A GENERALIZED QUADRATIC PROGRAMMING-BASED PHASE I-PHASE II METHOD FOR INEQUALITY-CONSTRAINED OPTIMIZATION

by

E. J. Wiest and E. Polak

Memorandum No. UCB/ERL M90/46

24 May 1990

A GENERALIZED QUADRATIC PROGRAMMING-BASED PHASE I-PHASE II METHOD FOR INEQUALITY-CONSTRAINED OPTIMIZATION

by

E. J. Wiest and E. Polak

Memorandum No. UCB/ERL M90/46

24 May 1990

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

TITLE PAGE

A GENERALIZED QUADRATIC PROGRAMMING-BASED PHASE I-PHASE II METHOD FOR INEQUALITY-CONSTRAINED OPTIMIZATION

by

E. J. Wiest and E. Polak

Memorandum No. UCB/ERL M90/46

24 May 1990

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

A GENERALIZED QUADRATIC PROGRAMMING-BASED PHASE I - PHASE II METHOD FOR INEQUALITY-CONSTRAINED OPTIMIZATION*

E. J. Wiest† and E. Polak†

ABSTRACT

We present a globally convergent phase I - phase II algorithm for inequality-constrained minimization, which computes search directions by approximating the solution to a generalized quadratic program. In phase II, these search directions are feasible descent directions. The algorithm is shown to converge linearly under convexity assumptions. Both theory and numerical experiments suggest that it generally converges faster than the Polak-Mayne-Trahan method of centers.

KEY WORDS

Methods of feasible directions, constrained optimization, linear convergence, rate of convergence, generalized quadratic program.

*The research reported herein was sponsored in part by the Air Force Office of Scientific Research (grant AFOSR-90-0068) and a Howard Hughes Doctoral Fellowship (Hughes Aircraft Co.).

†Department of Electrical Engineering and Computer Science, University of California, Berkeley, California 94720

1. INTRODUCTION

We consider the inequality-constrained nonlinear programming problem,

$$\min_{x \in \mathbb{R}^n} \{ f^0(x) \mid f^j(x) \le 0 \quad \forall j \in \underline{p} \} \ , \tag{1.1}$$

where $\underline{p}$ denotes the set of natural numbers $\{1, \ldots, p\}$ and the functions $f^j : \mathbb{R}^n \to \mathbb{R}$, $j \in \underline{p} \cup \{0\}$, are continuously differentiable. [Pol.4] proposed algorithms for the solution of problem (1.1) which obtain a search direction at each iteration by solving a natural approximation to (1.1) in which each function $f^j(\cdot)$ is replaced by the quadratic approximation $f^j(x) + \langle \nabla f^j(x), h \rangle + ½\langle h, H_j h \rangle$, for some $H_j \in \mathbb{R}^{n \times n}$. The resulting subproblem is a quadratic program with quadratic constraints, which we will call a *generalized quadratic program* (GQP):

$$\min_{h \in \mathbb{R}^n} \{ f^0(x) + \langle \nabla f^j(x), h \rangle + ½\langle h, H_j h \rangle \mid f^j(x) + \langle \nabla f^j(x), h \rangle + ½\langle h, H_j h \rangle \le 0 \quad \forall j \in \underline{p} \} \ . \tag{1.2}$$

The use of GQP subproblems in algorithms for the solution of (1.1) offers some potential advantages over the use of quadratic programs. For example, information about the curvature of individual constraints can be incorporated directly into the constraints of the GQP subproblem. If the matrices H_j are positive definite and the current iterate is feasible, the resulting search direction is a feasible descent direction. This paper presents the first thorough analysis of convergence and rate of convergence of an *implementable* GQP-based algorithm.

There has been some theoretical analysis of GQP-based algorithms. The convergence of conceptual phase II algorithms is treated in [Pol.4]. Rates of convergence are obtained for GQP-based minimax algorithms in [Pol.5-6] under assumptions of uniform convexity. It is shown in [Pan.3] that, on uniformly convex problems, the norms of the search directions constructed by a conceptual GQP-based algorithm converge superlinearly to zero as the iterates approach a solution.[1]

The GQP-based algorithms proposed in [Pol.4, Pan.3-4] were conceptual, that is, they assumed that the GQP subproblem is solved exactly. These algorithms were not implemented (to our knowledge) because no finite step procedures for solving problem (1.2) were known [Pol.4, Pan.4]. Furthermore, (1.2) may not have feasible solutions if x is infeasible for (1.1). In this paper, we resolve these difficulties for the case of first-order information, where each H_j is taken to be a multiple of the identity.

Our GQP-based method approximates the solution to (1.2) by adding a correction to the search direction of the Polak-Mayne-Trahan algorithm [Pol.2, Pir.1]. The approximation is exact under certain conditions, and requires the solution of only one quadratic program and a projection operation. The method uses the Polak-Mayne-Trahan search direction when no solution to (1.2) exists.

Because we set each H_j in (1.2) to a multiple of the identity, the search direction at each feasible point is a feasible descent direction. Hence, once the algorithm constructs a feasible

---

[1] Quadratic constraints have also appeared in the subproblems of trust region algorithms [More.1]. However, in these algorithms, they function to limit the search direction, rather than to represent the constraints of the problem.

point, x_{i_*}, the inequalities

$$f^j(x_i) \le 0 \ \ \forall j \in \underline{p} \quad \text{and} \quad f^0(x_{i+1}) < f^0(x_i) \ , \tag{1.3}$$

hold for all subsequent iterates, $\{ x_i \}_{i > i_*}$. This property is important in engineering design problems for which function evaluations are extremely costly and for which designs failing to satisfy specifications are unacceptable [Nye.1]. Other first-order algorithms satisfying these requirements include [Hua.1, Mey.1, Mif.1, Pir.1, Pol.1-2, Top.1, Her.1].

We compare the efficiency of our GQP-based algorithm with that of the Polak-Mayne-Trahan algorithm, because the GQP-based algorithm can be viewed as a modification of the Polak-Mayne-Trahan algorithm and because the Polak-Mayne-Trahan algorithm satisfies (1.3) and has been shown to converge linearly in Phase II [Pir.1, Cha.1] under convexity assumptions. We show that the GQP-based algorithm converges linearly with a smaller bound on the *cost convergence ratio*[2] than that obtained for the Polak-Mayne-Trahan algorithm. Numerical experiments also show the new algorithm to be generally superior to the Polak-Mayne-Trahan algorithm, and competitive with the feasible descent algorithm of [Her.1].

The GQP-based algorithm presented in this paper accepts infeasible starting points, and a linear rate of convergence obtains even if the sequence of iterates approaches feasibility only asymptotically.

In Section 3, convergence and rate of convergence results are obtained for a local, conceptual GQP-based algorithm. In Section 4, an implementation of the local, conceptual algorithm is developed. In Section 5, the convergence and rate of convergence results are obtained for the stabilized, implementable algorithm, and the results of numerical experiments are presented in Section 6. The properties of the Polak-Mayne-Trahan algorithm are reviewed in the next section.

2. PROPERTIES OF THE POLAK-MAYNE-TRAHAN ALGORITHM

The Polak-Mayne-Trahan (PMT) algorithm [Pol.2] is a phase I - phase II extension of the Pironneau-Polak algorithm [Pir.1], which, in turn, is an implementation of Huard's method of centers [Hua.1]. The PMT algorithm is one of very few first-order phase I - phase II methods for which the rate of convergence is known (see also [Pol.3]), and its computational behavior is quite competitive in this class. We will use the PMT algorithm as a benchmark for evaluating the new algorithm. The PMT algorithm solves problems of the form

$$\min \{ f^0(x) \mid f^j(x) \le 0 \ , \ j \in \underline{p} \ \} \ , \tag{2.1a}$$

under the assumption that the functions $f^j : \mathbb{R}^n \to \mathbb{R}$ are continuously differentiable and the constraint qualification that the function $\max_{j \in \underline{p}} f^j(x)$ not have any stationary points outside the interior of the feasible set.

We will use the following definitions. We denote the set of natural numbers $\{ 1 , \dots , p \}$ by $\underline{p}$, and the set $\{ 0 , 1 , \dots , p \}$ by $\underline{p} \cup 0$. The p smooth constraints $f^j(x) \le 0$ in (1.1) can be

[2]We define the *cost convergence ratio* of a sequence $\{ x_i \}_{i \in \mathbb{N}}$ which converges to $\hat{x}$ to be $\limsup_{i \to \infty} | f^0(x_{i+1}) - f^0(\hat{x}) | \, / \, | f^0(x_i) - f^0(\hat{x}) |$.

combined into a single nonsmooth constraint $\psi(x) \le 0$, where $\psi(x) \triangleq \max_{j \in \underline{p}} f^j(x)$. Constraint violation is indicated by the values of the function $\psi_+(x) \triangleq \max\{\psi(x), 0\}$. Finally, we define first-order convex approximations to the functions $f^j(\cdot)$ at x by

$$\tilde{f}^j(h \mid x) \triangleq \begin{cases} f^j(x) + \langle \nabla f^j(x), h \rangle + \tfrac{1}{2}\gamma |h|^2 & \text{if } j \in \underline{p} \\ \langle \nabla f^0(x), h \rangle + \tfrac{1}{2}\gamma |h|^2 & \text{if } j = \underline{0}, \end{cases} \tag{2.1b}$$

for some fixed $\gamma > 0$. Note that $\tilde{f}^0(0 \mid x) = 0$ and $\tilde{f}^j(0 \mid x) = f^j(x)$ for all $j \in \underline{p}$.

Algorithm 2.1

Data: x_0; $\alpha, \beta \in (0,1)$; $\gamma > 0$; $i = 0$.

Step 1: Compute the *search direction*,

$$h(x_i) \triangleq \operatorname*{argmin}_{h \in \mathbb{R}^n} \max_{j \in \underline{p} \cup 0} \tilde{f}^j(h \mid x_i), \tag{2.2a}$$

and evaluate the *optimality function*

$$\theta(x_i) \triangleq \max_{j \in \underline{p} \cup 0} \tilde{f}^j(h(x_i) \mid x_i) - \psi_+(x_i). \tag{2.2b}$$

Step 2: If $\psi(x_i) \le 0$, set

$$\lambda_i = \max\{\beta^k \mid f^0(x_i + \beta^k h(x_i)) - f^0(x_i) \le \alpha\beta^k\theta(x_i) \text{ and } \psi(x_i + \beta^k h(x_i)) \le 0\}, \tag{2.2c}$$

else set

$$\lambda_i = \max\{\beta^k \mid \psi(x_i + \beta^k h(x_i)) - \psi(x_i) \le \alpha\beta^k\theta(x_i)\}. \tag{2.2d}$$

Step 3: Set $x_{i+1} = x_i + \lambda_i h(x_i)$.

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

Step 2 ensures that, once a sequence generated by Algorithm 2.1 has entered the feasible region $X \triangleq \{x \in \mathbb{R}^n \mid f^j(x) \le 0 \; \forall j \in \underline{p}\}$, it can never leave it. Referring to [Pol.1] we see that the search direction vector $h(x_i)$ can be computed in two steps. First one solves the dual of (2.2a), i.e., the positive semi-definite quadratic program

$$\max_{\mu \in \Sigma_{p+1}} \min_{h \in \mathbb{R}^n} \sum_{j \in \underline{p} \cup 0} \mu^j \tilde{f}^j(h \mid x) - \psi_+(x)$$

$$= \max_{\mu \in \Sigma_{p+1}} \sum_{j \in \underline{p}} \mu^j f^j(x) - \psi_+(x) - \tfrac{1}{2}\gamma^{-1} \Big| \sum_{j \in \underline{p} \cup 0} \mu^j \nabla f^j(x) \Big|^2, \tag{2.3}$$

for any solution $\mu(x_i)$. We denote the set of solutions to (2.3) by $U_{PP}(x) \triangleq \operatorname*{argmax}_{\mu \in \Sigma_{p+1}} \sum_{j \in \underline{p}} \mu^j f^j(x) - \psi_+(x) - \tfrac{1}{2}\gamma^{-1} \big| \sum_{j \in \underline{p} \cup 0} \mu^j \nabla f^j(x) \big|^2$. This can be done using one of several methods [Gil.1, Hoh.1, Hig.1, Kiw.1-2, Rus.1]. The unique primal solution, $h(x_i)$, is then given by

$$h(x) = \operatorname*{argmin}_{h \in \mathbb{R}^n} \sum_{j \in \underline{q} \cup 0} \mu^j(x) \tilde{f}^j(h \mid x) = \frac{1}{\gamma} \sum_{j \in \underline{q} \cup 0} \mu^j(x) \nabla f^j(x). \tag{2.4a}$$

From (2.3), we can write

$$\theta(x) = \max_{\mu \in \Sigma_{p+1}} \sum_{j \in \underline{p}} \mu^j f^j(x) - \psi_+(x) - \tfrac{1}{2}\gamma^{-1} \| \sum_{j \in \underline{p} \cup 0} \mu^j \nabla f^j(x) \|^2 . \tag{2.4b}$$

The following theorem summarizes the properties of the optimality function $\theta: \mathbb{R}^n \to \mathbb{R}$, the search direction function $h: \mathbb{R}^n \to \mathbb{R}^n$ used in the above algorithm.

Theorem 2.1[Pol.1]

(a) *If $\hat{x}$ is a local minimizer for problem (1.1), then $\theta(\hat{x}) = 0$.*

(b) *For any $\bar{x} \in \mathbb{R}^n$, $\theta(\bar{x}) = 0$ if and only if there exists $\hat{\mu} \in \Sigma_{p+1}$ such that*

$$\sum_{j \in \underline{p} \cup 0} \hat{\mu}^j \nabla f^j(\bar{x}) = 0 , \tag{2.5a}$$

$$\sum_{j \in \underline{p}} \hat{\mu}^j f^j(\bar{x}) = \psi_+(\bar{x}) . \tag{2.5b}$$

(c) *Both $\theta(\cdot)$ and $h(\cdot)$ are continuous.* □

Note that if $\bar{x}$ satisfies (2.5a-b) for some $\mu \in \Sigma_{p+1}$, and $\psi_+(\bar{x}) > 0$, then $\mu^0 = 0$, and hence $\bar{x}$ satisfies the standard first order condition for a local minimizer of $\psi(\cdot)$. If $\hat{x}$ is a local minimizer of (1.1), then $U_{PP}(\hat{x})$ is the set of Fritz John multiplier vectors which, together with $\hat{x}$, satisfy (2.5a-b).

Theorem 2.2:[Pol.1] *If $\bar{x}$ is an accumulation point of a sequence $\{x_i\}_{i=0}^{\infty}$ constructed by Algorithm 2.1 in solving (1.1), then $\theta(\bar{x}) = 0$. Furthermore, if, for all $x \in \mathbb{R}^n$ such that $\psi(x) \geq 0$, $0 \notin \partial\psi(x)$ (where $\partial\psi(x)$ denotes the generalized gradient of $\psi(\cdot)$ at x [Cla.1]), then $\psi(\bar{x}) \leq 0$.* □

It was first shown in [Pir.1] that an algorithm based on the search direction rule (2.2a) converges linearly under convexity assumptions. Chaney [Cha.1] later established linear convergence under slightly weaker assumptions. The following theorem is a variant of Chaney's result, accounting for the fact that Algorithm 2.1 uses an Armijo-type line search [Arm.1] rather than an exact minimizing line search as in [Cha.1, Pir.1]. Let

$$F^j(x) \triangleq \partial^2 f^j(x)/\partial x^2 , \tag{2.6}$$

and $\underline{\mu}^0 \triangleq \min \{ \mu^0 \mid \mu \in U_{PP}(\hat{x}) \}$.

Theorem 2.3: *Suppose that*

(i) *the functions $f^j(\cdot)$, $j \in \underline{p} \cup 0$ are twice continuously differentiable,*

(ii) *the set $L \triangleq \{ x \in \mathbb{R}^n \mid \psi(x) \leq \psi_+(x_0) \}$ is bounded, and the necessary conditions* (2.5a-b) *are satisfied at a single point, $\hat{x} \in X$, at which the Mangasarian-Fromovitz constraint qualification holds (i.e. - there exist $\bar{h} \in \mathbb{R}^n$ and $\delta > 0$ such that $\langle \nabla f^j(\hat{x}), \bar{h} \rangle < -\delta$ for each $j \in \underline{p}$ such that $f^j(\hat{x}) = 0$),*

(iii) *for $\hat{x}$ as above, and with*

$$\hat{J} \triangleq \bigcup \{ J(\mu) \mid \mu \in U_{PP}(\hat{x}) \} , \tag{2.7a}$$

where for any $\mu \in \Sigma_{p+1}$, $J(\mu) \triangleq \{ j \in \underline{p} \mid \mu^j > 0 \}$, there exists $m \in (0, \gamma)$ such that

$$m\|h\|^2 < \langle h\,,\left[\sum_{j\in \underline{p}\cup 0}\mu^j F^j(\hat{x})\right]h\rangle\,, \tag{2.7b}$$

for every $\mu \in U_{PP}(\hat{x})$ *and for every nonzero* $h \in H$, *where*

$$H \triangleq \{\, h \mid \langle \nabla f^j(\hat{x})\,, h\rangle = 0\,,\ \forall j \in \hat{J}\ \}\,. \tag{2.7c}$$

If Algorithm 2.1 constructs a sequence $\{x_i\}_{i=0}^{\infty}$ *in solving problem (1.1), then* (a) $x_i \to \hat{x}$ *as* $i \to \infty$, *and* (b) *if* $\psi(x_i) \le 0$ *for any* $i \in \mathbb{N}$, *then*

$$\limsup_{i\to\infty} \frac{f^0(x_{i+1}) - f^0(\hat{x})}{f^0(x_i) - f^0(\hat{x})} \le 1 - \alpha\beta\mu^0 \frac{m}{M}\,, \tag{2.7d}$$

for any $M > \max_{j\in \underline{p}\cup 0}\{\,\|F^j(\hat{x})\|\,,\gamma\}$. □

Inequality (2.7d) then gives an upper bound on the cost convergence ratio of sequences constructed by Algorithm 2.1.

3. A CONCEPTUAL GQP-BASED ALGORITHM

We begin by considering a conceptual, local algorithm for solving (1.1) which computes a search direction at x_i by solving the *generalized quadratic program*,

$$\mathrm{GQP}(x):\ \min_{h\in\mathbb{R}^n}\{\,\tilde{f}^0(h \mid x) \mid \tilde{f}^j(h \mid x) \le 0 \quad \forall j \in \underline{p}\,\}\,, \tag{3.1a}$$

with $x = x_i$.

Local Algorithm 3.1:

Data: x_0; $\beta \in (0\,,1)$; $\gamma > 0$; $i = 0$.

Step 1: Compute the search direction,

$$h_i = h_{\mathrm{GQP}}(x_i) \triangleq \operatorname*{argmin}_{h\in\mathbb{R}^n}\{\,\tilde{f}^0(h \mid x) \mid \tilde{f}^j(h \mid x) \le 0 \quad \forall j \in \underline{p}\,\}\,. \tag{3.1b}$$

Step 2: Compute the step size,

$$\lambda_i = \max\{\,\beta^k \mid f^0(x_i + \beta^k h_i) - f^0(x_i) \le \beta^k \tilde{f}^0(h_i \mid x_i)\,,$$
$$\psi_+(x_i + \beta^k h_i) - \psi_+(x_i) \le \beta^k[\max_{j\in\underline{p}}\{\,\tilde{f}^j(h_i \mid x_i)\,, 0\,\} - \psi_+(x_i)]\,\} \tag{3.1c}$$

Step 3: Set $x_{i+1} = x_i + \lambda_i h_i$.

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

Lemma 3.1: *Suppose that assumptions* (i)-(iii) *of Theorem 2.3 hold, and let* $\hat{x}$ *be as defined in assumption* (ii) *of Theorem 2.3. Then there exists a neighborhood* V *of* $\hat{x}$ *such that* $\mathrm{GQP}(x)$ *has a continuous solution,* $h_{\mathrm{GQP}}(x)$, *for all* $x \in V$.

Proof: Suppose that $x \in \mathbb{R}^n$ is such that there exists an $h' \in \mathbb{R}^n$ satisfying $\tilde{f}^j(h' \mid x) < 0$ for all $j \in \underline{p}$. Then the set-valued map $G(x) \triangleq \{\, h \in \mathbb{R}^n \mid \tilde{f}^j(h \mid x) \le 0\,,\ \forall\, j \in \underline{p}\,\}$ is upper semicontinuous at x. $G(x)$ is compact since the functions $\tilde{f}^j(\cdot)$ are uniformly convex. Hence, by the Maximum Theorem [Ber.1], the set of solutions to GQP(x) is an upper semicontinuous set-valued map at x. Since GQP(x) is a strictly convex program, its solution set is a singleton, $\{h_{GQP}(x)\}$. Therefore, the solution, $h_{GQP}(x)$, to GQP(x) is continuous at any point x at which GQP(x) is strictly feasible.

By assumption (ii) of Theorem 2.3, there exist $\bar{h} \in \mathbb{R}^n$ and $\delta > 0$ such that $\langle \nabla f^j(\hat{x}), \bar{h} \rangle < -\delta$ for each $j \in J(\hat{x})$. Therefore, there exist $t > 0$ and a neighborhood, V, of $\hat{x}$ such that $\tilde{f}^j(t\bar{h} \mid x) < 0$ for all $x \in V$ and $j \in \underline{p}$. Therefore, GQP(x) is strictly feasible for all $x \in V$. In light of the previous paragraph, $h_{GQP}(x)$ exists and is continuous in V. □

For any $x \in \mathbb{R}^n$ such that GQP(x) has a solution, we will denote the set of *Fritz John* multiplier vectors associated with the unique solution, $h_{GQP}(x)$ by

$$U_{GQP}(x) \triangleq \{\mu \in \Sigma_{p+1} \mid \sum_{j \in \underline{p} \cup 0} \mu^j \nabla \tilde{f}^j(h_{GQP}(x) \mid x) = 0,$$

$$\sum_{j \in \underline{p}} \mu^j \tilde{f}^j(h_{GQP}(x) \mid x) = 0\} . \tag{3.1d}$$

Consider the l_∞ penalty function, $p_\varepsilon(x) \triangleq \varepsilon f^0(x) + \psi_+(x)$, where $\varepsilon > 0$. The proofs below exploit the correspondence between minimizers of the constrained problem (1.1) and those of the minimax problem,

$$\min_{x \in \mathbb{R}^n} p_\varepsilon(x) . \tag{3.2a}$$

As is shown in the following lemma, the solution to (1.1) is also a strict local minimizer of $p_\varepsilon(\cdot)$ for sufficiently small ε. Let

$$d_\varepsilon(x) \triangleq \operatorname*{argmin}_{h \in \mathbb{R}^n} \{\varepsilon \tilde{f}^0(h \mid x) + \max\{0, \tilde{f}^j(h \mid x)\}\} , \tag{3.2b}$$

and let

$$\theta_\varepsilon(x) \triangleq \varepsilon \tilde{f}^0(d_\varepsilon(x) \mid x) + \max\{0, \tilde{f}^j(d_\varepsilon(x) \mid x)\} - \psi_+(x) . \tag{3.2c}$$

Recall that $\underline{\mu}^0 \triangleq \min\{\mu^0 \mid \mu \in U_{PP}(\hat{x})\}$.

Lemma 3.2: *Suppose that assumptions* (i)-(iii) *of Theorem 2.3 hold, let $\hat{x}$ be as defined in assumption* (ii) *of Theorem 2.3, and let V be as defined in Lemma 3.1. Then, for any $\varepsilon \in (0, \underline{\mu}^0/(1-\underline{\mu}^0))$, there exists a neighborhood, $W_\varepsilon \subset V$, of $\hat{x}$, such that, for all $x \in W_\varepsilon$,* (a) $p_\varepsilon(x) \geq p_\varepsilon(\hat{x}) + \frac{1}{4}m|x - \hat{x}|^2$, *and* (b) $d_\varepsilon(x) = h_{GQP}(x)$.

Proof: (a) Assumptions (i)-(iii) of Theorem 2.3 ensure that the point $\hat{x}$ satisfies the standard second-order sufficiency conditions for problem (1.1) [McC.1]. In fact, they ensure that $\hat{x}$ satisfies these conditions for the problem,

$$\min_{x \in \mathbb{R}^n} \{f^0(x) - \tfrac{1}{4}m|x - \hat{x}|^2 \mid f^j(x) - \tfrac{1}{4}m|x - \hat{x}|^2 \leq 0\} . \tag{3.2d}$$

It follows from Theorem 4.6 of [Han.1] (see Theorem A.1 of the Appendix for a restatement), therefore, that $\hat{x}$ is a strict local minimizer of $p_\varepsilon(\cdot) - \frac{1}{4}m(\varepsilon+1)|\cdot - \hat{x}|^2$, provided that $1/\varepsilon > \sum_{j \in \underline{p}} u^j$ for some Kuhn-Tucker multiplier vector for the problem (3.2d), $u \in \mathbb{R}^p$, associated with $\hat{x}$. Since the Kuhn-Tucker multiplier vectors for (3.2d) associated with $\hat{x}$ are the same as those of (1.1), we can construct a Kuhn-Tucker multiplier vector for (3.2d) from any Fritz John multiplier vector, $\mu \in U_{PP}(\hat{x})$, as follows:

$$u_\mu = (\mu^1, \ldots, \mu^p)/\mu^0 , \tag{3.2e}$$

because the Mangasarian-Fromovitz constraint qualification (assumption (ii) of Theorem 2.3) ensures that $\mu^0 \geq \underline{\mu}^0 > 0$. Hence, if $1/\varepsilon > |u_\mu|_1 = (1-\underline{\mu}^0)/\underline{\mu}^0$, then $\hat{x}$ is a strict local minimizer of $p_\varepsilon(\cdot) - \frac{1}{4}m(\varepsilon+1)|\cdot - \hat{x}|^2$. This implies that $p_\varepsilon(\hat{x}) \leq p_\varepsilon(x) - \frac{1}{4}m(\varepsilon+1)|x - \hat{x}|^2$ for x in some neighborhood of $\hat{x}$.

(b) We recall that by Lemma 3.1, the solution $h_{GQP}(x)$ to GQP(x) exists for all x in a neighborhood V of $\hat{x}$. We will now prove that, for any $\varepsilon < \underline{\mu}^0/(1-\underline{\mu}^0)$, $d_\varepsilon(x) = h_{GQP}(x)$ for all x in a neighborhood of $\hat{x}$.

We first show that, for x near $\hat{x}$, the norm of some Kuhn-Tucker multiplier vector associated with the solution to GQP(x) is bounded from above by $(1-\underline{\mu}^0)/\underline{\mu}^0$. We denote the set of Kuhn-Tucker multiplier vectors for GQP(x) by

$$KT_{GQP}(x) \triangleq \{ u \in \mathbb{R}_+^p \mid \nabla \tilde{f}^0(h_{GQP}(x) \mid x) + \sum_{j \in \underline{p}} u^j \nabla \tilde{f}^j(h_{GQP}(x) \mid x) = 0 ,$$
$$\sum_{j \in \underline{p}} \mu^j \tilde{f}^j(h_{GQP}(x) \mid x) = 0 \} , \tag{3.3}$$

for $x \in V$. Since $h_{GQP}(\hat{x}) = 0$ and $\psi_+(\hat{x}) = 0$, an inspection of (3.1d) reveals that $U_{GQP}(\hat{x}) = U_{PP}(\hat{x})$. By assumption (ii) of Theorem 2.3, $\underline{\mu}^0 > 0$. Since $U_{GQP}(\cdot)$ is an upper semicontinuous, compact-valued set-valued map at $\hat{x}$, there exists, for any $\delta \in (0, \underline{\mu})$, a neighborhood, $W_\delta \subset V$, of $\hat{x}$, such that $\mu^0 > \underline{\mu}^0 - \delta$ for every $\mu \in U_{GQP}(W_\delta)$. Now, every Fritz John multiplier vector $\mu \in U_{GQP}(x)$ corresponds to a Kuhn-Tucker multiplier vector, $u_\mu \triangleq (\mu^1/\mu^0, \ldots, \mu^p/\mu^0) \in KT_{GQP}(x)$. For such Kuhn-Tucker multiplier vectors, $|u_\mu|_1 = (1-\mu^0)/\mu^0 < (1-\underline{\mu}^0+\delta)/(\underline{\mu}^0-\delta)$ for every $\mu \in U_{GQP}(W_\delta)$.

Because (i) for any $\delta \in (0, \underline{\mu})$, there exists a neighborhood W_δ of $\hat{x}$ such that $\min\{ |u|_1 \mid u \in KT_{GQP}(x)\} < (1-\underline{\mu}^0+\delta)/(\underline{\mu}^0-\delta)$ for $x \in W_\delta$ (from the previous paragraph), (ii) $\max_{j \in \underline{p}} \tilde{f}^j(h' \mid x) < 0$ for $x \in V$ and some $h' \in \mathbb{R}^n$ (from the proof of Lemma 3.1), and (iii) problem GQP(x) is a convex program, we can apply Theorem 4.9 of [Han.1] to conclude that, for $\varepsilon < (\hat{\mu}^0 - \delta)/(1-\hat{\mu}^0+\delta)$, $h_{GQP}(x)$ is the unique minimizer of the convex function $\min_{d \in \mathbb{R}^n} \varepsilon \tilde{f}^0(d \mid x) + \max\{0, \tilde{f}^j(d \mid x)\}$ for all $x \in W_\delta$. (See Theorem A.2 of the Appendix for a restatement of Theorem 4.9 of [Han.1].) Hence, $h_{GQP}(x) = d_\varepsilon(x)$ for all $x \in W_\delta$. Since δ was arbitrary, such a neighborhood exists for any $\varepsilon < \underline{\mu}^0/(1-\underline{\mu}^0)$. □

Theorem 3.1: *Suppose that assumptions* (i)-(iii) *of Theorem 2.3 hold, and let* $\hat{x}$ *be as defined in* *assumption* (ii) *of Theorem 2.3. Then, for any neighborhood, W, of $\hat{x}$, there exists a neighborhood $V_W \subset W$ of $\hat{x}$ such that, if $x_0 \in V_W$, the sequence $\{x_i\}_{i \in \mathbb{N}}$ constructed by Algorithm 3.1 remains in V_W and converges to $\hat{x}$.*

Proof: Let $A(\cdot)$ denote the *iteration map* of Algorithm 3.1. The function $A(\cdot)$ maps one iterate into the next, i.e., $x_{i+1} = A(x_i)$. The sequence $\{x_i\}_{i \in \mathbb{N}}$ will remain in a set V_W if the set V_W is invariant under $A(\cdot)$, i.e., $A(V_W) \subset V_W$. We now show that such a neighborhood $V_W \subset W$ of $\hat{x}$ exists.

Let $\varepsilon < \underline{\mu}^0 / (1 - \underline{\mu}^0)$ be arbitrary. By Lemma 3.2(a), there exists a neighborhood W_ε of $\hat{x}$ such that $p_\varepsilon(x) \geq p_\varepsilon(\hat{x}) + \frac{1}{4} m' |x - \hat{x}|^2$ for $x \in W_\varepsilon$. For small enough $\delta > 0$, therefore, the set $V_W \triangleq \{ x \in W_\varepsilon \mid p_\varepsilon(x) < p_\varepsilon(\hat{x}) + \delta \}$ is contained in W. By the continuity of $p_\varepsilon(\cdot)$, the set V_W is a neighborhood of $\hat{x}$.

By Step 2 of Algorithm 3.1, with $x_1 = A(x_0)$ for any $x_0 \in V$,

$$p_\varepsilon(x_1) - p_\varepsilon(x_0) = \varepsilon[f^0(x_1) - f^0(x_0)] + [\psi_+(x_1) - \psi_+(x_0)]$$

$$\leq \lambda_0[\varepsilon \tilde{f}^0(h_{\mathrm{GQP}}(x_0) \mid x_0) + \max_{j \in \underline{p}} \{ \tilde{f}^j(h_{\mathrm{GQP}}(x_0) \mid x_0), 0 \} - \psi_+(x_0)] . \tag{3.4a}$$

By Lemma 3.2(b), $h_{\mathrm{GQP}}(x_0) = h_\varepsilon(x_0)$ for $x_0 \in V_W$, and, hence,

$$p_\varepsilon(x_1) - p_\varepsilon(x_0) \leq \lambda_0[\varepsilon \tilde{f}^0(h_\varepsilon(x_0) \mid x_0) + \max_{j \in \underline{p}} \{ \tilde{f}^j(h_\varepsilon(x_0) \mid x_0), 0 \} - \psi_+(x_0)]$$

$$= \lambda_0 \theta_\varepsilon(x_0) \leq 0 . \tag{3.4b}$$

Therefore, $p_\varepsilon(x_1) \leq p_\varepsilon(x_0) \leq p_\varepsilon(\hat{x}) + \delta$, implying that $A(V_W) \subset V_W$.

Now we show that only $\hat{x}$ can be an accumulation point of the sequence $\{x_i\}_{i \in \mathbb{N}}$ constructed by Algorithm 3.1, from an $x_0 \in V_W$. Suppose that $\{x_i\}_{i \in K}$ converges to $\bar{x} \in V_W$, where $K \subset \mathbb{N}$ and $\bar{x} \neq \hat{x}$. Since, by assumption (ii) of Theorem 2.3, $\hat{x}$ is the only stationary point for (1.1) in V_W, $\bar{x}$ cannot be stationary for problem (3.2a). By Lemma 3.1, $\tilde{f}^0(h_{\mathrm{GQP}}(x) \mid x))$ is continuous in V_W, and therefore there exist $\delta > 0$ and a neighborhood, $W' \subset V_W$, of $\bar{x}$ such that

$$\theta_\varepsilon(x) < -\delta , \tag{3.4c}$$

for all $x \in W'$. Clearly, there exists an $i_0 \in K$, such that $x_i \in W'$ for all $i > i_0$, $i \in K$. Let $M' < \infty$ be such that $|F^j(x)| \leq M'$ for all $x \in W$. Then,

$$f^j(x_i + \lambda h_{\mathrm{GQP}}(x_i)) \leq f^j(x_i) + \lambda \langle \nabla f^j(x_i), h_{\mathrm{GQP}}(x_i) \rangle + \tfrac{1}{2} M' \lambda^2 |h_{\mathrm{GQP}}(x_i)|^2 , \tag{3.4d}$$

for all $i \in K$, $i > i_0$. Hence, for $j \in \underline{p}$ and $\lambda \leq 1$,

$$f^j(x_i + \lambda h_{\mathrm{GQP}}(x_i)) - \psi_+(x_i) \leq \lambda \{ f^j(x_i) + \langle \nabla f^j(x_i), h_{\mathrm{GQP}}(x_i) \rangle + \tfrac{1}{2} M' \lambda |h_{\mathrm{GQP}}(x_i)|^2 - \psi_+(x_i) \} , \tag{3.4e}$$

since $\psi_+(x_i) \geq f^j(x_i)$. For $\lambda \leq \gamma / M'$, then

$$f^j(x_i + \lambda h_{\mathrm{GQP}}(x_i)) - \psi_+(x_i) \leq \lambda \{ f^j(x_i) + \langle \nabla f^j(x_i), h_{\mathrm{GQP}}(x_i) \rangle + \tfrac{1}{2} \gamma |h_{\mathrm{GQP}}(x_i)|^2 - \psi_+(x_i) \}$$

$$= \lambda \{ \tilde{f}^{j}(h_{GQP}(x_i) \mid x_i) - \psi_+(x_i) \} . \tag{3.4f}$$

Taking the maximum over $j \in \underline{p}$,

$$\psi_+(x_i + \lambda h_{GQP}(x_i)) - \psi_+(x_i) \le \lambda \left\{ \max_{j \in \underline{p}} \{ 0 , \tilde{f}^{j}(h_{GQP}(x_i) \mid x_i) \} - \psi_+(x_i) \right\} , \tag{3.4g}$$

for all $\lambda \in (0 , \gamma / M']$ and $i > i_0 , i \in K$. Setting $j = 0$ in (3.4d),

$$f^0(x_i + \lambda h_{GQP}(x_i)) - f^0(x_i) \le \lambda \{ \langle \nabla f^j(x_i) , h_{GQP}(x_i) \rangle + \tfrac{1}{2}\gamma \| h_{GQP}(x_i) \|^2 \}$$

$$= \lambda \tilde{f}^0(h_{GQP}(x_i) \mid x_i) , \tag{3.4h}$$

for $\lambda \le \gamma / M'$ and $i \in K , i > i_0$. Inequalities (3.4g) and (3.4h) and Step 2 of Algorithm 3.1 imply that $\lambda_i \ge \beta\gamma / M'$. From Step 2 of Algorithm 3.1 and the fact that $h_{GQP}(x_i) = h_\varepsilon(x_i)$ for $i > i_0$,

$$p_\varepsilon(x_{i+1}) - p_\varepsilon(x_i) = p_\varepsilon(x_i + \lambda_i h_{GQP}(x_i)) - p_\varepsilon(x_i)$$

$$\le \lambda_i \left\{ \varepsilon \tilde{f}^0(h_{GQP}(x_i) \mid x_i) + \max_{j \in \underline{p}} \{ 0 , \tilde{f}^{j}(h_{GQP}(x_i) \mid x_i) \} - \psi_+(x_i) \right\}$$

$$\le \lambda_i \left\{ \varepsilon \tilde{f}^0(h_\varepsilon(x_i) \mid x_i) + \max_{j \in \underline{p}} \{ 0 , \tilde{f}^{j}(h_\varepsilon(x_i) \mid x_i) \} - \psi_+(x_i) \right\}$$

$$= \lambda_i \theta_\varepsilon(x_i) . \tag{3.4i}$$

Then for $i \in K , i > i_0$,

$$p_\varepsilon(x_{i+1}) - p_\varepsilon(x_i) = p \le \frac{\beta\gamma}{M} \theta_\varepsilon(x_i) \le -\frac{\beta\gamma\delta}{M} . \tag{3.4j}$$

Now $p_\varepsilon(x_{i+1}) \le p_\varepsilon(x_i)$ for *all* $i > i_0$ by (3.4i). Hence (3.4j) implies that $p_\varepsilon(x_i) \to -\infty$ as $i \to \infty$. However, this is impossible, since $\{ x_i \}_{i \in \mathbf{N}}$ is contained in the bounded set $V_{\bar{w}}$. Therefore, $\bar{x} \ne \hat{x}$ cannot be an accumulation point for the sequence.

Since $V_{\bar{w}}$ is compact, the sequence $\{ x_i \}_{i \in \mathbf{N}} \subset V_{\bar{w}}$ must converge to the set of its accumulation points. We have shown that $\hat{x}$ can be the only accumulation point for the sequence. Therefore, the sequence converges to $\hat{x}$. □

Let $\bar{\mu}^0 \triangleq \max \{ \mu^0 \mid \mu \in U_{PP}(\hat{x}) \}$.

Theorem 3.2: *Suppose that assumptions* (i)-(iii) *of Theorem 2.3 hold with* $\hat{x}$ *as defined there, that* $x_0 \in V_{\bar{w}}$, *with any* $V_{\bar{w}}$ *as defined in Theorem 3.1, and that Algorithm 3.1 constructs a sequence* $\{ x_i \}_{i=0}^{\infty}$ *in solving (1.1) starting from a point* $x_0 \in V_{\bar{w}}$. *Then,* (a) *for any* $\varepsilon < \underline{\mu}^0 / (1 - \underline{\mu}^0)$,

$$\limsup_{i \to \infty} \frac{p_\varepsilon(x_{i+1}) - p_\varepsilon(\hat{x})}{p_\varepsilon(x_i) - p_\varepsilon(\hat{x})} \le 1 - \beta \frac{m}{M} \min \{ \frac{\varepsilon}{\bar{\mu}^0(1 + \varepsilon)} , 1 \} , \tag{3.5a}$$

and (b) *if* $\psi(x_{i_0}) \le 0$ *for any* $i_0 \in \mathbf{N}$,

$$\limsup_{i \to \infty} \frac{f^0(x_{i+1}) - f^0(\hat{x})}{f^0(x_i) - f^0(\hat{x})} \leq 1 - \beta(\underline{\mu}^0/\bar{\mu}^0)\frac{m}{M} . \tag{3.5b}$$

□

Proof: (a) Let positive $\varepsilon \in (0, \underline{\mu}^0/(1-\underline{\mu}^0))$ be arbitrary. The proof of Theorem 3.1 gives us a relation between the decrease in the penalty function $p_\varepsilon(x)$ at iteration i and the decrease predicted by $\theta_\varepsilon(x)$,

$$p_\varepsilon(x_{i+1}) - p_\varepsilon(x_i) \leq \frac{\beta\gamma}{M}\,\theta_\varepsilon(x_i) . \tag{3.6}$$

for large i. Hence,

$$\limsup_{i \to \infty} \frac{p_\varepsilon(x_{i+1}) - p_\varepsilon(x_i)}{p_\varepsilon(x_i) - p_\varepsilon(\hat{x})} \leq \frac{\beta\gamma}{M} \limsup_{i \to \infty} \frac{\theta_\varepsilon(x_i)}{p_\varepsilon(x_i) - p_\varepsilon(\hat{x})} . \tag{3.7}$$

To complete our proof, we will make use of Theorem A.3 in the Appendix, which is a restatement of Lemma 3.3 of [Wie.1]. This result provides an upper bound on the right-hand side of (3.7). For this purpose, we will show that the assumptions of Theorem A.3 hold. Assumptions (i) and (ii) of Theorem 2.3 ensure that assumptions (i) and (ii) of Theorem A.3 hold with respect to the minimax problem (3.2a) at $\hat{x}$. Next we turn to assumption (iii) of Theorem A.3.

We associate with the minimax problem (3.2a) the set of multiplier vectors $U_\varepsilon(\hat{x})$ consisting of those $\mu \in \Sigma_{p+1}$ such that

$$\mu^0 \varepsilon \nabla f^0(\hat{x}) + \sum_{j \in \underline{p}} \mu^j \{ \varepsilon \nabla f^0(\hat{x}) + \nabla f^j(\hat{x}) \} = 0 , \tag{3.8a}$$

$$\mu^0 \varepsilon f^0(\hat{x}) + \sum_{j \in \underline{p}} \mu^j \{ \varepsilon f^0(\hat{x}) + f^j(\hat{x}) \} = p_\varepsilon(\hat{x}) . \tag{3.8b}$$

The sets $U_\varepsilon(\hat{x})$ and $U_{PP}(\hat{x})$ are related as follows. Since $\psi_+(\hat{x}) = 0$, (3.8a-b) can be rewritten as

$$\varepsilon \nabla f^0(\hat{x}) + \sum_{j \in \underline{p}} \mu^j \nabla f^j(\hat{x}) = 0 , \tag{3.9a}$$

$$\sum_{j \in \underline{p}} \mu^j f^j(\hat{x}) = 0 . \tag{3.9b}$$

Then, since $1 - \mu^0 = \sum_{j \in \underline{p}} \mu^j$, $(\varepsilon, \mu^1, \dots, \mu^p)/(\varepsilon + 1 - \mu^0) \in U_{GQP}(\hat{x})$, for any $\mu \in U_\varepsilon(\hat{x})$. Since $U_{GQP}(\hat{x}) = U_{PP}(\hat{x})$ as we showed in the proof of Lemma 3.2, $(\varepsilon, \mu^1, \dots, \mu^p)/(\varepsilon + 1 - \mu^0) \in U_{PP}(\hat{x})$ for any $\mu \in U_\varepsilon(\hat{x})$. It follows from assumption (iii) of Theorem 2.3, that, with H as defined in Theorem 2.3,

$$m\|h\|^2 < \langle h , \left[\frac{\varepsilon}{\varepsilon + 1 - \mu^0} F^0(\hat{x}) + \sum_{j \in \underline{p}} \frac{\mu^j}{\varepsilon + 1 - \mu^0} F^j(\hat{x}) \right] h \rangle , \quad \forall\, h \in H ,\ h \neq 0 , \tag{3.10a}$$

for any $\mu \in U_\varepsilon(\hat{x})$. Hence for any $\mu \in U_\varepsilon(\hat{x})$,

$$m_\varepsilon \|h\|^2 < \langle h \,,\, \left[\mu^0 \varepsilon F^0(\hat{x}) + \sum_{j \in \underline{p}} \mu^j \{ \varepsilon F^0(\hat{x}) + F^j(\hat{x}) \} \right] h \rangle \quad \forall\, h \in H \;,\, h \neq 0 \;, \tag{3.10b}$$

where $m_\varepsilon \triangleq \min \{ m\,(\varepsilon + 1 - \mu^0) \mid \mu \in U_\varepsilon(\hat{x}) \} = m\,(\varepsilon + 1 - \max\{ \mu^0 \mid \mu \in U_\varepsilon(\hat{x}) \})$. Hence, assumption (iii) of Theorem A.3 is satisfied at $\hat{x}$ for the minimax problem (5.1), and it therefore follows from Theorem A.3 that

$$\limsup_{i \to \infty} \frac{\theta_\varepsilon(x_i)}{p_\varepsilon(x_i) - p_\varepsilon(\hat{x})} \leq - \frac{\min\{ m_\varepsilon , (1+\varepsilon)\gamma \}}{(1+\varepsilon)\gamma} \;. \tag{3.11}$$

Combining (3.11) with (3.7) yields

$$\limsup_{i \to \infty} \frac{p_\varepsilon(x_{i+1}) - p_\varepsilon(x_i)}{p_\varepsilon(x_i) - p_\varepsilon(\hat{x})} \leq - \frac{\beta\gamma}{M} \frac{\min\{ m_\varepsilon , (1+\varepsilon)\gamma \}}{(1+\varepsilon)\gamma}$$

$$= -\beta \frac{\min\{ m_\varepsilon / (1+\varepsilon) , \gamma \}}{M} \;. \tag{3.12}$$

Next consider any $\mu \in U_\varepsilon(\hat{x})$. As mentioned above, $(\varepsilon, \mu^1, \ldots, \mu^p) / (\varepsilon + 1 - \mu^0) \in U_{PP}(\hat{x})$. Recall that $\bar{\mu}^0 \triangleq \max\{ \mu^0 \mid \mu \in U_{PP}(\hat{x}) \}$. Then

$$\frac{\varepsilon}{\varepsilon + 1 - \mu^0} \leq \bar{\mu}^0 \;, \tag{3.13}$$

and hence

$$m_\varepsilon = m(\varepsilon + 1 - \max\{ \mu^0 \mid \mu \in U_\varepsilon(\hat{x}) \}) \geq m \frac{\varepsilon}{\bar{\mu}^0} \;. \tag{3.14}$$

Substituting (3.14) into (3.12) yields

$$\limsup_{i \to \infty} \frac{p_\varepsilon(x_{i+1}) - p_\varepsilon(x_i)}{p_\varepsilon(x_i) - p_\varepsilon(\hat{x})} \leq - \frac{\beta}{M} \min\{ m\varepsilon / (\bar{\mu}^0(1+\varepsilon)) , \gamma \} \leq -\beta \frac{m}{M} \min\{ \frac{\varepsilon}{\bar{\mu}^0(1+\varepsilon)} , 1 \} \;. \tag{3.15}$$

Adding 1 to each side of the inequality in (3.15), we obtain (3.5a).

(b) Using the fact that $p_\varepsilon(x_i) = f^0(x_i)$ for $i > i_0$,

$$\limsup_{i \to \infty} \frac{f^0(x_{i+1}) - f^0(\hat{x})}{f^0(x_i) - f^0(\hat{x})} \leq 1 - \beta \frac{m}{M} \min\{ \frac{\varepsilon}{\bar{\mu}^0(1+\varepsilon)} , 1 \} \;. \tag{3.16}$$

Since $\varepsilon < \underline{\mu}^0 / (1 - \underline{\mu}^0)$ is arbitrary, (3.5b) holds. □

Unless $\hat{x}$ is also an *unconstrained* minimizer of $f^0(\cdot)$ (in which case, $\bar{\mu}^0 = 1$), the bound in (3.5b) on the cost convergence ratios of sequences constructed by Algorithm 3.1 is smaller than the bound in (2.7d) for sequences constructed by Algorithm 2.1,

$$1 - \alpha\beta(\underline{\mu}^0/\bar{\mu}^0) \frac{m}{M} \underline{\mu}^0 < 1 - \underline{\mu}^0 \frac{m}{M} \;.$$

4. GLOBALIZATION AND IMPLEMENTATION OF THE GQP SUBPROCEDURE

There are two issues associated with the use of the problem

$$\text{GQP}(x): \quad \min\{\tilde{f}^0(h \mid x) \mid \tilde{f}^j(h \mid x) \le 0,\ \forall j \in \underline{p}\}\ , \tag{4.1}$$

as a search direction subprocedure that must be resolved. The first is the issue of globalization. When x is not feasible for (1.1) and is far from a solution to (1.1), GQP(x) may not have any feasible solutions. The second is the issue of implementation. Unlike the search direction problem (2.2a) of Algorithm 2.1, GPQ(x) cannot be transformed into a quadratic program to be solved by known methods. We must find an efficient method for solving it in a neighborhood of any solution $\hat{x}$ of (1.1), where, by Lemma 3.1, GQP(x) is known to have a solution.

We will develop the globalized, implementable search direction subprocedure in three steps. First, we will show that GPQ(x) is equivalent to a problem $\text{GQ}\tilde{\text{P}}(x)$ with *linear* equality constraints and a *single* quadratic inequality constraint, determined by the constraints active at the solution to GPQ(x). Second, we will use the PMT search direction subprocedure to predict which constraints are active at the solution. This will allow us to construct a problem with linear equality constraints and a single quadratic inequality constraint which approximates $\text{GQ}\tilde{\text{P}}(x)$. We will show that, when the approximating problem has a solution, it can be easily obtained from the PMT search direction vector $h(x)$. Third, we will incorporate these observations in a search direction subprocedure which reverts to the PMT search direction when the approximating problem has no solution.

Because, the PMT search direction subprocedure correctly predicts the constraints active at the solution to GQP(x) when x is near a solution to (1.1) at which strict complimentary slackness holds, the globalized, implementable search direction subprocedure leads to a phase I - phase II algorithm which has the same robustness properties as the PMT algorithm and the same rate of convergence as the *conceptual* Algorithm 3.1.

Thus, we begin by developing an equivalent statement for GQP(x). For any $x \in \mathbb{R}^n$ and set $J \subset \underline{p}$, we define the problem

$$\text{P}(x\ ,J): \quad \min_{h \in \mathbb{R}^n}\{\tilde{f}^0(h \mid x) \mid \tilde{f}^{j_0}(h \mid x) \le 0\ ,\ \tilde{f}^j(h \mid x) = \tilde{f}^{j_0}(h \mid x)\ , \forall j \in J \setminus j_0\}\ , \tag{4.2a}$$

where $j_0 \in J$ is arbitrary. A brief inspection of (4.2a) reveals that the problem P(x , J) is independent of the selection of $j_0 \in J$. We will denote the solution to P(x , J) by $d(x\ ,J)$.

Since the functions $\tilde{f}^j(\cdot \mid x)$ all have the same quadratic term, $\frac{1}{2}\gamma|h|^2$, the equality constraints in (4.2a) are linear. Hence, problem (4.2a) requires the minimization of a quadratic function subject to *linear* equality constraints and a *single* positive-definite quadratic inequality constraint. A subproblem of this form appears in trust region methods, and efficient methods for solving it have been developed [Mor.1]. However, because $\tilde{f}^0(\cdot \mid x)$ and $\tilde{f}^{j_0}(\cdot \mid x)$ have the *same* quadratic term, a simpler technique can be used to solve (4.2a) for our choice of J (see Proposition 4.2).

Assuming that (4.1) is feasible, we define the *active constraint index set* by

$$J_{GQP}(x) \triangleq \{ j \in \underline{p} \mid \tilde{f}^j(h_{GQP}(x) \mid x) = 0 \} . \tag{4.2b}$$

(The set $J_{GQP}(x)$ may be empty.) A small amount of reflection confirms that the problem GQP(x) is equivalent to the problem P$(x\,, J_{GQP}(x))$. (Problem P$(x\,, J_{GQP}(x))$ is what we referred to above as GQ$\tilde{\mathrm{P}}(x)$.) Hence, when the set $J_{GQP}(x)$ is known, the problem GQP(x) is relatively easy to solve. Next, for any $\mu \in \Sigma_{p+1}$, let

$$J(\mu) \triangleq \{ j \in \underline{p} \mid \mu^j > 0 \} \tag{4.3a}$$

and let $\mu_{PP}(x)$ be any selection from $U_{PP}(x)$. In the following propositions, we will prove that the use of

$$J_{PP}(x) \triangleq J(\mu_{PP}(x)) \,, \tag{4.3b}$$

as an estimate of $J_{GQP}(x)$ has several desirable consequences.

The following proposition shows that $d(x\,, J_{PP}(x))$ can be obtained rather easily from $h(x)$. Recall that, for any $x \in \mathbb{R}^n$ such that GQP(x) has a solution, we denote the set of Fritz John multiplier vectors associated with the solution by $U_{GQP}(x)$ (see (3.1d)).

Proposition 4.2: *Suppose that problem* P$(x\,, J_{PP}(x))$ *has a solution* $d(x\,, J_{PP}(x))$. *Let* $j_0 \in J_{PP}(x)$ *be arbitrary, let* G_x *be a matrix with columns* $\nabla f^j(h(x) \mid x) - \nabla f^{j_0}(h(x) \mid x)$, $j \in J_{PP}(x) \setminus j_0$, *let* N_x *be a matrix whose columns form an orthonormal basis for the null space of* G_x^T, *and let* $P_x \triangleq N_x N_x^T$ *be the orthogonal projection operator whose range is the null space of* G_x^T. *Then there exists a* $\tau \in \mathbb{R}$ *such that*

$$d(x\,, J_{PP}(x)) = h(x) + \tau P_x \nabla \tilde{f}^0(h(x) \mid x) . \tag{4.4}$$

Proof: First, we rewrite P$(x\,, J_{PP}(x))$ in the form

$$\min \{ \tilde{f}^0(h \mid x) \mid \tilde{f}^{j_0}(h \mid x) \le 0 \,,\ g_x + G_x^T h = 0 \} \,, \tag{4.5a}$$

where $j_0 \in J_{PP}(x)$ is arbitrary, g_x is the vector with elements $f^j(h(x) \mid x) - f^{j_0}(h(x) \mid x)$, $j \in J_{PP}(x) \setminus j_0$. Since $g_x + G_x^T h(x) = 0$, it follows that if we set $h = h(x) + \delta h$ in (4.5a), then we must have $G_x^T \delta h = 0$, which implies that $\delta h = N_x y$ for some y. Hence, By substituting $\delta h = N_x y$ into (4.5a), the equality constraint in (4.5a) can be eliminated. Upon expansion of the functions $\tilde{f}^j(\cdot \mid x)$ around $h(x)$, (4.5a) becomes

$$\min \{ \tilde{f}^0(h(x) \mid x) + \langle \nabla \tilde{f}^0(h(x) \mid x) \,, N_x y \rangle + \tfrac{1}{2}\gamma |N_x y|^2 \mid$$

$$\tilde{f}^{j_0}(h(x) \mid x) + \langle \nabla \tilde{f}^{j_0}(h(x) \mid x) \,, N_x y \rangle + \tfrac{1}{2}\gamma |N_x y|^2 \le 0 \} . \tag{4.5b}$$

If $J_{PP}(x) = \emptyset$, then $\mu_{PP}^0 = 1$, then $\nabla \tilde{f}^0(h(x) \mid x) = 0$ and the optimal solution to (4.5a) is $\delta h(x) = 0$. Now suppose that $\nabla \tilde{f}^0(h(x) \mid x) \ne 0$. This implies that $\mu_{PP}^0(x) < 1$ and that $J_{PP}(x) \ne \emptyset$. Then the solution $\delta h(x)$ for problem (4.5b) satisfies the first-order condition

$$N_x^T \Big[\mu^0 \nabla \tilde{f}^0(h(x) \mid x) + (1 - \mu^0) \nabla \tilde{f}^{j_0}(h(x) \mid x) + \gamma \delta h(x) \Big] = 0 \,, \tag{4.5c}$$

for some $\mu^0 \in [0\,, 1]$. Since $N_x N_x^T \delta h(x) = P_x \delta h(x) = \delta h(x)$, we obtain from (4.5c) that

$$\delta h(x) = \gamma^{-1}\left[\mu^0 P_x \nabla \tilde{f}^{\,0}(h(x) \mid x) + (1-\mu^0) P_x \nabla \tilde{f}^{\,j'}(h(x) \mid x)\right]. \tag{4.5d}$$

Now, $h(x)$, the solution to (2.2a), satisfies the optimality condition $\sum_{j \in \underline{p} \cup 0} \mu^j_{PP}(x) \nabla \tilde{f}^{\,j}(h(x) \mid x) = 0$. Rearranging this equation (and dropping the dependence of μ_{PP} on x) yields

$$0 = \mu^0_{PP} \nabla \tilde{f}^{\,0}(h(x) \mid x) + (1-\mu^0_{PP}) \nabla \tilde{f}^{\,j'}(h(x) \mid x) + \sum_{j \in \underline{p}} \mu^j_{PP}\left[\nabla \tilde{f}^{\,j}(h(x) \mid x) - \nabla \tilde{f}^{\,j'}(h(x) \mid x)\right]. \tag{4.5e}$$

Applying P_x to both sides of (4.5e), we conclude that

$$0 = \mu^0_{PP} P_x \nabla \tilde{f}^{\,0}(h(x) \mid x) + (1-\mu^0_{PP}) P_x \nabla \tilde{f}^{\,j'}(h(x) \mid x) , \tag{4.5f}$$

since $P_x(\nabla \tilde{f}^{\,j}(h(x) \mid x) - \nabla \tilde{f}^{\,j'}(h(x) \mid x)) = 0$ for all $j \in \underline{p}$ by the definition of P_x. Since $\mu^0_{PP} < 1$,

$$P_x \nabla \tilde{f}^{\,j'}(h(x) \mid x) = -\frac{\mu^0_{PP}}{1-\mu^0_{PP}} P_x \nabla \tilde{f}^{\,0}(h(x) \mid x). \tag{4.5g}$$

Substituting (4.5g) into (4.5d) yields

$$\delta h(x) = \gamma^{-1}\left[\mu^0 - (1-\mu^0)\frac{\mu^0_{PP}}{1-\mu^0_{PP}}\right] P_x \nabla \tilde{f}^{\,0}(h(x) \mid x). \tag{4.5h}$$

□

The search direction $d(x, J_{PP(x)})$ may not be a feasible solution for GQP(x). The following subprocedure returns the Polak-Trahan-Mayne search direction in this case.

Search Direction Subprocedure 4.1:

Step 1: Compute the Polak-Trahan-Mayne search direction $h(x)$ and identify the set $J_{PP}(x)$.

Step 2: Compute the step, $\Delta h(x) = P_x \nabla \tilde{f}^{\,0}(h(x) \mid x)$.

Step 3: Compute $\tau \in \mathbb{R}$ by solving

$$\min \{ \tilde{f}^{\,0}(h(x) + \tau \Delta h(x) \mid x) \mid \tilde{f}^{\,j}(h(x) + \tau \Delta h(x) \mid x) \le 0 \ \ \forall j \in \underline{p} \} . \tag{4.6}$$

(If problem (4.6) is infeasible, set $\tau = 0$.)

Step 4: Set $d(x) = h(x) + \tau \Delta h(x)$. □

The minimization in Step 3 can be performed very quickly since it involves only quadratic functions of a single variable. Note that $\Delta h(x)$ of the Search Direction Subprocedure 4.1 is equal to $\tau \delta h(x)$, with $\delta h(x)$ as defined in the proof of Proposition 4.2. The following proposition summarizes the useful properties of $d(x)$.

We now prove that, if $h(x)$ is feasible for GQP(x), then $d(x)$ is a feasible direction promising as much decrease in the objective as $h(x)$. If $h(x)$ is not feasible for GQP(x), then $d(x)$ provides as much improvement in the constraint violation as $h(x)$.

Proposition 4.3:

(a) *If* $\tilde{f}^{\,j}(h(x) \mid x) \le 0$ *for each* $j \in \underline{p}$, *then* $\tilde{f}^{\,0}(d(x) \mid x) \le \tilde{f}^{\,0}(h(x) \mid x)$ *and* $\tilde{f}^{\,j}(d(x) \mid x) \le 0$ *for each* $j \in \underline{p}$.

(b) *If* $\max_{j \in \underline{p}} \tilde{f}^{\,j}(h(x) \mid x) > 0$, *then* $\max_{j \in \underline{p}} \tilde{f}^{\,j}(d(x) \mid x) \le \max_{j \in \underline{p}} \tilde{f}^{\,j}(h(x) \mid x)$.

(c) *If* GQP(x) *is feasible and* $J_{PP}(x) = J_{GQP}(x)$, *then* $d(x)$ *solves* GQP(x).

Lemma 5.1 shows that the assumptions of Proposition 4.3(c) hold in a neighborhood of a solution $\hat{x}$ to (1.1), provided that strict complementary slackness holds at $\hat{x}$.

Proof: (a) This follows from the fact $\tau = 0$ is feasible for the single-variable minimization in Step 4.

(b) If problem (4.6) is feasible, then

$$\max_{j \in \underline{p}} \tilde{f}^j(d(x) \mid x) = 0 \le \max_{j \in \underline{p}} \tilde{f}^j(h(x) \mid x). \tag{4.7}$$

If problem (4.6) is infeasible, $d(x) = h(x)$.

(c) Since $J_{PP}(x) = J_{GQP}(x)$, $d(x, J_{PP}(x))$ solves GQP(x). We show that Algorithm 4.1 computes $d(x, J_{PP}(x))$. Since $d(x, J_{PP}(x))$ minimizes $\tilde{f}^0(\cdot \mid x)$ over $\{h \in \mathbb{R}^n \mid \tilde{f}^j(h \mid x) \le 0, j \in \underline{p}\}$,

$$\tilde{f}^0(d(x, J_{PP}(x))) = \min_{h \in \mathbb{R}^n} \{\tilde{f}^0(h \mid x) \mid \tilde{f}^j(h \mid x) \le 0, j \in \underline{p}\}$$

$$\le \min_{\tau \in \mathbb{R}} \{\tilde{f}^0(h(x) + \tau\Delta h(x) \mid x) \mid \tilde{f}^j(h(x) + \tau\Delta h(x) \mid x) \le 0, j \in \underline{p}\}.$$

Since $d(x, J_{PP}(x))$ can be expressed as $h(x) + \tau_0\Delta h(x) \mid x)$ for some $\tau_0 \in \mathbb{R}$, problem (4.6) is feasible and has the solution τ_0. Therefore, $d(x) = d(x, J_{PP}(x))$. □

5. A STABILIZED IMPLEMENTABLE GQP-BASED ALGORITHM

We replace Step 2 of Algorithm 3.1 with the Search Direction Subprocedure 4.1 to obtain a global phase I - phase II method, and we establish its convergence properties.

Algorithm 5.1:

Data: x_0; $\beta \in (0, 1)$; $\gamma > 0$; $i = 0$.

Step 1: Compute a search direction $d_i = d(x_i)$ by means of Search Direction Subprocedure 4.1.

Step 2: Compute a step size,

$$\lambda_i = \max_{k \in \mathbb{N}} \{ \beta^k \mid f^0(x_i + \beta^k d_i) - f^0(x_i) \le \beta^k \tilde{f}^0(d_i \mid x_i),$$

$$\psi_+(x_i + \beta^k d_i) - \psi_+(x_i) \le \beta^k [\max_{j \in \underline{p}} \{\tilde{f}^j(d_i \mid x_i), 0\} - \psi_+(x_i)] \}. \tag{5.1}$$

Step 3: Set $x_{i+1} = x_i + \lambda_i d_i$.

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

The three cases listed in Theorem 5.1 are exhaustive. In case (b), $\theta(\bar{x}) = 0$ implies that $0 \in \partial\psi(\bar{x})$, where $\partial\psi(x)$ denotes the generalized gradient of $\psi(\cdot)$ at x. This case is normally ruled out by assumption. The convergence result obtained for Algorithm 5.1 is slightly weaker than that obtained for Algorithm 2.1 in Theorem 2.2. In case (c), where Algorithm 5.1 constructs a sequence which remains infeasible but has feasible accumulation points, not all of the accumulation points are guaranteed to be stationary points of problem (1.1).

Theorem 5.1: *Suppose that the functions $f^j(\cdot)$ in (1.1) have continuous derivatives, that Algorithm 5.1 constructs a sequence $\{x_i\}_{i=0}^{\infty}$ in solving (1.1), and that $\bar{x}$ is an accumulation point of the sequence.*

(a) *If there exists an $i_0 \in \mathbb{N}$ such that $\psi(x_{i_0}) \le 0$, then $\theta(\bar{x}) = 0$.*

(b) *If $\psi(x_i) > 0$ for all $i \in \mathbb{N}$ and $\psi(\bar{x}) > 0$, then $\theta(\bar{x}) = 0$.*

(c) *If $\psi(x_i) > 0$ for all $i \in \mathbb{N}$ and $\psi(\bar{x}) = 0$, then $\liminf_{i \to \infty} |\theta(x_i)| = 0$.*

Proof: First we derive bounds on $\|d(x)\|$ for use in the proof of parts (a) and (b). Suppose that the subsequence $\{x_i\}_{i \in K}$ converges to $\bar{x}$, for some subset $K \subset \mathbb{N}$, and that $\theta(\bar{x}) \neq 0$. By Theorem 2.1, $\theta(\cdot)$ is continuous, and, by (4.2b), $\theta(x) \le 0$ for all $x \in \mathbb{R}^n$. Therefore there exists a $\delta > 0$ and a neighborhood, W_0, of $\bar{x}$ such that

$$\theta(x) = \max_{j \in \underline{p} \cup 0} \{\tilde{f}^j(h(x) \mid x)\} - \psi_+(x) < -\delta\,, \tag{5.2a}$$

for all $x \in W_0$. We use this fact and Proposition 4.3 to show that $\|d(x)\| > 0$ for all x in a neighborhood of $\bar{x}$.

Suppose that $\psi(\bar{x}) \le 0$. In view of (5.2a), there exists a neighborhood, $W_1 \subset W_0$, of $\bar{x}$, such that $\psi(x) < \frac{1}{2}\delta$ for all $x \in W_1$. Then, for $x \in W_1$,

$$\max_{j \in \underline{p}} \tilde{f}^j(h(x) \mid x) \le \theta(x) + \psi_+(x) \le -\tfrac{1}{2}\delta < 0\,. \tag{5.2b}$$

From Proposition 4.3(a), we have that

$$\tilde{f}^0(d(x) \mid x) - \psi_+(x) \le \tilde{f}^0(h(x) \mid x) - \psi_+(x) \le \theta(x) < -\delta\,, \tag{5.2c}$$

for all $x \in W_0$. Since $\psi(x) < \frac{1}{2}\delta$, if follows from (5.2b) and (5.2c) that $\tilde{f}^0(d(x) \mid x) < -\frac{1}{2}\delta$ for all $x \in W_1$. Hence, since $\tilde{f}^0(0 \mid x) = 0$, and since $\tilde{f}^0(h \mid x)$ is continuous in h, uniformly in x, there exists $b' > 0$ such that $\|d(x)\| > b'$ for all $x \in W_1$.

Now suppose that $\psi(\bar{x}) > 0$. We proceed in a manner similar to that in the previous paragraph. There exists a neighborhood, $W_2 \subset W_0$, of $\bar{x}$, such that $\psi(x) > \frac{1}{2}\psi(\bar{x})$ for each $x \in W_2$. For each $x \in W_2$, either $\max_{j \in \underline{p}} \tilde{f}^j(h(x) \mid x) > 0$, or else $\max_{j \in \underline{p}} \tilde{f}^j(h(x) \mid x) \le 0$. In the former case, it follows from Proposition 4.3(b) that

$$\max_{j \in \underline{p}} \tilde{f}^j(d(x) \mid x) - \psi_+(x) \le \max_{j \in \underline{p}} \tilde{f}^j(h(x) \mid x) - \psi_+(x) < -\delta\,. \tag{5.2d}$$

In the latter case, it follows from Proposition 4.3(b) that

$$\max_{j \in \underline{p}} \tilde{f}^j(d(x) \mid x) - \psi_+(x) < 0 - \psi_+(x) = -\tfrac{1}{2}\psi(\bar{x})\,, \tag{5.2e}$$

for all $x \in W_2$. Therefore, for all $x \in W_2$, $\max_{j \in \underline{p}} \tilde{f}^j(d(x) \mid x) - \psi_+(x) < -\min\{\delta, \frac{1}{2}\psi(\bar{x})\}$. Hence, since $\max_{j \in \underline{p}} \tilde{f}^j(0 \mid x) - \psi_+(x) = 0$ for $x \in W_2$, and since the function $\max_{j \in \underline{p}} \tilde{f}^j(h \mid x)$ is continuous in h, uniformly in x, there exists $b \in (0, b')$ such that $\|d(x)\| > b$ for all $x \in W_2$.

Because the functions $\tilde{f}^k(\cdot \mid x)$ are strongly convex in h, uniformly in x, $\|d(x)\|$ is also bounded from above in W_2. Because $\|d(x)\|$ is bounded on W_2 and the gradients $\nabla f^j(\cdot)$ are continuous, there exist $\bar{\lambda} > 0$ and a neighborhood, W_3, of $\bar{x}$, such that $\|\int_0^1 \nabla f^j(x + s\lambda d(x))ds - \nabla f^j(x)\| < \frac{1}{2}\gamma b$ for all $x \in W_3$, $\lambda \in [0, \bar{\lambda}]$ and $j \in \underline{p} \cup 0$. (We assume without loss of generality that $W_3 \subset W_1$ if $\psi(\bar{x}) \le 0$ and that $W_3 \subset W_2$ if $\psi(\bar{x}) > 0$.)

(a) Suppose that $\psi(x_{i_0}) \le 0$ for some $i_0 \in \mathbb{N}$. (This implies that $\psi(x_i) \le 0$ for all $i \ge i_0$ and that $\psi(\bar{x}) \le 0$.) Then there exists an $i_1 \ge i_0$ such that $x_i \in W_3$ for all $i \ge i_1$, $i \in K$. For $i > i_1$, $i \in K$ and $\lambda \in (0, \bar{\lambda}]$,

$$
\begin{aligned}
f^0(x_i + \lambda d_i) - f^0(x_i) &= \langle \nabla f^0(x_i), \lambda d_i \rangle + \langle \int_0^1 [\nabla f^0(x_i + s\lambda d_i) - \nabla f^0(x_i)]ds , \lambda d_i \rangle \\
&\le \lambda \{ \langle \nabla f^0(x_i), d_i \rangle + \|d_i\| \|\int_0^1 [\nabla f^j(x_i + s\lambda d_i) - \nabla f^j(x_i)]ds\| \} \\
&\le \lambda \{ \langle \nabla f^0(x_i), d_i \rangle + \tfrac{1}{2}\gamma b \|d_i\| \} \\
&\le \lambda \{ \langle \nabla f^0(x_i), d_i \rangle + \tfrac{1}{2}\gamma \|d_i\|^2 \} = \lambda \tilde{f}^0(d_i \mid x_i) . \qquad (5.3a)
\end{aligned}
$$

Similarly, for $\lambda \in (0, \bar{\lambda}]$, $i > i_1$, $i \in K$, and $j \in \underline{p}$,

$$
\begin{aligned}
f^j(x_i + \lambda d_i) &\le \lambda \{ f^j(x_i) + \langle \nabla f^j(x_i), d_i \rangle + \|d_i\| \|\int_0^1 [\nabla f^j(x_i + s\lambda d_i) - \nabla f^j(x_i)]ds\| \} \\
&\le \lambda \{ f^j(x_i) + \langle \nabla f^j(x_i), d_i \rangle + \tfrac{1}{2}\gamma b \|d_i\| \} \le \lambda \tilde{f}^j(d_i \mid x_i) . \qquad (5.3b)
\end{aligned}
$$

Taking the maximum over $j \in \underline{p}$, and using the fact that $\psi_+(x_i) = 0$, we obtain from (5.3b) that

$$
\psi_+(x_i + \lambda d_i) - \psi_+(x_i) \le \lambda [\max_{j \in \underline{p}} \{ \tilde{f}^j(d_i \mid x_i), 0 \} - \psi_+(x_i)] , \qquad (5.3c)
$$

for $i > i_1$, $i \in K$ and $\lambda \in (0, \bar{\lambda}]$. It follows from (5.3a), (5.3c) and Step 2 of Algorithm 5.1 that $\lambda_i > \beta\bar{\lambda}$ for $i > i_1$, $i \in K$. By Proposition 4.3(a), $\tilde{f}^0(d_i \mid x_i) \le \theta(x_i)$ for $i > i_1$, $i \in K$, and hence

$$
f^0(x_i + \lambda_i d_i) - f^0(x_i) \le \lambda_i \tilde{f}^0(d_i \mid x_i) \le \lambda_i \theta(x_i) \le -\tfrac{1}{2}\beta\bar{\lambda}\delta , \qquad (5.3d)
$$

for $i > i_1$, $i \in K$.

However, this is impossible, since $f^0(x_i)$ is monotone decreasing for $i \ge i_1$ and $f^0(x_i) \xrightarrow{K} f^0(\bar{x})$, as $i \xrightarrow{K} \infty$. Thus, the necessary condition (2.5a-b), must be satisfied at $\bar{x}$ in this case.

(b) Now suppose that $\psi(x_i) > 0$ for all $i \in \mathbb{N}$ and that $\psi(\bar{x}) > 0$. Then there exists an $i_1 \in \mathbb{N}$ such that $x_i \in W_3$ for all $i \ge i_1$, $i \in K$. For any $x \in \mathbb{R}^n$ such that $\tilde{f}^j(h(x) \mid x) \le 0$ for each $j \in \underline{p}$, $\tilde{f}^j(d(x) \mid x) \le 0$ for each $j \in \underline{p}$ by Proposition 4.3(a). For any $x \in \mathbb{R}^n$ such that $\max_{j \in \underline{p}} \tilde{f}^j(h(x) \mid x) > 0$, $\max_{j \in \underline{p}} \tilde{f}^j(d(x) \mid x) \le \max_{j \in \underline{p}} \tilde{f}^j(h(x) \mid x) \le 0$ by Proposition 4.3(b). Therefore, $\tilde{f}^j(d(x) \mid x) \le \psi(x)$ for all $j \in \underline{p}$ and $x \in \mathbb{R}^n$. Hence,

$$
f^j(x_i + \lambda d_i) - \psi_+(x_i) = f^j(x_i) + \langle \nabla f^j(x_i), \lambda d_i \rangle + \langle \int_0^1 [\nabla f^j(x_i + s\lambda d_i) - \nabla f^j(x_i)]ds , \lambda h(x_i) \rangle - \psi_+(x_i)
$$

$$\leq \lambda \{ f^j(x_i) + \langle \nabla f^j(x_i), d_i \rangle + \tfrac{1}{2}\gamma b \|d_i\| - \psi_+(x_i) \}$$

$$\leq \lambda \{ f^j(x_i) + \langle \nabla f^j(x_i), d_i \rangle + \tfrac{1}{2}\gamma \|d_i\|^2 - \psi_+(x_i) \}$$

$$= \lambda \{ \tilde{f}^j(d_i \mid x_i) - \psi_+(x_i) \} \ , \tag{5.4a}$$

for all $i > i_1, i \in K, \lambda \in (0, \bar{\lambda}]$, and $j \in \underline{p}$. Taking the maximum over $j \in \underline{p}$, and using the fact that $0 - \psi_+(x_i) \leq \lambda[\max_{j \in \underline{p}} \tilde{f}^j(d_i \mid x_i) - \psi_+(x_i)]$,

$$\psi_+(x_i + \lambda d_i) - \psi_+(x_i) \leq \lambda[\max_{j \in \underline{p}} \{ \tilde{f}^j(d_i \mid x_i), 0 \} - \psi_+(x_i)] . \tag{5.4b}$$

Similarly,

$$f^0(x_i + \lambda d_i) - f^0(x_i) \leq \lambda \tilde{f}^0(d_i \mid x_i) \ , \tag{5.4c}$$

for all $i > i_1, i \in K, \lambda \in (0, \bar{\lambda}]$, and $j \in \underline{p}$. It follows from (5.4b), (5.4c) and Step 2 of Algorithm 5.1 that $\lambda_i > \beta\bar{\lambda}$ for $i > i_1, i \in K$.

From Proposition 4.3(b), if $\max_{k \in \underline{p}} \tilde{f}^k(h(x_i) \mid x_i) > 0$,

$$\max_{k \in \underline{p}} \tilde{f}^k(d_i \mid x_i) - \psi_+(x_i) \leq \max_{k \in \underline{p}} \tilde{f}^k(h_i \mid x_i) - \psi_+(x_i) \leq \theta(x_i) \leq -\delta . \tag{5.4d}$$

Otherwise, $\max_{k \in \underline{p}} \tilde{f}^k(h(x_i) \mid x_i) \leq 0$, which, together with Proposition 4.3(a), implies that

$$\max_{k \in \underline{p}} \tilde{f}^k(d_i \mid x_i) - \psi_+(x_i) \leq 0 - \psi_+(x_i) . \tag{5.4e}$$

There exists $i_2 > i_1$ such that $\psi_+(x_i) > \tfrac{1}{2}\psi_+(\bar{x})$ for $i > i_2, i \in K$. Substituting (5.4d) and (5.4e) into (5.4f),

$$\psi_+(x_i + \lambda d_i) - \psi_+(x_i) \leq -\lambda_i \min \{ \psi_+(x_i), \delta \} \leq -\beta\bar{\lambda}\min \{ \tfrac{1}{2}\psi_+(\bar{x}), \delta \} \ , \tag{5.4f}$$

for $i > i_1, i \in K$.

Since $\psi(x_i)$ is monotone decreasing, (5.4f) implies that $\psi(x_i) \xrightarrow{K} -\infty$ as $i \xrightarrow{K} \infty$. However, this is impossible, since $\psi(x_i) \to \psi(\bar{x})$ as $i \to \infty$, Therefore, the necessary condition (2.5a-b) must be satisfied at $\bar{x}$.

(c) Now suppose that $\psi(x_i) > 0$ for all $i \in \mathbf{N}$ and that $\psi(\bar{x}) = 0$. In this case, we do not show that $\theta(\bar{x}) = 0$, but merely that $\liminf_{i \to \infty} |\theta(x_i)| = 0$.

To obtain a contradiction, suppose that $\liminf_{i \to \infty} \theta(x_i) < -\delta' < 0$. Then there exists $i_1 \in \mathbf{N}$ such that $\theta(x_i) < -\delta'$ for all $i > i_1$. By Proposition 4.3(a-b),

$$\max_{j \in \underline{p}} \tilde{f}^j(d_i \mid x_i) \leq \max \{ 0, \max_{j \in \underline{p}} \tilde{f}^j(h(x_i) \mid x_i) \}$$

$$\leq \max \{ 0, \theta(x_i) + \psi_+(x_i) \} . \tag{5.5a}$$

Hence

$$\max_{j \in \underline{p}} \tilde{f}^j(d_i \mid x_i) - \psi_+(x_i) \le \max\{-\psi_+(x_i), \theta(x_i)\} \le \max\{-\psi_+(x_i), -\delta'\} < 0 , \tag{5.5b}$$

for all $i > i_1$. This implies that $\psi_+(x_i)$ is monotone decreasing, and, since $\psi(\bar{x}) = 0$, the sequence $\{\psi_+(x_i)\}_{i \in \mathbb{N}}$ converges to 0. Therefore, there exists $i_2 > i_1$ such that $\psi_+(x_i) < ½\,\delta'$ for all $i > i_2$. Hence,

$$\max_{j \in \underline{p}} \tilde{f}^j(h(x_i) \mid x_i) \le \theta(x_i) + \psi_+(x_i) \le -\delta' + ½\delta' < 0 , \tag{5.5c}$$

for all $i > i_2$. From Proposition 4.3(a), then,

$$\tilde{f}^0(d_i \mid x_i) \le \tilde{f}^0(h(x_i) \mid x_i) \le \theta(x_i) + \psi_+(x_i) \le -½\delta' , \tag{5.5d}$$

for all $i > i_2$. This implies that $f^0(x_i)$ is monotone decreasing for for $i > i_2$.

Now we use the fact that $\bar{x}$ is an accumulation point of the sequence $\{x_i\}_{i \in \mathbb{N}}$. It follows from an argument similar to the ones used in parts (a) and (b) that there exists $\underline{\lambda} > 0$ such that $\lambda_i > \underline{\lambda}$ for all $i > i_2, i \in K$. Combining this fact with (5.5d) and Step 2 of Algorithm 5.1,

$$f^0(x_{i+1}) - f^0(x_i) \le -½\delta'\underline{\lambda} , \tag{5.5e}$$

for $i > i_1, i \in K$. Since $f^0(x_i)$ is monotonically decreasing, (5.5e) implies that $f^0(x_i) \to -\infty$ as $i \to \infty$. This is impossible, however, since $f^0(x_i) \xrightarrow{K} f^0(\bar{x})$ as $i \to \infty$. The contradiction proves that $\liminf_{i \to \infty} |\theta(x_i)| = 0$. □

Recall the definitions of $J_{\text{GQP}}(x)$ and $J_{\text{PP}}(x)$ in (4.2b) and (4.3b) respectively, and that $h_{\text{GQP}}(x)$ denotes the solution to GQP(x).

Lemma 5.1: *Suppose that assumptions* (i)-(iii) *of Theorem 2.3 hold, and that* (iv) *strict complementary slackness holds at the solution, $\hat{x}$, of (1.1), (i.e. - for every $\mu \in U(\hat{x})$ and $j \in \underline{p}$, $\mu^j > 0$* if and only if $f^j(\hat{x}) = 0$). *Then, there exist a neighborhood, V'', of $\hat{x}$, $h' \in \mathbb{R}^n$ and $\delta > 0$ such that, for all $x \in V''$,* (a) $J_{\text{PP}}(x) = J_{\text{GQP}}(x)$, *and* (b) $d(x) = h_{\text{GQP}}(x)$.

Proof: First we observe that assumption (ii) of Theorem 2.3 implies that $\psi(\hat{x}) \le 0$. Assumption (iv), above, implies that $U_{\text{PP}}(\hat{x})$ is a singleton $\{\hat{\mu}\}$ for some $\hat{\mu} \in \Sigma_{p+1}$, and hence that $\hat{J} = J(\hat{\mu}) = \{ j \in \underline{p} \mid f^j(\hat{x}) = 0 \}$. Let V be as defined in Lemma 3.1.

(a) Because (i) $U_{\text{PP}}(\hat{x}) = \{\hat{\mu}\}$, (ii) $U_{\text{PP}}(\cdot)$ is an upper semicontinuous, compact-valued set-valued map, and (iii) $\hat{\mu}^j > 0$ for all $j \in \hat{J}$, there exists a neighborhood $W_0 \subset V$ of $\hat{x}$ such that $\mu^j > 0$ for every $j \in \hat{J}$ and $\mu \in U_{\text{PP}}(W_0)$. From the definition of $J_{\text{PP}}(x)$ in (4.3b), $J_{\text{PP}}(x) \supset \hat{J}$ for all $x \in W_0$. Now we show that $J_{\text{PP}} \subset \hat{J}$. By strict complementary slackness, $f^j(\hat{x}) < 0$ for every $j \notin \hat{J}$. Since $h(\hat{x}) = 0$ and $h(\cdot)$ is continuous [Pol.1], there exists a neighborhood, $W_1 \subset W_0$, of $\hat{x}$ such that $\tilde{f}^j(h(x) \mid x) - \psi_+(x) < 0$ for all $j \notin \hat{J}$ and $x \in W_1$. It follows from the definition of $U_{\text{PP}}(x)$ that $\mu^j = 0$ for every $j \notin \hat{J}$ and every $\mu \in U_{\text{PP}}(W_1)$. Hence $j \notin \hat{J}$ implies $j \notin J_{\text{PP}}(W_1)$. Therefore, $J_{\text{PP}}(x) = \hat{J}$ for every $x \in W_1$.

By a similar argument, we show that $J_{\text{GQP}}(x) = \hat{J}$ for all x contained in a neighborhood of

$\hat{x}$. (i) Since $h_{GQP}(\hat{x}) = 0$ and $\psi_+(\hat{x}) = 0$, an inspection of (3.1d) reveals that $U_{GQP}(\hat{x}) = U_{PP}(\hat{x}) = \{\hat{\mu}\}$. (ii) Lemma 3.1 implies that $h_{GQP}(x)$ is continuous in W_1, and hence $U_{GQP}(x)$ is an upper semicontinuous, compact-valued set-valued map. (iii) For all $j \in \hat{J}$, $\hat{\mu}^j > 0$. Hence, there exists a neighborhood, $W'_0 \subset V$, of $\hat{x}$ such that $\mu^j > 0$ for every $j \in \hat{J}$ and $\mu \in U_{GQP}(W'_0)$. From the definition of $U_{GQP}(x)$ in (3.1d), this implies that $\tilde{f}(h_{GQP}(x) \mid x) = 0$ for $j \in \hat{J}$ and $x \in W'_0$. Hence, by the definition of $J_{GQP}(x)$ in (4.2b) $J_{GQP}(x) \supset \hat{J}$ for every $x \in W'_0$. Now we show that $J_{GQP}(x) \subset \hat{J}$. By strict complementary slackness, $f^j(\hat{x}) < 0$ for every $j \notin \hat{J}$. Since $h_{GQP}(\hat{x}) = 0$ and $h_{GQP}(\cdot)$ is continuous, there exists a neighborhood $W'_1 \subset W'_0$ of $\hat{x}$ such that $\tilde{f}^j(h_{GQP}(x) \mid x) < 0$ for every $j \notin \hat{J}$ and $x \in W'_1$. From the definition of $U_{GQP}(x)$, $\mu^j = 0$ for every $j \notin \hat{J}$ and every $\mu \in U_{GQP}(W'_1)$. Hence $j \notin \hat{J}$ implies that $j \notin J_{GQP}(W'_1)$. Therefore, $J_{GQP}(x) = \hat{J}$ for every $x \in W'_1$. Statement (a) holds with $V'' = W_1 \cap W'_1$.

(b) This follows from (a) and Proposition 4.3(c). □

The following theorem asserts that, under an additional strict complementarity assumption, the implementable Algorithm 5.1 has the same asymptotic rate of convergence as Local Algorithm 3.1. Without the strict complementarity assumption, the bound on the cost convergence ratio which can be obtained for Algorithm 5.1 is the same as that obtained for Algorithm 2.1 in Theorem 2.3. However, an improved bound is not obtained for Algorithm 2.1 under this additional assumption. Under the strict complementarity assumption, $U_{PP}(\hat{x}) = \{\hat{\mu}\}$ for some $\hat{\mu} \in \Sigma_{p+1}$ and hence $\underline{\mu}^0 = \bar{\mu}^0 = \hat{\mu}^0$.

Theorem 5.2: *Suppose that assumptions* (i)-(iii) *of Theorem 2.3 hold, that* (iv) *strict complementary slackness holds at* $(\hat{x}, \hat{\mu})$ *for every* $\hat{\mu} \in U_{PP}(\hat{x})$, *(i.e. - for every* $j \in \underline{p}$, $\hat{\mu}^j > 0$ if and only if $f^j(\hat{x}) = 0$*), and that Algorithm 5.1 constructs a sequence* $\{x_i\}_{i=0}^{\infty}$ *in solving (1.1). Then,* (a) $x_i \to \hat{x}$ *as* $i \to \infty$, (b) *for any* $\varepsilon < \hat{\mu}^0 / (1 - \hat{\mu}^0)$,

$$\limsup_{i \to \infty} \frac{p_\varepsilon(x_{i+1}) - p_\varepsilon(\hat{x})}{p_\varepsilon(x_i) - p_\varepsilon(\hat{x})} \le 1 - \beta \frac{m}{M} \min\{ \frac{\varepsilon}{\hat{\mu}^0(1+\varepsilon)}, 1 \} , \tag{5.6a}$$

and (c) *if* $\psi(x_{i_0}) \le 0$ *for any* $i_0 \in \mathbf{N}$,

$$\limsup_{i \to \infty} \frac{f^0(x_{i+1}) - f^0(\hat{x})}{f^0(x_i) - f^0(\hat{x})} \le 1 - \beta \frac{m}{M} . \tag{5.6b}$$

□

Proof: (a) The sequence lies in the bounded set L defined in assumption (ii) of Theorem 2.3, and hence it converges to the set of its accumulation points. By Theorem 5.1, $\liminf_{i \to \infty} |\theta(x_i)| = 0$.

We prove that $\hat{x}$ must be an accumulation point. Suppose not. Then there exists a neighborhood W of $\hat{x}$ such that $\{x_i\}_{i \in \mathbf{N}} \subset L \setminus W$. By assumption (ii) of Theorem 2.3, there is no point in $L \setminus W$ which satisfies (2.5a-b). Since $L \setminus W$ is compact, this and Theorem 2.1(b) imply that $\inf\{ \theta(x) \mid x \in L \setminus W \} > 0$. But this contradicts the fact that $\liminf_{i \to \infty} |\theta(x_i)| = 0$. Hence

$\hat{x}$ must be an accumulation point.

Let V'' be as defined in Lemma 5.1. The iteration maps (see the proof of Theorem 3.1) of Algorithms 3.1 and 5.1 coincide for $x \in V''$. By Theorem 3.1, there exists a neighborhood $V''' \subset V''$ of $\hat{x}$ such that if the sequence $\{x_i\}_{i \in \mathbf{N}}$ enters V''', it remains in V''' and converges to $\hat{x}$. Since $\hat{x}$ is an accumulation point of the sequence, it must enter V'''. Hence, the sequence converges to $\hat{x}$.

(b) and (c) Since, by (a), $\{x_i\}_{i \in \mathbf{N}}$ converges to $\hat{x}$ and the iteration map of Algorithm 5.1 coincides with that of Algorithm 3.1 in the neighborhood V'' of $\hat{x}$, the results of Theorem 3.2 hold. Since $U_{\mathrm{PP}}(\hat{x}) = \{\hat{\mu}\}$, $\underline{\mu}^0 = \bar{\mu}^0 = \hat{\mu}^0$. □

6. NUMERICAL RESULTS

Algorithm 5.1 was compared with Algorithm 2.1 and the feasible descent algorithm in [Her.1] (which also satisfies (1.3)) on several well-known inequality-constrained problems. Table 1 summarizes the performances of the three algorithms on these problems. The results for the algorithm of [Her.1] are quoted from that paper. The abbreviations in the table have the following meanings:

NF: Number of objective function evaluations.

NG: Number of constraint function evaluations.

NDF: Number of gradient evaluations of the objective function.

NDG: Number of gradient evaluations of the constraints.

Each constraint was counted separately in the tabulation of NG and NDG. Bounds on the variables, i.e., $x^j \le 0$, were not included in the tabulation.

The algorithm parameters for both Algorithms 2.1 and 5.1 were set at $\alpha = 0.9$, $\beta = 0.9$, $\gamma = 1.0$ in the experiments. To reduce the number of trial step sizes tested in the Armijo step rule, quadratic interpolation was used at each iteration of both algorithms to determine the initial trial step size.

The Rosen-Suzuki problem is problem 43 in [Hoc.1]. See Figure 1 for a comparison of the performance of Algorithms 2.1 and 5.1. (The y-axis label "Cost Error" of the figures refers to the quantity, $f^0(x_i) - f^0(\hat{x})$). Colville's Test Problems One and Two are problems 86 and 117, respectively, in [Hoc.1].

Kuhn-Tucker Problem [Con.1]: This problem has a unique minimizer at which neither the Kuhn-Tucker constraint qualification nor the Mangasarian-Fromovitz constraint qualification holds. It serves as a test of algorithm robustness. The minimum value of -1 occurs at $\hat{x} = (0, 1)$. Both algorithms converged to the solution from the feasible initial point $x_0 = (0.25, 0.25)$. However, Algorithm 2.1 converged sublinearly, while Algorithm 5.1 converged linearly. See Figure 2.

Circular-Quadratic Problem: In this problem, the function approximations are exact for $\gamma = 1$, that is, $\tilde{f}^j(h \mid x) = f^j(x+h)$ for $j \in \underline{p} \cup 0$.

$$\min \{ \tfrac{1}{2}(x_1^2 + (x_2+4)^2) \mid \tfrac{1}{2}((x_1+1)^2 + x_1^2) - 2 \le 0 \ , \ \tfrac{1}{2}((x_1-1)^2 + x_2^2) - 2 \le 0 \} \ . \tag{6.3}$$

The minimum value of 4.5 occurs at $\hat{x} = (0, -1)$; the feasible initial point $x_0 = (1, 1)$ was used.

Infeasible Problem: This simple problem was constructed to demonstrate the behavior of the algorithms when the constraints cannot be satisfied.

$$\min \{ -x_1 \mid (x_1+10)^2 + x_2^2 \le 0 \ , (x_1-10)^2 + x_2^2 \le 0 \} \ . \tag{6.2}$$

The minimum value of 1 occurs at the origin. Both Algorithms 2.1 and 5.1 converged to the solution from the initial point $x_0 = (-10, -20)$.

7. CONCLUSION

We obtained a bound on the cost convergence ratio of sequences constructed by Algorithm 5.1 which is smaller than that obtained for Algorithm 2.1. On all of the standard problems on which they were tested, Algorithm 5.1 far surpassed the performance of Algorithm 2.1 and was competitive with the first-order feasible descent algorithm of [Her.1]. Search Direction Subprocedure 4.1 was developed as a method for approximating the solution to the GQP subproblem. The above facts show that the subprocedure can profitably be viewed as a speed-enhancing correction to the method of centers search direction (2.2a).

8. APPENDIX

The following two Theorems are special cases of Theorems 4.6 and 4.9 of [Han.1], used in the proof of Lemma 3.2.

Theorem A.1 [Han.1]: *Consider the problem*

$$\min_{x \in \mathbb{R}^n} \{ g^0(x) \mid g^j(x) \le 0, \ \forall j \in \underline{p} \} \ , \tag{A.1}$$

and suppose that the functions $g^j(\cdot)$ are twice continuously differentiable.

If $\bar{x} \in \mathbb{R}^n$, together with a Kuhn-Tucker multiplier vector $\bar{u} \in \mathbb{R}_+^m$, satisfies the standard second-order sufficiency conditions [McC.1], then, for any $\varepsilon < 1/|\bar{u}|_1$, $\bar{x}$ is a strict local minimizer of the function $\varepsilon g^0(\cdot) + \max_{j \in \underline{p}} g^j(\cdot)$. □

Theorem A.2 [Han.1]: *Consider the problem (A.1) and suppose that (i) the functions $g^j(\cdot)$ are convex and continuously differentiable, and (ii) there exists $x' \in \mathbb{R}^n$ such that $g^j(x') < 0$ for all $j \in \underline{p}$.*

If $\bar{x} \in \mathbb{R}^n$, together with a Kuhn-Tucker multiplier vector $\bar{u} \in \mathbb{R}_+^m$, satisfies the standard second-order sufficiency conditions [McC.1], then, for any $\varepsilon < 1/|\bar{u}|_1$, $\bar{x}$ is a global minimizer of the function $\varepsilon g^0(\cdot) + \max_{j \in \underline{p}} g^j(\cdot)$. □

The following theorem is a restatement of Lemma 3.3 in [Wie.1].

Theorem A.3 [Wie.1]: *Suppose that*

(i) *for* $j \in \underline{q}$, *the functions* $g^j : \mathbb{R}^n \to \mathbb{R}$ *are twice continuously differentiable, and that* $\varepsilon > 0$ *is given,*

(ii) *there exists* $T \in \mathbb{R}$ *such that the set* $S \triangleq \{ x \in \mathbb{R}^n \mid \phi(x) \le T , j \in \underline{q} \}$, *where* $\phi(x) \triangleq \max_{j \in \underline{q}} g^j(x)$, *is bounded and contains a single point* $\hat{x}$ *such that*

$$\sum_{j \in \underline{q}} \mu^j \nabla g^j(\hat{x}) = 0 \,, \tag{A.2a}$$

and

$$\sum_{j \in \underline{q}} \mu^j g^j(\hat{x}) = \phi(\hat{x}) \,, \tag{A.2b}$$

for some $\mu \in \Sigma_q$.

Let J^* *be the union of the sets* $J(\mu)$, *taken over all* $\mu \in \Sigma_q$ *which, together with* $\hat{x}$, *satisfy (A.2a-b). Let* B *denote the null space of the matrix with columns* $\{ \nabla g^j(\hat{x}) \}_{j \in J^*}$. *Suppose also that*

(iii) *there exists* $r > 0$ *such that, for all* $\hat{\mu} \in U(\hat{x})$,

$$r \| h \|^2 < \langle h \,, \left[\sum_{j \in \underline{q}} \hat{\mu}^j G^j(\hat{x}) \right] h \rangle \quad \forall h \in B \,, \tag{A.3}$$

where $G^j(x)$ *denotes the second derivative matrix of* $g^j(x)$. *Then,*

$$\limsup_{x \to \hat{x}} \frac{\Phi(x)}{\phi(x) - \phi(\hat{x})} \le - \frac{\min\{ r \,, \rho \}}{\rho} \,, \tag{A.4}$$

where $\Phi(x) \triangleq \min_{h \in \mathbb{R}^n} \max_{j \in \underline{q}} g^j(x) + \langle \nabla g^j(x) \,, h \rangle + \tfrac{1}{2} \rho \| h \|^2 - \phi(x)$ *with* $\rho > 0$. □

9. REFERENCES

[Arm.1] Armijo, L., "Minimization of Functions Having Continuous Partial Derivatives", *Pacific J. Math*, Vol. 16, pp. 1-3, 1966

[Asa.1] J. Asaadi, "A Computational Comparison of Some Non-Linear Programs", *Mathematical Programming*, Vol. 4, pp. 144-154, 1973.

[Ber.1] Berge, C., *Topological Spaces*, Macmillan, New York, N.Y., 1963. Wiley-Interscience, New York, N.Y., 1983.

[Cha.1] R. W. Chaney, "On the Pironneau-Polak Method of Centers", *J. Opt. Theory Appl.*, Vol. 20, No. 3, pp. 269-295, 1976.

[Cla.1] F. H. Clarke, Optimization and Nonsmooth Analysis, Wiley-Interscience, New York, 1983.

[Con.1] A. R. Conn, "Constrained Optimization Using a Nondifferentiable Penalty Function", *SIAM J. Num. Ana.*, vol. 10, No. 4, pp. 760-784, 1973.

[Con.2] A. R. Conn and T. Pietrzykowski, "A Penalty Function Method Converging Directly to a Constrained Optimum", *SIAM J. Num. Ana.*, Vol. 14, No. 2, pp. 348-375, 1977.

[Gil.1] P. E. Gill, S. J. Hammarling, W. Murray, M. A. Saunders, and M. H. Wright, "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.

[Han.1] S.-P. Han and O. L. Mangasarian, "Exact Penalty Functions in Nonlinear Programming", *Mathematical Programming*, Vol. 17, pp. 251-269, 1979.

[Her.1] J. N. Herskovits, "A Two-Stage Feasible Directions Algorithm for Nonlinear Constrained Optimization", *Mathematical Programming*, Vol. 36, pp. 19-38, 1986.

[Hig.1] J. E. Higgins and E. Polak, "Minimizing Pseudo-Convex Functions on Convex Compact Sets", *Electronics Research Laboratory, University of California, Berkeley*, Memorandum No. UCB/ERL M86/53, 1986.

[Hoc.1] W. Hock and K. Schittkowski, "Test Examples for Nonlinear Programming Codes", *Lecture Notes in Economics and Mathematical Systems 187*, Springer-Verlag, Berlin, 1981.

[Hoh.1] B. von Hohenbalken, "Simplical Decomposition in Nonlinear Programming Algorithms", *Mathematical Programming*, Vol. 13, pp. 49-68, 1977

[Hua.1] P. Huard, *Programmation mathematic convex*, Rev. Fr. Inform. Rech. Operation., 7 (1968) pp. 43-59.

[Joh.1] F. John, *Extremum problems with inequalities as side conditions*, in Courant Anniversary Volume, K. O. Friedrichs, O. E. Neugebauer and J. J. Stoker, eds., Wiley-Interscience, New York, 1948, pp. 187-204.

[Kiw.1] K. Kiwiel, "A Method for Solving Certain Quadratic Programming Problems Arising in Nonsmooth Optimization", *IMA Journal of Numerical Analysis*, Vol. 6, pp. 137-152, 1986.

[Kiw.2] K. Kiwiel, "A Dual Method for Solving Certain Positive Semi-Definite Quadratic Programming Problems", *SIAM Journal of Scientific and Statistical Computing*, Vol. 10, 1989.

[Lue.1] D. G. Luenberger, *Introduction to Linear and Nonlinear Systems*, Addison-Wesley, New York, 1983.

[May.1] D. Q. Mayne and E. Polak, "A Superlinearly Convergent Algorithm for Constrained Optimization Problems", *Mathematical Programming Study 16*, pp. 45-61, 1982.

[McC.1] G. McCormick, "Second Order Conditions for Constrained Minima", *SIAM Journal on Applied Mathematics*, Vol. 15, pp. 641-652, 1967.

[Mor.1] J. J. More and D. C. Sorensen, "Computing a Trust Region Step", *SIAM J. Sci. Stat.*

Comput., Vol.4, No. 3, pp. 553-572, 1983.

[Pan.1] E. R. Panier, A. L. Tits and J. N. Herskovits, "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.

[Pan.2] V. M. Panin, "Some Methods of Solving Convex Programming Problems", *USSR Comput. Maths. Math. Phys.*, Vol. 21, No. 2, pp. 57-72, 1981, (Zh. vychisl. Mat. mat. Fiz., Vol. 21, No. 2, pp. 315-328, 1981).

[Pan.3] V. M. Panin, "On a Second-Order Method for the Discrete Min-Max Problem", *USSR Comput. Maths. Math. Phys.*, Vol. 19, No. 1, pp. 90-100, (Zh. vychisl. Mat. mat. Fiz., Vol. 19, No. 1, pp. 88-98, 1979).

[Pan.4] V. M. Panin, "Some Methods of Solving Convex Programming Problems", *USSR Comput. Maths. Math. Phys.*, Vol. 21, No. 2, pp. 57-72, 1981, (Zh. vychisl. Mat. mat. Fiz., Vol. 21, No. 2, pp. 315-328, 1981).

[Pir.1] O. Pironneau and E. Polak, "On the Rate of Convergence of Certain Methods of Centers", *Mathematical Programming*, Vol. 2, No. 2, pp. 230-258, 1972.

[Pol.1] E. Polak, "On the Mathematical Foundations of Nondifferentiable Optimization in Engineering Design", *SIAM Review*, pp. 21-91, March 1987.

[Pol.2] E. Polak, R. Trahan and D. Q. Mayne, "Combined phase I - phase II methods of feasible directions", *Mathematical Programming*, Vol. 17, No. 1, pp. 32-61, 1979.

[Pol.3] E. Polak and L. He, "A Unified Phase I Phase II Method of Feasible Directions for Semi-infinite Optimization", *University of California, Electronics Research Laboratory*, Memo UCB/ERL M89/7, 3 February 1989, to appear in *JOTA*.

[Pol.4] E. Polak, *Computational Methods in Optimization: A Unified Approach*, Academic Press, 329 pages, 1971.

[Pol.5] E. Polak, D. Q. Mayne and J. Higgins, "A Superlinearly Convergent Algorithm for Min-Max Problems", *University of California, Berkeley, Electronics Research Laboratory* Memo No. M86/103, Nov. 15, 1986. Proc. *28th IEEE CDC*, Tampa, Florida, Dec. 1989.
Full length version to appear in *JOTA*

[Pol.6] E. Polak, D. Q. Mayne, and J. Higgins, "On the Extension of Newton's Method to Semi-Infinite Minimax Problems", *University of California, Electronics Research Laboratory*, Memo UCB/ERL M89/92, 1 August, 1989. Submitted to SICOPT

[Rus.1] A. Ruszczynski, "A Regularized Decomposition Method for Minimizing a Sum of Polyhedral Functions", *Mathematical Programming*, Vol. 35, pp. 46-61, 1986.

[Wie.1] E. J. Wiest and E. Polak, "On the Rate of Convergence of Two Minimax Algorithms", *University of California, Electronics Research Laboratory*, Memo UCB/ERL M89/111, 15 August 1989.

[Zou.1] G. Zoutendijk, *Mathematical Programming Methods*, North Holland, Amsterdam, 1976.

Problem	Algorithm	NF	NG	NDF	NDG	FV
Rosen-Suzuki	[Her.1]	7	27	7	21	-43.81453
	Algorithm 2.1	66	198	33	99	-43.83851
	Algorithm 5.1	6	18	3	9	-43.82342
	[Her.1]	15	54	15	45	-43.99907
	Algorithm 2.1	132	396	66	198	-43.99912
	Algorithm 5.1	20	60	10	30	-43.99927
Colville #1	[Her.1]	6	60	6	60	-32.03453
	Algorithm 2.1	265	2650	127	1270	-32.06142
	Algorithm 5.1	12	120	6	60	-32.21449
	[Her.1]	9	90	9	90	-32.34851
	Algorithm 2.1	884	8840	436	4360	-32.34851
	Algorithm 5.1	32	320	16	160	-32.34865
Colville #2	[Her.1]	36	190	36	180	32.81567
	Algorithm 2.1	1840	9200	872	4360	32.81530
	Algorithm 5.1	526	2630	246	1230	32.66952
	[Her.1]	53	320	53	265	32.34897
	Algorithm 5.1	1741	8705	324	1620	32.34906
Kuhn-Tucker	Algorithm 2.1	92	184	46	92	-0.9009127
	Algorithm 5.1	45	90	6	12	-0.92223418
	Algorithm 2.1	6116	12232	3058	6116	-0.9900006
	Algorithm 5.1	110	220	15	30	-0.9905035
Circular-Quadratic	Algorithm 2.1	10	20	5	10	4.526097
	Algorithm 5.1	2	4	1	2	4.530063
	Algorithm 2.1	54	108	27	54	4.500000
	Algorithm 5.1	4	8	2	4	4.500000

Table 1: Summary of Numerical Results

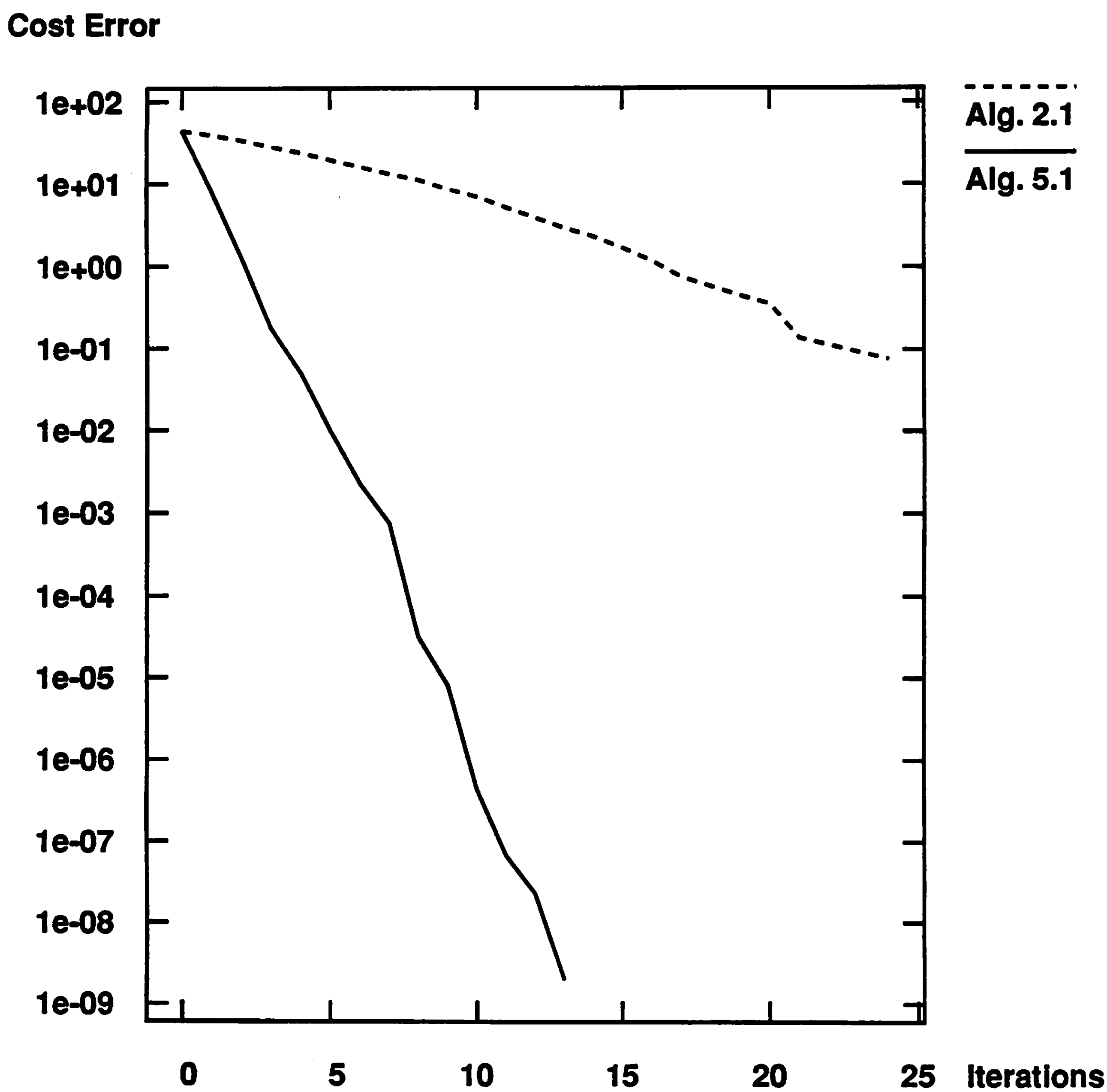

Figure 1: Rosen-Suzuki Problem

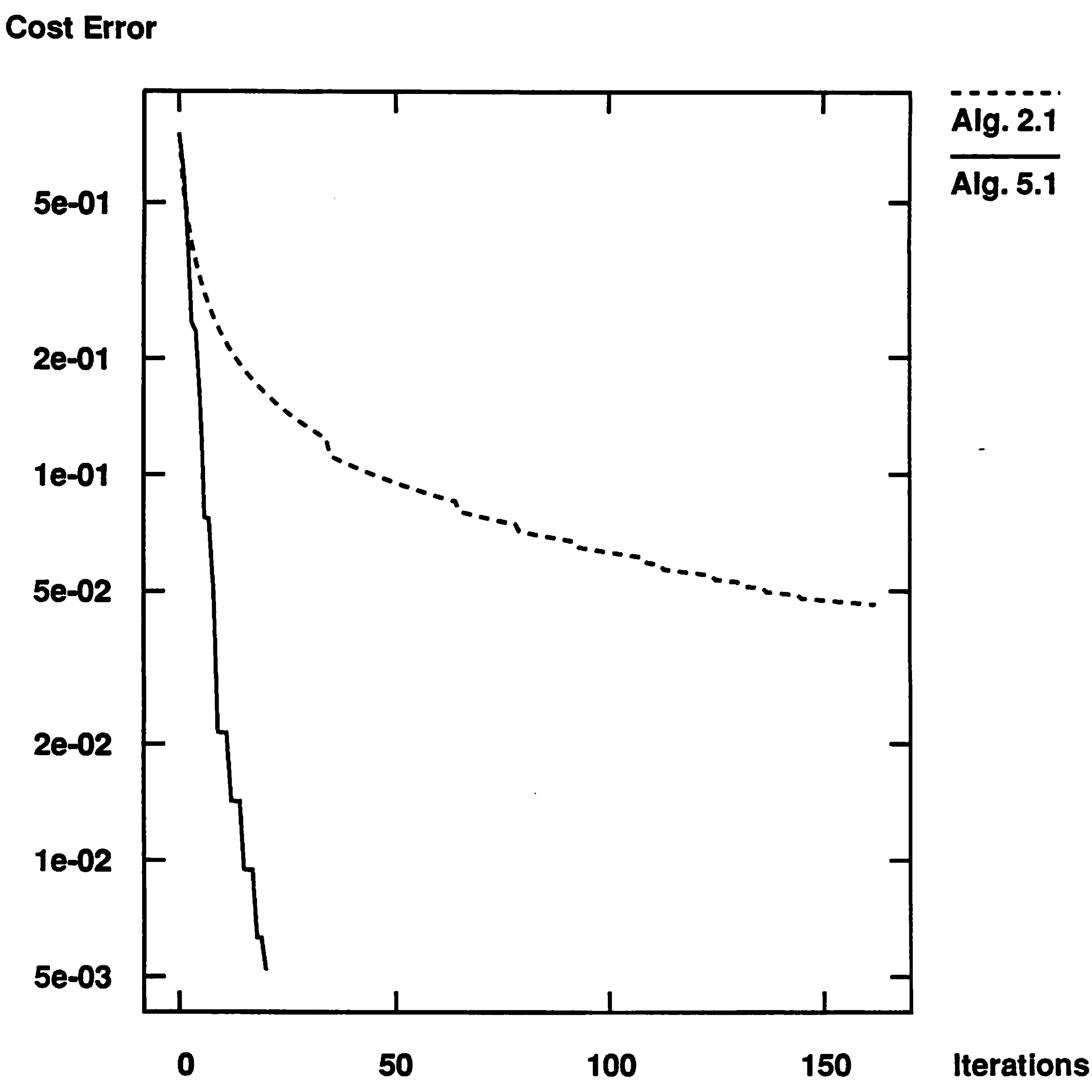

Figure 2: Kuhn-Tucker Problem