A UNIFIED PHASE I – PHASE II METHOD OF FEASIBLE DIRECTIONS FOR SEMI-INFINITE OPTIMIZATION

by

E. Polak and L. He

Memorandum No. UCB/ERL M89/7

3 February 1989

A UNIFIED PHASE I – PHASE II METHOD OF FEASIBLE DIRECTIONS FOR SEMI-INFINITE OPTIMIZATION

by

E. Polak and L. He

Memorandum No. UCB/ERL M89/7

3 February 1989

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

A UNIFIED PHASE I – PHASE II METHOD OF FEASIBLE DIRECTIONS FOR SEMI-INFINITE OPTIMIZATION

by

E. Polak and L. He

Memorandum No. UCB/ERL M89/7

3 February 1989

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

A UNIFIED PHASE I-PHASE II METHOD OF FEASIBLE DIRECTIONS FOR SEMI-INFINITE OPTIMIZATION*

E. Polak** and L. He**

ABSTRACT

In this paper we complete a cycle in the construction of methods of feasible directions for solving semi-infinite, constrained optimization problems. Earlier phase I - phase II methods of feasible directions used one search direction rule in all of $\mathbb{R}^n$ with two step-size rules: one for feasible points and one for infeasible points. The algorithm presented in this paper uses both a single search direction rule and a single step-size rule in all of $\mathbb{R}^n$. The new algorithm is simpler to analyze and performs somewhat better than existing, first order, phase I - phase II methods. The new algorithm is globally convergent, with linear rate.

KEY WORDS

semi-infinite constrained optimization, method of feasible directions, linear convergence.

* Research sponsored in part by the National Science Foundation Grant ECS-8121149, Air Force Office of Scientific Research Grant 83-0361, the State of California MICRO Program.

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

1. INTRODUCTION

Simple methods of feasible directions solve problems of the form

$$\mathbf{P}: \quad \min\{ f^0(x) \mid f^j(x) \le 0,\ j = 1,2,\ldots,m \}, \tag{1.1}$$

where the functions $f^j: \mathbb{R}^n \to \mathbb{R}$ are continuously differentiable. The first versions of methods of feasible directions (see e.g., [Zou.1, Zuk.1, Top.1, Pol.1]) were all phase II, i.e., they required a feasible starting point x_0. A starting point x_0 was computed by a phase I procedure which consisted of applying the phase II method to the following problem, in the augmented space $\mathbb{R}^{n+1}$,

$$\mathbf{P_I}: \quad \min\{ x^{n+1} \mid f^j(x) - x^{n+1} \le 0,\ j = 1,2,\ldots,m \}. \tag{1.2}$$

Given any $x^* \in \mathbb{R}^n$, it is clear that if we set $x^{*n+1} = \max_{j \in \mathbf{m}} f^j(x^*)$[1], then (x^*, x^{*n+1}) is feasible for $\mathbf{P_I}$ and hence the phase II method can be applied to its solution.

The main problem with this approach is that the initial point x_0 thus produced can turn out to be quite bad. Because of this, in [Pol.2], we have developed a class of phase I - phase II methods for solving **P**, which take the cost function into account even when the current iterate is infeasible. In [Pol.3], we find extensions of the methods in [Pol.2] to semi-infinite optimization problems of the form

$$\min\{ \psi^0(x) \mid \psi^j(x) \le 0 \}, \tag{1.3}$$

where $\psi^j(x) = \max_{y_j \in Y_j} \phi^j(x, y_j)$, $j = 0,1,\ldots,m$, with the $\phi^j(\cdot,\cdot)$ continuously differentiable in x and continuous in y_j, and the sets $Y_j \subset \mathbb{R}^{k_j}$ compact. The methods in [Pol.2], as well as their extensions, use a unified search direction procedure which always involves the cost gradient, but they use two distinct step size rules: one for the case where the current iterate is feasible and the other one for the case when the current iterate is infeasible. One consequence of this is that their convergence analysis is complicated by the fact that it requires the use of two cost functions: the actual cost function for sequences whose tails are feasible, and the constraint violation function ($\psi(x) = \max_{j \in \mathbf{m}} f^j(x)$) for sequences which remain infeasible. A second consequence is that while the sequence that these algorithms construct is infeasible, it moves towards an attractor point in the feasible set which need not be a local minimizer.

---

[1] We use the notation $\mathbf{m} \triangleq \{ 1,2,3,\ldots,m \}$.

In this paper, we present a new phase I - phase II method of feasible directions for solving semi-infinite optimization problems of the form (1.3). The new method uses both a unified search direction subprocedure and a unified step size subprocedure. The new method has two advantages over its predecessors. The first is that it is easier to establish its convergence, Q-linear rate of convergence of the cost or constraint violation functions, R-linear rate of convergence of the iterates as well as the range of the *steering* parameter, which ensures that a feasible point is attained in a finite number of iterations. The second advantage is that it does not switch attractor points in the transition from the infeasible region to the feasible region, and as a result tends to compute a better first feasible point than its predecessors. This fact is corroborated by our computational results, which show the new method to be somewhat more efficient.

2. PRELIMINARY RESULTS

We will consider semi-infinite constrained optimization problems of the form

$$\min\{\ \psi^0(x) \mid \psi^j(x) \leq 0,\ j \in \mathbf{m}\ \}\ , \tag{2.1a}$$

where $\mathbf{m} \triangleq \{\ 1,2,\ldots,m\ \}$, and, with $\mathbf{M} \triangleq \{\ 0,1,2,\ldots,m\ \}$,

$$\psi^j(x) = \max_{y_j \in \mathbf{Y}_j} \phi^j(x,y_j),\ \ \forall\ j \in \mathbf{M}, \tag{2.1b}$$

where $\phi^j : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$. We will assume that the functions $\phi^j(\cdot,\cdot)$ and their gradients $\nabla_x \phi^j(\cdot,\cdot)$ are Lipschitz continuous on bounded sets. In addition, we will assume that the intervals $\mathbf{Y}_j = [a_j,b_j] \subset \mathbb{R}$ are compact. We note that when $\mathbf{Y}_j$ contains only one point, i.e., $a_j = b_j$, the function $\psi^j(x) = \phi^j(x,a_j)$ is differentiable (otherwise it need not be); thus we see that the formulation (2.1a-b) allows that some of the $\psi^j(x)$ are ordinary differentiable functions. Let

$$\psi(x) \triangleq \max_{j \in \mathbf{m}} \psi^j(x), \tag{2.2a}$$

$$\psi_+(x) \triangleq \max\{\ 0, \psi(x)\ \}. \tag{2.2b}$$

Let the *steering parameter*[2] $\gamma > 0$ be given. Then, for any $z \in \mathbb{R}^n$, we define the parametrized

[2] An examination of (2.2c) shows that the value of γ and, in fact, the term $\psi_+(z)$ has no effect at *feasible points*. We shall see later that their inclusion enables us to construct a phase I - phase II algorithm which does not require a feasible starting point, and (see Theorem 4.1(iii)) that γ can be used to control the speed with which feasibility is achieved.

function $F_z(x)$ by

$$F_z(x) \triangleq \max\{ \psi^0(x) - \psi^0(z) - \gamma\psi_+(z), \psi^j(x) - \psi_+(z), j \in \mathbf{m} \}. \tag{2.2c}$$

Note that (i) for any $z \in \mathbb{R}^n$, $F_z(z) = 0$, and (ii) if x^* is a local minimizer for (2.1a), then, since $\psi(x) > 0$ when x is infeasible for (2.1a), and since $\psi^0(x) \geq \psi^0(x^*)$ for all feasible x in a ball about x^*, x^* must also be a local minimizer for the problem

$$\min_{x \in \mathbb{R}^n} F_{x^*}(x) . \tag{2.2d}$$

This fact is used in [Cla.1] to obtain the following optimality condition for problem (2.1a):

Proposition 2.1 [Cla.1] : If x^* is a local minimizer for (2.1a), then

$$dF_{x^*}(x^*;h) \geq 0, \ \forall \ h \in \mathbb{R}^n, \tag{2.3a}$$

where $dF_{x^*}(x^*;h)$ denotes the directional derivative of $F_{x^*}(\cdot)$ at x^* in the direction h. Equivalently,

$$0 \in \partial F_{x^*}(x^*), \tag{2.3b}$$

where $\partial F_x(x)$ denotes the *generalized gradient* of $F_x(\cdot)$ at x, and is given by

$$\partial F_x(x) = \text{co} \{ \partial\psi^j(x) \mid j \in J(x) \} , \tag{2.3c}$$

where $\partial\psi^j(x)$ is the generalized gradient of $\psi^j(\cdot)$ at x, and

$$J(x) \triangleq \begin{cases} I(x) & \psi(x) > 0 \\ I(x) \cup \{0\} & \psi(x) = 0, \\ \{0\} & \psi(x) < 0 \end{cases} \tag{2.3d}$$

with $I(x) \triangleq \{ j \mid \psi^j(x) = \psi(x), j \in \mathbf{m} \}$. ■

It is easy to see that (2.3b) can be restated in multiplier form, as follows:

Corollary 2.1 : If x^* is optimal for (2.1a), then there exists a multiplier vector $\mu^* \in \Sigma_{m+1} \triangleq \{ \mu \in \mathbb{R}^{m+1} \mid \mu^j \geq 0, j \in \mathbf{M}, \sum_{j \in \mathbf{M}} \mu^j = 1 \}$, such that

$$0 \in \sum_{j=0}^{m} \mu^{*j} \partial\psi^j(x^*) , \tag{2.3e}$$

$$0 = \sum_{j=1}^{m} \mu^{*j} \psi^j(x^*) \, . \tag{2.3f}$$

■

The following sufficient condition is fairly obvious:

Proposition 2.2 : Suppose that the functions $\psi^j(\cdot)$, $j \in \mathbf{M}$, are convex, and that x^* is such that $\psi_+(x^*) = 0$ and there exists a multiplier vector $\mu^* \in \Sigma_{m+1}$ such that (2.3e-f) hold and $\mu^{*0} \neq 0$, then x^* is a global minimizer for (2.1a). ■

Referring to [Pol.3], we see that for the purpose of constructing algorithms, it is useful to replace the linear first order approximation $dF_z(z;h)$ of $F_z(z+h) - F_z(z)$ in a neighborhood of z by a *convex* first order approximation, as follows.

Given $z \in \mathbb{R}^n$, we approximate each function $\phi^j(x,y_j)$, $j \in \mathbf{M}$, around z by the *first order convex* approximation:

$$\hat{\phi}^j_z(x,y_j) \triangleq \phi^j(z,y_j) + \langle \nabla_x \phi^j(z,y_j), (x - z) \rangle + \tfrac{1}{2} |x - z|^2 \, . \tag{2.4a}$$

Then $\psi^j(x)$ is approximated in a neighborhood of z by the *first order convex* approximation

$$\hat{\psi}^j_z(x) \triangleq \max_{y_j \in Y_j} \hat{\phi}^j_z(x,y_j), \quad j \in \mathbf{M}, \tag{2.4b}$$

and, in turn, $F_z(x)$ is approximated in a neighborhood of z by the *first order convex* approximation

$$\hat{F}_z(x) \triangleq \max\{\, \hat{\psi}^0_z(x) - \psi^0(z) - \gamma\psi_+(z), \hat{\psi}^j_z(x) - \psi_+(z), \, j \in \mathbf{m} \,\} . \tag{2.4c}$$

Referring to [Pol.3], we define the *optimality function* $\theta\colon \mathbb{R}^n \to \mathbb{R}$ and the *search direction map* $h\colon \mathbb{R}^n \to \mathbb{R}^n$ as follow:

$$\theta(x) \triangleq \min_{h \in \mathbb{R}^n} \hat{F}_x(x + h) \, , \tag{2.5a}$$

$$h(x) \triangleq \arg \min_{h \in \mathbb{R}^n} \hat{F}_x(x + h) \, . \tag{2.5b}$$

The fact that $h(x)$ is well defined can be established using either duality or the von Neumann theorem. Specifically, we get the following result by straightforward extension of the results in [Pol.3].

Lemma 2.1 : Let $G^j:\mathbb{R}^n \to 2^{\mathbb{R}^{n+1}}$, for $j \in \mathbf{M}$, and $G:\mathbb{R}^n \to 2^{\mathbb{R}^{n+1}}$ be defined as follows:

$$G^0(x) = \text{co}\left\{\begin{bmatrix}\psi^0(x) - \phi^0(x,y_0) + \gamma\psi_+(x) \\ \nabla_x\phi^0(x,y_0)\end{bmatrix}\right\}_{y_0 \in \mathbf{Y}_0}, \tag{2.6a}$$

$$G^j(x) = \text{co}\left\{\begin{bmatrix}\psi_+(x) - \phi^j(x,y_j) \\ \nabla_x\phi^j(x,y_j)\end{bmatrix}\right\}_{y_j \in \mathbf{Y}_j}, \quad j \in \mathbf{m}, \tag{2.6b}$$

$$G(x) = \text{co}\,\{G^j(x)\}_{j \in \mathbf{M}}. \tag{2.6c}$$

Then

(i) For any $x \in \mathbb{R}^n$,

$$\theta(x) = -\min\{\, \xi^0 + \tfrac{1}{2}|\xi|^2 \mid (\xi^0,\xi)^T \in G(x)\,\}. \tag{2.7a}$$

(ii) There is a unique $\bar{\xi}(x) \triangleq (\xi^0(x),\xi(x))^T \in G(x)$ such that

$$\begin{aligned}\theta(x) &= -(\xi^0(x) + \tfrac{1}{2}|\xi(x)|^2) \\ &= \max_{\bar{\xi} \in G(x)} (-\,\xi^0 + \langle\xi,h(x)\rangle + \tfrac{1}{2}|h(x)|^2) \\ &= \min_{h \in \mathbb{R}^n} (-\,\xi^0(x) + \langle\xi(x),h\rangle + \tfrac{1}{2}|h|^2)\,,\end{aligned} \tag{2.7b}$$

and

$$h(x) = -\,\xi(x)\,. \tag{2.7c}$$

■

It can be deduced from Proposition 5.4 and Proposition 5.5 in [Pol.3] that the following result holds:

Proposition 2.3 :

(i) The functions $\theta(\cdot)$ and $h(\cdot)$ are both continuous.

(ii) For any $x \in \mathbb{R}^n$, $\theta(x) \le 0$. Furthermore, $\theta(x) = 0$ if and only if $0 \in G(x)$, which holds if and only if either $\psi(x) \le 0$ and $0 \in \partial F_x(x)$ (i.e., x satisfies the first order optimality condition for problem (2.1a)), or $\psi(x) > 0$ and $0 \in \partial\psi(x)$, (i.e., x satisfies the first order optimality condition for the problem $\min_{z \in \mathbb{R}^n} \psi(z)$).

(iii) For any x such that $\theta(x) \ne 0$, $h(x)$ is a descent direction for $F_x(\cdot)$ at x, more precisely,

$$dF_x(x;h(x)) \triangleq \lim_{\lambda \to 0+} \frac{F_x(x+\lambda h(x))-F_x(x)}{\lambda} \leq \theta(x). \tag{2.8}$$

■

3. THE ALGORITHM

For the purpose of comparison, we recall that the two-step-size-rule semi-infinite phase I-phase II methods of feasible directions evolved from [Pir.1], in [Pol.2] and [Pol.3], have the following form:

Algorithm 3.1 (Original Phase I - Phase II Method of Feasible Directions) :

Parameters : $\gamma > 0$, α, $\beta \in (0,1)$.

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

Step 0 : Set $i = 0$.

Step 1 : Compute the the *optimality function* value $\theta_i = \theta(x_i)$, and the corresponding *search direction* $h_i = h(x_i)$.

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

$$\lambda_i = \max\{\ \beta^k \mid k \in \mathbb{N},\ \psi(x_i + \beta^k h_i) - \psi(x_i) \leq \beta^k \alpha \theta_i\ \}\ . \tag{3.1a}$$

else set

$$\lambda_i = \max\{\ \beta^k \mid k \in \mathbb{N},\ \psi^0(x_i + \beta^k h_i) - \psi^0(x_i) \leq \beta^k \alpha \theta_i\ ,\ \ \psi(x_i + \beta^k h_i) \leq 0\ \}. \tag{3.1b}$$

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

Remark 3.1: The infeasible points x_i generated by Algorithm 3.1 are attracted to the set $\{\ x \in \mathbb{R}^n \mid d\psi(x;h(x)) \geq 0\ \} \subset \{\ x \in \mathbb{R}^n \mid \psi_+(x) - \psi(x) \geq -\theta(x) + \frac{1}{2}\|h(x)\|^2\ \}$, while the feasible points x_i are attracted to the set $\{\ x \in \mathbb{R}^n \mid \theta(x) = 0\ \}$. The use of the attractor set $\{\ x \in \mathbb{R}^n \mid d\psi(x;h(x)) \geq 0\ \}$ in the infeasible region tends to detract from the ability of the algorithm to enter the feasible set at an advantageous point. ■

Next we state our new algorithm, which differs from Algorithm 3.1 only in its use of a unified step size rule in Step 2. Unlike Algorithm 3.1, the new algorithm has the same set of attractor points $\{\ x \in \mathbb{R}^n \mid \theta(x) = 0\ \}$, both while the sequence is infeasible and when the sequence is feasible. The advantages of this are demonstrated in Fig. 5.6.

Algorithm 3.2 (Unified Phase I - Phase II Method of Feasible Directions) :

Parameters : $\gamma > 0$, α, $\beta \in (0,1)$.

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

Step 0 : Set $i = 0$.

Step 1 : Compute the the *optimality function* value $\theta_i = \theta(x_i)$, and the corresponding *search direction* $h_i = h(x_i)$.

Step 2 : Compute the *step size* λ_i:

$$\lambda_i = \max\{\ \beta^k \mid k \in \mathbb{N},\ F_{x_i}(x_i + \beta^k h_i) \le \beta^k \alpha \theta_i\ \}\ . \tag{3.2}$$

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

We note that the step size computed by Algorithm 3.2 is always smaller than or equal to the step size computed by Algorithm 3.1, and that, on the average, Algorithm 3.2 uses slightly more work per iteration, because it evaluates the cost function $\psi^0(\cdot)$ at infeasible points, while Algorithm 3.1 does not. We will see in the Sections 4 and 5 that these seeming disadvantages do not handicap Algorithm 3.2.

Lemma 3.1 : If $\{\ x_i\ \}_{i=0}^{\infty}$ is a sequence constructed by Algorithm 3.2, then for $i \ge 0$,

$$\psi^0(x_{i+1}) \le \psi^0(x_i) + \gamma\psi_+(x_i); \tag{3.3a}$$

$$\psi_+(x_{i+1}) \le \psi_+(x_i)\ . \tag{3.3b}$$

Proof : By the construction of $\{\ x_i\ \}_{i=0}^{\infty}$, $F_{x_i}(x_{i+1}) \le F_{x_i}(x_i) = 0$. It now follows from of the definition of $F_z(x)$, that $\psi^0(x_{i+1}) - \psi^0(x_i) - \gamma\psi_+(x_i) \le 0$ and that $\psi(x_{i+1}) - \psi_+(x_i) \le 0$. Thus, (3.1a) and (3.1b) must hold. ■

Theorem 3.1 If $\{\ x_i\ \}_{i=0}^{\infty}$ is a sequence constructed by Algorithm 3.2, then any accumulation point x^* of the sequence $\{\ x_i\ \}_{i=0}^{\infty}$ satisfies $\theta(x^*)$.

Proof : Suppose that $x_i \xrightarrow{K} x^*$ and that $\theta(x^*) \ne 0$. Then, by Proposition 2.3(ii), $\theta(x^*) < 0$. Hence it follows from Proposition 2.3(iii) that there exist an $\alpha^* \in (\alpha,1)$ and a $k_0 \in \mathbb{N}$ such that for $0 < \lambda \le \beta^{k_0}$,

$$F_{x^*}(x^* + \lambda h(x^*)) - F_{x^*}(x^*) \le \alpha^* \lambda \theta(x^*) \,. \tag{3.4}$$

Since $F_x(x) = 0$ and $F_x(x)$, $\theta(x)$ and $h(x)$ are continuous, there exists a $\rho^* > 0$ such that for all $x \in B(x^*,\rho^*) \triangleq \{ x \in \mathbb{R}^n \mid \|x - x^*\| \le \rho^* \}$ and $0 < \lambda \le \beta^{k_0}$,

$$F_x(x + \lambda h(x)) \le \alpha \lambda \theta(x), \tag{3.5a}$$

and

$$\theta(x) \le \theta(x^*)/2 < 0 \,. \tag{3.5b}$$

Since $x_i \xrightarrow{K} x^*$, there exists an $i_0 \in \mathbb{N}$ such that $x_i \in B(x^*,\rho^*)$ for all $i \in K$, $i > i_0$. Hence, (3.2) is satisfied by $\lambda_i \ge \beta^{k_0}$ for all $i \in K$, $i > i_0$. Therefore, making use of (3.5a-b), we obtain that for all $i \in K$ and $i > i_0$,

$$F_{x_i}(x_{i+1}) \le \alpha \beta^{k_0} \theta(x_i) \le \alpha \beta^{k_0} \theta(x^*)/2 \,. \tag{3.6}$$

Now, we must consider two cases.

Case (i): There exists $i_1 \in \mathbb{N}$ such that $\psi(x_{i_1}) \le 0$. Then, it follows from Lemma 3.1 that $\psi(x_i) \le 0$ and $\psi^0(x_{i+1}) \le \psi^0(x_i)$ for all $i > i_1$. In addition, it follows from (3.6) that for all $i \in K$, $i > \max\{i_0, i_1\}$,

$$\psi^0(x_{i+1}) - \psi^0(x_i) \le F_{x_i}(x_{i+1}) \le \alpha \beta^{k_0} \theta(x^*)/2 \,. \tag{3.7}$$

Since the sequence $\{ \psi^0(x_i) \}_{i=i_1}^{\infty}$ is monotonic decreasing, we conclude from (3.7) that $\psi^0(x_i) \to -\infty$, as $i \to \infty$, which contradicts the fact that $\psi^0(x_i) \to \psi^0(x^*)$.

Case (ii): $\psi(x_i) > 0$ for all i. Then, it follows from Lemma 3.1 that $\psi(x_{i+1}) \le \psi(x_i)$ for all i. Making use of (3.6), we conclude that for all $i \in K$, $i > i_0$,

$$\psi(x_{i+1}) - \psi(x_i) \le F_{x_i}(x_{i+1}) \le \alpha \beta^{k_0} \theta(x^*)/2 \,. \tag{3.8}$$

Since the sequence $\{\psi(x_i)\}_{i=0}^{\infty}$ is monotonic decreasing, we conclude from (3.8) that $\psi(x_i) \to -\infty$, which contradicts the fact that $\psi(x_i) > 0$ for all i. ■

4. RATE OF CONVERGENCE AND STEERING

Next we turn to an analysis of the rate of convergence of Algorithm 3.2, and of the effects of the steering parameter γ.

Assumption 4.1 : We will assume that the functions $\phi^j(\cdot,\cdot)$, $j \in \mathbf{M}$, in Problem (2.1a) satisfy following hypotheses:

(i) There exist $0 < c < 1 < C < \infty$ such that for all $x \in \mathbb{R}^n$, $z \in \mathbb{R}^n$, $y_j \in \mathbf{Y}_j$ and for all $j \in \mathbf{M}$,

$$c\|z\|^2 \le \langle z, \frac{\partial^2 \phi^j(x,y_j)}{\partial x^2} z\rangle \le C\|z\|^2 . \tag{4.1}$$

(ii) The set $\{ x \mid \psi(x) < 0 \}$ is not empty. ∎

Lemma 4.1 : Suppose that Assumption 4.1(i) holds. Then

(i) For any $x, h \in \mathbb{R}^n$, $y_j \in \mathbf{Y}_j$ and $j \in \mathbf{M}$,

$$\phi^j(x,y_j) + \langle \nabla_x \phi^j(x,y_j), h\rangle + \tfrac{1}{2}c\|h\|^2 \le \phi^j(x+h,y_j) \le \phi^j(x,y_j) + \langle \nabla_x \phi^j(x,y_j), h\rangle + \tfrac{1}{2}C\|h\|^2 . \tag{4.2a}$$

(ii) For any x, $h \in \mathbb{R}^n$, $\xi \in \partial\psi^j(x)$, and $j \in \mathbf{M}$,

$$\psi^j(x) + \langle \xi, h\rangle + \tfrac{1}{2}c\|h\|^2 \le \psi^j(x+h) . \tag{4.2b}$$

(iii) For any x, $h \in \mathbb{R}^n$,

$$\max_{\xi \in G^0(x)} \{ -\xi^0 + \langle \xi, h\rangle + \tfrac{1}{2}c\|h\|^2 \} \le \psi^0(x+h) - \psi(x) - \gamma\psi_+(x) \le \max_{\xi \in G^0(x)} \{ -\xi^0 + \langle \xi, h\rangle + \tfrac{1}{2}C\|h\|^2 \} \tag{4.2c}$$

$$\max_{\xi \in G^j(x)} \{ -\xi^0 + \langle \xi, h\rangle + \tfrac{1}{2}c\|h\|^2 \} \le \psi^j(x+h) - \psi_+(x) \le \max_{\xi \in G^j(x)} \{ -\xi^0 + \langle \xi, h\rangle + \tfrac{1}{2}C\|h\|^2 \}, \ j \in \mathbf{m} \tag{4.2d}$$

Proof : (i) The inequality (4.2a) follows directly from the Taylor second order expansion in integral form and Assumption 4.1(i).

(ii) We recall that $\partial\psi^j(x) = \text{co} \{ \nabla_x \phi^j(x,y_j) \mid y_j \in \mathbf{Y}_j, \ \phi^j(x,y_j) = \psi^j(x) \}$. Hence we see that (4.2a) implies (4.2b).

(iii) It follows from (4.2a) that for $y_0 \in \mathbf{Y}_0$,

$$\phi^0(x,y_0) - \psi^0(x) - \gamma\psi_+(x) + \langle \nabla_x \phi^j(x,y_0), h\rangle + \tfrac{1}{2}c\|h\|^2 \le \phi^0(x+h,y_0) - \psi^0(x) - \gamma\psi_+(x)$$

$$\leq \phi^0(x,y_0) - \psi^0(x) - \gamma\psi_+(x) + \langle \nabla_x \phi^0(x,y_0), h \rangle + \tfrac{1}{2} C |h|^2 . \tag{4.3a}$$

Thus, we must have that

$$\max_{y_0 \in Y_0} \{ \phi^0(x,y_0) - \psi^0(x) - \gamma\psi_+(x) + \langle \nabla_x \phi^0(x,y_0), h \rangle + \tfrac{1}{2} c |h|^2 \} \leq \psi^0(x+h) - \psi^0(x) - \gamma\psi_+(x)$$

$$\leq \max_{y_0 \in Y_0} \{ \phi^0(x,y_0) - \psi^0(x) - \gamma\psi_+(x) + \langle \nabla_x \phi^0(x,y_0), h \rangle + \tfrac{1}{2} C |h|^2 \} . \tag{4.3b}$$

Making use of the definition of $G^0(x)$ and the fact that maximizing a linear function over a compact set is equivalent to maximizing the linear function over the convex hull of the compact set, we conclude that (4.3b) implies (4.2c). Similarly, we can obtain (4.2d). ■

Lemma 4.2 : Suppose that Assumption 4.1 holds. Then

(i) For all x such that $\psi(x) \geq 0$, $0 \notin \partial\psi(x)$.

(ii) Problem (2.1a) has a unique solution.

(iii) The unique solution of (2.1a) is the unique zero of $\theta(\cdot)$.

Proof : (i) Since the functions $\phi^j(\cdot,\cdot), j \in \mathbf{M}$, satisfy Assumption 4.1(i), the functions $\phi^j(\cdot,y_j)$, with $y_j \in \mathbf{Y}_j$, are strictly convex and have bounded level set. Hence, the functions $\psi^j(\cdot)$ and $\psi(\cdot)$ are also strictly convex and have bounded level sets. Therefore $\psi(\cdot)$ has only one local minimizer, $\bar{x}$, which is therefore the global minimizer of $\psi(\cdot)$. Since $\psi(\cdot)$ is strictly, it follows that $\bar{x}$ is the only point satisfying the relation $0 \in \partial\psi(x)$. By Assumption 4.1(ii), $\psi(\bar{x}) < 0$.

(ii) Because the functions $\psi^j(\cdot)$, $j = 0,1,...,m$, are strictly convex and the feasible set of Problem (2.1a) is not empty and bounded, Problem (2.1a) has a unique solution.

(iii) Suppose that x^* is such that $\theta(x^*) = 0$. Then, it follows from Lemma 4.2(i), Proposition 2.3(ii) and Corollary 2.1, that $\psi(x^*) \leq 0$ and $0 \in \partial F_{x^*}(x^*)$, which implies that there exists a multiplier vector $\mu^* \in \Sigma_{m+1}$, $\mu^* = (\mu^{*0},\mu^{*1},...,\mu^{*m})$ such that (2.3e-f) is satisfied. If $\mu^{*0} = 0$, then (2.3e-f) imply that $\psi(x^*) = 0$ and that $0 \in \partial\psi(x^*)$, which contradicts to Lemma 4.2(i). Therefore $\mu^{*0} \neq 0$. Hence, by Proposition 2.2, x^* is the solution of Problem (2.1a). Since the solution of (2.1a) is unique, we must have that the zero of $\theta(\cdot)$ is unique. ■

Let x^* denote the unique solution of (2.1a) and let $L(x^*) \subset \Sigma_{m+1}$ be the set of multiplier vectors satisfying (2.3e-f) at the optimal solution x^* of (2.1a), and, for any $x \in \mathbb{R}^n$, let U(x) be the set of dual variables associated with the optimality function and the search direction map, i.e.,

$$L(x^*) \triangleq \{ \mu \in \Sigma_{m+1} \mid 0 \in \sum_{j=0}^{m} \mu^j \partial\psi^j(x^*),\ \sum_{j=1}^{m} \mu^j \psi^j(x^*) = 0 \}, \tag{4.4a}$$

$$U(x) \triangleq \{ \mu \in \Sigma_{m+1} \mid \xi(x) \in \sum_{j=0}^{m} \mu^j G^j(x) \} . \tag{4.4b}$$

Since the generalized gradients $\partial\psi^j(x^*)$, $j \in$ **M**, are compact and convex, the set $L(x^*)$ is compact and convex. Let $\underline{\mu}^0 \triangleq \min\{ \mu^0 \mid \mu \in L(x^*) \}$ and let $\bar{\mu}^0 \triangleq \max\{ \mu^0 \mid \mu \in L(x^*) \}$, Then, we have following result.

Lemma 4.3 : Suppose that Assumption 4.1 holds and that x^* is the unique solution of (2.1a). Then

(i) $0 < \underline{\mu}^0 \le \bar{\mu}^0 \le 1$.

(ii) $U(x^*) = L(x^*)$.

(iii) For any $\varepsilon \in (0,1)$, there exists a $\rho^* > 0$ such that for any $x \in B(x^*,\rho^*) = \{ x \in \mathbb{R}^n \mid \|x - x^*\| \le \rho^* \}$ and $\mu = (\mu^0,\mu^1, \cdots ,\mu^m) \in U(x)$,

$$\underline{\mu}^0(1-\varepsilon) \le \mu^0 \le \bar{\mu}^0(1+\varepsilon) . \tag{4.5a}$$

(iv) For all $x \in \mathbb{R}^n$,

$$\bar{\mu}^0[\psi^0(x^*) - \psi^0(x)] \le (1-\bar{\mu}^0)\psi_+(x) , \tag{4.5b}$$

$$\frac{c}{2}\|x - x^*\|^2 \le \underline{\mu}^0[\psi^0(x) - \psi^0(x^*)] + (1-\underline{\mu}^0)\psi_+(x) . \tag{4.5c}$$

Proof : (i) It follows directly from the proof of Lemma 4.2(iii) and the compactness of $L(x^*)$ that $\underline{\mu}^0 > 0$.

(ii) Since $\theta(x^*) = 0$, $\xi(x^*) = 0$. Making use of the facts that $\partial\psi^j(x) = \text{co}\{ \nabla_x\phi^j(x,y_j) \mid y_j \in Y_j,\ \phi^j(x,y_j) = \psi^j(x) \}$ for $j \in$ **M** and that $\psi_+(x^*) = 0$, we conclude that $U(x^*) = L(x^*)$.

(iii) Since $\xi(\cdot)$ and $G^j(\cdot), j \in$ **M** are continuous, the set valued map $U(\cdot)$ must be upper semi-continuous, and hence for any sequence x_i and $\mu_i \in U(x_i)$ such that $x_i \to x^*$ and $\mu_i \to \mu^*$, then

$\mu^* \in U(x^*)$. Therefore, (iii) follows from Lemma 4.3(i)-(ii) and the upper semi-continuity of $U(\cdot)$.

(iv) For any $\mu = (\mu^0, \mu^1, \ldots, \mu^m) \in L(x^*)$, there exist $\xi_j \in \partial\psi^j(x^*)$ for $j \in \mathbf{M}$ such that

$$0 = \sum_{j=0}^{m} \mu^j \xi_j , \tag{4.6a}$$

$$0 = \sum_{j=1}^{m} \mu^j \psi^j(x^*) . \tag{4.6b}$$

Making use of the fact that $\psi_+(x) \geq \psi^j(x)$ for all $j \in \mathbf{m}$, of Lemma 4.1(ii) and of (4.6a-b), we obtain that for any $x \in \mathbb{R}^n$,

$$\begin{aligned} \mu^0\psi^0(x) + (1-\mu^0)\psi_+(x) &\geq \sum_{j=0}^{m} \mu^j\psi^j(x) \\ &\geq \sum_{j=0}^{m} \mu^j(\psi^j(x^*) + \langle \xi_j, (x-x^*)\rangle + \frac{c}{2}\|x-x^*\|^2) \\ &= \sum_{j=0}^{m} \mu^j\psi^j(x^*) + \frac{c}{2}\|x-x^*\|^2 \\ &= \mu^0\psi^0(x^*) + \frac{c}{2}\|x-x^*\|^2 . \end{aligned} \tag{4.7}$$

Replacing μ^0 by $\bar{\mu}^0$ and $\underline{\mu}^0$ respectively, we obtain (4.5b-c). ■

Lemma 4.4 : Suppose that Assumption 4.1 holds and that $\{ x_i \}_{i=0}^{\infty}$ is a sequence constructed by Algorithm 3.2. Then for $i \geq 0$ and $\mu_i = (\mu_i^0, \ldots, \mu_i^m) \in U(x_i)$,

(i) $$F_{x_i}(x_{i+1}) \leq \frac{\alpha\beta c \mu_i^0}{C}[\psi^0(x^*) - \psi^0(x_i)] - \frac{\alpha\beta c(1 + (\gamma-1)\mu_i^0)}{C}\psi_+(x_i) \ ; \tag{4.8a}$$

(ii) $$\psi^0(x_{i+1}) - \psi^0(x^*) \leq (1 - \frac{\alpha\beta c \mu_i^0}{C})[\psi^0(x_i) - \psi^0(x^*)] + [\gamma - \frac{\alpha\beta c(1 + (\gamma-1)\mu_i^0)}{C}]\psi_+(x_i) \ ; \tag{4.8b}$$

(iii) $$\psi(x_{i+1}) \leq \frac{\alpha\beta c \mu_i^0}{C}[\psi^0(x^*) - \psi^0(x_i)] + [1 - \frac{\alpha\beta c(1 + (\gamma-1)\mu_i^0)}{C}]\psi_+(x_i) \ . \tag{4.8c}$$

Proof : (i) First we obtain a bound on the decrease in $F_{x_i}(\cdot)$ at the i–th iteration. Making use of Lemma 4.1(iii), Lemma 2.1(ii) and the fact that $\xi^0 \geq 0$ for all $\xi \in G(x)$, we find that for all $\lambda \in [0, 1/C]$,

$$F_{x_i}(x_i + \lambda h(x_i)) \ = \ \max\{ \psi^0(x_i + \lambda h(x_i)) - \psi^0(x_i) - \gamma\psi_+(x_i), \psi^j(x_i + \lambda h(x_i)) - \psi_+(x_i); j \in \mathbf{m} \}$$

$$\leq \max\{ -\xi^0 + \langle \xi, \lambda h(x_i)\rangle + \tfrac{1}{2}C|\lambda h(x_i)|^2 \mid \bar{\xi} \in G(x_i) \}$$

$$\leq \lambda \max\{ -\xi^0 + \langle \xi, h(x_i)\rangle + \tfrac{1}{2}|h(x_i)|^2 \mid \bar{\xi} \in G(x_i) \}$$

$$= \lambda\theta(x_i) . \tag{4.9a}$$

Therefore (3.2) is satisfied with $\lambda_i > \beta/C$, and thus

$$F_{x_i}(x_{i+1}) \leq \alpha\lambda_i\theta(x_i) \leq \alpha\beta\theta(x_i)/C . \tag{4.9b}$$

Next we relate $\theta(x_i)$, to $\psi^0(x^*) - \psi^0(x_i)$ and to $\psi_+(x_i)$. For any $\mu_i \in \mathbf{U}(x_i)$, there exist $\bar{\xi}_i^j = (\xi_i^{j0}, \xi_i^j)^T \in G^j(x_i), j \in \mathbf{M}$, such that

$$\bar{\xi}(x_i) = \sum_{j=0}^{m} \mu_i^j \bar{\xi}_i^j . \tag{4.10a}$$

Making use of Lemma 2.1(ii), (4.10a) and Lemma 4.1(iii), we obtain that

$$\theta(x_i) = \min_{h \in \mathbb{R}^n} \{ -\xi^0(x_i) + \langle \xi(x_i), h\rangle + \tfrac{1}{2}|h|^2 \}$$

$$= \min_{h \in \mathbb{R}^n} \{ \sum_{j=0}^{m} \mu_i^j (-\xi_i^{j0} + \langle \xi_i^j, h\rangle + \tfrac{1}{2}\|h\|^2) \}$$

$$\leq c \min_{h \in \mathbb{R}^n} \{ \sum_{j=0}^{m} \mu_i^j(-\xi_i^{j0} + \langle \xi_i^j, h/c\rangle + \tfrac{1}{2}c|h/c|^2) \}$$

$$\leq c \min_{h \in \mathbb{R}^n} \{ \mu_i^0(\psi^0(x_i + h/c) - \psi^0(x_i) - \gamma\psi_+(x_i)) + \sum_{j=1}^{m} \mu_i^j(\psi^j(x_i + h/c) - \psi_+(x_i)) \}. \tag{4.10b}$$

Replacing h by $c(x^* - x_i)$ in (4.10b) and making use of the fact that $\psi(x^*) \leq 0$, we obtain that

$$\theta(x_i) \leq c\{\mu_i^0(\psi^0(x^*) - \psi^0(x_i) - \gamma\psi_+(x_i)) + \sum_{j=1}^{m} \mu_i^j(\psi^j(x^*) - \psi_+(x_i))\}$$

$$\leq c \{\mu_i^0[\psi^0(x^*) - \psi^0(x_i)] - [1 + (\gamma - 1)\mu_i^0]\psi_+(x_i) \} . \tag{4.10c}$$

Combining (4.9b) and (4.10c) together, we obtain (4.8a).

Finally, (ii) and (iii) follow from (4.8a) and the fact that both $\psi^0(x_{i+1}) - \psi^0(x_i) - \gamma\psi_+(x_i) \leq F_{x_i}(x_{i+1})$ and $\psi(x_{i+1}) - \psi_+(x_i) \leq F_{x_i}(x_{i+1})$. ■

Lemma 4.5 : Suppose that Assumption 4.1 holds and that x^* is the unique solution of (2.1a). Then for any $\varepsilon \in (0,1)$, there exists a $\rho > 0$ such that for any sequence $\{ x_i \}_{i=0}^{\infty}$ constructed by Algorithm

3.2, if $x_i \in B(x^*,\rho) = \{ x \in \mathbb{R}^n \mid \|x-x^*\| \le \rho \}$, then

(i) $\psi_+(x_{i+1}) \le \delta_1(\varepsilon)\, \psi_+(x_i)$; (4.11a)

(ii) $\max\{ 0, \psi^0(x_{i+1}) - \psi^0(x^*) \} \le \delta_2(\varepsilon)[\max\{ 0, \psi^0(x_i) - \psi^0(x^*) \} + \gamma\psi_+(x_i)]$; (4.11b)

where

$$\delta_1(\varepsilon) = \max(0, 1 - \varepsilon\gamma\underline{\mu}^0\alpha\beta c/C) \in [0,1), \tag{4.11c}$$

$$\delta_2(\varepsilon) = 1 - \varepsilon\underline{\mu}^0\alpha\beta c/C \in (0,1)\ . \tag{4.11d}$$

Proof : (i) For any fixed $\varepsilon \in (0,1)$, we can pick $\varepsilon_1 > 0$, small enough, so that

$$1 - \varepsilon_1 - \varepsilon_1/(\gamma\underline{\mu}^0) \ge \varepsilon\ . \tag{4.12a}$$

By Lemma 4.3(iii), there exists a $\rho > 0$ such that for any $x_i \in B(x^*,\rho)$ and $\mu_i \in \mathrm{U}(x_i)$,

$$\underline{\mu}^0(1 - \varepsilon_1) \le \mu_i^0 \le \bar{\mu}^0(1 + \varepsilon_1)\ . \tag{4.12b}$$

Making use of (4.8c), (4.5b) and (4.12a-b), we find that for $x_i \in B(x^*,\rho)$,

$$\begin{aligned}
\psi(x_{i+1}) &\le \frac{\alpha\beta c\mu_i^0(1 - \bar{\mu}^0)}{C\bar{\mu}^0}\psi_+(x_i) + [1 - \frac{\alpha\beta c(1 + (\gamma - 1)\mu_i^0)}{C}]\psi_+(x_i) \\
&= [\ 1 + \frac{\alpha\beta c}{C}(-1 + \mu_i^0/\bar{\mu}^0 - \gamma\mu_i^0)\ ]\psi_+(x_i) \\
&\le [\ 1 + \frac{\alpha\beta c}{C}(-1 + \bar{\mu}^0(1 + \varepsilon_1)/\bar{\mu}^0 - \gamma\underline{\mu}^0(1 - \varepsilon_1))\ ]\psi_+(x_i) \\
&= [\ 1 - \frac{\gamma\underline{\mu}^0\alpha\beta c}{C}(\ 1 - \varepsilon_1 - \varepsilon_1/(\gamma\underline{\mu}^0)\ )\ ]\ \psi_+(x_i) \\
&\le (1 - \varepsilon\gamma\underline{\mu}^0\alpha\beta c/C)\psi_+(x_i)\ .
\end{aligned} \tag{4.13a}$$

Therefore (4.11a) must hold.

(ii) Making use of (4.8b) and the fact that $1 + (\gamma - 1)\mu_i^0 \ge \gamma\mu_i^0$, we obtain that for $x_i \in B(x^*,\rho)$,

$$\begin{aligned}
\psi^0(x_{i+1}) - \psi^0(x^*) &\le (1 - \frac{\alpha\beta c\mu_i^0}{C})[\psi^0(x_i) - \psi^0(x^*)] + [\gamma - \frac{\alpha\beta c(1 + (\gamma - 1)\mu_i^0)}{C}]\psi_+(x_i) \\
&\le (1 - \frac{\alpha\beta c\mu_i^0}{C})\max(0, \psi^0(x_i) - \psi^0(x^*)) + \gamma(1 - \frac{\alpha\beta c\mu_i^0}{C})\psi_+(x_i)
\end{aligned}$$

$$\le (1 - \frac{\alpha\beta c\mu_i^0}{C})[\max(0,\psi^0(x_i) - \psi^0(x^*)) + \gamma\psi_+(x_i)] \ . \tag{4.13b}$$

Since $1 - \varepsilon_1 \ \ge \ \varepsilon$ and $\mu_i^0 \ \ge \ \underline{\mu}^0(1-\varepsilon_1)$, $\mu_i^0 \ \ge \ \underline{\mu}^0\varepsilon$. Thus, (4.11b) must hold. ∎

We are now ready to establish the linear convergence of the Algorithm 3.2.

Theorem 4.1 : Suppose that Assumption 4.1 holds and that x^* is the unique solution of (2.1a).

(i) Any sequence $\{ x_i \}_{i=0}^{\infty}$ constructed by Algorithm 3.2 converges to x^*.

(ii) Let $\delta_1(\varepsilon)$, $\delta_2(\varepsilon)$ be defined as in (4.11c), (4.11d), respectively. Then for any $\varepsilon \in (0,1)$, there exists a $\rho > 0$ such that for any sequence $\{ x_i \}_{i=0}^{\infty}$ constructed by Algorithm 3.2,

(a) If $\psi(x_i) > 0$ and $x_i \in B(x^*,\rho)$, then

$$\psi(x_{i+1}) \le \delta_1(\varepsilon)\ \psi(x_i) \ . \tag{4.14a}$$

(b) If $\psi(x_i) \le 0$ and $x_i \in B(x^*,\rho)$, then

$$\psi^0(x_{i+1}) - \psi^0(x^*) \le \delta_2(\varepsilon)[\psi^0(x_i) - \psi^0(x^*)\ ] \ . \tag{4.14b}$$

(iii) If $\gamma > C/(\underline{\mu}^0\alpha\beta c)$, then there exists a $\rho > 0$ such that for any sequence $\{ x_i \}_{i=0}^{\infty}$ constructed by Algorithm 3.2, if $x_i \in B(x^*,\rho)$, then $\psi(x_{i+1}) \ \le \ 0$.

Proof : (i) Since the functions $\phi^j(\cdot,\cdot)$, $j = 0,1,...,m$, satisfy Assumption 4.1(i), $\psi(\cdot)$ has bounded level sets. Making use of Lemma 4.1, we conclude that $\psi_+(x_i) \le \psi_+(x_0)$. Hence, $\{ x_i \}_{i=0}^{\infty}$ is bounded. Since x^* is the unique zero of $\theta(\cdot)$, it follows from Theorem 3.1 that $\{ x_i \}_{i=0}^{\infty}$ has only one accumulation point x^*. Therefore $\{ x_i \}_{i=0}^{\infty}$ converges to x^*.

(ii) This part follows directly from Lemma 4.5 and the fact that (a) $\psi_+(x) = \psi(x)$ when $\psi(x) > 0$ and (b) $\psi_+(x) = 0$ and $\psi^0(x) - \psi^0(x^*) > 0$ when $\psi(x) \le 0$.

(iii) Since $1 - \gamma\underline{\mu}^0\alpha\beta c/C < 0$, we can pick an $\varepsilon \in (0,1)$ such that $\delta_1(\varepsilon) = 0$. Then, the desired result follows from Lemma 4.5 for this particular ε. ∎

Corollary 4.1 : Suppose that Assumption 4.1 holds and that x^* is the unique solution of (2.1a). Then for any sequence $\{ x_i \}_{i=0}^{\infty}$ constructed by Algorithm 3.2,

(i) If $\psi(x_i) > 0$ for all $i \geq 0$, then

$$\limsup_{i\to\infty} \frac{\psi(x_{i+1})}{\psi(x_i)} \leq 1 - \frac{\gamma\mu^0\alpha\beta c}{C} . \tag{4.14c}$$

(ii) If there exists an $i_0 \in \mathbf{N}$ such that $\psi(x_{i_0}) \leq 0$, then

$$\limsup_{i\to\infty} \frac{\psi^0(x_{i+1}) - \psi^0(x^*)}{\psi^0(x_i) - \psi^0(x^*)} \leq 1 - \frac{\mu^0\alpha\beta c}{C} . \tag{4.14d}$$

■

The following theorem establishes the R-linear convergence of the iterates constructed by Algorithm 3.2.

Theorem 4.2 : Suppose that Assumption 4.1 holds and that x^* is the unique solution of (2.1a). Then for any $\varepsilon \in (0,1)$, there exist $\rho_1 \geq \rho_0 > 0$ such that for any sequence $\{ x_i \}_{i=0}^{\infty}$ constructed by Algorithm 3.2, if i_0 is such that $\mathring{x}_{i_0} \in B(x^*,\rho_0)$, then

$$\| x_i - x^* \| \leq \rho_1 [\delta_3(\varepsilon)^{1/2}]^{i - i_0}, \quad i = i_0, i_0 + 1, i_0 + 2, \ldots, \tag{4.15a}$$

where

$$\delta_3(\varepsilon) \triangleq \max \{\delta_1(\varepsilon), \delta_2(\varepsilon)\} = 1 - \min\{1,\gamma\}\varepsilon\mu^0\alpha\beta c/C, \tag{4.15b}$$

with $\delta_1(\varepsilon)$, $\delta_2(\varepsilon)$ defined by (4.11c), (4.11d), respectively.

Proof : For any fixed $\varepsilon \in (0,1)$, we pick a $\varepsilon_1 \in (\varepsilon,1)$. Then there exists $\rho_1 > 0$ such that Lemma 4.5 holds for $\varepsilon = \varepsilon_1$ and $\rho = \rho_1$. Since $\delta_3(\varepsilon_1) < \delta_3(\varepsilon)$, we claim that there exists a $K > 0$ such that for all $i \geq 0$,

$$\left[\frac{\delta_3(\varepsilon_1)}{\delta_3(\varepsilon)}\right]^i < K . \tag{4.16a}$$

Since $\psi^0(\cdot)$ and $\psi(\cdot)$ are continuous, we can find a $\rho_0 \in (0,\rho_1)$ such that

$$\max_{x \in B(x^*,\rho_0)} [\, \mu^0 \max(0,\psi^0(x) - \psi^0(x^*)) + (\mu^0\gamma K + 1 - \mu^0)\psi_+(x) \,] \leq \frac{c}{2}(\rho_1)^2 . \tag{4.16b}$$

We shall prove by induction that (4.15a) holds for any sequence $\{ x_i \}_{i=0}^{\infty}$ constructed by Algorithm 3.2. Let x_{i_0} be the first element of this sequence in the ball $B(x^*,\rho_0)$.

Since $\rho_0 \le \rho_1$, (4.15a) holds for $i = i_0$. Suppose that (4.15a) holds for $i = i_0, \cdots, k+i_0$. Then $x_i \in B(x^*, \rho_1)$ for $i = i_0, \cdots, k+i_0$. Thus, due to the selection of ρ_1, for $i = i_0, \cdots, k+i_0$, x_i satisfies inequalities (4.11a-b), where ε is replaced by ε_1, i.e.,

$$\psi_+(x_{i+1}) \le \delta_1(\varepsilon_1)\, \psi_+(x_i)\ , \tag{4.17a}$$

$$\max(0, \psi^0(x_{i+1}) - \psi^0(x^*)) \le \delta_2(\varepsilon_1)[\max(0, \psi^0(x_i) - \psi^0(x^*)) + \gamma\psi_+(x_i)\ ]\ . \tag{4.17b}$$

Therefore, we can recursively obtain that for $i = i_0, \cdots, k+i_0$,

$$\psi_+(x_{i+1}) \le [\delta_3(\varepsilon_1)]^{i+1-i_0}\psi_+(x_{i_0})\ , \tag{4.18a}$$

$$\max(0, \psi^0(x_{i+1}) - \psi^0(x^*)) \le [\delta_3(\varepsilon_1)]^{i+1-i_0}[\ \max(0, \psi^0(x_{i_0}) - \psi^0(x^*)) + \gamma(i+1-i_0)\psi_+(x_{i_0})\ ]\ . \tag{4.18b}$$

Making use of (4.5c), (4.18a-b) and the fact that $\delta_3(\varepsilon) > \delta_3(\varepsilon_1)$, we obtain that

$$\begin{aligned}
\frac{c}{2}\|x_{k+1+i_0} - x^*\|^2 &\le \underline{\mu}^0[\psi^0(x_{k+1+i_0}) - \psi^0(x^*)] + (1 - \underline{\mu}^0)\psi_+(x_{k+1+i_0}) \\
&\le [\delta_3(\varepsilon_1)]^{k+1}\ [\underline{\mu}^0 \max\ \{0, \psi^0(x_{i_0}) - \psi^0(x^*)\} + (\underline{\mu}^0\ \gamma\ (k+1) + 1 - \underline{\mu}^0\ )\ \psi_+(x_{i_0})] \\
&\le [\delta_3(\varepsilon)]^{k+1}\ [\ \underline{\mu}^0 \max\ \{0, \psi^0(x_{i_0}) - \psi^0(x^*)\} \\
&\qquad + (\underline{\mu}^0\ \gamma\ (k+1)\ (\delta_3(\varepsilon_1)/\delta_3(\varepsilon))^{k+1} + 1 - \underline{\mu}^0)\psi_+(x_{i_0})]\ .
\end{aligned} \tag{4.19a}$$

Since $x_{i_0} \in B(x^*, \rho_0)$, we obtain from (4.16a-b) and (4.19a) that

$$\begin{aligned}
\frac{c}{2}\|x_{k+1+i_0} - x^*\|^2 &\le [\delta_3(\varepsilon)]^{k+1}\ [\underline{\mu}^0 \max\ \{0, \psi^0(x_{i_0}) - \psi^0(x^*)\} + (\underline{\mu}^0\ \gamma\, K + 1 - \underline{\mu}^0\ )\ \psi_+(x_{i_0})] \\
&\le \frac{c}{2}[\delta_3(\varepsilon)]^{k+1}(\rho_1)^2\ .
\end{aligned} \tag{4.19b}$$

Consequently, (4.15a) holds for $i = k+1+i_0$. Therefore the induction proof of (4.15a) is completed. ■

Corollary 4.2 : Suppose that Assumption 4.1 holds and that x^* is the unique solution of (2.1a). Let $\delta_3(\varepsilon)$ be defined as in (4.15b). Then for any sequence $\{\ x_i\ \}_{i=0}^{\infty}$ constructed by Algorithm 3.2,

$$\limsup_{i \to \infty}\ (\|x_i - x^*\|)^{(1/i)} \le [\delta_3(1)]^{1/2}\ . \tag{4.20}$$

■

5. NUMERICAL RESULTS

Since the exact calculation of the global maxima of a function is not a numerically implementable operation, numerical methods for solving Problem (2.1a) must discretize the intervals $\mathbf{Y}_j$. The discretization may be either fixed, or variable, see, e.g., [He.1, Kle.1, Pol.1]. Assuming that the cost function was originally differentiable, i.e., that Y_0 contains only one point, the discretized problem is of the form of Problem (2.1a), but contains as many constraint functions $\psi^j(\cdot)$ as the number of discretization points used, with the associated sets $\mathbf{Y}_j$ containing only one point. When the cost is a max function, an additional variable has to be introduced on discretization in order to reduce the discretized problem to the form (2.1a).

In view of the above, we have implemented the Algorithms 3.1 and 3.2 for the case when all the functions $\psi^j(\cdot)$, $j = 0,1,2,..,m$, in Problem (2.1a), are differentiable, i.e., for the case when the sets $\mathbf{Y}_j$, $j = 0,1,2,..,m$, contain only one point.

Both Algorithm 3.1 and Algorithm 3.2 were coded in C and were executed on a SUN 3/140 Workstation. The search direction h_i and the optimality function value θ_i, in the Step 1 of the algorithms, were computed by solving the quadratic program (2.7a) by means of the method based on support functions in [Hig.1]. The stopping criterion was $\theta_i \leq -\varepsilon$, where ε is a given positive number. In the experiments below, the algorithm parameters were set at $\alpha = 0.9, \beta = 0.9, \gamma = 1.0, \varepsilon = 0.000001$, unless stated otherwise.

Algorithm 3.2 (Our unified phase I-phase II method) was compared with Algorithm 3.1 (phase I-phase II method proposed in [Pir.1], [Pol.1] and [Pol.2]) on several well-known problems. For each problem, two tests were carried out. The first test used a feasible initial point, while the second test used an infeasible initial point. Table 5.1 summarizes the performance of the two algorithms on these problems. We evaluate their performance by comparing the number of iterations, the number of function evaluations (one gradient evaluation was counted as n function evaluations), and the CPU time (seconds) which they required to achieve the given degree of accuracy , i.e, $\theta_i \leq -\varepsilon$. In Figure 5.1-5.5, the data generated by Algorithm 3.1 are plotted with dotted lines, while the data generated by Algorithm 3.2 are plotted with solid lines. The test problems and the detailed results are as follows:

Rosen-Suzuki's Problem [Con.1]:

$$\min \ x_1^2 + x_2^2 + 2x_3^2 + x_4^2 - 5x_1 - 5x_2 - 21x_3 - 7x_4$$

$$\text{s.t.} \ 2x_1^2 + x_2^2 + x_3^2 + 2x_1 - x_2 - x_4 - 5 \le 0 ,$$

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_1 - x_2 + x_3 - x_4 - 8 \le 0,$$

$$x_1^2 + 2x_2^2 + x_3^2 + 2x_4^2 - x_1 - x_4 - 10 \le 0 .$$

In this problem, all of the functions are strongly convex. The feasible initial point (0,0,0,0) and the infeasible point (2,4,8,1) were used. Both algorithms converged to the minimum value of -44 at (0,1,2,−1). In the first trial, Algorithm 3.2 performed about the same as Algorithm 3.1, see Figure 5.1. In the second trial, the optimality function in Algorithm 3.2 decreased faster, approximately by a factor of 10, than that in Algorithm 3.1, see Figure 5.2. Figure 5.3 and 5.4 plot the cost function values $\psi^0(x_i)$ and constraint function values $\psi(x_i)$ as functions of iteration number i, for the first 20 iterations. It was observed that the cost was increased a lot in the first iteration of Algorithm 3.1 though the feasible region was almost reached.

Wong's Problem [Asa.1]:

$$\min \ (x_1 - 10)^2 + 5(x_2 - 12)^2 + x_3^4 + 3(x_4 - 11)^2 + 10x_5^6 + 7x_6^2 + x_7^4 - 4x_6x_7 - 10x_6 - 8x_7$$

$$\text{s.t.} \ 2x_1^2 + 3x_2^4 + x_3 + 4x_4^2 + 5x_5 - 127 \le 0,$$

$$7x_1 + 3x_2 + 10x_3^2 + x_4 - x_5 - 282 \le 0 ,$$

$$23x_1 + x_2^2 + 6x_6^2 - 8x_7 - 196 \le 0,$$

$$4x_1^2 + x_2^2 - 3x_1x_2 + 2x_3^2 + 5x_6 - 11x_7 \le 0 .$$

In this problem, all the constraint functions are strongly convex, while the cost function has a single nonconvex term. The minimum value for the problem is 680.63 at (2.33,1.95,−0.48,4.37, −0.62,1.04,1.59). The feasible initial point (1,2,0,4,0,1,1) and the infeasible initial point (3,3,0,5,1,3,0) were used. It was observed that both algorithms generated the same iterate points in the first trial. Figure 5.5 shows that Algorithm 3.2 performed better in the second trial. Note that optimality function does not monotonous decrease, but it almost decrease linearly in the sense that $\log(-\theta(x_i))$ as a function of i is almost linear.

Quadratic Problem:

$$\min \; 3(x_1 - 1.4)^2 + (x_2 - 1)^2$$

$$\text{s.t.} \; (x_1 - 0.7)^2 + x_2^2 - 1 \leq 0$$

$$2(x_1 + 0.7)^2 + 0.5x_2^2 - 1 \leq 0 \, .$$

In this problem all the functions are strongly convex. The minimum value for the problem is 6.4235 which occurs at $\hat{x} = (-0.0202, 0.3896)$. The feasible initial point (-0.3,0.0) and the infeasible initial point (2.2,1.6) were used. Figure 5.6 displays the first three iterates constructed by the algorithms from the infeasible initial point $x_0 = (2.2, 1.6)$, where $\{ \bar{x}_1, \bar{x}_2 \}$ and $\{ x_1, x_2 \}$ are generated by Algorithm 3.1 and Algorithm 3.2, respectively, $\hat{x}$ is the optimum point, and the intersection area of the two ellipsoids is the feasible region. It is observed that x_1 is closer to the optimum point than $\bar{x}_1$, although x_1 is little bit further away from the feasible region than $\bar{x}_1$. The reason is that Algorithm 3.2 does not sacrifice as much of an increase in the cost in order to approach the feasible region. That is why Algorithm 3.2 performs better than Algorithm 3.1 on the Rosen-Suzuki's Problem, on Wong's Problem and the Quadratic Problem when an infeasible initial point is used.

	Algorithm 3.1			Algorithm 3.2		
Problem	iterations	function evaluations	time	iterations	function evaluations	time
Rosen-Suzuki.1	76	2417	2.88	77	2473	3.04
Rosen-Suzuki.2	68	2138	2.36	55	1689	1.86
Wong.1	157	23,286	21.48	157	23,286	21.54
Wong.2	171	24,697	22.48	151	22,241	20.88
Quadratic.1	48	586	0.76	49	601	0.76
Quadratic.2	50	620	0.72	43	550	0.68

Table 5.1: Summary of Numerical Results

6. CONCLUSION

This paper completes a cycle in which methods of feasible directions were taken from the situation where separate problem formulations had to be used for phase I and for phase II, to combined phase I-phase II methods, such as Algorithm 3.1, which used a unified search direction rule in all of

$\mathbb{R}^n$, but different step size rules, depending on whether the current iterate was feasible or not, to the unified phase I-phase II method of feasible directions (Algorithm 3.2) proposed in this paper, which uses the same search direction rule and step size rule in all of $\mathbb{R}^n$. The new algorithm is simpler theoretically and is somewhat more efficient numerically because it sacrifices less in cost increases in order to achieve feasibility.

Acknowledgement. The C code of Algorithm 3.1 was provided by Joseph Higgins.

7. REFERENCES

[Arm.1] L. Armijo, "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] C. Berge, *Topological Spaces*, Macmillan, New York, 1963.

[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.

[He.1] L. He and E. Polak, "An Optimal Diagonalization strategy for the Solution of a Class of Optimal Design Problems", *Electronics Research Laboratory, University of California, Berkeley*, Memorandum No. UCB/ERL M88/41, 1988. (to appear in IEEE Trans. Automat. Control)

[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. (to appear in J. of Optim. Theory and Appl.)

[Hua.1] P. Huard, "The Method of Centers", in *Nonlinear Programming*, McGraw-hill, New York, 1972.

[Kle.1] R. Klessig and E. Polak, "An adaptive precision gradient method for optimal control", *SIAM J. Control,* Vol. 11, No. 1, pp. 80-93, 1973.

[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, *Computational Methods in Optimization: A Unified Approach,* Academic Press, New York, 1971.

[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. 61-73, 1979.

[Pol.3] E. Polak, "On the Mathematical Foundations of Nondifferentiable Optimization in Engineering Design", *SIAM Review,* pp. 21-91, March 1987.

[Top.1] D. M. Topkis and A. F. Veinott, Jr., "On the Convergence of Some Feasible Direction Algorithms for Nonlinear Programming", *SIAM Journal on Control,* Vol. 5, No. 2, pp. 268-279, 1967.

[Zou.1] G. Zoutendijk, *Methods of Feasible Directions,* Elsevier, Amsterdam, 1960.

[Zuk.1] S. I. Zukhovitskii, R. A. Polyak and M. E. Primak "An Algorithm for the Solution of Convex Programming Problems", *DAN USSR,* Vol. 153, No. 5, pp. 991-1000, 1963.

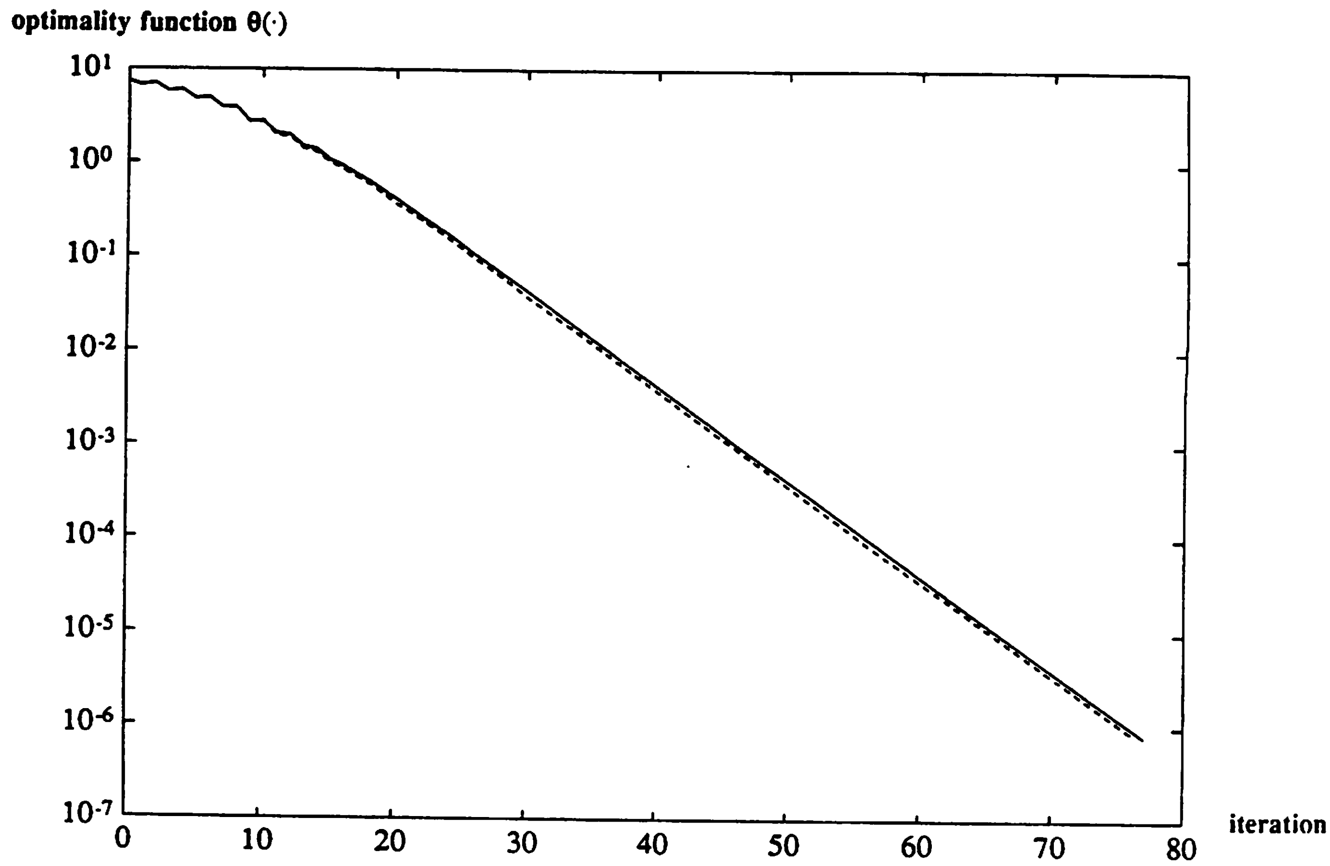

Figure 5.1. RosenSuzuki's Problem with feasible initial point

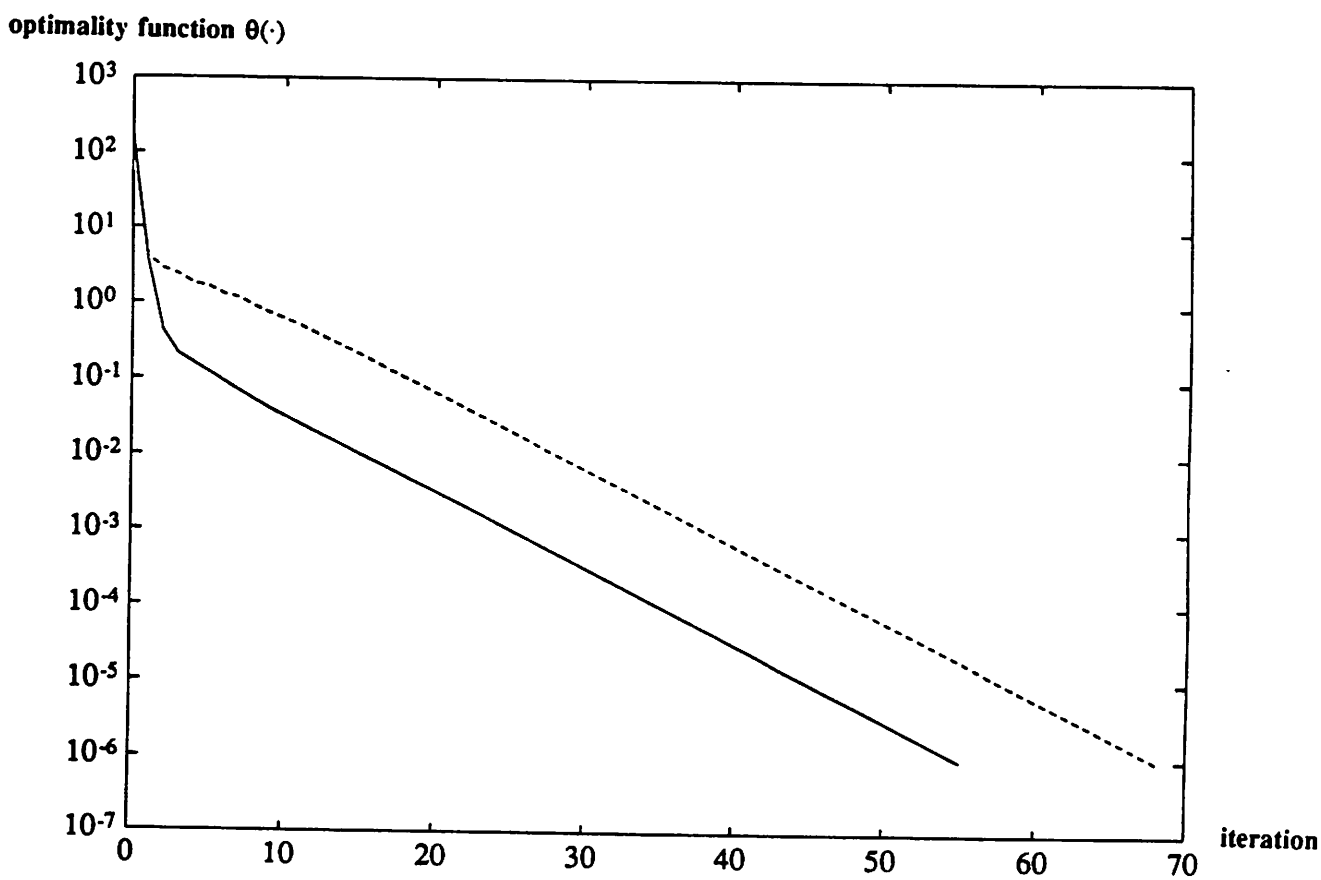

Figure 5.2. RosenSuzuki's Problem with infeasible initial point

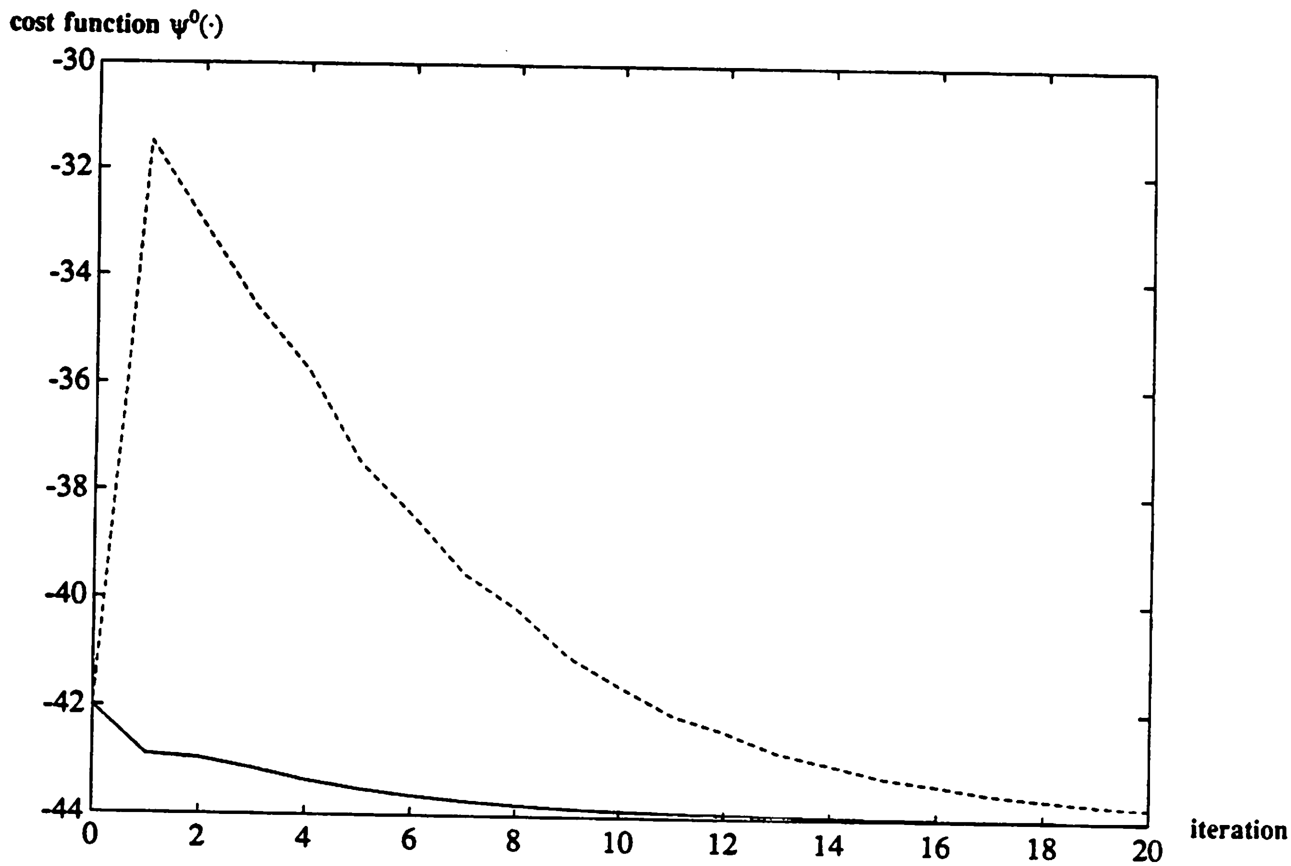

Figure 5.3. RosenSuzuki's Problem with infeasible initial point

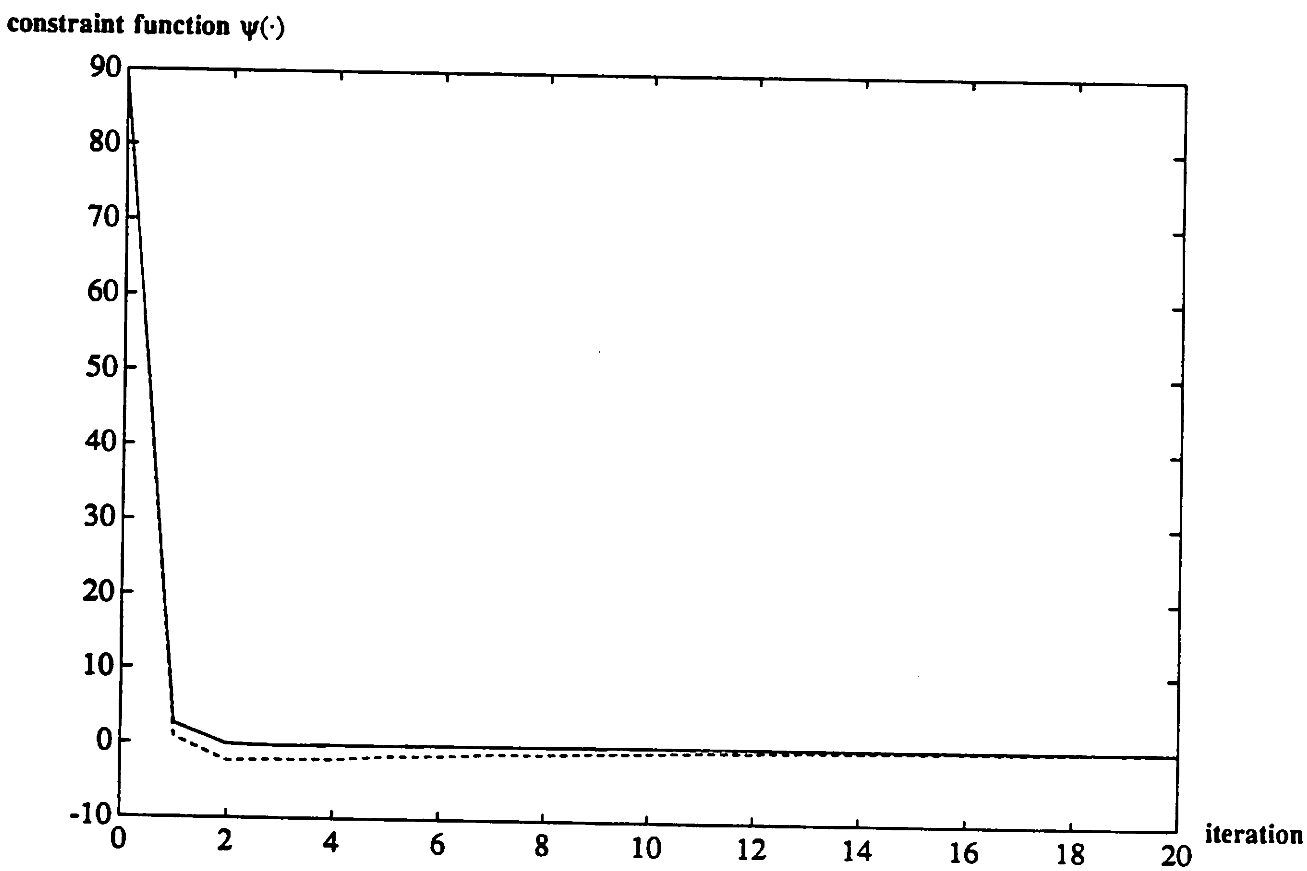

Figure 5.4. RosenSuzuki's Problem with infeasible initial point

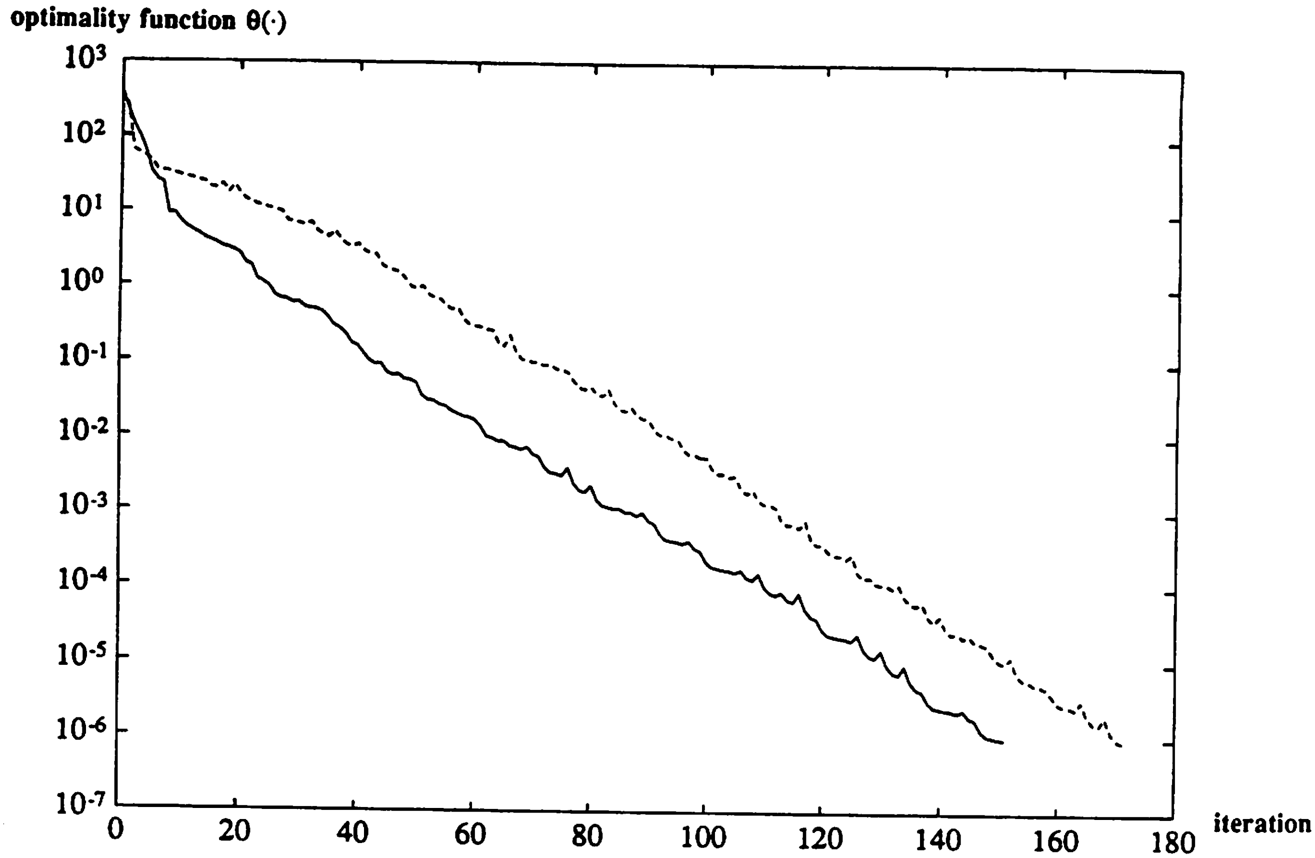

Figure 5.5. Wong's Problem with infeasible initial point

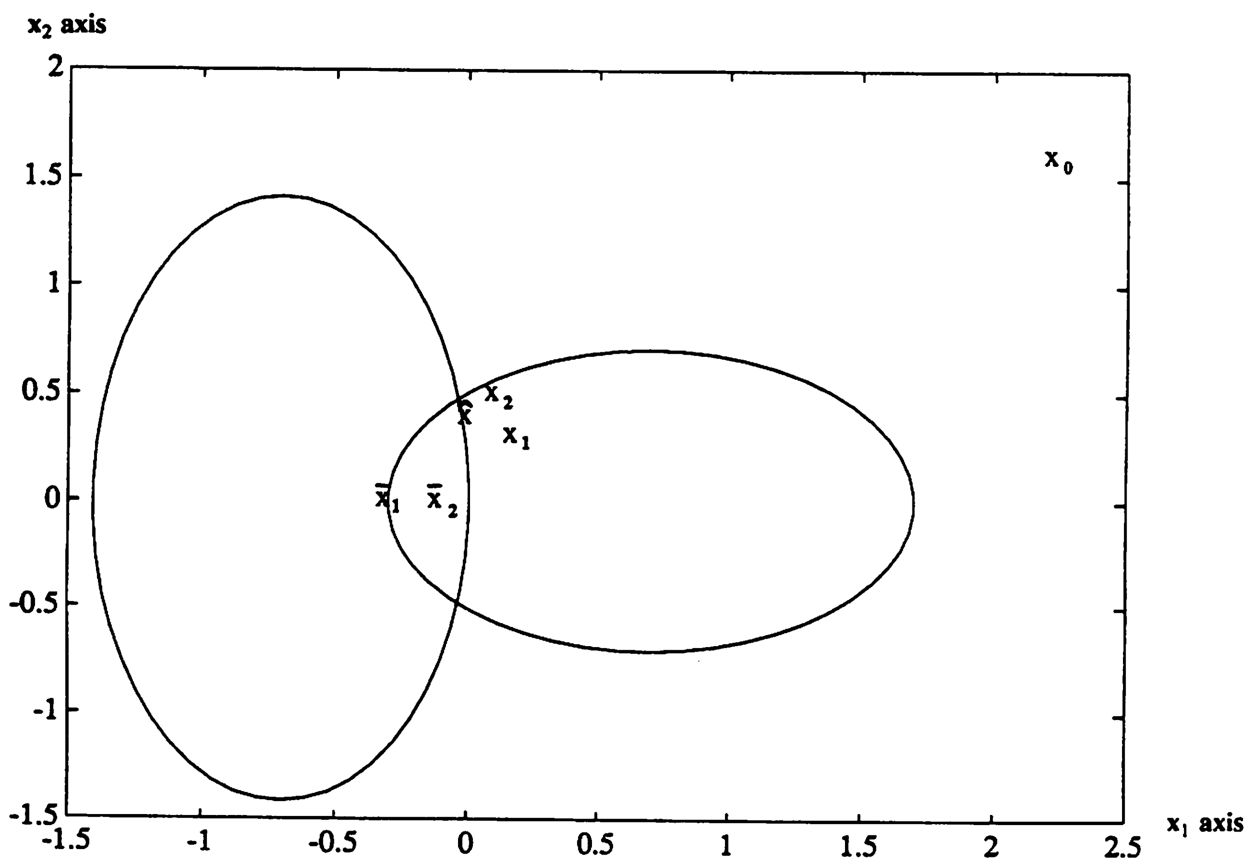

Figure 5.6. Quadratic Problem with infeasible initial point