ON THE LINEAR CONVERGENCE OF THE PSHENICHNYI METHOD OF LINEARIZATIONS

by

E. J. Wiest and E. Polak

Memorandum No. UCB/ERL M89/55

17 May 1989

ON THE LINEAR CONVERGENCE OF THE PSHENICHNYI METHOD OF LINEARIZATIONS

by

E. J. Wiest and E. Polak

Memorandum No. UCB/ERL M89/55

17 May 1989

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

ON THE LINEAR CONVERGENCE OF THE PSHENICHNYI METHOD OF LINEARIZATIONS

by

E. J. Wiest and E. Polak

Memorandum No. UCB/ERL M89/55

17 May 1989

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

On the Linear Convergence of the Pshenichnyi Method of Linearizations[1]

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

Abstract. It is shown that the sequence of function values constructed by the Pshenichnyi linearization method in solving a minimax problem converges linearly to the minimum. The proof does not assume strict complementary slackness, and applies to a version of the method employing a practical step size rule, which can be executed in a finite number of steps. The result is extended to minimax problems in which each function appearing in the max is composed with a different linear function.

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

We are concerned with the rate of convergence of algorithms for solving the minimax problem,

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

where the functions $f^j:\mathbb{R}^n \to \mathbb{R}$ are continuously differentiable and where $\underline{p}$ denotes the set of integers $\{ 1 , \ldots , p \}$. We denote the maximum function by $\psi(x) \triangleq \max_{j \in \underline{p}} f^j(x)$. Rates of convergence have been established for several minimax algorithms. Shor has shown that a sequence of iterates $\{ x_i \}_{i=0}^{\infty}$ constructed by a subgradient method on a convex problem converges linearly to a minimizer $\hat{x}$ (Ref. 1). In Ref. 2, Kiwiel has proved that sequences generated by a bundle algorithm also converge linearly on constrained minimax problems. Pevnyi has shown that a sequence of values $\{ \psi(x_i) \}_{i=0}^{\infty}$ constructed by the ε-active set method of Dem'yanov converges linearly to a value less than $\psi(\hat{x}) + \varepsilon$ (Ref. 3). These methods use only first order information. If every subset of $\{ \nabla f^j(\hat{x}) \mid f^j(\hat{x}) = \psi(\hat{x}) \}$ has maximal rank, then $\hat{x}$ is said to be a *Haar point*. The sequential linear programming algorithm of Ref. 4 was shown to converge quadratically to Haar points. Superlinear rates of convergence to minimizers meeting less stringent requirements have been obtained for algorithms using second order information, see, for example, Refs. 5-8. The algorithm in Ref. 7 switches between a first-order trust region method, in which a linear program is solved to obtain the search direction, and a second-order active set approach, in which the search direction is obtained from a quasi-Newton iteration on the necessary conditions for optimality for the minimax problem. Convergence was shown to be superlinear in the quasi-Newton phase, and was shown to be quadratic in the sequential linear programming phase, if the sequence converges to a Haar point.

The minimax problem (1) can be converted into a nonlinear program. An extra variable x_0 representing the value of $\psi(\cdot)$ is chosen as the cost, which is subject to the differentiable constraints, $f^j(x) \le x_0$ for $j \in \underline{p}$. A number of minimax algorithms are based on this fact (including those in Refs. 5-9). The obstacle to achieving superlinear convergence by applying standard nonlinear programming techniques to this formulation of problem (1) is the practical insistence on a decrease in the value of $\psi(\cdot)$ at each iteration. This translates into the requirements that the value of the cost of the nonlinear program decrease at each iteration and that the iterates remain feasible. Standard nonlinear programming techniques (see, for example, Refs. 10 and 11) do not satisfy these requirements, but a recently developed algorithm does (Ref. 12), by solving a corrected subproblem (Ref. 13) and using a curvilinear step size (Ref. 14).

In most of the convergence rate theorems cited, the restrictions imposed on the problems and on the choice of step size are unrealistic. Ref. 4 makes the very strong assumption that the sequence of iterates converges to a Haar point. Non-degeneracy assumptions are also made in Refs. 5, 7, 8 and 12. Superlinear convergence of the algorithm in Refs. 5 and 8 require that a step size of one be chosen in the tail of any sequence constructed. Except in the neighborhood of a Haar point (Ref. 15), a step size of one does not always reduce $\psi(\cdot)$. However, in practice and in convergence theorems, step size rules which guarantee a decrease in $\psi(\cdot)$ are generally required.[1] The step size rule used to prove linear convergence of an algorithm in Ref. 1 does not satisfy the assumptions of the convergence theorem proved for the algorithm.

Of the results cited, only that of Pevnyi (Ref. 3) regarding the Dem'yanov algorithm specifies a realistic step size rule for which a strong global convergence result obtains and which does not make overly restrictive assumptions on the problem, such as convergence to a Haar point or strict complementary slackness. Pevnyi's result assumes that the Hessians of all of the functions maximal at $\hat{x}$ are positive definite.

In this paper, we discuss a variant of the Pshenichnyi linearization method (Ref. 16) for solving the minimax problem (1). The method obtains a direction of descent for $\psi(\cdot)$ by solving an approximation to (1) in which the functions $f^j(x+h)$ are replaced by *first-order, quadratic* approximations,

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

for fixed $\gamma_j > 0$. The search direction is chosen to be the unique solution to the problem $\min_{h \in \mathbb{R}^n} \max_{j \in I_\delta(x)} \phi^j(h \mid x)$, where

[1] The algorithm in Ref. 7 allows $\psi(\cdot)$ to *increase* during the quasi-Newton phase, but a rather weak convergence result is obtained as a result.

$$I_\delta(x) \triangleq \{ j \in \underline{p} \mid f^j(x) \geq \psi(x) - \delta \} \ , \tag{3}$$

for some $\delta > 0$. In Ref. 16, it is assumed that $\gamma_j = 1$ for every $j \in \underline{p}$, and it is shown that a sequence $\{ x_i \}_{i=0}^{\infty}$ constructed by the algorithm will converge linearly to the solution $\hat{x}$ of problem (1). In Refs. 15 and 16, it is shown that, if the iterates constructed by the Pshenichnyi algorithm using an Armijo-type step size rule converge to Haar point, then the convergence is quadratic.

However, rate of convergence results for the Pshenichnyi algorithm suffer from some of the deficits described above. The linear convergence result in Ref. 16 is obtained for step sizes lying below a threshold which is not generally known *a priori*. A non-degeneracy assumption is also made. In Section 3, we show that the sequence of values $\{ \psi(x_i) \}_{i=0}^{\infty}$ generated by the Pshenichnyi algorithm converges linearly to the minimum value. Our result is obtained for versions of the algorithm employing an exact minimizing line search and, in Section 4, an implementable Armijo-type step size rule. These step size rules can be used in practice, unlike the unknown step size of Pshenichnyi's result and the unity step size of other results. In these sections, we make no assumption of non-degeneracy. Finally, our result requires that the Hessian of the Lagrangian function at $\hat{x}$ be positive definite only on the subspace orthogonal to the span of the gradients of functions maximal at $\hat{x}$.

In Section 5, we extend these results to the *composite* minimax problem,

$$\min_{x \in \mathbb{R}^n} \max_{j \in \underline{p}} g^j(A_j x) \ , \tag{4}$$

in which each continuously differentiable function $g^j : \mathbb{R}^{l_j} \rightarrow \mathbb{R}$ is composed with a different linear function $A_j : \mathbb{R}^n \rightarrow \mathbb{R}^{l_j}$.

2. The Pshenichnyi Algorithm

We begin by stating a version of the Pshenichnyi algorithm which uses an exact minimizing line search, and a convergence result.

Algorithm 2.1 : (see Algorithm 5.2 and Corollary 5.1 in Ref.17)

Data: $x_0 \in \mathbb{R}^n$; $\gamma_j > 0$, $\forall\, j \in \underline{p}$; $i = 0$.

Step 1: Compute the search direction,

$$h(x_i) \triangleq \arg \min_{h \in \mathbb{R}^n} \max_{j \in \underline{p}} \phi^j(h \mid x_i) . \tag{5}$$

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. ■

When $\gamma_j = \gamma$ for all $j \in \underline{p}$, the search direction finding problem,

$$\theta(x) \triangleq \min_{h \in R^n} \max_{j \in \underline{p}} \phi(h \mid x) - \psi(x) \ , \tag{6}$$

can be solved by converting it to dual form as follows. Let the standard unit simplex be denoted by $\Sigma_p \triangleq \{ \mu \in \mathbb{R}^p \mid \mu^j \geq 0 \ , \ \sum_{j \in \underline{p}} \mu^j = 1 \}$. We replace the max over $j \in \underline{p}$ by a max over the standard unit simplex,

$$\theta(x) = \min_{h \in R^n} \max_{\mu \in \Sigma_p} \sum_{j \in \underline{p}} \mu^j \phi(h \mid x) - \psi(x) . \tag{7}$$

By an extension to von Neumann's Minimax Theorem (Ref. 17), the max and min in (8) can be interchanged,

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

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

$$= \max_{\mu \in \Sigma_p} \sum_{j \in \underline{p}} \mu^j \left(f^j(x) - \psi(x) \right) - \frac{1}{2\gamma} \| \sum_{j \in \underline{p}} \mu^j \nabla f^j(x) \|^2 \ , \tag{8}$$

by solving the inner minimization problem. Several methods exist for solving this positive semi-definite quadratic program (see, for example, Refs. 18-22). The solution is not always unique, and we denote the solution set by

$$U(x) \triangleq \arg \max_{\mu \in \Sigma_p} \sum_{j \in \underline{p}} \mu^j \left(f^j(x) - \psi(x) \right) - \frac{1}{2\gamma} \| \sum_{j \in \underline{p}} \mu^j \nabla f^j(x) \|^2 . \tag{9}$$

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

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

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

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

$$= \min_{h \in \mathbb{R}^n} \sum_{j \in \underline{p}} \bar{\mu}^j \phi^j(h \mid x) \,. \tag{10}$$

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

$$h(x) = \arg \min_{h \in \mathbb{R}^n} \sum_{j \in \underline{p}} \bar{\mu}^j \phi^j(h \mid x) = -\frac{1}{\gamma} \sum_{j \in \underline{p}} \bar{\mu}^j \nabla f^j(x) \;, \tag{11}$$

to the primal problem (8), which is unique since the function $\max_{j \in \underline{p}} \phi^j(\cdot \mid x)$ is strictly convex. Note that when $p = 1$, $\psi(x) = f^1(x)$ and $h(x)$ reduces to the search direction employed by the method of steepest descent.

The following necessary condition for optimality can be found in Ref. 17.

Theorem 2.1: (Ref. 17) *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*

$$\sum_{j \in \underline{p}} \hat{\mu}^j \nabla f^j(\hat{x}) = 0 \;, \tag{12a}$$

$$\sum_{j \in \underline{p}} \hat{\mu}^j [f^j(\hat{x}) - \psi(\hat{x})] = 0 \,. \tag{12b}$$

■

If the functions $f^j(\cdot)$ are convex, equations (12a, 12b) are a sufficient condition for optimality. We denote the minimum value for problem (1) and the set of minimizers by $\hat{\psi} \triangleq \min_{x \in \mathbb{R}^n} \psi(x)$ and $\hat{G} \triangleq \arg\min_{x \in \mathbb{R}^n} \psi(x)$, respectively. For any $\hat{x} \in \hat{G}$, the set of multiplier vectors $\hat{\mu} \in \Sigma_p$ which satisfy equations (12a, 12b) together with $\hat{x}$ is exactly $U(\hat{x})$.

Theorem 2.2: (Ref.17) *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 (12a, 12b).* ■

By analogy with nonlinear programming, we shall say that *strict complementary slackness* holds at a solution $\hat{x}$ if, for every multiplier vector $\hat{\mu}$ satisfying (12a, 12b) with $\hat{x}$,

$$\hat{\mu}^j > 0 \text{ if and only if } f^j(\hat{x}) = \psi(\hat{x}) \ . \tag{13}$$

If strict complementary slackness holds at a minimizer $\hat{x}$, the multiplier vector $\hat{\mu}$ satisfying (12a, 12b) with $\hat{x}$ is unique. For any $\hat{x} \in \hat{G}$, we define

$$J(\hat{x}) \triangleq \{ j \in \underline{p} \mid \exists \mu \in U(\hat{x}) : \mu^j > 0 \} \ . \tag{14}$$

If strict complementary slackness holds at $\hat{x}$, $J(\hat{x}) = \{ j \in \underline{p} \mid f^j(\hat{x}) = \psi(\hat{x}) \}$.

3. Linear Convergence of the Pshenichnyi Algorithm

We restate the Pshenichnyi convergence rate theorem in Ref.16. Let $F^j(\cdot)$ denote the second derivative matrix of $f^j(\cdot)$.

Theorem 3.1: (Ref. 16) *Suppose that*

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

(ii) $\hat{x}$ *is a local minimizer of* $\psi(\cdot)$,

(iii) the gradients $\{ \nabla f^j(\hat{x}) \}_{j \in J(\hat{x})}$ *are affinely independent,*[2]

(iv) strict complementary slackness holds at $\hat{x}$,

(v) the second derivative matrix of the Lagrangian at $(\hat{x}, \hat{\mu})$, $\sum_{j \in \underline{p}} \hat{\mu}^j F^j(\hat{x})$, *is positive definite.*

Then there exists a neighborhood W of $\hat{x}$ *such that the process*

$$x_{k+1} = x_k + \alpha h(x_k) \ , \ k \in \mathbb{N} \ , \tag{15}$$

converges linearly to $\hat{x}$, *provided that* $x_1 \in W$ *and* $\alpha > 0$ *and* $\delta > 0$ *are sufficiently small.*

If, in addition, the affine hull of $\{ \nabla f^j(\hat{x}) \}_{j \in J(\hat{x})}$ *contains* $\mathbb{R}^n$, $\{ x_i \}_{i=0}^{\infty}$ *converges to* $\hat{x}$ *quadratically.* ■

We now prove that the sequence of values, $\{ \psi(x_i) \}_{i=0}^{\infty}$, converges linearly to the minimum value under weaker assumptions than those of Theorem 3.1. In particular, our theorem omits assumptions (iii) and (iv) of Theorem 3.1 and relaxs the convexity requirement of assumption (v). Its proof draws on ideas which appear in the proofs of linear convergence of the Pironneau-Polak algorithm for inequality

[2]The vectors $\{ v_j \}_{j \in \underline{p}}$ are said to be *affinely independent* if, for any $j_0 \in \underline{p}$, the vectors $\{ v_j - v_{j_0} \}_{j \in \underline{p} \setminus j_0}$ are linearly independent.

constrained minimization in Refs. 23 and 24. We make the following assumptions, denoting the initial point of Algorithm 2.1 by x_0.

Hypothesis 3.1: *Suppose that*

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

(ii) *the set* $S \triangleq \{ x \in \mathbb{R}^n \mid \psi(x) \le \psi(x_0) \}$ *is bounded, and the necessary conditions* (12a, 12b) *are satisfied at a single point,* $\hat{x} \in S$,

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

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.

Hypothesis 3.2: *Suppose that* (i) *there exists* $m > 0$ such that, for all $\hat{\mu} \in U(\hat{x})$,

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

(ii) m and M are chosen so that $m < \gamma_j < M$ *for all* $j \in \underline{p}$. ■

The convexity assumption (i) is analogous to the weak convexity assumption used in Ref. 23. The proof of linear convergence requires several technical lemmas. 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 $B^\perp$. For any $y \in \mathbb{R}^n$ and $\mu \in \Sigma_p$, we define

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

The function $R(\cdot \, , \cdot)$ is continuous, and, by Hypothesis 3.2(i), $R(0 \, , \hat{\mu})$ is negative definite on the subspace $B^\perp$ for any $\hat{\mu} \in U(\hat{x})$. We use the notation $z_i \to Z$ to represent the convergence of a sequence $\{ z_i \}_{i=0}^\infty \subset \mathbb{R}^n$ to a set $Z \subset \mathbb{R}^n$, *i.e.* - $\lim_{i \to \infty} \min_{y \in Z} \|z_i - y\| = 0$.

Lemma 3.1: *If Hypotheses 3.1 and 3.2 hold, then there exists* $K > 0$ *such that*

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

Proof: See the Appendix. ■

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

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

Proof: See the Appendix. ■

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

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

$$\limsup_{x \to \hat{x}} \frac{\theta(x)}{\psi(x) - \hat{\psi}} \le - \frac{m}{\max_{\mu \in U(\hat{x})} \sum_{j \in \underline{p}} \mu^j \gamma_j} \,. \tag{20}$$

Proof: Since $\sum_{j \in \underline{p}} \mu^j \phi^j(h \mid x) \to \infty$ as $h \to \infty$, uniformly for $x \in S$, and μ lies in the compact set Σ_p, an extension of von Neumann's Minimax Theorem (Ref. 17) permits the interchange of max and min below,

$$\begin{aligned} \theta(x) &\triangleq \min_{h \in \mathbb{R}^n} \max_{\mu \in \Sigma_p} \sum_{j \in \underline{p}} \mu^j \phi^j(h \mid x) - \psi(x) \\ &= \max_{\mu \in \Sigma_p} \min_{h \in \mathbb{R}^n} \sum_{j \in \underline{p}} \mu^j \phi^j(h \mid x) - \psi(x) \,. \end{aligned} \tag{21}$$

For any $\bar{\mu} \in U(x)$,

$$\theta(x) = \min_{h \in \mathbb{R}^n} \sum_{j \in \underline{p}} \bar{\mu}^j \phi^j(h \mid x) - \psi(x) \,. \tag{22}$$

Setting $s(x) \triangleq m\left[\max_{\mu \in U(x)} \sum_{j \in \underline{p}} \mu^j \gamma_j\right]^{-1}$, we have $s(x) < 1$ by Hypothesis 3.2(ii). Substituting $h = s(x)(\hat{x} - x)$ in (22) and using the definition of $\phi^j(\cdot \mid \cdot)$ in (2),

$$\begin{aligned} \theta(x) &\le \sum_{j \in \underline{p}} \bar{\mu}^j \phi^j(s(x)(x - \hat{x}) \mid x) - \psi(x) \\ &= \sum_{j \in \underline{p}} \bar{\mu}^j \left[f^j(x) - \psi(x) + \langle \nabla f^j(x) \,, s(x)(\hat{x} - x) \rangle + \tfrac{1}{2} \gamma_j \, s(x)^2 \, |\hat{x} - x|^2 \right] \\ &\le s(x) \left\{ \sum_{j \in \underline{p}} \bar{\mu}^j f^j(x) + \langle \sum_{j \in \underline{p}} \bar{\mu}^j \nabla f^j(x) \,, \hat{x} - x \rangle + \tfrac{1}{2} m |\hat{x} - x|^2 - \psi(x) \right\} \,, \end{aligned} \tag{23}$$

since $s(x) \in (0\,,1)$ and $f^j(x) \le \psi(x)$. Adding and subtracting the term $\langle x - \hat{x}\,, \left[\int_0^1 (1-t) \sum_{j \in P} \mu^j F^j(\hat{x} + (1-t)(x-\hat{x}))dt\right](x-\hat{x})\rangle$,

$$\theta(x) \le s(x)\left\{\sum_{j \in P} \bar{\mu}^j f^j(x) + \langle \sum_{j \in P} \bar{\mu}^j \nabla f^j(x)\,, \hat{x} - x\rangle + \langle x - \hat{x}\,, \left[\int_0^1 (1-t) \sum_{j \in P} \bar{\mu}^j F^j(\hat{x} + (1-t)(x-\hat{x}))dt\right](x-\hat{x})\rangle - \psi(x) \right.$$

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

Using Taylor's Theorem,

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

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

Dividing both sides of (25) by $\psi(x) - \hat{\psi}$,

$$\frac{\theta(x)}{\psi(x) - \psi(\hat{x})} \le s(x)\left\{-1 + \frac{\langle x - \hat{x}\,, R(x-\hat{x}\,, \bar{\mu})(x-\hat{x})\rangle}{\psi(x) - \psi(\hat{x})}\right\}. \tag{26}$$

By Lemma 3.1,

$$\limsup_{x \to \hat{x}} \max_{\mu \in U(x)} \frac{\langle x - \hat{x}\,, R(x-\hat{x}\,, \mu)(x-\hat{x})\rangle}{|x - \hat{x}|\,|P(x-\hat{x})|} < K\ , \tag{27}$$

and, by Lemma 3.2,

$$\limsup_{x \to \hat{x}} \frac{|x - \hat{x}|\,|P(x-\hat{x})|}{\psi(x) - \hat{\psi}} = 0\ . \tag{28}$$

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

$$\liminf_{x \to \hat{x}} s(x) \ge \frac{m}{\max_{\mu \in U(\hat{x})} \mu^j \gamma_j}\ . \tag{29}$$

Taking the lim sup of (26) as $x \to \hat{x}$ and using (27), (28) and (29) 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.

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

$$\limsup_{\substack{x \to \hat{x} \\ x \neq \hat{x}}} \min_{\lambda \in \mathbb{R}} \frac{\psi(x + \lambda h(x)) - \hat{\psi}}{\psi(x) - \hat{\psi}} \leq 1 - \frac{m}{M} \frac{\min_{j \in p} \gamma_j}{\max_{\mu \in U(\hat{x})} \sum_{k \in p} \mu^k \gamma_k} . \tag{30}$$

Proof: If there existed a point $x \in S$, distinct from $\hat{x}$, such that $\psi(x) = \hat{\psi}$, then x would meet the necessary conditions (12a, 12b). However, this would contradict Hypothesis 3.1(ii), and thus $\psi(\hat{x}) = \hat{\psi}$ and $\psi(x) > \hat{\psi}$ for all $x \neq \hat{x}$. Since $\theta(x)$ is zero if and only if the necessary conditions (12a, 12b) are met at x, $\theta(x) < 0$ for all $x \neq \hat{x}$.

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

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

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

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

Setting $\bar{\lambda} = \min_{j \in p} \gamma_j / M$ and using Hypothesis 3.2(ii),

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

Hence, for all x,

$$\min_{\lambda \in \mathbb{R}} \frac{\psi(x + \lambda h(x)) - \psi(x)}{\theta(x_i)} \geq \bar{\lambda} = \frac{\min_{j \in p} \gamma_j}{M} . \tag{34}$$

Applying inequality (34) and Lemma 3.3 to the right hand side of (30),

$$\limsup_{\substack{x \to \hat{x} \\ x \neq \hat{x}}} \min_{\lambda \in \mathbb{R}} \frac{\psi(x + \lambda h(x)) - \psi(x)}{\psi(x) - \hat{\psi}} = \limsup_{\substack{x \to \hat{x} \\ x \neq \hat{x}}} \min_{\lambda \in \mathbb{R}} \frac{\psi(x + \lambda h(x)) - \psi(x)}{\theta(x)} \quad \frac{\theta(x)}{\psi(x) - \hat{\psi}}$$

$$\leq \frac{\min_{j \in \underline{p}} \gamma_j}{M} \limsup_{\substack{x \to \hat{x} \\ x \neq \hat{x}}} \frac{\theta(x)}{\psi(x) - \hat{\psi}}$$

$$\leq \frac{\min_{j \in \underline{p}} \gamma_j}{M} \left[\frac{-m}{\max_{\mu \in U(\hat{x})} \sum_{j \in \underline{p}} \mu^k \gamma_k} \right]$$

$$= \frac{-m}{M} \frac{\min_{j \in \underline{p}} \gamma_j}{\max_{\mu \in U(\hat{x})} \sum_{j \in \underline{p}} \mu^k \gamma_k} \,. \tag{35}$$

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* $\{ x_i \}_{i=0}^{\infty}$*, then (a)* $x_i \to \hat{x}$*, the unique solution* $\hat{x}$*, and (b) either the sequence terminates in a finite number of steps at* $\hat{x}$ *or*

$$\limsup_{i \to \infty} \frac{\psi(x_{i+1}) - \hat{\psi}}{\psi(x_i) - \hat{\psi}} \leq 1 - \frac{m}{M} \frac{\min_{j \in \underline{p}} \gamma_j}{\max_{\mu \in U(\hat{x})} \sum_{k \in \underline{p}} \mu^k \gamma_k} \,. \tag{36}$$

The left hand side of the inequality is known as the *convergence ratio* of the sequence $\{ \psi(x_i) \}_{i=0}^{\infty}$.

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 (12a, 12b). Since, by Hypothesis 3.1(ii), only $\hat{x} \in S$ satisfies (12a, 12b), the sequence converges to $\hat{x}$.

(b) Follows from (a) and Lemma 3.5. ■

4. Linear Convergence Using an Armijo Step-Size Rule

The step size rule used in Algorithm 2.1 calls for the exact minimization of a single variable function. 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. 17) of the Armijo rule for differentiable functions (Ref. 25), Step 2': Compute the step size,

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

with fixed parameters α, $\beta \in (0, 1)$. The convergence result, Theorem 2.2, holds for the algorithm which substitutes Step 2' for Step 2 in Algorithm 2.1 (Ref. 17). 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 Algorithm 2.1 generates a sequence* $\{ x_i \}_{i=0}^{\infty}$ *using the step size rule Step 2', then (a)* $x_i \to \hat{x}$, *the unique solution* $\hat{x}$, *and (b) either the sequence terminates in a finite number of steps at* $\hat{x}$ *or*

$$\limsup_{i \to \infty} \frac{\psi(x_{i+1}) - \hat{\psi}}{\psi(x_i) - \hat{\psi}} \le 1 - \alpha\beta\frac{m}{M}\frac{\min_{j \in \underline{p}} \gamma_j}{\max_{\mu \in U(\hat{x})} \sum_{k \in \underline{p}} \mu^k\gamma_k} \ . \tag{38}$$

Proof: (a) Same as the proof of Theorem 3.2(a).

(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 $i \in \mathbb{N}$ and $k \ge 0$,

$$\begin{aligned} \psi(x_i + \beta^k h_i) - \psi(x_i) &= \max_{j \in \underline{p}} f^j(x_i + \beta^k h_i) - \psi(x_i) \\ &\le \max_{j \in \underline{p}} f^j(x_i) + \langle \nabla f^j(x_i), \beta^k h_i\rangle - \psi(x_i) + \tfrac{1}{2}M\beta^{2k}|h_i|^2 \\ &\le \beta^k \left[\max_{j \in \underline{p}} f^j(x_i) + \langle \nabla f^j(x_i), h_i\rangle - \psi(x_i) + \tfrac{1}{2}M\beta^k |h_i|^2\right] \ , \end{aligned} \tag{39}$$

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

$$\begin{aligned} \psi(x_i + \beta^k h_i) - \psi(x_i) &\le \beta^k \left[\max_{j \in \underline{p}} f^j(x_i) + \langle \nabla f^j(x_i), h_i\rangle - \psi(x_i) + \tfrac{1}{2}\gamma_j|h_i|^2\right] \\ &= \beta^k\theta(x_i) < \alpha\beta^k\theta(x_i) < 0 \ . \end{aligned} \tag{40}$$

By (37), $\lambda_i \ge \beta\min_{j \in \underline{p}}\gamma_j / M$ and

$$\psi(x_{i+1}) - \psi(x_i) \le \alpha\beta\frac{\min_{j \in \underline{p}} \gamma_j}{M}\, \theta(x_i) \ . \tag{41}$$

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

■

5. Linear Convergence on Composite Minimax Problems

We consider the composite minimax problem,

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

where $g^j : \mathbb{R}^{l_j} \rightarrow \mathbb{R}$ is continuously differentiable and A_j is an $l_j \times n$ real matrix. We note that if the matrices A_j have a common null space, problem (42) will *not* have a unique minimum and therefore will not satisfy Hypothesis 3.1(ii).

Nor will the problem satisfy the convexity requirement of Hypothesis 3.2(i). For problem (42), $f^j \triangleq g^j \circ A_j$ and the second derivative of the Lagrangian at a minimizer $\hat{x}$ has the form,

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

where $G^j(\cdot)$ denotes the second derivative matrix of $g^j(\cdot)$. If the A_j matrices have a common null space, the second derivative matrix may be only positive *semi*-definite on the subspace $B^{\perp}$, so that $m = 0$. This yields a convergence ratio bound of one, allowing for sublinear convergence. However, we have observed in computational experiments that linear convergence of the values $\{ \psi(x_i) \}_{i=0}^{\infty}$ to the minimum value is not lost in this circumstance. We derive a bound on the rate of convergence of Pshenichnyi's method which requires that the Lagrangian Hessian be positive definite only on the orthogonal complement of the common null space. We assume in this section that the Pshenichnyi algorithm employs the exact minimizing line search for determining the step size at each iteration, and that $\gamma_j = \gamma$ for each $j \in \underline{p}$ and some $\gamma > 0$.

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

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

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

Then $\mu_t \triangleq \mu_1 + t(\mu_2 - \mu_1) \in \Sigma_q$, satisfies (12a, 12b) with $\hat{x}$, and hence $\mu_t \in U(\hat{x})$. Necessarily, $\mu_1^{j_0} > 0$, and hence $j_0 \in J(\hat{x})$ by (13). However, $\mu_t^{j_0} = 0$ by construction, implying that $j_0 \notin J(\hat{x})$. 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 some $\hat{x}'$, $\hat{x}'' \in \hat{G}$. 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 (12b) that $\mu_t^{j_1} = 0$ for all $t \in (0\ ,1)$. Now, by the Maximum Theorem in Ref. 26, $U(\cdot)$ 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 (13), 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 $t \in [0\ ,1]$ 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 (12a, 12b) with $\hat{x}'$ satisfies (12a, 12b) with $\hat{x}''$. This fact and (i) 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 $\mu \in U(x)$ and $j \notin \hat{J}$.*

Proof: (a) Since $h(x)$ is the solution of the primal problem (6), it satisfies the optimality conditions (12a, 12b) with respect to problem (6). Every $\mu \in U(x)$ satisfies equations (12a, 12b) together with $h(x)$, and hence the second of those equations yields

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

By Proposition 5.5 in Ref. 17, $h(\hat{x}) = 0$ 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, $\phi^j(h(x) \mid x) \to g^j(A_j\hat{x})$, implying that

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

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

We show below that the Pshenichnyi algorithm converges linearly on some problems of the form (42) which do not satisfy the assumptions of Theorem 3.2. Letting $j_1 < \ldots < j_b$ be the indices comprising $\hat{J}$, we define $\hat{A}^T \triangleq [A_{j_1}^T, \ldots, A_{j_b}^T]$. First we show that the tail of a sequence $\{ x_i \}_{i=0}^{\infty}$ generated by the Pshenichnyi algorithm is contained in a translation of the range of $\hat{A}^T$. Let $a \triangleq rank(\hat{A}^T)$ and let Z be an $n \times a$ matrix, the columns of which form an orthonormal basis for $Range(\hat{A}^T)$. We then show that the sequence corresponds to that constructed on a restriction of problem (42) to a tranlation of the range of $\hat{A}^T$,

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

Finally, we show that the restricted problem satisfies the assumptions of Theorem 3.2. We use $\sigma^+[X]$ 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 $j \in \underline{p}$,

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

(iii) strict complementary slackness holds for all $\hat{x} \in \hat{G}$,

(iv) l and L are chosen so that the scaling parameter, γ, satisfies

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

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

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

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

If Algorithm 2.1 constructs a sequence $\{x_i\}_{i=0}^{\infty}$ *in solving problem (42), then, (b)* $x_i \to \hat{x}$ *for some* $\hat{x} \in \hat{G}$ *as* $i \to \infty$*, and (c) either the sequence terminates in a finite number of steps at* $\hat{x}$ *or*

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

■

The assumption of strict complementary slackness is necessary only if the matrices A_j have different null spaces. In particular, the linear convergence result holds without this assumption if each A_j is the $n \times n$ identity matrix.

Proof of Theorem 5.1: First, we show that $x_i \to \hat{G}$ as $i \to \infty$. Let $\bar{A}^T \triangleq [A_1^T, \dots, A_p^T]$. From (11), every h_i is of the form $\sum_{j \in \underline{p}} A_j^T z_j$, with $z_j \in \mathbb{R}^{l_j}$. Thus, the sequence $\{ x_i \}_{i=0}^{\infty}$ is contained in the closed and convex set $Q \triangleq \{ x_0 + Range(\bar{A}^T) \} \cap \{ x \in \mathbb{R}^n \mid \psi(x) \leq \psi(x_0) \}$. Suppose that Q is unbounded. Then there exists a nonzero $u \in Range(\bar{A}^T)$, such that, with $x_t \triangleq x_0 + tu$, $\psi(x_t) \leq \psi(x_0)$ for all $t \geq 0$. If $A_{j_0} u \neq 0$ for some $j_0 \in \underline{p}$, then the strong convexity of $g^j(\cdot)$, which follows from assumption (ii) of this theorem, implies that $\lim_{t \to \infty} \psi(x_t) = +\infty$. This contradicts $\psi(x_t) \leq \psi(x_0)$, and thus $\bar{A}u = 0$. But $u \in Range(\bar{A}^T) \cap Null(\bar{A}) = \{ 0 \}$ contradicts the assumption that $u \neq 0$. 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 in Ref. 17, any accumulation point $\hat{x}$ of a sequence generated by Algorithm 2.1 must satisfy $\theta(\hat{x}) = 0$, and therefore, by Proposition 3.1, $\hat{x} \in \hat{G}$. Since Q is compact, it follows that $x_i \to \hat{G}$ as $i \to \infty$.

Now we apply Theorem 3.2 to a restricted problem. From Proposition 5.2, there exists a neighborhood $W \supset \hat{G}$ such that $\mu^j = 0$ for all $j \notin \hat{J}$ and $\mu \in U(W)$. From relation (11),

$h(x) = \sum_{j \in p} \mu^j A_j^T \nabla g^j(A_j x_i)$ for any $\mu \in U(x)$. For all $x \in W$, then,

$$h(x) \in Range(\hat{A}^T) \; . \tag{49}$$

Since $x_i \to \hat{G}$, there exists $i_0 \in \mathbb{N}$ such that $x_i \in W$ for all $i > i_0$. Hence, $\{ x_i \}_{i = i_0}^{\infty} \subset x_{i_0} + Range(\hat{A}^T)$. We consider the restricted problem,

$$\min_{y \in \mathbb{R}^a} \psi(x_{i_0} + Zy) \; , \tag{50}$$

formed by substituting $x = x_{i_0} + Zy$ into problem (42). By its definition above, Z has orthonormal columns spanning the range of $\hat{A}^T$. The search direction $d(y)$ constructed by the Pshenichnyi algorithm at a point $y \in \mathbb{R}^a$ with respect to problem (55) satisfies

$$\begin{aligned} d(y) &\triangleq \arg \min_{d \in \mathbb{R}^a} \max_{j \in p} g^j(A_j(x_{i_0} + Zy)) + \langle Z^T A_j^T \nabla g^j(A_j(x_{i_0} + Zy)) \, , d \rangle + \tfrac{1}{2}\gamma \| d \|^2 \\ &= \arg \min_{d \in \mathbb{R}^a} \max_{j \in p} g^j(A_j(x_{i_0} + Zy)) + \langle A_j^T \nabla g^j(A_j(x_{i_0} + Zy)) \, , Zd \rangle + \tfrac{1}{2}\gamma \| Zd \|^2 \\ &= \arg \min_{d \in \mathbb{R}^a} \max_{j \in p} \phi^j(Zd \mid x_{i_0} + Zy) \; , \end{aligned} \tag{51}$$

since $Z^T Z = I_a$ and $\phi^j(h \mid x) \triangleq g^j(x) + \langle \nabla g^j(x), h \rangle + \tfrac{1}{2}\gamma \| h \|^2$. Now, by Proposition 5.2, if $x_{i_0} + Zy \in W$, $h(x_{i_0} + Zy) \in Range(\hat{A}^T)$. Hence, equation (51) implies that

$$h(x_{i_0} + Zy) = Zd(y) \; . \tag{52}$$

Also, for y such that $x_{i_0} + Zy \in W$,

$$\begin{aligned} \arg \min_{\lambda \in \mathbb{R}} \psi(x_{i_0} + Z(y + \lambda d(y))) &= \arg \min_{\lambda \in \mathbb{R}} \psi(x_{i_0} + Zy + \lambda Zd(y)) \\ &= \arg \min_{\lambda \in \mathbb{R}} \psi(x_{i_0} + Zy + \lambda h(x_{i_0} + Zy)) \triangleq \lambda(x_{i_0} + Zy) \; . \end{aligned} \tag{53}$$

Suppose the Pshenichnyi algorithm is applied to problem (50) with initial point $y_0 = 0$, and generates sequences of iterates $\{ y_k \}_{k=0}^{\infty}$, search directions $\{ d(y_k) \}_{k=0}^{\infty}$ and step sizes $\{ \kappa_k \}_{k=0}^{\infty}$. Relations (52) and (53) imply that the search directions satisfy $h(x_{i_0} + Zy_k) = Zd(y_k)$ and that the step sizes satisfy $\kappa_k = \lambda(x_{i_0} + Zy_k)$. Hence, $x_i = x_{i_0} + Zy_{i - i_0}$ for all $i \geq i_0$.

Now we verify that Hypothesis 3.1 is satisfied for the restricted problem (50) and apply Theorem 3.2. The functions $g^j(A_j(x_{i_0} + Zy))$ are twice continuously differentiable in y by assumption (i) of this theorem. The set $\overline{Q} \triangleq \{ y \in \mathbb{R}^a \mid \psi(x_{i_0} + Zy) \le \psi(x_{i_0}) \}$ lies in the set $Z^T(Q - x_{i_0})$, where $Q = (x_0 + Range(\overline{A}^T)) \cap \{ x \in \mathbb{R}^n \mid \psi(x) \le \psi(x_0) \}$, as defined in the proof of Theorem 5.1(a). The set $\overline{Q}$ is bounded since Q is bounded.

Establishing that only a single point in $\overline{Q}$ satisfies the necessary conditions for optimality for problem (50) is slightly involved. Let $\hat{D}$ denote the minimizing set of the restricted problem (55). If $x_{i_0} + Zy \in \hat{G}$, then $y \in \hat{D}$, and hence $\hat{D} \supset Z^T(\hat{G} \cap Q - x_{i_0})$. Now consider any $y' \in \mathbb{R}^a$ such that $x_{i_0} + Zy' \notin \hat{G}$. Since $x_i \to \hat{G}$ and $\{ x_i \}_{i=i_0}^{\infty}$ lies in the closed subspace $x_{i_0} + Range(Z)$, $\hat{G} \cap [x_{i_0} + Range(\hat{A}^T)] \ne \phi$, and hence, there exists $y'' \in \mathbb{R}^a$ such that $x_{i_0} + Zy'' \in \hat{G}$. Then $\psi(x_{i_0} + Zy') > \psi(x_{i_0} + Zy'')$, contradicting the assumption that $y' \in \hat{D}$. Therefore, $\hat{D} = Z^T(\hat{G} \cap Q - x_{i_0})$.

Now consider the set of multipliers which, together with y, satisfy the equations (12a, 12b) corresponding to the optimality conditions for problem (50),

$$U_Y(y) \triangleq \left\{ \mu \in \Sigma_p \;\middle|\; \begin{array}{l} \sum_{j \in \mathbf{p}} \mu^j Z^T A_j^T \nabla g^j(A_j(x_{i_0} + Zy)) = 0 \ , \\ \sum_{j \in \mathbf{p}} \mu^j \left(g^j(A_j(x_{i_0} + Zy)) - \psi(x_{i_0} + Zy)\right) = 0 \end{array} \right\} . \tag{54}$$

For any $\overline{y} \in \hat{D}$, we have $x_{i_0} + Z\overline{y} \in \hat{G}$, and thus $g^j(A_j(x_{i_0} + Z\overline{y})) < \psi(x_{i_0} + Z\overline{y})$ for all $j \notin \hat{J}$. This implies that, for any $\overline{\mu} \in U_Y(\overline{y})$, $\overline{\mu}^j = 0$ for all $j \notin \hat{J}$. Therefore,

$$\sum_{j \in \mathbf{p}} \overline{\mu}^j A_j^T \nabla g^j(A_j(x_{i_0} + Zy)) \in Range(\hat{A}^T) = Range(Z) \ . \tag{55}$$

Equation (55) and $\sum_{j \in \mathbf{p}} \overline{\mu}^j Z^T A_j^T \nabla g^j(A_j(x_{i_0} + Zy)) = 0$ together imply that $\sum_{j \in \mathbf{p}} \overline{\mu}^j A_j^T \nabla g^j(A_j(x_{i_0} + Zy)) = 0$. Hence, $\overline{\mu}$, together with $x_{i_0} + Z\overline{y}$, satisfies the necessary conditions for the original problem (42). Thus, $U_Y(y) \subset U(x_{i_0} + Zy)$ for $y \in \hat{D}$, and $U_Y(\hat{D}) = \{ \hat{\mu} \}$, where $\hat{\mu}$ is the sole member of $U(\hat{G})$.

Suppose that y_1 , $y_2 \in \overline{Q}$ satisfy the optimality conditions corresponding to the for problem (50). Since $\psi(x_{i_0} + Zy)$ 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 $\hat{D}$. Since $U_Y([y_1 , y_2]) = \{ \hat{\mu} \}$ and $\hat{\mu}^j > 0$ for all $j \in \hat{J}$, $g^j(A_j(x_{i_0} + Zy)) = \psi(x_{i_0} + Zy) = \hat{\psi}$ for all $y \in [y_1 , y_2]$ and all $j \in \hat{J}$. Because the functions $g^j(\cdot)$ are strictly convex, this implies that $A_jZy_1 = A_jZy_2$ for all $j \in \hat{J}$, or $y_1 - y_2 \in Null(\hat{A}Z)$. Since $Range(Z) = Range(\hat{A}^T)$, $Null(\hat{A}Z) = \{ 0 \}$, implying that $y_1 = y_2$. Therefore, the necessary conditions are satisfied at a unique point $\hat{y} \in \overline{Q}$ and Hypothesis 3.1(ii) holds. Since $x_i \to \hat{G} \cap [x_{i_0} + Range(\hat{A}^T)]$, this proves part (a) of the theorem, $x_i \to x_{i_0} + Z\hat{y}$.

Letting $\underline{\sigma}[X]$ denote the minimum eigenvalue value of any real symmetric matrix X,

$$\underline{\sigma}\left[\sum_{j \in \underline{P}} \hat{\mu}^j \, \partial^2 g^j(A_j(x_{i_0} + Z\hat{y})) \, / \, \partial y^2\right] = \underline{\sigma}\left[\sum_{j \in \underline{P}} \hat{\mu}^j \, Z^T A_j^T G^j(A_j(x_{i_0} + Z\hat{y}))A_j Z\right] \tag{56}$$

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

$$= l \, \sigma^+\left[\sum_{j \in \underline{P}} \hat{\mu}^j \, A_j^T A_j\right] , \tag{57}$$

since the columns of Z span $Range(\hat{A}^T) = Null(\hat{A})^\perp$ and $Null(\sum_{j \in \underline{P}} \hat{\mu}^j A_j^T A_j) = Null(\hat{A})$. Thus Hypothesis 3.2(i) holds.

Assumption (ii) of Theorem 5.1 ensures that Hypothesis 3.1(iii) holds. Assumption (iv) of Theorem 3.2 holds with $M = L\max_{j \in \underline{P}} |Z^T A_j^T A_j Z|$. Therefore, Theorem 3.2 applies to the restricted problem (50), yielding (a) $y_k \to \hat{y}$ and $x_i \to x_{i_0} + Z\hat{y} \in \hat{G}$. Lemma 3.5 applies as well, and letting $\hat{x} \triangleq x_{i_0} + Z\hat{y}$,

$$\limsup_{\substack{x \to \hat{x} \\ x \in x_{i_0} + Range(Z) \\ x \neq \hat{x}}} \min_{\lambda \in \mathbb{R}} \frac{\psi(x + \lambda h(x)) - \hat{\psi}}{\psi(x) - \hat{\psi}} = \limsup_{\substack{y \to \hat{y} \\ y \neq \hat{y}}} \min_{\lambda \in \mathbb{R}} \frac{\psi(x_{i_0} + Z(y + \lambda d(y)) - \hat{\psi}}{\psi(x_{i_0} + Zy) - \hat{\psi}}$$

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

which is part (b). Part (c) of Theorem 5.1 follows directly from parts (a) and (b). ∎

6. Conclusion

We have shown that a sequence $\{ \psi(x_i) \}_{i=0}^{\infty}$ generated by the Pshenichnyi algorithm converges linearly to the minimum value under weaker conditions than were assumed in Pshenichnyi's convergence rate analysis. Our result applies to an implementable version of the algorithm, not just to a conceptual version requiring that the step sizes lie below some unknown threshold. Furthermore, we have shown that linear convergence is achieved also on a class of composite minimax problems which arise in optimal design problems.

The Pshenichnyi algorithm can be generalized in a straightforward way to solve *semi-infinite* composite minimax problems (Ref.17) which arise in control system design,

$$\min_{x \in \mathbb{R}^n} \max_{j \in \underline{p}} \max_{y_j \in Y_j} \phi^j(A_j x, y_j) \ , \tag{59}$$

where the sets $Y_j \subset \mathbb{R}^{s_j}$ are compact, and the functions $\phi^j : \mathbb{R}^{l_j} \times \mathbb{R}^{s_j} \rightarrow \mathbb{R}$, $j \in \underline{p}$ 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, a linear rate of convergence has been obtained for the semi-infinite case (Refs. 27 and 28).

7. Appendix

Proof of Lemma 3.1: Writing $y = Py + P^{\perp}y$ for any $y \in \mathbb{R}^n$,

$$\begin{aligned} \langle y \, , R(y \, , \mu)y \rangle &= \langle Py + P^{\perp}y \, , R(y \, , \mu)(Py + P^{\perp}y) \rangle \\ &= \langle P^{\perp}y \, , R(y \, , \mu)P^{\perp}y \rangle + \langle Py \, , R(y \, , \mu)(Py + 2P^{\perp}y) \rangle \\ &\leq \langle Py \, , R(y \, , \mu)(Py + 2P^{\perp}y) \rangle \ , \end{aligned} \tag{60}$$

for μ near $U(\hat{x})$ and y small by Hypothesis 3.2(i). Using the Schwarz inequality and the fact that $\| Py + 2P^{\perp}y \| \leq 2\| y \|$,

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

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

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

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

Proof of Lemma 3.2: Using Taylor's Theorem,

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

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

Since $f^j(\hat{x}) = \psi(\hat{x})$ for all $j \in J(\hat{x})$ and by Hypothesis 3.1(iii),

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

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

Since $\langle \nabla f^j(\hat{x})\,,\,P^{\perp}(x-\hat{x})\rangle = 0$ for all $j \in J(\hat{x})$,

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

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

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

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

Suppose not. Then, there exists a nonzero $\bar{u} \in B$ such that $\max_{j\in J(\hat{x})} \langle \nabla f^j(\hat{x})\,,\,\bar{u}\rangle \le 0$. By (12b), $\mu^j = 0$ for all $j \notin J(\hat{x})$ and for any $\mu \in U(\hat{x})$. Therefore,

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

by (12a). Since $U(\hat{x})$ is convex, there exists a $\hat{\mu} \in U(\hat{x})$ such that $\hat{\mu}^j > 0$ for all $j \in J(\hat{x})$. Then, equation (66) implies that there is a convex combination of the nonpositive numbers $\{ \langle \nabla f^j(\hat{x}), \bar{u} \rangle \}_{j \in J(\hat{x})}$ with nonzero coefficients is zero. This implies that $\langle \nabla f^j(\hat{x}), \bar{u} \rangle = 0$ 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 $\sum_{j \in \underline{p}} \hat{\mu}^j \nabla f^j(\hat{x}) = 0$,

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

By Hypothesis 3.1(iii) and Hypothesis 3.2(i),

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

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

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

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

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

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

We use (71) to show that

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

which is equivalent to (19). Given any integer $k > 0$, there exists a real number $r > 0$ such that the first term in the max in (70) is greater than k if $r(x) \le r$. 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 (70) exceeds k for all $x \in W_k$, and, therefore, (19) holds. ■

8. References

1. SHOR, N. Z., *Minimization Methods for Non-Differentiable Functions*, Springer-Verlag, Berlin, 1985.
2. KIWIEL, K. C, *A Phase I - Phase II Method for Inequality Constrained Minimax Problems*, Control and Cybernetics, Vol. 12, pp. 55-75, 1983.
3. 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.
4. MADSEN, K. and SCHJAER-JACOBSEN, H., *Linearly Constrained Minimax Optimization*, Mathematical Programming, Vol. 14, pp. 208-223, 1978.
5. WOMERSLEY, R. S. and FLETCHER R., *An Algorithm for Composite Nonsmooth Optimization Problems*, Journal of Optimization Theory and Applications, Vol. 48, No. 3, pp. 493-523, 1986.
6. FLETCHER, R., *A Model Algorithm for Composite Nondifferentiable Optimization Problems*, Mathematical Programming Study, Vol. 17, pp. 67-76, 1982.
7. HALD, J. and MADSEN, K., *Combined LP and Quasi-Newton methods for minimax optimization*, Mathematical Programming, Vol. 20, pp. 49-62 (1981).
8. MURRAY, W. and OVERTON, M. L., *A Projected Lagrangian Algorithm for Nonlinear Minimax Optimization*, SIAM Journal for Scientific and Statistical Computing, Vol. 1, No. 3, pp. 345-370, 1980.
9. HAN, S. P., *Variable Metric Methods for Minimizing a Class of Nondifferentiable Functions*, Mathematical Programming, Vol. 20, pp. 1-13, 1981.
10. HAN, S. P., *Superlinearly Convergent Variable Metric Algorithms for General Nonlinear Programming Problems*, Mathematical Programming, Vol. 11, pp. 263-282, 1976.
11. POWELL, M. J. D., *The Convergence of Variable Metric Methods for Nonlinearly Constrained Programming Problems*, in O. L. Mangasarian, R. R. Meyer and S. M. Robinson, eds., *Nonlinear Programming 3*, Academic Press, New York, N. Y., 1978.

12. PANIER, E. R., TITS, A. L. and HERSKOVITS, J. N., *A QP-Free, Globally Convergent, Locally Superlinearly Convergent Algorithm for Inequality Constrained Optimization*, SIAM J. Control and Opt., Vol. 26, No. 4, pp. 788-811, 1988.

13. J. N. Herskovits, "A Two-Stage Feasible Directions Algorithm for Nonlinear Constrained Optimization", *Mathematical Programming*, Vol. 36, pp. 19-38, 1986.

14. MAYNE, D. Q. and POLAK, E., *A Superlinearly Convergent Algorithm for Constrained Optimization Problems*, Mathematical Programming Study 16, pp. 45-61, 1982.

15. 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.

16. PSHENICHNYI, B. N. and DANILIN, YU. M., *Numerical Methods in Extremal Problems*, Nauka, Moscow, 1975.

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

18. 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.

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

20. 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.

21. 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.

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

23. CHANEY, R. W., *On the Pironneau-Polak Method of Centers*, Journal of Optimization Theory and Applications, Vol. 20, No. 3, pp. 269-295, 1976.

24. 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.

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

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

27. 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.

28. POLAK, E. and WIEST, E. J., *A Variable Metric Technique for the Solution of Affinely Parametrized Nondifferentiable Optimal Design Problems*, University of California, Berkeley, Electronics Research Laboratory Memo No. UCB/ERL M88/42, 10 June 1988. To appear in the Journal of Optimization Theory and Applications.