On multi-directional search in Optimization

J.A.Gomez (CEMAFIT), W. Gomez (UH)

October 30th, 1995.

Abstract

The paper studies convergence properties of numerical optimization algorithms using the natural idea of searching in several directions and where one-processor computation is assumed. Global convergence theorems (with exact arithmetic) are proved for two classes of "descent" methods in $\mathbb{R}^n$ and for one class of such methods in Hilbert spaces in the framework of unconstrained problems and where inexact multidirectional search is also considered. Some ilustrative examples of multidirectional algorithms for Nonlinear Programming and Optimal Control problems are formulated and their theoretical convergence discussed.

1 Introduction

We consider the following optimization problem :

$$\min \ \{f(x) : x \in X\}, \tag{1}$$

which is finite dimensional if $X \subset \mathbb{R}^n$ and infinite dimensional if $X \subset H$, where H is a functional space (mainly a Hilbert space). We say the problem (1) is *unconstrained* if $X = \mathbb{R}^n$ or $X = H$ and *constrained* if X is a proper subset of the correponding space. In all the paper we assume f is a continuously differentiable function in X.

One motivation of this work is given by large scale optimization problems. Special methods have been developped for unconstrained problems (see and for linearly constrained (sparse matrix) problems (see [5]) in finite dimension. The general methodology presented in this paper is another approach to handle nonlinear large-scale problems.

The common way in optimization algorithms is to reduce the n-dimensional problem to the solution of a sequence of 1-dimensional problems (line search). This idea is natural to generalize to p-dimensional (multidirectional search). This is a different approach to the results presented in [1] where 1-dimensional optimization is always supposed.

We call *Descent algorithm with exact search in a linear variety* the following iterative process (Ap) :

1. Choose $p \in I\!N$, $x_0 \in X$, and set $k = 0$.
2. Determine p directions $d_k^1, ..., d_k^p$ which will be *feasible* and *descent,* i.e. satisfying the following conditions :

$$\exists \Lambda \subset I\!R^p,\ \Lambda \neq \emptyset : x_k + \lambda_1 d_k^1 + ... + \lambda_p d_k^p \in X,\ \forall (\lambda_1, ..., \lambda_p) \in \Lambda \tag{2}$$
$$\exists\, (\bar\lambda_1, ..., \bar\lambda_p) \in \Lambda : f(x_k + \bar\lambda_1 d_k^1 + ... + \bar\lambda_p d_k^p) < f(x_k) \tag{3}$$

If such directions do not exist, the algorithm stops and x_k is taken as a final point.
If such directions do exist, then continue to step 3.
3. Find a vector $\lambda_k = (\lambda_{k1}, ..., \lambda_{kp}) \in \Lambda$ such that :

$$f(x_k + \lambda_{k1} d_k^1 + ... + \lambda_{kp} d_k^p) = \min_{\lambda \in \Lambda}\ f(x_k + \lambda_1 d_k^1 + ... + \lambda_p d_k^p) \tag{4}$$

4. Take $x_{k+1} = x_k + \lambda_{k1} d_k^1 + ... + \lambda_{kp} d_k^p$, $k = k + 1$, and return to step 2.
Global convergence properties of this process will be obtained in section 2, and some examples will be given in section 4. Large scale problems can then be solved by means of a sequence of less dimensional subproblems $(p << n)$.
Another motivation for this point of view is some popular numerical methods for the solution of optimal control problems (or other optimization problems with differential equations constraints), where, at each iteration, the problem is reduced to a finite dimensional subspace.
We call *Descent algorithm with exact search in a family of linear varieties* the following process (Ap_k) :
1. Choose $x_0 \in X$ and set $k = 0$.
2. Choose $p_k \in I\!N$ and determine p_k directions $d_k^1, ..., d_k^{p_k}$ which will be *feasible* and *descent,* i.e. satisfying the following conditions :

$$\exists \Lambda \subset I\!R^p,\ \Lambda \neq \emptyset : x_k + \lambda_1 d_k^1 + ... + \lambda_{p_k} d_k^{p_k} \in X,\ \forall (\lambda_1, ..., \lambda_{p_k}) \in \Lambda \tag{5}$$
$$\exists\, (\bar\lambda_1, ..., \bar\lambda_{p_k}) \in \Lambda : f(x_k + \bar\lambda_1 d_k^1 + ... + \bar\lambda_{p_k} d_k^{p_k}) < f(x_k) \tag{6}$$

If such p_k and directions do not exist, the algorithm stops and x_k is taken as a final point.
If such p_k and directions do exist, then continue to step 3.
3. Find a vector $\lambda_k = (\lambda_{k1}, ..., \lambda_{kp_k}) \in \Lambda$ such that :

$$f(x_k + \lambda_{k1} d_k^1 + ... + \lambda_{kp_k} d_k^{p_k}) = \min_{\lambda \in \Lambda}\ f(x_k + \lambda_1 d_k^1 + ... + \lambda_{p_k} d_k^{p_k}) \tag{7}$$

4. Take $x_{k+1} = x_k + \lambda_{k1} d_k^1 + ... + \lambda_{kp_k} d_k^{p_k}$, $k = k + 1$, and return to step 2.
In the next sections we will give some convergence results for this process.

2 Basic ConvergenceTheory

2.1 Global convergence Theorem

The general definitions and results of global convergence can be essentially found in [7]. The concepts of closed point-to-set maps, algorithms, composition of point-to-set maps, and descent functions related to an algorithm A and a solution set Γ will be supposed to be known. For language economy we introduce the following :

Definition 1 *A point-to-set map from X to Y is called open at $x \in X$ if X, Y are topological spaces and*

$$\begin{array}{c} \textit{For any two sequences } \{x_k\} \subset X, \{y_k\} \subset Y \textit{, satisfying:} \\ (x_k \longrightarrow x, y_k \in A(x_k)) \\ \textit{there exists } y \in Y \textit{ and a subsequence } \{y_{k_j}\} \textit{ such that } y_{k_j} \longrightarrow y \end{array} \tag{8}$$

A point-to-set map from X to Y is called continuous at $x \in X$ if it is open and closed at x.
Note that a point-to-point continuous function is open and closed in this sense, and then is also continuous as a point-to-set map.

Proposition 1 *Let $A : X \longrightarrow Y$, and $B : Y \longrightarrow W$ be point-to-set maps defined on the topological spaces X, Y, W. Supose A is closed at $x \in X$, B is closed at $A(x) \subset Y$, and one of the following additional conditions holds :*
i) Y is a sequentially compact space, or
ii) A is open (therefore continuous) at $x \in X$,
then the composition map $C = B \circ A$ is closed at $x \in X$.

Definition 2 *We will call globally convergent, an algorithm A for the problem :*

$$\textit{Find} \quad \bar{x} \in \Gamma. \tag{9}$$

which satisfies that any acumulation point of each bounded sequence generated by A belongs to Γ.

As a corollary of the global convergence theorem (see [7]), any algorithm A for (9) is globally convergent if A is closed and there exists a descent function Z related to Γ and A.

2.2 Descent methods in a linear variety

We deal with iteration processes of the form (Ap) or (Ap_k) . The theoretical statement of the corresponding algorithms follows the classical formulation of considering steps 2) and 3) as separated point-to-set maps : G (for direction selector) and S (for new point selector) , and define the overall algorithm as the composition $S \circ G$. We must define continuous G and closed S in order to apply the global convergence theorem and for this end we introduce some definitions.

Definition 3 *The symbol $\mathcal{M}_{(n,p)}$ (or $\mathcal{M}$ in short) denotes the normed space of matrices of order $n \times p$ with the infinite norm :*

$$\|M\|_{\infty} = \max_{1 \leq i \leq n,\ 1 \leq j \leq p} |M_{ij}|\,. \tag{10}$$

Choosing p directions in $I\!R^n$ is equivalent to choose a matrix $M \in \mathcal{M}_{(n,p)}$. We introduce the classical descent ρ-uniform condition for directions and gradients and the boundedness condition for directions, but they are not enough to apply the global convergence theorem in the multidimensional case, as we shall see.

Definition 4 *Consider the problem (1) with $X = I\!R^n$, $\rho > 0$ and let h_1, h_2 be non-negative continuous functions on X. The descent ρ condition for a matrix (of directions) $D \in \mathcal{M}_{(n,p)}$, at the point $x \in X$, is the following :*

$$\exists \lambda \in I\!R^p, \lambda \neq 0 :\ \nabla f(x) D \lambda \leq -\rho \|\nabla f(x)\| \, \|D\lambda\|\,. \tag{11}$$

The h_1-boundedness condition for $D \in \mathcal{M}_{(n,p)}$ at the point $x \in X$ is the following :

$$\|D\|_{\infty} \leq h_1(x). \tag{12}$$

The h_2-full-rank condition for $D \in \mathcal{M}_{(n,p)}$ at the point $x \in X$ is the following :

$$\det(D^{\mathrm{T}} D) \geq h_2(x). \tag{13}$$

Definition 5 *The point-to-set direction map $G_1 : X \longrightarrow \mathcal{P}(X \times \mathcal{M})$ is defined by :*

$$G_1(x) = \{(x, D) \in \{x\} \times \mathcal{M} \mid (x, D) \textit{ satisfies (11) and (12)}\} \tag{14}$$

The point-to-set direction map $G_2 : X \longrightarrow \mathcal{P}(X \times \mathcal{M})$ is defined by :

$$G_2(x) = G_1(x) \cap \{(x, D) \in \{x\} \times \mathcal{M} \mid (x, D) \textit{ satisfies (13)}\} \tag{15}$$

The point-to-set map G_2 includes what we called the *full-rank condition* of directions and it plays a fundamental role in global convergence. In the one dimensional case this condition is always satisfied if $d \neq 0$, but a matrix can be non zero and be composed by linearly dependent vectors. In this degenerated situation the general algorithm $A = S \circ G_1$ could be non closed. We face this problem in two ways :
- restricting the search directions to those linearly independent, or
- restricting the optimization search only to bounded lambdas.

Definition 6 *The point-to-set search map $S_1 : X \times \mathcal{M}_{(n,p)} \longrightarrow \mathcal{P}(X)$ is defined by :*

$$S_1(x, D) = \left\{ y \in X \mid y = x + D\bar{\lambda},\ f(y) = \min_{\lambda \in I\!\!R^p} f(x + D\lambda) \right\} \tag{16}$$

The point-to-set search map $S_2 : X \times \mathcal{M}_{(n,p)} \longrightarrow \mathcal{P}(X)$ is defined by :

$$S_2(x, D) = \left\{ y \in X \mid y = x + D\bar{\lambda},\ f(y) = \min_{\lambda \in \mathcal{B}(x)} f(x + D\lambda) \right\}, \tag{17}$$

where :

$$\mathcal{B}(x) = \{\lambda \in I\!\!R^p \mid ||\lambda|| \leq h_3(x)\}, \tag{18}$$

and h_3 is a continuous non-negative function.

Lemma 1 *G_1 is a continuous point-to-set map in all $x \in X$.*

Proof Suppose $x \in X$, and let be $x_n \longrightarrow x$, $D_n \longrightarrow D$, $(x_n, D_n) \in G_1(x_n)$ for all $n \in I\!\!N$. Then there exist $\lambda_n \in I\!\!R^p, \lambda_n \neq 0$, such that :

$$\begin{aligned} \nabla f(x_n) D_n \lambda_n &\leq -\rho ||\nabla f(x_n)|| \, ||D_n \lambda_n|| && (19) \\ ||D_n|| &\leq h_1(x_n). && (20) \end{aligned}$$

If rank $D < p$, there exists $\hat{\lambda} \in I\!\!R^p, \hat{\lambda} \neq 0$, such that $D\hat{\lambda} = 0$. Therefore, the relations (11) and (12) hold trivially and we can then suppose that :

$$\text{rank } D = p. \tag{21}$$

By (21) almost all (for $n \geq n_0$) matrices D_n have full rank p. As a consequence, we can suppose (without lost of generality) that there exists a convergent subsequence $\{\lambda_{n_k}\} \longrightarrow \lambda \in I\!\!R^p$, satisfying (19). In fact, the assumption $d_n = D_n \lambda_n = 0$ implies $\lambda_n = 0$, $n \geq n_0$ (full rank!) and then a contradiction.

The sequence $\{d_n\}$ satisfies $\|d_n\| \neq 0$, $n \geq n_0$, and we can divide (19) by $\|d_n\| = \|D_n \lambda_n\|$ and redefine $\bar{\lambda}_n = \frac{\lambda_n}{\|d_n\|}$ which satisfies (19). The sequence $D_n \bar{\lambda}_n$ has unit norm and then admits a nonzero limit. From the full rank condition of D_n and D we obtain that $\bar{\lambda}_n$ admits a subsequence converging to a non zero $\lambda \in I\!R^p$. has unit norm and then admits a subsequence converging to some non zero $\lambda \in I\!R^p$.
In addition $D_{n_k} \lambda_{n_k}$ converges to $D\lambda$, since $D_{n_k} \longrightarrow D$ and $\lambda_{n_k} \longrightarrow \lambda$. Therefore, we can write (19) and (20) for x_{n_k}, D_{n_k} and λ_{n_k} and take limits in k . From continuity we obtain that G_1 is closed.
Let now be two sequences $\{x_n\}, \{D_n\}$ such that $x_n \longrightarrow x$, $(x_n, D_n) \in G_1(x_n)$, then by (12) the sequence $\|D_n\|_\infty$ is bounded, i.e. every component sequence $\{(D_{ij})_n\}$ is bounded, and then there exists a subsequence $\{D_{n_k}\}$ converging to $D \in \mathcal{M}$. The subsequence (x_{n_k}, D_{n_k}) converges to (x, D) and this shows that G_1 is open and then continuous. □

Lemma 2 *G_2 is a continuous point-to-set map for all $x \in X$.*

Proof Let be $x_n \longrightarrow x$, $D_n \longrightarrow D$, $(x_n, D_n) \in G_2(x_n)$. Since determinant is a continuous function of the components of the matrix, from continuity of h_2 and the inequality :

$$\det\left(D_n^{\mathrm{T}} D_n\right) \geq h_2(x_n)$$

we obtain :

$$\det\left(D^{\mathrm{T}} D\right) \geq h_2(x). \tag{22}$$

Continuity of G_1 and (22) gives that G_2 is also continuous.□

Lemma 3 *S_1 is a closed point-to-set map for all $(x, D) \in X \times \mathcal{M}_{(n,p)}$, satisfying the full-rank condition (13) with :*

$$h_2(x) > 0. \tag{23}$$

Proof Let be sequences $(x_n, D_n) \longrightarrow (x, D)$, $y_n \longrightarrow y$, $y_n \in S_1(x_n, D_n)$. We must show that $y \in S_1(x, D)$. As D has full rank p, we can suppose D_n has full rank p for all $n \geq n_0$. Furthemore, for each n there exist $\lambda_n \in I\!R^p$ such that :

$$y_n = x_n + D_n \lambda_n \tag{24}$$

$$f(y_n) \leq f(x_n + D_n \lambda), \forall \lambda \in I\!R^p, \tag{25}$$

From (24) and full rank, we have :

$$\lambda_n = (D_n^{\mathrm{T}} D_n)^{-1} D_n^{\mathrm{T}} (y_n - x_n), \;\; n \geq n_0, \tag{26}$$

and by continuity of matrix inversion, we can take limit in (26) and obtain that the sequence λ_n converges to $\lambda = (D^{\mathrm{T}}D)^{-1}D^{\mathrm{T}}(y-x)$. From (24) we have $y = x + D\lambda$. If in (25) we fix λ and take limit in n we will have :

$$f(y) \leq f(x + D\lambda), \forall \lambda \in I\!\!R^p \tag{27}$$

which in turns proves that $y \in S_1(x, D)$, and S_1 is closed.$\square$
The Lemma also includes what we called the *full-rank condition* of directions. In the case of matrix D is rank defficient, the map S_1 is not closed, as the following example shows :
EXAMPLE:
Take $X = I\!\!R^2$, $f(x_1, x_2) = (x_1 - 1)^2 + (x_2 - 2)^2$, $x_n = (0,0)^{\mathrm{T}}$, (constant),

$$D_n = \left(\begin{array}{cc} \frac{1}{n} & 0 \\ 0 & 1 \end{array} \right).$$

In this case, $x_n \longrightarrow (0,0)$ and $D_n \longrightarrow D = \left(\begin{array}{cc} 0 & 0 \\ 0 & 1 \end{array} \right)$. For all $n \in I\!\!N$ the column vectors of D_n are linearly independent and generate $I\!\!R^2$, then the search function gives :

$$S_1(x_n, D_n) = \arg \min_{\lambda \in I\!\!R^2} f(x_n + D_n\lambda) = y_n = (1,2)^{\mathrm{T}}, \text{ constant.}$$

Therefore, we have $x_n \longrightarrow x = (0,0)^{\mathrm{T}}$, $D_n \longrightarrow D = \left(\begin{array}{cc} 0 & 0 \\ 0 & 1 \end{array} \right)$, $y_n \longrightarrow y = (1,2)^{\mathrm{T}}$, $y_n \in S_1(x_n, D_n)$, but $(1,2)^{\mathrm{T}} \notin S_1(x, D)$ since :

$$S_1(x, D) = \arg \min_{\lambda \in I\!\!R^2} f(x + D\lambda) = y = (0,2)^{\mathrm{T}}.$$

But in this example, the optimal parameter sequence $\hat{\lambda}_n$ corresponding to y_n is not bounded because a simple calculation gives $\hat{\lambda}_n = (n, 2)^{\mathrm{T}}$. For this reason, we have another possibility :

Lemma 4 *S_2 is a closed point-to-set map for all $(x, D) \in X \times \mathcal{M}_{(n,p)}$.*

Proof Let be sequences $(x_n, D_n) \longrightarrow (x, D)$, $y_n \longrightarrow y$, $y_n \in S_2(x_n, D_n)$. We must show that $y \in S_2(x, D)$. For each n there exist $\lambda_n \in I\!\!R^p$, $\|\lambda_n\| \leq h_3(x_n)$, such that :

$$\begin{aligned} y_n &= x_n + D_n\lambda_n & (28) \\ f(y_n) &\leq f(x_n + D_n\lambda), \forall \lambda : \|\lambda\| \leq h_3(x_n) & (29) \end{aligned}$$

By continuity of h_3 the sequence λ_n is bounded and then, there exists a vector $\lambda \in I\!R^p$, such that $\|\lambda\| \leq h_3(x)$, and a subsequence λ_{n_k} converging to λ. The subsequence $D_{n_k}\lambda_{n_k}$ also converges to $D\lambda$ and from (28) we have $y = x + D\lambda$. In addition, from continuity of h_3 we can write :

$$\forall \varepsilon > 0,\ \exists m_\varepsilon \in I\!N : |h_3(x_n) - h_3(x)| \leq \varepsilon,\ \forall n \geq m_\varepsilon. \tag{30}$$

From (29) and (30), for arbitrary fixed $\varepsilon > 0$ there exists k_0 such that for all $k \geq k_0$ we have $n_k \geq m_\varepsilon$ and the following inequalities hold :

$$\min_{\|\lambda\| \leq h_3(x)+\varepsilon} f(x_{n_k} + D_{n_k}\lambda) \leq f(y_{n_k}) \leq \min_{\|\lambda\| \leq h_3(x)-\varepsilon} f(x_{n_k} + D_{n_k}\lambda). \tag{31}$$

Take $k \longrightarrow \infty$ in (31) and we have :

$$\min_{\|\lambda\| \leq h_3(x)+\varepsilon} f(x + D\lambda) \leq f(y) \leq \min_{\|\lambda\| \leq h_3(x)-\varepsilon} f(x + D\lambda).$$

Since ε is arbitrary, it can be easily obtained that $y \in S_2(x, D)$ and S_2 is closed. $\square$

An example of convergence theorem for the process (Ap) in $\mathcal{M}$ can be written easily.

3 Extensions

3.1 Variable number of directions

For the process (Ap_k), where the number of directions varies at each iteration, in the finite dimensional case $X = I\!R^n$, we first make the natural assumption that the sequence $\{p_k\}$ is bounded (for example by the dimension n of the space X). In that case, we can work in the normed space $\mathcal{M}_{(n,n)}$ of $n \times n$ matrices with the same norm (10) with the following identification :

$$M \in \mathcal{M}_{(n,p)} \Longrightarrow M_{n \times p} \simeq [M\ 0]_{n \times n},$$

i.e. fill in the missing $(n-p)$ columns with zeros.
When $\{p_k\}$ is bounded, the basic G and S-Lemmas and the Convergence Theorem of process (Ap_k) are also true, taking in account that there must be a subsequence $\{p_{k_i}\}$ which is *constant* and then the proofs can be reduced to the constant case. Note that the full-rank assumption (13) forces the sequence $\{p_k\}$ to be bounded by n.

3.2 Inexact search

Let be $X = I\!R^n$ for the problem (1). In our next analysis of the process (Ap) we will consider that the subproblem (4) can't be solved exactly, and then some type of stoppig-rule has to be taken. The most popular is that defined by the Wolfe-Powell conditions (see for example [3]).

Definition 7 *Let be $\alpha \in (0,1), \beta \in (\alpha, 1)$. We say the vector $d \in I\!R^n$ satisfies Armijo condition if :*

$$f(x+d) - f(x) \leq \alpha \nabla f(x) d. \tag{32}$$

The vector $d \in I\!R^n$ satisfies Wolfe-Powell (W-P) conditions if in addition to (32) the following inequality holds :

$$\nabla f(x+d)d \geq \beta \nabla f(x) d. \tag{33}$$

We call (Apw) the process (Ap) with the step 3) changed by the following:
3w) Find a vector $\lambda_k = (\lambda_{k1}, ..., \lambda_{kp}) \in \Lambda$ such that :

$$D_k \lambda_k = \lambda_{k1} d_k^1 + ... + \lambda_{kp} d_k^p \text{ satisfies W-P conditions.} \tag{34}$$

For the convergence theory it is only necessary the analysis of definitions and Lemmas of the search maps :

Definition 8 *The point-to-set search map $S_1 : X \times \mathcal{M} \longrightarrow \mathcal{P}(X)$ is defined by :*

$$S_1(x, D) = \left\{ y \in I\!R^n \;\middle|\; \begin{array}{c} \exists\, \lambda \in I\!R^p, \lambda \neq 0 : y = x + D\lambda, \\ D\lambda \textit{ satisfies (32),(33) and (11)} \end{array} \right\}, \tag{35}$$

i.e. $d = D\lambda$ satisfies the Wolfe-Powell conditions and the descent ρ-condition. The point-to-set search map $S_2 : X \times \mathcal{M} \longrightarrow \mathcal{P}(X)$ is defined by :

$$S_2(x, D) = \left\{ y \in I\!R^n \;\middle|\; \begin{array}{c} \exists \lambda \in \mathcal{B}(x), \lambda \neq 0 : y = x + D\lambda, \\ D\lambda \textit{ satisfies (32),(33) and (11)} \end{array} \right\}, \tag{36}$$

where $\mathcal{B}(x)$ is the ball with center at 0 and radius $h_3(x)$ as before, where h_3 is again a continuous non-negative function.

Lemma 5 *S_1 is a closed point-to-set map in all $(x, D) \in X \times \mathcal{M}$ satisfying $\nabla f(x) \neq 0$ and the full-rank condition (13).*

Proof The proof follows the same lines described in the first S_1-Lemma. There is only one new condition, namely $\lambda \neq 0$. Using full-rank we obtain that the sequence $\{\lambda_n\}$ is convergent, non-zero and

$$\lambda_n = (D_n^{\mathrm{T}} D_n)^{-1} D_n^{\mathrm{T}} (y_n - x_n) \longrightarrow \lambda = (D^{\mathrm{T}} D)^{-1} D^{\mathrm{T}} (y - x),$$

where we use the same notation as before. As $\|D_n \lambda_n\| \neq 0, n \geq n_0$, we define:

$$d_n = \frac{D_n \lambda_n}{\|D_n \lambda_n\|}$$

which is a unit norm sequence, and there exists a subsequence $\{d_{n_k}\}$ converging to $d \in I\!\!R^n, d \neq 0$.
Using (11) and taking limits in the inequalities :

$$\nabla f(x_{n_k}) d_{n_k} \leq -\rho \|\nabla f(x_{n_k})\|$$

we have, since $\nabla f(x) \neq 0$, that d satisfies :

$$\nabla f(x) d \leq -\rho \|\nabla f(x)\| < 0. \tag{37}$$

On the other hand, by Wolfe-Powell condition (33) we have :

$$\nabla f(y_{n_k}) d_{n_k} \geq \beta \nabla f(x_{n_k}) d_{n_k}, \tag{38}$$

then if we suppose $\lambda = 0 = \lim \lambda_{n_k}$, it implies $y_{n_k} \longrightarrow x$, and therefore we obtain from (38):

$$\nabla f(x) d \geq \beta \nabla f(x) d,$$

which with (37) contradicts the assumption $\beta \in (0,1)$ and then $\lambda \neq 0$.
The relations (32),(33) and (11) for $D\lambda$ is obtained by continuity as in the preceedings proofs. □
Analogous Lemmas can be proved for the others S_2-maps but they are very similar to the preceding ones and we won't give the details. Also the convergence theorem for (Ap) in $\mathcal{M}$ can be given for the process (Apw) if we select convenient point-to-set direction maps G.

3.3 Hilbert spaces

For the process (Ap) we suppose now that the space X is a real Hilbert space H with scalar product :

$$\prec x, y \succ, \ \ x, y \in H$$

and the induced norm :

$$\|x\|_H = \sqrt{\prec x, x \succ},$$

for which H is a Banach space.
The space where we select the directions will be H^p, the set of all p-tuples $(d_1,...,d_p)$, with $d_i \in H, i = 1,...,p$, but we consider the weak topology in H^p, i.e. the product of the pointwise convergence topology, which coincide with the weak-* topology because reflexivility, and for which any norm-bounded ball is compact (Alaouglu's Theorem)(see for example [8]). We will write D for an element of H^p and write also D_i for its components, $i = 1,...,p$. A sequence $\{D_n\}$ in H^p is weakly convergent to D, denoted by $D_n \rightharpoonup D$, if each component sequence $\{(D_n)_i\}$ converges weakly (pointwise) to D_i, i.e.

$$D_n \rightharpoonup D \Longleftrightarrow \lim_n \prec (D_n)_i,\ x \succ = \prec D_i,\ x \succ,\ \forall x \in H, i = 1,p \tag{39}$$

Definition 9 *Consider the problem (1) with $X = H$, $\rho > 0$ and let h_1, h_2, h_3 be non-negative norm-continuous functions on X. The descent ρ-conditions for a matrix (of directions) $D \in H^p$, at the point $x \in X$, are the following :*

$$\exists \lambda \in I\!R^p : \|\lambda\| = h_1(x),\ \prec \nabla f(x), D\lambda \succ \leq -\rho \|\nabla f(x)\|_H \|D\lambda\|_H\ , \tag{40}$$
$$\exists \lambda \in I\!R^p,\ \lambda \neq 0 : \prec \nabla f(x), D\lambda \succ \leq -\rho \|\nabla f(x)\|_H \|D\lambda\|_H\ . \tag{41}$$

The h_2-boundedness condition for $D \in H^p$ at the point $x \in X$ is the following:

$$\|D\|_\infty = \max_{1\leq i\leq p} \|D_i\|_H \leq h_2(x). \tag{42}$$

The h_3-full-rank condition for $D \in H^p$ at the point $x \in X$ is the following:

$$\left|\det(D^{\mathrm{T}} D)\right| \geq h_3(x), \tag{43}$$

where $D^{\mathrm{T}} D$ denote the Gramm matrix of the components of D :

$$(\prec D_i, D_j \succ), 1 \leq i,j \leq p.$$

The inequality (43) (for $h_3(x) > 0$) is equivalent to linear independence of vectors $D_i, i = 1,...,p$, and for this reason we keep the same "full-rank" name; but even this conditions is not enough to obtain the result with (41) since, for example, from weak convergence of D_n to D does not follows the convergence of $D_n^{\mathrm{T}} D_n$ to $D^{\mathrm{T}} D$. Then we need to add the norm-equality (to h_1) of λ in (40) and make a substantial reformulation of the proofs of G and S Lemmas.

Definition 10 *The point-to-set direction map $G_1 : X \longrightarrow \mathcal{P}(X \times H^p)$ is defined by :*

$$G_1(x) = \{(x, D) \in \{x\} \times H^p \mid (x, D) \text{ satisfies (41)-(42)}\}\ . \tag{44}$$

The point-to-set direction map $G_{11} : X \longrightarrow \mathcal{P}(X \times H^p)$ *is defined by :*

$$G_{11}(x) = \{(x, D) \in \{x\} \times H^p \mid (x, D) \text{ satisfies (40)-(42)}\}. \tag{45}$$

The point-to-set direction map $G_2 : X \longrightarrow \mathcal{P}(X \times H^p)$ *is defined by :*

$$G_2(x) = G_1(x) \cap \{(x, D) \in \{x\} \times H^p \mid (x, D) \text{ satisfies (43)}\} \tag{46}$$

The point-to-set search map $S_1 : X \times H^p \longrightarrow \mathcal{P}(X)$ *is defined by :*

$$S_1(x, D) = \left\{ y \in X \mid y = x + D\bar{\lambda},\ f(y) = \min_{\lambda \in I\!R^p} f(x + D\lambda) \right\} \tag{47}$$

The point-to-set search map $S_2 : X \times H^p \longrightarrow \mathcal{P}(X)$ *is defined by :*

$$S_2(x, D) = \left\{ y \in X \mid y = x + D\bar{\lambda},\ f(y) = \min_{\lambda \in \mathcal{B}(x)} f(x + D\lambda) \right\} \tag{48}$$

where $\mathcal{B}(x)$ *is defined through a non-negative norm-continuous function* h_4 *as before.*

Lemma 6 G_{11} *is a continuous point-to-set map in all* $x \in X$ *.*

Proof Suppose $x_0 \in X$ and let be $x_n \longrightarrow x_0$, $D_n \rightharpoonup D_0$, $(x_n, D_n) \in G_{11}(x_n)$ for all $n \in I\!N$. Then there exist $\lambda_n \in I\!R^p$, such that :

$$\begin{aligned} \|\lambda_n\| &= h_1(x_n), & (49)\\ \prec \nabla f(x_n), D_n\lambda_n \succ &\leq -\rho \|\nabla f(x_n)\|_H \|D_n\lambda_n\|_H . & (50) \end{aligned}$$

From continuity of h_2 and (42) we also have :

$$\forall \varepsilon > 0, \exists n_0 \in I\!N : \forall n \geq n_0, \|D_n\|_\infty \leq h_2(x_0) + \varepsilon. \tag{51}$$

Since any bounded ball is convex and norm-closed, then weakly closed, from (51) we obtain :

$$\forall \varepsilon > 0, \|D_0\|_\infty \leq h_2(x_0) + \varepsilon \Longrightarrow \|D_0\|_\infty \leq h_2(x_0). \tag{52}$$

Furthermore, the sequence $\{\lambda_n\}$ is bounded, since h_1 is norm-continuous, and there exists a subsequence $\{\lambda_{n_k}\}$ converging to some $\lambda_0 \in I\!R^p$. From continuity and (49) we have also :

$$\|\lambda_0\| = h_1(x_0) \tag{53}$$

By the Cauchy-Schwartz inequality :

$$|\prec D_{n_k}\lambda_{n_k} - D_0\lambda_0, x \succ| \leq \prec D_{n_k}\lambda_{n_k} - D_{n_k}\lambda_0, x \succ + \prec D_{n_k}\lambda_0 - D_0\lambda_0, x \succ \leq$$

$$\leq \|D_{n_k}\|_\infty \|\lambda_{n_k} - \lambda_0\| \|x\|_H + \prec (D_{n_k} - D_0)\lambda_0, x \succ, \forall\ x \in H, \tag{54}$$

and from (54) and assumptions we obtain the sequence $D_{n_k}\lambda_{n_k}$ converges weakly to $D_0\lambda_0$,. In particular :

$$\lim_k \prec \nabla f(x_0), D_{n_k}\lambda_{n_k} \succ = \prec \nabla f(x_0), D_0\lambda_0 \succ . \tag{55}$$

From (49) and (51) follows the norm-boundedness of $D_{n_k}\lambda_{n_k}$. Using norm-continuity of $\nabla f(x)$ and norm-boundedness of $D_{n_k}\lambda_{n_k}$ we have :

$$\begin{aligned} &|\prec (\nabla f(x_{n_k}) - \nabla f(x_0)), D_{n_k}\lambda_{n_k} \succ| \\ \leq &|\ \nabla f(x_{n_k}) - \nabla f(x_0)\ |_H |\ D_{n_k}\lambda_{n_k}\ |_H \longrightarrow 0. \end{aligned} \tag{56}$$

In addition, from (50) we can write :

$$\prec \nabla f(x_{n_k}) - \nabla f(x_0), D_{n_k}\lambda_{n_k} \succ + \prec \nabla f(x_0), D_{n_k}\lambda_{n_k} \succ$$

$$\leq -\rho \|\nabla f(x_{n_k})\|_H \|D_{n_k}\lambda_{n_k}\|_H$$

which gives, taking limit in k and using (55) and (56), the inequality (see [2], Proposition III.5):

$$\begin{aligned} \prec \nabla f(x_0), D_0\lambda_0 \succ \leq -\rho \lim\nolimits_k \inf \|\nabla f(x_{n_k})\|_H \|D_{n_k}\lambda_{n_k}\|_H \leq \\ -\rho \|\nabla f(x_0)\|_H \|D_0\lambda_0\|_H . \end{aligned} \tag{57}$$

The relations (52), (53) and (57) means that G_1 is closed.
On the other hand, if $x_n \longrightarrow x_0$, $(x_n, D_n) \in G_1(x_n)$, the sequence D_n is norm-bounded, for example by :

$$K = \sup_n\ h_2(x_n). \tag{58}$$

This implies (see [2]) the existence of a subsequence (x_{n_k}, D_{n_k}) and a vector $(x_0, D_0) \in X \times H^p$ such that $(x_{n_k}, D_{n_k}) \longrightarrow (x_0, D_0)$, and then G_1 is open and continuous. □
The simplest way to obtain the continuity of the point-to-set map G_2 is the Lemma follows. Nevertheless it will be useful for establish global convergence of some optimal control algorithms, as we shall see in the next section.

Lemma 7 *If there exists a finite dimensional subspace $H_0^p \subset H^p$ such that :*

$$G_2(x) \subset \{x\} \times H_0^p, \ \forall x \in X, \tag{59}$$

then G_2 is a continuous point-to-set map at all $x \in X$ satisfying :

$$h_3(x) > 0. \tag{60}$$

Proof Let be $x_0 \in X$ which satisfies (60). If $x_n \longrightarrow x_0$, $D_n \rightharpoonup D_0$, $(x_n, D_n) \in G_2(x_n)$, we have the inequalities :

$$\det\left(D_n^{\mathrm{T}} D_n\right) \geq h_3(x_n), \ \forall n \in \mathbb{N}, \tag{61}$$

and we will show that :

$$\det\left(D_0^{\mathrm{T}} D_0\right) \geq h_3(x_0). \tag{62}$$

By assumption H_0^p is a finite dimensional subspace and then is norm-closed (see [14]). For this reason H_0^p itself is also a Hilbert space for the induced scalar product. In addition, it is also a weak-closed subspace by convexity, and therefore $D_0 \in H_0^p$.
Each component $(D_n)_j \in H_0^p$ converges weakly to $(D_0)_j$, $j = 1, ..., p$, but in H_0^p it is equivalent to strong-convergence since H_0^p is finite dimensional, i.e.

$$\prec (D_n)_j - (D_0)_j, (D_n)_j - (D_0)_j \succ \longrightarrow 0, \ j = 1, ..., p. \tag{63}$$

Then, from inequalities :

$$\begin{aligned} &|\prec (D_n)_j, (D_n)_j \succ - \prec (D_0)_j, (D_0)_j \succ| \leq \\ &\quad \leq \|(D_n)_j - (D_0)_j\|_H \, \|(D_n)_j\|_H + \\ &\|(D_n)_j - (D_0)_j\|_H \, \|(D_0)_j\|_H \,, \ j = 1, ..., p, \end{aligned} \tag{64}$$

(63) and norm-boundedness we have :

$$\lim_n D_n^{\mathrm{T}} D_n = D_0^{\mathrm{T}} D_0. \tag{65}$$

Since determinant is a continuous function of the elements of the matrix and using continuity of h_3 and modulo, we can take limits in (61) to get (62). The continuity of G_2 now can be obtained like preceeding proofs, since the relation (65) gives linearly independence of $D_n^{\mathrm{T}} D_n$, $n \geq n_0$, and then we can find $\lambda_0 \neq 0$ satisfying (41). $\square$

Lemma 8 *If f is a weakly continuous function on X, then S_1 is a closed point-to-set map in all $(x, D) \in X \times H^p$ satisfying the full-rank condition (43) with (60) :*

Proof Let be $x_0 \in X$ satisfying (43) and (60) and sequences $x_n \longrightarrow x_0$, $D_n \rightharpoonup D_0$, $y_n \longrightarrow y_0$, $y_n \in S_1(x_n, D_n)$. We must show that $y_0 \in S_1(x_0, D_0)$.
There exists a sequence $\lambda_n \in I\!\!R^p$ such that :

$$\begin{aligned} y_n &= x_n + D_n \lambda_n, & (66)\\ f(y_n) &\leq f(x_n + D_n \lambda), \ \forall \lambda \in I\!\!R^p. & (67) \end{aligned}$$

Weak convergence gives the relations :

$$\lim_n \prec (D_0)_i, (D_n)_j \succ = \prec (D_0)_i, (D_0)_j \succ, \ 1 \leq i, j \leq p,$$

and then we obtain :

$$\lim_n D_0^{\mathrm{T}} D_n = D_0^{\mathrm{T}} D_0. \tag{68}$$

The full-rank condition (43), (68) and continuity of h_3 give us :

$$\forall \varepsilon > 0, \ \exists n_0 \in I\!\!N : \forall n \geq n_0, \left|\det(D_0^{\mathrm{T}} D_n)\right| \geq h_3(x) - \varepsilon. \tag{69}$$

From (66), (69) and $\varepsilon > 0$ small enough, we have :

$$\lambda_n = \left(D_0^{\mathrm{T}} D_n\right)^{-1} D_0^{\mathrm{T}}(y_n - x_n), \ \forall n \geq n_0, \tag{70}$$

which means that the sequence $\{\lambda_n\}$ is convergent with limit :

$$\lambda_0 = \lim_n \lambda_n = \left(D_0^{\mathrm{T}} D_0\right)^{-1} D_0^{\mathrm{T}}(y_0 - x_0).$$

We now take limit in (66) and obtain :

$$y_0 = x_0 + D_0 \lambda_0.$$

For any fixed $\lambda \in I\!\!R^p$ the sequence $x_n + D_n\lambda$ converges weakly to $x_0 + D_0\lambda$, then using the weak continuity of f and (67) we obtain :

$$f(y_0) \leq f(x_0 + D_0\lambda), \ \forall \lambda \in I\!\!R^p,$$

which means that S_1 is closed. □

Lemma 9 *If f is a weakly continuous function on X, then S_2 is a closed point-to-set map in all $(x, D) \in X \times H^p$.*

Proof The continuity of h_4 gives boundedness of the sequence λ_n and convergence of some subsequence λ_{n_k} to $\lambda \in I\!R^p$ such that $\|\lambda\| \leq h_4(x)$. The rest of the proof follows the same arguments as in the preceeding Lemma. $\square$
Convergence Theorems can be formulated analogously for $S_1 \circ G_2$ and $S_2 \circ G_{11}$, as we did before, but we won't repeat it. Furthermore, as we pointed out, the remarks given for a process (Ap_k) in finite dimension are still valid for Hilbert space while the sequence $\{p_k\}$ remains bounded.

4 Applications

In this section we will give ilustrative examples of multidirectional algorithms for optimization problems and establish their global convergence with the tools developped in the preceeding sections. We analyse also the global convergence of a simple class of algorithms for optimal control problems. A systemathic design and theoretical and computational study of multidirectional algorithm (specially for infinite-dimensional problems) is desired but it must be also a matter of another paper.

4.1 Gradient and Newton-like direction methods

The two not only more popular but also more important search directions in minimization algorithms is the (negative of) gradient ($d_1 = -\nabla f^{\mathrm{T}}(x)$) and Newton ($d_2 = -\nabla^{-2} f(x) \nabla f^{\mathrm{T}}(x)$) directions, where $\nabla^2 f(x)$ denote the Hessian matrix of f at point x, and $\nabla^{-2} f(x)$ its inverse. In Quasi-Newton approach, some positively definite approximations of the Hessian $H(x)$, and the setting $d_2 = -H^{-1}(x)\nabla f^{\mathrm{T}}(x)$ is also used instead, where we don't need the second derivatives of f. The non-familiar reader can see [3], for example. It is interesting to note that the family of directions defined by :

$$d(x) = -H \nabla f^{\mathrm{T}}(x), \tag{71}$$

where H is a positively definite matrix, satisfies a "uniform " property given by the following :

Definition 11 *Let be $d(x)$ a continuous function $d : X \longrightarrow X$, and $\rho \in (0,1)$. We say that $d(x)$ is uniform descent if it satisfies the descent ρ-uniform condition for all $x \in X$, i.e.*

$$\nabla f(x) d(x) \leq -\rho \, \|\nabla f(x)\| \, \|d(x)\| \, , \ \forall x \in X. \tag{72}$$

Proposition 2 *If H is a positively definite matrix, then the function $d(x)$, given by (71), is a uniform descent function.*

Proof Denote by σ_l and σ_h the lowest and highest eigenvalue of the matrix H, respectively. The inequality :

$$\nabla f(x)d(x) \leq -\frac{\sigma_l}{\sigma_h} \|\nabla f(x)\| \left\| H\nabla f^{\mathrm{T}}(x) \right\|,$$

can be stablish easily, which means $d(x)$ is uniform descent for $\rho = \sigma_l/\sigma_h$. □
Therefore, any multidirectional algorithm whose matrix of search directions $D = D(x)$ are defined in such a way that at least one of its columns (say the first column) is a uniform descent function, then automatically (no matter the rest of the directions would be) it satisfies (11) (set $\lambda = (1, 0, ..., 0) \in \mathbb{R}^p$), and then to apply the global convergence theorem it is only necessary to verify (12) and (13). This will be the case, in any algorithm that define the matrix of search directions with any function of the Gradient or Newton-like type (71).
A known example of this kind of algorithms is the classical idea of combining the Newton and Gradient directions :

$$D = D(x) = (D_1(x)\ D_2(x))_{n\times 2} = \left(-\nabla f(x),\ -\nabla f^{-2}(x)\nabla f(x)\right), \qquad (73)$$

and where the subproblems (4) are two-dimensional. We call NEWYG this algorithm but with inexact two-dimensional search and Wolfe-Powell conditions as stopping rule. However, the third step is not completely defined, unless we give a rule to generate a new trial point from a failed one. This can be done considering, at step k, the function :

$$\varphi(\lambda_1, \lambda_2) = f(x_k + \lambda_1 d_1^k + \lambda_2 d_2^k), \qquad (74)$$

where $d_1^k = -\nabla f(x_k)$, $d_2^k = -\nabla f^{-2}(x_k)\nabla f(x_k)$, and first try the point $\lambda = (0, 1)$ which corresponds to a full step in the Newton direction and which guarantees second order convergence near optimum. If the Wolfe-Powell conditions are not satisfied then we can consider, for instance, the second-degree and two-variables polynomial :

$$P(\lambda_1, \lambda_2) = a_2\lambda_1^2 + b_2\lambda_2^2 + a_1\lambda_1 + b_1\lambda_2 + b_0, \qquad (75)$$

which interpolate $\varphi(\lambda_1, \lambda_2)$ in convenient points and generate new trial-points in the same way the quadratic fit do in line search.
In some very initial computational experiences, NEWYG behaves very similar to those algorithms who use trust-region search and rather better to those who use classical line-search.

4.2 Methods for Optimal Control problems

We will consider now optimal control problems of the form :

$$\begin{aligned} \dot{x}(t) &= f(x(t), u(t), t), \quad t \in [0, T] && (76)\\ x(0) &= x_0 && (77)\\ u(t) &\in [-M, M]^r && (78)\\ \min J(u) &= \varphi\left[x(T)\right] && (79) \end{aligned}$$

A theoretical convergence theorem for Hilbert spaces could be applied directly but the usual procedure in optimal control computations does not follows that general algorithm. In order to solve the problem in real computations, they always make four fundamental choices :
a) a finite dimensional variety $\mathcal{U} \subset L_2\left([0,T], \mathbb{R}^m\right)$ in which the problem will be solved,
b) a numerical optimization method which will be used,
c) a numerical method with the ODE constraint will be integrated in order to evaluate the objective function, and
d) a method for calculating the objective function gradients.
The answers of these questions influence all the behaviour of computations but we will consider only some theoretical aspects of convergence of certain class of algorithms defined as follows :
Algorithmic Process for Optimal Control (OCp) :
1) Choose $p \in \mathbb{N}$, $\rho > 0$, a feasible function $u_0(.) \in L_2$, and set $k = 0$.
2) Select $p_k \in \mathbb{N}$, such that $p = mp_k$, for some non-zero $m \in \mathbb{N}$, and step $h_k = \frac{T}{p_k} > 0$, defining a partition :

$$t_0^k = 0,\ t_{i+1}^k = t_i^k + h_k,\ i = 0, 1, ..., p_k - 1,\ t_{p_k}^k = T, \tag{80}$$

and a subspace :

$$\begin{aligned} \mathcal{U}_{p_k} = \mathcal{U}(t_1^k, ..., t_{p_k}^k) = \{u(., \lambda) \in L_2 \mid u(t) = u(t, \lambda) = \\ = \textstyle\sum_{i=0}^{p_k - 1} \lambda_i\ \mathbf{1}_{[t_i^k, t_{I+1}^k)}(t)\} \end{aligned} \tag{81}$$

such that, there exists a vector $\lambda = (\lambda_1, ..., \lambda_{p_k})$, i.e. a function $u(., \lambda) \in \mathcal{U}_{p_k}$, satisfying :

$$\begin{aligned} \prec \nabla J(u_k), u(., \lambda) \succ \ &\leq\ -\rho\ \|\nabla J(u_k)\|_{L_2} \|\lambda\|, && (82)\\ -M \leq u_k(t) + u(t, \lambda)\ &\leq\ M,\ \forall t \in [0, T]. && (83) \end{aligned}$$

If such p_k and function $u(.)$ do not exist, then we take $u_k(.)$ as a final solution. If such p_k and function $u(.)$ do exist, then continue to step 3.

3) Find $\hat{\lambda} = (\hat{\lambda}_1, ..., \hat{\lambda}_{p_k})$ such that :

$$J(u_k + u(.,\hat{\lambda})) = \min \ \{J(u_k + u(.,\lambda) \mid \lambda \in I\!R^p, u(.,\lambda) \text{ satisfies } (83)\} \quad (84)$$

4) Take $u_{k+1} = u_k + u(.,\hat{\lambda})$, $k = k+1$, and go to step 2).
It is clear that (OCp) is an (Ap_k) process in $H = L_2$ with bounded $\{p_k\}$. The step functions $v_i^k(t) = \mathbf{1}_{[t_i^k, t_{I+1}^k)}(t)$, $i = 1, ..., p_k$, play the role of search directions and are always linearly independent. Note that we also have, from (82), the relation :

$$\prec \nabla J(u_k), u(.,\lambda) \succ \leq -\frac{\rho}{T^{1/2}} \ \|\nabla J(u_k)\|_{L_2} \|u(.,\lambda)\|_{L_2}, \tag{85}$$

since

$$\|u(.,\lambda)\|_{L_2} = \left(\sum_{i=0}^{p_k-1} \int_{t_i}^{t_{i+1}} \lambda_i^2 dt\right)^{1/2} = \left(\sum_{i=0}^{p_k-1} \lambda_i^2 (t_{i+1} - t_i)\right)^{1/2} \leq T^{1/2} \ \|\lambda\|,$$

and therefore the conditions for global convergence are fulfilled for the point-to-set maps : $G(u) =$

$$\left\{ \begin{array}{c} (u,D) \in \{u\} \times H^p \mid D_i = v_i, i = 1, ..., p, \text{and} \exists \lambda \in I\!R^p : \lambda \neq 0, \\ \prec \nabla J(u), D\lambda \succ = \prec \nabla J(u), d(.,\lambda) \succ \leq -\frac{\rho}{T^{1/2}} \ \|\nabla J(u)\|_{L_2} \|d(.,\lambda)\|_{L_2} \end{array} \right\},$$

$$S(u,D) = \{v \in H \mid v = u_k + D\lambda, \text{ and } v \text{ solves } (84)\}, \tag{86}$$

where $d(t,\lambda) = D\lambda(t) = \sum_{i=0}^{p-1} \lambda_i v_i(t)$, and $v_i(t) = \mathbf{1}_{[t_i, t_{I+1})}(t)$, $t_i = ih, i = 1, ..., p-1$, $T = hp$.
In fact, the problem in (84) is finite dimensional, D is full rank, and under the assumption of exact calculation of J and ∇J, the map S is closed (since in finite dimensions weak and strong continuity of J are equivalent). It is also possible to applied a usual optimization routine for its solution.
Furthermore G is continuous with image in a finite dimensional subspace $\mathcal{U}$ of H :

$$\mathcal{U} = \mathcal{U}_p = \left\{ u(.,\lambda) \in L_2 \mid u(t) = u(t,\lambda) = \sum_{i=0}^{p-1} \lambda_i \ \mathbf{1}_{[t_i, t_{I+1})}(t) \right\}.$$

Note that the functions belonging to $\mathcal{U}_{p_k}$ also belongs to $\mathcal{U}$, since p is a multiple of p_k. Therefore we have two different cases :
i) The algorithm (OCp) generates an infinite sequence u_k, all of its limit points $\tilde{u}$ are solutions (they satisfies $\nabla J(\tilde{u}) = 0$). If some $\tilde{p} \in I\!N$, $\tilde{p} \leq p$, is infinitely many repeated in $\{p_k\}$ then the corresponding limit point $\tilde{u}$ belongs to $\mathcal{U}_{\tilde{p}}$, and

it means that there would be solutions of the control problem belonging to a (possible less dimensional) subspace of $\mathcal{U}$.
ii) at some finite step N the algorithm stops at step 2), and we obtain a final point u_N for which $G(u_N) = \emptyset$. It means that in the hold subspace $\mathcal{U}$ we can't improve J and then we only can say that $\prec \nabla J(u_N), u \succ \geq 0$, for all feasible u belonging to $u_N + \mathcal{U}$.
Case *ii)* gives points that are not solutions of the original problem (even with exact calculation of J and ∇J). This is not a surprise because the process rigidly choose the form of the search directions, and its global convergence strongly depends on the partition defined by p and the initial solution u_0, i.e. we will have convergence if and only if a local minimum u^* belongs to the linear variety $u_0 + \mathcal{U}$.
In the general case, to obtain a global convergent process is sufficient to add a column to the matrix D, which be a uniform descent function of u, like $(-\nabla J(u))$ for example, but this procedure is not commonly found in practice.
As a final note to this part, the convergence obtained is in the L_2-norm, and this means that we don't have exact information about how close are the partition points t_i to the real jumping-points (discontinuities) of the optimal control, when the process is stopped after a finite number of steps (as usual).
Now we suppose that the calculations of J and ∇J can't be done exactly and, for instance, a multi-step integrator scheme (of variable order) is used, i.e. at each step k, the continuous-time optimal control problem is dropped and changed by the following approximated discrete-time optimal control problem :

$$\min \; J_{\mathsf{h}_k(u)}(u) = \varphi\left[x_N\right], \tag{87}$$

$$x_{i+1} = x_i + \mathsf{h}_i \Phi_i \left[x_i, u(\tau_i), ..., x_{i-\nu_i}, u(\tau_{i-\nu_i})\right], \; i = 0, 1, .., N-1, \tag{88}$$

$$\tau_{i+1} = \tau_i + \mathsf{h}_i, \; i = 0, 1, ..., N-2, \tag{89}$$

$$x_0, \; \tau_0, \mathsf{h}_i, \; \nu_i \text{ - given, } 0 \leq \nu_i \leq i, \; i = 0, 1, ..., N-1, \tag{90}$$

$$u(.) \in u_k(.) + \mathcal{U}. \tag{91}$$

The τ_i define the partition of integration and ν_i the order of the scheme used. The steps h_i could be fixed constants, or a priori given at each algorithmic step k, but we suppose they are resulting of an step generator, which can be modelled by the following equations :

$$\mathsf{h}_{i,j} = \Theta(\mathsf{h}_{i,j-1}, \epsilon_{i,j-1}), i = 0, ..., N-1, \; j = 1, ..., j_0, \tag{92}$$

$$\mathsf{h}_{0,0}\text{- given, } \mathsf{h}_{i+1,0} = \mathsf{h}_i, i = 0, ..., N-2 \tag{93}$$

$$\epsilon_{i,j} = \Omega_i \left[x_i, u(\tau_i), ..., x_{i-\nu_i}\right],$$

$$\left[u(\tau_{i-\nu_i}), \mathsf{h}_{i,j}\right], i = 0, ..., N-1, j = 1, ..., j_0,$$

$$\epsilon_{0,0} = 0,$$

$$\mathsf{h}_{i+1} = \sum_{j=1}^{j_0} \mathsf{h}_{i,j} \Pi_j(\epsilon_{i,1}, ..., \epsilon_{i,j}), \tag{94}$$

Equation (92) define a trial-step generator depending on the previous trial-step and error. It is supposed that a fixed number j_0 of trial-step is calculated (in fact, the common integration routines generate a decreasing sequence of trial-steps $\mathsf{h}_{i,j}$, $j = 0, 1, ...,$ and has a minimum step length $\mathsf{h}_{\min}$; when this minimum is attained the overall computation process is aborted). Equation (93) means that the initial trial-step $\mathsf{h}_{i+1,0}$ at $i+1$ is precisely the step length h_i, just selected previously at step i. Equation (4.2) define an approximated local error calculator, using the trial-step $\mathsf{h}_{i,j}$, and equation (94) gives the rule to choose the successfull step which depends on calculated trial-steps and all the error history. A model for the order selections ν_i could be added also, but we didn't want to extend more this example. Note, however, that the selection of steps h_i strongly depends on the error calculator functions Ω_i which, in turns, are always dependent on u. For this reason we denote the discrete objective function as $J_{\mathsf{h}_k(u)}(u)$, which approximately gives the real objective function value $J(u)$.
We assume the gradient $\nabla J(u)$ is approximated in the following (very popular) way : Forget the dependence of h on u, and *after* the calculation of the complete discrete solution x_i, when we already know the used steps h_i, calculate the gradient $\nabla J_{\mathsf{h}_k}(u)$ of the function $J_{h_k}(u) = \varphi\left[x_N\right]$, where x_N is given by (88) to (91). Formulas for this gradient are well known (see [13] or [9]) but is precisely here where the global convergence of any optimization algorithm, applied for the solution of the subproblem (84), can't be ensured. At each step k and each time when we calculate approximated values and gradients of J, we change the discrete model (because different steps $\mathsf{h}_i(u)$ correspond, in general, to different u), and we cannot appy the global convergence theorems to the situation of *variable objective function and variable constraints.*
Nevertheless, we can consider the *complete model* (87) to (94) as the definition of the subproblem (84) and in this case, where we didn't drop the dependence of h on u, from general theory we can obtain the global convergence of many optimization algorithm applied to (84), but always if we calculate the gradient of the correct function $\nabla J_{\mathsf{h}_k(u)}(u)$. However in many practical situations, from roundoff errors, both gradients $\nabla J_{\mathsf{h}_k(u)}(u)$ and $\nabla J_{\mathsf{h}_k}(u)$ are very near, and

in other usual and important cases (of step generators and selectors) they are theoretically coincidents. One example of this last situation is the commonly used functions :

$$\begin{aligned} \Theta(\mathsf{h}_{i,j-1}, \epsilon_{i,j-1}) &= \theta \ \mathsf{h}_{i,j-1} \\ \Pi(\epsilon_{i,1}, ..., \epsilon_{i,j}) &= \mathbf{1}_{[0,\varepsilon]^{[j]}}(\epsilon_{i,1}, ..., \epsilon_{i,j}), \end{aligned}$$

where $\theta \in (0,1)$ and using the notation :

$$[0,\varepsilon]^{[1]} = [0,\varepsilon]\,;\ [0,\varepsilon]^{[j+1]} = [0,\varepsilon]^c \times [0,\varepsilon]^{[j]}\,,\ j = 1, ..., j_0 - 1,$$

$[0,\varepsilon]^c$ is the complement set $I\!R \setminus [0,\varepsilon]$ and the symbol $\times$ denote the cartesian product.

This means that we construct the next trial step multiplying the old one by a fixed factor θ (avoiding local error considerations), and we select as h_{i+1} the first trial-step $\mathsf{h}_{i,j}$, $j = 0, 1, ..., j_0$, for which the local error $\epsilon_{i,j}$ satisfies a given error tolerance $\varepsilon > 0$. In this case, it is easy to see that h_{i+1} is a sum of constant step-functions and $\epsilon_{i,j}$ does not depends on u, therefore its derivative respect to u is zero, and then $\nabla J_{\mathsf{h}_k(u)}(u) = \nabla J_{\mathsf{h}_k}(u)$. This explain why the common users of many ODE integration routines can work much time without computational convergence problems in step 3) of (OCp) while solving optimal control problems.

A great number of interesting questions remain open.

References

[1].- Bazaraa, M.; G.M. Sheety, J.J., *On the convergence of a class of algorithms using linearly independent search directions.* Mathematical Programming ,Vol. 18, No. 1, 1980, 89-93.

[2].- Brézis, H., *Análisis funcional. Teoría y aplicaciones.* Alianza Editorial, Madrid (in spanish), 1984.

[3].- Dennis, J.E.; R. Schnabel, *Numerical methods for unconstrained optimization and nonlinear equations.* Prentice Hall, London, 1983.

[4] Evtuchenko, Y.G., *Methods for the solution of extremal problems and its application on optimization systems.* Nauka, Moscow (in russian), 1982.

[5].- Gill, P.E.; W. Murray, M.H. Wright, *Practical optimization.* Academic Press, London, 1981.

[6].- Gómez, J.A.; W. Gómez, *Multidirectional search for Nonlinear Programming and Optimal Control problems.* Research Report CEMAFIT, Havana (In print), 1995.

[7].- Luenberger, D., *Introduction to linear and nonlinear programming.* Addison-Wesley Pub., Massachusets, 1984.

[8] Luenberger, D., *Optimization by vector space methods.* John Wiley and Sons, New York, 1969.
[9].- Marrero, A.; J.A. Gómez, *Problema de estimación de parámetros en modelos dinámicos no lineales.* Ciencias Matemáticas, (in spanish) Vol. 15, No. 1, 1994, 41-54.
[10].- Nocedal, J.; D.C. Lui, *On the limited memory BFGS method for large scale optimization.* Mathematical Programming, Vol. 45, No. 3, 1989, 503-528.
[11].- Polak, E., *Computational methods in Optimization : A unified approach.* Academic Press, New York, 1971.
[12].- 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.
[13].- Roemish, W.; M.L. Baguer, *Computing gradients in parametrization- discretization schemes for constrained optimal control problems.* Technical report No. 508, Humboldt University-Berlin, 1994.
[14].- Rudin, W. *Functional Analysis*, Mac.Graw-Hill, 1973.