Global convergence of multidirectional algorithms for unconstrained optimization in normed spaces

J. A. Gómez and M. Romero

Center of Mathematics and Theoretical Physics.
ICIMAF, Havana, Cuba.

Abstract

Global convergence theorems for a class of descent methods for unconstrained optimization problems in normed spaces, using multidirectional search, are proved. Exact and inexact search are considered and the results allow to define a globally convergent algorithm for an unconstrained optimal control problem which operates, at each step, on discrete approximations of the original continuous problem.

AMS Subject Classification: 49M10, 49M37, 65K10.

1 INTRODUCTION

Global convergence theorems of multidirectional search algorithms for finite dimensional optimization problems were derived in [11]. The general global convergence theorem of Zangwill (see [13]) was systematically used and applications to numerical algorithms for unconstrained finite dimensional problems were given. Infinite dimensional optimization problems in Hilbert space were also considered in [11], but the heavy assumption about the closedness of the point-to-set map, which is essential in the Zangwill theorem, limited the applications, in optimal control problems, only to algorithms defined by point-to-set maps with range in a finite dimensional subspace.

In this paper we consider general infinite dimensional optimization problem in a real normed space. Global convergence theorems are proved for a class of descent multidirectional algorithms, using the global convergence theorem of Polak (see [17]) without any finite dimensional assumption. Exact and inexact search are considered and in this last case we use the well known Wolfe conditions for global convergence, which is a common instrument in almost all the Quasi-Newton methods for unconstrained optimization (see for example [5] or [10]).

The results are used to design an algorithm for unconstrained optimal control problems which is globally convergent to local minima of the continuous problem but which operates, at each iteration, with finite dimensional approximation problems which are discretizations of the continuous one. In some sense the convergence result of this algorithm justifies the usual procedure for solving unconstrained optimal control problems, which takes the optimal solution of a convenient finite dimensional discrete problem as a good approximation of the original one.

At each iteration, the algorithm find a new point (i.e. a control function) which satisfies Wolfe's conditions for the continuous problem at the current (function) point, but this is performed finding a new point (i.e. a control sequence) which satisfies Wolfe's conditions for the discrete problem at the current (sequence) point. This point of view allows to use, in the implementation, all the common Quasi-Newton optimization methods and subroutines for the discrete problems. In addition, it also differs from the usual convergence results in common publications, in which it is always proved the convergence of the *optimal solution* of the discrete problem to the optimal solution of the continuous problem (see for example [6]). Finally, the convergence results of the algorithm, given in the paper, are related with recent developments which study the influence of the discretization step size in the algorithmic convergence (see [2]), since each iteration can be seen as an attempt to select not only a direction, but also a convenient step size to improve the current solution.

2 GLOBAL CONVERGENCE THEORY

We consider the following problem:

$$\begin{array}{l} \min\ f(x) \\ x \in X, \end{array} \tag{1}$$

where X is a real normed space, and we will study the global convergence of multidirectional descent algorithms, with exact and inexact search. We suppose $f \in C^1$, a continuously differentiable function, and denote by $\nabla f(x) \in X^*$ the Frechet derivative of f at x, where X^* is the topological dual space of X. In all the paper, the symbol $\langle y^*, x\rangle$ denotes the evaluation of the functional $y^* \in X^*$ at point x.

Definition 1 *An algorithmic map in X is a point-to-set map A defined in X with values in the class $\mathcal{P}(X)$ of all subsets of X. A function $c : X \to \Re$ is a stopping rule for the set $\Gamma \subset X$, respect to the map A in X, if for all $x \notin \Gamma$ we have:*

$$c(x^{'}) < c(x), \forall x^{'} \in A(x).$$

Definition 2 *An algorithm $\mathcal{A}$ on X for finding points of a set $\Gamma \subset X$ (shortly an algorithm for Γ), is given by a sequence of pairs $(A_i, c_i)_{i\in\aleph}$, where A_i are algorithmic maps satisfying:*

$$A_i(z) \neq \emptyset,\ \forall z \notin \Gamma,\ \forall i \in \aleph,$$

and $c_i(.)$ are stopping rule functions for the set Γ, and the following steps:

A1) Choose $x_0 \in X$, and set $i = 0$,

A2) Compute $A_i(x_i)$,

A3) Choose $y_i \in A_i(x_i)$,

A4) If $c_i(y_i) \geq c_i(x_i)$, stop and take x_i as the last point of the sequence,
If $c_i(y_i) < c_i(x_i)$, go to step A5),

A5) Take $x_{i+1} = y_i$, set $i + 1 \to i$, and go to step A2).

When both sequences are constant $A_i = A$, $c_i = c$, $\forall i \in \aleph$, we say that we have a uniform algorithm $\mathcal{A} = (A, c)$.

Definition 3 *A (finite or infinite) sequence generated by an algorithm $\mathcal{A} = (A_n, c_n)_{n\in\aleph}$, from the starting point $\bar{x} \in X$, is any sequence $\{x_n,\ n = 1, 2, ...\} \subset$*

X which is obtained following the steps A1) - A5) in the last definition, with $x_0 = \bar{x}$.

An algorithm $\mathcal{A} = (A_n, c_n)_{n\in\aleph}$ *for* $\Gamma \subset X$ *is called globally convergent to* Γ, *if for all* $x_0 \in X$ *and any sequence* $\{x_n\}$ *generated by* $\mathcal{A}$ *from the starting point* x_0, *we have that the last element of* $\{x_n\}$ *belongs to* Γ, *if* $\{x_n\}$ *is finite, or that every accumulation point of* $\{x_n\}$ *belongs to* Γ, *if* $\{x_n\}$ *is infinite.*

Theorem 1 *(Polak): Let* X *be a metric space and* $\mathcal{A} = (A, c)$ *a uniform algorithm in* X *for* $\Gamma \subset X$, *and suppose the maps* $A(.)$ *and* $c(.)$ *satisfy the following conditions:*

$$\begin{array}{l} i)\ c(.) \text{ is continuous or bounded below in } X, \\ ii)\ \forall z \notin \Gamma, \exists \varepsilon = \varepsilon(z) > 0, \exists \gamma = \gamma(z) < 0 : \\ c(z'') - c(z') < \gamma, \forall z'' \in A(z'), \forall z' \in B(z, \varepsilon) \end{array} \tag{2}$$

Then the algorithm $\mathcal{A}$ *is globally convergent to* Γ.

For a proof see ([17]).

Definition 4 *Let be* α, $\beta \in (0,1)$. *We say the direction* $d \in X$ *satisfies Wolfe's conditions respect to* $f \in C^1$ *at* $x \in X$ *if the following inequalities hold:*

$$\begin{array}{r} f(x+d) \le f(x) + \alpha \langle \bigtriangledown f(x), d \rangle, \\ \langle \bigtriangledown f(x+d), d \rangle \ge \beta \langle \bigtriangledown f(x), d \rangle. \end{array} \tag{3}$$

We call $\alpha-$*condition (respectively* $\beta-$*condition) the inequality corresponding to* α *(respectively to* β*) in (3). We will also say that the point* $y = x + d$ *satisfies the Wolfe conditions.*

Lemma 1 *Let* X *be a normed space and* f *a Frechet differentiable function. Let's consider* $\alpha, \beta \in (0,1)$, $\beta > \alpha$. *Let be* $x \in X$ *and* d *a descent direction in* x, *i.e.* $\langle \bigtriangledown f(x), d \rangle < 0$. *Suppose the set* $\{f(x + \lambda d), \lambda \ge 0\}$ *is bounded below.*

Then there exists $y \in X$, *with the form* $y = x + \bar{\lambda} d$, *which strictly satisfies the Wolfe conditions:*

$$\begin{array}{r} f(y) < f(x) + \alpha\bar{\lambda} \langle \bigtriangledown f(x), d \rangle, \\ \langle \bigtriangledown f(y), d \rangle > \beta \langle \bigtriangledown f(x), d \rangle, \end{array} \tag{4}$$

and therefore, there exists an entire interval $(\bar{\lambda} - \varepsilon, \bar{\lambda} + \varepsilon)$ *where the Wolfe conditions hold.*

Proof. We have:

$$f(x+dl) = f(x) + \langle \nabla f(x), dl \rangle + o(|l|) < f(x) + \alpha \langle \nabla f(x), dl \rangle \qquad (5)$$

for small $l > 0$.

Since $\{f(x+dl), l \in \Re\}$ is bounded below and $l \to f(x+dl)$ is continuous, there exists the least $l_0 > 0$ such that:

$$f(x+dl_0) = f(x) + \alpha \langle \nabla f(x), dl_0 \rangle .$$

In fact, for l near 0 we have (5) and for $l \to +\infty$, the function $f(x+dl)$ is bounded and $f(x) + \alpha \langle \nabla f(x), dl \rangle \to -\infty$.

On the other hand, we have:

$$f(x+dl_0) - f(x) = \left\langle \nabla f(x+d\tilde{l}), dl_0 \right\rangle, \text{ for some } \tilde{l} \in (0, l_0),$$

therefore:

$$\left\langle \nabla f(x+d\tilde{l}), dl_0 \right\rangle = \alpha \langle \nabla f(x), dl_0 \rangle > \beta \langle \nabla f(x), dl_0 \rangle ,$$

since $\beta > \alpha$ and $\langle \nabla f(x), d \rangle < 0$, then:

$$\left\langle \nabla f(x+d\tilde{l}), d \right\rangle > \beta \langle \nabla f(x), d \rangle .$$

Taking $y = x + d\tilde{l}$, and recalling that $\tilde{l} < l_0$ and the definition of l_0, we have:

$$\begin{aligned} f(x+d\tilde{l}) &= f(y) < f(x) + \alpha \tilde{l} \langle \nabla f(x), d \rangle , \\ \left\langle \nabla f(x+d\tilde{l}), d \right\rangle &= \langle \nabla f(y), d \rangle > \beta \langle \nabla f(x), d \rangle . \end{aligned}$$

The existence of an interval is a consequence of the continuity of the functions $l \to f(x+dl)$ and $l \to \nabla f(x+dl)$. □

Lemma 2 *Let X be a normed space, f a continuously Frechet differentiable function and $\beta \in (0,1)$. Let be $\{x_n\}_{n \in \aleph} \subset X$, with $x_n \underset{n \to +\infty}{\to} x$ and $\nabla f(x) \neq 0$, and $\{y_n\}_{n \in \aleph} \subset X$ with $y_n = x_n + \alpha_n d_n$, $\alpha_n \geq 0, \|d_n\| = 1$, such that, for all $n \in \aleph$, the inequalities:*

$$\begin{aligned} \langle \nabla f(x_n), d_n \rangle &\leq -\rho \|\nabla f(x_n)\| , \\ \langle \nabla f(y_n), d_n \rangle &\geq \beta \langle \nabla f(x_n), d_n \rangle , \end{aligned}$$

hold, and suppose $y_n \underset{n \to +\infty}{\to} y$. Then $x \neq y$.

Proof. We have:

$$\begin{aligned} &\langle [\nabla f(x_n) - \nabla f(y_n)], d_n \rangle \leq \langle \nabla f(x_n), d_n \rangle - \beta \langle \nabla f(x_n), d_n \rangle = \\ &= (1-\beta) \langle \nabla f(x_n), d_n \rangle \leq -\rho(1-\beta) \| \nabla f(x_n) \| < 0, \end{aligned} \tag{6}$$

and

$$|\langle [\nabla f(x_n) - \nabla f(y_n)], d_n \rangle| \leq \| \nabla f(x_n) - \nabla f(y_n) \| \, \| d_n \| = \| \nabla f(x_n) - \nabla f(y_n) \| .$$

If we suppose $x = y$, then there exists the limit of the sequence:

$$|\langle [\nabla f(x_n) - \nabla f(y_n)], d_n \rangle| ,$$

which, by continuity of ∇f, satisfies:

$$\lim_{n \to +\infty} |\langle [\nabla f(x_n) - \nabla f(y_n)], d_n \rangle| = 0, \tag{7}$$

and therefore, the sequence of real numbers $\langle [\nabla f(x_n) - \nabla f(y_n)], d_n \rangle$ converges to 0; but on the other hand:

$$\begin{aligned} &\lim_{n \to +\infty} \langle [\nabla f(x_n) - \nabla f(y_n)], d_n \rangle \leq \\ &\leq \lim_{n \to +\infty} -\rho(1-\beta) \| \nabla f(x_n) \| = -\rho(1-\beta) \| \nabla f(x) \| < 0. \end{aligned}$$

We have a contradiction. □

Usually, a descent uniform algorithm $\mathcal{A}$ for the problem (1) is defined through an algorithmic point-to-set map A, which is the composition of two maps $A = S \circ G$, representing a "selector of directions" map and a "selector of new points" map respectively. We recall that the composition map of the point-to-set maps S and G is defined by:

$$(S \circ G)(x) = \bigcup_{y \in G(x)} S(y).$$

The stopping rule function $c(.)$ of a descent uniform algorithm $\mathcal{A}$ for the problem (1), is almost always chosen as the objective function $f(.)$. In defining our first algorithm, we will use this common point of view.

Definition 5 *Let be $p \in \aleph$ and $\rho \in (0,1)$. The point-to-set map of ρ-non orthogonal descent directions $G_\rho^p(z) : X \to X \times \mathcal{P}(X^p)$ is defined by:*

$$\begin{aligned} G_\rho^p(z) = \; &\{(z, D) \in \{z\} \times X^p \mid \exists \lambda \in \Re^p : \| D\lambda \| \neq 0, \\ &\langle \nabla f(z), D\lambda \rangle \leq -\rho \| \nabla f(z) \| \, \| D\lambda \| \} \end{aligned} \tag{8}$$

where D is a $p-$vector $D = (D_1, ..., D_p)$, with $D_i \in X, i = 1, ..., p$, and the product $D\lambda$ is defined by the linear combination : $D\lambda = \sum_{i=1}^{p} \lambda_i D_i$. The norm in X^p is the usual product norm:

$$\|D\|_\infty = \max_{i=1,...,p} \|D_i\| .$$

Definition 6 *Let be $p \in \aleph$. The point-to-set map of exact search $S^p(z, D) : X \times X^p \to X$ is defined by:*

$$S^p(z, D) = \left\{ y \in X \mid y = z + D\bar{\lambda},\ f(y) = \min_{\lambda \in \Re^p} f(z + D\lambda) \right\} . \tag{9}$$

Definition 7 *The descent uniform algorithm $\mathcal{A}_p = (A_p, c)$ for the problem (1), with multidirectional and exact search, is defined by the algorithmic point-to-set map $A_p = S^p \circ G_\rho^p$, and the stopping rule function $c(x) = f(x)$.*

In words, the algorithm $\mathcal{A}_p$ selects, at step k, a $p-$vector of directions $D = (d_1, ..., d_p)$ belonging to X^p, in such a way that the linear variety generated by those vectors contains a descent direction of the objective function f at the current point x_k. In the next step, a new point x_{k+1} is chosen as the minimum of the objective function $f(z)$ in the linear variety $x_k + \mathcal{S}[d_1, ..., d_p]$.

Theorem 2 *Let X be a normed space, f a continuously Frechet differentiable and bounded below function, and $\rho \in (0, 1)$. For any $p \in \aleph$, the descent uniform algorithm with multidirectional and exact search, $\mathcal{A}_p = (A_p, f)$, is globally convergent to the set:*

$$\Gamma = \{x \in X \mid \bigtriangledown f(x) = 0\} .$$

Proof. If $x \notin \Gamma \Leftrightarrow \nabla f(x) \neq 0$, there exists a $\varepsilon_0 > 0$, such that:

$$\nabla f(x') \neq 0,\ \forall x' \in B(x, \varepsilon),\ \forall \varepsilon \in (0, \varepsilon_0),$$

then a $\rho-$descent direction always exist and therefore, Lemma (1) ensures that:

$$A_p(x') \neq \emptyset,\ \forall x' \in B(x, \varepsilon),\ \forall \varepsilon \in (0, \varepsilon_0).$$

. We will verify the conditions (2) of Polak's global convergence theorem 1 for the set Γ.

i) $c(.) = f(.)$ is continuous.

ii) By contradiction, suppose it is not true. Then $\exists\, x \in X$ with $\bigtriangledown f(x) \neq 0$ such that $\forall \varepsilon_x > 0$ and $\forall \gamma_x < 0$, $\exists x^{'} \in B(x, \varepsilon_x)$ and $\exists x^{''} \in A_p(x^{'})$ such that $\gamma_x \leq f(x^{''}) - f(x^{'}) < 0$. Taking $(\gamma_x)_n = -\frac{1}{n}$ and $(\varepsilon_x)_n = \frac{1}{n}$, $\forall n \in \aleph$, we have $\exists \left\{ x_n^{'} \right\}_{n\in\aleph}$ with $x_n^{'} \underset{n\to+\infty}{\to} x$ and $\exists \left\{ x_n^{''} \right\}_{n\in\aleph}$ with $x_n^{''} \in A_p(x_n^{'})$ such that $f(x_n^{''}) - f(x_n^{'}) \underset{n\to+\infty}{\to} 0$. But, by definition $x_n^{''} = x_n^{'} + D_n^{'} \lambda_n^{'}$, and denoting $\bar{\lambda}_n^{'}$ the vector of $\Re^p$ which satisfies the $\rho-$condition in (8), we will have:

$$f(x_n^{''}) - f(x_n^{'}) = f(x_n^{'} + D_n^{'} \lambda_n^{'}) - f(x_n^{'}) \leq f(x_n^{'} + D_n^{'}(\bar{\lambda}_n^{'} l)) - f(x_n^{'}),\ \forall l \in \Re,$$

by optimality of $x_n^{''} \in S^p(x_n^{'}, D_n^{'})$. But for all $n \in \aleph$, $\left\| D_n^{'} \bar{\lambda}_n^{'} \right\| \neq 0$, and the vectors $d_n^{'} = \frac{D_n^{'} \bar{\lambda}_n^{'}}{\left\| D_n^{'} \bar{\lambda}_n^{'} \right\|}$, $\left\| d_n^{'} \right\| = 1$, satisfy the inequalities:

$$\left\langle \bigtriangledown f(x_n^{'}), d_n^{'} \right\rangle \leq -\rho \left\| \bigtriangledown f(x_n^{'}) \right\|.$$

By Lema 1, for any $\alpha, \beta \in (0,1)$, $\beta > \alpha$, and for all $n \in \aleph$, there exists $z_n^{'} \in X$, $z_n^{'} = x_n^{'} + d_n^{'} l_n^{'}$, such that:

$$\begin{aligned} f(z_n^{'}) &\leq f(x_n^{'}) + \alpha \left\langle \bigtriangledown f(x_n^{'}), d_n^{'} l_n^{'} \right\rangle, \\ \left\langle \bigtriangledown f(z_n^{'}), d_n^{'} \right\rangle &\geq \beta \left\langle \bigtriangledown f(x_n^{'}), d_n^{'} \right\rangle. \end{aligned}$$

Hence, there is not a subsequence $l_{n_k}^{'}$ converging to 0, since we would have:

$$\begin{aligned} & x_{n_k}^{'} \underset{k\to+\infty}{\to} x,\ \bigtriangledown f(x) \neq 0, \\ & z_{n_k}^{'} \in X,\ z_{n_k}^{'} = x_{n_k}^{'} + d_{n_k}^{'} l_{n_k}^{'}, \\ & \left\langle \bigtriangledown f(x_{n_k}^{'}), d_{n_k}^{'} \right\rangle \leq -\rho \left\| \bigtriangledown f(x_{n_k}^{'}) \right\|, \\ & \left\langle \bigtriangledown f(z_{n_k}^{'}), d_{n_k}^{'} \right\rangle \geq \beta \left\langle \bigtriangledown f(x_{n_k}^{'}), d_{n_k}^{'} \right\rangle, \\ & z_{n_k}^{'} \underset{k\to+\infty}{\to} x, \end{aligned}$$

which is a contradiction with Lema 2. In addition, by the definition of A_p, we have the inequalities:

$$f(x_n^{''}) - f(x_n^{'}) \leq f(z_n^{'}) - f(x_n^{'}) \leq \alpha \left\langle \bigtriangledown f(x_n^{'}), d_n^{'} l_n^{'} \right\rangle < 0,\ \forall n \in \aleph,$$

and then:

$$f(z_n^{'}) - f(x_n^{'}) \underset{n\to+\infty}{\rightarrow} 0 \Rightarrow \left\langle \bigtriangledown f(x_n^{'}), d_n^{'} l_n^{'} \right\rangle \underset{n\to+\infty}{\rightarrow} 0. \tag{10}$$

On the other hand,

$$\lim_{n\to+\infty} \left\langle \bigtriangledown f(x_n^{'}), d_n^{'} l_n^{'} \right\rangle \leq \lim \sup_{n\to+\infty} (-\rho) \left\| \nabla f(x_n^{'}) \right\| l_n^{'} = -\rho \left\| \nabla f(x) \right\| \lim \inf_{n\to+\infty} l_n^{'} < 0,$$

which is a contradiction with (10). □

The other descent uniform algorithm requires a non composite definition:

Definition 8 *Let be $p \in \aleph$ and $\rho, \alpha, \beta \in (0,1)$, $\beta > \alpha$. The point-to-set map $\bar{A}_p : X \to X$, given by:*

$$\begin{array}{rl} \bar{A}_p(z) = \{ y \in X \mid & \exists \bar{\lambda} \in \Re^p, D \in X^p : \left\| D\bar{\lambda} \right\| \neq 0, \\ & y = z + D\bar{\lambda}, \\ & \left\langle \nabla f(z), D\bar{\lambda} \right\rangle \leq -\rho \left\| \nabla f(z) \right\| \left\| D\bar{\lambda} \right\|, \\ & f(z + D\bar{\lambda}) \leq f(z) + \alpha \left\langle \bigtriangledown f(z), D\bar{\lambda} \right\rangle, \\ & \left\langle \bigtriangledown f(z + D\bar{\lambda}), D\bar{\lambda} \right\rangle \geq \beta \left\langle \bigtriangledown f(z), D\bar{\lambda} \right\rangle \}, \end{array} \tag{11}$$

and the stopping rule objective function $c(.) = f(.)$ define the descent uniform algorithm $\mathcal{A}_{p,w}$ with multidirectional inexact search..

In words, $\mathcal{A}_{p,w}$ find, in the subspace generated by the directions belonging to D, a linear combination $D\bar{\lambda}$ which satisfies the $\rho-$, $\alpha-$ and $\beta-$conditions at the same time. We can consider also an algorithm with variable number of directions, i.e. we use multidirectional search but in subspaces with different dimensions at each step:

Definition 9 *Let $\{p_n\} \subset \aleph$ be a sequence of natural numbers. For $\rho, \alpha, \beta \in (0,1)$, $\beta > \alpha$, let's consider the sequences of point-to-set maps $\bar{A}_{p_n}$, defined, for $p = p_n$, as in (11). The descent algorithm with variable multidirectional inexact search is defined by $\mathcal{A}_{\{p_n\},w} = (\bar{A}_{p_n}, c)_{n\in\aleph}$, where the stopping rule sequence is constant, $c(.) = f(.)$.*

Theorem 3 *Let X be a normed space and f a continuously Frechet differentiable and bounded below function. Let be $\rho, \alpha, \beta \in (0,1)$, $\beta > \alpha$. For any sequence $\{p_n\} \subset \aleph$, the descent uniform algorithm $\mathcal{A}_{\{p_n\},w}$, with variable multidirectional and inexact search, is globally convergent to the set:*

$$\Gamma = \{ x \in X \mid \bigtriangledown f(x) = 0 \}.$$

Proof. Lemma (1) ensures again that:

$$\forall x \notin \Gamma, \exists \varepsilon_x > 0 : \forall n \in \aleph,\ \bar{A}_{p_n}(x') \neq \emptyset,\ \forall x' \in B(x,\varepsilon),\ \forall \varepsilon \in (0,\varepsilon_x).$$

We will use similar arguments as in the preceding proof to verify the conditions (2) of the Polak's global convergence theorem.

i) $c(.) = f(.)$ is continuous.

ii) By contradiction, suppose it is not true. Then $\exists\ x \in X$ with $\bigtriangledown f(x) \neq 0$ such that $\forall \varepsilon_x > 0$ and $\forall \gamma_x < 0$, $\exists x' \in B(x,\varepsilon_x)$ and $\exists x'' \in \bar{A}_{p_n}(x')$ such that $\gamma_x \leq f(x'') - f(x') < 0$. Taking $(\gamma_x)_n = -\frac{1}{n}$ and $(\varepsilon_x)_n = \frac{1}{n}$, $\forall n \in \aleph$, we have $\exists \left\{x'_n\right\}_{n\in\aleph}$ with $x'_n \underset{n\to+\infty}{\to} x$ and $\exists \left\{x''_n\right\}_{n\in\aleph}$ with $x''_n \in \bar{A}_{p_n}(x'_n)$ such that $f(x''_n) - f(x'_n) \underset{n\to+\infty}{\to} 0$.

But $x''_n = x'_n + l'_n d'_n$, with $d'_n = \frac{D_n\lambda_n}{\|D_n\lambda_n\|}$, $D_n \in X^{p_n}$, $l'_n = \|D_n\lambda_n\|$, $\left\langle \bigtriangledown f(x'_n), d'_n \right\rangle \leq -\rho \left\|\bigtriangledown f(x'_n)\right\|$, $\left\langle \bigtriangledown f(x''_n), d'_n \right\rangle \geq \beta \left\langle \bigtriangledown f(x'_n), d'_n \right\rangle$, $f(x''_n) \leq f(x'_n) + \alpha \left\langle \bigtriangledown f(x'_n), d'_n l'_n \right\rangle$, $\left\|d'_n\right\| = 1$, $l'_n > 0$, $\forall n \in \aleph$; then, there is not a subsequence $\left\{l'_{n_k}\right\}_{k\in\aleph}$ such that $l'_{n_k} \underset{k\to+\infty}{\to} 0$, since we would have $x'_{n_k} \underset{k\to+\infty}{\to} x$ and $x''_{n_k} = x'_{n_k} + l'_{n_k} d'_{n_k} \underset{k\to+\infty}{\to} x$, and this is a contradiction with Lema 2. Furthermore:

$$f(x''_n) - f(x'_n) \leq \alpha \left\langle \bigtriangledown f(x'_n), d'_n l'_n \right\rangle < 0$$

and if we have $f(x''_n) - f(x'_n) \underset{n\to+\infty}{\to} 0$ we obtain:

$$\alpha \left\langle \bigtriangledown f(x'_n), d'_n l'_n \right\rangle \underset{n\to+\infty}{\to} 0.$$

On the other hand

$$\alpha \left\langle \bigtriangledown f(x'_n), d'_n l'_n \right\rangle \leq -\alpha\rho \left\|\bigtriangledown f(x'_n)\right\| l'_n,$$

and then:

$$0 = \lim_{n\to+\infty} \alpha \left\langle \bigtriangledown f(x'_n), d'_n l'_n \right\rangle \leq \limsup_{n\to+\infty} -\alpha\rho \left\|\bigtriangledown f(x'_n)\right\| l'_n = -\alpha\rho \left\|\bigtriangledown f(x)\right\| \liminf_{n\to+\infty} l'_n < 0,$$

since there is no subsequence $\left\{l'_{n_k}\right\}_{k\in\aleph}$ converging to 0. This contradiction proves the theorem. □.

3 APPLICATION TO OPTIMAL CONTROL PROBLEMS

3.1 Discretization of Unconstrained Optimal Control Problems

We consider the following optimal control problem:

$$\begin{aligned} \min J(u(.)) &= \varphi[x(t_1)], \\ s.t.\ \dot{x}(t) &= f(x(t),u(t)),\ a.a.\ t \in [t_0.t_1], \\ x(t_0) &= \hat{x}_0, \\ u(t) &\in U,\ a.a.\ \ t \in [t_0.t_1], \end{aligned} \tag{12}$$

where $u(.) \in L_2^m = L_2\left([t_0.t_1], \Re^m\right)$,the set of square integrable functions with image in $\Re^m$ and $x(.) \in \mathcal{C} = C_a\left([t_0.t_1], \Re^n\right)$, the set of absolutely continuous functions with image in $\Re^n$.

The maps $\varphi : \Re^n \to \Re$ and $f : \Re^n \times \Re^m \to \Re^n$, are supposed continuously differentiable respect to its arguments, the functional $J(u)$ is assumed to be Frechet differentiable respect to the L_2^m-norm and the set $U \subset \Re^m$, for our purposes, will be considered the whole space $\Re^m$.

If we apply Euler's integration scheme we obtain a discrete approximation of the problem (12) by the following finite dimensional optimal control problem:

$$\begin{aligned} \min J_N(\bar{u}_N) &= \varphi[x_N], \\ s.t.\ x_{i+1} &= x_i + hf(x_i,u_i),\ i=0,1,...,N-1, \\ x_0 &= \hat{x}_0, \\ \bar{u}_N &= \left\{u_0,u_1,...,u_{N-1}\right\}, \\ u_i &\in U,\ i=0,1,...,N-1. \end{aligned} \tag{13}$$

where $h = \frac{(t_1-t_0)}{N}$ is the integration step, which defines a partition $\{\tau_i,\ i = \overline{0,N}\}$ of $[t_0,t_1]$ in N subintervals:

$$\tau_i = t_0 + ih,\ i=0,...,N,$$

which in turn allows us to define, from a given feasible control sequence $\bar{u}_N = \{u_0,u_1,...,u_{N-1}\} \in \mathcal{L}_N^m$ of the discrete problem (13), a feasible control function $\bar{u}_N(.) \in L_2^m$ of the continuous problem (12), with the classical piecewise constant form:

$$\bar{u}_N(t) = u_j, \text{ for } t \in [\tau_j, \tau_{j+1}), \; j = 0, 1, ..., N-1. \tag{14}$$

We can make this definition from any given vector sequence defined on any partition $\{\tau_i\}$. The piecewise constant function $\bar{u}_N(.)$, defined in (14), from a vector sequence $\bar{u}_N = \{u_0, u_1, ..., u_{N-1}\}$ given on a partition $\{\tau_i, \; i = \overline{0, N}\}$, will be called "the constant canonical extension" of the sequence to a function in L_2^m.

¿From the corresponding discrete trajectory, $\bar{x}_N = \{x_0, x_1, ..., x_N\} \in \mathcal{L}_{N+1}^n$, we can define the continuous function:

$$\bar{x}_N[t] = x_j + (t - \tau_j) f(x_j, u_j), \; t \in [\tau_j, \tau_{j+1}], \tag{15}$$

which is the classical polygonal Euler approximation of the solution $x_N(.)$ of the differential equation system of (12), corresponding to the constant canonical extension $\bar{u}_N(.)$, and satisfying:

$$\dot{x}_N(t) = f(x_N(t), \bar{u}_N(t)), \; t \in [t_0.t_1], \; x_N(t_0) = \hat{x}_0.$$

The continuous function defined in (15), from a vector sequence $\bar{x}_N$ given in a partition $\{\tau_i\}$, will be called "the polygonal canonical extension" of the sequence to a function in $\mathcal{C}$.

We can define, reciprocally, from any piecewise continuous function $z(t)$ on $[t_0, t_1]$ with image in $\Re^k$, a vector sequence $\bar{z}_N = \{z_0, z_1, ..., z_{N-1}\} \in \mathcal{L}_N^k$, defined on a partition $\{\tau_i, \; i = \overline{0, N}\}$ of $[t_0, t_1]$, in the trivial way:

$$z_i = z(\tau_i), \; i = 0, ..., N-1, \tag{16}$$

and we will call (16) "the sequential canonical reduction" of $z(.)$ to a sequence in $\mathcal{L}_N^k$.

As final remarks about notation, we are denoting by $\mathcal{L}_N^k$ the set of finite sequence of N vectors in $\Re^k$, which is isomorphic to $\Re^{kN}$;the vector sequence associated with a partition $\{\tau_i, i = 1, ..., N\}$, with elements$\{z_i, i = 1, ..., N\}$, is denoted by $\bar{z}_N$; for the constant canonical extension to L_2^m we add parenthesis $\bar{z}_N(.)$ and for the polygonal canonical extension to $\mathcal{C}$ we add square brackets $\bar{z}_N[.]$; finally, for the sequential canonical reduction of $z(.)$ to $\mathcal{L}_N^k$, we suppress the parenthesis, add an over bar to z and a subindex with the number of vectors defined in the partition: $\bar{z}_N$.

We note also that a vector sequence $\bar{u}_N \in \Re^{mN}$, given on the partition $\{\tau_i\}$, can be considered defined in any partition $\{\tau_i^{'}\}$ of $[t_0, t_1]$ which contains $\{\tau_i\}$ (with $N^{'}$ intervals, $N^{'} > N$), defining first the constant canonical extension $\bar{u}_N(.)$ of $\bar{u}_N$ to L_2^m and then taking the sequential canonical reduction $\bar{u}_{N'}$ of $\bar{u}_N(.)$ to $\{\tau_i^{'}\}$. It's easy to see that we also have equality for the constant canonical extension of $\bar{u}_N$ and $\bar{u}_{N'}$:

$$\bar{u}_N(t) = \bar{u}_{N'}(t), \forall\, t \in [t_0.t_1], \forall N^{'} \geq N,$$

therefore, from now on we will identify $\bar{u}_N(.)$ with $\bar{u}_{N'}(.)$ for any $N^{'} > N$ which corresponds to a partition $\{\tau_i^{'}\}$ containing $\{\tau_i\}$.

There are many publications about the convergence of the optimal solution of the discrete problem (13) to an optimal solution of the continuous problem (12) when $N \to +\infty$. There are even quantitative results in the speed of convergence of error estimates of optimal trajectories, controls and adjoint variables, and also many generalized results in several directions. As examples, it can be mentioned the works of Alt [1], Daniel [4], Dontchev [6], Evtuschenko [8], Hager [12], Malanowski [14], Mordukhovich [16], Teo [20] and many others.

In our opinion, all these results are more in the "stability of optimal solution" framework than in the "convergence of an algorithm" context. We haven't seen that the concepts of "descent and feasible directions" or "inexact line search", appearing naturally in the context of finite dimensional optimization algorithm (see for example [5]), have been sufficiently exploited in the design of "finite dimensional approximation" algorithms for optimal control problems and in the proof of they global convergence. In this paper we will present an example of this other point of view.

To any control sequence $\bar{u}_N = \{u_0, u_1, ..., u_{N-1}\} \in \mathcal{L}_N^m$, which is feasible to the problem (13), corresponds a piecewise constant function $\bar{u}_N(.) \in L_2^m$, through the constant canonical extension, which is feasible to the problem (12) and reciprocally.

¿From a current point $\bar{u}_N^k \in \Re^{mN}$, the $k+1-$step in a classical descent algorithm for the discrete optimization problem (13) consists of finding a direction of decrease $\triangle\bar{u}_N^k$ in $\bar{u}_N^k$ and then performing an inexact line search to find a step length λ_k and a new point $\bar{u}_N^{k+1} = \bar{u}_N^k + \lambda_k \triangle\bar{u}_N^k$ which satisfies the $\alpha-$ and $\beta-$ global convergence conditions of Wolfe. This $k+1-$step can be viewed as one step of a single-direction optimization algorithm for the continuous problem (12), identifying $\bar{u}_N$ and $\triangle\bar{u}_N^k$ with their canonical extensions to L_2^m. If any of the Wolfe conditions fails to hold or even when they both hold, we can perform several iteration now in a non classical algorithm for the discrete problem and this can be viewed as one step of a multidirectional optimization algorithm for the continuous problem. Increasing of N implies an increment of the number of variables in the discrete problem and a reduction of the integration step in Euler's formula for the continuous problem. It also can be considered as the start of searching in a new direction of a multidirectional optimization algorithm for (12).

The number of iteration made by the optimization algorithm in each approximating discrete problem until a satisfactory point was found would be the number of directions that we take at that step of the algorithm for the continuous problem. This number can be the same or can be varied in different iteration during computation, but in practice it is always bounded.

Therefore, since we have global convergence theorems to local minima for unconstrained multidirectional descent methods with inexact line search (using Wolfe conditions) and with possible variable number of direction at each iteration, we have the conditions to model an algorithm for the continuous problem which is based on iterations in the discrete one. Hence, we will examine the following questions:

a) When $\rho-$non orthogonal conditions in the discrete problems implies the same condition in the continuous problem?

b) When the global convergence $\alpha-$ and $\beta-$Wolfe conditions in the discrete problem implies the same conditions in the continuous problem?

c) Is it possible to design a global convergence algorithm for the continuous problem, only ensuring the (possible varying) Wolfe conditions in the discrete problems?

In the next section we will answer these questions positively, and the idea is quite simple:

1) The descent $\rho-$condition for the direction $\triangle\bar{u}_N^k(.)$ depends on the gradient of $J(.)$ at the current point $\bar{u}_N^k(.)$:

$$\left\langle \nabla J(\bar{u}_N^k(.)), \triangle\bar{u}_N^k(.)\right\rangle_{L_2^m} \leq -\rho \left\| \nabla J(\bar{u}_N^k(.))\right\|_{L_2^m} \left\|\triangle\bar{u}_N^k(.)\right\|_{L_2^m}$$

therefore we should have the discrete gradient $\nabla J_N(\bar{u}_N^k)$ close, in some sense, to the continuous gradient $\nabla J(\bar{u}_N^k(.))$ at the current point,

2) The $\beta-$condition for the direction $\triangle\bar{u}_N^k(.)$ depends on the gradient of

$J(.)$ at the current point $\bar{u}_N^k(.)$ and at the new point $\bar{v}_N^k(.) = \bar{u}_N^k(.) + \triangle\bar{u}_N^k(.)$:

$$\left\langle \nabla J(\bar{v}_N^k(.)), \triangle\bar{u}_N^k(.)\right\rangle_{L_2^m} \geq \beta \left\langle \nabla J(\bar{u}_N^k(.)), \triangle\bar{u}_N^k(.)\right\rangle_{L_2^m},$$

therefore we should also have the discrete gradient $\nabla J_N(\bar{v}_N^k)$ close, in the same sense as before, to the continuous gradient $\nabla J(\bar{v}_N^k(.))$ at the new point,

3) The $\alpha-$condition for the direction $\triangle\bar{u}_N^k(.)$ depends on the increment of $J(.)$ at the new point respect to the current one, and on the gradient of $J(.)$ at the current point :

$$J(\bar{v}_N^k(.)) - J(\bar{u}_N^k(.)) \leq \alpha \left\langle \bigtriangledown J(\bar{u}_N^k(.)), \triangle\bar{u}_N^k(.)\right\rangle_{L_2^m},$$

therefore we should have also the discrete increment $J_N(\bar{v}_N^k) - J_N(\bar{u}_N^k)$ close to the continuous increment $J(\bar{v}_N^k(.)) - J(\bar{u}_N^k(.))$.

We can't expect that the direction $\triangle\bar{u}_N^k$ verify both the discrete and the continuous global convergence conditions for the same parameters ρ, α and β. Then, the main difficulties are:

- first, to find conditions on the parameters and on the closeness of the required quantities in such a way that the corresponding $\rho-, \alpha-$ or $\beta-$condition is satisfied at each problem,
- second, to prove that it is possible to choose the parameter values in such a way that they satisfies all the conditions together, and
- third, to design a globally convergent algorithm for the continuous problem, using the above results.

3.2 Relations between Wolfe's Conditions

We need first to point out some relations between the scalar products and the norm of the sequence of controls and their canonical extensions to L_2^m.

In any feasible control sequence $\bar{u}_N = \{u_0, u_1, ..., u_{N-1}\} \in \mathcal{L}_N^m$ of (13), each u_j is a vector in $\Re^m$ with euclidean norm:

$$\|u_j\|_m = \sqrt{\langle u_j, u_j\rangle_{\Re^m}} = \sqrt{\|u_j\|_m^2} = \sqrt{\sum_{i=1}^m u_{ji}^2},$$

and hence, we can define the norm of the sequence $\bar{u}_N$ as the ℓ_2-norm:

$$\|\bar{u}_N\|_{\ell_2} = \sqrt{\langle \bar{u}_N, \bar{u}_N\rangle_{\mathcal{L}_N^m}} = \sqrt{\langle \bar{u}_N, \bar{u}_N\rangle_{\Re^{mN}}} = \sqrt{\sum_{j=0}^{N-1} \langle u_j, u_j\rangle_{\Re^m}} =$$

$$= \sqrt{\sum_{j=0}^{N-1} \|u_j\|_m^2} = \sqrt{\sum_{j=0}^{N-1} \sum_{i=1}^{m} u_{ji}^2} = \|\bar{u}_N\|_{mN} ,$$

i.e. the euclidean norm of the vector $(u_{01}, ..., u_{0m}, ..., u_{N-1,1}, ..., u_{N-1.m}) \in \Re^{mN}$.

For any $\bar{u}_N \in \mathcal{L}_N^m$, the L_2^m-norm of the function $\bar{u}_N(.)$ can be calculated:

$$\|\bar{u}_N(.)\|_{L_2^m} = \sqrt{\int_{t_0}^{t_1} \|\bar{u}_N(t)\|_m^2 \, dt} = \sqrt{\sum_{j=0}^{N-1} \int_{\tau_j}^{\tau_{j+1}} \|u_j\|_m^2 \, dt} = \sqrt{h} \, \|\bar{u}_N\|_{mN} ,$$

and for any $v(.) \in L_2^m$, we have:

$$\langle v(.), \bar{u}_N(.)\rangle_{L_2^m} = \int_{t_0}^{t_1} v^{\mathrm{T}}(t)\bar{u}_N(t)dt = \sum_{j=0}^{N-1} \left(\int_{\tau_j}^{\tau_{j+1}} v^{\mathrm{T}}(t)dt \right) u_j.$$

If, for example, $v(.)$ is the function of L_2^m representing the gradient $\nabla J(\bar{u}_N(.))$ of the objective function of the continuous problem (12) at the point $\bar{u}_N(.)$, and since L_∞^m is dense in L_2^m, it should have the following well known formula: (see [17]):

$$\nabla J(\bar{u}_N(.))(t) = -\psi_N^{\mathrm{T}}(t) f_u(x_N(t), \bar{u}_N(t)), \; a.a. \; t \in [t_0.t_1],$$

where $\psi_N(.)$ is the solution of the adjoint differential system:

$$\begin{aligned} \dot{\psi}_N(t) &= -f_x^{\mathrm{T}}(x_N(t), \bar{u}_N(t))\psi_N(t), \; a.a. \; t \in [t_0.t_1], \\ \psi_N(t_1) &= -\varphi_x^{\mathrm{T}}[x_N(t_1)], \end{aligned}$$

and $x_N(.)$ is the continuous trajectory of the problem (12) corresponding to $\bar{u}_N(.)$, then we have:

$$\langle \nabla J(\bar{u}_N(.))(.), \bar{u}_N(.)\rangle_{L_2^m} = -\sum_{j=0}^{N-1} \left(\int_{\tau_j}^{\tau_{j+1}} \psi_N^{\mathrm{T}}(t) f_u(x_N(t), \bar{u}_N(t))dt \right) u_j.$$

Another important example is when $v(.)$ is the constant canonical extension to L_2^m of the vector sequence $\nabla J_N(\bar{u}_N)$, i.e. the vector sequence of the gradient of the discrete problem objective function (13) at the point

$\bar{u}_N = \{u_0, ..., u_{N-1}\}$. The formula for the discrete gradient is also well known (see [17]):

$$\nabla J_N(\bar{u}_N) = \{-h\psi_1^{\mathrm{T}} f_u(x_0, u_0), ..., -h\psi_N^{\mathrm{T}} f_u(x_{N-1}, u_{N-1})\} \in \mathcal{L}_N^m \simeq \Re^{mN},$$

where $\bar{\psi}_N = \{\psi_1, ..., \psi_N\}$ is the solution sequence of the following adjoint difference system of equations:

$$\begin{aligned}
\psi_j &= \psi_{j+1} + hf_x^{\mathrm{T}}(x_j, u_j)\psi_{j+1}, \; j = N-1, N-2, ..., 1, \\
\psi_N &= -\varphi_x^{\mathrm{T}}[x_N],
\end{aligned}$$

and where $\bar{x}_N = \{x_1, ..., x_N\}$ is the discrete trajectory sequence of the problem (13) corresponding to $\bar{u}_N$. In this case we will have:

$$\nabla J_N(\bar{u}_N)(t) = -h\psi_{j+1}^{\mathrm{T}} f_u(x_j, u_j), \quad \text{for } t \in [\tau_j, \tau_{j+1}), \; j = 0, 1, ..., N-1,$$

$$\begin{aligned}
\langle \nabla J_N(\bar{u}_N)(.), \bar{u}_N(.)\rangle_{L_2^m} &= \int_{t_0}^{t_1} \nabla J_N^{\mathrm{T}}(\bar{u}_N)(t)\bar{u}_N(t)dt = \\
&= \sum_{j=0}^{N-1} \left(-\int_{\tau_j}^{\tau_{j+1}} h\psi_{j+1}^{\mathrm{T}} f_u(x_j, u_j)dt \right) u_j = \\
&= -h^2 \sum_{j=0}^{N-1} \psi_{j+1}^{\mathrm{T}} f_u(x_j, u_j)u_j = \\
&= h \langle \nabla J_N(\bar{u}_N), \bar{u}_N \rangle_{\Re^{mN}} .
\end{aligned}$$

In addition, note that if $\bar{u}_{N_1}(.)$ is a piecewise constant fixed function and $\bar{u}_N(.)$ denotes the same function but with N subintervals ($N \geq N_1$) corresponding to the integration step h, and if $h \to 0$ ($N \to \infty$), under standard hypothesis the polygonal canonical extensions $\bar{x}_N[.], \bar{\psi}_N[.]$, and the constant canonical extensions $\bar{x}_N(.), \bar{\psi}_N(.)$, converge pointwise to the respectively continuous solutions $x_{N_1}(.)$ and $\psi_{N_1}(.)$, on $[t_0, t_1]$. Furthermore, by uniform continuity of f_u and Lebesgue's dominated convergence theorem, we have:

$$\lim_{N\to\infty} \int_{t_0}^{t_1} \left\| \psi_{N_1}^{\mathrm{T}}(t) f_u(x_{N_1}(t), \bar{u}_{N_1}(t)) - \bar{\psi}_N^{\mathrm{T}}[t] f_u(\bar{x}_N[t], \bar{u}_N(t)) \right\|^2 dt =$$

$$\lim_{N\to\infty} \int_{t_0}^{t_1} \left\| \psi_{N_1}^{\mathrm{T}}(t) f_u(x_{N_1}(t), \bar{u}_{N_1}(t)) - \bar{\psi}_N^{\mathrm{T}}(t) f_u(\bar{x}_N(t), \bar{u}_N(t)) \right\|^2 dt =$$

$$\lim_{N\to\infty}\sum_{j=0}^{N-1}\left(\int_{\tau_j}^{\tau_{j+1}}\left\|\psi_{N_1}^{\mathrm{T}}(t)f_u(x_{N_1}(t),\bar{u}_{N_1}(t))-\psi_{j+1}^{\mathrm{T}}f_u(x_j,u_j)\right\|^2 dt\right)=$$

$$\int_{t_0}^{t_1}\lim_{N\to\infty}\left\|\psi_{N_1}^{\mathrm{T}}(t)f_u(x_{N_1}(t),\bar{u}_{N_1}(t))-\psi_N^{\mathrm{T}}(t)f_u(\bar{x}_N(t),\bar{u}_N(t))\right\|^2 dt=0.$$

that is,

$$\frac{1}{h}\nabla J_N(\bar{u}_N)(.)\underset{L_2^m}{\rightarrow}\nabla J(\bar{u}_N(.))=\nabla J(\bar{u}_{N_1}(.)),\ \text{for } h\to 0\ (N\to\infty).$$

Thus, we can write the fundamental relations:

$$\begin{array}{l}\|\bar{u}_N(.)\|_{L_2^m}=\sqrt{h}\,\|\bar{u}_N\|_{mN}\,,\\ \langle\nabla J_N(\bar{u}_N)(.),\bar{u}_N(.)\rangle_{L_2^m}=h\,\langle\nabla J_N(\bar{u}_N),\bar{u}_N\rangle_{\Re^{mN}}\ .\\ \left\|\nabla J(\bar{u}_N(.))-\frac{1}{h}\nabla J_N(\bar{u}_N)(.)\right\|_{L_2^m}\underset{N\to\infty}{\rightarrow}0.\end{array}\tag{17}$$

Let's introduce the following notation:

Definition 10 *We consider the problem (12) and its discrete approximation (13) with step h,* $h=\frac{t_1-t_0}{N}$. *For* $\bar{v}(.),\bar{u}(.)\in L_2^m$, *we denote by:*

$$\triangle J(\bar{v}(.),\bar{u}(.))=J(\bar{v}(.))-J(\bar{u}(.)),$$

the objective function increment of the continuous problem (12), respect to the pair $(\bar{v}(.),\bar{u}(.))$. *Analogously, for the increment of the objective function of the discrete problem (13), respect to the pair* $(\bar{v}_N,\bar{u}_N)\in\Re^{mN}\times\Re^{mN}$, *we will use:*

$$\triangle J_N(\bar{v}_N,\bar{u}_N)=J_N(\bar{v}_N)-J_N(\bar{u}_N).$$

We denote by:

$$\eta_N=\eta_N(\bar{v}_N,\bar{u}_N)=|\triangle J(\bar{v}_N(.),\bar{u}_N(.))-\triangle J_N(\bar{v}_N,\bar{u}_N)|$$

the error of the increment of the continuous objective function $J(.)$ *at* $(\bar{v}_N(.),\bar{u}_N(.))$ *respect to the increment of the discrete objective function* $J_N(.)$ *at* $(\bar{v}_N,\bar{u}_N)$ *and by:*

$$\varepsilon_N=\varepsilon_N(\bar{u}_N)=\left\|\nabla J(\bar{u}_N(.))-\frac{1}{h}\nabla J_N(\bar{u}_N)(.)\right\|_{L_2^m}$$

the error (in L_2^m-*norm) of the continuous gradient* $\nabla J(.)$ *at* $\bar{u}_N(.)\in L_2^m$, *respect to its approximating (canonical extension to* L_2^m*) discrete gradient* $\frac{1}{h}\nabla J_N(\bar{u}_N)(.)$.

Lemma 3 *Let be* $\bar{u}_N, \bar{v}_N \in \Re^{mN}$, *then we have the inequalities:*

$$\begin{array}{l} |\triangle J(\bar{v}_N(.), \bar{u}_N(.))| \leq |\triangle J_N(\bar{v}_N, \bar{u}_N)| + \eta_N, \\ |\triangle J_N(\bar{v}_N, \bar{u}_N)| \leq |\triangle J(\bar{v}_N(.), \bar{u}_N(.))| + \eta_N \end{array} \tag{18}$$

$$\begin{array}{l} \frac{1}{h} \left\| \nabla J_N(\bar{u}_N)(.) \right\|_{L_2^m} \leq \left\| \nabla J(\bar{u}_N(.)) \right\|_{L_2^m} + \varepsilon_N, \\ \left\| \nabla J(\bar{u}_N(.)) \right\|_{L_2^m} \leq \frac{1}{h} \left\| \nabla J_N(\bar{u}_N)(.) \right\|_{L_2^m} + \varepsilon_N. \end{array} \tag{19}$$

Proof. Trivial by the triangular inequality. □

We shall find first the relations between the $\rho-$conditions for the problems (12) and (13).

Lemma 4 *Let be* $\rho, \rho_1 \in (0,1)$, $\rho > \rho_1$, *and* $\bar{u}_{N_1} \in \Re^{mN_1}$. *Suppose we choose* $N \geq N_1 \in \aleph$, *large enough, such that the following gradient error condition at* $\bar{u}_N$ *holds:*

$$0 < \varepsilon_N(\bar{u}_N) = \left\| \nabla J(\bar{u}_N(.)) - \frac{1}{h} \nabla J_N(\bar{u}_N)(.) \right\|_{L_2^m} \leq \left(\frac{\rho - \rho_1}{1 + \rho} \right) \left\| \nabla J(\bar{u}_N(.)) \right\|_{L_2^m}. \tag{20}$$

If $\triangle \bar{u}_N \in \Re^{mN}$ *satisfies the* $\rho-$*condition for* $J_N(.)$ *at* $\bar{u}_N$:

$$\langle \nabla J_N(\bar{u}_N), \triangle \bar{u}_N \rangle_{\Re^{mN}} \leq -\rho \left\| \nabla J_N(\bar{u}_N) \right\|_{mN} \left\| \triangle \bar{u}_N \right\|_{mN}, \tag{21}$$

then, $\triangle \bar{u}_N(.) \in L_2^m$ *satisfies the* ρ_1-*condition for* $J(.)$ *at* $\bar{u}_N(.)$:

$$\langle \nabla J(\bar{u}_N(.)), \triangle \bar{u}_N(.) \rangle_{L_2^m} \leq -\rho_1 \left\| \nabla J(\bar{u}_N(.)) \right\|_{L_2^m} \left\| \triangle \bar{u}_N(.) \right\|_{L_2^m}. \tag{22}$$

Proof. Using (17) and Schwarz's inequality we have:

$$\begin{array}{l} \langle \nabla J(\bar{u}_N(.)), \triangle \bar{u}_N(.) \rangle_{L_2^m} = \\ = \left\langle \frac{1}{h} \nabla J_N(\bar{u}_N)(.), \triangle \bar{u}_N(.) \right\rangle_{L_2^m} + \left\langle \nabla J(\bar{u}_N(.)) - \frac{1}{h} \nabla J_N(\bar{u}_N)(.), \triangle \bar{u}_N(.) \right\rangle_{L_2^m} \\ \leq \langle \nabla J_N(\bar{u}_N), \triangle \bar{u}_N \rangle_{\Re^{mN}} + \left\| \nabla J(\bar{u}_N(.)) - \frac{1}{h} \nabla J_N(\bar{u}_N)(.) \right\|_{L_2^m} \left\| \triangle \bar{u}_N(.) \right\|_{L_2^m}, \end{array}$$

with the assumption (21) and (17) we also have:

$$\begin{array}{l} \langle \nabla J(\bar{u}_N(.)), \triangle \bar{u}_N(.) \rangle_{L_2^m} \leq \\ \leq -\rho \left\| \nabla J_N(\bar{u}_N) \right\|_{mN} \left\| \triangle \bar{u}_N \right\|_{mN} + \varepsilon_N \left\| \triangle \bar{u}_N(.) \right\|_{L_2^m} \\ \leq -\rho \frac{1}{h} \left\| \nabla J_N(\bar{u}_N)(.) \right\|_{L_2^m} \left\| \triangle \bar{u}_N(.) \right\|_{L_2^m} + \varepsilon_N \left\| \triangle \bar{u}_N(.) \right\|_{L_2^m}. \end{array}$$

Now we use inequality (19) from Lemma 3 and assumption (20), and then:

$$\begin{aligned}
\langle \nabla J(\bar{u}_N(.)), \triangle \bar{u}_N(.) \rangle_{L_2^m} &\leq -\rho \left(\|\nabla J(\bar{u}_N(.))\|_{L_2^m} - \varepsilon_N \right) \|\triangle \bar{u}_N(.)\|_{L_2^m} + \varepsilon_N \|\triangle \bar{u}_N(.)\|_{L_2^m} \\
&\leq -\rho \|\nabla J(\bar{u}_N(.))\|_{L_2^m} \|\triangle \bar{u}_N(.)\|_{L_2^m} + (1+\rho)\varepsilon_N \|\triangle \bar{u}_N(.)\|_{L_2^m} \\
&\leq -\rho_1 \|\nabla J(\bar{u}_N(.))\|_{L_2^m} \|\triangle \bar{u}_N(.)\|_{L_2^m} . \ \square
\end{aligned}$$

The next Lemma deals with the $\beta-$condition.

Lemma 5 *Let be $\rho, \rho_1, \beta \in (0,1)$, such that $\gamma = \frac{1}{\rho_1}\left(\frac{\rho-\rho_1}{1+\rho}\right) \in (0,1)$, and*

$$\frac{1-\beta}{1+\beta} > \gamma. \tag{23}$$

Let be $\bar{u}_{N1} \in \Re^{mN_1}$ and let's suppose we choose $N \geq N_1$ such that the gradient error condition at $\bar{u}_N$ (20) holds.

Let $\triangle \bar{u}_N \in \Re^{mN}$ be a descent direction which satisfies the $\rho-$condition (21), the $\beta-$condition for $J_N(.)$ at the new point $\bar{v}_N = \bar{u}_N + \triangle \bar{u}_N :$

$$\langle \nabla J_N(\bar{v}_N), \triangle \bar{u}_N \rangle_{\Re^{mN}} \geq \beta \langle \nabla J_N(\bar{u}_N), \triangle \bar{u}_N \rangle_{\Re^{mN}} \ , \tag{24}$$

and the gradient error condition at $\bar{v}_N = \bar{u}_N(.) + \triangle \bar{u}_N(.) :$

$$\begin{aligned}
\varepsilon_N(\bar{v}_N) = \left\| \nabla J(\bar{v}_N) - \tfrac{1}{h} \nabla J_N(\bar{v}_N)(.) \right\|_{L_2^m} \leq \\
\leq \left(\tfrac{\rho-\rho_1}{1+\rho}\right) \|\nabla J(\bar{u}_N(.))\|_{L_2^m} .
\end{aligned} \tag{25}$$

Then the direction $\triangle \bar{u}_N(.) \in L_2^m$ satisfies the β_1-condition for $J(.)$ at the new point $\bar{v}_N(.) = \bar{u}_N(.) + \triangle \bar{u}_N(.) :$

$$\langle \nabla J(\bar{v}_N(.)), \triangle \bar{u}_N(.) \rangle_{L_2^m} \geq \beta_1 \langle \nabla J(\bar{u}_N(.)), \triangle \bar{u}_N(.) \rangle_{L_2^m} , \tag{26}$$

for all $\beta_1 \in (0,1)$ satisfying:

$$\frac{\beta_1 - \beta}{1+\beta} \geq \gamma. \tag{27}$$

Proof. With the Schwarz inequality and (17) we have:

$$\begin{aligned}&\langle \nabla J(\bar{v}_N(.)), \triangle \bar{u}_N(.)\rangle_{L_2^m} = \\ &= \left\langle \tfrac{1}{h}\nabla J_N(\bar{v}_N)(.), \triangle \bar{u}_N(.)\right\rangle_{L_2^m} + \left\langle \nabla J(\bar{v}_N(.)) - \tfrac{1}{h}\nabla J_N(\bar{v}_N)(.), \triangle \bar{u}_N(.)\right\rangle_{L_2^m} \geq \\ &\geq \langle \nabla J_N(\bar{v}_N), \triangle \bar{u}_N\rangle_{\Re^{mN}} - \left\| \nabla J(\bar{v}_N(.)) - \tfrac{1}{h}\nabla J_N(\bar{v}_N)(.)\right\|_{L_2^m} \|\triangle \bar{u}_N(.)\|_{L_2^m}.\end{aligned}$$

Using the assumptions (24) and (25), Schwarz's inequality and (17), we can write:

$$\begin{aligned}&\langle \nabla J(\bar{v}_N(.)), \triangle \bar{u}_N(.)\rangle_{L_2^m} \geq \beta \langle \nabla J_N(\bar{u}_N), \triangle \bar{u}_N\rangle_{\Re^{mN}} - \varepsilon_N \|\triangle \bar{u}_N(.)\|_{L_2^m} \\ &\geq \beta \tfrac{1}{h} \langle \nabla J_N(\bar{u}_N)(.), \triangle \bar{u}_N(.)\rangle_{L_2^m} - \varepsilon_N \|\triangle \bar{u}_N(.)\|_{L_2^m} \\ &\geq \beta \langle \nabla J(\bar{u}_N(.)), \triangle \bar{u}_N(.)\rangle_{L_2^m} + \beta \left\langle \nabla J_N(\bar{u}_N)(.) - \tfrac{1}{h}\nabla J(\bar{u}_N(.)), \triangle \bar{u}_N(.)\right\rangle_{L_2^m} - \\ &\qquad -\varepsilon_N \|\triangle \bar{u}_N(.)\|_{L_2^m} \geq \\ &\geq \beta \langle \nabla J(\bar{u}_N(.)), \triangle \bar{u}_N(.)\rangle_{L_2^m} - (1+\beta)\varepsilon_N \|\triangle \bar{u}_N(.)\|_{L_2^m}.\end{aligned}$$

¿From (20), (21) and (17), we obtain the inequalities:

$$\begin{aligned}&\langle \nabla J(\bar{v}_N(.)), \triangle \bar{u}_N(.)\rangle_{L_2^m} \geq \\ &\geq \beta \langle \nabla J(\bar{u}_N(.)), \triangle \bar{u}_N(.)\rangle_{L_2^m} - (1+\beta)\left(\tfrac{\rho-\rho_1}{1+\rho}\right) \|\nabla J(\bar{u}_N(.))\|_{L_2^m} \|\triangle \bar{u}_N(.)\|_{L_2^m} \\ &\geq \beta \langle \nabla J(\bar{u}_N(.)), \triangle \bar{u}_N(.)\rangle_{L_2^m} + (1+\beta)\left(\tfrac{\rho-\rho_1}{1+\rho}\right) \tfrac{1}{\rho_1} \langle \nabla J(\bar{u}_N(.)), \triangle \bar{u}_N(.)\rangle_{L_2^m}\end{aligned}$$

$$\Rightarrow \langle \nabla J(\bar{v}_N(.)), \triangle \bar{u}_N(.)\rangle_{L_2^m} \geq [\beta + (1+\beta)\gamma] \langle \nabla J(\bar{u}_N(.)), \triangle \bar{u}_N(.)\rangle_{L_2^m}. \quad (28)$$

¿From (28) and (27) we obtain the β_1-condition (26). Note that the assumption (23) allows to choose $\beta_1 \in (0,1)$. □

The $\alpha-$condition is the subject of the next Lemma.

Lemma 6 *Let be $\rho, \rho_1, \alpha \in (0,1)$, such that $\gamma = \frac{1}{\rho_1}\left(\frac{\rho-\rho_1}{1+\rho}\right) \in (0, \frac{1}{2})$, and:*

$$\frac{\alpha}{1+\alpha} > \gamma. \quad (29)$$

Let be $\bar{u}_{N1} \in \Re^{mN_1}$, and suppose we choose $N \geq N_1$ such that the gradient error condition at $\bar{u}_N$ (20) holds.

Let $\triangle\bar{u}_N \in \Re^{mN}$ be a direction satisfying the $\rho-$ condition (21) for $J_N(.)$ at $\bar{u}_N$, the $\alpha-$condition for $J_N(.)$ at the new point $\bar{v}_N = \bar{u}_N + \triangle\bar{u}_N$:

$$J_N(\bar{v}_N) \leq J_N\left(\bar{u}_N\right) + \alpha\left\langle \bigtriangledown J_N\left(\bar{u}_N\right), \triangle\bar{u}_N\right\rangle_{\Re^{mN}}, \tag{30}$$

and the following increment error condition at the new point $\bar{v}_N = \bar{u}_N + \triangle\bar{u}_N$:

$$\eta_N(\bar{v}_N, \bar{u}_N) \leq \left(\frac{\rho-\rho_1}{1+\rho}\right)\left\|\nabla J(\bar{u}_N(.))\right\|_{L_2^m}\left\|\triangle\bar{u}_N\right\|_{L_2^m}. \tag{31}$$

Then $\triangle\bar{u}_N(.) \in L_2^m$ satisfies the α_1-condition for $J(.)$ at the new point $\bar{v}_N(.) = \bar{u}_N(.) + \triangle\bar{u}_N(.)$:

$$J(\bar{v}_N(.)) \leq J(\bar{u}_N(.)) + \alpha_1\left\langle \bigtriangledown J(\bar{u}_N(.)), \triangle\bar{u}_N(.)\right\rangle_{L_2^m}, \tag{32}$$

for all $\alpha_1 \in (0,1)$ satisfying:

$$\frac{\alpha-\alpha_1}{1+\alpha} \geq \gamma. \tag{33}$$

Proof. Put $\xi_N = \left(\frac{\rho-\rho_1}{1+\rho}\right)\left\|\nabla J(\bar{u}_N(.))\right\|_{L_2^m}$. From inequality (31) and the $\alpha-$condition (30), we have:

$$\begin{aligned}\triangle J(\bar{v}_N(.), \bar{u}_N(.)) &= J(\bar{v}_N(.)) - J(\bar{u}_N(.)) \leq \triangle J_N(\bar{v}_N, \bar{u}_N) + \eta_N(\bar{v}_N, \bar{u}_N) \\ &\leq \alpha\left\langle \bigtriangledown J_N\left(\bar{u}_N\right), \triangle\bar{u}_N\right\rangle_{\Re^{mN}} + \xi_N\left\|\triangle\bar{u}_N\right\|_{L_2^m},\end{aligned}$$

and using (17), Schwarz's inequality and (20), we obtain:

$$\triangle J(\bar{v}_N(.), \bar{u}_N(.)) \leq \alpha\frac{1}{h}\left\langle \bigtriangledown J_N(\bar{u}_N)(.), \triangle\bar{u}_N(.)\right\rangle_{L_2^m} + \xi_N\left\|\triangle\bar{u}_N(.)\right\|_{L_2^m}$$

$$\leq \alpha\left\langle \bigtriangledown J(\bar{u}_N(.)), \triangle\bar{u}_N(.)\right\rangle_{L_2^m} + \alpha\left\|\tfrac{1}{h}\bigtriangledown J_N(\bar{u}_N)(.) - \bigtriangledown J(\bar{u}_N(.))\right\|_{L_2^m}\left\|\triangle\bar{u}_N(.)\right\|_{L_2^m} + \\ +\xi_N\left\|\triangle\bar{u}_N(.)\right\|_{L_2^m} \leq$$

$$\begin{aligned}&\leq \alpha\left\langle \bigtriangledown J(\bar{u}_N(.)), \triangle\bar{u}_N(.)\right\rangle_{L_2^m} + \alpha\varepsilon_N\left\|\triangle\bar{u}_N(.)\right\|_{L_2^m} + \xi_N\left\|\triangle\bar{u}_N(.)\right\|_{L_2^m} \\ &\leq \alpha\left\langle \bigtriangledown J(\bar{u}_N(.)), \triangle\bar{u}_N(.)\right\rangle_{L_2^m} + (\alpha+1)\,\xi_N\left\|\triangle\bar{u}_N(.)\right\|_{L_2^m}.\end{aligned}$$

¿From definition and the ρ_1-condition (22), given by the application of Lemma 4, we have:

$$\begin{aligned}\triangle J(\bar{v}_N(.), \bar{u}_N(.)) \; & \leq \; \alpha \left\langle \bigtriangledown J(\bar{u}_N(.)), \triangle \bar{u}_N(.) \right\rangle_{L_2^m} + \\ & + (\alpha+1) \left(\frac{\rho - \rho_1}{1+\rho} \right) \left\| \nabla J(\bar{u}_N(.)) \right\|_{L_2^m} \left\| \triangle \bar{u}_N(.) \right\|_{L_2^m},\end{aligned}$$

$$\Rightarrow \quad \triangle J(\bar{v}_N(.), \bar{u}_N(.)) \leq \alpha \left\langle \bigtriangledown J(\bar{u}_N(.)), \triangle \bar{u}_N(.) \right\rangle_{L_2^m} - \\ - (\alpha+1) \left(\frac{\rho - \rho_1}{1+\rho} \right) \frac{1}{\rho_1} \left\langle \bigtriangledown J(\bar{u}_N(.)), \triangle \bar{u}_N(.) \right\rangle_{L_2^m},$$

and finally, we have the inequalities:

$$J(\bar{v}_N(.)) - J(\bar{u}_N(.)) \leq [\alpha - (\alpha+1)\,\gamma] \left\langle \bigtriangledown J(\bar{u}_N(.)), \triangle \bar{u}_N(.) \right\rangle_{L_2^m}.$$

¿From (33) we deduce the α_1-condition (32) for $J(.)$ at $\bar{v}_N(.)$. Note that the assumptions (29) and (33) allow us to choose $\alpha, \alpha_1 \in (0,1)$. □

3.3 Global convergence results and Algorithm

The Lemmas 4 to 6 give us quite important relations between the $\rho-, \alpha-$ and $\beta-$conditions for the continuous and discrete problems, but the principal question remains open:

- if we have a new point $\bar{v}_N \in \Re^{mN}$ which satisfies all the conditions for the discrete problem, the corresponding constant canonical extension $\bar{v}_N(.)$ to L_2^m also satisfies analogous conditions for the continuous problem?, or more precisely:

- for a parameter triple (ρ, α, β), with $\alpha < \beta$, and for a new point $\bar{v}_N \in \Re^{mN}$ which satisfies the $\rho-, \alpha-$, and $\beta-$ conditions, under which assumptions there exists a parameter triple $(\rho_1, \alpha_1, \beta_1)$, with $\alpha_1 < \beta_1$, such that the constant canonical extension $\bar{v}_N(.)$ to L_2^m satisfies the ρ_1-, α_1- and β_1- conditions?.

The following theorem offers a satisfactory answer:

Theorem 4 *Let be $u_{N_1} \in \Re^{mN_1}$, $\rho, \alpha, \beta \in (0,1)$, $\beta > \alpha$. Choose $\varepsilon \in (0,1)$ satisfying:*

$$\varepsilon < \min \left\{ \frac{\rho}{2+\rho}, \frac{\rho(1-\beta)}{2+\rho(1-\beta)}, \frac{\rho\alpha}{1+\alpha(2+\rho)} \right\}, \tag{34}$$

and suppose we take $N \geq N_1$ large enough such:

$$\varepsilon_N(\bar{u}_N) = \left\| \nabla J(\bar{u}_N) - \frac{1}{h} \nabla J_N(\bar{u}_N)(.) \right\|_{L_2^m} \leq \varepsilon \left\| \nabla J(\bar{u}_N(.)) \right\|_{L_2^m}, \qquad (35)$$

Denote $\bar{v}_N = \bar{u}_N + \triangle \bar{u}_N$ and let's suppose the direction $\triangle \bar{u}_N \in \Re^{mN}$ is chosen such that the following inequalities hold:

$$\varepsilon_N(\bar{v}_N) = \left\| \nabla J(\bar{v}_N) - \frac{1}{h} \nabla J_N(\bar{v}_N)(.) \right\|_{L_2^m} \leq \varepsilon \left\| \nabla J(\bar{u}_N(.)) \right\|_{L_2^m}, \qquad (36)$$

$$\eta_N(\bar{v}_N, \bar{u}_N) = \left| \triangle J(\bar{v}_N(.), \bar{u}_N(.)) - \triangle J_N(\bar{v}_N, \bar{u}_N) \right| \leq \varepsilon \left\| \nabla J(\bar{u}_N(.)) \right\|_{L_2^m} \left\| \triangle \bar{u}_N \right\|_{L_2^m}, \qquad (37)$$

and the $\rho-, \beta-$ and $\alpha-$conditions (21), (24), (30) respectively are satisfied for the discrete problem (13).

Then, there exists numbers $\rho_1, \alpha_1, \beta_1 \in (0,1)$, $\beta_1 > \alpha_1$ such that the direction $\triangle \bar{u}_N(.) \in L_2^m$ satisfies the ρ_1-, β_1- and α_1-conditions (22), (26), (32) respectively, for the continuous problem (12).

Furthermore, for given $\rho, \alpha, \beta \in (0,1)$, $\beta > \alpha$, and $\varepsilon \in (0,1)$ satisfying (34), we can choose the parameters $\rho_1, \alpha_1, \beta_1$ from the following intervals:

$$\rho_1 \in \left(\hat{\rho}_1, \rho - \varepsilon(1+\rho)\right), \qquad (38)$$

where:

$$\hat{\rho}_1 = \max \left\{ \frac{\rho(1+\beta)}{2+\rho-\beta\rho}, \frac{\rho(1+\alpha)}{1+2\alpha+\alpha\rho} \right\}, \qquad (39)$$

$$\beta_1 \in \left(\beta + \gamma(1+\beta), 1\right), \qquad (40)$$

where:

$$\gamma = \frac{1}{\rho_1} \left(\frac{\rho - \rho_1}{1+\rho} \right),$$

and:

$$\alpha_1 \in \left(0, \alpha - \gamma(1+\alpha)\right). \qquad (41)$$

Proof. From definition of ε we have, first:

$$\varepsilon < \frac{\rho}{2+\rho} \Leftrightarrow \varepsilon < \frac{\rho(1+\rho)}{(2+\rho)(1+\rho)} \Leftrightarrow \varepsilon(1+\rho) < \frac{\rho(1+\rho)}{(2+\rho)} \Leftrightarrow$$

$$\varepsilon(1+\rho) < \rho\left[1-\frac{1}{2+\rho}\right] \Leftrightarrow$$

$$\rho-\varepsilon(1+\rho) > \frac{\rho}{2+\rho} > 0,$$

second:

$$\varepsilon < \frac{\rho(1-\beta)}{2+\rho(1-\beta)} \Leftrightarrow \varepsilon(1+\rho) < \frac{\rho(1-\beta)(1+\rho)}{2+\rho(1-\beta)} \Leftrightarrow$$

$$\varepsilon(1+\rho) < \rho\left[\frac{1+\rho-\beta-\beta\rho}{2+\rho-\beta\rho}\right] \Leftrightarrow \varepsilon(1+\rho) < \rho\left[1-\frac{1+\beta}{2+\rho-\beta\rho}\right] \Leftrightarrow$$

$$\rho-\varepsilon(1+\rho) > \frac{\rho(1+\beta)}{2+\rho-\beta\rho} > 0,$$

and third:

$$\varepsilon < \frac{\rho\alpha}{1+\alpha(2+\rho)} \Leftrightarrow \varepsilon(1+\rho) < \frac{\rho\alpha(1+\rho)}{1+\alpha(2+\rho)} \Leftrightarrow$$

$$\varepsilon(1+\rho) < \rho\left[1-\frac{1+\alpha}{1+\alpha(2+\rho)}\right] \Leftrightarrow$$

$$\rho-\varepsilon(1+\rho) > \frac{\rho(1+\alpha)}{1+2\alpha+\alpha\rho} > 0.$$

Then, it is possible to choose $\rho_1 \in (0,1)$ satisfying (38) and (39), i.e.

$$\rho_1 < \rho-\varepsilon(1+\rho) \Leftrightarrow \varepsilon < \left(\frac{\rho-\rho_1}{1+\rho}\right), \tag{42}$$

$$\rho_1 > \frac{\rho(1+\beta)}{2+\rho-\beta\rho}, \tag{43}$$

$$\rho_1 > \frac{\rho(1+\alpha)}{1+2\alpha+\alpha\rho}, \tag{44}$$

and therefore, from (42), (35), (36) and (37) the essential assumptions (20), (25) and (31) of Lemmas 4, 5 and 6 hold. Hence, from Lemma 4 $\triangle\bar{u}_N(.)$ satisfies the ρ_1-condition (22).

Furthermore, from (43) we have:

$$\rho_1\left[\frac{2+\rho-\beta\rho}{(1+\beta)(1+\rho)}\right] > \frac{\rho}{1+\rho} \Leftrightarrow \rho_1\left[\frac{1-\beta}{1+\beta}+\frac{1}{1+\rho}\right] > \frac{\rho}{1+\rho} \Leftrightarrow$$

$$\frac{1-\beta}{(1+\beta)} > \frac{1}{\rho_1}\left(\frac{\rho-\rho_1}{1+\rho}\right) = \gamma,$$

and then, from Lemma 5 $\triangle \bar{u}_N(.)$ satisfies the β_1-condition (26) for all β_1 satisfying (40).

Finally, from (44) we have:

$$\rho_1\left[\frac{1+2\alpha+\alpha\rho}{(1+\alpha)(1+\rho)}\right] > \frac{\rho}{1+\rho} \Leftrightarrow \rho_1\left[\frac{\alpha}{1+\alpha}+\frac{1}{1+\rho}\right] > \frac{\rho}{1+\rho} \Leftrightarrow$$

$$\frac{\alpha}{1+\alpha} > \frac{1}{\rho_1}\left(\frac{\rho-\rho_1}{1+\rho}\right) = \gamma,$$

and then, from Lemma 6 $\triangle \bar{u}_N(.)$ satisfies the α_1-condition (32) for all α_1 satisfying (41). □

Remark 1 *Note that if $\varepsilon \to 0$ we can take $\rho_1 \to \rho$ and therefore $\gamma \to 0$, then we can take $\alpha_1 \to \alpha$ and $\beta_1 \to \beta$. This reflects the natural fact that, for $h \to 0$, the iteration in the continuous problem mimic the iteration in the discrete problem.*

Remark 2 *Some instances of possible values for the parameters can be calculated from the corresponding formulas. Some examples follow for illustration:*

$$\textit{For: } \rho = \frac{1}{2},\ \alpha = \frac{1}{10},\ \beta = \frac{9}{10},\ \textit{ we can choose:}$$

$$\varepsilon \le \frac{1}{50},\ \rho_1 = \frac{77}{164},\ \alpha_1 \le \frac{5}{100},\ \beta_1 \ge \frac{99}{100},\ \textit{ where } \gamma = \frac{10}{231},$$

$$\textit{or } \varepsilon \le \frac{1}{100},\ \rho_1 = \frac{39}{82},\ \alpha_1 \le \frac{5}{100},\ \beta_1 \ge \frac{97}{100},\ \textit{ where } \gamma = \frac{4}{117}.$$

$$\textit{For: } \rho = \frac{75}{100},\ \alpha = \frac{1}{100},\ \beta = \frac{7}{10},\ \textit{ we can choose:}$$

$$\varepsilon \le \frac{5}{1000},\ \rho_1 = \frac{74}{100},\ \alpha_1 \le \frac{1}{1000},\ \beta_1 \ge \frac{9}{10},\ \textit{ where } \gamma = \frac{2}{259},$$

Discrete approximation algorithm (Daa) *for the problem* (12) *with* $U = \Re^m$:

0) Choose $N_{inic} \in \aleph, \bar{u}_{N_{inic}} \in \Re^{mN_{inic}},\ \rho, \alpha, \beta, \hat{\varepsilon}, \hat{\rho} \in (0,1)$, such that: $\beta > \alpha,\ \hat{\varepsilon} < \min\left\{\frac{\rho}{2+\rho}, \frac{\rho(1-\beta)}{2+\rho(1-\beta)}, \frac{\rho\alpha}{1+\alpha(2+\rho)}\right\},\ \hat{\rho} \geq \max\left\{\frac{\rho(1+\beta)}{2+\rho-\beta\rho}, \frac{\rho(1+\alpha)}{1+2\alpha+\alpha\rho}\right\}$,

1) Take $\varepsilon_0 \in (0, \hat{\varepsilon}),\ \rho_1 \in (\hat{\rho}, \rho - \varepsilon(1+\rho)),\ \beta_1 \in (\beta + \gamma(1+\beta), 1)$, $\alpha_1 \in (0, \alpha - \gamma(1+\alpha))$, where $\gamma = \frac{1}{\rho_1}\left(\frac{\rho-\rho_1}{1+\rho}\right)$.

2) Put $\bar{N}_0 = N_{inic},\ u_0(.) = \bar{u}_{\bar{N}_0}(.) \in L_2$ and $l = 0$.

3) Set $N_0 \geq \bar{N}_l$, such that:

$$h_0 = \frac{t_1 - t_0}{N_0} \in (0,1),$$

$$h_0 < \varepsilon_0.$$

4) Compute the gradient $\nabla J(\bar{u}_{N_l}(.)) = \nabla J(u_l(.))$.

- If $\|\nabla J(\bar{u}_{N_l}(.))\|_{L_2} = 0$, stop and take $u_l(.)$ as a local minima.
- If $\|\nabla J(\bar{u}_{N_l}(.))\|_{L_2} > 0$, put $k = 0$ and go to step 5)

5) Repeat for $k = 1, 2, \ldots$

5.0) Put $\varepsilon_k = \frac{\varepsilon_{k-1}}{2}$.

5.1) Verify the inequality:

$$\left\|\nabla J(\bar{u}_{N_k}(.)) - \frac{1}{h}\nabla J_{N_k}(\bar{u}_{N_k})(.)\right\|_{L_2} < \varepsilon_k \left\|\nabla J(\bar{u}_{N_k}(.))\right\|_{L_2}.$$

- If it is satisfied then go to step 5.2).
- If not, choose a new $N_k^{'} > N_k$ such that the inequality holds, set $N_k = N_k^{'}$ and go to step 5.2).

5.2) Repeat for $r = 1, 2, \ldots$,rmax

5.2a) Denote $\Sigma_r \triangle\bar{u}_{N_k} = \Sigma_{r-1}\triangle\bar{u}_{N_k} + \triangle\bar{u}_{N_k}$, with $\Sigma_0\triangle\bar{u}_{N_k} = 0$.

5.2b) Choose a direction $\triangle\bar{u}_{N_k} \in \Re^{mN_k}$ which satisfies the following $\rho-, \alpha-$ and $\beta-$ conditions for $J_{N_k}(.)$ at $\bar{u}_{N_k}$, where $\bar{v}_{N_k} = \bar{u}_{N_k} + \Sigma_r\triangle\bar{u}_{N_k}$:

$$\begin{aligned}&\langle\nabla J_{N_k}(\bar{u}_{N_k}), \Sigma_r\triangle\bar{u}_{N_k}\rangle_{\Re^{mN_k}} \leq \\ &\leq -\rho\left\|\nabla J_{N_k}(\bar{u}_{N_k})\right\|_{mN_k}\left\|\Sigma_r\triangle\bar{u}_{N_k}\right\|_{mN_k},\end{aligned}$$

$$J_{N_k}(\bar{v}_{N_k}) \leq J_{N_k}(\bar{u}_{N_k}) + \alpha\left\langle\nabla J_{N_k}(\bar{u}_{N_k}), \Sigma_r\triangle\bar{u}_{N_k}\right\rangle_{\Re^{mN_k}},$$

$$\left\langle\nabla J_{N_k}(\bar{v}_{N_k}), \Sigma_r\triangle\bar{u}_{N_k}\right\rangle_{\Re^{mN_k}} \geq \beta\left\langle\nabla J_{N_k}(\bar{u}_{N_k}), \Sigma_r\triangle\bar{u}_{N_k}\right\rangle_{\Re^{mN_k}},$$

5.2c) Verify if the following inequalities hold:

$$\left\| \nabla J(\bar{v}_{N_k}(.)) - \frac{1}{h} \nabla J_{N_k}(\bar{v}_{N_k})(.) \right\|_{L_2} \leq \varepsilon_k \left\| \nabla J(\bar{u}_{N_k}(.)) \right\|_{L_2},$$

$$\begin{array}{l} |\triangle J(\bar{v}_{N_k}(.), \bar{u}_{N_k}(.)) - \triangle J_{N_k}(\bar{v}_{N_k}, \bar{u}_{N_k})| \leq \\ \leq \varepsilon_k \left\| \nabla J(\bar{u}_{N_k}(.)) \right\|_{L_2} \left\| \Sigma_r \triangle \bar{u}_{N_k} \right\|_{mN_k}. \end{array}$$

- If they are satisfied, put:

$$\begin{array}{rcl} u_{l+1}(.) & = & \bar{u}_{N_k}(.) + \Sigma_r \triangle \bar{u}_{N_k}(.), \\ \bar{N}_{l+1} & = & N_k, \ l+1 = l, \end{array}$$

and go to step 3).

- If not, and
  - if $r <$ rmax, go to step 5.2a)
  - if $r \geq$ rmax, put $k = k+1$, and go to step 5.0).

Remark 3 *In step 5.2b) we can proceed in two ways:*

- find first a direction satisfying the $\rho-$condition and then perform a usual line search to satisfy the Wolfe $\alpha-$ and $\beta-$conditions, or

- iterate in the $m \times N_k-$dimensional discrete problem (13) and use the $\rho-, \alpha-$ and $\beta-$ conditions as a stopping rule.

Remark 4 *Algorithm (Daa) contains three cycles in l, k and r. Index l corresponds to the main iteration, which find the control function sequence. Index k reduces the error that is accepted in the approximation of the continuous gradient and of the continuous increment by they corresponding discrete approximations. Index r iterates inside the discrete problem looking for a convenient overall step. This last cycle produces a multidirectional search in the continuous problem.*

Theorem 5 *We consider the unconstrained optimal control problem (12), with $U = \Re^m$, and its discrete approximation (13). Suppose that for all $N_1 \in \aleph$ and all $\xi > 0$ there exists $K \in \aleph$ such that, for all $k \geq K$ and $N_k = 2^k N_1$, the following inequalities hold:*

$$\left\| \nabla J(\bar{u}_{N_k}(.)) - \frac{1}{h} \nabla J_{N_k}(\bar{u}_{N_k})(.) \right\|_{L_2^m} < \xi,$$

$$|\triangle J(\bar{v}_{N_k}(.), \bar{u}_{N_k}(.)) - \triangle J_{N_k}(\bar{v}_{N_k}, \bar{u}_{N_k})| < \xi \left\| \bar{v}_{N_k} - \bar{u}_{N_k} \right\|_{mN_k},$$

for any $\bar{v}_{N_1}, \bar{u}_{N_1} \in \Re^{mN_1}$ belonging to the ball:

$$\bar{B}\left(0, \sqrt{mN_1^3}\right) = \left\{ \bar{w}_{N_1} \in \Re^{mN_1} \mid \|\bar{w}_{N_1}\|_{mN_1} \leq \sqrt{mN_1^3} \right\}.$$

Then, the algorithm (Daa) is globally convergent, i.e. the sequence $\{u_l\}$ is finite and its last element is a (piecewise constant) local minima or every accumulation point of the sequence is a (Riemann integrable) local minima.

Proof. Is a consequence of Theorem 4 and Theorem 3. □

3.4 Final Remarks

1) The assumptions in theorem 5 need the L_2-convergence of the discrete gradient to the continuous gradient, at every piecewise constant function of L_2, when the step size $h \to 0$. Furthermore, the convergence of the increments of the discrete and continuous objective function, at the same piecewise constant L_2-functions, is also needed. They are natural hypothesis, and it is not very hard to get conditions about $J(.)$ such that these assumptions hold.

2) Some generalizations would also be desirable:

a) The use of other types of numerical integration, instead of Euler scheme, with different partitions for u and x.

b) The inclusion of constraints, for example, $U \neq \Re^m$, and others.

ACKNOWLEDGMENTS

The results in this paper were partially obtained during the stay of the first author at Humboldt University in Berlin, from November 1997 to January 1998, with a grant of DAAD Organization. The first author wish to thank also to Prof. W. Roemisch, from Humboldt University, for valuable discussions and advice.

References

[1] Alt, W., On the approximation of infinite dimensional optimization problems with an application to optimal control problems. Journal on Computational Optimization and Applications, 2 (1993), 77-100.

[2] ——, Discretization and mesh-independence of Newton's methods for generalized equations, in Mathematical Programming with Data Perturbation. A. V. Fiacco, Eds. Marcel Dekker Inc., New York (1998).

[3] Cuesta, L.E.; M. Goldbrich, Convergence rates of discretization for constrained nonlinear optimal control problems with C^1 data. Technical report No. 284, University of California, Santa Barbara (1994).

[4] Daniel, J.W., On the convergence of a numerical method for optimal control problems, Journal of Optimization Theory and Applications, 4 (1969), 330-342.

[5] Dennis, J.E., R.B. Schnabel, Numerical methods for Unconstrained Optimization and Nonlinear Equations, Prentice Hall, Inc., New Jersey (1983).

[6] Dontchev, A.L. Discrete approximation in optimal control, in Nonsmooth Analysis and Geometric Methods in Optimal Control, B.S. Morkhudovich, R.T. Rockafellar and H. Sussman, Eds., Proceedings IMA, Springer Verlag (1997).

[7] Dontchev, A.L., W.W. Hager, Lipschitz stability in nonlinear control and optimization, SIAM Journal on Control and Optimization, 31 (1993), 569-603.

[8] Evtuschenko, Yu. G., Numerical Optimization Techniques, Optimization Software Inc., New York (1985).

[9] Fiacco, A.V., Introduction to Sensitivity and Stability Analysis in Nonlinear Programming. Academic Press (1993).

[10] Gill, P.E., W. Murray, M.H. Wright, Practical Optimization. Academic Press, London (1983).

[11] Gómez, J.A.; W. Gómez, On multidirectional search in Optimization. Research Report ICIMAF 96-26, (1996).

[12] Hager, W.W., The Ritz-Troffitz method for state and control constrained optimal control problems. SIAM Journal on Numerical Analysis, 12 (1975), 854-867.

[13] Luenberger, D., Introduction to Linear and Nonlinear Programming, Addison-Wesley, Massachussets (1984).

[14] Malanowski, K. K., Finite difference approximation to constrained optimal control problems, in Optimization and Optimal Control, A. Auslander, W. Oettli and J. Stoer, Eds., Lecture Notes in Control and Information Sciences, Vol. 30, Berlin (1981), 243-254.

[15] ——, C. Buskel, H. Maurer, Convergence of approximation to nonlinear optimal control problems, in Mathematical Programming with Data Perturbation. A. V. Fiacco, Eds. Marcel Dekker Inc., New York (1998).

[16] Mordukhovich, B., On difference approximations of optimal control systems, Journal of Applied Mathematics and Mechanics, 42 (1979), 452-461.

[17] Polak, E., Computational Methods in Optimization: A Unified Approach. Academic Press, New York (1971).

[18] Richter, C., B. Renner, Efficient methods for solving optimal control problems, in Numerical Methods of Nonlinear Programming and their Implementations, C. Richter, H. Hollatz and D. Pallaschke, Eds., Academie Verlag, Berlin (1991).

[19] Sachs, E.W., Topics on modern computational methods for optimal control problems, in Modern Methods of Optimization, W. Krabs and J. Zowe, Eds., Springer Verlag, Berlin (1992), 313-348.

[20] Teo, K.L., C.J. Goh, K.H. Wong, A Unified Computational Approach to Optimal Control Problems. Pitman monographs and survey in Pure and Applied Mathematics Series, John Wiley (1991).

[21] von Styrk, O., Numerical solution of optimal control problems by direct collocation, in Optimal Control, R. Burlisch, A. Miele, J. Stoer and K.H. Well, Eds., Birkhauser, Basel (1993), 129-143.