THE METHOD OF DUBOVITSKII-MILYUTIN IN MATHEMATICAL PROGRAMMING*

Hubert Halkin**

Department of Mathematics
University of California at San Diego
La Jolla, California

1. INTRODUCTION

I want to give here a brief description of a very attractive formalism in optimization theory: the method of Dubovitskii and Milyutin [1] and the relate some recent extensions of that method, Halkin [2], with the necessary condition of Fritz John[3].

The first step in the method of Dubovitskii-Milyutin is to notice that, in any optimization problem, to say that some solution is optimal is equivalent of saying that a certain family of sets $\{S_i : i \in I\}$ have no points in common i.e. $\bigcap_{i\in I} S_i = \Phi$. Consider for example the optimization problem consisting in minimizing a function f over all point of the plane R^2 where a function g is nonpositive. To say that an element $\hat{x}$ in R^2 with $g(\hat{x}) \leq 0$ is an optimal solution for that problem is equivalent of saying that the sets $S_1 = \{x : x \in R^2, f(x) < f(\hat{x})\}$ and $S_2 = \{x : x \in R^2, g(x) \leq 0\}$ have no points in common. In general we can do very little with two disjoint sets S_1 and S_2. But if S_1 and S_2 happen to be convex and nonempty then we can separate them by an hyperplane, i.e. we can find a nonzero vector p in R^2 such that

$$\sup_{x\in S_1} p\cdot x \leq \inf_{x\in S_2} p\cdot x .$$

Here the scalar product of two vectors p and x is denoted by $p\cdot x$. We shall show below that this separation, when possible, leads to interesting results.

The second step in the method of Dubovitskii-Milyutin is to repace each set S_i by a set Ω_i which is convex and such a good approximation of S_i (in a sense to be precised later) that from the fact that the sets S_i have no point in common, i.e. $\bigcap_{i\in I} S_i = \Phi$, we can prove that the sets Ω_i will have no point in common, i.e. $\bigcap_{i\in I}\Omega_i = \Phi$.

* This research was supported by the Air Force Office of Scientific Research, under Grant No. AFOSR-68-1529C.

**On leave for the academic year 1971-1972 at CORE, University of Louvain, Belgium.

In the simple example given above, if f and g are differentiable at x, we could consider the convex sets

$\Omega_1=\{x : x\in R^2, f(\hat{x})+(x-\hat{x})\cdot \text{grad } f(\hat{x})<f(\hat{x})\}=\{x : x\in R^2, (x-\hat{x})\cdot \text{grad } f(\hat{x})<0\}$ and

$\Omega_2=\{x : x\in R^2, g(\hat{x})+(x-\hat{x})\cdot \text{grad } g(\hat{x})<0\}$.

The third step in the method of Dubovitskii-Milyutin is to go from the fact that the sets Ω_i have no point in common, i.e. $\bigcap_{i\in I} \Omega_i=\Phi$ to the fact that the sets $\{\Omega_i : i\in I\}$ can be separated. We shall define later what we mean by separating more than two convex sets. For the time being let us go back to our simple example where we have only two sets Ω_1 and Ω_2 and let us see how the separation of the two convex sets Ω_1 and Ω_2 is equivalent to the known necessary condition for that problem: if $\hat{x}$ is optimal and grad $g(\hat{x})\neq 0$, then for some $\lambda \leqslant 0$ we have $\lambda g(\hat{x}) = 0$ and $-\text{grad } f(\hat{x}) + \lambda \text{grad } g(\hat{x}) = 0$. The set Ω_1 will always be empty if grad $f(\hat{x}) \neq 0$ and the set Ω_2 will always be empty if grad $g(\hat{x}) \neq 0$. We shall temporarily assume that grad $f(\hat{x})$ and grad $g(\hat{x})$ are different from zero. Since for i=1 and 2 the point $\hat{x}$ belongs to $\overline{\Omega}_i$, (the closure of Ω_i), i.e. since there are points in Ω_i arbitrarily close to $\hat{x}$, then the hyperplane separating Ω_1 and Ω_2 must pass through $\hat{x}$, i.e. there exists a nonzero vector p such that $\sup_{x\in\Omega_1} p\cdot x = p\cdot\hat{x} = \inf_{x\in\Omega_2} p\cdot x$. In other words, for all $x\in\Omega_1$, i.e. for all x such that $(x-\hat{x})\cdot \text{grad } f(\hat{x}) < 0$ we have $p\cdot x \leqslant p\cdot\hat{x}$, i.e. $(x-\hat{x})\cdot p \leqslant 0$. Since $p \neq 0$, this means that for some $a > 0$ we must have grad $f(\hat{x}) = ap$. Moreover, for all $x \in \Omega_2$, i.e. for all x such that $g(\hat{x}) + (x-\hat{x})\cdot \text{grad } g(\hat{x}) < 0$, we have $p\cdot x \geqslant p\cdot\hat{x}$; i.e. $(x-\hat{x})\cdot p \geqslant 0$. Since $p \neq 0$, this means that $g(\hat{x})=0$ and that for some $b > 0$, we have $p=-b$ grad $g(\hat{x})$ and $bg(\hat{x})=0$. Combining those two results, we obtain that, under the assumptions grad $f(\hat{x})\neq 0$ and grad $g(\hat{x})\neq 0$, we have $-\text{grad } f(\hat{x}) + \lambda\text{grad } g(\hat{x}) = 0$ and $\lambda g(\hat{x})=0$ for some $\lambda=-ab < 0$. In the case where grad $g(\hat{x}) \neq 0$ and grad $f(\hat{x}) = 0$, we can state trivially that $-\text{grad } f(\hat{x}) + \lambda\text{grad } g(\hat{x}) = 0$ and $\lambda g(\hat{x}) = 0$ by letting $\lambda=0$. We have thus obtained the classical result already mentioned above: if $\hat{x}$ is optimal and grad $g(\hat{x}) \neq 0$, then for some $\lambda\leqslant 0$ we ve $\lambda g(\hat{x}) = 0$ and $-\text{grad } f(\hat{x}) + \lambda\text{grad } g(\hat{x}) = 0$.

Without the assumption grad $g(\hat{x}) \neq 0$, this classical result would be incorrect as one can see in the following simple optimization problem on the real line R^1 : minimize $f(t)\equiv t$ subject to the constraint $g(t)\equiv t^2<0$. The optimal solution is obviously $\hat{t}=0$. But since grad $f(0)=1$ and grad $g(0)=0$, it is impossible to find a real number $\lambda\leqslant 0$ such that $-\text{grad } f(0) + \lambda \text{grad } g(0)=0$, i.e. such that $-1 + \lambda\cdot 0 = 0$. The reader should realize that in this simple pathologival problem the sets Ω_1 and Ω_2 are nevertheless disjoint since Ω_1 is the set $\{t:t < 0\}$ and Ω_2 is the empty set.

The assumption grad $g(\hat{x}) \neq 0$ is the most benign form of a general class of assumptions known as constraint qualifications; I shall come back to that topic in Section 5. If we do not assume that grad $g(\hat{x})\neq 0$, then the necessary condition for that problem takes the form: if $\hat{x}$ is optimal,then for some $(\alpha,\beta)\neq 0$ with α and $\beta\leqslant 0$, we have α grad $f(\hat{x})$ + βgrad $g(\hat{x})=0$ and $\beta g(\hat{x})=0$. Indeed if grad $g(\hat{x})\neq 0$, then we let $\alpha=-1$ and $\beta=\lambda$, where λ is the number given earlier; if on the other hand we have grad $g(\hat{x})=0$, then either $g(\hat{x})=0$ or $g(\hat{x})<0$. If $g(\hat{x})=0$, we let $\alpha=0$ and $\beta=-1$. If $g(\hat{x})<0$, then the fact that the sets Ω_1 and Ω_2 are disjoint implies that grad $f(\hat{x})=0$ and in that case we let $\alpha=-1$ and $\beta=0$.

In this paper, I shall always use the n-dimensional Euclidean space R^n as the basic reference space. Although this is sufficient for most applications to mathematical programming, this is not the case in the theory of optimal control in which we must consider infinite-dimensional spaces of trajectories. However, the reader should be aware that everything stated here can be extended to general normed linear spaces, Halkin [2], in which optimal control problems can be treated.

2. SEPARATING SEVERAL CONVEX SETS

If p is a vector in R^n and if α is a real number, then the function f defined over R^n by the relation $f(x)=p\cdot x + \alpha$ is called an affine function on R^n. Colloquially speaking, an affine function is a linear-plus-a-constant function. If $p\neq 0$, then the affine function $p\cdot x + \alpha$ is said to be an affine nonconstant function. A finite family $\{\Omega_i : i\in I\}$ of convex sets will be said to be separated, if there exists a finite family of affine functions $\{\omega_i : i\in I\}$ such that

(i) $\Sigma_{i\in I}\, \omega_i =0$,

(ii) $\omega_i(x) \geqslant 0$ for all $i\in I$ and all $x\in \Omega_i$,

(iii) ω_i is nonconstant for some $i\in I$.

Let us show that in the case of two sets Ω_1 and Ω_2, this new definition of separation coincides with the classical definition of separation. If Ω_1 and Ω_2 are (classically) separated, then there exists a nonzero vector p such that $\sup_{x\in\Omega_1} p\cdot x \leqslant \inf_{x\in\Omega_2} p\cdot x$. If we let $\omega_1(x)=-p\cdot x + \sup_{x\in\Omega_1} p\cdot x$ and $\omega_2(x)=p\cdot x - \sup_{x\in\Omega_1} p\cdot x$, we thus obtain

(i) $\omega_1 + \omega_2 = 0$,

(ii) $\omega_i(x) \geqslant 0$ for all $i\in\{1,2\}$ and all $x\in\Omega_i$,

(iii) both ω_1 and ω_2 are nonconstant.

In other words, Ω_1 and Ω_2 are separated according to the new definition. Conversely, if Ω_1 and Ω_2 are separated according to the new definition, i.e. if for some $\omega_1(x)=p_1\cdot x + \alpha_1$ and $\omega_2(x) = p_2\cdot x + \alpha_2$, we have

(i) $\omega_1 + \omega_2=0$,

(ii) $\omega_i(x) \geqslant 0$ for all $i\in\{1,2\}$ and all $x\in\Omega_i$,

(iii) at least one of the vectors p_1 or p_2 is different from zero.

Then, by (i), we have $p_2 = -p_1$, $\alpha_2 = -\alpha_1$ and, by (iii), we have $p_2 = -p_1 \neq 0$. If we let $p = p_2 = -p_1$, we obtain $\sup_{x\in\Omega_1} p\cdot x \leqslant \inf_{x\in\Omega_2} p\cdot x$ and the two sets Ω_1 and Ω_2 are separated according to the classical definitions.

We know that two disjoint nonempty convex sets in R^n can be separated, but it is not correct to say that two separated convex sets in R^n are disjoint. For instance the sets $\Omega_1 = \{(x_1,x_2): x_1\leqslant 0\}$ and $\Omega_2 =\{(x_1,x_2): x_1\geqslant 0\}$ in the plane R^2 are not disjoint, since $\Omega_1 \cap \Omega_2 = \{(x_1,x_2): x_1=0\} \neq \Phi$ but they are separated since for $p=(1,0) \neq 0$, we have $\sup_{x\in\Omega_1} x\cdot p = \inf_{x\in\Omega_2} x\cdot p$. If either Ω_1 or Ω_2 is open however, we know that the fact that Ω_1 and Ω_2 are separated will imply that Ω_1 and Ω_2 are disjoint. The same result can be extended to several convex sets in the following manner:

Theorem 2.1. If $\{\Omega_i : i\in I\}$ is a finite family of nonempty convex sets in R^n such that $\bigcap_{i\in I} \Omega_i$ is empty, then the family $\{\Omega_i : i\in I\}$ can be separated. Conversely, if $\{\Omega_i : i\in I\}$ is a finite separated family of convex sets in R^n and if at most one of them fails to be open, then $\bigcap_{i\in I} \Omega_i$ is empty.

The proof of Theorem 2.1 can be found in Halkin [2].

In several applications, we shall assume that $0\in \overline{\Omega}_i$ for each $i\in I$. In that case, if the family $\{\Omega_i : i\in I\}$ of convex sets is separated by the family $\{\omega_i : i \in I\}$ of affine functions, then we shall have $\omega_i(0) \geqslant 0$ for each $i \in I$ and $\Sigma_{i \in I}\ \omega_i(0)=0$ which imply $\omega_i(0)=0$ for each $i \in I$ and hence the functions ω_i are not only affine but linear, i.e. of the form $\omega_i(x)=p_i \cdot x$. We can thus state that a family $\{\Omega_i : i \in I\}$ of convex sets with $0 \in \overline{\Omega_i}$ for each $i \in I$ is separated if and only if there exists a finite set of vectors $\{p_i : i \in I\}$ such that

(i) $\Sigma_{i \in I}\ p_i = 0$, (ii) $p_i \cdot x \geqslant 0$ whenever $i \in I$ and $x \in \Omega_i$,

(iii) $p_i \neq 0$ for some $i \in I$.

3. CONVEX APPROXIMATIONS OF SETS

We shall consider three different types of convex approximations of sets: (i) the interior convex approximation (ii) the tangent convex approximation and (iii) the simplicial^k convex approximation where k is a positive integer. In mathematical programming, the two concepts of interior convex approximation (asociated with the objective function and the inequality constraints) and of tangent convex approximation (associated with the equality constraints) are the most useful. The simplicial^k convex approximation is used chiefly in optimal control theory and is associated with operator constraints (i.e. when one requires a trajectory to satisfy some differential equations). However, in the case k=1, i.e. in the case of the simplicial^1 convex approximation, this concept is also used in mathematical programming under the form of the Abadie Sequential Constraint Qualification, Abadie [4].

We should normally give all the definitions under the form: the set Ω is an interior (resp. tangent or simplicial^k convex approximation around a point $\hat{x}$ to a set S, if .... For the sake of simplicity of notation, we shall give all those definitions with respect to the point $\hat{x}=0$. To go back to the general case, we shall use the following convention: the set Ω is an interior (resp. tangent or simplicial^k) convex approximation around a point $\hat{x}$ to the set S, if the set $\Omega-\hat{x}$ is an interior (resp. tangent or simplicial^k convex approximation to the set $S-\hat{x}$). If A is the set in R^n and a is a vector in R^n, then we use the notation A-a to denote the set $\{x-a : x \in A\}$. Let us specify some further notations. If $x \in R^n$, then $|x|$ will be the Euclidean length of x. If $A \in R^n$, then coA will be the convex hull of A. A set $\{x_1, \ldots, x_\ell\}$ in R^n is said to be in general position if the vectors $x_2-x_1, x_3-x_1, \ldots, x_\ell-x_1$ are linearly independent.

Definition 3.1. A subset Ω of R^n is an interior convex approximation to a subset S of R^n if (i) Ω is open, (ii) Ω is convex, (iii) $0\in\bar{\Omega}$ and (iv) for all $\bar{x}\in\Omega$ there exists an $\varepsilon>0$ such that $\eta x\in S$ whenever $|x-\bar{x}|<\varepsilon$ and $\eta\in(0,\varepsilon)$.

Definition 3.2. A subset Ω of R^n is a tangent convex approximation to a subset S of R^n if there exists a neighborhood V of 0 and a continuous real-valued function ϕ defined on V, differentiable at x=0 and such that (i) grad $\phi(0)\neq 0$, (ii) $\phi(0)=0$, (iii) $\Omega=\{x:x\in R^n, x\cdot \text{grad}\ \phi(0)=0\}$ and (iv) $\{x:x\in V,\phi(x)=0\}\subset S$.

Definition 3.3. If k is a positive integer, we shall say that a subset Ω of R^n is a simplicial^k convex approximation to a subset S of R^n, if (i) Ω is convex, (ii) $0\in\bar{\Omega}$ and (iii) for any set $\{x_1,\ldots,x_\ell\}$ with $\ell\leq k$ elements in general position in Ω and for any real number $\varepsilon>0$ there exists a number $\eta\in(0,\varepsilon)$ and a continuous function ζ from $\text{co}\{x_1,\ldots,x_\ell\}$ into R^n such that $|\zeta(x)-x|\leq\varepsilon$ and $\eta\zeta(x)\in S$ whenever $x\in\text{co}\{x_1,\ldots,x_\ell\}$.

Remark: As I mentioned before,the concept of simplicial^1 convex approximation is related to Abadie Sequential Constraint Qualification. Indeed from Definition 3.3,we have: a subset Ω of R^n is a simplicial^1 convex approximation to a subset S of R^n if (i) Ω is convex,(ii) $0\in\bar{\Omega}$ and (iii) for each $\bar{x}\in\Omega$ and each real number $\varepsilon>0$, there exists a number $\eta\in(0,\varepsilon)$ and an element $y\in R^n$ such that $|y-\bar{x}|\leq\varepsilon$ and $\eta y\in S$. The last condition can be rewritten as: for each $\bar{x}\in\Omega$, there exists a sequence of positive real numbers $\eta_1,\eta_2,\ldots$ and a sequence of elements $y_1,y_2,\ldots$ in R^n such that $\lim_{i\to\infty}|y_i-\bar{x}|=0$, $\lim_{i\to\infty}\eta_i=0$ and $\eta_i y_i\in S$ for all $i=1,2,\ldots$.

Examples of Convex Approximations. If ϕ is a real-valued function defined on R^n such that (i) $\phi(0)\leq 0$ and (ii) grad $\phi(0)$ exists and is different from zero, then $\Omega=\{x:x\in R^n, x\cdot\text{grad}\ \phi(0)<0\}$ is an interior convex approximation to each of the sets $S=\{x:x\in R^n, \phi(x)<0\}$ and $\bar{S}=\{x:x\in R^n,\phi(x)\leq 0\}$, (the proof of that fact is not too hard). If moreover,$\phi(0)=0$ and ϕ is continuous in some neighborhood of 0, then $\Omega^*=\{x:x\in R^n, x\cdot\text{grad}\ \phi(0)=0\}$ is a tangent convex approximation to the set $S^*=\{x:x\in R^n,\phi(x)=0\}$,(there is nothing to prove here, just apply the definition). As I mentioned before,simplicial^k convex approximations are used in optimal control theory to handle operator constraint of the type $x\in S$ where S is the set of all trajectories which are solutions of a given family of ordinary differential equations. It is very hard to express operator constraint in terms of inequality and/or equality constraint(s) and even when it is possible the function describing those constraints are not "smooth" enough to apply the concepts of interior convex approximation and/or tangent convex approximation. This is the reason why it is convenient to keep operator constraints under their given forms and to define a special type of convex approxiamtion adapted to those operator constraints. This special type of convex approximation is the simplicial^k convex approximation. In optimal control theory, the simplicial^k convex approximation Ω to the set S will be the set of all solutions of a certain linearization of the given family of ordinary differential equations. For more details, see Halkin [5] and Halkin-Neustadt [6].

4. THE THEOREM OF DUBOVITSKII AND MILYUTIN

Theorem 4.1. Let $S_{-\mu},\dots,S_{-1},S_0,S_1$ be subsets of R^n such that $\bigcap_{i=-\mu}^{+1} S_i=\Phi$. Assume that we have convex sets $\Omega_{-\mu},\dots,\Omega_1$ such that Ω_i is an interior convex approximation to S_i for each $i=-\mu,\dots,0$ and such that Ω_1 is a simplicial^1 convex approximation to S_1. Then, the sets $\Omega_{-\mu},\dots,\Omega_1$ are disjoint and hence separated.

The proof of Theorem 4.1 is particularly simple. If the sets $\Omega_{-\mu},\dots,\Omega_1$ are not disjoint, then there exists an element $\bar{x}$ which belongs to each of them. Since for each $i\in\{-\mu,\dots,0\}$ the set Ω_i is an interior convex approximation to the set S_i, we know that there exists an $\varepsilon_i>0$ such that $\eta x\in S_i$ whenever $|x-\bar{x}|<\varepsilon_i$ and $\eta\in(0,\varepsilon_i)$. If we let $\varepsilon=\min\{\varepsilon_{-\mu},\dots,\varepsilon_0\}$ we see that $\eta x\in S_i$ whenever $i\in\{-\mu,\dots,0\}$, $|x-\bar{x}|<\varepsilon$ and $\eta\in(0,\varepsilon)$. Since $\bar{x}\in\Omega_1$ and since Ω_1 is a simplicial^1 convex approximation to S_1, we know that there exists a number $\eta\in(0,\varepsilon)$ and an element $y\in R^n$ such that $|y-\bar{x}|<\varepsilon$ and $\eta y\in S_1$. Since $\eta\in(0,\varepsilon)$, the element ηy belongs to every S_i for all $i\in\{-\mu,\dots,0\}$. This contradiction concludes the proof of Theorem 4.1.

The results of Theorem 4.1 can be applied directly to the following mathematical programming problem: given a set $S_1\subset R^n$ and given functions $\phi_{-\mu},\dots,\phi_{-1},\phi_0$ defined over R^n, find an element $x\in S_1$ which minimizes $\phi_0(x)$ subject to constraints $\phi_i(x)\leq 0$ for $i=-\mu,\dots,-1$. We assume that an optimal solution $\hat{x}$ exists for that problem, that Ω_1 is a simplicial^1 convex approximation to S_1 around $\hat{x}$ and that $\phi_{-\mu},\dots,\phi_0$ are differentiable at $\hat{x}$. We then obtain

Theorem 4.2. Under the preceeding assumptions, there exist numbers $\lambda_{-\mu},\dots,\lambda_0$, not all zero, such that

(i) $\lambda_i \leq 0$ for each $i \in \{-\mu,\ldots,0\}$,

(ii) $\lambda_i \phi_i(\hat{x}) = 0$ for each $i \in \{-\mu,\ldots,-1\}$,

(iii) $\Sigma_{i=-\mu}^{0} \lambda_i$ grad $\phi_i(\hat{x})\cdot(x-\hat{x}) \leq 0$ for all $x \in \Omega_1$.

Proof of Theorem 4.2: Without loss of generality, we assume that $\hat{x}=0$ and that $\phi_0(0)=0$. For each $i \in \{-\mu,\ldots,0\}$, let $S_i=\{x:\phi_i(x)<0\}$ and let $\Omega_i=\{x:\phi_i(0)+\text{grad } \phi_i(0)\cdot x<0\}$. According to Theorem 4.1, the sets $\{\Omega_i : i=-\mu,\ldots,1\}$ are disjoint and hence separated. By construction, we have $0 \in \overline{\Omega}_i$ for all $i \in \{-\mu,\ldots,+1\}$ and hence there exists a set of vectors $\{p_{-\mu},\ldots,p_1\}$, not all zero, such that

(i) $p_i \cdot x \geq 0$ whenever $x \in \Omega_i$ and $i \in \{-\mu,\ldots,+1\}$,

(ii) $\Sigma_{i=-\mu}^{+1} p_i=0$.

Since $\Sigma_{i=-\mu}^{+1} p_i=0$ and since at least one of the vectors $p_{-\mu},\ldots,p_1$ is different from zero, we must have at least two of the vectors $p_{-\mu},\ldots,p_1$ which are different from zero, and hence at least one of the vectors $p_{-\mu},\ldots,p_0$ must be different from zero. Since $p_i \cdot x \geq 0$ whenever $\phi_i(0)$ + grad $\phi_i(0)\cdot x<0$, it follows that for some $\lambda_i \leq 0$, we have $p_i=\lambda_i$ grad $\phi_i(0)$. We remark here that we may choose $\lambda_i=0$ for all $i \in \{-\mu,\ldots,-1\}$ such that $\phi_i(0)<0$, since in that case we have $p_i=0$. We have $p_1= -\Sigma_{i=-\mu}^{0} p_i$ and hence the relation $p_i \cdot x \geq 0$ for all $x \in \Omega_1$ may be written under the form $(\Sigma_{i=-\mu}^{0} p_i)\cdot x \leq 0$ for all $x \in \Omega_1$. The last inequality is equivalent to relation (iii) under the assumption $\hat{x}=0$. This concludes the proof of Theorem 4.2.

Remark 1. The inequality (iii) of Theorem 4.2 may be written under the form of a Maximum Principle:

$$(\Sigma_{i=-\mu}^{0} \lambda_i \text{ grad } \phi_i(\hat{x}))\cdot\hat{x} \geq (\Sigma_{i=-\mu}^{0} \lambda_i \text{ grad } \phi_i(\hat{x}))\cdot x \text{ for all } x \in \Omega_1.$$

Remark 2. If the point $\hat{x}$ is an interior point of Ω_1, then the inequality (iii) becomes

$$(\text{iii})^* \quad \Sigma_{i=-\mu}^{0} \lambda_i \text{grad } \phi_i(\hat{x}) = 0.$$

This will always be the case for the problems where $S_1=\Omega_1=R^n$. This last form of Theorem 4.2 is known as Fritz John Theorem [3].

5. CONSTRAINT QUALIFICATIONS

We remark here that Theorem 4.2 contains no information about λ_0 besides the fact that $\lambda_0 \leqslant 0$. If we would know that $\lambda_0 < 0$, then we could multiply the entire vector λ by the positive number $-1/\lambda_0$ and we would obtain the same type of necessary conditions with some vector λ^* for which we would have $\lambda_0^* = -1$. A great variety of conditions (Constraint Qualifications) can be imposed on the problem which would allow us to guarantee that there exists some vector λ with $\lambda_0 < 0$. One of the major shortcomings of the traditional presentation of necessary conditions in the mathematical programming literature is, in my opinion, that the concept and the choice of those Constraint Qualifications influence the entire development of the theory of necessary conditions instead of being introduced at the last minute and used only to prove a variety of, practically important but theoretically easy, corollaries to Theorem 4.2. For example, a very general Constraint Qualification for the problem of Section 4 is to assume that $\overline{\Omega}_{-\mu} \cap \overline{\Omega}_{-\mu+1} \cap \cdots \cap \overline{\Omega}_{-1} \cap \Omega_1$ is a simplicial[1] convex approximation to $S_{-\mu} \cap S_{-\mu+1} \cap \cdots \cap S_{-1} \cap S_{+1}$. In the case $\Omega_1 = S_1 = R^n$, this Constraint Qualification is known as the Abadie Sequential Constraint Qualification. (See the remark following Definition 3.3).

6. LIMITATIONS OF THE METHOD OF DUBOVITSKII AND MILYUTIN

The method of Dubovitskii-Milyutin is not well adapted to problems with equality constraints. I shall examplify those difficulties by considering the following optimization problem in the plane R^2: minimize $\phi_0(x_1,x_2) \equiv x_1$ subject to the constraints $\phi_1(x_1,x_2) \equiv x_2 = 0$ and $\phi_2(x_1,x_2) \equiv x_2 - x_1^2 = 0$. The point $(\hat{x}_1,\hat{x}_2) = (0,0)$ is the obvious optimal solution of that problem. The sets $S_0 = \{(x_1,x_2) : \phi_0(x_1,x_2) < \phi_0(\hat{x}_1,\hat{x}_2)\} = \{(x_1,x_2) : x_1 < 0\}$, $S_1 = \{(x_1,x_2) : \phi_1(x_1,x_2) = 0\} = \{(x_1,x_2) : x_2 = 0\}$ and $S_2 = \{(x_1,x_2) : \phi_2(x_1,x_2) = 0\} = \{(x_1,x_2) : x_2 = x_1^2\}$ have no point in common, but the sets Ω_0, Ω_1 and Ω_2 have points in common (here, Ω_0 is the interior convex approximation to S_0, and Ω_i is the tangent convex approximation to S_i for i=1 and 2). Indeed, $\Omega_0 = \{(x_1,x_2) : x_1 < 0\}$ and $\Omega_1 = \Omega_2 = \{(x_1,x_2) : x_2 = 0\}$ and we have $\bigcap_{i=0,1,2} \Omega_i = \{(x_1,x_2) : x_1 < 0, x_2 = 0\} \neq \Phi$. Of course such "accidents" could be ruled out by conditions resembling some Constraint Qualifications. In the simple example given above for instance, we could require that the set of gradients of the equality constraints be linearly independent at the optimal point. If operator constraints are present in the problem, the situation would still be more complex and one would need further types of Constraint Qualifications. In the next section, I will present a theory of necessary conditions for optimization problems with equality and operator constraints (and also inequality constraints, but they never present any difficulties) which will be independent of any sort of constraint qualifications.

7. THE CASE OF INEQUALITY, EQUALITY AND OPERATOR CONSTRAINTS

The central part of this section is the following result.

Theorem 7.1. If $I=\{-\mu,\dots,m+1\}$ and if $\{S_i:i\in I\}$ and $\{\Omega_i:i\in I\}$ are families of subsets of R^n such that (i) $\bigcap_{i\in I}S_i=\Phi$, (ii) for $i=-\mu,\dots,0$, the set Ω_i is an interior convex approximation to the set S_i, (iii) for $i=1,\dots,m$, the set Ω_i is a tangent convex approximation to the set S_i, and (iv) Ω_{m+1} is an $(m+1)$-convex approximation to the set S_{m+1}, then the family $\{\Omega_i:i\in I\}$ is separated.

We remark immediately that in the case $m=0$, Theorem 7.1 coincides with Theorem 4.1. From the counterexample given in Section 6, we know that under the assumptions of Theorem 7.1, it would be incorrect to say (as in Theorem 4.1) that $\bigcap_{i\in I} S_i=\Phi$ implies that $\bigcap_{i\in I}\Omega_i=\Phi$,but we can still assert that the family $\{\Omega_i:i\in I\}$ is separated and this last statement is all that is needed to obtained appropriate necessary conditions. The proof of Theorem 7.1, given in Halkin [2], makes a critical use of Brouwer Fixed Point Theorem.

Let us now assume that we are faced with the following optimization problem: given a subset S of R^n and functions $\phi_{-\mu},\dots,\phi_m$ defined over R^n, find an element $\hat{x}\in R^n$ which minimizes $\phi_0(\hat{x})$ subject to

(α) the inequality constraints $\phi_i(\hat{x})\leqslant 0$ for $i=-\mu,\dots,-1$,

(β) the equality constraints $\phi_i(\hat{x})=0$ for $i=1,\dots,m$,

(γ) the operator constraint $\hat{x}\in S$.

The optimality of such element $\hat{x}$ can be expressed by writing that $\bigcap_{i=-\mu}^{m+1}S_i=\Phi$, where the sets $S_{-\mu},\dots,S_{m+1}$ are defined by

$$S_i=\{x:x\in R^n,\phi_i(x)\leqslant 0\} \quad \text{for } i=-\mu,\dots,-1,$$

$$S_0=\{x:x\in R^n,\phi_o(x)<\phi_o(\hat{x})\},$$

$$S_i=\{x:x\in R^n,\phi_i(x)=0\} \quad \text{for } i=1,\dots,m$$

and $S_{m+1}=S$.

Let us assume that the functions $\phi_{-\mu},\dots,\phi_m$ are differentiable at $\hat{x}$ and that the functions $\phi_1,\dots,\phi_m$ are continuous in a neighborhood of $\hat{x}$. We then define convex sets $\Omega_{-\mu},\ldots,\Omega_m$ by the relations

$$\Omega_i = \{x : x \in R^n, \phi_i(\hat{x}) + \text{grad } \phi_i(\hat{x})\cdot(x-\hat{x})<0\} \text{ if } i=-\mu,\ldots,-1,$$

$$\Omega_0 = \{x : x \in R^n, \text{grad } \phi_0(\hat{x})\cdot(x-\hat{x})<0\}$$

and

$$\Omega_i = \{x : x \in R^n, \text{ grad } \phi_i(\hat{x})\cdot(x-\hat{x})=0\} \quad \text{for } i=1,\ldots,m.$$

As was mentioned in Section 3, the sets $\Omega_{-\mu},\ldots,\Omega_0$ are interior convex approximations around the point $\hat{x}$ to the sets $S_{-\mu},\ldots,S_0$ and the sets $\Omega_1,\ldots,\Omega_m$ are tangent convex approximations around the point $\hat{x}$ to the sets $S_1,\ldots,S_m$. Let us assume that we are given a set Ω_{m+1} which is a simplicial m+1 convex approximation around the point $\hat{x}$ to the set $S_{m+1}=S$. From Theorem 7.1, we know that the sets $\Omega_{-\mu},\ldots,\Omega_{m+1}$ will be separated. If we translate this last result in terms of the functions $\phi_{-\mu},\ldots,\phi_m$ and their gradients at the point $\hat{x}$, then, by an argument similar to the argument followed in the proof of Theorem 4.2, we obtain

Theorem 7.2. If $\hat{x}$ is an optimal solution of the given problem, then there exist numbers $\lambda_{-\mu},\ldots,\lambda_m$, not all zero, such that

(i) $\lambda_i \leq 0$ for $i=-\mu,\ldots,0$;

(ii) $\lambda_i\phi_i(\hat{x})=0$ for $i=-\mu,\ldots,-1$;

(iii) $\Sigma_{i=-\mu,\ldots,m}\ \lambda_i \text{grad } \phi_i(\hat{x})\cdot(x-\hat{x}) \leq 0$ for all $x \in \Omega_{m+1}$.

We conclude by making two remarks similar to the remarks made at the end of Section 4.

Remark 1. The inequality (iii) of Theorem 7.2 may be written under the form a Maximum Principle:

$$\Sigma_{i=-\mu,\ldots,m}\ \lambda_i \text{ grad } \phi_i(\hat{x})\cdot\hat{x} \geq \Sigma_{i=-\mu,\ldots,m}\ \lambda_i \text{grad } \phi_i(\hat{x})\cdot x \quad \text{for all } x \in \Omega_1.$$

Remark 2. If the point $\hat{x}$ is an interior point of Ω, then the inequality (iii) becomes

$$\Sigma_{i=-\mu,\ldots,m}\ \lambda_i \text{ grad } \phi_i(\hat{x}) = 0.$$

This will always be the case for the problems where $S_1=\Omega_1=R^n$.

REFERENCES

[1] Dubovitskii, A.Ya, and Milyutin, A.A., Extremum Problems in the Presnece of Restrictions, U.S.S.R. Computational Mathematics and Mathematical Statistics, 5, 1965, 1-79.

[2] Halkin, H., A Satisfactory Treatment of Equality and Operator Constraints in the Dubovitskii-Milyutin Optimization Formalism, Journal of Optimization Theory and Applications, 6, 1970, 138-149.

[3] John,F., Extremum Problems with Inequalities as Subsidiary Conditions, in "Studies and Essays:Courant Anniversary Volume",(K.O.Friedricks,O.E.Neugebauer,and J.J. Stoker,(eds.)),pp.187-204,Interscience Publishers,New York,1948.

[4] Abadie,J., On the Kuhn-Tucker Theorem, in "Nonlinear Programming",J.Abadie(ed.), pp.19-36,North-Holland,1967.

[5] Halkin,H., Optimal Control as Programming in Infinite Dimensional Spaces, in "C.I.M.E.:Calculus of Variations,Classical and Modern",pp.179-192,Eidizioni Cremonese,Roma,1966.

[6] Halkin,H. and Neustadt,L.W., Control as Programming in General Normed Linear Spaces, Lecture Notes in Operations Research and Mathematical Economics,Springer Verlag,11,1969,23-40.