RATE PRESERVING DISCRETIZATION STRATEGIES FOR SEMI-INFINITE PROGRAMMING AND OPTIMAL CONTROL

by

E. Polak and L. He

Memorandum No. UCB/ERL M89/112

25 September 1989

RATE PRESERVING DISCRETIZATION STRATEGIES FOR SEMI-INFINITE PROGRAMMING AND OPTIMAL CONTROL

by

E. Polak and L. He

Memorandum No. UCB/ERL M89/112

25 September 1989

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

RATE PRESERVING DISCRETIZATION STRATEGIES FOR SEMI-INFINITE PROGRAMMING AND OPTIMAL CONTROL

by

E. Polak and L. He

Memorandum No. UCB/ERL M89/112

25 September 1989

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

RATE PRESERVING DISCRETIZATION STRATEGIES FOR SEMI-INFINITE PROGRAMMING AND OPTIMAL CONTROL*

E. Polak** and L. He**

ABSTRACT

Neither semi-infinite programming nor optimal control problems can be solved without discretization: i.e., decomposition of the original problems into an infinite sequence of finite dimensional, finitely described optimization problems. We present three sets of discretization refinement rules: (i) for unconstrained semi-infinite minimax problems, (ii) for constrained semi-infinite problems, and (iii) for unconstrained optimal control problems. These rules are built into a master algorithm which calls certain linearly converging algorithms for finite dimensional, finitely described optimization problems. The discretization refinement rules ensure that the sequences constructed by the overall scheme converge to a solution of the original problem with the same rate constant as applies for the finite dimensional, finitely described approximations. Hence the resulting scheme is more efficient than fixed discretization.

KEY WORDS

approximation theory, minimax, semi-infinite programming, optimal control, linear convergence.

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

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

1. INTRODUCTION

The numerical solution of a semi-infinite optimization or optimal control problem always involves some form of discretization which decomposes the original optimization problem, **P**, into an infinite sequence of *finite dimensional, finitely described* optimization problems, $\mathbf{P}_q$, $q = 1,2,3,\ldots,$ which approximate **P** more and more closely and which are therefore of increasing computational complexity[1]. There is considerable empirical evidence to suggest that the computationally most efficient approach is to increase discretization gradually, using a process which constructs iterates that approach a stationary point for a problem $\mathbf{P}_q$ until some test is satisfied and then carry over the last iterate as a starting point for problem $\mathbf{P}_{q+1}$, until the value of q is increased to some preassigned maximum value q^*, rather than to solve $\mathbf{P}_{q^*}$ directly. The heuristic explanation of the success of this approach is that far from a solution, coarse discretization does not appear to interfere with progress towards a solution, while resulting in considerable computational savings per iteration over fine discretization.

There are two basic approaches to the construction of discretization refinements tests. The first is based on the concept of *diagonalization* (see [Dan.1, He.1]) which can be used under minimal consistency conditions. For example, suppose there is a family of continuous, negative-valued functions $\{ \theta_q(\cdot) \}$ such that if $\hat{x}_q$ is optimal for $\mathbf{P}_q$, then $\theta_q(\hat{x}_q) = 0$, and suppose that there is a continuous function $\theta(\cdot)$ such that (i) $\theta_q(x) \to \theta(x)$ as $q \to \infty$, uniformly in x (in a bounded set), and (ii) if $\hat{x}$ is a solution to **P**, then $\theta(\hat{x}) = 0$. Then a diagonalization scheme would consist of computing points x_q such that $\theta_q(x_q) \geq -1/q$. The main disadvantage of a diagonalization approach is that all the convergence statements are in terms of the points x_q only (discarding the intermediate points constructed on the way to x_q), e.g., "all the accumulation points of the sequence $\{ x_q \}_{q=q_0}^{\infty}$ are stationary points for the problem **P**". The second approach is more subtle: it starts with a *conceptual* algorithm for solving **P** (see [Kle.1, Pol.1]) and uses the problems $\mathbf{P}_q$ in the construction of an *implementation* of this algorithm. It requires stronger consistency conditions than diagonalization; these, fortunately, are frequently satisfied in practice. In return, it yields the considerable advantage that all the convergence statements are in terms of

---

[1] By this we mean that one iteration of a particular algorithm on problem $\mathbf{P}_{q+1}$ is more costly than one iteration of the same algorithm on problem $\mathbf{P}_q$.

the *entire* sequence constructed by the algorithm, e.g, "all the accumulation points of the entire sequence $\{ x_i \}_{i=0}^{\infty}$ constructed by the implementation are stationary points for the problem **P**". In [Kle.1, Pol.1] we find an abstract theory for the construction of *implementable* algorithms. As a particular application, we find in [Kle.1] an implementation of the method of steepest descent for solving continuous optimal control problems. In [Dun.1] this theory was used to implement a conditional gradient method for optimal control problems with ODE dynamics.

The discretization adjustment tests described in [Kle.1, Pol.1] are very basic: they make no use of optimality functions or rate of convergence properties of the conceptual algorithm. In this paper, we present three master discretization algorithms which differ from those based on the theory in [Kle.1, Pol.1] in two respects: (i) they use tests, based on optimality functions and a measure of the accuracy with which the discretized problems approximate the original problem, for increasing discretization, and (ii) unlike master algorithms based on the theory in [Kle.1, Pol.1], they demonstrably preserve the convergence rate constants of the conceptual algorithms that they implement. The reason for their superiority can be seen, as follows. First-order algorithms are characterized by a *rate constant*, $\eta \in (0,1)$, which appears in formulas of the form $e_{i+1} \leq \eta e_i$. The rate constant of first-order algorithms, applicable to discretized problems $\mathbf{P}_q$, depends basic constants such as bounds on second order derivatives and the John multiplier associated with the cost function. An examination of particular examples of discretized problems $\mathbf{P}_q$ shows that the entire family $\{ \mathbf{P}_q \}$ shares the *same* values of these constants which are inherited from the original problem **P**. Hence a particular first-order algorithm converges with the *same* rate constant on *every* member of the family $\{ \mathbf{P}_q \}$. Now suppose that a master discretization algorithm (say M) which calls a particular first-order algorithm (say A) as a subroutine, is linearly converging on **P**, with the *same* rate constant as A has on the problems $\mathbf{P}_q$. Then, given an initial point x_0, k iterations of **M** on **P**, yield an end point x_k^{M} and discretization parameter q_k, while k iterations of A on the problem $\mathbf{P}_{q_k}$ yield an end point x_k^{A}. Because of the same rate of convergence, we can expect that x_k^{M} and x_k^{A} are equally good approximations to a solution of **P**. However, the total computing time used to produce x_k^{M} must be less because the early iterations of A, as called by M, face coarser discretizations than those encountered by A in solving $\mathbf{P}_{q_k}$ directly. Hence the master discretization algorithm is more efficient than a fixed discretization scheme.

In Section 2, we present a master discretization algorithm, to be used in conjunction with the Pironneau-Polak-Pshenichnyi (PPP) minimax algorithm [Pir.1, Pol.3, Psh.1] for solving unconstrained semi-infinite minimax problems. Its rate of convergence constant is shown to be the same as that of the PPP algorithm, established in [Pir.1, Pol.3]. In Section 3, we present a master discretization algorithm, to be used in conjunction with the Polak-He unified steerable phase I - phase II method of feasible directions (USFD)[2] [Pol.4], for solving constrained semi-infinite optimization problems. its rate of convergence constant is shown to be the same as that of USFD. Finally, in Section 4, we present a master discretization algorithm, to be used in conjunction with the Armijo gradient method [Arm.1], for solving unconstrained optimal control problems. Its rate of convergence constant is shown to be the same as that of the Armijo method on composite function problems.[3]

We use standard notation: thus $L_\infty^m[0,T]$ denotes the space of equivalence classes of essentially bounded, measurable functions from $[0,T]$ into $\mathbb{R}^m$, $L_2^m[0,T]$ denotes the space of equivalence classes of square integrable functions from $[0,T]$ into $\mathbb{R}^m$, and $\|\cdot\|$, $\langle\cdot,\cdot\rangle$ denote the Euclidean norm and scalar product, respectively, in $\mathbb{R}^n$. For $A \in \mathbb{R}^{m\times n}$, $\|A\| \triangleq \max_{\|x\|=1}\|Ax\|$, for $u,v \in L_2^m[0,T]$, $\|u\|_2^2 \triangleq \int_0^T \|u(t)\|^2 dt$, $\langle u,v\rangle_2 \triangleq \int_0^T \langle u(t),v(t)\rangle dt$, for $u \in L_\infty^m[0,T]$, $\|u\|_\infty \triangleq \text{ess sup}_{t\in[0,T]} \|u(t)\|$, for $U \in L_2^{m\times n}[0,T]$, $\|U\|_2^2 \triangleq \int_0^T \|U(t)\|^2 dt$, and for $U \in L_\infty^{m\times n}[0,T]$, $\|U\|_\infty \triangleq \text{ess sup}_{t\in[0,T]}\|U(t)\|$.

2. MINIMAX PROBLEMS

We begin with minimax problems of the form

$$\text{MMP}: \min_{x\in\mathbb{R}^n} \psi(x) \tag{2.1a}$$

where

[2] This algorithm is the only phase I - phase II method of feasible directions that we were able to implement in such a way that once a feasible point for a problem P_{q_o} was found, terminating phase I on P_{q_o}, the algorithm remained in phase II for this and all the following problems P_q.

[3] Discrete optimal control problems have cost functions of the form $f(u) = g(x(N,u,x_o))$, and the Hessian of $f(\cdot)$ is only positive semi-definite, at best. Hence "standard" rate of convergence theory (see [Lue.1, Pol.1]) leads to the conclusion that the Armijo method converges on these problems sublinearly. The results, for linear dynamics, in [Pol.5], to be extended in this paper, show that this is not so.

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

where $\mathbf{l} \triangleq \{ 1, 2, \ldots, l \}$, $\phi^j : \mathbb{R}^n \times \mathbb{R}^{p_j} \rightarrow \mathbb{R}$ and $\mathbf{Y}_j$ is a compact subset of $\mathbb{R}^{p_j}$. We will assume that the functions $\phi^j(\cdot,\cdot)$ and their gradients $\nabla_x \phi^j(\cdot,\cdot)$ are Lipschitz continuous. Since the exact calculation of the global maxima of $\phi^j(\cdot,\cdot)$ over the compact set $\mathbf{Y}_j$ is not a numerically implementable operation, numerical methods for solving Problem (2.1a) must discretize the compact sets $\mathbf{Y}_j$. Hence, we introduce a family of approximating problems, parametrized by the discretization parameter $q \in \mathbb{N}$:

$$\mathbf{MMP}_q : \min_{x \in \mathbb{R}^n} \psi_q(x) \ , \tag{2.2a}$$

where

$$\psi_q(x) \triangleq \max_{j \in \mathbf{l}} \max_{y_j \in \mathbf{Y}_j} \phi_q^j(x, y_j) \ , \tag{2.2b}$$

and the functions $\phi_q^j(\cdot,\cdot)$ are constructed by linear interpolation of the $\phi^j(\cdot,\cdot)$ over a "triangulated" (uniform) grid in the sets Y_j. Thus, for example, when $Y_j = [0,1] \subset \mathbb{R}$, we divide this interval into q subintervals, and then we define $\phi_q^j(x,y)$ to be linear on each interval, so that $\phi_q^j(x,y) = \lambda\phi^j(x, i/q) + (1-\lambda)\phi^j(x, (i+1)/q)$ for $y = \lambda i/q + (1-\lambda)(i+1)/q$ and $\lambda \in [0,1]$ and $i = 0, 1, 2, \ldots, q-1$. When $Y_j = [0,1] \times [0,1] \subset \mathbb{R}^2$, we first break up the plane into small squares, with sides of length $1/q$, and each square is then divided into two triangles, using parallel diagonals. The function $\phi_q^j(x,y)$ is then defined as a continuous linear interpolation of $\phi^j(x,y)$ on this triangulated grid. We note that when defined in this manner, the evaluation of $\psi_q(x)$ is a finite process.

The master adaptive discretization algorithm that we will shortly introduce, calls the Pironneau-Polak-Pshenichnyi (PPP) minimax algorithm [Pir.1, Pol.2, Psh.1] as a subroutine. This algorithm computes search directions by evaluating an optimality function. Hence, proceeding as in [Pol.2], we define $\theta(x)$, $\theta_q(x)$ to be the *optimality functions* for problems MMP and problem $\mathbf{MMP}_q$, respectively, as follows:

$$\theta(x) \triangleq \min_{h \in \mathbb{R}^n} \max_{j \in \mathbf{l}} \max_{y_j \in \mathbf{Y}_j} \{ \phi^j(x, y_j) + \langle \nabla_x \phi^j(x, y_j), h \rangle + \tfrac{1}{2}\|h\|^2 - \psi(x) \} \ , \tag{2.3a}$$

$$\theta_q(x) \triangleq \min_{h \in \mathbb{R}^n} \max_{j \in \mathbf{l}} \max_{y_j \in \mathbf{Y}_j} \{ \phi_q^j(x, y_j) + \langle \nabla_x \phi_q^j(x, y_j), h \rangle + \tfrac{1}{2}\|h\|^2 - \psi_q(x) \} \ . \tag{2.3b}$$

When the functions $\phi_q^j(\cdot,\cdot)$ are defined by linear interpolation on a triangulated grid, (2.3b) is an ordinary quadratic programming problem which can be solved finitely using standard quadratic programming subroutines.

Let $h, h_q : \mathbb{R}^n \to \mathbb{R}$ be *search direction functions* defined by

$$h(x) \triangleq \arg \min_{h \in \mathbb{R}^n} \max_{j \in \mathbf{I}} \max_{y_j \in \mathbf{Y}_j} \{ \phi^j(x,y_j) + \langle \nabla_x \phi^j(x,y_j), h \rangle + \tfrac{1}{2}\|h\|^2 - \psi(x) \} , \tag{2.3c}$$

$$h_q(x) \triangleq \arg \min_{h \in \mathbb{R}^n} \max_{j \in \mathbf{I}} \max_{y_j \in \mathbf{Y}_j} \{ \phi_q^j(x,y_j) + \langle \nabla_x \phi_q^j(x,y_j), h \rangle + \tfrac{1}{2}\|h\|^2 - \psi_q(x) \} . \tag{2.3d}$$

The PPP minimax algorithms for solving MMP and MMP_q, use one of the above search direction functions (as appropriate) and an Armijo type step size rule, which requires two parameters $\alpha,\ \beta \in (0,1)$. Thus, the *conceptual* PPP algorithm, for solving MMP, described in [Pol.2], constructs iterates according to the rule

$$x_{i+1} = x_i + \lambda_i h(x_i), \tag{2.3e}$$

where

$$\lambda_i = \max\{ \beta^k \mid k \in \mathbb{N} ,\ \psi(x_i + \beta^k h(x_i)) - \psi(x_i) \le \alpha\beta^k\theta(x_i) \} , \tag{2.3f}$$

while the *implementable* PPP algorithm, for solving MMP_q, described in [Pir.1, Pol.2, Psh.1]), constructs iterates according to (2.3e), (2.3f), with $h(\cdot)$, $\psi(\cdot)$, and $\theta(\cdot)$ replaced by $h_q(\cdot)$, $\psi_q(\cdot)$, and $\theta_q(\cdot)$, respectively.

No matter how the approximating functions $\phi_q^j(\cdot,\cdot)$ are constructed, we will require that the functions $\phi_q^j(\cdot,\cdot)$ together with the functions $\phi^j(\cdot,\cdot)$ satisfy the following assumption[4]:

Assumption 2.1 :

(i) There exist constants $0 < K < \infty$ and $\tau > 0$ such that for all $x \in \mathbb{R}^n$ and all $q \in \mathbb{N}$,

$$| \psi(x) - \psi_q(x) | \le K/q^{\tau} . \tag{2.4a}$$

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

[4] This assumption is satisfied, with $\tau = 1$, by the two examples we gave using a *uniform* discretization grid, when the functions $\phi^j(x,\cdot)$ are at least Lipschitz continuous. When the functions $\nabla_y \phi^j(x,\cdot)$ are Lipschitz continuous, then our assumption is satisfied with $\tau = 2$.

$$\lim_{\substack{x' \to x \\ q \to \infty}} \phi_q(x', y_j) = \phi^j(x, y_j) \,, \quad \lim_{\substack{x' \to x \\ q \to \infty}} \nabla_x \phi_q(x', y_j) = \nabla_x \phi^j(x, y_j) \,. \tag{2.4b}$$

■

Lemma 2.1 : (i) For any $x \in \mathbb{R}^n$,

$$\theta(x) = -\min\{\, \xi^0 + \tfrac{1}{2}|\xi|^2 \mid (\xi^0, \xi)^T \in G(x) \,\} \,, \tag{2.5a}$$

$$\theta_q(x) = -\min\{\, \xi^0 + \tfrac{1}{2}|\xi|^2 \mid (\xi^0, \xi)^T \in G_q(x) \,\} \,, \tag{2.5b}$$

where

$$G(x) \triangleq \text{co}\left\{ \bigcup_{j \in \mathbf{I}} \bigcup_{y_j \in \mathbf{Y}_j} \begin{bmatrix} \psi(x) - \phi^j(x, y_j) \\ \nabla_x \phi^j(x, y_j) \end{bmatrix} \right\}, \tag{2.5c}$$

$$G_q(x) \triangleq \text{co}\left\{ \bigcup_{j \in \mathbf{I}} \bigcup_{y_j \in \mathbf{Y}_j} \begin{bmatrix} \psi_q(x) - \phi_q^j(x, y_j) \\ \nabla_x \phi_q^j(x, y_j) \end{bmatrix} \right\}. \tag{2.5d}$$

(ii) For any $x \in \mathbb{R}^n$, $\theta(x) = 0$ if and only if $0 \in \partial\psi(x)$[5], and $\theta_q(x) = 0$ if and only if $0 \in \partial\psi_q(x)$, i.e., the zeros of these functions are the stationary points of the corresponding problems.

(iii) Suppose that Assumption 2.1 holds. Then, for any $x \in \mathbb{R}^n$,

$$\lim_{\substack{x' \to x \\ q \to \infty}} \theta_q(x') = \theta(x) \,. \tag{2.6}$$

Proof : Both (i) and (ii) were established in [Pol.2].

(iii) Since the sets $\mathbf{Y}_j$ are compact, it follows from Assumption 2.1(ii) and (2.5c-d) that

$$\lim_{\substack{x' \to x \\ q \to \infty}} G_q(x') = G(x) \,. \tag{2.7}$$

Hence, (2.6) follows from (2.7) and the definitions (2.5a-b). ■

We can now state a master adaptive discretization algorithm which calls the PPP minimax algorithm as a subroutine, for solving the problem MMP.

[5] We denote the Clarke generalized gradient [Cla.1] of a locally Lipschitz function $f: \mathbb{R}^n \to \mathbb{R}$ at $x \in \mathbb{R}^n$ by $\partial f(x)$.

Adaptive Discretization Algorithm 2.1 (for MMP):

Data: $x_0 \in \mathbb{R}^n$, $q_{-1} \in \mathbb{N}$, $\alpha \in (0,1)$, $\beta \in (0,1)$, $D > 0$ and $\sigma > 1$.

Step 0: Set $i = 0$.

Step 1: Computer $q_i \in \mathbb{N}$, $\theta_i = \theta_{q_i}(x_i)$, and $h_i = h_{q_i}(x_i)$ such that $q_i \geq q_{i-1}$ and

$$D/q_i^{\tau} \leq [-\theta_{q_i}(x_i)]^{\sigma} . \tag{2.8}$$

Step 2: Computer the step size λ_i :

$$\lambda_i = \max\{ \beta^k \mid k \in \mathbb{N} \ , \ \psi_{q_i}(x_i + \beta^k h_i) - \psi_{q_i}(x_i) \leq \alpha\beta^k\theta_i \} . \tag{2.9}$$

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

Remark 2.1 : It follows from Lemma 2.1(iii) whenever $\theta(x_i) \neq 0$, Step 1 of Algorithm 2.1 yields a finite discretization parameter q_i. For simplicity, in the rest of this section, we assume that Algorithm 2.1 does not produce an iterate x_i such that $\theta(x_i) = 0$, so that the resulting value of q_i, in Step 1, is finite. ■

Lemma 2.2 : Suppose that $\psi(\cdot)$ is bounded from below and that the sequence of iterates $\{ x_i \}_{i=0}^{\infty}$ and corresponding sequence of discretization parameters $\{ q_i \}_{i=0}^{\infty}$ were constructed by Algorithm 2.1. Then $q_i \to \infty$ as $i \to \infty$.

Proof: For $i \in \mathbb{N}$, let q_i, θ_i, h_i and λ_i be defined as in Algorithm 2.1, and suppose that $q_i \to \infty$ as $i \to \infty$ does not hold. Then, since $\{ q_i \}_{i=0}^{\infty}$ is a nondecreasing sequence of integers, it follows that there exist $i_0, \hat{q} \in \mathbb{N}$, such that for all $i \geq i_0$, $q_i = \hat{q}$, and hence, in conjunction with (2.8), that there exists an $\varepsilon > 0$, such that $\theta_i \leq -\varepsilon$ for all $i \geq i_0$. Making use of (2.9) and the assumption that $\psi(\cdot)$ is bounded from below, we conclude that $\psi_{\hat{q}}(\cdot)$ is also bounded from below. Hence we obtain that $-\infty < \sum_{i=i_0}^{\infty} [\psi_{\hat{q}}(x_{i+1}) - \psi_{\hat{q}}(x_i)] \leq \sum_{i=i_0}^{\infty} \alpha\lambda_i\theta_i$. Referring to (2.5b), we see that if $(\xi_i^0, \xi_i) = \arg\min\{ \xi^0 + \frac{1}{2}\|\xi\|^2 \mid (\xi^0, \xi) \in G_{q_i}(x_i) \}$, then $h_i = -\xi_i$. Hence $\|h_i\|^2 \leq -2\theta_i$ and $-\theta_i \geq \varepsilon$ for $i \geq i_0$, we deduce that $\|h_i\|^2 \leq 2\theta_i^2/\varepsilon$ for all $i \geq i_0$. Hence for any $j > i \geq i_0$,

$$\|x_j - x_i\| \le \sum_{k=i}^{j-1} \|x_{k+1} - x_k\| \le \sum_{k=i}^{j-1} \lambda_k \|h_k\| \le \sum_{k=i}^{\infty} (2/\varepsilon)^{1/2} \lambda_k(-\theta_k) . \tag{2.10}$$

Therefore, $\{ x_i \}_{i=0}^{\infty}$ is a Cauchy sequence in $\mathbb{R}^n$, and hence it follows from Theorem 5.2b and Corollary 5.1 in [Pol.2] (which show that any accumulation point x^*, of $\{ x_i \}_{i=0}^{\infty}$, constructed by the PPP algorithm, satisfies $\theta_{\hat{q}}(x^*) = 0$) that $\theta_{\hat{q}}(x_i) \to 0$, contradicting the construction in (2.8). ■

Theorem 2.1 : Suppose that Assumption 2.1 holds, that $\psi(\cdot)$ is bounded from below, that the second derivatives $\partial^2\phi_q^j(x,y_j)/\partial x^2$ exist for all $q \in \mathbb{N}$, and that there exists an $M \in [1,\infty)$ such that for all $x \in \mathbb{R}^n$, $z \in \mathbb{R}^n$, $j \in \mathbf{I}$ and $q \in \mathbb{N}$,

$$\langle z, \frac{\partial^2\phi_q^j(x,y_j)}{\partial x^2} z\rangle \le M\|z\|^2 , \quad \forall \;\; y_j \in \mathbf{Y}_j . \tag{2.11}$$

Then any accumulation point $\hat{x}$ of the sequence of iterates $\{ x_i \}_{i=0}^{\infty}$, generated by Algorithm 2.1, satisfies $\theta(\hat{x}) = 0$.

Proof: First, we obtain a bound on the decrease in $\psi(\cdot)$ at the i-th iteration. Using (2.11), we obtain that

$$\psi_{q_i}(x_i + \lambda h_i) - \psi_{q_i}(x_i) = \max_{j \in \mathbf{I}} \max_{y_j \in \mathbf{Y}_j} \{ \phi_{q_i}^j(x_i+\lambda h_i, y_j) - \psi_{q_i}(x_i) \}$$

$$\le \max_{j \in \mathbf{I}} \max_{y_j \in \mathbf{Y}_j} \{ \phi_{q_i}^j(x_i, y_j) - \psi_{q_i}(x_i) + \langle \nabla_x \phi_{q_i}^j(x_i, y_j), \lambda h_i\rangle + \tfrac{1}{2}M\|\lambda h_i\|^2 \}. \tag{2.12}$$

In view of (2.3b) and the fact that $\phi_{q_i}^j(x_i, y_j) - \psi_{q_i}(x_i) \le 0$ and that $M \ge 1$, we find that for all $\lambda \in [0,1/M]$,

$$\psi_{q_i}(x_i + \lambda h_i) - \psi_{q_i}(x_i) \le \lambda\theta_{q_i}(x_i) . \tag{2.13}$$

Therefore (2.9) is satisfied with $\lambda_i \ge \beta/M$, and thus

$$\psi_{q_i}(x_{i+1}) - \psi_{q_i}(x_i) \le \alpha\lambda_i\theta_i \le \alpha\beta\theta_i/M . \tag{2.14}$$

Hence, it follows from the Assumption 2.1(i) that

$$\psi(x_{i+1}) - \psi(x_i) \le \alpha\beta\theta_i/M + 2K/q_i^{\tau} . \tag{2.15}$$

Next, since $\theta_i \le -D^{1/\sigma}/(q_i^\tau)^{1/\sigma}$, we have that

$$\psi(x_{i+1}) - \psi(x_i) \le -\frac{\alpha\beta D^{1/\sigma}}{M(q_i^\tau)^{1/\sigma}}[1 - \frac{2KM}{\alpha\beta D^{1/\sigma}(q_i^\tau)^{(\sigma-1)/\sigma}}] . \tag{2.16}$$

Since $\sigma > 1$ and since by Lemma 2.2, $q_i \to \infty$ as $i \to \infty$, there exists an $i_0 \in \mathbb{N}$ such that for all $i \ge i_0$,

$$\psi(x_{i+1}) - \psi(x_i) \le -\frac{\alpha\beta D^{1/\sigma}}{2M(q_i^\tau)^{1/\sigma}} . \tag{2.17}$$

Hence,

$$\psi(x_{i+1}) - \psi(x_{i_0}) \le -\sum_{k=i_0}^{i} \frac{\alpha\beta D^{1/\sigma}}{2M(q_k^\tau)^{1/\sigma}} . \tag{2.18}$$

Because $\psi(\cdot)$ is bounded from below, the left hand side of the above inequality is bounded from below, which leads to the conclusion that $\sum_{k=0}^{\infty} 1/(q_k^\tau)^{1/\sigma} < \infty$. Consequently, $\sum_{k=0}^{\infty} 1/(q_k^\tau) < \infty$. Next, returning to (2.15), we conclude that

$$\psi(x_{i+1}) - \psi(x_0) \le -\sum_{k=0}^{i} \alpha\beta(-\theta_k)/M + \sum_{k=0}^{i} 2K/q_k^\tau . \tag{2.19}$$

Since both $\psi(x_{i+1}) - \psi(x_0)$ and $\sum_{k=0}^{i} 2K/q_k^\tau$ are bounded, it follows that $\sum_{k=0}^{\infty} (-\theta_k) < \infty$. The desired result now follows from Lemma 2.1(iii). ■

Theorem 2.3 : Suppose that Assumption 2.1 holds, that the second derivatives $\partial^2\phi^j(x,y_j)/\partial x^2$, $\partial^2\phi_q^j(x,y_j)/\partial x^2$ exist for all $q \in \mathbb{N}$, and that there exist constants $0 < m < 1 < M < \infty$, such that for all $x \in \mathbb{R}^n$, $z \in \mathbb{R}^n$ and $j \in \mathbf{I}$,

$$m\|z\|^2 \le \langle z, \frac{\partial^2\phi_q^j(x,y_j)}{\partial x^2} z\rangle \le M\|z\|^2 , \quad \text{for all } y_j \in \mathbf{Y}_j , \ q \in \mathbf{N} , \tag{2.20a}$$

$$m\|z\|^2 \le \langle z, \frac{\partial^2\phi^j(x,y_j)}{\partial x^2} z\rangle \le M\|z\|^2 , \quad \text{for all } y_j \in \mathbf{Y}_j . \tag{2.20b}$$

Then any sequence $\{ x_i \}_{i=0}^{\infty}$, generated by Algorithm 2.1, converges to the unique solution $\hat{x}$ of problem MMP, and

$$\overline{\lim_{i \to \infty}} \frac{\psi(x_{i+1}) - \psi(\hat{x})}{\psi(x_i) - \psi(\hat{x})} \leq 1 - \alpha\beta m/M . \tag{2.21}$$

Proof: It follows from (2.20a), (2.20b) that the functions $\psi_q(\cdot)$ $q \in \mathbb{N}$, $\psi(\cdot)$ are strongly convex, and hence that they have unique minimizers. For any $q \in \mathbb{N}$, let $\hat{x}_q$ be the unique solution of the problem $\mathbf{MMP}_q$. First, we deduce from Assumption 2.1(i) that

$$|\psi_q(\hat{x}_q) - \psi(\hat{x})| \leq K/q^{\tau} . \tag{2.22}$$

Next, referring to [Pol.3], [Pol.5], we see that for all $x \in \mathbb{R}^n$ and $q \in \mathbb{N}$,

$$m[\psi_q(x) - \psi_q(\hat{x}_q)] \leq -\theta_q(x) \leq M[\psi_q(x) - \psi_q(\hat{x}_q)] . \tag{2.23}$$

Combining (2.14) and (2.23), we get

$$\psi_{q_i}(x_{i+1}) - \psi_{q_i}(x_i) \leq \frac{\alpha\beta m}{M}[\psi_{q_i}(\hat{x}_{q_i}) - \psi_{q_i}(x_i)] . \tag{2.24}$$

Adding $\psi_{q_i}(\hat{x}_{q_i})$ to both sides in (2.24) and rearranging terms, we obtain that

$$\psi_{q_i}(x_{i+1}) - \psi_{q_i}(\hat{x}_{q_i}) \leq [1 - \alpha\beta m/M][\psi_{q_i}(x_i) - \psi_{q_i}(\hat{x}_{q_i})] . \tag{2.25}$$

It now follows from (2.22) and Assumption 2.1(i) that

$$\psi(x_{i+1}) - \psi(\hat{x}) \leq [1 - \alpha\beta m/M][\psi(x_i) - \psi(\hat{x})] + 4K/q_i^{\tau} . \tag{2.26}$$

Next, making use of Assumption 2.1(i), (2.22), (2.23) and the fact that $-\theta_i \geq D^{1/\sigma}/(q_i^{\tau})^{1/\sigma}$, we obtain that

$$\begin{aligned} M[\psi(x_i) - \psi(\hat{x})] &\geq M[\psi_{q_i}(x_i) - \psi(\hat{x}_{q_i})] - 2MK/q_i^{\tau} \\ &\geq -\theta_i - 2MK/q_i^{\tau} \\ &\geq D^{1/\sigma}/(q_i^{\tau})^{1/\sigma} - 2MK/q_i^{\tau} \end{aligned} \tag{2.27}$$

Since $\sigma > 1$ and since by Lemma 2.2 $q_i \to \infty$ as $i \to \infty$, we conclude that there exists i_0 such that for all $i \geq i_0$,

$$M[\psi(x_i) - \psi(\hat{x})] \geq D^{1/\sigma}/[2(q_i^{\tau})^{1/\sigma}] . \tag{2.28}$$

It now follows from (2.26) and (2.28) that

$$\psi(x_{i+1}) - \psi(\hat{x}) \leq [1 - \frac{\alpha\beta m}{M} + \frac{8KM}{D^{1/\sigma}(q_i^{\tau})^{(\sigma-1)/\sigma}}][\psi(x_i) - \psi(\hat{x})] \ , \quad \text{for } i \geq i_0 \ . \tag{2.29}$$

Therefore, (2.21) follows from (2.29) and the fact that by Theorem 2.1, $q_i \to \infty$ as $i \to \infty$. Since $\psi(x_i) \to \psi(\hat{x})$ and $\hat{x}$ is the unique minimizer of $\psi(\cdot)$, $\{ x_i \}_{i=0}^{\infty}$ must converge to $\hat{x}$. ■

For comparison, referring to [Pol.3], we find the following result for the *conceptual* PPP minimax algorithm:

Theorem 2.4 [6] : Suppose that the second derivatives $\partial^2\phi^j(x,y_j)/\partial x^2$ exist and that there exist constants $0 < m < 1 < M < \infty$, such that for all $x \in \mathbb{R}^n$, $z \in \mathbb{R}^n$ and $j \in \mathbf{l}$, (2.20b) is satisfied. Then any sequence $\{ x_i \}_{i=0}^{\infty}$, generated by the PPP minimax algorithm defined, by (2.3e), (2.3f), converges to the unique solution $\hat{x}$ of problem **MMP**, and (2.21) holds. ■

We thus see that our Adaptive Discretization Algorithm 2.1 converges with exactly the same linear rate constant as the PPP minimax algorithm, whether applied to **MMP** or to any **MMP**$_q$. This leads us to the following observation as to the relative efficiency of our Adaptive Discretization Algorithm, under the assumptions in Theorem 2.4. Suppose that we are given an initial point x_0 (sufficiently close to $\hat{x}$, the solution of **MMP**), and that we perform k iterations, ending with a point x_k and discretization parameter q_k. If, instead, we perform k iterations on the problem **MMP**$_{q_k}$, ending at a point x'_k, we cannot expect to have $\psi(x'_k) < \psi(x_k)$. However, the total computing time used on solving **MMP**$_{q_k}$ must be longer, because the early iterations of Algorithm 2.1 use a coarser discretization. Hence Algorithm 2.1 is more efficient than a fixed discretization scheme.

3. CONSTRAINED SEMI-INFINITE OPTIMIZATION PROBLEMS

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

$$\mathbf{CSP} : \ \min\{ \psi^0(x) \mid \psi^j(x) \leq 0, \, j \in \mathbf{l} \, , \, x \in \mathbb{R}^n \ \} \ , \tag{3.1a}$$

where $\mathbf{l} \triangleq \{ 1,2,\dots,l \}$, and, with $\mathbf{L} \triangleq \{ 0,1,2,\dots,l \}$,

[6] It should be obvious that, under analogous assumptions, the conclusions of Theorem 2.4 remain valid for the implementable PPP minimax algorithm which can be used to solve **MMP**$_q$.

$$\psi^j(x) = \max_{y_j \in Y_j} \phi^j(x, y_j), \quad \forall\, j \in \mathbf{L}, \tag{3.1b}$$

where $\phi^j : \mathbb{R}^n \times \mathbb{R}^{p_j} \to \mathbb{R}$ and $\mathbf{Y}_j$ is a compact set in $\mathbb{R}^{p_j}$. We will assume that the functions $\phi^j(\cdot,\cdot)$ and their gradients $\nabla_x \phi^j(\cdot,\cdot)$ are Lipschitz continuous.

Using the interpolation techniques mentioned in the preceding section, we can construct a family of approximating problems, parametrized by the discretization parameter $q \in \mathbb{N}$:

$$\mathbf{CSP}_q : \ \min\{ \psi_q^0(x) \mid \psi_q^j(x) \le 0,\ j \in \mathbf{l}\ ,\ x \in \mathbb{R}^n \ \}\ , \tag{3.2a}$$

where

$$\psi_q^j(x) = \max_{y_j \in Y_j} \phi_q^j(x, y_j), \quad \forall\, j \in \mathbf{L}\ . \tag{3.2b}$$

In [Pol.4] we find a unified steerable phase I - phase II method of feasible directions which uses a *steering parameter* $\gamma > 0$. We will refer to this algorithm as the USFD algorithm. The steering parameter controls the speed with which infeasible iterates approach the feasible set. When this parameter is greater than a certain value, the algorithm in [Pol.4] constructs a feasible point in a finite number of iterations. We will call this algorithm as a subroutine from the master adaptive discretization algorithm that we will describe shortly. For the original problem CSP, the algorithm in [Pol.4] requires the following functions:

$$\psi(x) \triangleq \max_{j \in \mathbf{l}} \psi^j(x)\ , \tag{3.3a}$$

$$\psi_+(x) \triangleq \max\{\ 0, \psi(x)\ \}\ , \tag{3.3b}$$

$$F_z(x) \triangleq \max\{\ \psi^0(x) - \psi^0(z) - \gamma\psi_+(z), \psi(x) - \psi_+(z)\ \}, \tag{3.3c}$$

$$\theta(x) \triangleq \min_{h \in \mathbb{R}^n} \ \max\{ \max_{y_0 \in Y_0} \{\ \phi^0(x, y_0) + \langle \nabla_x \phi^0(x, y_0), h \rangle + \tfrac{1}{2}\|h\|^2 - \psi^0(x) - \gamma\psi_+(x)\ \},$$
$$\max_{j \in \mathbf{l}} \ \max_{y_j \in Y_j} \{\ \phi^j(x, y_j) + \langle \nabla_x \phi^j(x, y_j), h \rangle + \tfrac{1}{2}\|h\|^2 - \psi_+(x)\ \}\ \}, \tag{3.3d}$$

$$h(x) \triangleq \arg\min_{h \in \mathbb{R}^n} \ \max\{ \max_{y_0 \in Y_0} \{\ \phi^0(x, y_0) + \langle \nabla_x \phi^0(x, y_0), h \rangle + \tfrac{1}{2}\|h\|^2 - \psi^0(x) - \gamma\psi_+(x)\ \}$$
$$\max_{j \in \mathbf{l}} \ \max_{y_j \in Y_j} \{\ \phi^j(x, y_j) + \langle \nabla_x \phi^j(x, y_j), h \rangle + \tfrac{1}{2}\|h\|^2 - \psi_+(x)\ \}\ \}\ . \tag{3.3e}$$

For the problems $\mathbf{CSP}_q$, the algorithm in [Pol.4] uses corresponding quantities, $\psi_q(\cdot)$, $\psi_{q+}(\cdot)$, $F_{z,q}(x)$,

$\theta_q(x)$ and $h_q(x)$, resulting from the replacement of $\phi^j(\cdot,\cdot)$, $\nabla_x\phi^j(\cdot,\cdot)$, $\psi^j(\cdot)$, $\psi(\cdot)$ and $\psi_+(\cdot)$ by $\phi_q^j(\cdot,\cdot)$, $\nabla_x\phi_q^j(\cdot,\cdot)$, $\psi_q^j(\cdot)$, $\psi_q(\cdot)$ and $\psi_{q+}(x)$ in (3.3a-e), respectively.

In addition to the steering parameter γ, the USFD algorithm in [Pol.4] uses two Armijo step size parameters: $\alpha,\ \beta \in (0,1)$. In solving CSP, this algorithm constructs iterates according to the rule

$$x_{i+1} = x_i + \lambda_i h(x_i), \tag{3.3f}$$

where

$$\lambda_i = \max\{\ \beta^k \mid k \in \mathbb{N}\ ,\ F_{x_i}(x_i + \beta^k h(x_i)) - F_{x_i}(x_i) \le \alpha\beta^k\theta(x_i)\ \}\ . \tag{3.3g}$$

For solving $\mathbf{CSP}_q$ one makes the obvious substitutions in the above rule.

We will assume that the following relationship between the functions defining the problems $\mathbf{CSP}_q$ and problem **CSP**:

Assumption 3.1 : (i) There exist constant $0 < K < \infty$ and $\tau > 0$ such that for all $x \in \mathbb{R}^n$ and all $q \in \mathbb{N}$,

$$|\ \psi^0(x) - \psi_q^0(x)\ | \le K/q^\tau\ . \tag{3.4a}$$

$$|\ \psi(x) - \psi_q(x)\ | \le K/q^\tau\ . \tag{3.4b}$$

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

$$\lim_{\substack{x' \to x \\ q \to \infty}} \phi_q^j(x',y_j) = \phi^j(x,y_j)\ , \quad \lim_{\substack{x' \to x \\ q \to \infty}} \nabla_x\phi_q^j(x',y_j) = \nabla_x\phi^j(x,y_j)\ . \tag{3.4c}$$

■

Lemma 3.1 : (i) For any $x \in \mathbb{R}^n$,

$$\theta(x) = -\ \min\{\ \xi^0 + \tfrac{1}{2}\|\xi\|^2 \mid (\xi^0,\xi)^T \in G(x)\ \}\ , \tag{3.5a}$$

$$\theta_q(x) = -\ \min\{\ \xi^0 + \tfrac{1}{2}\|\xi\|^2 \mid (\xi^0,\xi)^T \in G_q(x)\ \}\ , \tag{3.5b}$$

where

$$G(x) \triangleq \text{co}\left\{ \bigcup_{y_0 \in \mathbf{Y}_0} \begin{bmatrix} \psi^0(x) - \phi^0(x,y_0) + \gamma\psi_+(x) \\ \nabla_x\phi^0(x,y_0) \end{bmatrix}, \bigcup_{j \in \mathbf{I}} \bigcup_{y_j \in \mathbf{Y}_j} \begin{bmatrix} \psi(x) - \phi^j(x,y_j) \\ \nabla_x\phi^j(x,y_j) \end{bmatrix} \right\}, \tag{3.5c}$$

$$G_q(x) \triangleq \text{co}\left\{ \bigcup_{y_0 \in \mathbf{Y}_0} \begin{bmatrix} \psi_q^0(x) - \phi_q^0(x,y_0) + \gamma\psi_{q+}(x) \\ \nabla_x\phi^0(x,y_0) \end{bmatrix}, \bigcup_{j \in \mathbf{I}} \bigcup_{y_j \in \mathbf{Y}_j} \begin{bmatrix} \psi_q(x) - \phi_q^j(x,y_j) \\ \nabla_x\phi_q^j(x,y_j) \end{bmatrix} \right\}. \tag{3.5d}$$

(ii) For any $x \in \mathbb{R}^n$, $\theta(x) = 0$ if and only if either $\psi(x) \leq 0$ and $0 \in \partial F_x(x)$ (i.e., x satisfies the first order optimality condition for problem (3.1a)), or $\psi(x) > 0$ and $0 \in \partial\psi(x)$, (i.e., x satisfies the first order optimality condition for the problem $\min_{x \in \mathbb{R}^n} \psi(x)$).

(iii) Suppose that Assumption 3.1 holds. Then, for any $x \in \mathbb{R}^n$,

$$\lim_{\substack{x' \to x \\ q \to \infty}} \theta_q(x') = \theta(x) . \tag{3.6}$$

Proof : (i) The relations (3.5a), (3.5b) were established in [Pol.4] using the von Neuman's minimax Theorem.

(ii) This part can be deduced from Propositions 5.4 and 5.5 in [Pol.2].

(iii) Since $\mathbf{Y}_j$ are compact sets, it follows from Assumption 3.1(ii) and (3.5c-d) that

$$\lim_{\substack{x' \to x \\ q \to \infty}} G_q(x') = G(x) . \tag{3.7}$$

Hence, (3.6) follows from (3.7) and (3.5a-b). ■

We are now ready to state an adaptive discretization scheme, based on the USFD algorithm in [Pol.4], for solving the problem CSP.

Adaptive Discretization Algorithm 3.1 (for CSP):

Data: $x_0 \in \mathbb{R}^n$, $q_{-1} \in \mathbb{N}$, $\alpha \in (0,1)$, $\beta \in (0,1)$, $\gamma > 0$, $D > 0$ and $\sigma > 1$.

Step 0: Set $i = 0$.

Step 1: Compute $q_i \in \mathbb{N}$, $\theta_i = \theta_{q_i}(x_i)$, and $h_i = h_{q_i}(x_i)$ such that $q_i \geq q_{i-1}$ and

$$D/q_i^\tau \leq [-\theta_{q_i}(x_i)]^\sigma \tag{3.8}$$

Step 2: Compute the step size λ_i :

$$\lambda_i = \max\{ \beta^k \mid k \in \mathbb{N} \ , \ F_{x_i, q_i}(x_i + \beta^k h_i) - F_{x_i, q_i}(x_i) \ \leq \ \alpha \beta^k \theta_i \ \} \ . \tag{3.9}$$

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 : It follows from Lemma 3.1(ii-iii) that whenever $\theta(x_i) \neq 0$, Step 1 of Algorithm 3.1 yields a finite discretization parameter q_i. For simplicity, in the rest of this section, we assume that Algorithm 3.1 does not produce an iterate x_i such that $\theta(x_i) = 0$, so that the resulting value of q_i, in Step 1, is finite. ■

Lemma 3.2 : Suppose that $\psi^0(\cdot)$ is bounded from below and that the sequence of iterates $\{ x_i \}_{i=0}^{\infty}$ and the corresponding sequence of discretization parameters $\{ q_i \}_{i=0}^{\infty}$ were constructed by Algorithm 3.1. Then $q_i \to \infty$ as $i \to \infty$.

Proof: Suppose that $q_i \to \infty$ as $i \to \infty$ does not hold. Then, since $\{ q_i \}_{i=0}^{\infty}$ is a nondecreasing sequence of integers, it follows that there exists an i_0, $\hat{q} \in \mathbb{N}$, such that for all $i \geq i_0$, $q_i = \hat{q}$, and hence, in view of (3.8) that there exists an $\varepsilon > 0$ such that $\theta_i \leq -\varepsilon$ for all $i \geq i_0$. It now follows from the properties of the algorithm defined by (3.3f), (3.3g) (see [Pol.4]), that there are two possibilities: either $\psi_{\hat{q}}(x_i) > 0$ for all $i \geq i_0$, or there exists an $i_1 \geq i_0$ such that $\psi_{\hat{q}}(x_i) \leq 0$ for all $i \geq i_1$. In the former case, $\psi_{\hat{q}}(x_{i+1}) - \psi_{\hat{q}}(x_i) \leq \alpha\lambda_i\theta_i$ for all $i \geq i_0$. In the latter case, $\psi^0_{\hat{q}}(x_{i+1}) - \psi^0_{\hat{q}}(x_i) \leq \alpha\lambda_i\theta_i$. Making use of the (3.4a) and the assumption that $\psi^0(\cdot)$ is bounded from below, we conclude that $\psi^0_{\hat{q}}(\cdot)$ is also bounded from below. Hence we obtain that either $-\infty < \sum_{i=i_0}^{\infty} [\psi_{\hat{q}}(x_{i+1}) - \psi_{\hat{q}}(x_i)] \leq \sum_{i=i_0}^{\infty} \alpha\lambda_i\theta_i$, or that $-\infty < \sum_{i=i_0}^{\infty} [\psi^0_{\hat{q}}(x_{i+1}) - \psi^0_{\hat{q}}(x_i)] \leq \sum_{i=i_0}^{\infty} \alpha\lambda_i\theta_i$. In either event, we are led to the conclusion that $\sum_{i=i_0}^{\infty} \alpha\lambda_i\theta_i > -\infty$. Since $\|h_i\|^2 \leq 2(-\theta_i)$, we conclude, applying the reasoning used in proving Lemma 2.2, that $\{ x_i \}_{i=0}^{\infty}$ is a Cauchy sequence in $\mathbb{R}^n$. Since by Lemma 3.2 in [Pol.4] (for algorithm (3.3f), (3.3g)), any accumulation point x^* of $\{ x_i \}_{i=0}^{\infty}$ satisfies $\theta_{\hat{q}}(x^*) = 0$, it follows that $\theta_{\hat{q}}(x_i) \to 0$, which contradicts our assumption. Therefore $\overline{\lim}_{i \to \infty} \theta_i = 0$. It now follows from (3.8) and the fact that $q_i \geq q_{i-1}$ that $q_i \to \infty$ as $i \to \infty$. ■

Theorem 3.1 : Suppose that Assumption 3.1 holds, that $\psi^0(\cdot)$ is bounded from below, and that there exists a constant $M \in [1, \infty)$ such that for all $x \in \mathbb{R}^n$, $z \in \mathbb{R}^n$, $j \in \mathbf{L}$ and $q \in \mathbb{N}$,

$$\langle z, \frac{\partial^2 \phi_q^j(x,y_j)}{\partial x^2} z\rangle \leq M\|z\|^2 \ , \quad \text{for } y_j \in \mathbf{Y}_j \ . \tag{3.10}$$

Then any accumulation point $\hat{x}$ of the sequence of iterates $\{ x_i \}_{i=0}^{\infty}$, generated by Algorithm 3.1, satisfies $\theta(\hat{x}) = 0$.

Proof: It follows from the definitions of $F_{z,q}(\cdot)$ and $F_z(\cdot)$, and Assumption 3.1(i) that

$$|F_{z,q}(x) - F_z(x)| \leq (2+\gamma)K/q^{\tau} \ , \quad \text{for all } x \in \mathbb{R}^n \ , \ z \in \mathbb{R}^n \ , \ q \in \mathbb{N} \ . \tag{3.11}$$

Next, we observe the i-th iteration of Algorithm 3.1 consists of one iteration of the PPP minimax algorithm on the problem $\min_{x \in \mathbb{R}^n} F_{x_i,q_i}(x)$ starting with x_i. Hence, replacing $\psi_{q_i}(\cdot)$, $\psi(\cdot)$ by $F_{x_i,q_i}(\cdot)$, $F_{x_i}(\cdot)$, we conclude from the proof of Theorem 2.1 that there exists an i_0 such that for all $i \geq i_0$,

$$F_{x_i}(x_{i+1}) - F_{x_i}(x_i) \leq \alpha\beta\theta_i/M + 2(2+\gamma)K/q_i^{\tau} \leq -\frac{\alpha\beta D^{1/\sigma}}{2M(q_i^{\tau})^{1/\sigma}} \ . \tag{3.12}$$

It now follows from the definition of $F_{x_i}(\cdot)$, (3.3c), that for all $i \geq i_0$,

$$\psi^0(x_{i+1}) - \psi^0(x_i) - \gamma\psi_+(x_i) \leq -\alpha\beta\theta_i/M + (4+2\gamma)K/q_i^{\tau} \leq -\frac{\alpha\beta D^{1/\sigma}}{2M(q_i^{\tau})^{1/\sigma}} < 0 \ , \tag{3.13a}$$

$$\psi(x_{i+1}) - \psi_+(x_i) \leq \alpha\beta\theta_i/M + (4+2\gamma)K/q_i^{\tau} \leq -\frac{\alpha\beta D^{1/\sigma}}{2M(q_i^{\tau})^{1/\sigma}} < 0 \ . \tag{3.13b}$$

We must consider two cases.

Case (i): There exists an integer $i_1 > i_0$ such that $\psi(x_{i_1}) \leq 0$. It then follows from (3.13b) that $\psi(x_i) \leq 0$ for all $i \geq i_1$ and from (3.13a) that

$$\psi^0(x_{i+1}) - \psi^0(x_i) \leq -\alpha\beta\theta_i/M + (4+2\gamma)K/q_i^{\tau} \leq -\frac{\alpha\beta D^{1/\sigma}}{2M(q_i^{\tau})^{1/\sigma}} \ , \ \text{for all } i \geq i_1 \ . \tag{3.14}$$

Hence, by the same reasoning used in the proof of Theorem 2.1, we conclude that $\sum_{k=i_1}^{\infty} \theta_k > -\infty$, which leads to the desired result.

Case (ii): $\psi(x_i) > 0$ for all $i > i_0$. It then follows from (3.13b) that for all $i > i_0$,

$$\psi(x_{i+1}) - \psi(x_i) \leq -\alpha\beta\theta_i/M + (4+2\gamma)K/q_i^{\tau} \leq -\frac{\alpha\beta D^{1/\sigma}}{2M(q_i^{\tau})^{1/\sigma}} \ . \tag{3.15}$$

Since $\psi(x_i) > 0$ of all $i \geq i_0$, the reasoning used in the proof of Theorem 2.1 leads to the conclusion that $\sum_{k=i_0}^{\infty} \theta_k > -\infty$. Hence the desired result follows from Lemma 3.1(iii). ■

To establish the linear convergence of Algorithm 3.1, we need following assumption and results which we borrow from [Pol.4].

Assumption 3.2 : (i) There exist $0 < m \leq 1 \leq M < \infty$ such that for all $x \in \mathbb{R}^n$, $z \in \mathbb{R}^n$ and $j \in \mathbf{L}$,

$$m|z|^2 \leq \langle z, \frac{\partial^2 \phi_q^j(x,y_j)}{\partial x^2} z \rangle \leq M|z|^2 , \quad \text{for all } y_j \in \mathbf{Y}_j \text{ , } q \in \mathbb{N} \text{ ,} \tag{3.16a}$$

$$m|z|^2 \leq \langle z, \frac{\partial^2 \phi^j(x,y_j)}{\partial x^2} z \rangle \leq M|z|^2 , \quad \text{for all } y_j \in \mathbf{Y}_j . \tag{3.16b}$$

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

Lemma 3.3 : (Lemma 4.2, Lemma 4.3 in [Pol.4]) : Suppose that Assumption 3.2 holds. Then

(i) Problem (3.1a) has a unique solution, $\hat{x}$.

(ii) $\hat{x}$ is the unique zero of $\theta(\cdot)$.

(iii) Let $\underline{\mu}^0 \triangleq \min\{ \mu^0 \mid \mu \in L(\hat{x}) \}$ and $\overline{\mu}^0 \triangleq \max\{ \mu^0 \mid \mu \in L(\hat{x}) \}$, where

$$L(\hat{x}) \triangleq \{ \mu = (\mu^0, \mu^1, ...\mu^l) \mid 0 \in \sum_{j=0}^{l} \mu^j \partial\psi^j(\hat{x}), \ \sum_{j=1}^{l} \mu^j \psi^j(\hat{x}) = 0, \ \sum_{j=0}^{l} \mu^j = 1, \ \mu^j \geq 0 \} . \tag{3.17a}$$

Then $0 < \underline{\mu}^0 \leq \overline{\mu}^0 \leq 1$.

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

$$\overline{\mu}^0[\psi^0(\hat{x}) - \psi^0(x)] \leq (1 - \overline{\mu}^0)\psi_+(x) , \tag{3.17b}$$

■

Theorem 3.2 : Suppose that Assumptions 3.1 and 3.2 hold. Then any sequence of iterates $\{ x_i \}_{i=0}^{\infty}$, generated by Algorithm 3.1, converges to $\hat{x}$, and there exists an integer i_0 such that either

(i) $\psi(x_i) \leq 0$ for all $i \geq i_0$ and

$$\overline{\lim_{i \to \infty}} \frac{\psi^0(x_{i+1}) - \psi^0(\hat{x})}{\psi^0(x_i) - \psi^0(\hat{x})} \leq 1 - \underline{\mu}^0 \alpha\beta m/M . \tag{3.18a}$$

or (ii) $\psi(x_i) > 0$ for all $i \geq i_0$ and

$$\overline{\lim_{i \to \infty}} \frac{\psi(x_{i+1})}{\psi(x_i)} \leq 1 - \gamma \underline{\mu}^0 \alpha \beta m / M \ , \tag{3.18b}$$

where $\underline{\mu}^0$ is defined in Lemma 3.3 (iii).

Proof: Let $\hat{x}_{x,q}$, $\hat{x}_x$ be the unique solutions of $\min_{x' \in \mathbb{R}^n} F_{x,q}(x')$ and $\min_{x' \in \mathbb{R}^n} F_x(x')$, respectively. First we show that both $\{ x_i \}_{i=0}^{\infty}$ and $\{ \hat{x}_{x_i} \}_{i=0}^{\infty}$ converge to $\hat{x}$. It follows from (3.13b), that $\psi_+(x_{i+1}) \leq \psi_+(x_i)$ when i is sufficiently large. Since $\psi(\cdot)$ has bounded level sets, the sequence $\{ x_i \}_{i=0}^{\infty}$, constructed by Algorithm 3.1, is bounded. It therefore follows from Lemma 3.3 (i-ii) and Theorem 3.1 that $x_i \to \hat{x}$ as $i \to \infty$. Hence, making use of the fact that $F_{x_i}(\hat{x}_{x_i}) \leq F_{x_i}(x_i) = 0$, we deduce that $\{ \hat{x}_{x_i} \}_{i=0}^{\infty}$ is bounded, and that, because $F_x(x')$ is continuous in (x,x'), any accumulation point $\tilde{x}$ of $\{ \hat{x}_{x_i} \}_{i=0}^{\infty}$, must satisfy $F_{\hat{x}}(\tilde{x}) \leq 0$, which implies that $\tilde{x} = \hat{x}$. Therefore, $\hat{x}_{x_i} \to \hat{x}$ as $i \to \infty$.

Now the i-th iteration of Algorithm 3.1, consists of one iteration of the PPP minimax algorithm on $\min_{x \in \mathbb{R}^n} F_{x_i,q_i}(x)$, starting with x_i. Therefore, replacing $\psi_{q_i}(\cdot)$, $\psi(\cdot)$, $\hat{x}_{q_i}$ and $\hat{x}$ by $F_{x_i,q_i}(\cdot)$, $F_{x_i}(\cdot)$, $\hat{x}_{x_i,q_i}$ and $\hat{x}_{x_i}$, respectively, in the proof for Theorem 2.1, we conclude that there exists an integer i_1 such that for all $i \geq i_1$,

$$F_{x_i}(x_{i+1}) - F_{x_i}(\hat{x}_{x_i}) \ \leq \ [1 - \nu_i][F_{x_i}(x_i) - F_{x_i}(\hat{x}_{x_i})] \ , \tag{3.19a}$$

where

$$\nu_i \triangleq \frac{\alpha \beta m}{M} - \frac{8(2+\gamma)KM}{D^{1/\sigma}(q_i^{\tau})^{(\sigma-1)/\sigma}} > 0 \ . \tag{3.19b}$$

Since $F_{x_i}(x_i) = 0$, we get

$$F_{x_i}(x_{i+1}) \leq \nu_i F_{x_i}(\hat{x}_{x_i}) \ . \tag{3.20}$$

Now, since $\hat{x}_i$ is the minimizer of $F_{x_i}(\cdot)$, there exists a multiplier $\hat{\mu}_i = (\hat{\mu}_i^0, \hat{\mu}_i^1, \ldots, \hat{\mu}_i^l)$ such that

$$0 \in \sum_{j=0}^{l} \hat{\mu}_i^j \partial \psi^j(\hat{x}_{x_i}) \,, \tag{3.21a}$$

$$\hat{\mu}_i^0[\psi^0(\hat{x}_{x_i}) - \psi^0(x_i) - \gamma\psi_+(x_i)] + \sum_{j=1}^{l} \hat{\mu}_i^j[\psi^j(\hat{x}_{x_i}) - \psi_+(x_i)] = F_{x_i}(\hat{x}_{x_i}) \,, \tag{3.21b}$$

$$\sum_{j=0}^{l} \hat{\mu}_i^j = 1 \,, \quad \hat{\mu}_i^j \geq 0 \,, \text{ for } j \in \mathbf{L} \,. \tag{3.21c}$$

Since both sequences $\{ x_i \}_{i=0}^{\infty}$ and $\{ \hat{x}_{x_i} \}_{i=0}^{\infty}$ converge to $\hat{x}$, any accumulation point of $\{ \hat{\mu}_i \}_{i=0}^{\infty}$ is in $L(\hat{x})$. Hence

$$\underline{\mu}^0 \leq \underline{\lim}_{i \to \infty} \hat{\mu}_i^0 \leq \overline{\lim}_{i \to \infty} \hat{\mu}_i^0 \leq \bar{\mu}^0 \,. \tag{3.22}$$

Let $\tilde{F}_i(\cdot)$ be defined by

$$\tilde{F}_i(x) = \hat{\mu}_i^0[(\psi^0(x) - \psi^0(x_i) - \gamma\psi_+(x_i)] + \sum_{j=1}^{l} \hat{\mu}_i^j[\psi^j(x) - \psi_+(x_i)] \,. \tag{3.23}$$

It follows from (3.21a)-(3.21c) that $0 \in \partial\tilde{F}_i(\hat{x}_{x_i})$. Since $\tilde{F}_i(\cdot)$ is strictly convex, $\hat{x}_{x_i}$ must be its unique minimizer. Hence, making use of (3.21c) and the fact that $\tilde{F}_i(\hat{x}_{x_i}) = F_{x_i}(\hat{x}_{x_i})$ and that $\psi(\hat{x}) \leq 0$, we obtain that

$$F_{x_i}(\hat{x}_{x_i}) \leq \tilde{F}_i(\hat{x}) \leq \hat{\mu}_i^0[(\psi^0(\hat{x}) - \psi^0(x_i)] - [1 + (\gamma - 1)\hat{\mu}_i^0]\psi_+(x_i) \,. \tag{3.24}$$

Combining (3.20) and (3.24), and rearranging terms, we get

$$\psi^0(x_{i+1}) - \psi^0(\hat{x}) \leq (1 - \hat{\mu}_i^0 \nu_i)[\psi^0(x_i) - \psi^0(\hat{x})] + [\gamma - (1+(\gamma-1)\hat{\mu}_i^0)\nu_i]\psi_+(x_i) \,, \tag{3.25a}$$

$$\psi(x_{i+1}) \leq \hat{\mu}_i^0 \nu_i[\psi^0(\hat{x}) - \psi^0(x_i)] + [1 - (1 + (\gamma - 1)\hat{\mu}_i^0)\nu_i]\psi_+(x_i) \,. \tag{3.25b}$$

By Lemma 3.3, $q_i \to \infty$ as $i \to \infty$. Hence, we deduce from (3.19b) that

$$\lim_{i \to \infty} \nu_i = \frac{\alpha\beta m}{M} \,. \tag{3.26}$$

Now, it was shown in the proof of Theorem 3.1 that there exists an integer $i_0 > i_1$ such that either (i) $\psi(x_i) \leq 0$ for all $i \geq i_0$ or (ii) $\psi(x_i) > 0$ for all $i \geq i_0$. In the former case, (3.18a) follows from (3.22), (3.25a), (3.26) and the fact that $\psi_+(x_i) = 0$ for $i \geq i_0$. In the latter case, we obtain from (3.25b) and (3.17b) that

$$\psi(x_{i+1}) \le [\hat{\mu}_i^0 \nu_i (1 - \bar{\mu}^0)/\bar{\mu}^0 + 1 - (1 + (\gamma - 1)\hat{\mu}_i^0)\nu_i]\psi_+(x_i)$$

$$= [1 + (\hat{\mu}_i^0/\bar{\mu}^0 - 1 - \gamma\hat{\mu}_i^0)\nu_i]\psi(x_i) \,. \tag{3.27}$$

Hence, (3.18b) follows from (3.22), (3.26) and (3.27). ■

For comparison, we reproduce the rate of convergence theorem for the USFD algorithm described in [Pol.4]:

Theorem 3.3 : Suppose that the relevant part of Assumption 3.2 holds. Then any sequence of iterates $\{ x_i \}_{i=0}^{\infty}$, generated by the USFD algorithm, converges to $\hat{x}$, and there exists an integer i_0 such that either

(i) $\psi(x_i) \le 0$ for all $i \ge i_0$ and

$$\overline{\lim_{i\to\infty}} \frac{\psi^0(x_{i+1}) - \psi^0(\hat{x})}{\psi^0(x_i) - \psi^0(\hat{x})} \le 1 - \mu^0\alpha\beta m/M \,. \tag{3.28a}$$

or (ii) $\psi(x_i) > 0$ for all $i \ge i_0$ and

$$\overline{\lim_{i\to\infty}} \frac{\psi(x_{i+1})}{\psi(x_i)} \le 1 - \gamma\mu^0\alpha\beta m/M \,, \tag{3.28b}$$

where μ^0 is defined in Lemma 3.3 (iii). ■

Thus we see that Algorithm 3.1 has the same rate of convergence as the USFD algorithm and hence, by the same arguments used at the end of Section 2, we conclude that using adaptive discretization in the form of Algorithm 3.1 should result in savings in computing time over the use of the USFD algorithm on a single high precision approximation to the original problem.

4. OPTIMAL CONTROL PROBLEMS

Finally we turn to unconstrained optimal control problems of the form

$$\text{OCP} : \min_{u \in G(\infty)} c(u) = g(\bar{x}(u)), \tag{4.1a}$$

where $g: \mathbb{R}^n \to \mathbb{R}$, $G(\omega) \triangleq L_2^m[0,T] \cap \{ u \mid \max_{t \in [0,T]} |u(t)| < \omega \}$, T is a given time period, and $\bar{x}: G(\infty) \to \mathbb{R}^n$ is defined by $\bar{x}(u) \triangleq x(T,u,x_0)$, with $x(\cdot,u,x_0)$ the solution of the differential equation:

$$\dot{x}(t) = f(x(t), u(t), t), \quad t \in [0,T], \quad x(0) = x_0, \tag{4.1b}$$

where x_0 is a given vector in $\mathbb{R}^n$. In this section, the $L_2^m[0,T]$ topology will be used on $G(\infty)$ and its subspaces, unless we stated otherwise. We will assume that functions $g: \mathbb{R}^n \to \mathbb{R}$ and $f: \mathbb{R}^n \times \mathbb{R}^m \times [0,T] \to \mathbb{R}^n$ have the following properties:

Assumption 4.1 : (i) The function $g(\cdot)$ and its gradient $\nabla g(\cdot)$ are locally Lipschitz continuous.

(ii) The function $f(\cdot,\cdot,\cdot)$ and its partial derivatives $\frac{\partial f(\cdot,\cdot,\cdot)}{\partial x}$ and $\frac{\partial f(\cdot,\cdot,\cdot)}{\partial u}$ are locally Lipschitz continuous.

(iii) for every $\omega > 0$, there exists a constant $K_1(\omega)$ such that for all $x \in \mathbb{R}^n$, $u \in \{ u' \in \mathbb{R}^m \mid \|u'\| \le \omega \}$ and $t \in [0,T]$,

$$\|f(x,u,t)\| \le K_1(\omega)[\|x\| + 1]. \tag{4.2}$$

■

Lemma 4.1 : ([Kle.1]) Suppose that Assumption 4.1 is satisfied. Then

(i) The differential equation (4.1b) has a unique solution for every $u \in G(\infty)$.

(ii) The functions $\bar{x}(\cdot)$ and $c(\cdot)$ are continuously Frechet differentiable on $G(\infty)$. ■

Since the functions $\bar{x}(\cdot)$ and $c(\cdot)$ are continuously Frechet differentiable on $G(\infty)$, there exist continuous functions, a "jacobian" $\frac{\partial \bar{x}}{\partial u} : G(\infty) \to G(\infty)^n$ and a "gradient" $\nabla c : G(\infty) \to G(\infty)$, such that

$$\lim_{\substack{\delta u \in G(\infty) \\ \|\delta u\|_2 \to 0}} \frac{\| \bar{x}(u + \delta u) - \bar{x}(u) - \int_0^T \frac{\partial \bar{x}(u)}{\partial u}(t) \delta u(t) dt \|}{\|\delta u\|_2} = 0, \tag{4.3a}$$

$$\lim_{\substack{\delta u \in G(\infty) \\ \|\delta u\|_2 \to 0}} \frac{| c(u + \delta u) - c(u) - \int_0^T \langle \nabla c(u)(t), \delta u(t) \rangle dt |}{\|\delta u\|_2} = 0. \tag{4.3b}$$

Lemma 4.2 : ([Kle.1]) Suppose that Assumption 4.1 is satisfied. Then

(i) for every $u \in G(\infty)$ and $t \in [0,T]$,

$$\frac{\partial \bar{x}(u)}{\partial u}(t) = \Phi_u(T,t) \frac{\partial f(x(u,t), u(t), t)}{\partial u}, \tag{4.4a}$$

$$\nabla c(u)(t) = \left[\frac{\partial \bar{x}(u)}{\partial u}(t)\right]^T \nabla g(\bar{x}(u)) \ , \tag{4.4b}$$

where $\Phi_u(t,\tau)$ is the state transition matrix for the linear differential equation

$$\dot{y}(t) = \frac{\partial f(x(t,u),u(t),t)}{\partial x} y(t) \quad \text{on } t \in [0,T] \ . \tag{4.4c}$$

(ii) for any $\omega > 0$, there exists a $K_2(\omega)$ such that for all $u, \delta u \in G(\omega)$,

$$\|\bar{x}(u)\| \leq K_2(\omega) \ , \quad \|\frac{\partial \bar{x}(u)}{\partial u}\|_\infty \leq K_2(\omega) \ , \quad \|\nabla c(u)\|_\infty \leq K_2(\omega) \ , \tag{4.4d}$$

$$\|\bar{x}(u + \delta u) - \bar{x}(u)\| \leq K_2(\omega)\|\delta u\|_2 \ , \tag{4.4e}$$

$$\|\bar{x}(u + \delta u) - \bar{x}(u) - \int_0^T [\frac{\partial \bar{x}(u)}{\partial u}(t)]\delta u(t)dt\| \leq K_2(\omega)\|\delta u\|_2^2 \ . \tag{4.4f}$$

■

As a first step towards the numerical solution of the infinite dimensional problem OCP, we define a sequence of finite dimensional subspaces $G_q(\infty)$ of $G(\infty)$, $q \in \mathbb{N}$. Thus, for any $q \in \mathbb{N}$, let $\Delta_q \triangleq T/2^q$ and, for any $\omega \in (0,\infty)$ let $G_q(\omega) \triangleq G(\omega) \cap \{ u \mid u(t) = u^i \in \mathbb{R}^m$ for $t \in [j\Delta_q,(j+1)\Delta_q)$, $j = 0,1,\ldots,2^q - 1 \}$. Next, for any $q \in \mathbb{N}$, and any $u \in G_q(\omega)$, let $\bar{x}_q(u)$ be an approximation to $\bar{x}(u)$, obtained by solving the differential equation (4.1b) by means of a numerical method, such as the Euler-Cauchy method, the Modified Euler method, the Runge-Kutta method, etc. We can now define a family of finite dimensional approximating problems, parametrized by the discretization parameter $q \in \mathbb{N}$:

$$\mathbf{OCP}_q : \quad \min_{u \in G_q(\infty)} c_q(u) \ . \tag{4.5}$$

where $c_q(u \triangleq g(\bar{x}_q(u))$. We will assume that $\bar{x}_q(\cdot)$ approximates $\bar{x}(u)$ in the following sense.

Assumption 4.2 : (i) For every $\omega > 0$, there exist constants $K_3(\omega)$, $\tau \in (0,\infty)$ such that for any $q \in \mathbb{N}$,

$$| x(u) - \bar{x}_q(u)| \leq K_3(\omega)/(2^q)^\tau \ , \quad \text{for all } u \in G_q(\omega) \ . \tag{4.6a}$$

(ii) $\bar{x}_q(\cdot)$ is continuously Frechet differentiable on $G_q(\infty)$. We will denote its "jacobian" by $\dfrac{\partial \bar{x}_q(u)}{\partial u}(t)$.

(iii) For any ω, $\varepsilon \in (0,\infty)$, there exists a $\hat{q} \in \mathbb{N}$ such that for all $q \geq \hat{q}$ and $u \in G_q(\omega)$,

$$\| \frac{\partial \bar{x}_q(u)}{\partial u} - \frac{\partial \bar{x}(u)}{\partial u} \|_\infty \leq \varepsilon \, , \tag{4.6b}$$

(iv) For any $\omega > 0$, there exists $K_4(\omega) \in (0,\infty)$ such that for all $q \in \mathbb{N}$ and all u, $\delta u \in G_q(\omega)$,

$$\| \bar{x}_q(u + \delta u) - \bar{x}_q(u) \| \leq K_4(\omega) \| \delta u \|_2 \, , \tag{4.6c}$$

$$\| \bar{x}_q(u + \delta u) - \bar{x}_q(u) - \int_0^T [\frac{\partial \bar{x}_q(u)}{\partial u}(t)] \delta u(t) dt \| \leq K_4(\omega) \| \delta u \|_2^2 \, . \tag{4.6d}$$

■

Remark 4.1 : Referring to [Kle.1], we see that when the Euler-Cauchy method is used to define $\bar{x}_q(u)$, Assumption 4.2 is satisfied with $\tau = 1$. It is easy to show that when the Modified Euler method or Runge-Kutta method are used to define $\bar{x}_q(u)$, Assumption 4.2 is satisfied with $\tau = 3$ and $\tau = 5$, respectively. ■

Lemma 4.3 : Suppose that Assumptions 4.1 and 4.2 are satisfied. Then

(i) The function $c_q(\cdot)$ is continuously Frechet differentiable on $G_q(\infty)$, and for any $u \in G_q(\infty)$ and $t \in [0,T]$,

$$\nabla c_q(u)(t) = \left[\frac{\partial \bar{x}_q(u)}{\partial u}(t) \right]^T \nabla g(\bar{x}_q(u)) \, . \tag{4.7a}$$

(ii) For any $\omega > 0$, there exists a $K_5(\omega) \in (0,\infty)$ such that for $q \in \mathbb{N}$ and all $u \in G_q(\omega)$,

$$| c(u) - c_q(u) | \leq K_5(\omega)/(2^q)^\tau \, . \tag{4.7b}$$

Proof : Part (i) follows from the Assumption 4.2(ii), (4.6c) and the local Lipschitz continuity of $\nabla g(\cdot)$. The inequality (4.7b) follows from Assumption 4.2(i), (4.4d) and the local Lipschitz continuity of $g(\cdot)$.■

Since $G_q(\infty)$ is isomorphic to $\mathbb{R}^{2^q}$, each problem OCP_q, defined by (4.5), can be solved by the Armijo gradient method [Arm.1] which uses two parameters α, $\beta \in (0,1)$ and which, for OCP_q constructs iterates according to the rule:

$$u_{i+1} = u_i - \lambda_i \nabla c_q(u_i) \, , \tag{4.8a}$$

with

$$\lambda_i = \max\{ \beta^k \mid k \in \mathbb{N} \ , \ c_q(u_i - \beta^k \nabla c_q(u_i)) - c_q(u_i) \le -\alpha\beta^k \| \nabla c_q(u_i) \|_2^2 \} \ . \tag{4.8b}$$

We can now state an adaptive discretization scheme, based on the Armijo gradient method (4.8a), (4.8b), for solving the problem OCP.

Adaptive Discretization Algorithm 4.1 (for OCP)

Data: $u_0 \in G_{q_{-1}}(\infty)$, $q_{-1} \in \mathbb{N}$, $\alpha \in (0,1)$, $\beta \in (0,1)$, $D > 0$ and $\sigma > 1$.

Step 0: Set $i = 0$.

Step 1: Compute $q_i \in \mathbb{N}$, $h_i = -\nabla c_{q_i}(u_i)$, and $\theta_i = -\| \nabla c_{q_i}(u_i) \|_2^2$, such that $q_i \ge q_{i-1}$ and

$$D/(2^{q_i})^\tau \le [-\theta_i]^\sigma \ . \tag{4.9a}$$

Step 2: Compute the step size λ_i :

$$\lambda_i = \max\{ \beta^k \mid k \in \mathbb{N} \ , \ c_{q_i}(u_i + \beta^k h_i) - c_{q_i}(u_i) \le \alpha\beta^k\theta_i \} \ . \tag{4.9b}$$

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

Remark 4.2 : It follows from Assumption 4.2(iii), (4.4b) and (4.7a), that whenever $\nabla c(u_i) \ne 0$ (i.e., u_i does not satisfy a first order optimality condition for the problem OCP), Step 1 of Algorithm 4.1 yields a finite q_i. For simplicity, in the rest of this section, we will assume that Algorithm 4.1 does not construct a u_i such that $\nabla c(u_i) = 0$, for any finite i. ■

Lemma 4.4 : Suppose that Assumption 4.1 is satisfied, that $g(\cdot)$ is bounded from below, and that the sequence of controls $\{ u_i \}_{i=0}^\infty$ and the corresponding sequence of discretization parameters $\{ q_i \}_{i=0}^\infty$ are constructed by Algorithm 4.1. Then $q_i \to \infty$ as $i \to \infty$.

Proof : Suppose that $q_i \to \infty$ as $i \to \infty$ does not hold. Then, using the same reasoning as in the proof of Lemma 2.2 and the fact that $\| h_i \|_2^2 = -\theta_i$, we conclude that there exist i_0 and $\hat{q} \in \mathbb{N}$ such that $q_i = \hat{q}$ for all $i \ge i_0$, and that $\{ u_i \}_{i=0}^\infty$ is a Cauchy sequence in $L_2^m[0,T]$. Consequently, since $u_i \in G_{\hat{q}}(\infty)$, which is a finite dimensional space, $\{ u_i \}_{i=0}^\infty$ is bounded in the L_∞ norm, and hence it converges also in this norm. Since for any L_∞ accumulation point $\hat{u}$, of a sequence $\{ u_i \}_{i=0}^\infty$ constructed by the Armijo gradient method (4.8a), (4.8b) in the finite dimensional space, $G_{\hat{q}}(\infty)$, $\nabla c_{\hat{q}}(\hat{u}) = 0$ (see [Arm.1]),

$\theta_{\hat{q}}(u_i) = -\|\nabla c_{\hat{q}}(u_i)\|_2^2 \to 0$, which contradicts the test (4.9a). Thus we must have that $q_i \to \infty$ as $i \to \infty$. ∎

Theorem 4.1 : Suppose that Assumptions 4.1 and 4.2 are satisfied, that $g(\cdot)$ is twice continuously differentiable and bounded from below, and that there exists a constant $M_g \in (0,\infty)$ such that for all $x \in \mathbb{R}^n$, $z \in \mathbb{R}^n$,

$$\langle z, \frac{\partial^2 g(x)}{\partial x^2} z \rangle \le M_g \|z\|^2 . \tag{4.10}$$

Then any L_2 accumulation point $\hat{u} \in G(\infty)$ of an L_∞ bounded sequence of controls $\{ u_i \}_{i=0}^{\infty}$, generated by Algorithm 4.1, satisfies $\nabla c(\hat{u}) = 0$.

Proof : Since $\|u_i\|_\infty$ is bounded and, by Lemma 4.3, $q_i \to \infty$ as $i \to \infty$, it follows from Assumptions 4.1 and 4.2, and Lemmas 4.2 and 4.3 that there exists an $\omega > 0$ such that for all $i \ge 0$ and $\lambda \in [0,1]$,

$$\|u_i\|_\infty \le \omega , \quad \|h_i\|_\infty \le \omega , \tag{4.11a}$$

$$\|\bar{x}_{q_i}(u_i + \lambda h_i)\| \le \omega , \quad \left\| \frac{\partial \bar{x}_{q_i}(u_i + \lambda h_i)}{\partial u} \right\|_\infty \le \omega , \quad \|\bar{x}(u_i + \lambda h_i)\| \le \omega . \tag{4.11b}$$

Making use of (4.10), (4.7a), (4.6c-d) and the fact that $h_i = -\nabla c_{q_i}(u_i)$ and that $\theta_i = -\|h_i\|_2^2$, we obtain that for all $\lambda \in [0,1]$,

$$\begin{aligned} c_q(u_i + \lambda h_i) - c_q(u_i) &= g(\bar{x}_{q_i}(u_i + \lambda h_i)) - g(\bar{x}_{q_i}(u_i)) \\ &\le \langle \nabla g(\bar{x}_{q_i}(u_i)), (\bar{x}_{q_i}(u_i + \lambda h_i) - \bar{x}_{q_i}(u_i)) \rangle + \frac{M_g}{2} \|\bar{x}_{q_i}(u_i + \lambda h_i) - \bar{x}_{q_i}(u_i)\|^2 \\ &\le \langle \nabla g(\bar{x}_{q_i}(u_i)), \int_0^T \frac{\partial \bar{x}_{q_i}(u_i)}{\partial u}(t)\, \lambda h_i(t)\, dt \rangle + K_4(\omega) \| \nabla g(\bar{x}_{q_i}(u_i)) \| \| \lambda h_i \|_2^2 + \frac{M_g}{2} (K_4(\omega) \|\lambda h_i\|_2)^2 \\ &= \lambda\theta_i - \lambda^2 [K_4(\omega) \|\nabla g(\bar{x}_{q_i}(u_i))\| + M_g K_4^2(\omega)/2] \theta_i . \end{aligned} \tag{4.12}$$

Since $\bar{x}_{q_i}(u_i)$ is bounded and $\nabla g(\cdot)$ is continuous, there exists a $K_6(\omega) \in (1,\infty)$ such that

$$c_q(u_i + \lambda h_i) - c_q(u_i) \le (\lambda - \lambda^2 K_6(\omega))\theta_i = \alpha\lambda\theta_i + \lambda(1 - \alpha - K_6(\omega)\lambda)\theta_i , \ \forall \ \lambda \in [0,1] . \tag{4.13}$$

Hence (4.9b) is satisfied with $\lambda_i \ge (1 - \alpha)\beta / K_6(\omega)$, and thus

$$c_{q_i}(u_{i+1}) - c_{q_i}(u_i) \le \alpha\lambda_i\theta_i \le \alpha(1 - \alpha)\beta\theta_i / K_6(\omega) . \tag{4.14}$$

It therefore follows from (4.11a) and (4.7b) that

$$c(u_{i+1}) - c(u_i) \le \alpha(1-\alpha)\beta\theta_i/K_6(\omega) + 2K_5(\omega)/(2^{q_i})^{\tau}\,. \tag{4.15}$$

Resorting to the reasoning used in the proof of Theorem 2.1, with $\psi(\cdot)$, x_i, q_i, M and K replaced by $c(\cdot)$, u_i, 2^{q_i}, $K_6(\omega)/(1-\alpha)$ and $K_5(\omega)$, we can show that $\sum_{k=0}^{\infty}\theta_k > -\infty$. Hence $\theta_i \to 0$ as $i \to \infty$. Now, for all $i \in \mathbb{N}$,

$$\|\nabla c(\hat{u})\|_2 \le \|\nabla c(\hat{u}) - \nabla c(u_i)\|_2 + \|\nabla c(u_i) - \nabla c_{q_i}(u_i)\|_2 + \|\nabla c_{q_i}(u_i)\|_2$$

$$\le \|\nabla c(\hat{u}) - \nabla c(u_i)\|_2 + T^{1/2}\,\|\nabla c(u_i) - \nabla c_{q_i}(u_i)\|_\infty + (-\theta_i)^{1/2}\,. \tag{4.16}$$

Consequently, the desired result follows from the continuity of $\nabla c(\cdot)$, (4.7b) and the fact that $\theta_i \to 0$ and $q_i \to \infty$ as $i \to \infty$. ■

Lemma 4.5 : Suppose that Assumptions 4.1 and 4.2 hold and that there exist $0 < m_g < M_g < \infty$ and $0 < m_c < M_c < \infty$ such that for all $x \in \mathbb{R}^n$, $z \in \mathbb{R}^n$ and $u \in G(\infty)$,

$$m_g\|z\|^2 \le \langle z, \frac{\partial^2 g(x)}{\partial x^2} z \rangle \le M_g\|z\|^2, \tag{4.17a}$$

$$m_c\|z\|^2 \le \langle z\,, \left\{ \int_0^T \left[\frac{\partial \bar{x}(u)}{\partial u}(t)\right]\left[\frac{\partial \bar{x}(u)}{\partial u}(t)\right]^T dt \right\} z \rangle \le M_c\|z\|^2\,. \tag{4.17b}$$

Then for any $\omega > 0$ and $\varepsilon \in (0, m_c)$, there exists a $\hat{q}$ such that for any $q \ge \hat{q}$ and $u \in G_q(\omega)$,

$$(m_c - \varepsilon)\|z\|^2 \le \langle z\,, \left\{ \int_0^T \left[\frac{\partial \bar{x}_q(u)}{\partial u}(t)\right]\left[\frac{\partial \bar{x}_q(u)}{\partial u}(t)\right]^T dt \right\} z \rangle \le (M_c + \varepsilon)\|z\|^2\,, \tag{4.18a}$$

$$(m_c - \varepsilon)\|\nabla g(\bar{x}_q(u))\|^2 \le \|\nabla c_q(u)\|_2^2 \le (M_c + \varepsilon)\|\nabla g(\bar{x}_q(u))\|^2\,, \tag{4.18b}$$

$$\frac{1}{2M_g(M_c+\varepsilon)}\|\nabla c_q(u)\|_2^2 \le c_q(u) - g(\hat{x}) \le \frac{1}{2m_g(m_c-\varepsilon)}\|\nabla c_q(u)\|_2^2\,. \tag{4.18c}$$

where $\hat{x}$ is the unique minimizer of $g(\cdot)$.

Proof : Inequality (4.18a) follows directly from (4.4d), (4.17b) and Assumption 4.2 (iii), while (4.18b) follows from (4.7a) and (4.18a). Making use of (4.18b) and the fact that for all $x \in \mathbb{R}^n$,

$$\frac{1}{2M_g}\|\nabla g(x)\|^2 \le g(x) - g(\hat{x}) \le \frac{1}{2m_g}\|\nabla g(x)\|^2 , \qquad (4.19)$$

we get (4.18c). ■

Remark 4.3 : The matrix $\int_0^T \left[\frac{\partial \overline{x}(u)}{\partial u}(t)\right]\left[\frac{\partial \overline{x}(u)}{\partial u}(t)\right]^T dt$ is the controllability Gramian of the linearization of the system (4.1b). When the dynamical system (4.1b) is linear, the inequality (4.17b) holds if and only if (4.1b) is completely controllable on $[0,T]$. When the dynamical system (4.1b) is nonlinear, condition (4.17b) is a sufficient condition for the complete controllability of the system (4.1b) on $[0,T]$ (see [Sas.1]). ■

Theorem 4.2 : Suppose that Assumptions 4.1 and 4.2 are satisfied and that there exist $0 < m_g < M_g < \infty$ and $0 < m_c < M_c < \infty$ such that $M_g M_c \ge 2$ and (4.17a), (4.17b) hold. Let $\hat{x}$ be the unique minimizer of $g(\cdot)$. If a sequence of controls $\{ u_i \}_{i=0}^{\infty}$, generated by Algorithm 4.1, is bounded in the L_∞ norm, then

(i) $\lim_{i\to\infty} \nabla g(\overline{x}_{q_i}(u_i)) = 0$. (4.20a)

(ii) $\lim_{i\to\infty} c(u_i) = g(\hat{x})$, (4.20b)

(iii) $\overline{\lim_{i\to\infty}} \dfrac{c(u_{i+1}) - g(\hat{x})}{c(u_i) - g(\hat{x})} \le 1 - \dfrac{4(1-\alpha)\alpha\beta m_g m_c}{M_g M_c}$. (4.20c)

(iv) The sequence $\{ \overline{x}_{q_i}(u_i) \}_{i=0}^{\infty}$ converges R-linearly to $\hat{x}$.

(v) There exists a $\hat{u} \in G(\infty)$ such that $\nabla c(\hat{u}) = 0$ and the sequence $\{ \|u_i - \hat{u}\|_\infty \}_{i=0}^{\infty}$ converges R-linearly to 0.

Proof : Since the sequence $\{ u_i \}_{i=0}^{\infty}$ is L_∞ bounded, there exists an $\omega > 0$ such that (4.11a), (4.11b) hold. Hence, making use of Lemmas 4.4, and 4.5, we conclude that for every $\varepsilon \in (0,m_c)$, there exists an i_ε such that for all $i \ge i_\varepsilon$, (4.18a) - (4.18c) hold for $q = q_i$ and all $u \in G_{q_i}(\omega)$.

(i) By Theorem 4.1, $\|\nabla c_{q_i}(u_i)\|_2 \to 0$ as $i \to \infty$. Hence, it follows from (4.18b) that $\nabla g(\overline{x}_{q_i}(u_i)) \to 0$ as $i \to \infty$.

(ii) (4.20) follows from (4.7b), (4.18c) and the fact that $\|\nabla c_{q_i}(u_i)\|_2 \to 0$ and $q_i \to \infty$ as $i \to \infty$.

(iii) First we will obtain a bound on the step size λ_i. Making use of (4.6d), (4.18a) and the fact that $h_i(t) = -\left[\frac{\partial \bar{x}_{q_i}(u_i)}{\partial u}(t)\right]^T \nabla g(\bar{x}_{q_i}(u_i))$ and that $\langle v, A^2 v\rangle \leq \|A\| \langle v, Av\rangle$ for all symmetric, positive definite matrices A and vectors v, we obtain that for all $i \geq i_\varepsilon$ and $\lambda \in [0,1]$,

$$\|\bar{x}_{q_i}(u_i + \lambda h_i) - \bar{x}_{q_i}(u_i)\| \leq \left\| \int_0^T \left[\frac{\partial \bar{x}_{q_i}(u_i)}{\partial u}(t)\right] \lambda h_i(t) dt \right\| + K_4(\omega)\|\lambda h_i\|_2^2$$

$$= \lambda \left[\langle \nabla g(\bar{x}_{q_i}(u_i)), \left[\int_0^T \left[\frac{\partial \bar{x}_{q_i}(u_i)}{\partial u}(t)\right]\left[\frac{\partial \bar{x}_{q_i}(u_i)}{\partial u}(t)\right]^T dt \right]^2 \nabla g(\bar{x}_{q_i}(u_i))\rangle \right]^{1/2} + \lambda^2 K_4(\omega)\|h_i\|_2^2$$

$$\leq \lambda \left[(M_c + \varepsilon) \langle \nabla g(\bar{x}_{q_i}(u_i)), \left[\int_0^T \left[\frac{\partial \bar{x}_{q_i}(u_i)}{\partial u}(t)\right]\left[\frac{\partial \bar{x}_{q_i}(u_i)}{\partial u}(t)\right]^T dt \right] \nabla g(\bar{x}_{q_i}(u_i))\rangle \right]^{1/2} + \lambda^2 K_4(\omega)\|h_i\|_2^2$$

$$\leq [(M_c + \varepsilon)^{1/2} + \lambda K_4(\omega)\|h_i\|_2]\lambda \|h_i\|_2 . \tag{4.21}$$

Next, we deduce from (4.17a), (4.6d), (4.7a), (4.21) and the fact that $h_i = -\nabla c_{q_i}(u_i)$, that for all $\lambda \in [0,1]$,

$$c_q(u_i + \lambda h_i) - c_q(u_i) = g(\bar{x}_{q_i}(u_i + \lambda h_i)) - g(\bar{x}_{q_i}(u_i))$$

$$\leq \langle \nabla g(\bar{x}_{q_i}(u_i)), (\bar{x}_{q_i}(u_i + \lambda h_i) - \bar{x}_{q_i}(u_i))\rangle + \frac{M_g}{2}\|\bar{x}_{q_i}(u_i + \lambda h_i) - \bar{x}_{q_i}(u_i)\|^2$$

$$\leq -\lambda\|h_i\|_2^2 + \lambda^2 K_4(\omega)\|\nabla g(\bar{x}_{q_i}(u_i))\|\|h_i\|_2^2 + \frac{M_g}{2}\lambda^2((M_c + \varepsilon)^{1/2} + \lambda K_4(\omega)\|h_i\|_2)^2\|h_i\|_2^2 . \tag{4.22}$$

Since $\|h_i\|_2 \to 0$ and $\|\nabla g(\bar{x}_{q_i}(u_i))\| \to 0$ as $i \to \infty$, it follows from (4.22) that there exists $i'_\varepsilon \geq i_\varepsilon$ such that for $i \geq i'_\varepsilon$ and $\lambda \in [0,1]$,

$$c_{q_i}(u_i + \lambda h_i) - c_{q_i}(u_i) \leq -\lambda\|h_i\|_2^2 + M_g(M_c + 2\varepsilon)\lambda^2\|h_i\|_2^2/2$$

$$= \lambda\theta_i - M_g(M_c + 2\varepsilon)\lambda^2\theta_i/2 . \tag{4.23}$$

Hence, (4.9b) is satisfied with $\lambda_i \geq 2(1-\alpha)\beta/[M_g(M_c + 2\varepsilon)]$ for all $i \geq i'_\varepsilon$, and thus

$$c_{q_i}(u_{i+1}) - c_{q_i}(u_i) \leq \alpha\lambda_i\theta_i \leq 2\alpha(1-\alpha)\beta\theta_i/[M_g(M_c + 2\varepsilon)] , \quad \forall\ i \geq i'_\varepsilon . \tag{4.24}$$

Combining (4.24) and (4.18c), and rearranging terms, we obtain that for all $i \geq i'_\varepsilon$

$$c_{q_i}(u_{i+1}) - g(\hat{x}) \leq [1 - \frac{4(1-\alpha)\alpha\beta m_g(m_c - \varepsilon)}{M_g(M_c + 2\varepsilon)}][c_{q_i}(u_i) - g(\hat{x})] \tag{4.25}$$

Hence it follows from (4.7b) that for $i \geq i'_\varepsilon$,

$$c(u_{i+1}) - g(\hat{x}) \leq [1 - \frac{4(1-\alpha)\alpha\beta m_g(m_c - \varepsilon)}{M_g(M_c + 2\varepsilon)}][c(u_i) - g(\hat{x})] + 2K_5(\omega)/(2^{q_i})^\tau . \tag{4.26}$$

Finally, making use of (4.7b), (4.18c) and the fact that $-\theta_i \geq D^{1/\sigma}/((2^{q_i})^\tau)^{1/\sigma}$, we obtain that for $i \geq i'_\varepsilon$,

$$\begin{aligned} 2M_g(M_c + \varepsilon)[c(u_i) - g(\hat{x})] &\geq 2M_g(M_c + \varepsilon)[c_{q_i}(u_i) - g(\hat{x})] - 2M_g(M_c + \varepsilon)K_5(\omega)/(2^{q_i})^\tau \\ &\geq -\theta_i - 2M_g(M_c + \varepsilon)K_5(\omega)/(2^{q_i})^\tau \\ &\geq D^{1/\sigma}/((2^{q_i})^\tau)^{1/\sigma} - 2M_g(M_c + \varepsilon)K_5(\omega)/(2^{q_i})^\tau . \end{aligned} \tag{4.27}$$

Since $\sigma > 1$ and, by Lemma 4.3, $q_i \to \infty$ as $i \to \infty$, we claim that there exists $i''_\varepsilon \geq i'_\varepsilon$ such that for $i \geq i''_\varepsilon$,

$$4M_g M_c[c(u_i) - g(\hat{x})] \geq \frac{1}{2} D^{1/\sigma}/((2^{q_i})^\tau)^{1/\sigma} . \tag{4.28}$$

Thus, (4.20c) follows from (4.26), (4.28) and the arbitrary choice of ε in $(0, m_c)$.

(iv) Note that

$$\|\bar{x}_{q_i}(u_i) - \hat{x}\|^2 \leq \frac{2}{m_g}[g(\bar{x}_{q_i}(u_i)) - g(\hat{x})] = \frac{2}{m_g}[c_{q_i}(u_i) - g(\hat{x})] . \tag{4.29}$$

This, together with (4.25), leads to the desired result.

(v) Making use of (4.11b), (4.19), (4.25) and the fact that $\lambda_i \leq 1$, we obtain that for $i \geq i'_\varepsilon$,

$$\begin{aligned} \|u_{i+1} - u_i\|_\infty &= \lambda_i \|h_i\|_\infty \\ &\leq \lambda_i \|\frac{\partial \bar{x}_{q_i}(u_i)}{\partial u}\|_\infty \|\nabla g(\bar{x}_{q_i}(u_i))\| \\ &\leq 2\omega M_g[g(\bar{x}_{q_i}(u_i)) - g(\hat{x})] \end{aligned}$$

$$\leq 2\omega M_g[c_{q_{i'_\varepsilon}}(u_{i'_\varepsilon}) - g(\hat{x})][1 - \frac{4(1-\alpha)\alpha\beta m_g(m_c - \varepsilon)}{M_q(M_c + 2\varepsilon)}]^{i - i'_\varepsilon} . \tag{4.30}$$

Let $\eta \triangleq [1 - \frac{4(1-\alpha)\alpha\beta m_g(m_c - \varepsilon)}{M_q(M_c + 2\varepsilon)}]$. Then, for all $j > i \geq i'_\varepsilon$,

$$\| u_j - u_i \|_\infty \leq \sum_{k=i}^{j-1} \| u_{k+1} - u_k \|_\infty$$

$$\leq 2\omega M_g[c_{q_{i'_\varepsilon}}(u_{i'_\varepsilon}) - g(\hat{x})] \sum_{k=i}^{\infty} \eta^{k - i'_\varepsilon}$$

$$\leq 2\omega M_g[c_{q_{i'_\varepsilon}}(u_{i'_\varepsilon}) - g(\hat{x})]\eta^{i - i'_\varepsilon}/(1 - \eta) . \tag{4.31}$$

Therefore $\{ u_i \}_{i=0}^{\infty}$ is a Cauchy sequence in $L_\infty^m[0,T]$. Since $L_\infty^m[0,T]$ is a complete space in the L_∞ norm, there exists a $\hat{u} \in L_\infty^m[0,T]$ such that $u_i \to \hat{u}$ as $i \to \infty$ in L_∞ norm. Let j go to ∞ in (4.31), then for all $i \geq i'_\varepsilon$,

$$\| \hat{u} - u_i \|_\infty \leq 2\omega M_g[c_{q_{i'_\varepsilon}}(u_{i'_\varepsilon}) - g(\hat{x})]\eta^{i - i'_\varepsilon}/(1 - \eta) . \tag{4.32}$$

Thus, $\{ \| \hat{u} - u_i \|_\infty \}_{i=0}^{\infty}$ converges to 0 R-linearly. Finally, by Theorem 4.1, $\nabla c(\hat{u}) = 0$. ■

For comparison, we present a rate of convergence result for the Armijo algorithm (4.8a), (4.8b) as applied to composite functions. The proof of this theorem follows by generalization of the Armijo algorithm rate of convergence theorem for affine-composite functions presented in [Pol.5]:

Theorem 4.3 : Consider the problem

$$\min_{u \in \mathbb{R}^N} c(u), \tag{4.33a}$$

where $c(u) \triangleq g(\bar{x}(u))$, with $g : \mathbb{R}^n \to \mathbb{R}$ a twice continuously differentiable function satisfying (4.17a), with $0 < m_g \leq M_g < \infty$, and $\bar{x} : \mathbb{R}^N \to \mathbb{R}^n$ is a Lipschitz continuously differentiable function such that for some $0 < m_c \leq M_c < \infty$

$$m_c \| z \|^2 \leq \langle z, \left[\frac{\partial \bar{x}(u)}{\partial u} \right] \left[\frac{\partial \bar{x}(u)}{\partial u} \right]^T z \rangle \leq M_c \| z \|^2, \ \forall \ z \in \mathbb{R}^n, \ u \in \mathbb{R}^N . \tag{4.33b}$$

Suppose that $M_c M_g \geq 2$, that $\hat{x}$ is the unique minimizer of $g(\cdot)$, and that $\{ u_i \}_{i=0}^{\infty}$ is a sequence constructed by the Armijo method in solving the problem (4.33a). Then

(i) $\lim_{i \to \infty} \nabla g(\bar{x}(u_i)) = 0$. (4.34a)

(ii) $\lim_{i \to \infty} c(u_i) = g(\hat{x})$, (4.34b)

(iii) $\overline{\lim_{i \to \infty}} \frac{c(u_{i+1}) - g(\hat{x})}{c(u_i) - g(\hat{x})} \leq 1 - \frac{4(1-\alpha)\alpha\beta m_g m_c}{M_g M_c}$. (4.34c)

(iv) The sequence $\{ \bar{x}(u_i) \}_{i=0}^{\infty}$ converges R-linearly to $\hat{x}$.

(v) There exists a $\hat{u} \in \mathbb{R}^N$ such that $\nabla c(\hat{u}) = 0$ and the sequence $\{ \| u_i - \hat{u} \| \}_{i=0}^{\infty}$ converges R-linearly to 0. ■

Again we see that the use of adaptive discretization is preferable to fixed discretization.

5. CONCLUSIONS

There is an accumulation of empirical evidence to support the claim that, in skillful hands, adaptive discretization schemes can produce considerable computational savings in the solution of optimization problems which must be discretized. However, prior to the work presented in this paper, there was no automatic discretization scheme whose computational savings could be predicted on the basis of analysis and whose overall rate of convergence could be established. We expect that the discretization techniques presented in this paper will prove to be of practical importance in engineering design and optimal control.

6. REFERENCES

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

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

[Dan.1] J. W. Daniel, *The Approximation Minimization of Functionals*, Prentice-Hall, Englewood Cliffs, 1971.

[Dun.1] J. C. Dunn, "Diagonally Modified Conditional Gradient Methods for Input Constrained Optimal Control Problems", *SIAM J. Control and Optimization*, Vol. 24, pp. 1177-1191, 1986.

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

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

[Pol.3] E. Polak, "Basics of Minimax Algorithms", in *Nonsmooth Optimization and Related Topics*, F. H. Clarke, V. F. Dem'yanov and F. Giannessi eds., pp. 343-367, Plenum Press, New York, 1989.

[Pol.4] E. Polak and L. He, "A Unified Phase I-Phase II Method of Feasible Directions for Semi-infinite Optimization", *Electronics Research Laboratory, University of California, Berkeley*, Memorandum No. UCB/ERL No. M89/7, Feb. 1989, *J. of Optim. Theory and Appl.* in press.

[Pol.5] E. Polak and E. J. Wiest, "Domain Rescaling Techniques for the Solution of Affinely Parametrized Nondifferentiable Optimal Design Problems", Proc. *27th IEEE Conference on Dec. and Contr.*, Austin, Tx., Dec. 7-9 1988. Dec. 1988.

[Psh.1] B. N. Pshenichnyi and Yu. M. Danilin, *Numerical methods in extremal problems (Chislennye metody v ekstremal'nykh zadachakh)*, Nauka, Moscow, 1975.

[Sas.1] S. S. Sastry and C. A. Desoer, "The Robustness of Controllability and Observability of Linear Time-Varying Systems", *IEEE Trans. on Automatic Control*, Vol. 27, No.4, pp 933-939,1982.