline{p}} \bar{\mu}^j f^j(x) + \langle \sum_{j \in \underline{p}} \bar{\mu}^j \nabla f^j(x), \hat{x} - x\rangle + \tfrac{1}{2} m |\hat{x} - x|^2 - \psi(x) \right\},
\end{aligned} \tag{22}$$

since $s(x) \in (0,1)$ and $f^j(x) \leq \psi(x)$. Adding and subtracting the term $\langle x - \hat{x}, \left[\int_0^1 (1 - t) \sum_{j \in \underline{p}} \mu^j F^j(\hat{x} + (1 - t)(x - \hat{x}))dt \right] (x - \hat{x})\rangle$ to the right hand side of (22), we find that

$$\begin{aligned}
\theta(x) \leq s(x) \Bigg\{ & \sum_{j \in \underline{p}} \bar{\mu}^j f^j(x) + \langle \sum_{j \in \underline{p}} \bar{\mu}^j \nabla f^j(x), \hat{x} - x\rangle + \langle x - \hat{x}, \left[\int_0^1 (1 - t) \sum_{j \in \underline{p}} \bar{\mu}^j F^j(\hat{x} + (1 - t)(x - \hat{x}))dt \right] (x - \hat{x})\rangle \\
& - \psi(x) + \langle x - \hat{x}, R(x - \hat{x}, \bar{\mu})(x - \hat{x})\rangle \Bigg\}
\end{aligned} \tag{23}$$

The first three terms in the right hand side of (23) constitute the second-order Taylor expansion of

$\sum_{j \in p} \bar{\mu}^j f^j(\hat{x})$. Hence,

$$\theta(x) \le s(x)\left\{ \sum_{j \in p} \bar{\mu}^j f^j(\hat{x}) - \psi(x) + \langle x - \hat{x}, R(x - \hat{x}, \bar{\mu})(x - \hat{x})\rangle \right\}$$

$$\le s(x)\left\{ \psi(\hat{x}) - \psi(x) + \langle x - \hat{x}, R(x - \hat{x}, \bar{\mu})(x - \hat{x})\rangle \right\}. \tag{24}$$

Dividing both sides of (24) by $\psi(x) - \hat{\psi}$, we get

$$\frac{\theta(x)}{\psi(x) - \psi(\hat{x})} \le s(x)\left\{ -1 + \frac{\langle x - \hat{x}, R(x - \hat{x}, \bar{\mu})(x - \hat{x})\rangle}{\psi(x) - \psi(\hat{x})} \right\}. \tag{25}$$

By Lemma 3.1,

$$\limsup_{x \to \hat{x}} \max_{\mu \in U(x)} \frac{\langle x - \hat{x}, R(x - \hat{x}, \mu)(x - \hat{x})\rangle}{|x - \hat{x}|\, |P(x - \hat{x})|} < K\,, \tag{26}$$

and, by Lemma 3.2,

$$\limsup_{x \to \hat{x}} \frac{|x - \hat{x}|\, |P(x - \hat{x})|}{\psi(x) - \hat{\psi}} = 0\,. \tag{27}$$

Since the set-valued map $U(x)$ is upper semicontinuous, $s(x)$ is lower semicontinuous and

$$\liminf_{x \to \hat{x}} s(x) \ge \frac{m}{\gamma}\,. \tag{28}$$

Taking the lim sup of (25) as $x \to \hat{x}$ and using (26), (27) and (28) yields (20). ■

We combine Lemma 3.3 with a relation between the decrease predicted by $\theta(x)$ and the actual decrease obtained at x in the direction $h(x)$ using an exact line search. Let

$$M \triangleq \max\{\, M', \gamma \,\}\,. \tag{29}$$

Lemma 3.4: *If Hypotheses 3.1 and 3.2 hold, then*

$$\limsup_{\substack{x \to \hat{x} \\ x \ne \hat{x}}} \min_{\lambda \in \mathbb{R}} \frac{\psi(x + \lambda h(x)) - \hat{\psi}}{\psi(x) - \hat{\psi}} \le 1 - \frac{\min\{\, m', \gamma \,\}}{\max\{\, M', \gamma \,\}}\,. \tag{30}$$

Proof: Since by Hypothesis 3.1(ii), $\hat{x}$ is the only point in S satisfying the necessary condition for optimality (13a, 13b), it follows that $\psi(\hat{x}) = \hat{\psi}$ and $\psi(x) > \hat{\psi}$ for all $x \in S \setminus \hat{x}$. Since $\theta(x)$ is zero if and only if the necessary conditions (13a, 13b) are met at x, $\theta(x) < 0$ for all $x \in S \setminus \hat{x}$.

The second derivative bound of Hypothesis 3.1(iii) implies that for each $f^j(\cdot)$,

$$f^j(y+z) - f^j(y) - \langle \nabla f^j(y), z\rangle \le \tfrac{1}{2} M \|z\|^2 , \quad \forall\, y,z \in \mathbb{R}^n . \tag{31}$$

Thus, for any $\bar{\lambda} \in (0,1)$ and $x \in S \setminus \hat{x}$,

$$\begin{aligned}
\min_{\lambda \in \mathbb{R}} \psi(x+\lambda h(x)) - \psi(x) &\le \psi(x+\bar{\lambda} h(x)) - \psi(x) \\
&\le \max_{j \in \underline{p}} f^j(x) - \psi(x) + \langle \nabla f^j(x), \bar{\lambda} h(x)\rangle + \tfrac{1}{2} M \|\bar{\lambda} h(x)\|^2 , \\
&\le \bar{\lambda}\left[\max_{j \in \underline{p}} f^j(x) - \psi(x) + \langle \nabla f^j(x), h(x)\rangle + \tfrac{1}{2}\bar{\lambda} M \|h(x)\|^2 \right].
\end{aligned} \tag{32}$$

Setting $\bar{\lambda} = \gamma / M$ and using the definition of M,

$$\min_{\lambda \in \mathbb{R}} \psi(x+\lambda h(x)) - \psi(x) \le \bar{\lambda}\left[\max_{j \in \underline{p}} f^j(x) - \psi(x) + \langle \nabla f^j(x), h(x)\rangle + \tfrac{1}{2}\gamma \|h(x)\|^2 \right] = \bar{\lambda}\theta(x) . \tag{33}$$

Since $\theta(x) < 0$ for all $x \ne \hat{x}$,

$$\min_{\lambda \in \mathbb{R}} \frac{\psi(x+\lambda h(x)) - \psi(x)}{\theta(x)} \ge \bar{\lambda} = \frac{\gamma}{M} . \tag{34}$$

Applying inequality (34) and Lemma 3.3 to the left hand side of (30), we obtain

$$\begin{aligned}
\limsup_{\substack{x \to \hat{x} \\ x \ne \hat{x}}} \min_{\lambda \in \mathbb{R}} \frac{\psi(x+\lambda h(x)) - \psi(x)}{\psi(x) - \hat{\psi}} &= \limsup_{\substack{x \to \hat{x} \\ x \ne \hat{x}}} \min_{\lambda \in \mathbb{R}} \frac{\psi(x+\lambda h(x)) - \psi(x)}{\theta(x)} \; \frac{\theta(x)}{\psi(x) - \hat{\psi}} \\
&\le \frac{\gamma}{M} \limsup_{\substack{x \to \hat{x} \\ x \ne \hat{x}}} \frac{\theta(x)}{\psi(x) - \hat{\psi}} \\
&\le \frac{\gamma}{M}\left[\frac{-m}{\gamma}\right] \\
&= -\frac{m}{M} = -\frac{\min\{ m', \gamma \}}{\max\{ M', \gamma \}} .
\end{aligned} \tag{35}$$

The second step holds because $\theta(x) < 0$ and $\psi(x) > \hat{\psi}$. Adding 1 to both sides yields the desired result.

■

Theorem 3.2: *If Hypotheses 3.1 and 3.2 hold and Algorithm 2.1 generates a sequence* $\{ x_i \}_{i=0}^{\infty}$, *starting from a point* $x_0 \in S$, *then (a)* $x_i \to \hat{x}$ *as* $i \to \infty$, *and (b) either the sequence terminates in a finite number of steps at* $\hat{x}$ *or*

$$\limsup_{i \to \infty} \frac{\psi(x_{i+1}) - \hat{\psi}}{\psi(x_i) - \hat{\psi}} \le 1 - \frac{\min\{ m', \gamma \}}{\max\{ M', \gamma \}} . \tag{36}$$

Proof: (a) The sequence $\{ x_i \}_{i=0}^{\infty}$ lies in the compact set S, and hence it converges to the set of its accumulation points. By Theorem 2.2, each accumulation point must satisfy the necessary conditions (13a, 13b). Since, by Hypothesis 3.1(ii), only $\hat{x} \in S$ satisfies (13a, 13b), the sequence converges to $\hat{x}$.

(b) Follows from (a) and Lemma 3.5. ■

Following Luenberger (Ref. 26), we refer to the quantity $\limsup_{i \to \infty} (\psi(x_{i+1}) - \hat{\psi})/(\psi(x_i) - \hat{\psi})$ as the *convergence ratio* of the sequence $\{ \psi(x_i) \}_{i=0}^{\infty}$. The right-hand side of inequality (36) bounds the convergence ratio of any sequence constructed by the PPP-ELS Algorithm in solving any problem in the class defined by Hypotheses 3.1 and 3.2.

Remark 3.1: The functions $\phi^j(\cdot \mid \cdot)$, could have been defined in (6) using different values of γ, say γ_j. This would have had two effects. First, the search direction finding problem would have been considerably more difficult to solve. Second, the convergence ratio bound in the right-hand side of inequality (36) could turn out to be larger; certainly it would not be smaller. However, if individual bounds on the $\|F^j(x)\|$ are assumed, one may be able to establish lower bounds when using different γ_j. ■

Remark 3.2: From the definition of m and M in (17a) and (29), the ratio m/M appearing in the convergence ratio bound is independent of γ provided that $\gamma \in [m', M']$. However, for γ outside this range, m/M is smaller and the convergence ratio bound is greater. The following example shows that this dependence of the convergence ratio bound on γ is not an artifact of our proof technique, but that it reflects the dependence of the actual convergence ratios on γ. We applied the PPP-Armijo Algorithm (see Section 4) to the problem of minimizing the maximum of $f^1(x) \triangleq -6x_0 + 4(x_0^2 + x_1^2)$ and $f^2(x) \triangleq x_0 + \frac{1}{2}(x_0^2 + x_1^2)$ using a variety of values for γ. For this problem $m' = 2$ and $M' = 8$. We started the algorithm from the point (1,1), and used $\gamma \in \{ 2^{-3}, 2^{-2}, 2^{-1}, 2^0, 2^1, 2^2, 2^4, 2^5, 2^6 \}$. Figure 1 displays both the convergence ratio bounds computed from the right-hand side of (38) and the convergence ratios which were observed. ■

4. Rate of Convergence of the PPP-Armijo Algorithm

The step size rule used in Algorithm 2.1 calls for the exact minimization of a function of a single variable. In practice, we use a step size rule which can be executed in a finite number of steps. A suitable replacement for Step 2 in Algorithm 2.1 is the following generalization (Ref. 13) of the Armijo rule for differentiable functions (Ref. 19),

Step 2': Compute the step size,

$$\lambda_i = \max_{k \in \mathbb{Z}} \{ \beta^k \mid \psi(x_i + \beta^k h_i) - \psi(x_i) - \alpha \beta^k \theta(x_i) \leq 0 \} , \tag{37}$$

with fixed parameters $\alpha, \beta \in (0,1)$. We will call the resulting algorithm the PPP-Armijo algorithm. The convergence result, Theorem 2.2, holds for the PPP-Armijo algorithm (Ref. 13). We show that a rate of convergence result very similar to Theorem 3.2 holds as well.

Theorem 4.1: *If Hypotheses 3.1 and 3.2 hold and the PPP-Armijo algorithm generates a sequence $\{ x_i \}_{i=0}^{\infty}$, starting from a point $x_0 \in S$, then (a) $x_i \to \hat{x}$, as $i \to \infty$, and (b) either the sequence terminates in a finite number of steps at $\hat{x}$ or*

$$\limsup_{i \to \infty} \frac{\psi(x_{i+1}) - \hat{\psi}}{\psi(x_i) - \hat{\psi}} \leq 1 - \alpha\beta \frac{\min\{ m', \gamma \}}{\max\{ M', \gamma \}} . \tag{38}$$

Proof: (a) The sequence $\{ x_i \}_{i=0}^{\infty}$ is contained in the compact set S, and hence it converges to the set of its accumulation points. Referring to Ref. 13 and using the fact that the functions $f^j(\cdot)$ are continuously differentiable, we conclude that any accumulation point must satisfy the necessary conditions (13a, 13b). Since the only point in S satisfying these conditions is $\hat{x}$, the sequence must converge to $\hat{x}$.
(b) We obtain a bound on the decrease in $\psi(\cdot)$ obtained at each iteration, assuming that the sequence does not terminate in a finite number of steps at $\hat{x}$. The second derivative bounds again imply relation

(31), and so, for all $i \in \mathbf{N}$ and $k \geq 0$,

$$\psi(x_i + \beta^k h_i) - \psi(x_i) = \max_{j \in \underline{p}} f^j(x_i + \beta^k h_i) - \psi(x_i)$$

$$\leq \max_{j \in \underline{p}} f^j(x_i) + \langle \nabla f^j(x_i), \beta^k h_i \rangle - \psi(x_i) + \tfrac{1}{2} M \beta^{2k} |h_i|^2$$

$$\leq \beta^k \left[\max_{j \in \underline{p}} f^j(x_i) + \langle \nabla f^j(x_i), h_i \rangle - \psi(x_i) + \tfrac{1}{2} M \beta^k |h_i|^2 \right], \qquad (39)$$

because $\beta^k \leq 1$ and $f^j(x) \leq \psi(x)$. Therefore, if $\beta^k \leq \min_{j \in \underline{p}} \gamma / M$,

$$\psi(x_i + \beta^k h_i) - \psi(x_i) \leq \beta^k \left[\max_{j \in \underline{p}} f^j(x_i) + \langle \nabla f^j(x_i), h_i \rangle - \psi(x_i) + \tfrac{1}{2} \gamma |h_i|^2 \right]$$

$$= \beta^k \theta(x_i) < \alpha \beta^k \theta(x_i) < 0 . \qquad (40)$$

It follows from (37) that $\lambda_i \geq \beta\gamma / M$ and hence that

$$\psi(x_{i+1}) - \psi(x_i) \leq \frac{\alpha\beta\gamma}{M} \theta(x_i) . \qquad (41)$$

Combining inequality (41) with Lemma 3.3 yields the desired result. ■

5. Rate of Convergence of the PPP Algorithms on Composite Minimax Problems

Next we will establish the rate of convergence of the PPP-ELS and PPP-Armijo algorithms on a class of composite minimax problems of the form

$$\min_{x \in \mathbb{R}^n} \max_{j \in \underline{p}} g^j(A_j x) , \qquad (42)$$

where $g^j : \mathbb{R}^{l_j} \to \mathbb{R}$ is continuously differentiable and A_j is an $l_j \times n$ real matrix. We note that (42) is a problem of the form (1a), with the functions $f^j(\cdot)$ defined by $f^j \triangleq g^j \circ A_j$. In conformity with the previous sections, we will use the notation $\psi(x) = \max_{j \in \underline{p}} g^j(A_j x)$. We note that when the null spaces of the matrices A_j have a nontrivial intersection, which we will call their *common null space*, problem (42) does *not* have a unique minimum and therefore does not satisfy Hypothesis 3.1(ii). In this case, problem (42) may also fail to satisfy the convexity requirement of Hypothesis 3.2. To see this, note that for problem (42), the second derivative of the Lagrangian at a minimizer $\hat{x}$ has the form,

$$\sum_{j \in \underline{p}} \hat{\mu}^j A_j^T G^j(A_j \hat{x}) A_j , \qquad (43)$$

where $G^j(\cdot)$ denotes the second derivative matrix of $g^j(\cdot)$. Continuing to denote by B the subspace spanned by the vectors $\{ \nabla f^j(\hat{x}) \}_{j \in J(\hat{x})}$, we find that the second derivative matrix will only be positive *semi*-definite on the subspace $B^{\perp}$. However, we have observed in computational experiments that linear convergence of the values $\{ \psi(x_i) \}_{i=0}^{\infty}$ constructed by PPP-ELS and PPP-Armijo is not lost in this circumstance. In this section we will derive a bound on the rate of convergence of PPP-ELS and the PPP-Armijo algorithms under the assumption that the Lagrangian Hessian is positive definite only on the orthogonal complement of the common null space of the matrices A_j.

By analogy with nonlinear programming, we shall say that *strict complementary slackness* holds at a solution $\hat{x}$ of (1a) (or (42)) if, for *every* multiplier vector $\hat{\mu}$ satisfying (13a, 13b) with $\hat{x}$,

$$\hat{\mu}^j > 0 \ \text{ if and only if } \ f^j(\hat{x}) = \psi(\hat{x}) \,. \tag{14}$$

When strict complementary slackness holds at a minimizer $\hat{x}$, the multiplier vector $\hat{\mu}$ satisfying (13a, 13b) with $\hat{x}$ is unique and hence $J(\hat{x}) = \{ j \in \underline{p} \mid f^j(\hat{x}) = \psi(\hat{x}) \}$.

Proposition 5.1: *Suppose that the functions $g^j(\cdot)$, are strictly convex and that strict complementary slackness holds for every $\hat{x} \in \hat{G}$. Then, (a) there is a unique $\hat{\mu}$ such that $U(\hat{x}) = \{ \hat{\mu} \}$ for all $\hat{x} \in \hat{G}$, and (b) the set $\hat{J} \triangleq J(\hat{x})$ is independent of $\hat{x}$ for all $\hat{x} \in \hat{G}$.*

Proof: We show (b) first. Suppose that $\mu_1, \mu_2 \in U(\hat{x})$ for some $\hat{x} \in \hat{G}$, and that $\mu_1 \neq \mu_2$. Let t and j_0 be defined by

$$t \triangleq \min_{j \in \underline{p}} \{ \mu_1^j / (\mu_1^j - \mu_2^j) \mid \mu_1^j > \mu_2^j \} > 0 \,, \tag{44a}$$

$$j_0 \triangleq \arg \min_{j \in \underline{p}} \{ \mu_1^j / (\mu_1^j - \mu_2^j) \mid \mu_1^j > \mu_2^j \}. \tag{44b}$$

Then $\mu_t \triangleq \mu_1 + t(\mu_2 - \mu_1) \in \Sigma_p$ satisfies (13a, 13b) with $\hat{x}$, and hence $\mu_t \in U(\hat{x})$. By construction, $\mu_t^{j_0} = 0$. Hence, it follows from the strict complementary slackness assumption that $f^j(\hat{x}) < \psi(\hat{x})$. However, $\mu_1^{j_0} > 0$, and hence, again by strict complementary slackness, $f^j(\hat{x}) = \psi(\hat{x})$. This contradiction shows that $U(\hat{x})$ is a singleton for each $\hat{x} \in \hat{G}$.

Suppose that $j_1 \in J(\hat{x}')$ but $j_1 \notin J(\hat{x}'')$ for distinct points $\hat{x}', \hat{x}'' \in \hat{G}$. Then $g^{j_1}(A_{j_1}\hat{x}'') < \psi(\hat{x}'')$. Let $\hat{x}_t \triangleq t\hat{x}' + (1-t)\hat{x}''$. Then $\hat{x}_t \in \hat{G}$ for all $t \in [0,1]$, and, by the convexity of $g^{j_1}(\cdot)$, $g^{j_1}(A_{j_1}\hat{x}_t) < \psi(\hat{x}'') = \psi(\hat{x}_t)$ for all $t \in (0,1)$. It follows from (i) above that $U(\hat{x}_t) = \{ \mu_t \}$, a singleton, and from (13b) that $\mu_t^{j_1} = 0$ for all $t \in (0,1)$. Now, by the Maximum Theorem in Ref. 20, $U(\cdot)$ is an upper semicontinuous set-valued map. Since $U(\hat{x}') = \{ \hat{\mu}' \}$, a singleton, $U(\cdot)$ is continuous at $\hat{x}'$. Hence $\mu_t \to \hat{\mu}'$ as $t \to 1$, which implies that $\hat{\mu}'^{j_1} = 0$. Since $j_1 \in J(\hat{x}')$, this contradicts our strict complementary slackness assumption, and we conclude that (b) holds.

Now we prove (a). Suppose that $\hat{x}', \hat{x}'' \in \hat{G}$. From (b), $g^j(A_j(\hat{x}' + t(\hat{x}'' - \hat{x}')))$ is constant for all $t \in [0,1]$ and all $j \in \hat{J}$. Since each $g^j(\cdot)$ is strictly convex, we conclude that $A_j(\hat{x}' - \hat{x}'') = 0$ for each $j \in \hat{J}$. Therefore, for all $j \in \hat{J}$, $A_j^T \nabla g^j(A_j\hat{x}') = A_j^T \nabla g^j(A_j\hat{x}'')$ and hence any $\hat{\mu}$ satisfying (13a, 13b) with $\hat{x}'$ satisfies (13a, 13b) with $\hat{x}''$. This and the fact that $U(\hat{G})$ is a singleton imply (a). ■

Proposition 5.2: *There exists a neighborhood, W, of $\hat{G}$ such that, for all $x \in W$, $\mu^j = 0$ for all $\mu \in U(\hat{x})$ and $j \notin J(\hat{x})$.*

Proof: (a) Since $h(x)$ is the solution of the primal problem (6), it satisfies the optimality conditions (13a, 13b) with the functions $f^j(\cdot)$ replaced by $\phi^j(\cdot \mid x)$. Every $\mu \in U(x)$ satisfies equations (13a, 13b) together with $h(x)$, and hence the second of those equations yields

$$\sum_{j \in \underline{p}} \mu^j \left[\phi^j(h(x) \mid x) - \psi(x) - \theta(x) \right] = 0 . \tag{45}$$

By Proposition 5.5 in Ref. 13, $h(\hat{x}) = 0$ and $\theta(\hat{x}) = 0$ for every $\hat{x} \in \hat{G}$. Since both functions are continuous, $h(x) \to 0$ and $\theta(x) \to 0$ as $x \to \hat{G}$. Therefore, $\phi^j(h(x) \mid x) \to g^j(A_j\hat{x})$ as $x \to \hat{x} \in \hat{G}$ for all j, implying that

$$\phi^j(h(x) \mid x) - \psi(x) - \theta(x) < 0 \tag{46}$$

for every $j \notin J(\hat{x})$ in some neighborhood, W, of $\hat{G}$. It follows from (45) and (46) that, for all $x \in W$, $\mu^j = 0$ for all $j \notin J(\hat{x})$ for all $\mu \in U(x)$.

■

We now proceed to show that the PPP-ELS and PPP-Armijo algorithms converge linearly on some problems of the form (42) which do not satisfy the assumptions of Theorem 3.2. Letting $j_1 < ... < j_b$ be the indices constituting $\hat{J}$, with $\hat{J}$ defined as in Proposition 5.1, we define $\hat{A}^T \triangleq [A_{j_1}^T, ..., A_{j_b}^T]$. First we will show that the tail of a sequence $\{ x_i \}_{i=0}^{\infty}$ generated either by PPP-ELS or by PPP-Armijo is contained in a translation of the range of $\hat{A}^T$. We will then show that the sequence corresponds to that constructed by the corresponding PPP algorithm on the following restriction of problem (42) to a translation of the range of $\hat{A}^T$:

$$\min_{y \in \mathbb{R}^a} \psi(\bar{x} + Zy) , \tag{47}$$

where $a \triangleq rank(\hat{A}^T)$, and Z is a matrix, the columns of which form an orthonormal basis for $Range(\hat{A}^T)$. Finally, we will show that the restricted problem (47) satisfies the assumptions of Theorem 3.2. We will use the notation $\sigma^+[X]$ to denote the minimum *positive* eigenvalue of any symmetric, positive semi-definite matrix X.

Theorem 5.1: *Suppose that*

(i) the functions $g^j(\cdot)$ are twice continuously differentiable,

(ii) there exist constants $0 < l' \leq L'$ such that, for all $j \in \underline{p}$,

$$l' \|h\|^2 < \langle h, G^j(z)h \rangle \leq L' \|h\|^2 , \quad \forall \; h, z \in \mathbb{R}^{l_j} , \tag{48a}$$

(iii) strict complementary slackness holds at all $\hat{x} \in \hat{G}$,[7]

(iv) Let $l \leq l'$ and $L \geq L'$ be such that

$$l\sigma^+[\sum_{j \in \underline{p}} \hat{\mu}^j A_j^T A_j] < \gamma < L \max_{j \in \underline{p}} \|Z^T A_j^T A_j Z\| , \tag{48b}$$

where $\hat{\mu}$ is the sole member of $U(\hat{G})$. Under these assumptions:

(a) For any $\hat{x} \in \hat{G}$,

[7]The assumption of strict complementary slackness is necessary only if the matrices A_j have different null spaces. For example, the linear convergence result holds without this assumption if the matrices A_j are identical.

$$\limsup_{\substack{x \to \hat{x} \\ x \in \hat{x} + Range(Z) \\ x \neq \hat{x}}} \min_{\lambda \in \mathbb{R}} \frac{\psi(x + \lambda h(x)) - \hat{\psi}}{\psi(x) - \hat{\psi}} \leq 1 - \frac{l}{L} \frac{\sigma^{+}[\sum_{j \in \underline{p}} \hat{\mu}^{j} A_j^T A_j]}{\max_{k \in \underline{p}} \| Z^T A_k^T A_k Z \|} . \tag{48c}$$

(b) If the PPP-ELS Algorithm constructs a sequence $\{x_i\}_{i=0}^{\infty}$ *in solving problem (42), then, the sequence converges to* $\hat{x}$ *for some* $\hat{x} \in \hat{G}$, *and either the sequence terminates in a finite number of steps at* $\hat{x}$ *or*

$$\limsup_{i \to \infty} \frac{\psi(x_{i+1}) - \hat{\psi}}{\psi(x_i) - \hat{\psi}} \leq 1 - \frac{l}{L} \frac{\sigma^{+}[\sum_{j \in \underline{p}} \hat{\mu}^{j} A_j^T A_j]}{\max_{k \in \underline{p}} \| Z^T A_k^T A_k Z \|} . \tag{48d}$$

Proof: (a) To prove this part, we will (i) show that it is sufficient consider the restriction of problem (42) to an affine space, (ii) verify that Hypotheses 3.1 and 3.2 hold for the restricted problem, and (iii) apply Lemma 3.4.

It follows from Proposition 5.2, that there exists a neighborhood $W \supset \hat{G}$ such that $\mu^j = 0$ for all $j \notin \hat{J}$ and $\mu \in U(W)$, and from (12), that $h(x) = \sum_{j \in \underline{p}} \mu^j A_j^T \nabla g^j(A_j x_j)$ for any $\mu \in U(x)$. Hence, for all $x \in W$,

$$h(x) \in Range(\hat{A}^T) = Range(Z) , \tag{49}$$

by the definition of Z above. Let us fix $\hat{x} \in \hat{G}$, and suppose that $x \in W$. If $x \in \hat{x} + Range(Z)$, then $x + \lambda\, h(x) \in \hat{x} + Range(Z)$. This suggests that we consider the restriction of problem (42) to $Range(Z)$, viz.,

$$\min_{y \in \mathbb{R}^a} \psi_r(y) , \tag{50a}$$

where

$$\psi_r(y) \triangleq \psi(\hat{x} + Zy) , \tag{50b}$$

so that $\psi_r(y) = \max_{j \in \underline{p}} f_r^j(y)$, with $f_r^j(y) \triangleq f^j(\hat{x} + Zy)$. The search direction $d(y)$ constructed by the PPP-ELS algorithm at a point $y \in \mathbb{R}^a$ for problem (50) is given by

$$d(y) \triangleq \arg \min_{d \in \mathbb{R}^a} \max_{j \in \underline{p}} g^j(A_j(x+Zy)) + \langle Z^T A_j^T \nabla g^j(A_j(x+Zy)), d\rangle + \tfrac{1}{2}\gamma|d|^2$$

$$= \arg \min_{d \in \mathbb{R}^a} \max_{j \in \underline{p}} g^j(A_j(x+Zy)) + \langle A_j^T \nabla g^j(A_j(x+Zy)), Zd\rangle + \tfrac{1}{2}\gamma|Zd|^2$$

$$= \arg \min_{d \in \mathbb{R}^a} \max_{j \in \underline{p}} \phi^j(Zd \mid x+Zy)\ , \tag{51}$$

since $Z^T Z = I_a$ and $\phi^j(h \mid x) \triangleq g^j(A_j x) + \langle A_j^T \nabla g^j(x), h\rangle + \tfrac{1}{2}\gamma|h|^2$. By (49), $h(\hat{x} + Zy) \in Range(Z)$. Hence, referring to (51), we see that

$$h(\hat{x}+Zy) = \arg \min_{h \in Range(Z)} \max_{j \in \underline{p}} \phi^j(h \mid x+Zy)$$

$$= Zd(y)\ . \tag{52}$$

Also, for y such that $\hat{x} + Zy \in W$,

$$\arg \min_{\lambda \in \mathbb{R}} \psi(\hat{x} + Z(y + \lambda d(y))) = \arg \min_{\lambda \in \mathbb{R}} \psi(\hat{x} + Zy + \lambda Zd(y))$$

$$= \arg \min_{\lambda \in \mathbb{R}} \psi(\hat{x} + Zy + \lambda h(\hat{x} + Zy))\ . \tag{53}$$

We conclude from (52) and (53) that

$$\limsup_{\substack{x \to \hat{x} \\ x \in \hat{x} + Range(Z) \\ x \neq \hat{x}}} \min_{\lambda \in \mathbb{R}} \frac{\psi(x + \lambda h(x)) - \hat{\psi}}{\psi(x) - \hat{\psi}} = \limsup_{\substack{y \to \hat{y} \\ y \neq \hat{y}}} \min_{\lambda \in \mathbb{R}} \frac{\psi(\hat{x} + Z(y + \lambda d(y)) - \hat{\psi}}{\psi(\hat{x} + Zy) - \hat{\psi}}\ . \tag{53b}$$

Hence, provided problem (50) satisfies Hypothesis 3.1 and Hypothesis 3.2, we can establish (48c) by applying Lemma 3.4 to problem (50) to obtain an upper bound on the right hand side of (53b).

We now verify that Hypothesis 3.1 is satisfied by problem (50). (i) The functions $g^j(A_j(\hat{x} + Zy))$ are twice continuously differentiable in y by assumption (i) of this theorem.

(ii) Let $S_r \triangleq \{ y \in \mathbb{R}^a \mid \psi(\hat{x} + Zy) \leq T \}$, with $T > \psi(\hat{G})$. Since the functions $g^j(\cdot)$ are uniformly convex by assumption (ii) of this theorem, the set $\hat{A}ZS_r$ is bounded. Since $Range(Z) = Range(\hat{A}^T)$ and $Null(Z) = \{ 0 \}$, $Null(\hat{A}Z) = \{ 0 \}$.[8] Hence S_r is bounded.

---

[8]Suppose that $\hat{A}Zy = 0$. Since $Null(\hat{A}) \cap Range(\hat{A}^T) = \{ 0 \}$, $Zy = 0$. But then $Null(Z) = \{ 0 \}$ implies that $y = 0$.

Showing that only one point in S_r satisfies the necessary conditions for optimality (13a, 13b) for problem (50) is slightly involved. Let $\hat{G}_r$ denote the minimizing set for problem (50). If $\hat{x} + Zy \in \hat{G}$, then $y \in \hat{G}_r$, and hence $\hat{G}_r \supset Z^T(\hat{G} - \hat{x}) \cap S_r$. Since $\hat{x} \in \hat{G}$, $0 \in \hat{G}_r$. Now suppose that there is a $y' \in \hat{G}_r$ such that $\hat{x} + Zy' \notin \hat{G}$. Then $\psi(\hat{x} + Zy') > \psi(\hat{x} + Z0)$, which contradicts the assumption that $y' \in \hat{G}_r$. Therefore, $\hat{G}_r = Z^T(\hat{G} - \hat{x}) \cap S_r$.

Now consider the set of multipliers,

$$U_r(y) \triangleq \left\{ \mu \in \Sigma_p \;\middle|\; \begin{array}{l} \sum_{j \in \underline{p}} \mu^j Z^T A_j^T \nabla g^j(A_j(\hat{x} + Zy)) = 0 \\ \sum_{j \in \underline{p}} \mu^j (g^j(A_j(\hat{x} + Zy)) - \psi(\hat{x} + Zy)) = 0 \end{array} \right\}, \tag{54}$$

which, together with y, satisfy the the optimality conditions (13a, 13b), when the functions functions $f^j(\cdot)$ are replaced by the functions $f_r^j(\cdot)$. For any $\bar{y} \in \hat{G}_r$, we have $\hat{x} + Z\bar{y} \in \hat{G}$, and hence $g^j(A_j(\hat{x} + Z\bar{y})) < \psi(\hat{x} + Z\bar{y})$ for all $j \notin \hat{J}$. Consequently, $\bar{\mu}^j = 0$ for all $j \notin \hat{J}$ and for any $\bar{\mu} \in U_r(\bar{y})$, and therefore

$$\sum_{j \in \underline{p}} \bar{\mu}^j A_j^T \nabla g^j(A_j(\hat{x} + Zy)) \in Range(\hat{A}^T) = Range(Z)\,. \tag{55}$$

For any $\bar{\mu} \in U_r(\bar{y})$, it follows from (54) that $\sum_{j \in \underline{p}} \bar{\mu}^j Z^T A_j^T \nabla g^j(A_j(\hat{x} + Zy)) = 0$, hence, making use of (55), we conclude that $\sum_{j \in \underline{p}} \bar{\mu}^j A_j^T \nabla g^j(A_j(\hat{x} + Z\bar{y})) = 0$. Hence, $\bar{\mu}$ together with $\hat{x} + Z\bar{y}$ satisfy the necessary conditions (13a, 13b) for the original problem (42). Thus, $U_r(\hat{G}_r) \subset U(\hat{x} + Z\hat{G}_r)$, and hence $U_r(\hat{G}_r) = \{ \hat{\mu} \}$, where $\hat{\mu}$ is the only member of $U(\hat{G})$.

Suppose that $y_1, y_2 \in S_r$ satisfy the optimality conditions (13a, 13b) for problem (50). Since $\psi_r(Zy)$ is convex in y, these necessary conditions are sufficient for optimality, and, furthermore, the entire line segment between y_1 and y_2, $[y_1, y_2]$, lies in $\hat{G}_r$. Since $U_r([y_1, y_2]) = \{ \hat{\mu} \}$ and $\hat{\mu}^j > 0$ for all $j \in \hat{J}$, $g^j(A_j(\hat{x} + Zy)) = \psi(\hat{x} + Zy) = \hat{\psi}$ for all $y \in [y_1, y_2]$ and all $j \in \hat{J}$. Because the functions $g^j(\cdot)$ are strictly convex, it follows that $A_j Z y_1 = A_j Z y_2$ for all $j \in \hat{J}$, and hence that $y_1 - y_2 \in Null(\hat{A}Z)$. As mentioned above, $Null(\hat{A}Z) = \{ 0 \}$, implying that $y_1 = y_2$. Therefore, the necessary conditions are satisfied at a unique point $\hat{y} \in S_r$ and Hypothesis 3.1(ii) holds.

(iii) It follows from Assumption (ii) of this theorem that for all $y \in \mathbb{R}^a$, $|F^j_r(y)| \le L \max_{j \in \underline{p}} |Z^T A_j^T A_j Z|$, where $F^j_r(\cdot)$ denotes the second derivative matrix of $f^j_r(\cdot)$.

Now we verify that Hypothesis 3.2 holds. Letting $\underline{\sigma}[X]$ denote the minimum eigenvalue value of any real symmetric matrix X, we obtain that

$$\underline{\sigma}\left[\sum_{j \in \underline{p}} \hat{\mu}^j G^j(A_j(\hat{x} + Z\hat{y}))\right] = \underline{\sigma}\left[\sum_{j \in \underline{p}} \hat{\mu}^j Z^T A_j^T G^j(A_j(\hat{x} + Z\hat{y}))A_j Z\right]$$

$$\ge l\, \underline{\sigma}\left[\sum_{j \in \underline{p}} \hat{\mu}^j Z^T A_j^T A_j Z\right]$$

$$= l\, \sigma^+\left[\sum_{j \in \underline{p}} \hat{\mu}^j A_j^T A_j\right], \tag{56}$$

since the columns of Z span $Range(\hat{A}^T) = Null(\hat{A})^\perp$ and $Null(\sum_{j \in \underline{p}} \hat{\mu}^j A_j^T A_j) = Null(\hat{A})$.

Assumption (iv) of this theorem ensures that

$$l\, \sigma^+\left[\sum_{j \in \underline{p}} \hat{\mu}^j A_j^T A_j\right] < \gamma < L \max_{j \in \underline{p}} |Z^T A_j^T A_j Z|. \tag{57}$$

Letting the left-hand and right-hand sides of the double inequality (57) correspond to m and M respectively, we can apply Lemma 3.4 to problem (50) to obtain

$$\limsup_{\substack{y \to \hat{y} \\ y \ne \hat{y}}} \min_{\lambda \in \mathbb{R}} \frac{\psi(\hat{x} + Z(y + \lambda d(y))) - \hat{\psi}}{\psi(\hat{x} + Zy) - \hat{\psi}}$$

$$\le 1 - \frac{l}{L} \frac{\sigma^+[\sum_{j \in \underline{p}} \hat{\mu}^j A_j^T A_j]}{\max_{k \in \underline{p}} |Z^T A_k^T A_k Z|}, \tag{58}$$

which, combined with (53b), gives part (a).

To show (b), we first show that $x_i \to \hat{x}$ as $i \to \infty$ for some $\hat{x} \in \hat{G}$, and then apply part (a) of this theorem. Let $\bar{A}^T \triangleq [A_1^T, \dots, A_p^T]$. From (12), every h_i, constructed by the PPP-ELS algorithm, is of the form $\sum_{j \in \underline{p}} A_j^T z_j$, with $z_j \in \mathbf{R}^{l_j}$. Thus, the sequence $\{ x_i \}_{i=0}^{\infty}$ is contained in the closed and convex set

$Q \triangleq \{ x_0 + Range(\bar{A}^T) \} \cap \{ x \in \mathbb{R}^n \mid \psi(x) \le \psi(x_0) \}$. Suppose that Q is unbounded. Then, since Q is convex, there exists a nonzero $u \in Range(\bar{A}^T)$, such that, with $x_t \triangleq x_0 + tu$, $\psi(x_t) \le \psi(x_0)$ for all $t \ge 0$. If $A_{j_0}u \ne 0$ for some $j_0 \in \underline{p}$, then the uniform convexity of $g^j(\cdot)$, which follows from assumption (ii) of this theorem, implies that $\lim_{t \to \infty} \psi(x_t) = +\infty$. Since this contradicts our assumption that $\psi(x_t) \le \psi(x_0)$, we must have that $\bar{A}u = 0$. Hence $u \in Range(\bar{A}^T) \cap Null(\bar{A}) = \{ 0 \}$, which contradicts the assumption that $u \ne 0$. Therefore, the set Q is bounded, and hence compact. Consequently, the sequence $\{ x_i \}_{i=0}^{\infty}$ must have an accumulation point, $\hat{x}$. From Corollary 5.1 and Proposition 5.5 in Ref. 13, any accumulation point $\hat{x}$ of a sequence generated by the PPP-ELS algorithm must satisfy the first-order necessary conditions for optimality (13a, 13b). Since $\psi(\cdot)$ is convex, this implies that $\hat{x} \in \hat{G}$. Since Q is compact, it follows that $x_i \to \hat{G}$ as $i \to \infty$.

Since $x_i \to \hat{G}$ as $i \to \infty$, there exists $i_0 \in \mathbb{N}$ such that $x_i \in W$ for all $i > i_0$. Hence, $\{ x_i \}_{i=i_0}^{\infty} \subset x_{i_0} + Range(\hat{A}^T) = x_{i_0} + Range(Z)$. Since the functions $g^j(\cdot)$ are uniformly convex, $\hat{G} \cap (x_{i_0} + Range(\hat{A}^T)$ is a singleton. Hence, the sequence $\{ x_i \}_{i=0}^{\infty}$ converges to $\hat{x} = \hat{G} \cap x_{i_0} + Range(\hat{A}^T)$.

Inequality (48d) follows directly from convergence of the sequence to $\hat{x}$ and part (a). ■

The corresponding result for the PPP-Armijo algorithm can be obtained by following the steps used in Section 4 and above.

Theorem 5.2: *Suppose that the assumptions of Theorem 5.1 hold.*

(a) For any $\hat{x} \in \hat{G}$,

$$\limsup_{\substack{x \to \hat{x} \\ x \in \hat{x} + Range(Z) \\ x \ne \hat{x}}} \min_{\lambda \in \mathbb{R}} \frac{\psi(x + \lambda h(x)) - \hat{\psi}}{\psi(x) - \hat{\psi}} \le 1 - \alpha\beta \frac{l}{L} \frac{\sigma^+[\sum_{j \in \underline{p}} \hat{\mu}^j A_j^T A_j]}{\max_{k \in \underline{p}} \| Z^T A_k^T A_k Z \|} . \tag{59a}$$

(b) If the PPP-Armijo algorithm constructs a sequence $\{x_i\}_{i=0}^{\infty}$ *in solving problem (42), then, the sequence converges to* $\hat{x}$ *for some* $\hat{x} \in \hat{G}$, *and either the sequence terminates in a finite number of steps*

at $\hat{x}$ *or*

$$\limsup_{i \to \infty} \frac{\psi(x_{i+1}) - \hat{\psi}}{\psi(x_i) - \hat{\psi}} \le 1 - \alpha\beta \frac{l}{L} \frac{\sigma^+[\sum_{j \in \underline{p}} \hat{\mu}^j A_j^T A_j]}{\max_{k \in \underline{p}} \| Z^T A_k^T A_k Z \|} \,. \tag{59b}$$

6. Conclusion

We have shown that sequences $\{ \psi(x_i) \}_{i=0}^{\infty}$ generated by two PPP minimax algorithms converge linearly to the minimum value under weaker conditions than those assumed in previous analyses of the rate of convergence of PPP algorithms (Refs. 1, 7-9). Although composite minimax problems which have nonunique, nonisolated minimizers, do not satisfy the second-order sufficiency conditions that we had to assume to establish linear convergence on *general* minimax problems, we were able to show that these PPP algorithms converge linearly on these problems provided that strong convexity and strict complementary slackness conditions are satisfied.

PPP algorithms can be generalized in a straightforward way to solve *semi-infinite* composite minimax problems (Ref.13) which arise in control system design,

$$\min_{x \in \mathbb{R}^n} \max_{j \in \underline{p}} \max_{y_j \in Y_j} \phi^j(A_j x, y_j) \,, \tag{60}$$

where the sets $Y_j \subset \mathbb{R}^{s_j}$ are compact, and the functions $\phi^j : \mathbb{R}^{l_j} \times \mathbb{R}^{s_j} \to \mathbb{R}$, $j \in \underline{p}$ and $\nabla_1 \phi^j(\cdot,\cdot)$ are continuous. As before, each A_j is an $l_j \times n$ matrix. Under assumptions analogous to those of Theorem 5.1, it can be shown that the semi-infinite versions of the PPP algorithms, considered in this paper, also converge linearly (see Ref. 8).

7. Appendix

Proof of Lemma 3.1: For any $y \in \mathbb{R}^n$, $y = Py + P^{\perp}y$. Hence, since $R(\cdot,\cdot)$ is continuous and $R(0,\hat{\mu})$ is negative definite for any $\hat{\mu} \in U(\hat{x})$ by Hypothesis 3.2,

$$\begin{aligned}\langle y, R(y,\mu)y \rangle &= \langle Py + P^{\perp}y, R(y,\mu)(Py + P^{\perp}y) \rangle \\ &= \langle P^{\perp}y, R(y,\mu)P^{\perp}y \rangle + \langle Py, R(y,\mu)(Py + 2P^{\perp}y) \rangle\end{aligned}$$

$$\le \langle Py, R(y,\mu)(Py + 2P^{\perp}y)\rangle , \tag{61a}$$

for μ near $U(\hat{x})$ and y small. Using the Schwarz inequality and the fact that $\|Py + 2P^{\perp}y\| \le 2\|y\|$,

$$\langle y, R(y,\mu)y\rangle \le \|R(y,\mu)\| \, \|Py\| \, \|Py + 2P^{\perp}y\|$$

$$\le 2\|R(y,\mu)\| \, \|Py\| \, \|y\|$$

$$\le 3 \max_{\mu \in U(\hat{x})} \|R(0,\hat{\mu})\| \, \|Py\| \, \|y\| , \tag{61b}$$

for μ near $U(\hat{x})$ and y small, since $\|R(\cdot,\cdot)\|$ is continuous. ■

Proof of Lemma 3.2: Using Taylor's Theorem, we obtain that for any $x \in \mathbb{R}^n$,

$$\psi(x) - \psi(\hat{x}) \ge \max_{j \in J(\hat{x})} f^j(x) - \psi(\hat{x})$$

$$= \max_{j \in J(\hat{x})} f^j(\hat{x}) + \langle \nabla f^j(\hat{x}), x - \hat{x} \rangle + \langle x - \hat{x}, \left[\int_0^1 (1-s)F^j(\hat{x} + s(x - \hat{x}))ds\right](x - \hat{x}) \rangle - \psi(\hat{x}) . \tag{62}$$

Since $f^j(\hat{x}) = \psi(\hat{x})$ for all $j \in J(\hat{x})$, it follows from Hypothesis 3.1(iii) that

$$\psi(x) - \psi(\hat{x}) \ge \max_{j \in J(\hat{x})} \langle \nabla f^j(\hat{x}), x - \hat{x} \rangle + \langle x - \hat{x}, \left[\int_0^1 (1-s)F^j(\hat{x} + s(x - \hat{x}))ds\right](x - \hat{x}) \rangle$$

$$\ge \max_{j \in J(\hat{x})} \langle \nabla f^j(\hat{x}), x - \hat{x} \rangle - M\|x - \hat{x}\|^2 . \tag{63}$$

Since $\langle \nabla f^j(\hat{x}), P^{\perp}(x - \hat{x})\rangle = 0$ for all $j \in J(\hat{x})$ and $h = Ph + P^{\perp}h$ for any $h \in \mathbb{R}^n$,

$$\max_{j \in J(\hat{x})} \langle \nabla f^j(\hat{x}), x - \hat{x}\rangle = \max_{j \in J(\hat{x})} \langle \nabla f^j(\hat{x}), P(x - \hat{x})\rangle . \tag{64}$$

We will to show that there exists an $\eta > 0$ such that

$$\max_{j \in J(\hat{x})} \langle \nabla f^j(\hat{x}), P(x - \hat{x})\rangle \ge \eta \|P(x - \hat{x})\| . \tag{65}$$

Suppose not. Then, there exists a nonzero $\bar{u} \in B$ such that $\max_{j \in J(\hat{x})} \langle \nabla f^j(\hat{x}), \bar{u} \rangle \le 0$. Since $U(\hat{x})$ is convex, there exists a $\hat{\mu} \in U(\hat{x})$ such that $\hat{\mu}^j > 0$ for all $j \in J(\hat{x})$. By (13b), $\hat{\mu}^j = 0$ for $j \notin J(\hat{x})$. Therefore, by (13a),

$$\sum_{j \in J(\hat{x})} \hat{\mu}^j \langle \nabla f^j(\hat{x}), \bar{u} \rangle = \langle \sum_{j \in \underline{p}} \hat{\mu}^j \nabla f^j(\hat{x}), \bar{u} \rangle = \langle 0, \bar{u} \rangle = 0 \,, \tag{66}$$

Equation (66) states that a convex combination of the nonpositive numbers, $\{ \langle \nabla f^j(\hat{x}), \bar{u} \rangle \}_{j \in J(\hat{x})}$, with nonzero coefficients, $\{ \hat{\mu}^j \}_{j \in J(\hat{x})}$, is zero. Hence $\langle \nabla f^j(\hat{x}), \bar{u} \rangle = 0$ for all $j \in J(\hat{x})$. But then $\bar{u} \in B \cap B^{\perp} = \{ 0 \}$, contradicting the assumption that $\bar{u} \neq 0$. Hence, let $\eta > 0$ be such that (65) holds. Substituting (64) into (65) and (65) into (63) yields

$$\psi(x) - \psi(\hat{x}) \geq \eta \| P(x - \hat{x}) \| - M \| x - \hat{x} \|^2 \,, \tag{67}$$

for x in some neighborhood of $\hat{x}$.

Now we derive *another* lower bound on $\psi(x) - \psi(\hat{x})$. For any $\hat{\mu} \in U(\hat{x})$, using Taylor's Theorem and the fact that $\sum_{j \in \underline{p}} \hat{\mu}^j \nabla f^j(\hat{x}) = 0$,

$$\begin{aligned}
\psi(x) - \psi(\hat{x}) &\geq \sum_{j \in \underline{p}} \hat{\mu}^j f^j(x) - \psi(\hat{x}) \\
&= \langle x - \hat{x}, \left[\int_0^1 (1 - s) \sum_{j \in \underline{p}} \hat{\mu}^j F^j(\hat{x} + s(x - \hat{x})) ds \right] (x - \hat{x}) \rangle \\
&= \langle P^{\perp}(x - \hat{x}), \left[\int_0^1 (1 - s) \sum_{j \in \underline{p}} \hat{\mu}^j F^j(\hat{x} + s(x - \hat{x})) ds \right] P^{\perp}(x - \hat{x}) \rangle \\
&\quad + \langle P(x - \hat{x}), \left[\int_0^1 (1 - s) \sum_{j \in \underline{p}} \hat{\mu}^j F^j(\hat{x} + s(x - \hat{x})) ds \right] (2P^{\perp}(x - \hat{x}) + P(x - \hat{x})) \rangle .
\end{aligned} \tag{68}$$

Making use of Hypothesis 3.1(iii) and Hypothesis 3.2, (68) leads to

$$\begin{aligned}
\psi(x) - \psi(\hat{x}) &\geq \tfrac{1}{2} m \| P^{\perp}(x - \hat{x}) \|^2 - \tfrac{1}{2} M \| P(x - \hat{x}) \| \, \| 2P^{\perp}(x - \hat{x}) + P(x - \hat{x}) \| \\
&\geq \tfrac{1}{2} m \| P^{\perp}(x - \hat{x}) \|^2 - M \| P(x - \hat{x}) \| \, \| x - \hat{x} \| \,,
\end{aligned} \tag{69}$$

for x in a neighborhood of $\hat{x}$.

Combining (67) with (69) and dividing by $\| P(x - \hat{x}) \| \| x - \hat{x} \|$ yields

$$\frac{\psi(x)-\psi(\hat{x})}{\|P(x-\hat{x})\|\|x-\hat{x}\|} \geq \max\left\{\frac{\frac{1}{2}m\|P^{\perp}(x-\hat{x})\|^2}{\|P(x-\hat{x})\|\|x-\hat{x}\|} - M\,,\ \frac{\eta}{\|x-\hat{x}\|} - M\frac{\|x-\hat{x}\|}{\|P(x-\hat{x})\|}\right\}, \tag{70}$$

for x in a neighborhood of $\hat{x}$. Using the fact that $\|x\| \leq \|Px\| + \|P^{\perp}x\|$, and defining $r(x) = \|P(x-\hat{x})\| \,/\, \|P^{\perp}(x-\hat{x})\|$, we obtain that

$$\frac{\psi(x)-\psi(\hat{x})}{\|P(x-\hat{x})\|\|x-\hat{x}\|} \geq \max\left\{\frac{\frac{1}{2}m}{r(x)^2 + r(x)} - M\,,\ \frac{\eta}{\|x-\hat{x}\|} - M(\frac{1}{r(x)} + 1)\right\}, \tag{71}$$

We use (71) to show that

$$\liminf_{x \to \hat{x}} \frac{\psi(x)-\psi(\hat{x})}{\|P(x-\hat{x})\|\,\|x-\hat{x}\|} = \infty\,, \tag{72}$$

which is equivalent to (19). Thus, given any integer $k > 0$, there exists a real number $r > 0$ such that the first term in the max in (71) is greater than k if $r(x) \leq r$. For x such that $r(x) > r$, the second term term in the max is greater than $\eta/\|x-\hat{x}\| - M(1/r + 1)$. Hence, there exists a neighborhood, W_k, of $\hat{x}$ such that the max in (71) exceeds k for all $x \in W_k$, and, therefore, (19) holds. ■

8. References

1. PSHENICHNYI, B. N. and DANILIN, YU. M., *Numerical Methods in Extremal Problems*, Nauka, Moscow, 1975.
2. POLYAK, B. T., *Introduction to Optimization*, (in Russian), Nauka, Moscow (1983).
3. SHOR, N. Z., *Minimization Methods for Non-Differentiable Functions*, Springer-Verlag, Berlin, 1985.
4. KIWIEL, K. C, *A Phase I - Phase II Method for Inequality Constrained Minimax Problems*, Control and Cybernetics, Vol. 12, pp. 55-75, 1983.
5. PEVNYI, A. B., *The Convergence Speed of Some Methods of Finding Minimaxes and Saddle Points*, Cybernetics, (A translation of Kibernetika) Vol. 8, No. 4, pp. 635-640, 1972.
6. MADSEN, K. and SCHJAER-JACOBSEN, H., *Linearly Constrained Minimax Optimization*, Mathematical Programming, Vol. 14, pp. 208-223, 1978.
7. DAUGAVET, V. A. and MALOZEMOV, V. N., *Quadratic Rate of Convergence of a Linearization Method for Solving Discrete Minimax Problems*, USSR Comput. Maths. Phys., Vol. 21, No. 4, pp. 19-28, 1981.
8. POLAK, E. and WIEST, E. J., *Domain Rescaling Techniques for the Solution of Affinely Parametrized Nondifferentiable Optimal Design Problems*, Proceedings of the 27th IEEE Conference on Decision and Control, Austin, Texas, 1988.
9. POLAK, E. P., *Basics of Minimax Algorithms*, Proc., Summer School on Nonsmooth Optimization and Related Topics, Erice, Italy, June 19 - July 8 1988.

10. PIRONNEAU, O. and POLAK, E., *On the Rate of Convergence of Certain Methods of Centers*, Mathematical Programming, Vol. 2, No. 2, pp. 230-258, 1972.

11. HUARD, P., *Method of Centers*, Nonlinear Programming, J. Abadie, ed., North-Holland Publishing Company, Amsterdam, Holland, 1967.

12. HUARD, P., *Programoation Mathematique Convexe*, Revue Francaise d'Informatique et de Recherche Operationnelle, Vol. 7, pp. 43-59, 1968.

13. POLAK, E., *On the Mathematical Foundations of Nondifferentiable Optimization in Engineering Design*, SIAM Review, Vol. 29, No.1, pp. 21-91, March 1987.

14. CHANEY, R. W., *On the Pironneau-Polak Method of Centers*, J. Opt. Theory Appl., Vol. 20, No. 3, pp. 269-295, 1976.

15. DEMYANOV, V. F., *On the Solution of Certain Minimax Problems I, II*, Kibernetika, No. 6, 1966.

16. DEMYANOV, V. F., *Algorithms for Some Minimax Problems*, Journal of Computer System Sciences, Nos. 2, 4, pp. 342-380, 1968.

17. ZOUTENDIJK, G., *Mathematical Programming Methods*, North Holland, Amsterdam, 1976.

18. WOMERSLEY, R. S., *Optimality Conditions for Piecewise Smooth Functions*, Mathematical Programming Study 17, pp. 13-27, 1982.

19. ARMIJO, L., *Minimization of Functions Having Continuous Partial Derivatives*, Pacific Journal of Mathematics, Vol. 16, pp. 1-3, 1966

20. BERGE, C., *Topological Spaces*, Macmillan, New York, New York, 1963.

21. GILL, P. E., HAMMARLING, J., MURRAY, W., SAUNDERS, M. A. and WRIGHT, M. H., *User's Guide for LSSOL (Version 1.0): A Fortran Package for Constrained Linear Least-squares and Convex Quadratic Programming*, Technical report SOL 86-1, Department of Operations Research, Stanford University, Stanford, January 1986.

22. VON HOHENBALKEN, B., *Simplical Decomposition in Nonlinear Programming Algorithms*, Mathematical Programming, Vol. 13, pp. 49-68, 1977

23. KIWIEL, K. C., *A Method for Solving Certain Quadratic Programming Problems Arising in Nonsmooth Optimization*, IMA Journal of Numerical Analysis, Vol. 6, pp. 137-152, 1986.

24. KIWIEL, K. C., *A Dual Method for Solving Certain Positive Semi-Definite Quadratic Programming Problems*, SIAM Journal of Scientific and Statistical Computing, Vol. 10, 1989.

25. RUSCZYNSKI, A., *A Regularized Decomposition Method for Minimizing a Sum of Polyhedral Functions*, Mathematical Programming, Vol. 35, pp. 46-61, 1986.

26. LUENBERGER, D. G., *Introduction to Linear and Nonlinear Programming*, Addison-Wesley, Reading, Massachusetts, 1973.