Copyright © 1992, by the author(s).
All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

ON THE USE OF CONSISTENT APPROXIMATIONS IN THE SOLUTION OF SEMI-INFINITE OPTIMIZATION AND OPTIMAL CONTROL PROBLEMS

by

E. Polak

Memorandum No. UCB/ERL M92/36

13 April 1992
(Revised 17 December 1992)

COVER PAGE

ON THE USE OF CONSISTENT APPROXIMATIONS IN THE SOLUTION OF SEMI-INFINITE OPTIMIZATION AND OPTIMAL CONTROL PROBLEMS

by

E. Polak

Memorandum No. UCB/ERL M92/36

13 April 1992
(Revised 17 December 1992)

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

ON THE USE OF CONSISTENT APPROXIMATIONS IN THE SOLUTION OF SEMI-INFINITE OPTIMIZATION AND OPTIMAL CONTROL PROBLEMS

by

E. Polak

Memorandum No. UCB/ERL M92/36

13 April 1992
(Revised 17 December 1992)

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

ON THE USE OF CONSISTENT APPROXIMATIONS IN THE SOLUTION OF SEMI-INFINITE OPTIMIZATION AND OPTIMAL CONTROL PROBLEMS†

by

E. Polak
Department of Electrical Engineering
and Computer Sciences
University of California
Berkeley, CA 94720

ABSTRACT

We define consistency of approximations to an optimization problem in terms of epiconvergence of the cost functions and convergence of stationary points of the approximating problems. We then show that standard discretization techniques decompose semi-infinite optimization and optimal control problems into families of finite dimensional problems, which, together with our optimality functions, are consistent discretizations to the original problems. We then present two types of techniques for using consistent approximations in obtaining an approximate solution of the original problems. The first is a "filter" type technique, similar to that used in conjunction with penalty functions, the second one is an adaptive technique that can be viewed as an implementation of a conceptual algorithm for solving the original problems.

Key Words: semi-infinite optimization, optimal control, discretization theory, epiconvergence, consistent approximations, algorithm convergence theory.

† The research reported herein was sponsored by the Air Force Office of Scientific Research contract AFOSR-90-0068, the National Science Foundation grant ECS-8916168.

1. INTRODUCTION

The vast majority of semi-infinite optimization and continuous optimal control problems cannot be solved without resorting to some form of discretization: domain discretization in semi-infinite optimization and numerical integration in optimal control. There are basically two, not altogether disjoint, discretization techniques in current use.

The first can be viewed as that of *implementation* of conceptual algorithms, and is characterized by a theoretically justified numerical implementation of operations in a conceptual algorithm. There is a moderate size literature dealing with the implementation of conceptual algorithms, see e.g. [14, 15, 21, 22, 23, 27]. In fact, one can even find a theory of implementation of conceptual algorithms, see [14, 18]. Properly constructed implementations of conceptual superlinearly converging algorithms remain superlinearly converging (see, e.g., [15, 23, 27]), but, in our experience, implementations of first order algorithms perform poorly. A particular aspect of implementations is that the approximations used in computing function values and gradients need not be coordinated, which permits the use of implicit methods of integration of differential equations in the solution of optimal control problems. However, the resulting approximate gradients are not gradients for the approximate functions, which explains the degradation of first order optimization algorithms.

The second technique, sometimes called *diagonalization* (see [7, 9]), emulates the use of differentiable penalty functions in nonlinear programming and hence is characterized by the fact that it decomposes a semi-infinite optimization problem into an infinite sequence of nonlinear programming problems, by domain discretization, and a continuous optimal control problem into an infinite sequence of discrete optimal control problems, by explicit numerical integration of the differential equations. Discrete optimal control problems are nonlinear programming problems with special structure. Obviously, the gradients computed for the approximating problems are gradients of the functions appearing in the approximating problems, which prevents the degradation of first order methods. As with penalty methods, the solution of an optimization problem via diagonalization can be viewed as a diagonal progression across minimizing sequences for the approximating problems, i.e, one solves an approximating problem until some test is satisfied, and then uses the resulting end point to start the solution of the next approximating problem. The choice of termination tests is important, since it has a considerable impact on computational effort. In [13] we find some results on the construction of optimal discretization strategies, while in [22, 23] rate-preserving strategies are presented for use with first order algorithms for convex problems.

In this paper we examine two major issues associated with the use of diagonalization in the solution of a semi-infinite optimization and optimal control problems. The first is the establishment of the concept of consistency for the approximating problems, while the second one is the expansion of available "cross-over" tests for use with diagonalization.

In Section 2 we introduce two concepts of consistency based on epiconvergence of the approximating problems, as well as on the convergence of stationary points, characterized as zeros of optimality functions. As we will see, unless some constraint qualification is satisfied, optimality functions may have zeros outside the feasible set. Hence the two definitions make distinctions between whether a constraint qualification is satisfied or not. In [8] it is shown that epiconvergence implies that sequences of global minimizers of the approximating problems converge to global minimizers of the original problem. We strengthen this result by showing that, in addition, sequences of "uniformly" strict local minimizers of the approximating problems converge to a local minimizer of the original problem. To conclude Section 2, we show that differentiable penalty functions in nonlinear programming, the most analyzed form of problem approximation, are consistent approximations in our sense.

In Section 3 and 4 we define sequences of approximating problems for semi-infinite optimization and optimal control problems and show they are consistent in the sense of our definitions. Finally, in Section 5 we present four new master algorithm models for use in solving optimization problems via diagonalization.

2. PRELIMINARIES

Since we intend to examine more than one type of approximation effect, it is simpler, at first, to deal with consistent approximations in abstract form. Thus let $\mathcal{B}$ be a normed linear space, with norm $\|\cdot\|$, and consider the problem

$$\mathbf{P} \qquad \min_{x \in X} f(x) \tag{2.1a}$$

where $f : \mathcal{B} \to \mathbb{R}$ is (at least) lower semicontinuous, and $X \subset \mathcal{B}$ is the constraint set. Next, let $\{ \mathcal{B}_N \}_{N=1}^{\infty}$ be a family of finite dimensional subspaces of $\mathcal{B}$ such that $\mathcal{B}_N = \mathcal{B}$ if $\mathcal{B}$ is finite dimensional ($\mathbb{R}^n$) and $\mathcal{B}_N \subset \mathcal{B}_{N+1}$, for all N, otherwise. Let $\mathbb{N} \triangleq \{ 1,2,3,\dots \}$, and consider the family of approximating problems

$$\mathbf{P}_N \qquad \min_{x \in X_N} f_N(x), \quad N \in \mathbf{N}, \tag{2.1b}$$

where $f_N : \mathcal{B}_N \to \mathbb{R}$ is (at least) lower semicontinuous, and $X_N \subset \mathcal{B}_N$.

The relationship between the $\mathbf{P}_N$ and $\mathbf{P}$ becomes clearer if we restate them all in epigraphical terms. Thus, let the epigraphs (actually subsets of epigraphs) $E \subset \mathbb{R}\times\mathcal{B}$ and $E_N \subset \mathbb{R}\times\mathcal{B}_N$ be defined by

$$E \triangleq \{ (x^0,x) \mid x \in X,\ x_0 \geq f(x) \}, \tag{2.1c}$$

$$E_N \triangleq \{ (x^0, x) \mid x \in X_N,\ x_0 \geq f_N(x) \}. \tag{2.1d}$$

Then the problems **P** and $\mathbf{P}_N$ can be restated in the following, equivalent form:

$$\mathbf{P}_N \qquad \min_{(x^0,x) \in E} x^0, \tag{2.1e}$$

$$\mathbf{P}_N \qquad \min_{(x^0,x) \in E_N} x^0. \tag{2.1f}$$

In the form (2.1e,f), we see that the problems $\mathbf{P}_N$ differ from the problem **P** only in the constraint set. Hence, it is intuitively clear that for the $\mathbf{P}_N$ to be of any use to us at all, the epigraphs E_N must converge to the epigraph E, in the sense that $\underline{\mathrm{Lim}}\, E_N = \overline{\mathrm{Lim}}\, E_N = E$. Because of the form of (2.1c,d), this requirement can be rephrased as follows (see [3, 8, 26]).

Definition 2.1. We will say that the problems in the family $\{ \mathbf{P}_N \}_{N=1}^{\infty}$ *converge epigraphically* to **P** ($\mathbf{P}_N \rightarrow^{Epi} \mathbf{P}$) if

(a) for every $x \in X$, there exists a sequence $\{ x_N \}_{N=1}^{\infty}$, with $x_N \in X_N$, such that $x_N \rightarrow x$ and $\overline{\lim}\, f_N(x_N) \leq f(x)$;

(b) for every infinite sequence $\{ x_N \}_{N \in K}$, where $K \subset \mathbb{N}$, satisfying $x_N \in X_N$, for all $N \in K$ and $x_N \rightarrow^K x$, we have that $x \in X$ and $\underline{\lim}\, f_N(x_N) \geq f(x)$. □

The main consequences of epiconvergence are contained in the theorem below, which requires the following definition.

Definition 2.2. A sequence $\{ x_k \}_{k=k^*}^{\infty}$ of local minimizers for the $\mathbf{P}_k$ is *uniformly strict*, if there exists a $\rho > 0$ such that $f_k(x_k) < f_k(x)$ for all $x \in X_k$, $x \neq x_k$, such that $\|x - x_k\| \leq \rho$ for all $k \geq k^*$. □

Theorem 2.3. Suppose that $\mathbf{P}_N \rightarrow^{Epi} \mathbf{P}$.

(a) If $\{ \hat{x}_N \}_{N=1}^{\infty}$ is a sequence of global minimizers of the $\mathbf{P}_N$, and $\hat{x}$ is any accumulation point of $\{ \hat{x}_N \}_{N=1}^{\infty}$, then $\hat{x}$ is a global minimizer of **P**.

(b) If $\{ \hat{x}_N \}_{N=1}^{\infty}$ is a sequence of uniformly strict local minimizers of the $\mathbf{P}_N$, and $\hat{x}$ is any accumulation point of $\{ \hat{x}_N \}_{N=1}^{\infty}$, then $\hat{x}$ is a local minimizer of **P**.

Proof. (a) A proof of this result can be found in [3, 8, 26], and is therefore omitted.

(b) Suppose that for some infinite subset $K \subset \mathbf{N}$, we have that $\hat{x}_N \rightarrow^K \hat{x}$. Let $\rho > 0$ be a common radius of attraction for the sequence $\{ \hat{x}_N \}_{N \in K}$. If $\hat{x}$ is not a local minimizer for **P**, then there must exist an $x^* \in X$, such that $\|x^* - \hat{x}\| \leq \rho/2$ and $f(x^*) = f(\hat{x}) - 3\delta$, with $\delta > 0$. By Definition 2.1 *(a)*, there exists a sequence $\{ x^*_N \}_{N=1}^{\infty}$, with $x^*_N \in X_N$, such that $x^*_N \rightarrow x^*$ and

$\overline{\lim}_{N \in K} f(x^*_N) \le \overline{\lim} f(x^*_N) \le f(x^*)$, and by Definition 2.1 *(b)*, we must have that $\underline{\lim}_{N \in K} f_N(\hat{x}_N) \ge \underline{\lim} f_N(\hat{x}_N) \ge f(\hat{x})$. Hence there exists an N_o such that for all $N \ge N_o$, $N \in K$, $\|x^*_N - \hat{x}_N\| < \rho$, $f(x^*_N) \le f(\hat{x}) - 2\delta$ and $f(\hat{x}_N) \ge f(\hat{x}) - \delta$, which contradicts the local optimality of the x_N. Hence the theorem is true. □

In the absence of convexity, nonlinear programming algorithms can only be shown to compute stationary points that are, hopefully, local minimizers of the $\mathbf{P}_N$, but not necessarily global minimizers of the $\mathbf{P}_N$. The worst outcome of such a process is illustrated in Fig. 2.1, where a sequence of local minimizers converges to a global maximizer. In view of Theorem 2.3, we note that epiconvergence ensures, at least, that uniformly strict local minimizers of the $\mathbf{P}_N$ cannot converge to anything but local minimizers of $\mathbf{P}$.

It is sometimes useful (as in the case of penalty function methods) to reformulate either the problem $\mathbf{P}$ or the problems $\mathbf{P}_N$, $N = 1, 2, \ldots$, or both, in the following equivalent forms:

$$\mathbf{P}^0 \qquad \min_{x \in \mathcal{B}} f^0(x), \tag{2.1g}$$

$$\mathbf{P}_N^0 \qquad \min_{x \in \mathcal{B}_N} f_N^0(x), \tag{2.1h}$$

where $f^0 : \mathcal{B} \to \overline{\mathbb{R}}$ and $f_N^0 : \mathcal{B} \to \overline{\mathbb{R}}$ are defined as follows: $f^0(x) = f(x)$ for all $x \in X$, and $f^0(x) = +\infty$ otherwise, and, similarly, $f_N^0(x) = f_N(x)$ for all $x \in X_N$, and $f_N^0(x) = +\infty$ otherwise. The following result should be obvious.

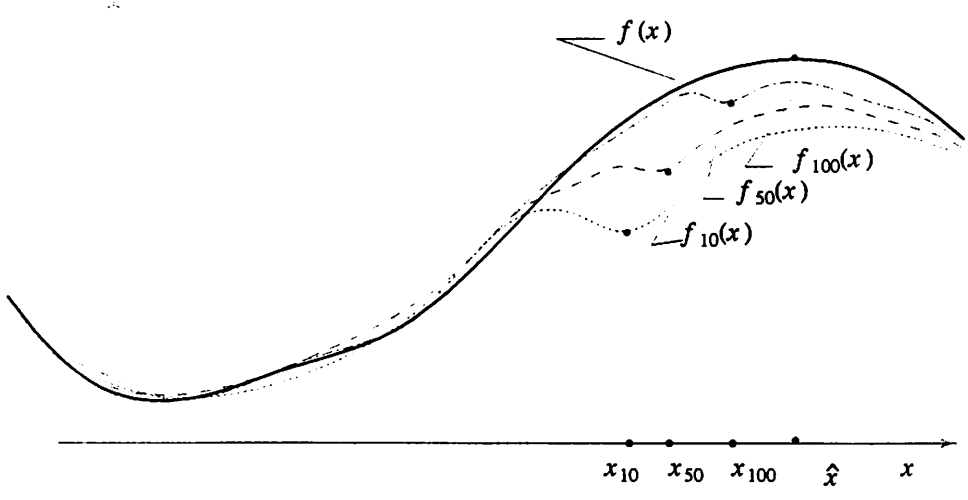

Fig. 2.1. Convergence of Local Minimizers x_N to a Global Maximizer $\hat{x}$

Corollary 2.4. Suppose that one of the following four statements is true: *(i)* $\mathbf{P}_N \to^{Epi} \mathbf{P}$, *(ii)* $\mathbf{P}_N \to^{Epi} \mathbf{P}^0$, *(iii)* $\mathbf{P}_N^0 \to^{Epi} \mathbf{P}$, *(iv)* $\mathbf{P}_N^0 \to^{Epi} \mathbf{P}^0$.

(a) If $\{\hat{x}_N\}_{N=1}^{\infty}$ is a sequence of global minimizers of the $\mathbf{P}_N$, and $\hat{x}$ is any accumulation point of $\{\hat{x}_N\}_{N=1}^{\infty}$, then $\hat{x}$ is a global minimizer of $\mathbf{P}$.

(b) If $\{\hat{x}_N\}_{N=1}^{\infty}$ is a sequence of uniformly strict local minimizers of the $\mathbf{P}_N$, and $\hat{x}$ is any accumulation point of $\{\hat{x}_N\}_{N=1}^{\infty}$, then $\hat{x}$ is a local minimizer of $\mathbf{P}$. □

As we have already mentioned, nonlinear programming algorithms can only be shown to compute stationary points, often characterized as the zeros of an optimality function (see, e.g. [18, 19]). We will therefore characterize stationarity of points with respect to the problems $\mathbf{P}$, $\mathbf{P}_N$, in terms of the zeros of optimality functions, $\theta : \mathcal{D} \to \mathbb{R}$ for $\mathbf{P}$ and $\theta_N : \mathcal{D}_N \to \mathbb{R}$ for $\mathbf{P}_N$, $N \in \mathbb{N}$, where $\mathcal{D} \subset \mathcal{B}$ and $\mathcal{D}_N \subset \mathcal{B}_N$, i.e., the optimality functions may not be defined on the entire space. Quite commonly (see, e.g., Section 4), we have that $\mathcal{D}_N = \mathcal{D} \cap \mathcal{B}_N$.

Definition 2.5. We will say that a function $\theta : \mathcal{D} \to \mathbb{R}$ is an *optimality function* for $\mathbf{P}$ if *(i)* $X \subset \mathcal{D}$, *(ii)* $\theta(\cdot)$ is upper semicontinuous, *(iii)* $\theta(x) \le 0$ for all $x \in \mathcal{D}$, and *(iv)* for $\hat{x} \in X$, $\theta(\hat{x}) = 0$, if and only if $\hat{x}$ is a stationary point for $\mathbf{P}$. Similarly, we will say that a function $\theta_N : \mathcal{D}_N \to \mathbb{R}$ is an *optimality function* for $\mathbf{P}_N$ if *(i)* $X_N \subset \mathcal{D}_N$, *(ii)* $\theta_N(\cdot)$ is upper semicontinuous, *(iii)* $\theta_N(x) \le 0$ for all $x \in \mathcal{D}_N$, and *(iv)* for $\hat{x}_N \in X_N$, $\theta_N(\hat{x}_N) = 0$, if and only if $\hat{x}_N$ is a stationary point for $\mathbf{P}_N$. □

While all the optimality functions that we will see in this paper are continuous, there are minimax and feasible directions algorithms that are based on upper semicontinuous optimality functions (see, e.g. [18, 19]). Hence our assumption of upper semicontinuity in the definition of optimality functions is inspired by practical considerations, rather than a search for generality.

The introduction of optimality functions leads us to two concepts of consistency of approximation, as we see below.

Definition 2.6. Let $\theta(\cdot)$, $\theta_N(\cdot)$, $N \in \mathbb{N}$, be optimality functions for $\mathbf{P}$, $\mathbf{P}_N$, respectively. We will say that the pairs $(\mathbf{P}_N, \theta_N)$, in the sequence $\{(\mathbf{P}_N, \theta_N)\}_{N=1}^{\infty}$ are *weakly consistent approximations* to the pair $(\mathbf{P}, \theta)$, if *(i)* $\mathbf{P}_N \to^{Epi} \mathbf{P}$, or *(ii)* $\mathbf{P}_N \to^{Epi} \mathbf{P}^0$, or *(iii)* $\mathbf{P}_N^0 \to^{Epi} \mathbf{P}$, or *(iv)* $\mathbf{P}_N^0 \to^{Epi} \mathbf{P}^0$, and for any sequence $\{x_N\}_{N \in K}$, $K \subset \mathbb{N}$, with $x_N \in X_N$ for all $N \in K$, such that $x_N \to x$, $\overline{\lim}\, \theta_N(x_N) \le \theta(x)$. □

The next definition implies that a constraint qualification is satisfied.

Definition 2.7. Let $\theta(\cdot)$, $\theta_N(\cdot)$, $N \in \mathbb{N}$, be optimality functions for $\mathbf{P}$, $\mathbf{P}_N$, respectively. We will

say that the pairs $(\mathbf{P}_N, \theta_N)$, in the sequence $\{ (\mathbf{P}_N, \theta_N) \}_{N=1}^{\infty}$ are *consistent approximations* to $(\mathbf{P}, \theta)$, if they are weakly consistent approximations, and, in addition $\theta(x) < 0$ for all $x \notin X$ and $\theta_N(x) < 0$ for all $x \notin X_N, N \in \mathbb{N}$. □

The best known examples of consistent approximations are not those used in semi-infinite programming and optimal control, but those found in nonlinear programming, in the form of various penalty function methods. It is useful to digress for a moment from our original charge and examine what can be said about penalty methods, so as to establish a yardstick for comparisons. Thus, consider the simple case where

$$\mathbf{P} \qquad \min \{ f(x) \mid g(x) = 0 \}, \tag{2.2a}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^l$, with $l < n$ are both continuously differentiable. Clearly, for the above problem, $X = \{ x \in \mathbb{R}^n \mid g(x) = 0 \}$. The simplest approximations using penalty functions have the form

$$\mathbf{P}_N \qquad \min_{x \in \mathbb{R}^n} f_N(x), \quad N \in \mathbb{N}, \tag{2.2b}$$

where $f_N : \mathcal{B}_N \to \mathbb{R}$ are defined by

$$f_N(x) \triangleq f(x) + \tfrac{1}{2} c_N \|g(x)\|^2, \tag{2.2c}$$

with $\{ c_N \}_{N=1}^{\infty}$ a strictly increasing sequence of positive penalties that diverges to infinity.

To obtain consistency results, we must restate $\mathbf{P}$ in the equivalent form

$$\mathbf{P}^0 \qquad \min_{x \in \mathbb{R}^n} f^0(x), \tag{2.2d}$$

where $f^0 : \mathbb{R} \to \overline{\mathbb{R}}$ is defined by $f^0(x) = f(x)$ for all $x \in \mathbb{R}^n$ such that $g(x) = 0$, and $f^0(x) = +\infty$, otherwise.

Theorem 2.8. The problems in the sequence $\{ \mathbf{P}_N \}_{N=1}^{\infty}$, defined in (2.2b), converge epigraphically to $\mathbf{P}^0$, defined in (2.2d).

Proof. First, since for any $\hat{x} \in \mathbb{R}^n$, $f_N(\hat{x}) \le f^0(\hat{x})$, it follows that $\overline{\lim} f_N(\hat{x}) \le f^0(\hat{x})$. Hence setting $x_N = \hat{x}$ for all $N \in \mathbb{N}$, we see that part *(a)* of Definition 2.1 is satisfied. Next, suppose that the sequence $\{ x_N \}_{N=1}^{\infty}$ converges to the point $\hat{x}$. If $g(\hat{x}) \neq 0$, then we must have that $\infty = \underline{\lim} f_N(x_N) = f^0(\hat{x})$. If $g(\hat{x}) = 0$, then we must have $\underline{\lim} f_N^0(x_N) \ge \lim f(x_N) = f^0(\hat{x})$. Hence we see that part *(b)* of Definition 2.1 is satisfied. □

Next we will introduce optimality functions for the problems $\mathbf{P}$ and $\mathbf{P}_N$. Let $\theta : \mathbb{R}^n \to \mathbb{R}$ be defined by

$$\theta(x) \triangleq -\min \{ \|\xi^0 \nabla f(x) - g_x(x)^T \xi\|^2 \mid (\xi^0)^2 + \|\xi\|^2 = 1 \} - \|g_x(x)^T g(x)\|^2, \tag{2.3a}$$

where $\xi \in \mathbb{R}^l$, and, for any $N \in \mathbb{N}$, let $\theta_N : \mathbb{R}^n \to \mathbb{R}$ be defined by

$$\theta_N(x) \triangleq -\left\| \frac{1}{\sqrt{1 + c_N^2 \|g(x)\|^2}} \nabla f(x) + g_x(x)^T \frac{c_N g(x)}{\sqrt{1 + c_N^2 \|g(x)\|^2}} \right\|^2 - \left\| \frac{1}{c_N} \nabla f(x) + g_x(x)^T g(x) \right\|^2,$$

$$= -\left(\frac{1}{1 + c_N^2 \|g(x)\|^2} + \frac{1}{c_N^2} \right) \nabla f_N(x). \tag{2.3b}$$

Clearly, $\theta(x) = 0$ at any point that satisfies the constraint $g(x) = 0$ and the F. John condition of optimality; while $\theta_N(x) = 0$ if and only if $\nabla f_N(x) = 0$. Since the continuity and sign properties of these functions are obvious, it follows that they are optimality functions.

Theorem 2.9. The pairs in the sequence $\{ (\mathbf{P}_N, \theta_N) \}_{N=1}^{\infty}$, defined by (2.2b) and (2.3b), are weakly consistent approximations to $(\mathbf{P}, \theta)$, defined by (2.2a), (2.3a). Furthermore, if $g_x(x)$ has maximum row rank for all $x \in \mathbb{R}^n$, then they are consistent approximations to $\mathbf{P}$.

Proof. First, by Theorem 2.8, the problems $\mathbf{P}_N$ converge epigraphically to $\mathbf{P}^0$. Next, let $\{ x_N \}_{N=1}^{\infty}$ be any sequence that has a limit point, say $\hat{x}$. Then, because for all x_N we must have that

$$-\left\| \frac{1}{\sqrt{1 + c_N^2 \|g(x_N)\|^2}} \nabla f(x_N) + g_x(x_N)^T \frac{c_N g(x_N)}{\sqrt{1 + c_N^2 \|g(x_N)\|^2}} \right\|$$

$$\leq -\min \{ \|\xi^0 \nabla f(x_N) + g_x(x_N)^T \xi\|^2 \mid (\xi^0)^2 + \|\xi\|^2 = 1 \}, \tag{2.4a}$$

and because

$$\left\| \frac{1}{c_N} \nabla f(x_N) + g_x(x_N)^T g(x_N) \right\|^2 \to \|g_x(\hat{x})^T g(\hat{x})\|^2, \tag{2.4b}$$

as $N \to \infty$, it follows that $\overline{\lim}\, \theta_N(x_N) \leq \theta(\hat{x})$, which shows that we have weak consistency.

Now suppose that $g_x(x)$ has maximum rank for all $x \in \mathbb{R}^n$. Then $\theta(x) = 0$ implies that $g(x) = 0$, i.e., that $x \in X$. Since $X_N = \mathbb{R}^n$, it now follows that we have consistency. □

We will now proceed to show that we can construct consistent approximations to semi-infinite optimization and optimal control problems.

3. CONSISTENT APPROXIMATIONS FOR SEMI-INFINITE OPTIMIZATION

To avoid excessively burdensome notation, we will restrict ourselves to the following two simple examples of semi-infinite optimization problems. The first is an unconstrained minimax problem:

MMP
$$\min_{x \in \mathbb{R}^n} \psi^0(x), \tag{3.1a}$$

while the second one is an inequality constrained minimax problem:

ICP
$$\min_{x \in X} \psi^0(x), \tag{3.1b}$$

where

$$X \triangleq \{ x \in \mathbb{R}^n \mid \psi^1(x) \le 0 \}. \tag{3.1c}$$

In (3.1abc), for $j = 0, 1$, the functions $\psi^j : \mathbb{R}^n \to \mathbb{R}$, are assumed to be of the form

$$\psi^j(x) \triangleq \max_{y \in Y} \phi^j(x, y), \tag{3.1d}$$

with $\phi^j : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ twice continuously differentiable, and the set $Y \triangleq [0, 1]$.

Now, for $N = 1, 2, 3, \ldots$, let $Y_N \triangleq \{ 0, 1/N, 2/N, \ldots, N/N \}$, and let

$$\psi_N^j(x) \triangleq \max_{y \in Y_N} \phi^j(x, y), \quad j = 0, 1. \tag{3.2a}$$

For $N = 1, 2, 3, \ldots$, we now define the approximating problems $\mathbf{MMP}_N$ and $\mathbf{ICP}_N$ by

$\mathbf{MMP}_N$
$$\min_{x \in \mathbb{R}^n} \psi_N^0(x), \tag{3.2b}$$

$\mathbf{ICP}_N$
$$\min_{x \in X_N} \psi_N^0(x), \tag{3.2c}$$

where

$$X_N \triangleq \{ x \in \mathbb{R}^n \mid \psi_N^1(x) \le 0 \}. \tag{3.2d}$$

Lemma 3.1. For any bounded set $S \subset \mathbb{R}^n$, there exists a constant $L < \infty$ such that for all $N = 1, 2, 3, \ldots$, and $x \in S$,

$$-\frac{L}{N} \le \psi_N^j(x) - \psi^j(x) \le 0, \quad j = 0, 1. \tag{3.3}$$

Proof. Let $j \in \{ 0, 1 \}$. First, since $Y_N \subset Y$, we always have that $\psi_N^j(x) \le \psi^j(x)$. Next, let $y^x \in Y$ be such that $\psi^j(x) = \phi^j(x, y^x)$. Then there exists a $y_N^x \in Y_N$ such that $| y^x - y_N^x | \le 1/N$. Hence

$$\psi_N^j(x) \geq \phi^j(x, y_N^x) \geq \phi^j(x, y^x) - \frac{L}{N}, \tag{3.4}$$

where $L < \infty$ is a Lipschitz constant for $\phi^j(\cdot,\cdot)$ on $S \times Y$, $j = 0, 1$. □

Theorem 3.2. The problems $\mathbf{MMP}_N$ and $\mathbf{ICP}_N$ converge epigraphically to the problems **MMP** and **ICP**, respectively.

Proof. We only need to consider the problems $\mathbf{ICP}_N$ and **ICP**, because if we set $\phi^1(x, y) \equiv 0$, then these problems degenerate to $\mathbf{MMP}_N$ and **MMP**, respectively.

Our first observation is that because of (3.3), $\psi_N^0(x) \leq \psi^0(x)$. Hence, since $X \subset X_N$ for all N, given any $x \in X$, we can define the sequence $\{x_N\}_{N=1}^{\infty}$ by $x_N = x$ for all N, and we immediately obtain that $x_N \in X_N$ for all N and $\overline{\lim}\, \psi_N^0(x_N) \leq \psi^0(x)$, which shows that part *(a)* of Definition 2.1 is satisfied.

Next, suppose that $\{x_N\}_{N=1}^{\infty}$ is a sequence such that $x_N \in X_N$ and $x_N \to x$ as $N \to \infty$. It now follows from the fact that $\psi_N^1(x_N) \leq 0$ and (3.3) that $\psi^1(x_N) \leq K/N$ for all N. Because $\psi^1(\cdot)$ is continuous, we conclude that $\psi^1(x) \leq 0$, i.e., that $x \in X$. Furthermore, again by (3.3), $\underline{\lim}\, \psi_N^0(x_N) \leq \underline{\lim}\, \psi^0(x_N) = \psi^0(x)$, which shows that part *(b)* of Definition 2.1 is satisfied. Hence our proof is complete. □

Before we can deal with the question of consistency, we need to introduce optimality functions for the problems **MMP**, $\mathbf{MMP}_N$, **ICP**, and $\mathbf{ICP}_N$. Optimality conditions for **ICP** ($\mathbf{ICP}_N$) can be obtained from those for **MMP** ($\mathbf{MMP}_N$), by making use of the parametrized functions $F_{x'} : \mathbb{R}^n \to \mathbb{R}$, and $F_{N,x'} : \mathbb{R}^n \to \mathbb{R}$, $N = 1, 2, \ldots$, with the parameter $x' \in \mathbb{R}^n$, defined by

$$F_{x'}(x) \triangleq \max\{\psi^0(x) - \psi^0(x') - \gamma\psi^1(x')_+, \psi^1(x) - \psi^1(x')_+\}, \tag{3.5a}$$

$$F_{N,x'}(x) \triangleq \max\{\psi_N^0(x) - \psi_N^0(x') - \gamma\psi_N^1(x')_+, \psi_N^1(x) - \psi_N^1(x')_+\}, \tag{3.5b}$$

where $\gamma > 0$, and $\psi^1(x)_+ \triangleq \{\max\psi^1(x), 0\}$, and $\psi_N^1(x)_+ \triangleq \{\max\psi_N^1(x), 0\}$. It should be obvious that if $\hat{x}$ is a local mininimizer for **ICP** ($\mathbf{ICP}_N$), then it is also a local unconstrained minimizer for $F_{\hat{x}}(\cdot)$ ($F_{N,\hat{x}}(\cdot)$). Hence, As in [19], for $\gamma \geq 0$, let the set valued maps $\overline{G}_\gamma^0(x)$, $\overline{G}_{N,\gamma}^0(x)$, $\overline{G}^1(x)$, $\overline{G}_N^1(x)$, with values in $\mathbb{R}^{n+1}$, be defined as follows:[‡]

$$\overline{G}_\gamma^0(x) \triangleq \operatorname*{co}_{y \in Y}\left\{\begin{bmatrix} \psi^0(x) - \phi^0(x, y) + \gamma\psi^1(x)_+ \\ \nabla_x \phi^0(x, y) \end{bmatrix}\right\}, \tag{3.5c}$$

[‡] The parameter γ is not needed for the optimality conditions, but will be needed in the algorithms that we will describe in Section 5.

$$\bar{G}^0_{N,\gamma}(x) \triangleq \underset{y \in Y_N}{\mathrm{co}} \left\{ \begin{bmatrix} \psi^0_N(x) - \phi^0(x,y) + \gamma\psi^1_N(x)_+ \\ \nabla_x \phi^0(x,y) \end{bmatrix} \right\}, \tag{3.5d}$$

$$\bar{G}^1(x) \triangleq \underset{y \in Y}{\mathrm{co}} \left\{ \begin{bmatrix} \psi^1(x)_+ - \phi^1(x,y) \\ \nabla_x \phi^1(x,y) \end{bmatrix} \right\}, \quad \bar{G}^1_N(x) \triangleq \underset{y \in Y_N}{\mathrm{co}} \left\{ \begin{bmatrix} \psi^1_N(x)_+ - \phi^1(x,y) \\ \nabla_x \phi^1(x,y) \end{bmatrix} \right\}, \tag{3.5e}$$

We will denote the elements of these sets by $\bar{\xi} = (\xi^0, \xi)$, with $\xi \in \mathbb{R}^n$. For the problems **MMP** and $\mathbf{MMP}_N$, we set $\gamma = 0$ and we define the optimality functions $\theta_{\mathbf{MMP}}, \theta_{\mathbf{MMP}_N}$, by

$$\theta_{\mathbf{MMP}}(x) \triangleq -\min_{\bar{\xi} \in \bar{G}^0_0(x)} \xi^0 + \tfrac{1}{2}\|\xi\|^2, \quad \theta_{\mathbf{MMP}_N}(x) \triangleq -\min_{\bar{\xi} \in \bar{G}^0_{N,0}(x)} \xi^0 + \tfrac{1}{2}\|\xi\|^2 . \tag{3.6a}$$

For the problems **ICP** and $\mathbf{ICP}_N$, we set $\gamma > 0$ and we define the optimality functions $\theta_{\mathbf{ICP}}, \theta_{\mathbf{ICP}_N}$, by

$$\theta_{\mathbf{ICP}}(x) \triangleq -\min_{\bar{\xi} \in \mathrm{co}\{\bar{G}^0_\gamma(x), \bar{G}^1(x)\}} \xi^0 + \tfrac{1}{2}\|\xi\|^2, \quad \theta_{\mathbf{ICP}_N}(x) \triangleq -\min_{\bar{\xi} \in \mathrm{co}\{\bar{G}^0_{N,\gamma}(x), \bar{G}^1_N(x)\}} \xi^0 + \tfrac{1}{2}\|\xi\|^2 . \tag{3.6b}$$

Theorem 3.3.

(a) If $\hat{x}$ is a local minimizer for **MMP** (x_N is a local minimizer for $\mathbf{MMP}_N$), then $0 \in \partial\psi^0(\hat{x})$ ($0 \in \partial\psi_N(x_N)$) (where $\partial\psi^0(\cdot)$, $\partial\psi^0_N(\cdot)$ denote the Clarke generalized gradients [6]).

(b) If $\hat{x}$ is a local minimizer for **ICP** ($\hat{x}_N$ is a local minimizer for $\mathbf{ICP}_N$), then (with $\gamma > 0$) $0 \in \partial F_{\hat{x}}(\hat{x})$ ($0 \in \partial F_{N,\hat{x}_N}(\hat{x}_N)$).

(c) For any $x \in \mathbb{R}^n$, $0 \in \partial\psi^0(x) \iff 0 \in \bar{G}^0_0(x) \iff \theta_{\mathbf{MMP}}(x) = 0$; $0 \in \partial\psi^0_N(x) \iff 0 \in \bar{G}^0_{N,0}(x) \iff \theta_{\mathbf{MMP}_N}(x) = 0$.

(d) Let *gammag* > 0. Then, for any $x \in \mathbb{R}^n$, $0 \in \partial F_x(x) \iff 0 \in \mathrm{co}\{\bar{G}^0_\gamma(\hat{x}), \bar{G}^1(\hat{x})\} \iff \theta_{\mathbf{ICP}}(x) = 0$: $0 \in \partial F_{N,x}(x) \iff 0 \in \mathrm{co}\{\bar{G}^0_{N,\gamma}(\hat{x}), \bar{G}^1_N(\hat{x})\} \iff \theta_{\mathbf{ICP}_N}(x) = 0$.

(e) For $\gamma \geq 0$, the set valued maps $\bar{G}^0_\gamma(\cdot)$, $\bar{G}^0_N(\cdot)$, $\bar{G}^1(\cdot)$, $\bar{G}^1_N(\cdot)$, $N = 1,2,3,\ldots$, and the corresponding optimality functions $\theta_{\mathbf{MMP}}(\cdot)$, $\theta_{MMP_N}(\cdot)$, $\theta_{\mathbf{ICP}}(\cdot)$, $\theta_{\mathbf{ICP}_N}(\cdot)$, $N = 1,2,3,\ldots$, are all continuous.

(f) For every bounded set $S \subset \mathbb{R}^n$, there exists a $K < \infty$ such that for all $x \in S$ and all $N = 1,2,3,\ldots$,

$$|\theta_{\mathbf{MMP}_N}(x) - \theta_{\mathbf{MMP}}(x)| \leq \frac{K}{N}, \tag{3.7a}$$

$$|\theta_{\mathbf{ICP}_N}(x) - \theta_{\mathbf{ICP}}(x)| \leq \frac{K}{N} . \tag{3.7b}$$

Proof. The proofs of *(a)* - *(e)* can be found in Examples 5.2 and 5.5 in [19]. Hence we only need to deal with *(f)*. Thus, suppose that for $x \in S$, $\bar{\xi}_N \in \bar{G}^0_{N,0}(x)$ is such that $\theta_{\mathbf{MMP}_N}(x) = -(\xi^0_N + \tfrac{1}{2}\|\xi_N\|^2)$.

Then the vector $(\xi_N^0 + \psi^0(x) - \psi_N^0(x), \xi_N) \in \bar{G}_0^0(x)$. It therefore follows from (3.3) that

$$-\theta_{\mathbf{MMP}}(x) \le \xi_N^0 + \psi^0(x) - \psi_N^0(x) + \tfrac{1}{2}\|\xi_N\|^2 \le -\theta_{\mathbf{MMP}_N}(x) + \frac{K}{N}. \tag{3.8a}$$

Next suppose that $\bar{\xi}_* \in \bar{G}_0^0(x)$ is such that $\theta_{\mathbf{MMP}}(x) = -(\xi_*^0 + \tfrac{1}{2}\|\xi_*\|^2)$. Then, by Caratheodory's Theorem, there exist barycentric coordinates $\mu^j \ge 0$, $j = 1,\ldots,n+1$, such that $\sum_{j=1}^{n+1} \mu^j = 1$, $\xi_*^0 = \psi^0(x) - \sum_{j=1}^{n+1} \mu^j \phi^0(x, y_j)$, and $\xi_* = \sum_{j=1}^{n+1} \mu^j \nabla_x \phi^0(x, y_j)$, with $y_j \in Y$. Clearly, there exist $y_{N,j} \in Y_N$, $j = 1,\ldots,n+1$, such that $|y_j - y_{N,j}| \le 1/N$. Let $\bar{\xi}_{N*} \in \bar{G}_{N,0}^0(x)$ be defined by $\xi_{N*}^0 = \psi_N^0(x) - \sum_{j=1}^{n+1} \mu^j \phi^0(x, y_{N,j})$, and $\xi_{N*} = \sum_{j=1}^{n+1} \mu^j \nabla_x \phi^0(x, y_{Nj})$. Then we must have that

$$|\xi_{N*}^0 - \xi_*^0 - \psi_N^0(x) + \psi^0(x)| \le \frac{L}{N}, \tag{3.8b}$$

$$\|\xi_{N*} - \xi_*\| \le \frac{L}{N}, \tag{3.8c}$$

where we assume that $L < \infty$ is a common local Lipschitz constant for $\phi(\cdot,\cdot)$ and $\nabla\phi(\cdot,\cdot)$ on $S \times Y$. Now, (3.8b), together with (3.3), implies that $|\xi_{N*}^0 - \xi_*^0| \le 2L/N$. Since the set valued maps $\bar{G}_0^0(x)$, $\bar{G}_{N,0}^0(x)$ are bounded on bounded sets, we now conclude that (3.7a) holds for some $K < \infty$. A similar proof applies to (3.7b). □

It is obvious from Theorem 3.3 that the functions $\theta_{\mathbf{MMP}}(\cdot)$, $\theta_{\mathbf{MMP}_N}(\cdot)$ are optimality functions for the problems **ICP** and $\mathbf{ICP}_N(x)$, respectively; similarly, it is obvious from Theorem 3.3 that the functions $\theta_{\mathbf{ICP}}(\cdot)$, $\theta_{\mathbf{ICP}_N}(\cdot)$ are optimality functions for the problems **ICP** and $\mathbf{ICP}_N(x)$, respectively. We are now ready to state our final result, which is obvious in view of Theorem 3.2 and Theorem 3.3 (see parts *(d), (e)*). Referring to Proposition 5.5 in [19], we see that $\theta_{\mathbf{ICP}}(x) = 0$ if and only if $0 \in \partial\psi^1(x)$, and similarly, $\theta_{\mathbf{ICP}_N}(x) = 0$ if and only if $0 \in \partial\psi_N^1(x)$. The requirement that $0 \notin \partial\psi^1(x)$ for all $x \in X$ is known as the generalized Mangasarian-Fromowitz constraint qualification (see [17]) and we invoke it to ensure consistency (i.e., to ensure that whenever $\psi^1(x) > 0$, $\theta_{\mathbf{ICP}}(x) < 0$, etc.).

Theorem 3.4.

(a) Consider the problems **MMP**, $\mathbf{MMP}_N$, with the assumptions stated. Then the pairs in the sequence $\{(\mathbf{MMP}_N, \theta_{\mathbf{MMP}_N})\}_{N=1}^{\infty}$ are consistent approximations to $(\mathbf{MMP}, \theta_{\mathbf{MMP}})$.

(b) Consider the problems **ICP**, $\mathbf{ICP}_N$, with the assumptions stated. Then the pairs in the sequence $\{(\mathbf{ICP}_N, \theta_{\mathbf{ICP}_N})\}_{N=1}^{\infty}$ are weakly consistent approximations to $(\mathbf{ICP}, \theta_{\mathbf{ICP}})$. Furthermore, if for all x such that $\psi^1(x) > 0$, $0 \notin \partial\psi^1(x)$, and, in addition, for all $N \in \mathbb{N}$, and x_N such that $\psi_N^1(x_N) > 0$, $0 \notin \partial\psi_N^1(x_N)$, then the pairs in the sequence $\{(\mathbf{ICP}_N, \theta_{\mathbf{ICP}_N})\}_{N=1}^{\infty}$ are consistent approximations to $(\mathbf{ICP}, \theta_{\mathbf{ICP}})$. □

4. CONSISTENT APPROXIMATIONS FOR OPTIMAL CONTROL

We can illustrate most of the issues related to optimal control problems by considering two fixed time optimal control problems. The first is an unconstrained optimal control problem, while the second one is an optimal control problem with control and inequality end point constraints.

Optimal control problems always involve the controls and trajectories of a dynamical system. We will assume that this dynamical system is described by the differential equation

$$\frac{d}{dt}x(t) = h(x(t), u(t)),\ t \in [0,1],\ x(0) = \xi, \tag{4.1}$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^m$, and hence $h : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$. Since we will keep the initial condition constant and only vary the control, we will denote the solution of (4.1) by $x^u(t)$.

The following assumption is standard:

Assumption 4.1. Let $\rho_{\max} \in (0,\infty)$ be a given, very large number. The function $h(\cdot,\cdot)$ in (4.1) is continuously differentiable, and there exists a constant $K \in [1,\infty)$ such that

(i) for all $x', x'' \in \mathbb{R}^n$, and $v', v'' \in B(0,\rho_{\max})$ the following three relations hold:

$$\|h(x',v') - h(x'',v'')\| \le K[\|x' - x''\| + \|v' - v''\|], \tag{4.2a}$$

$$\|h_x(x',v') - h_x(x'',v'')\| \le K[\|x' - x''\| + \|v' - v''\|], \tag{4.2b}$$

$$\|h_u(x',v') - h_u(x'',v'')\| \le K[\|x' - x''\| + \|v' - v''\|], \tag{4.2c}$$

(ii) for all $x \in \mathbb{R}^n$, $v \in B(0,\rho_{\max})$,

$$\|h(x,v)\| \le K[\|x\| + 1]\,. \tag{4.2d}$$

□

Referring to [1, 16 pp. 136-143], we see that under Assumption 4.1, the solution $x^u(\cdot)$ is Lipschitz continuously Frechet differentiable in u on the interior of the bounded subset

$$\mathbf{U} \triangleq \{\, u \in L_\infty^m[0,1] \mid \|u\|_\infty \le \rho_{\max} \,\}, \tag{4.3a}$$

Now $L_\infty^m[0,1]$ is not a Hilbert space, while $\mathbb{R}^n$, on which the approximating problems will be defined, is a Hilbert space, a fact that causes considerable technical difficulties, because of the form of the optimality functions that we use in $\mathbb{R}^n$. This difficulty can be removed by introducing the pre-Hilbert space:

$$L_{\infty,2}^m[0,1] \triangleq (L_\infty^m[0,1], \langle\cdot,\cdot\rangle_2, \|\cdot\|_2), \tag{4.3b}$$

i.e., the elements of the space $L_{\infty,2}^m[0,1]$ are functions $u \in L_\infty^m[0,1]$, but it is endowed with the scalar product and norm used on $L_2^m[0,1]$. The space $L_{\infty,2}^m[0,1]$ is not complete; however, it is dense in

$L_2^m[0,1]$.

It is reasonably straightforward to deduce from [1, 16] that the solution of our differential equation (4.1), $x^u(\cdot)$, is also Lipschitz continuously Frechet differentiable in u (in the $L_2^m[0,1]$ topology) on the following subset of $L_{\infty,2}^m[0,1]$:

$$\mathbf{U}^o \triangleq \{ u \in L_{\infty,2}^m[0,1] \mid \|u\|_\infty < \delta\rho_{\max} \}, \tag{4.3c}$$

where $\delta \in (0,1)$ is near unity. Clearly, $\mathbf{U}^o \subset \mathbf{U}$. For each $t \in [0,1]$, the Frechet differential $Dx^u(t\,;\cdot)$ is defined on $L_{\infty,2}^m[0,1]$, and takes values in $\mathbb{R}^n$.

For $j = 0,1,\ldots,q$, let $g^j : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz continuously differentiable function, and let

$$f^j(u) \triangleq g^j(x^u(1)), \quad j = 0,1,\ldots,q, \tag{4.4a}$$

$$\psi(u) \triangleq \max_{j \in \mathbf{q}} f^j(u), \tag{4.4b}$$

where $\mathbf{q} \triangleq \{1,2,\ldots,q\}$. We will consider the following two problems:

UP
$$\min_{u \in \mathbf{U}^o} f^0(u), \tag{4.4c}$$

CP
$$\min_{u \in \mathbf{U}_c} \{ f^0(u) \mid \psi(u) \le 0 \}, \tag{4.4d}$$

where

$$\mathbf{U}_c \triangleq \{ u \in L_{\infty,2}^m[0,1] \mid u(t) \in U, \forall t \in [0,1] \}, \tag{4.4d}$$

with $U \subset \mathbb{R}^m$ a compact, convex set contained in the interior of the ball $B(0,\delta\rho_{\max}) \triangleq \{ v \in \mathbb{R}^m \mid \|v\| \le \delta\rho_{\max} \}$; and

$$f^j(u) \triangleq g^j(x^u(1)). \tag{4.4e}$$

Problem (4.4c) can be restated in the canonical form (2.1a), as follows. Let

$$\mathcal{U}_c \triangleq \{ u \in \mathbf{U}_c \mid \psi(u) \le 0 \}, \tag{4.4f}$$

then we can rewrite (4.4c) in the equivalent form

CP
$$\min_{u \in \mathcal{U}_c} f^0(u). \tag{4.4g}$$

Computationally, the control constraint $u \in \mathbf{U}_c$ causes nontrivial complications because it is not differentiable in the pre-Hilbert space $L_{\infty,2}^m[0,1]$, and hence prevents expressing optimality functions in the rather convenient dual form (3.6b).

Since both the functions $g^j(\cdot)$ and the solutions $x^u(\cdot)$ are locally Lipschitz continuously differentiable, the following theorem is deduced from the chain rule and the linearization of the differential equation (4.1) (for a proof see [4]):

Theorem 4.2. Suppose that Assumption 4.1 is satisfied. Then the functions $f^j(\cdot)$, $j = 0, 1, 2, \ldots, q$, defined in (4.4a), have continuous Frechet differentials $Df^j : \mathbf{U}^o \times L_{\infty,2} \to \mathbb{R}^n$ that have the form $Df^j(u, \delta u) = \langle \nabla f^j(u), \delta u \rangle_2$, where the gradients, $\nabla f^j(u) \in L^m_{\infty,2}[0,1]$ are locally Lipschitz continuous on $\mathbf{U}^o$ and are given by

$$\nabla f^j(u)(t) = h_u(x^u(t), u(t))^T p^{j,u}(t), \; t \in [0,1], \tag{4.5a}$$

with $p^{j,u}(t) \in \mathbb{R}^n$ the solution of the adjoint equation

$$\dot{p}(t) = -h_x(x^u(t), u(t))^T p(t), \quad t \in [0,1], \quad p(1) = \nabla g^j(x^u(1)), \tag{4.5b}$$

□

Because there is no satisfactory Maximum Principle for discrete optimal control, the Pontryagin Maximum Principle [24] is not a useful optimality condition in the context of establishing the consistency of discrete approximations. Hence we propose to use the following, rather basic, first order optimality conditions and corresponding optimality functions to define stationary points.

Theorem 4.3. Suppose that Assumption 4.1 satisfied.

(a) Suppose that $\hat{u}$ is optimal for **UP**. Then

$$df^0(\hat{u}, \delta u) \geq 0, \; \forall \delta u \in L^m_{\infty,2}[0,1], \tag{4.6a}$$

where $df^0(\cdot,\cdot)$ denotes the directional derivative.

(b) Let $\theta_{\mathbf{UP}} : \mathbf{U}^o \to \mathbb{R}$ be defined by

$$\theta_{\mathbf{UP}}(u) = -\|\nabla f^0(u)\|_2^2 . \tag{4.6b}$$

Then $\theta_{\mathbf{UP}}(\cdot)$ is continuous in the $L^m_{\infty,2}[0,1]$ topology, and, for any $\hat{u} \in \mathbf{U}^o$ (4.6a) holds if and only if $\theta_{\mathbf{UP}}(\hat{u}) = 0$, i.e., $\theta_{\mathbf{UP}}(\cdot)$ is an optimality function for **UP**.

(c) Let $\gamma > 0$. For any $u \in L^m_{\infty,2}[0,1]$, let

$$\psi(u)_+ \triangleq \max\{0, \psi(u)\}, \tag{4.6c}$$

and for any $u, u' \in L^m_{\infty,2}[0,1]$, let

$$F_{u'}(u) \triangleq \max\{f^0(u) - f^0(u') - \gamma\psi(u')_+, \psi(u) - \psi(u')_+\}. \tag{4.6d}$$

If $\hat{u}$ is a local minimizer for **CP**, then

$$dF_{\hat{u}}(\hat{u}, u - \hat{u}) \geq 0, \ \forall u \in \mathbf{U}_c \ . \tag{4.6e}$$

(d) *Let* $\gamma > 0$, and let $\theta_{\mathbf{CP}} : \mathbf{U}_c \to \mathbb{R}$ be defined by †

$$\theta_{\mathbf{CP}}(u) \triangleq \min_{u+\delta u \in \mathbf{U}_c} \left\{ \tfrac{1}{2}\|\delta u\|_2^2 + \max_{j \in \mathbf{q}} \{ \langle \nabla f^0(u), \delta u \rangle_2 - \gamma\psi(u)_+ , f^j(u) - \psi(u)_+ + \langle \nabla f^j(u), \delta u \rangle_2 \} \right\} , \tag{4.6f}$$

Then *(i)* $\theta_{\mathbf{CP}}(\cdot)$ is negative valued, *(ii)* continuous in the $L^m_{\infty,2}[0,1]$ topology, and, *(iii)* any $\hat{u} \in \mathbf{U}_c$ satisfies (4.6e) if and only if $\theta_{\mathbf{CP}}(\hat{u}) = 0$, i.e., $\theta_{\mathbf{CP}}(\cdot)$ is an optimality function for **CP**.

Proof. Since $df^0(u, \delta u) = \langle \nabla f^0(u), \delta u \rangle_2$, and since $\nabla f^0(\cdot)$ is continuous, parts *(a)* and *(b)* are obvious.

(c) Since any local minimizer of **CP** is a local minimizer for the problem $\min_{u \in \mathbf{U}_c} F_{\hat{u}}(u)$, (4.6e) follows directly.

(d). First, since $\delta u = 0$ is admissible in (4.6c), it is obvious that $\theta_{\mathbf{CP}}(u) \leq 0$ for all $u \in \mathbf{U}_c$. Next we will show that $\theta_{\mathbf{CP}}(\cdot)$ is continuous. Let $\tilde{F} : \mathbf{U}_c \times L^m_{\infty,2}[0,1]c \to \mathbb{R}$ be defined by

$$\tilde{F}_u(\delta u) \triangleq \tfrac{1}{2}\|\delta u\|_2^2 + \max_{j \in \mathbf{q}} \{ \langle \nabla f^0(u), \delta u \rangle_2 - \gamma\psi(u)_+ , f^j(u) - \psi(u)_+ + \langle \nabla f^j(u), \delta u \rangle_2 \} , \tag{4.7a}$$

Then we can rewrite (4.6f) as

$$\theta_{\mathbf{CP}}(u) = \min_{u+\delta u \in \mathbf{U}_c} \tilde{F}_u(\delta u) \ . \tag{4.7b}$$

Note that $\tilde{F}_{u'}(u - u')$ is Lipschitz continuous in $(u', u) \in \mathbf{U}_c \times \mathbf{U}_c$, in the $L^m_{\infty,2}[0,1]$ topology. We will denote the Lipschitz constant by L. Now suppose that $\{ u_i \}_{i=1}^{\infty}$ is a sequence in $\mathbf{U}_c$ that converges to u, in the $L^m_{\infty,2}[0,1]$ topology. Let $u' \in \mathbf{U}_c$ be such that $\theta_{\mathbf{CP}}(u) = \tilde{F}_u(u' - u)$, and let $u'_i \in \mathbf{U}_c$ be such that $\theta_{\mathbf{CP}}(u_i) = \tilde{F}_{u_i}(u'_i \ u_i)$, for all $i \in \mathbb{N}$. Then we must have that

$$\theta_{\mathbf{CP}}(u_i) \leq \tilde{F}_{u_i}(u' - u_i), \ \forall i \in \mathbf{N}, \tag{4.7c}$$

and hence $\overline{\lim}\, \theta_{\mathbf{CP}}(u_i) \leq \lim \tilde{F}_{u_i}(u' - u_i) = \theta_{\mathbf{CP}}(u)$, i.e., $\theta_{\mathbf{CP}}(\cdot)$ is u.s.c. Next, we must have that for all $i \in \mathbb{N}$,

$$\begin{aligned} \theta_{\mathbf{CP}}(u) &\leq \tilde{F}_u(u'_i - u) \\ &= [\tilde{F}_u(u'_i - u) - \tilde{F}_{u_i}(u'_i - u_i)] + \tilde{F}_{u_i}(u'_i - u_i) \end{aligned}$$

† The fact that $\theta_{\mathbf{CP}}(u)$ is well defined follows directly from Corrolary III.20 in Haim Brezis, *Analyse Fonctionelle: Theorie et Applications*, Masson, Paris, 1983.

$$\leq L \| u - u_i \|_2 + \tilde{F}_{u_i}(u'_i u_i) . \tag{4.7d}$$

Hence we conclude that $\theta_{CP}(u) \leq \underline{\lim}\, \theta_{CP}(u_i)$, which shows that $\theta_{CP}(\cdot)$ is l.s.c., and hence continuous.

Next, we will show that $\theta_{CP}(\hat{u}) = 0$ if and only if (4.6e) holds. Since for any $u, \delta u$,

$$\tilde{F}_u(\delta u) \geq \tfrac{1}{2}\|\delta u\|_2^2 + dF_u(u, \delta u), \tag{4.7e}$$

it follows that if $\theta_{CP}(\hat{u}) < 0$, then (4.6e) cannot hold, and hence, by contraposition, if $\hat{u}$ satisfies (4.6e) then we must have that $\theta_{CP}(\hat{u}) = 0$. Now suppose that (4.6) does not hold. Then there must exist a $u \in \mathbf{U}_c$ such that $dF_{\hat{u}}(\hat{u}, u - \hat{u}) < 0$. It is not difficult to deduce that there must exist a $\lambda \in (0, 1]$, such that $\tilde{F}_{\hat{u}}(\lambda(u - \hat{u}) - \hat{u}) < 0$. Hence, again by contraposition, we see that $\theta_{CP}(\hat{u}) = 0$ implies that (4.6e) holds, which concludes our proof that $\theta_{CP}(\cdot)$ is an optimality function. □

The simplest set of consistent approximations to the problems **UP** and **CP** are obtained by integrating the differential equation (4.1) using Euler's forward method. This approach turns out be computationally efficient when the differential equation (4.1) is not stiff. We begin by constructing finite dimensional subspaces of $L^m_{\infty,2}[0, 1]$ on which the precision of Euler's method is easily established. For any integer $N \geq 1$, let $\Gamma(N) \triangleq 2^N$. Then, for any integer $N \geq 1$ and $k = 0, 1, 2, \ldots, \Gamma(N)$, we define $t_{N,k} \triangleq k/\Gamma(N)$, and for $k = 0, 1, 2, \ldots, \Gamma(N) - 1$, we define $\pi_{N,k} : \mathbb{R} \to \mathbb{R}$ by

$$\pi_{N,k}(t) \triangleq \begin{cases} 1 \text{ for all } t \in [t_{N,k}, t_{N,k+1}), \text{ if } k \leq \Gamma(N) - 2, \\ 1 \text{ for all } t \in [t_{N,k}, t_{N,k+1}], \text{ if } k = \Gamma(N) - 1, \\ 0 \text{ otherwise}. \end{cases} \tag{4.8a}$$

Next, for any integer $N \geq 1$, we define the subspace $L^m_N[0, 1] \subset L^m_{\infty,2}[0, 1]$, by

$$L^m_N[0, 1] \triangleq \{ u \in L^m_{\infty,2}[0, 1] \mid u(t) \triangleq \sum_{k=0}^{\Gamma(N)-1} u_k \pi_{N,k}(t) \}, \tag{4.8b}$$

where $\{ u_i \}_{i=0}^{\Gamma(N)-1}$ is a sequence in $\mathbb{R}^m$. Note that the union of the subspaces $L^m_N[0, 1]$ is dense in $L^m_2[0, 1]$. Since the functions $\pi_{N,k}(\cdot)$ are linearly independent, we see that $L^m_N[0, 1]$ is in one-to-one correspondence with the finite dimensional space

$$\bar{L}_N \triangleq \mathbb{R}^{\Gamma(N) \times m}, \tag{4.8c}$$

so that any $u \in L^m_N[0, 1]$, with $u(\cdot) = \sum_{i=0}^{\Gamma(N)-1} u_k \pi_{N,k}(\cdot)$, corresponds to $\bar{u} \in \bar{L}_N$, with $\bar{u} = (u_0, u_1, \ldots, u_{\Gamma(N)-1})$. Thus, for $N = 1, 2, 3, \ldots$, we can define the linear, invertible map $W_N : L^m_N[0, 1] \to \bar{L}_N$ by

$$W_N\Big(\sum_{k=0}^{\Gamma(N)-1} u_k \pi_{N,k}(\cdot)\Big) \triangleq (u_0, u_1, \ldots, u_{\Gamma(N)-1}) . \tag{4.8d}$$

Now, for any $u \in L_N^m[0,1]$,

$$\|u\|_2 = \frac{1}{\Gamma(N)}\Big(\sum_{k=0}^{\Gamma(N)-1} \|u_k\|^2\Big)^{1/2} . \tag{4.8e}$$

Hence, to retain a proper scaling balance between the continuous and discrete time problems, we define the scalar product $\langle \cdot, \cdot \rangle_{\bar{L}_N}$ and norm $\|\cdot\|_{\bar{L}_N}$, on $\bar{L}_N$, by

$$\langle \bar{u}, \bar{u}' \rangle_{\bar{L}_N} \triangleq \frac{1}{\Gamma(N)} \langle \bar{u}, \bar{u}' \rangle, \tag{4.8f}$$

$$\|\bar{u}\|_{\bar{L}_N} = \frac{1}{\Gamma(N)}\Big(\sum_{k=0}^{\Gamma(N)-1} \|u_k\|^2\Big)^{1/2}, \tag{4.8g}$$

where the scalar product, $\langle \cdot, \cdot \rangle$, in (4.8f), is the usual Euclidean scalar product, and the norm $\|\cdot\|$, in (4.8g), is the usual Euclidean norm. Consequently, if u, $u' \in L_N^m[0,1]$ and $\bar{u} = W_N u$, $\bar{u}' = W_N u'$, then we always have that $\langle \bar{u}, \bar{u}' \rangle_{\bar{L}_N} = \langle u, u' \rangle_2$ and $\|u\|_2 = \|\bar{u}\|_{\bar{L}_N}$.

In addition, we will use the notation

$$\mathbf{U}_N^o \triangleq \mathbf{U}^o \cap L_N^m[0,1], \quad \mathbf{U}_{c,N} \triangleq \mathbf{U}_c \cap L_N^m[0,1] . \tag{4.8h}$$

Clearly, whenever $N'' > N'$, we must have that $\mathbf{U}_N^o \subset \mathbf{U}_{N''}^o$, and $\mathbf{U}_{c,N'} \subset \mathbf{U}_{c,N''}$.

Next, given any $u \in L_N^m[0,1]$, where $u(\cdot) = \sum_{k=0}^{\Gamma(N)-1} u_k \pi_{N,k}(\cdot)$, we replace the continuous dynamics (4.1) by the discrete dynamics resulting from the use of the Euler integration formula:

$$\bar{x}(t_{N,k+1}) = \bar{x}(t_{N,k}) + \Delta(N) h(\bar{x}(t_{N,k}), u_k), \; k = 0, 1, \ldots, \Gamma(N) - 1, \; \bar{x}(0) = \xi, \tag{4.9a}$$

where

$$\Delta(N) \triangleq \frac{1}{\Gamma(N)}, \tag{4.9b}$$

so that $t_{N,k} = k\Delta(N)$. Clearly, (4.9a) has a unique solution for any $\bar{u} \in \bar{L}_N$. We will denote the solution of (4.9), corresponding to any $\bar{u} = W_N u$, with $u \in L_N^m[0,1]$, by $\{\bar{x}_N^u(t_{N,k})\}_{k=0}^{\Gamma(N)}$. We associate with the sequence $\{\bar{x}_N^u(t_{N,k})\}_{k=0}^{\Gamma(N)}$, of vectors in $\mathbb{R}^n$ the time function

$$\bar{x}_N^u(t) = \sum_{k=0}^{\Gamma(N)-1} \bar{x}_N^u(t_{N,k}) \pi_{N,k}(t) . \tag{4.9c}$$

Making use of Theorem 3.1.6 in [6] one can show that exists a constant $K_x < \infty$ such that

$$\|\bar{x}_N^u(t) - x^u(t)\| \le K_x \Delta(N), \; \forall t \in [0,1] . \tag{4.9d}$$

Also, it should be obvious that the solution $\bar{x}_N^u(t_{N,k})$ is continuously differentiable in u.

Next, for $N = 1, 2, 3, \ldots$, we define the functions $f_N^j : L_N^m[0,1] \to \mathbb{R}$, $j = 0, 1, \ldots, q$, and $\psi_N : L_N^m[0,1] \to \mathbb{R}$ by

$$f_N^j(u) \triangleq g^j(\bar{x}_N^u(1)), \quad \psi_N(u) = \max_{j \in \mathbf{q}} f_N^j(u). \tag{4.10a}$$

Then we define the approximating problems as follows:

$\mathbf{UP}_N$
$$\min_{u \in \mathbf{U}_N^\delta} f_N^0(u), \tag{4.10b}$$

$\mathbf{CP}_N$
$$\min_{u \in \mathbf{U}_{c,N}} \{ f_N^0(u) \mid \psi_N(u) \le 0q \}. \tag{4.10c}$$

If we now define

$$\mathcal{U}_{c,N} \triangleq \{ u \in \mathbf{U}_{c,N} \mid \psi_N(u) \le 0 \}, \tag{4.10d}$$

then we can transcribe (4.10c) into the canonical form (2.1b), as follows:

$\mathbf{CP}_N$
$$\min_{u \in \mathcal{U}_{c,N}} f_N^0(u). \tag{4.10e}$$

As in the continuous case, it follows from the chain rule that the gradients $\nabla f_N^j(u)(\cdot) \in L_N^m[0,1]$, $j = 0, 1, \ldots, q$, exist and are locally Lipschitz continuous, uniformly in $N \in \mathbb{N}$. They can be expressed as follows:

$$\nabla f_N^j(u)(t) = \sum_{k=0}^{\Gamma(N)-1} h_u(\bar{x}_N^u(t_{N,k}), u_k)^T \bar{p}^{u,j}(t_{N,k+1}) \pi_{N,k}(t), \quad t \in [0,1], \tag{4.10f}$$

where, for $k = 0, 1, \ldots, N$, $\bar{p}^{u,j}(t_{N,k})$ is determined by the adjoint equation

$$\bar{p}(t_{N,k}) - \bar{p}(t_{N,k+1}) = \Delta(N) h_x(\bar{x}_N^u(t_{N,k}), u_k)^T \bar{p}(t_{N,k+1}), k = 0, 1, \ldots, \Gamma(N) - 1, \tag{4.10g}$$

$$\bar{p}(1) = \nabla g^j(\bar{x}_N^u(1)). \tag{4.10h}$$

Let

$$\bar{p}_N^{u,j}(t) = \sum_{k=0}^{\Gamma(N)-1} \bar{p}_N^{u,j}(t_{N,k}) \pi_{N,k}(t). \tag{4.10i}$$

Making use of Theorem 3.1.6 in [6], one can show that there exists a $K_p < \infty$ such that

$$\| \bar{p}_N^{u,j}(t) - p^{u,j}(t) \| \le K_p \Delta(N). \tag{4.10j}$$

It should be clear that the following theorem is just a special case of Theorem 4.3.

Theorem 4.4. Suppose that Assumption 4.1 is satisfied.

(a) If $\hat{u}_N$ is a local minimizer for $\mathbf{UP}_N$, then

$$df_N^0(\hat{u}_N, \delta u) \geq 0, \ \ \forall \delta u \in L_N^m[0,1], \tag{4.11a}$$

where $df_N^0(\cdot,\cdot)$ denotes the directional derivative of $f_N^0(\cdot)$.

(b) For $N = 1, 2, \ldots$, let $\theta_{\mathbf{UP}_N} : \mathbf{U}_N^o \to \mathbb{R}$ be defined by

$$\theta_{\mathbf{UP}_N}(u) \triangleq -\|\nabla f_N^0(u)\|_2^2 . \tag{4.11b}$$

Then $\theta_{\mathbf{UP}_N}(\cdot)$ is continuous in the $L_{\infty,2}^m[0,1]$ topology, and, for any $\hat{u} \in \mathbf{U}^o$ (4.11a) holds if and only if $\theta_{\mathbf{UP}_N}(\hat{u}) = 0$, i.e., $\theta_{\mathbf{UP}_N}(\cdot)$ is an optimality function for $\mathbf{UP}_N$.

(c) Let $\gamma > 0$. For any $u \in L_{\infty,2}^m[0,1]$, let

$$\psi_N(u)_+ \triangleq \max\{0, \psi_N(u)\}, \tag{4.11c}$$

and for any $u, u' \in L_N^m[0,1]$, let

$$F_{N,u'}(u) \triangleq \max\{f_N^0(u) - f_N^0(u') - \gamma\psi_N(u')_+, \psi_N(u) - \psi_N(u')_+\}. \tag{4.11d}$$

If $\hat{u}_N$ is a local minimizer for $\mathbf{CP}_N$, then

$$dF_{N,\hat{u}}(\hat{u}, u - \hat{u}) \geq 0, \ \ \forall u \in \mathbf{U}_{c,N}, \tag{4.11e}$$

where $dF_{N,\hat{u}}(\cdot,\cdot)$ denotes the directional derivative of $F_{N,\hat{u}}(\cdot)$.

(d) Let $\gamma > 0$, and, for any $N \in \mathbb{N}$, let $\theta_{\mathbf{CP}_N} : \mathbf{U}_{c,N} \to \mathbb{R}$ be defined by

$$\theta_{\mathbf{CP}_N}(u) \triangleq \min_{u+\delta u \in \mathbf{U}_{c,N}} \Big\{ \tfrac{1}{2}\|\delta u\|_2^2$$

$$+ \max_{j \in \mathbf{q}} \{ \langle \nabla f_N^0(u), \delta u \rangle_2 - \gamma\psi_N(u)_+, f_N^j(u) - \psi_N(u)_+ + \langle \nabla f_N^j(u), \delta u \rangle_2 \} \Big\}. \tag{4.11f}$$

Then *(i)* $\theta_{\mathbf{CP}_N}(\cdot)$ is negative valued, *(ii)* continuous in the $L_N^m[0,1]$ topology, and, *(iii)* any $\hat{u}_N \in \mathbf{U}_{c,N}$ satisfies (4.11e) if and only if $\theta_{\mathbf{CP}_N}(\hat{u}_N) = 0$, i.e., $\theta_{\mathbf{CP}_N}(\cdot)$ is an optimality function for $\mathbf{CP}_N$. □

Next we obtain the following approximation results.

Lemma 4.5. Suppose that Assumption 4.1 is satisfied. Then there exists a constant $K_f < \infty$ such that for all $u \in \mathbf{U}_N^o$, and $N \in \mathbf{N}$ (with $N \geq 1$),

$$|f_N^j(u) - f^j(u)| \leq K_f \Delta(N), \ \ j = 0, 1, \ldots, q, \tag{4.12a}$$

$$|\psi_N(u) - \psi(u)| \le K_f \Delta(N), \tag{4.12b}$$

$$\|\nabla f_N^j(u) - \nabla f^j(u)\|_2 \le K_f \Delta(N), \quad j = 0, 1, \ldots, q \ . \tag{4.12c}$$

Proof. The existence of a $K_f < \infty$ such that (4.12a) holds, follows directly from the Lipschitz continuity of the $g^j(\cdot)$, in (4.4a), and (4.9d). Hence,

$$\psi_N(u) \le \max_{j \in \mathbf{q}} f^j(u) + K_f \Delta(N) = \psi(u) + K_f \Delta(N) \ . \tag{4.12d}$$

Reversing the roles of $\psi_N(u)$ and $\psi(u)$ in (4.13), we obtain (4.12b). Next, the existence of a $K_f < \infty$ (possibly larger than needed for (4.12a)), such that (4.12c) holds, follows from (4.10j) and the formulae for $\nabla f_N^j(u)$ and $\nabla f^j(u)$. □

In proving consistency, we will need two assumptions. The first is that δ and $\rho_{\max}$ have been chosen to be sufficiently large to ensure that the function $f^0(u)$ has no minimizers on the boundary of the set $\mathbf{U}^o$. The second consists of a constraint qualification which, among other things, rules out conversion of equality constraints into inequality constraints, and is closely related to the Mangasarian-Fromowitz constraint qualification [17]:

Assumption 4.6.

(a) Let $\mathcal{U}$ denote the closure of $\mathbf{U}_o$. We will assume that all the global minimizers of the problem

$\overline{\mathbf{UP}}$
$$\min_{u \in \mathcal{U}} f^0(u), \tag{13}$$

are in $\mathbf{U}^o$, i.e., that the problems **UP** and $\overline{\mathbf{UP}}$ are equivalent.

(b) For every $u \in \mathbf{U}_c$ such that $\psi(u) \le 0$, there exists a sequence $\{ u_N \}_{N=1}^{\infty}$, such that for all N, $u_N \in \mathbf{U}_{c,N}$, $\psi(u_N) < 0$, and $u_N \to u$ as $N \to \infty$. □

Theorem 4.7. Suppose that Assumptions 4.1 and 4.6 are satisfied. Then for $N = 1, 2, 3, \ldots$, the problems $\mathbf{UP}_N$ and $\mathbf{CP}_N$ converge epigraphically to the problems $\overline{\mathbf{UP}}$ and **CP**, respectively, in the $L_{\infty,2}^m[0, 1]$ topology.

Proof. We begin with the problems $\mathbf{UP}_N$. Since the union of the subspaces $L_N^m[0, 1]$ is dense in $L_{\infty,2}[0, 1]$, it is clear that for any $u \in \mathbf{U}^o$ there exists a sequence $\{ u_N \}_{N=1}^{\infty}$, with $u_N \in \mathbf{U}_N^o$, such that $u_N \to u$ as $N \to \infty$. It now follows from (4.12a) that $\lim f_N^0(u_N) = f^0(u)$, which shows that part *(a)* of Definition 2.1 is satisfied. Clearly, if $\{ u_N \}_{N=1}^{\infty}$, with $u_N \in \mathbf{U}_N^o$, is such that $u_N \to u$ as $N \to \infty$, then $u \in \overline{\mathbf{U}}^o$ and, again by (4.12a), $\lim f_N^0(u_N) = f^0(u)$, which shows that part *(b)* of Definition 2.1 is satisfied.

Next consider the problems $\mathbf{CP}_N$. We can infer from Assumption 4.6 that for any $u \in \mathcal{U}_c$

there exists a sequence $\{ u_N \}_{N=1}^{\infty}$, with $u_N \in \mathcal{U}_{c,N}$, such that $u_N \to u$ as $N \to \infty$, and by (4.12a), $\lim f_N^0(u_N) = f^0(u)$, which shows that part *(a)* of Definition 2.1 is satisfied. Clearly, if $\{ u_N \}_{N=1}^{\infty}$, with $u_N \in \mathcal{U}_{c,N}$, is such that $u_N \to u$ as $N \to \infty$, then $u \in \mathcal{U}_c$ and, again by (4.12a), $\lim f_N^0(u_N) = f^0(u)$, which shows that part *(b)* of Definition 2.1 is satisfied. Hence our proof is complete. □

Theorem 4.8.

(a) Suppose that $\{ u_N \}_{N=1}^{\infty}$ is such that for all $N \in \mathbb{N}$, $u_N \in \mathbf{U}^o$, and $u_N \to \hat{u}$, as $N \to \infty$, then $\theta_{\mathrm{UP}_N}(u_N) \to \theta_{\mathrm{UP}}(\hat{u})$, as $N \to \infty$.

(b) Suppose that $\{ u_N \}_{N=1}^{\infty}$ is such that for all $N \in \mathbb{N}$, $u_N \in \mathbf{U}_{c,N}$, and $u_N \to \hat{u}$, as $N \to \infty$, then $\theta_{\mathrm{CP}_N}(u_N) \to \theta_{\mathrm{CP}}(\hat{u})$, as $N \to \infty$.

Proof.

(a) This part follows directly from (4.12c).

(b) For any $N \in \mathbb{N}$ and u, $u' \in \mathbf{U}_{c,N}$, let

$$\tilde{F}_{N,u}(u') \triangleq \tfrac{1}{2}\|u' - u\|_2^2$$

$$+ \max_{j \in \mathbf{q}} \{ \langle \nabla f_N^0(u), u' - u \rangle_2 - \gamma\psi_N(u)_+, f_N^j(u) - \psi_N(u)_+ + \langle \nabla f_N^j(u), u' - u \rangle_2 \}, \qquad (4.14a)$$

where γ is as in (4.6d,f) and (4.11d,f). Without loss of generality, we will assume that $\gamma \geq 1$. Now suppose that the sequence $\{ u_N \}_{N=1}^{\infty}$ is such that for all $N \in \mathbb{N}$, $u_N \in \mathbf{U}_{c,N}$, and $u_N \to u$, as $N \to \infty$. For all N, let $u'_N \in \mathbf{U}_{c,N}$ be such that $\theta_{\mathrm{CP}_N}(u_N) = \tilde{F}_{N,u_N}(u'_N)$. Then $\theta_{\mathrm{CP}}(u_N) \leq \tilde{F}_{u_N}(u'_N)$, where $\tilde{F}_{u'}(u)$ is defined in (4.7a). Now, *(i)* because of (4.12b) $|\psi_N(u_N)_+ - \psi(u_N)_+| \leq K\Delta(N)$ for all N, and *(ii)* because $\mathbf{U}_c$ is bounded in $L_\infty^m[0,1]$, there exists a $b < \infty$, such that $\|u'_N - u_N\|_2 \leq b$ for all N. Hence making use of (4.12a,b,c) and the fact that $\psi_N(u_N)_+ = 0$, we find that

$$\theta_{\mathrm{CP}}(u_N) \leq \tilde{F}_{u_N}(u'_N)$$

$$= \tfrac{1}{2}\|u'_N - u_N\|_2^2$$

$$+ \max_{j \in \mathbf{q}} \{ \langle \nabla f^0(u), u'_N - u_N \rangle_2 - \gamma\psi(u_N)_+, f^j(u_N) - \psi(u_N)_+ + \langle \nabla f^j(u_N), u'_N - u_N \rangle_2 \},$$

$$= \tfrac{1}{2}\|u'_N - u_N\|_2^2$$

$$+ \max_{j \in \mathbf{q}} \{ \langle \nabla f_N^0(u), u'_N - u_N \rangle_2 + \langle \nabla f^0(u) - \nabla f_N^0(u), u'_N - u_N \rangle_2 - \gamma\psi(u_N)_+, f_N^j(u_N)$$

$$+ [f^j(u_N) - f_N^j(u_N)] - \psi(u_N)_+ + \langle \nabla f_N^j(u_N), u'_N - u_N \rangle_2 + \langle \nabla f^j(u_N) - \nabla f_N^j(u_N), u'_N - u_N \rangle_2 \}$$

$$\leq \tilde{F}_{N,u_N}(u'_N) + K(1+\gamma+b)\Delta(N) . \tag{4.14b}$$

Hence, since $\theta_{\mathbf{CP}}(\cdot)$ is continuous, we conclude that

$$\theta_{\mathbf{CP}}(\hat{u}) = \underline{\lim}\, \theta(u_N) \leq \underline{\lim}\, \theta_{\mathbf{CP}_N}(u_N) . \tag{4.14c}$$

Now, let $\hat{u}' \in \mathbf{U}_c$ be such that $\theta_{\mathbf{CP}}(\hat{u}) = \tilde{F}_{\hat{u}}(\hat{u}')$, and let $\hat{u}'_N \in \mathbf{U}_{c,N}$ be such that $\hat{u}'_N \to \hat{u}'$, as $N \to \infty$. Then for every N, $\theta_{\mathbf{CP}_N}(u_N) \leq F_{N,u_N}(\hat{u}'_N)$. Proceeding as for (4.14b), we conclude that

$$\theta_{\mathbf{CP}_N}(u_N) \leq \tilde{F}_{u_N}(\hat{u}'_N) + K(1+\gamma+b)\Delta(N) . \tag{4.14d}$$

Consequently, since $\tilde{F}_{u'}(u)$ is continuous in (u', u),

$$\overline{\lim}\, \theta_{\mathbf{CP}_N}(u_N) \leq \overline{\lim}\, [\tilde{F}_{u_N}(\hat{u}'_N) + K(1+\gamma+b)\Delta(N)] = \theta(\hat{u}) . \tag{4.14e}$$

Combining (4.14c) and (4.14d) we conclude that $\theta_N(u_N) \to \theta(\hat{u})$, which completes our proof. □

At this point, the following result is obvious:

Corollary 4.9.

(a) The pairs in the the sequence $\{ (\mathbf{UP}_N, \theta_{\mathbf{UP}_N}) \}_{N=1}^{\infty}$ are consistent approximations to $(\mathbf{UP}, \theta_{\mathbf{UP}})$.

(b) The pairs in the the sequence $\{ (\mathbf{CP}_N, \theta_{\mathbf{CP}_N}) \}_{N=1}^{\infty}$ are consistent approximations to $(\mathbf{CP}, \theta_{\mathbf{UP}})$. □

5. MASTER ALGORITHM MODELS FOR USE WITH CONSISTENT APPROXIMATIONS

Now that we have seen that we can construct consistent approximations for both semi-infinite optimization and optimal control problems, we need to address the question of how such approximations are to be used in the construction of an approximate solution of the original problem. We recall that the experience with penalty functions in nonlinear programming indicates that it is a bad idea to simply select a large penalty and solve the resulting unconstrained problem. The reason for this is that large penalties produce serious ill-conditioning. Hence the commonly used strategy is to solve approximately a sequence of progressively more severely penalized problems, which produces starting points for the successive problems from which Newton's method converges quadratically, and hence overcomes the ill-conditioning. While increasing discretization of semi-infinite optimization and optimal control problems does not lead to ill-conditioning, it does increase the computational complexity of the resulting problems. Referring to the literature (see, e.g. [9, 13, 14, 22]) we find reports that in the case of semi-infinite optimization and optimal control problems, there is also a

considerable benefit to be obtained from increasing the discretization in a preplanned manner. We will now describe two strategies, in the form of algorithm models, for increasing discretization in solving semi-infinite optimization and optimal control problems via consistent approximations.

The constraint set X in problem $\mathbf{P}$ can have a variety of characterizations. We will deal with only two: the first is when $X = \mathcal{X}$, where $\mathcal{X}$ is a "simple" convex set, as in minimax problems on $\mathbb{R}^n$ and control problems with or without control constraints, but no trajectory constraints, while the second is more complex, and has the form $X = \{ x \in \mathcal{X} \mid \psi(x) \leq 0 \}$, where $\mathcal{X}$ is a "simple" convex set and $\psi(\cdot)$ is a continuous function. To make this distinction explicit, we define the two cases as follows:

$$\mathbf{P}_u \qquad \min_{x \in \mathcal{X}} f(x) . \tag{5.1a}$$

$$\mathbf{P}_c \qquad \min_{x \in \mathcal{X}} \{ f(x) \mid \psi(x) \leq 0 \} . \tag{5.1b}$$

Similarly, for $N = 1, 2, 3, \ldots$, the approximating problems $\mathbf{P}_N$ acquire the following form

$$\mathbf{P}_{u,N} \qquad \min_{x \in \mathcal{X}_N} f_N(x) . \tag{5.2a}$$

$$\mathbf{P}_{c,N} \qquad \min_{x \in \mathcal{X}_N} \{ f_N(x) \mid \psi_N(x) \leq 0 \} . \tag{5.2b}$$

In view of the results in the preceding two sections, we make the following assumption.

Assumption 5.1.

(i) The functions $f, \psi : \mathcal{B} \to \mathbb{R}$ as well as the functions $f_N, \psi_N : \mathcal{B}_N \to \mathbb{R}$, $N = 1, 2, 3, \ldots$, are continuous.

(ii) The set $\mathcal{X}$ is either a convex, closed subset of $\mathcal{B}$, or $\mathcal{X} = \mathcal{B}$, and, for $N = 1, 2, 3, \ldots$, $\mathcal{X}_N = \mathcal{X} \cap \mathcal{B}_N$.

(iii) There exist continuous optimality functions $\theta_u : \mathcal{X} \to \mathbb{R}$, for $\mathbf{P}_u$, $\theta_c : \mathcal{X} \to \mathbb{R}$, for $\mathbf{P}_c$, as well as continuous optimality functions $\theta_{u,N} : \mathcal{X}_N \to \mathbb{R}$, for $\mathbf{P}_{u,N}$ and $\theta_{c,N} : \mathcal{X}_N \to \mathbb{R}$, for $\mathbf{P}_{c,N}$, $N = 1, 2, 3, \ldots$.

(iv) There exist a strictly positive valued, strictly monotone decreasing function $\Delta : \mathbb{N} \to \mathbb{R}$, such that $\Delta(N) \to 0$ as $N \to \infty$, and constants $K \in (0, \infty)$, $N_o \in \mathbb{N}$, such that for all $N \geq N_o$, and all $x \in \mathcal{X}_N$, (or at least for all x in a sufficiently large, bounded open subset of $\mathcal{X}_N$),

$$| f_N(x) - f(x) | \leq K \Delta(N), \tag{5.3a}$$

$$|\psi_N(x) - \psi(x)| \le K\Delta(N). \tag{5.3b}$$

(v) If $\{x_N\}_{N=1}^{\infty}$ is such that $x_N \in \mathcal{X}_N$ for all N, and $x_N \to \hat{x}$ as $N \to \infty$, then $\theta_{u,N}(x_N) \to \theta_u(\hat{x})$, and $\theta_{c,N}(x_N) \to \theta_c(\hat{x})$, as $N \to \infty$.

(vi) For every $x \in \mathcal{X}$ such that $\psi(x) \le 0$, there exists a sequence $\{x_N\}_{N=1}^{\infty}$ such that for all N, $x_N \in \mathcal{X}_N$, $\psi_N(x_N) \le 0$ and $x_N \to x$ as $N \to \infty$. □

It should be obvious that Assumption 5.1 ensures that the pairs $(\mathbf{P}_{u,N}, \theta_{u,N})$, $N = 1,2,3,\ldots$, are weakly consistent approximations to $(\mathbf{P}_u, \theta_u)$, and similarly, that the pairs $(\mathbf{P}_{c,N}, \theta_{c,N})$, $N = 1,2,3,\ldots$, are weakly consistent approximations to $(\mathbf{P}_c, \theta_c)$. Hence the following theorem is a direct consequence of Theorem 2.4 and Assumption 5.1.

Theorem 5.2. Suppose that Assumption 5.1 is satisfied.

(i) If $\{x_N\}_{N=1}^{\infty}$ is a sequence of global minimizers of $\mathbf{P}_{u,N}$ $(\mathbf{P}_{c,N})$ such that $x_N \to \hat{x}$ as $N \to \infty$, then $\hat{x}$ is a global minimizer of $\mathbf{P}_u$ $(\mathbf{P}_c)$.

(ii) If $\{x_N\}_{N=1}^{\infty}$ is a sequence of strict local minimizers of $\mathbf{P}_{u,N}$ $(\mathbf{P}_{c,N})$, with radius of attraction $\rho_N \ge 0$, such that $x_N \to \hat{x}$ as $N \to \infty$, and there exists an infinite subset $K \subset \mathbb{N}$, such that $\rho_N \ge \rho > 0$, for all $N \in K$, then $\hat{x}$ is a local minimizer of $\mathbf{P}_u$ $(\mathbf{P}_c)$.

(iii) If $\{x_N\}_{N=1}^{\infty}$ is a sequence of local minimizers of $\mathbf{P}_{u,N}$ $(\mathbf{P}_{c,N})$, such that $x_N \to \hat{x}$ as $N \to \infty$, then $\theta_u(\hat{x}) = 0$ $(\theta_c(\hat{x}) = 0)$. □

We will now describe our first strategy for increasing discretization in solving "conceptual problems" such as $\mathbf{P}_u$ and $\mathbf{P}_c$ via consistent approximations satisfying the conditions of Assumption 5.1. This strategy has the advantage that it can be used with a very broad class of nonlinear programming algorithms. Its disadvantage is that convergence results can be stated only about rather sparse, "filtered" subsequences of all the points constructed. We will present our strategies for solving the problems $\mathbf{P}_u$ and $\mathbf{P}_c$ in the form of algorithm models in which we will define the "outer" iterations. The "inner" iterations are defined by user supplied iteration maps $A_{u,N}, A_{c,N} : \mathcal{X}_N \to 2^{\mathcal{X}_N}$, that define one iteration of a nonlinear programming algorithm that can be used for solving the problems $\mathbf{P}_{u,N}$ and $\mathbf{P}_{c,N}$. We begin with the unconstrained problem $\mathbf{P}_u$.

Master Algorithm Model 5.3.

Data. $N_0 \in \mathbb{N}, x_0 \in \mathcal{X}_{N_0}$.

Step 0. Set $i = 0$, $N = N_0$.

Step 1. Compute a $x_{i+1} \in A_{u,N}(x_i)$.

Step 2. If $\theta_{u,N}(x_{i+1}) \geq -1/N$, set $x^*_N = x_{i+1}$, and replace N by $N+1$.

Step 3. Replace i by $i+1$ and go to Step 1. □

In view of Assumption 5.1, the following result is obvious:

Theorem 5.4. Suppose that *(a)* Assumption 5.1 is satisfied, and *(b)* that every accumulation point $\hat{x}$ of a sequence $\{x_i\}_{i=0}^{\infty}$, constructed according to the rule $x_{i+1} \in A_{u,N}(x_i)$, satisfies $\theta_{u,N}(\hat{x}) = 0$. Consider the the sequences $\{x_i\}$ and $\{x^*_N\}$ constructed by Algorithm Model 5.3.

(i) If the sequence $\{x^*_N\}$ is finite, then the sequence $\{x_i\}$ has no accumulation points.

(ii) If the sequence $\{x^*_N\}$ is infinite, then every accumulation point $\hat{x}$ of $\{x^*_N\}$, satisfies $\theta_u(\hat{x}) = 0$. □

For the constrained problem $\mathbf{P}_c$ we modify the above as follows:

Master Algorithm Model 5.5.

Data. $N_0 \in \mathbb{N}, x_0 \in \mathcal{X}_{N_0}$.

Step 0. Set $i = 0$, $N = N_0$.

Step 1. Compute a $x_{i+1} \in A_{c,N}(x_i)$.

Step 2. If $\theta_{c,N}(x_{i+1}) \geq -1/N$, and $\psi(x_{i+1}) \leq 1/N$, set $x^*_N = x_{i+1}$, and replace N by $N+1$.

Step 3. Replace i by $i+1$ and go to Step 1. □

Again because of Assumption 5.1, the following result is obvious:

Theorem 5.6. Suppose that *(a)* Assumption 5.1 is satisfied, and *(b)* that for every $N \geq N_o$, every accumulation point $\hat{x}$ of a sequence $\{x_i\}_{i=0}^{\infty}$, constructed according to the rule $x_{i+1} \in A_{c,N}(x_i)$, satisfies $\theta_{u,N}(\hat{x}) = 0$, and $\psi(\hat{x}) \leq 0$. Consider the the sequences $\{x_i\}$ and $\{x^*_N\}$ constructed by Master Algorithm Model 5.6.

(i) If the sequence $\{x^*_N\}$ is finite, then the sequence $\{x_i\}$ has no accumulation points.

(ii) If the sequence $\{x^*_N\}$ is infinite, then every accumulation point $\hat{x}$ of $\{x^*_N\}$, satisfies $\theta_u(\hat{x}) = 0$ and $\psi(\hat{x}) \leq 0$. □

We now turn to our alternative approach, which we believe to be computationally more efficient, and which can be used with almost all unconstrained nonlinear programming algorithms. However, for constrained problems, only the unified method of feasible directions, in [21] has so far been shown to be compatible with our alternative approach. Again we begin with the unconstrained

problem P_u. For this problem we require that the nonlinear algorithms used for solving the problems $\mathbf{P}_{u,N}$ satisfy the following *monotone uniform descent* condition:

Assumption 5.7. For every $x \in \mathcal{X}$, such that $\theta_u(x) < 0$, there exist $\rho_x > 0$, $N_x \in \mathbb{N}$, and $\delta_x < 0$ such that

$$f_N(x'') - f_N(x') \leq \delta_x, \tag{5.4}$$

for all $x' \in B(x, \rho_x) \cap \mathcal{X}_N$, for all $x'' \in A_{u,N}(x')$, and for all $N \geq N_x$. □

Referring to Theorem 1.3.10 in [18], we find that Assumption 5.7 is a generalization of the assumption in the following theorem.

Theorem 5.8. Suppose that Assumption 5.1 is satisfied. Let N be given and suppose that $\{x_i\}_{i=0}^{\infty}$ is a sequence in $\mathcal{X}_N$ constructed using the recursion $x_{i+1} \in A_{u,N}(x_i)$, $i \in \mathbb{N}$, in solving $\mathbf{P}_{u,N}$. If for every $x \in \mathcal{X}_N$, such that $\theta_{u,N}(x) < 0$, there exist $\rho_x > 0$, $\delta_x < 0$ such that

$$f_N(x'') - f_N(x') \leq \delta_x, \tag{5.5}$$

for all $x' \in B(x, \rho_x) \cap \mathcal{X}_N$, for all $x'' \in A_{u,N}(x')$, then every accumulation point $\hat{x}_N$, of $\{x_i\}_{i=0}^{\infty}$, satisfies $\theta_{u,N}(\hat{x}_N) = 0$. □

The assumptions of Theorem 5.8 are satisfied by most unconstrained optimization algorithms, including the Armijo gradient method [2, 18], the Polak-Ribière method of conjugate directions [18], Newton's method [12, 18], the BFGS method with back-stepping step-size rule [5], and the Pshenichnyi - Pironneau - Polak minimax algorithm [19, 20, 25]; however, there is no proof that the Fletcher - Reeves method of conjugate directions satisfies these assumptions. Thus, to show that Assumption 5.7 is satisfied, one only needs to show that one can find a ρ_x and a δ_x that are the same for all $N \geq N_x$. This is relatively easy to show both for semi-infinite optimization problems and for optimal control problems.

Now consider the following master algorithm for solving $\mathbf{P}_u$, which uses the strictly monotonically decreasing function $\Delta : \mathbb{N} \to \mathbb{R}$ introduced in Assumption 5.1.

Master Algorithm Model 5.9.

Parameter: $\beta \in (0, 1)$.

Data. $N_{-1} \in \mathbb{N}$, and $x_0 \in \mathcal{X}_{N_{-1}}$

Step 0. set $i = 0$.

Step 1. compute N_i and x_{i+1} such that $N_i \geq N_{i-1}$, $x_{i+1} \in A_{u,N_i}(x_i)$ and

$$f_{N_i}(x_{i+1}) - f_{N_i}(x_i) \le -\Delta(N_i)^\beta . \tag{5.6}$$

Step 2. Replace i by $i+1$ and go to Step 1. □

Lemma 5.10. Suppose that Assumption 5.1 is satisfied, and that Master Algorithm Model 5.9 has constructed an infinite sequence $\{x_i\}_{i=0}^{\infty}$ that has an accumulation point $\hat{x}$. Then the accompanying sequence $\{N_i\}_{i=0}^{\infty}$ is such that $N_i \to \infty$ as $i \to \infty$.

Proof. For the sake of contradiction, suppose that the monotone increasing sequence $\{N_i\}_{i=0}^{\infty}$ is bounded. Then there exists an $i_0 \in \mathbb{N}$, such that $N_i = N_{i_0} \triangleq N^* < \infty$ for all $i \ge i_0$. Then, by the test (5.6), for all $i \ge i_0$,

$$f_{N^*}(x_{i+1}) - f_{N^*}(x_i) \le -\Delta(N^*)^\beta, \tag{5.7}$$

which implies that $f_{N^*}(x_i) \to -\infty$, as $i \to \infty$. However, since $f_{N^*}(\cdot)$ is continuous and since by assumption, $x_i \to^I \hat{x}$ as $i \to \infty$, for some infinite subset $I \subset \mathbb{N}$, $f_{N^*}(x_i) \to^I f_{N^*}(\hat{x})$ as $i \to \infty$, which is a contradiction. Hence we must have that $N_i \to \infty$ as $i \to \infty$. □

Theorem 5.11. Suppose that Assumptions 5.1 and 5.7 are satisfied. If $\{x_i\}_{i=0}^{\infty}$ is a sequence constructed by the Master Algorithm Model 5.9, then every accumulation point $\hat{x}$ of $\{x_i\}_{i=0}^{\infty}$ satisfies $\theta_u(\hat{x}) = 0$.

Proof. Suppose that $x_i \to^I \hat{x}$ as $i \to \infty$, for some infinite subset $I \subset \mathbb{N}$. For the sake of contradiction, suppose that $\theta_u(\hat{x}) < 0$. Then, by Assumption 5.7, there exist $\delta_{\hat{x}} < 0$, $N_{\hat{x}} \in \mathbb{N}$, and $\rho_{\hat{x}} > 0$ such that

$$f_N(x'') - f_N(x') \le \delta_{\hat{x}}, \tag{5.8a}$$

for all $x' \in B(\hat{x}, \rho_{\hat{x}}) \cap \mathcal{X}_N$, for all $x'' \in A_{u,N}(x')$, and for all $N \ge N_{\hat{x}}$.

Since $x_i \to^I \hat{x}$ as $i \to \infty$, it follows from Lemma 5.10 that $N_i \to \infty$, as $i \to \infty$ and hence that $\Delta(N_i) \to 0$ as $i \to \infty$. Let $i_0 \in \mathbb{N}$ be such that for all $i \ge i_0$, with K as in (5.3.a,b), *(i)* $2K\Delta(N_i) \le -\frac{1}{2}\delta_{\hat{x}}$, *(ii)* $2K\Delta(N_i)^{1-\beta} \le 1$, *(iii)* $x_i \in B(\hat{x}, \rho_{\hat{x}})$, if $i \in I$. Next, let i_1 be such that $N_{i_1} \ge \max\{N_{\hat{x}}, N_{i_0}\}$. Then, for all $i \ge i_1$, $i \in I$, because of (5.3a), (5.8a), and the fact that $\Delta(N_{i+1}) \le \Delta(N_i) \le \Delta(N_{i_1})$,

$$f(x_{i+1}) - f(x_i) \le -\delta_{\hat{x}} + 2K\Delta(N_i) \le -\tfrac{1}{2}\delta_{\hat{x}} . \tag{5.8b}$$

Now, for all $i \geq i_1$, because $\Delta(N_{i+1}) \leq \Delta(N_i) \leq \Delta(N_{i_1})$, we obtain, making use of (5.3a) and (5.6), that

$$f(x_{i+1}) - f(x_i) \leq 2K\Delta(N_i) - \Delta(N_i)^{\beta}$$

$$= -\Delta(N_i)^{\beta}\{1 - 2K\Delta(N_i)^{1-\beta}\} \leq 0, \tag{5.8c}$$

and hence we see that the sequence $\{f(x_i)\}_{i=i_1}^{\infty}$ monotone decreasing. Since, by continuity of $f(\cdot)$, this sequence has an accumulation point, $f(\hat{x})$, it follows that the entire sequence $\{f(x_i)\}_{i=i_1}^{\infty}$ converges to $f(\hat{x})$. Since this is contradicted by (5.8b), our proof is complete. □

Next, we will construct a natural extension of the Master Algorithm 5.9. First, we define the parametrized function $F_{x'} : \mathcal{X} \to \mathbb{R}$, with $x' \in \mathcal{X}$, by

$$F_{x'}(x) \triangleq \max\{f(x) - f(x') - \gamma\psi(x')_+, \psi(x) - \psi(x')_+\}, \tag{5.9a}$$

where $\gamma > 0$ is a preselected parameter. Similarly, for every $N \geq N_o$, we define the parametrized function $F_{N,x'} : \mathcal{X}_N \to \mathbb{R}$, with $x' \in \mathcal{X}_N$, by

$$F_{N,x'}(x) \triangleq \max\{f_N(x) - f_N(x') - \gamma\psi_N(x')_+, \psi_N(x) - \psi_N(x')_+\}. \tag{5.9b}$$

We need the following extension of Assumption 5.7.

Assumption 5.12. Consider the problems $\mathbf{P}_{c,N}$ and suppose that for any $N \geq N_o$, $A_{c,N} : \mathcal{X}_N \to 2^{\mathcal{X}_N}$ is an algorithm map for $\mathbf{P}_{c,N}$. We assume that for every $x \in \mathcal{X}$ such that $\theta_c(x) < 0$, there exist $\rho_x > 0$, $N_x \in \mathbb{N}$, and $\delta_x < 0$ such that

$$F_{N,x'}(x'') \leq \delta_x, \tag{5.10}$$

for all $x' \in B(x, \rho_x) \cap \mathcal{X}_N$, for all $x'' \in A_{c,N}(x')$, and for all $N \geq N_x$. □

Now consider the following master algorithm which uses a strictly monotone decreasing function $\Delta : \mathbf{N} \to \mathbb{R}$, satisfying the conditions of Assumption 5.1.

Master Algorithm Model 5.13.

Parameters. $\beta \in (0, 1)$.

Data. $N_{-1} \in \mathbf{N}$, and $x_0 \in \mathcal{X}_{N_{-1}}$.

Step 0. set $i = 0$.

Step 1. compute N_i and x_{i+1} such that $N_i \geq N_{i-1}$, $x_{i+1} \in A_{c,N_i}(x_i)$ and

$$F_{N_i,x_i}(x_{i+1}) \leq -\Delta(N_i)^{\beta}. \tag{5.11}$$

Step 2. Replace i by $i+1$ and go to Step 1. □

Lemma 5.14. Suppose that Assumptions 5.1 and 5.12 are satisfied, and that Master Algorithm Model 5.13 has constructed an infinite sequence $\{x_i\}_{i=0}^{\infty}$ that has an accumulation point $\hat{x}$. Then the accompanying sequence $\{N_i\}_{i=0}^{\infty}$ is such that $N_i \to \infty$ as $i \to \infty$.

Proof. For the sake of contradiction, suppose that the monotone increasing sequence $\{N_i\}_{i=0}^{\infty}$ is bounded. Then there exists an $i_0 \in \mathbf{N}$, such that $N_i = N_{i_0} \triangleq N^* < \infty$ for all $i \geq i_0$. Then, by the test (5.11)

$$F_{N^*,x_i}(x_{i+1}) \leq -\Delta(N^*)^{\beta} \tag{5.12}$$

for all $i \geq i_0$. Since $\psi_{N^*}(x_{i+1}) - \psi_{N^*}(x_i)_+ \leq F_{N^*,x_i}(x_{i+1})$, for all $i \geq i_0$, it follows from (5.9b) and (5.11) that there must exist an $i_1 \geq i_0$, such that $\psi_{N^*}(x_i) \leq 0$ for all $i \geq i_1$. Hence for all $i \geq i_1$, $\psi_{N^*}(x_i)_+ = 0$, and therefore, in view of (5.9b), $f_{N^*}(x_{i+1}) - f_{N^*}(x_i) \leq F_{N^*,x_i}(x_{i+1})$. Taking into account (5.12) we now conclude that $f_{N^*}(x_i) \to -\infty$ as $i \to \infty$. However, since by continuity, $f_{N^*}(x_i) \to^K f_{N^*}(\hat{x})$, as $i \to \infty$, where $K \subset \mathbf{N}$ is such that $x_i \to^K \hat{x}$ as $i \to \infty$, we have a contradiction. Hence we must have that $N_i \to \infty$ as $i \to \infty$. □

Theorem 5.15. Suppose that Assumptions 5.1 and 5.12 are satisfied, and that Master Algorithm Model 5.13 has constructed an infinite sequence $\{x_i\}_{i=0}^{\infty}$ that has an accumulation point $\hat{x}$. Then $\theta(\hat{x}) = 0$.

Proof. First we note that for all $N \geq N_o$, because of (5.3a,b),

$$F_{x_i}(x_{i+1}) \leq F_{N_i,x_i}(x_{i+1}) + (2+\gamma)K\Delta(N_i)\,. \tag{5.13a}$$

Hence, because of the imposed condition (5.11),

$$F_{x_i}(x_{i+1}) \leq -\Delta(N_i)^{\beta} + (2+\gamma)K\Delta(N_i) = -\Delta(N_i)^{\beta}(1-(2+\gamma)K\Delta(N_i)^{(1-\beta)})\,. \tag{5.13b}$$

Since $1-\beta > 0$, it follows from (5.13b) and the fact that by Lemma 5.15, $\Delta(N_i) \to 0$ as $i \to \infty$, that there exists an i_0 such that for all $i \geq i_0$,

$$F_{x_i}(x_{i+1}) \leq 0\,. \tag{5.13c}$$

Consequently, if $\psi(x_i) > 0$ for all $i \geq i_0$, then $\{\psi(x_i)\}_{i=i_0}^{\infty}$ is a monotone decreasing sequence with an accumulation point $\psi(\hat{x})$. It therefore follows that $\psi(x_i) \to \psi(\hat{x})$ as $i \to \infty$. Alternatively, if there exists an $i_1 \geq i_0$ such that $\psi(x_{i_1}) \leq 0$, then, because of (5.13c), $\psi(x_i) \leq 0$ for all $i \geq i_1$ and $\{f(x_i)\}_{i=i_1}^{\infty}$ is a monotone decreasing sequence with an accumulation point $f(\hat{x})$, and hence that $f(x_i) \to f(\hat{x})$ as $i \to \infty$.

Now, for the sake of contradiction, suppose that $\theta(\hat{x}) < 0$, and that $K \subset \mathbf{N}$ is such that $x_i \to^K \hat{x}$ as $i \to \infty$. Then, because of Assumption 5.8, there exists an i_2, and a $\delta_{\hat{x}} < 0$, such that for all $i \in K, i \geq i_2$,

$$F_{N_i,x_i}(x_{i+1}) \leq \delta_{\hat{x}} < 0, \tag{5.13d}$$

and hence, because of (5.13a) and (5.13c),

$$F_{x_i}(x_{i+1}) \leq (2+\gamma)K\,\Delta(N_i) + \delta_{\hat{x}} < 0\,. \tag{5.13e}$$

Since by Lemma 5.14, $\Delta(N_i) \to 0$ as $i \to \infty$, it follows from (5.13e) that here exists an $i_3 \geq i_2$, such that for all $i \in K$, $i \geq i_3$, $F_{x_i}(x_{i+1}) \leq \frac{1}{2}\delta_{\hat{x}}$. But this contradicts the fact that either $\psi(x_i) \to \psi(\hat{x})$ as $i \to \infty$, or $f(x_i) \to f(\hat{x})$ as $i \to \infty$. Hence we must have that $\theta(\hat{x}) = 0$. □

6. CONCLUSION

We have addressed three issues related to the use of discretizations in the solution of semi-infinite optimization and optimal control problems. We have shown that discretizations of semi-infinite optimization and optimal control problems are consistent approximations to the original problems in the same sense as penalty functions are consistent approximations to constrained nonlinear programming problems, viz., they converge epigraphically to the original problems, and hence that their global minimizers can converge only to a global minimizer of the original problem and their uniformly strict local minimizers converge to a local minimizer of the original problem. Next we have shown that if we express stationarity in terms of zeros of continuous optimality functions, then the stationary points of discretizations of semi-infinite optimization and optimal control problems converge to stationary points of the original problem. Finally, we have proposed several master algorithm models that can be used in constructing algorithms, based on consistent approximations, for solving semi-infinite optimization and optimal control problems.

7. ACKNOWLEDGEMENT

Sincere thanks are due to Prof. R. J.-B. Wets for his explanation of the relevance of epiconvergence to the material treated in this paper, and to the referees whose comments have resulted in considerable revisions of the original manuscript.

8. REFERENCES

[1] ALEKSEEV, B. M., TIKHOMIROV, V. M., and FOMIN, S. V., *Optimalnoye Upravleniye (Optimal Control)*, Nauka, Moscow, 1979.

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

[3] ATTOUCH, H., *Variational Convergence for Functions and Operators*, Pitman, London, 1984.

[4] BAKER, T. E. and POLAK, E., "On the Optimal Control of Systems Described by Evolution Equations", *University of California, Electronics Research Laboratory*, Memo UCB/ERL M89/113, 27 September 1989. To appear in *SIAM Journal on Control and Optimization.*

[5] BYRD, R. H., and NOCEDAL, J., "A Tool for the Analysis of Quasi-Newton Methods with Applications to Unconstrained Optimization", *SIAM Journal on Numerical Analysis* Vol. 26, No.3, pp. 729-739, 1989.

[6] CLARKE, F. H., *Optimization and Nonsmooth Analysis*, Wiley-Interscience, New York, 1983.

[7] DANIEL, J. W., *The Approximation Minimization of Functionals*, Prentice-Hall, Englewood Cliffs, 1971.

[8] DOLECKI, S., SALINETTI, G. and WETS, R. J-B., *Convergence of functions: Equi-semicontinuity*, Transactions of the American Mathematical Society, vol. 276, No. 1, pp.409-429, 1983.

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

[10] FIACCO, A., and McCORMACK, G., *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, John Wiley and Sons, Inc. New York, 1968.

[11] FLETCHER, R., and REEVES, C. M., "Function Minimization by Conjugate Gradients, *Computer Journal*, Vol.7, No.2, pp. 149-154, 1964.

[12] GOLDSTEIN, A., *Constructive Real Analysis*, Harper, New York, 1967.

[13] HE L. and POLAK, E., "An Optimal Diagonalization Strategy for the Solution of a Class of Optimal design Problems", *IEEE Transactions on Automatic Control*, Vol. 35, No.3, pp 258-267, 1990.

[14] KLESSIG, R., and POLAK, E., "An Adaptive Precision Gradient Method for Optimal Control", *SIAM Journal on Control and Optimization*, Vol. 11, No. 1, pp. 80-93, 1973.

[15] KELLEY, C. T., and NORTHRUP J. L., "A Pointwise Quasi-Newton Method for Integral Equations", *SIAM Journal on Numerical Analysis*, Vol. 25, No. 5, pp. 1138-1155, 1988.

[16] LANG, S., *Real Analysis*, Second Edition, Addison Wesley Co, Reading, Mass, 1983.

[17] MANGASARIAN, O. L., and FROMOWITZ, S., "The Fritz John Necessary Optimality Conditions in the Presence of Equality Constraints", *J. Mathematical Analysis and Applications*, Vol. 17, pp. 34-47, 1967.

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

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

[20] POLAK, E., "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.

[21] POLAK, E., and HE, L., "A Unified Phase I-Phase II Method of Feasible Directions for Semi-infinite Optimization", *Journal of Optimization Theory and Applications*, Vol. 69, No.1, pp 83-107, 1991.

[22] POLAK, E., and HE, L., "Rate Preserving Discretization Strategies for Semi-infinite Programming and Optimal Control", *SIAM Journal on Control and Optimization*, Vol. 30, No. 3, pp 548-572, 1992

[23] POLAK, E., MAYNE, D. Q., and HIGGINS, J., "On the Extension of Newton's Method to Semi-Infinite Minimax Problems", *SIAM Journal on Control and Optimization*, Vol. 30, No.2, pp. 376-389, 1992.

[24] PONTRYAGIN, L. S., V. G. BOLTYANSKII, GAMKRELIDZE, R. V., and MISCHENKO, E. F., *The Mathematical Theory of Optimal Processes*, Wiley-Interscience, New York, 1962.

[25] PSHENICHNYI, B. N., and DANILIN, Yu. M., *Numerical Methods in Extremal Problems (Chislennye Metody v Ekstremal'nykh Zadachakh)*, Nauka, Moscow, 1975.

[26] ROCKAFELLAR, R. T. and WETS, R. J.-B., "Variational Systems, an Introduction", in *Multifunctions and Integrands*, G. Salinetti Ed., Springer Lecture Notes in Mathematics 1091, 1983, 1984.

[27] SACHS, E., *Rates of Convergence for Adaptive Newton Methods*, Journal of Optimization Theory and Applications, Vol. 48, No.1, pp. 175-190, 1986.