APPROXIMATION AND CONVERGENCE IN NONLINEAR OPTIMIZATION

Hedy Attouch, Paris-Orsay
Roger J.-B. Wets

October 1980
WP-80-142

Supported in part by the National Foundation
Grant
ENG-7903731

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

ABSTRACT

We show that the theory of e-convergence, originally developed to study approximation techniques, is also useful in the analysis of the convergence properties of algorithmic procedures for nonlinear optimization problems.

APPROXIMATION AND CONVERGENCE IN NONLINEAR OPTIMIZATION

Hedy Attouch
Roger J.-B. Wets

INTRODUCTION

In the late 60's, motivated by the need to approximate difficult (infinite dimensional) problems instatistics [1], [2], stochastic optimization [3], variational inequalities [4], [5], [6] and control of systems, there emerged a new concept of convergence, called here e-convergence, for functions and operators. Since then a number of mathematicians have made substantial contributions to the general theory and have exploited the properties e-convergence to study a wide variety of problems, in nonlinear analysis [7], convex analysis [8], [9], partial differential equations [10], homogenization problems [11], (classical) variational problems [12], [13], optimal control problems [14] and stochastic optimization problems [15]. Some parts of this theory are now well understood, especially the convex case, see [32] for a survey of the finite dimensional results.

The objective of this paper is to exhibit the connections between e-convergence--basically an approximation scheme for unconstrained optimization--and the convergence of some algorithmic procedures for nonlinear optimization problems. Since we are mostly interested in the conceptual aspects of this relationship, it is convenient to view a constrained (or unconstrained) optimization problem, as the minimization of a function f defined on R^n and taking its values in the extended reals. Typically,

$$f(x) = \begin{cases} g_o(x) & \text{if } g_i(x) \leq 0 \qquad i = 1, \ldots, m, \\ +\infty & \text{otherwise;} \end{cases}$$

where for i = 0, 1,..., m, the functions g_i are (continuous and) finite-valued.

In section 2, we introduce and review the main properties of e-convergence in the nonconvex case. In particular we show that e-convergence of a collection of functions $\{f_\nu, \nu \in N\}$ to a function f, implies the convergence of the optimal solutions in a sense made precise in the second part of that section. The result showing that the set of optimal solutions is the limit inferior of the set of ε-optimal solutions of the approximating problems appears here for the first time. In section 3, we show that the so-called barrier functions, engender a sequence of functions that e-converge to f. From this all the known convergence results for barrier methods follow readily.

The relation between pointwise-convergence and e-convergence is clarified in section 4. It is shown that if the family $\{f_\nu, \nu \in N\}$ satisfies an equi-semicontinuity condition then e- and pointwise-convergence coincide. This equivalence is exploited in section 5 to give a (new) blitzproof of the convergence results for penalty methods. We also consider exact penalty methods.

Finally, in section 6, we introduce the notion of e/h-convergence for bivariate functions. It implies, in a sense made precise in section 6, the convergence of the saddle points. The theory and its application is not yet fully developed but as is sketched out in section 7, it can be used to obtain convergence results for multiplier methods.

It should be emphasized that we exploit here this approximation theory for optimization problems to obtain--and in some case slightly generalize--some convergence results for constrained optimization. There are many other connections that are worth investigating, in particular between e-convergence and sensitivity analysis [16-19], and the convergence conditions for algorithms modeled by point-to-set maps, see e.g., [20], [21] and the references given therein.

2. e-CONVERGENCE

Let f be a function defined on R^n and with values in the extended reals. By epi f, we denote the epigraph of f, i.e.,

$$\text{epi } f = \{(x,a) \in R^{n+1} \mid f(x) \leq a\} ,$$

by dom f, the effective domain of f, i.e.,

$$\text{dom } f = \{x \mid f(x) < +\infty\} .$$

Its hypograph is $\{(x,a) \mid a \leq f(x)\}$ or equivalently $\{(x,a) \mid (x,-a) \in \text{epi}(-f)\}$. The function f is l.sc. (lower semicontinuous) if epi f is closed or equivalently if to every $x \in R^n$ and $\varepsilon > 0$, there corresponds a neighborhood V of x such that for all $y \in V$,

$$f(y) \geq f(x) - \varepsilon .$$

The function is u.sc. (upper semicontinuous) if -f is l.sc.

Let $\{f_\nu, \nu \in N\}$ be a countable family of extended real-valued functions defined on R^n. The e-limit inferior, denoted by $li_e f_\nu$, is defined by: for $x \in R^n$,

$$(2.1) \qquad (li_e f_\nu)(x) = \inf_{\substack{M \subset N \\ \{x_\mu \to x, \mu \in M\}}} \liminf_{\mu \in M} f_\mu(x_\mu) ,$$

where M will always be an infinite (countable) subset of N. The e-limit superior, denoted by $ls_e f_\nu$, is defined similarly: for $x \in R^n$,

$$(2.2) \qquad (ls_e f_\nu)(x) = \inf_{\{x_\nu \to x, \nu \in N\}} \limsup_{\nu \in N} f_\nu(x_\nu)$$

Since $N \subset N$, and liminf $\leqslant$ limsup, we have that

(2.3) $$li_e f_\nu \leqslant ls_e f_\nu \quad .$$

Also, since $\{x_\nu = x, \nu \in N\} \subset \{x_\nu \to x, \nu \in N\}$ we have that

(2.4) $$li_e f_\nu \leqslant li\ f_\nu \text{ and } ls_e f_\nu \leqslant ls\ f_\nu$$

where $li\ f_\nu$, the <u>pointwise-limit inferior</u> of the family $\{f_\nu, \nu \in N\}$, is defined by

(2.5) $$(li\ f_\nu)(x) = \liminf_{\nu \in N} f_\nu(x)$$

and $ls\ f_\nu$, the <u>pointwise-limit superior</u>, is given by

(2.6) $$(ls\ f_\nu)(x) = \limsup_{\nu \in N} f_\nu(x) \quad .$$

Finally, we note that

(2.7) $$\mathrm{epi}(li_e f_\nu) = \underline{Ls}\ \mathrm{epi}\ f_\nu \quad ,$$

and

(2.8) $$\mathrm{epi}(ls_e f_\nu) = \underline{Li}\ \mathrm{epi}\ f_\nu \quad ,$$

where $\underline{Li}$ epi f_ν and $\underline{Ls}$ epi f_ν are respectively the limits inferior and superior of the family of sets $\{\mathrm{epi}\ f_\nu, \nu \in N\}$, i.e.,

(2.9) $$\underline{Li}\ \mathrm{epi}\ f_\nu = \{(x,a) = \lim_{\nu \in N}(x_\nu, a_\nu) \mid a_\nu \geqslant f_\nu(x_\nu)\} \quad ,$$

and

(2.10) $$\underline{Ls}\ \mathrm{epi}\ f_\nu = \{(x,a) = \lim_{\mu \in M}(x_\mu, a_\mu) \mid a_\mu \geqslant f_\mu(x_\mu), M \subset N\} \quad .$$

The properties of these limit sets are elaborated in [22, sect. 25]; in particular we note that they are closed. This means that both $li_e f_\nu$ and $ls_e f_\nu$ have closed epigraphs or equivalently are <u>lower semicontinuous</u> (l.sc.).

We say that the family $\{f_\nu, \nu \in N\}$ <u>p-converges</u> (converges pointwise) to a function f, written $f_\nu \to_p f$, if

$$(2.11) \qquad ls\, f_\nu \leq f \leq li\, f_\nu \quad .$$

It <u>e-converges</u>, written $f_\nu \to_e f$, if

$$(2.12) \qquad ls_e f_\nu \leq f \leq li_e f_\nu \quad ,$$

or equivalently, in view of (2.3) if

$$ls_e f_\nu = f = li_e f_\nu \quad .$$

In this case, from (2.7) and (2.8) it follows that

$$(2.13) \qquad \underline{Ls}\ \text{epi}\ f_\nu = \text{epi}\ f = \underline{Li}\ \text{epi}\ f_\nu \quad ,$$

i.e., the epigraph of f is the limit of the epigraphs. This is why we refer to this type of convergence, as e-convergence.

Our interest in e-convergence is spurred on by the fact that it essentially implies the convergence of the minima, this is made precise here below. Let

$$(2.14) \qquad A_\nu = \text{argmin}\ f_\nu = \{x \in R^n \mid f_\nu(x) = \inf f_\nu\}$$

and A = argmin f. Then, if $f_\nu \to_e f$

$$(2.15) \qquad \underline{Ls}\ A_\nu \subset A \ .$$

The relation is trivially satisfied if $\underline{\text{Ls}}\ A_\nu$ is empty--this occurs if and only if for any bounded subset D of R^n, $A_\nu \cap D = \phi$ for all ν is sufficiently large. Otherwise, suppose that for some $M \subset N$,

$$x_\mu \in A_\mu \quad \text{and} \quad x_\mu \to x \quad .$$

We need to show that $x \in A$. To the contrary suppose that there exists $\bar{x}$ such that $f(\bar{x}) < f(x)$. Hence, by e-convergence

$$(\text{ls}_e f_\nu)(\bar{x}) = f(\bar{x}) < f(x) = (\text{li}_e f_\nu)(x) \leq \liminf f_\mu(x_\mu) \quad .$$

Thus for some sequence $\{\bar{x}_\nu, \nu \in N, \bar{x}_\nu \to x\}$ and μ sufficiently large

$$f_\mu(\bar{x}_\mu) < f_\mu(x_\mu) \quad ,$$

contradicting the hypothesis that $x_\mu \in A_\mu$.

For $\varepsilon > 0$, we denote by ε-A, the set of points that are within ε of m, the infimum of f. Similarly for $\nu \in N$, let

$$m_\nu = \inf f_\nu \quad ,$$

and

$$\varepsilon\text{-}A_\nu = \{x \mid f_\nu(x) - \varepsilon \leq m_\nu\} \quad .$$

If $f_\nu \to_e f$ and $m_\nu \to m$, then

$$\text{(2.16)} \qquad \underline{\text{Li}}\ \varepsilon\text{-}A_\nu \subset \underline{\text{Ls}}\ \varepsilon\text{-}A_\nu \subset \varepsilon\text{-}A \quad ,$$

and whenever m is finite

$$\text{(2.17)} \qquad A = \bigcap_{\varepsilon > 0} \underline{\text{Li}}\ \varepsilon\text{-}A_\nu \quad .$$

Clearly to verify (2.16), it suffices to check the second inclusion. Suppose $x \in \underline{Ls}\ \varepsilon\text{-}A_\nu$, then by definition of $\underline{Ls}$, there exists $M \subset N$ and $\{x_\mu \to x, \mu \in M\}$ such that

$$f_\mu(x_\mu) \leq m_\mu + \varepsilon \ .$$

From this and the hypotheses, it follows that

$$f(x) \leq (li_e f_\mu)(x) \leq \liminf_{\mu \in M} f_\mu(x_\mu) \leq \lim m_\mu + \varepsilon = m + \varepsilon$$

and consequently $x \in \varepsilon\text{-}A$.

In view of (2.16) and the fact that $A = \bigcap_{\varepsilon>0} \varepsilon\text{-}A$, to verify (2.17), it suffices to derive the inclusion $A \subset \bigcap_{\varepsilon>0} \underline{Li}\ \varepsilon\text{-}A_\nu$. If $A = \phi$ the inclusion is trivially satisfied. Thus, suppose that $x \in A \neq \phi$. Since $f_\nu \to_e f$, it follows from (2.13) and (2.8) that there exists $\{(x_\nu, a_\nu) \in \text{epi } f_\nu, \nu \in N\}$ such that $(x_\nu, a_\nu) \to (x, m)$. The statement will be proved if given any $\varepsilon>0$, for ν sufficiently large $x_\nu \in \varepsilon\text{-}A_\nu$ or equivalently $a_\nu \leq m_\nu + \varepsilon$. To the contrary, suppose that for some $\varepsilon>0$, there exists $M_\varepsilon \subset N$ such that for all $\mu \in M_\varepsilon$,

$$m_\mu + \varepsilon < f_\mu(x_\mu) \leq a_\mu \ .$$

From this it would follow that

$$\lim m_\mu + \varepsilon = m + \varepsilon \leq m = \lim a_\mu \ ,$$

contradicting the working hypothesis.

It is noteworthy that although e-convergence always implies (2.15), in general this is not sufficient to imply that $m_\nu \to m$; even if all the quantities involved are finite, the functions $\{f_\nu, \nu \in N\}$ and f are convex and continuous, and the $\{A_\nu, \nu \in N\}$

and A are nonempty. The following example illustrates that situation: Let

$$f_\nu(x) = \begin{cases} -1 & \text{if } x \leq -\nu \\ \nu^{-1}x & \text{if } -\nu \leq x \leq 0 \\ x & \text{if } x \geq 0 \end{cases},$$

and

$$f(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ x & \text{if } x \geq 0 \end{cases}.$$

Then $m_\nu \equiv -1 \not\to m = 0$, $A_\nu =]-\infty,-\nu]$ and,

$$\underline{\text{Ls}}\, A_\nu = \phi \subset A =]-\infty,0] \quad .$$

(A variant of this example defines f_ν as $\nu^{-1}x$ on $x \leq 0$, with the same f as the e-limit function. Then $m_\nu \equiv -\infty \not\to m = 0$; here $A_\nu \equiv \phi$.)

However, if A is nonempty and m is finite, then e-convergence always implies that

$$(2.18) \qquad m \geq \limsup m_\nu \quad .$$

To see this, simply note that $(x,m) \in \text{epi } f$ implies, via (2.13) and the definition of $\underline{\text{Li}}$, that there exists $\{(x_\nu,a_\nu) \in \text{epi } f_\nu, \nu \in N\}$ such that $(x_\nu,a_\nu) \to (x,m)$. Since $a_\nu \geq m_\nu$ for all $\nu \in N$, we obtain (2.18) by taking limsup on both sides.

If in addition $A = \underline{\text{Li}} A_\nu$, or more generally if (2.17) is satisfied, then $m = \lim m_\nu$. From (2.17) and the definition of $\underline{\text{Li}}$, we have that to each $x \in A$ and $\varepsilon > 0$, there corresponds a sequence $\{x_\nu \in \varepsilon\text{-}A_\nu, \nu \in N\}$ converging to x. Hence

$$m = f(x) = (\text{li}_e f)(x) \leq \liminf_{\nu \in N} f_\nu(x_\nu) \leq \varepsilon + \liminf_{\nu \in N} m_\nu \quad ,$$

which with (2.18) implies that $m = \lim m_\nu$. Observe that we have shown that if m is finite and $f_\nu \to_e f$, then $m_\nu \to m$ if and only if (2.17) is satisfied.

Finally, even if $m = \pm\infty$ it is possible to obtain variants of (2.17) that are genuine to those cases. The development is somewhat technical and would lead us too far astray from the main subject.

3. BARRIER METHODS

To illustrate some of the implications of e-convergence, we derive (and slightly generalize) the standard convergence results for barrier methods as a consequence of the properties of e-convergence. (A. Fiacco has recently published an interesting and comprehensive survey of barrier methods[23].) We consider the nonlinear optimization problem

(3.1) $$\text{Minimize } g_0(x) \text{ subject to } g_i(x) \leqslant 0 \quad i = 1,\dots,m,$$

where for $i = 0,\dots,m$, the g_i are continuous real-valued functions defined on R^n. We assume that

$$\text{cl int } S = S = \{x | g_i(x) \leqslant 0, i=1,\dots,m\} \quad ,$$

i.e., S is the closure of its interior. Define

(3.2) $$f(x) = \begin{cases} g_o(x) & \text{if } x \in S \\ +\infty & \text{otherwise} \end{cases}$$

and

(3.3) $$f_\nu(x) = g_o(x) + q(\theta_\nu, x)$$

where the $\theta_\nu > 0$ are strictly increasing to $+\infty$ with ν, and

$$q\colon]0,\infty[\times R^n \to]0,\infty]$$

is continuous, finite if $x \in \text{int } S$ and $+\infty$ otherwise, and if $x \in \text{int } S$, $\theta \to q(\theta, x)$ is strictly decreasing to 0. In particular these properties of q imply that given any $x \in S$ and $\varepsilon > 0$, (3.4) $\exists (x_\nu \to x \text{ and } \nu_\varepsilon)$ such that $\forall \nu \geq \nu_\varepsilon, q(\theta_\nu, x_\nu) \leq \varepsilon$.

To see this, for a given $\varepsilon>0$, let $S_\nu = \{x | q(\theta_\nu, x) \leq \varepsilon\}$. The family of sets $\{S_\nu, \nu \in N\}$ are nested under inclusion and $\text{cl} \cup_{\nu\in N} S_\nu = S$, as follows from our assumptions. Hence ($\underline{\text{Ls}}\, S_\nu =$) $\underline{\text{Li}}\, S_\nu = S$, see e.g., [24, Prop. 1] and thus every x in S is the limit of a sequence $\{x_\nu \in S_\nu, \nu \in N\}$ from which (3.4) follows immediately.

The function q is called the barrier function. The most commonly used barrier functions are:

(3.5) $$q(\theta,x) = -\theta^{-1}\Sigma_{i=1}^{m}[\min(0,g_i(x))]^{-1}$$

(3.6) $$q(\theta,x) = \theta^{-2}\Sigma_{i=1}^{m}[\min(0,g_i(x)]^{-2}$$

(3.7) $$q(\theta,x) = -\theta^{-1}\Sigma_{i=1}^{m}\ln[\min(.5,-g_i(x))]$$

with the understanding that $\ln a = -\infty$ if $a \leq 0$. It is easy to see that these functions and many variants thereof satisfy the assumptions laid out here above.

Next, we show that $f_\nu \to_e f$. We begin with $\text{ls}_e f_\nu \leq f$. The inequality is clearly valid if $x \notin S$. If $x \in S$, from (2.14) and the continuity of g_o, it follows that given any $\varepsilon>0$, we can always find $\{x_\nu, \nu\in N\}$ converging to x, such that for ν sufficiently large

$$g_o(x_\nu) - g_o(x) \leq \varepsilon .$$

Thus

$$(\text{li}_e f_\nu)(x) \leq \limsup_{\nu\in N} f_\nu(x_\nu) \leq \limsup_{\nu\in N} g_o(x_\nu) + \limsup_{\nu\in N} q(\theta_\nu,x_\nu) \leq 2\varepsilon + f(x)$$

which yields the desired inequality since ε is arbitrary. Again $f \leq \text{li}_e f_\nu$ is trivially satisfied if $x \notin S$. If $x \in S$, let $\{x_\mu, \mu\in M\subset N\}$ be arbitrary sequence converging to x. By continuity of g_o, we have that for any $\varepsilon>0$ and sufficiently large,

$g_0(x) - \varepsilon \leqslant g_0(x_\mu)$. A fortiori, since $q(\theta,x) > 0$

$$f(x) - \varepsilon = g_0(x) - \varepsilon \leqslant g_0(x_\mu) + q(\theta_\mu,x_\mu) = f_\mu(x_\mu) \quad ,$$

Thus

$$f(x) - \varepsilon \leqslant \limsup f_\mu(x_\mu) \quad .$$

This holds for every $\varepsilon>0$ and every sequence $\{x_\mu, \mu\in M \subset N\}$ converging to x, hence $f(x) \leqslant li_e f_\nu$.

Since the f_ν e-converge to f, it follows from (2.15) that if for each ν, x^*_ν minimize f_ν and x^* is any cluster point of the sequence $\{x^*_\nu, \nu\in N\}$, then x^* minimize f, i.e., solves (3.1). Note that if f is inf-compact--i.e., if for some $a \in R$, the set $S_a = S \cap \{g_0(x) \leqslant a\}$ is nonempty and bounded--then not only is A nonempty but also for every ν, $\phi \neq A_\nu \subset S_a$. Thus in this case, we are guaranteed to find approximate solutions to (3.1) by minimizing the "unconstrained" functionals f_ν. (The unconstrained minimization of the f_ν, must start from a feasible point, there are a number of ways to do this. A. Fiacco [23, p.400-401] has suggested a method that can be viewed as a phase I barrier method.)

Also, the convergence of parameter-free barrier methods can be handled in this framework. For example, consider the sequence of functions

$$(3.8) \qquad f_\nu(x) = g_0(x) + [g_0(x^*_{\nu-1}) - g_0(x)]^{-2} \sum_{i=1}^{m} [\min(0,g_i(x))]^{-1}$$

where $x^*_{\nu-1}$ minimizes $f_{\nu-1}$. Under some regularity conditions [25] these penalty functions have the same properties as those considered at the beginning of this section.

4. e-CONVERGENCE AND p-CONVERGENCE

Sometimes it might be easier to verify p-convergence (pointwise) than e-convergence. It is thus useful to make explicit the relationship between these two types of convergence. Unfortunately, neither implies the other. To see this simply consider the collection (of l.sc. convex) functions.

$$f_\nu(x_1,x_2) = \nu x_1 \text{ on dom } f_\nu = \{(x_1,x_2) \mid x_1 \leqslant 0, \nu x_1 \geqslant x_2\} ,$$

that e-converges to

$$f(x_1,x_2) = x_1 \text{ on dom } f = \{(x_1,x_2) \mid x_1 \leqslant 0, x_2=0\}$$

and p-converges to

$$f'(x_1,x_2) = 0 \text{ on dom } f' = \text{dom } f .$$

However, if the collection is equi-l.sc. then e- and p-convergence imply the other [26,4_p and 5_p]. The family $\{f_\nu, \nu \in N\}$ is <u>equi-l.sc.</u> if there exists a subset of $D \subset R^n$ such that conditions (4.1) and (4.2) are satisfied:

(4.1) To each $x \in D$, and $\varepsilon > 0$, there corresponds a neighborhood V of x and ν_ε such that for all $y \in V$ and all $\nu \geqslant \nu_\varepsilon$

$$f_\nu(y) \geqslant f_\nu(x) - \varepsilon ,$$

(4.2) To each $x \notin D$, and $\eta \in R$, there corresponds a neighborhood V of x and ν_η such that for all $y \in V$ and $\nu \geqslant \nu_\eta$

$$f_\nu(y) \geqslant \eta .$$

If the functions are finite-valued then equi-continuity--and a fortiori equi-Lipschitz--will imply equi-l.sc. but for our purposes those conditions are too restrictive since we view the f_ν as representing optimization problems, possibly involving constraints, and thus at best l.sc. and usually taking on the value $+\infty$. The equi-l.sc. condition is in some sense minimal since $f_\nu \to_e f$ and $f_\nu \to_p f$ imply (4.1) and (4.2) with $D = \operatorname{dom} f$ [26, 3_p].

5. (EXTERIOR) PENALIZATION METHODS

The relation between p- and e-convergence can be exploited to yield the convergence of penalization methods. The results are not new but the proof should help in coming to grips with the concept of equi-lower semicontinuity. We consider the nonlinear optimization problem:

$$\text{(5.1)} \qquad \begin{array}{lll} \text{Minimize } g_0(x) & & \\ \text{Subject } g_i(x) \leq 0 & & i = 1,\ldots, m \\ \qquad\quad\;\; g_i(x) = 0 & & i = m+1,\ldots, \bar{m} \end{array}$$

where for $i = 0,\ldots, m$, the g_i are continuous real-valued functions defined on R^n. By S we denote the set of feasible solutions. Define

$$\text{(5.2)} \qquad f(x) = \begin{cases} g_0(x) & \text{if } x \in S \\ +\infty & \text{otherwise} \end{cases}$$

and

$$\text{(5.3)} \qquad f_\nu(x) = g_0(x) + p(\theta_\nu, x)$$

where the θ_ν are strictly increasing with ν to $+\infty$, and

$$p :]0,\infty[\times R^n \to [0,\infty[$$

is continuous, nonnegative and finite; if $x \in S$ then $p(\theta,x,) = 0$, otherwise $\theta \to p(\theta,x)$ is increasing uniformly to $+\infty$ on compact subsets of $R^n \setminus S$.

All common (exterior) penalty functions satisfy these conditions, as can easily be verified. For example

$$(5.4) \qquad p(\theta,x) = \theta \Sigma_{i-1}^{m} [\max (0,g_i(x))]^{\alpha} + \theta \Sigma_{i=m+1}^{\bar{m}} | g_i(x) |^{\beta}$$

with $\alpha \geqslant 1$ and $\beta \geqslant 1$.

It is obvious that the collection $\{f_\nu, \nu \in N\}$ is equi-l.sc. - (4.1) and (4.2) are trivially satisfied with D = dom f--and that $f_\nu \to_p f$, hence by the results alluded to in the previous section $f_\nu \to_e f$. From (2.15) it follows that if the x_ν^* minimize the f_ν, then any cluster point x^* of the sequence $\{x_\nu^*, \nu \in N\}$ solves (5.1). As for barrier methods, the inf-compactness of f will grarantee the existence of the x_ν^* and of some cluster point x^* that solves the original problem.

Some results for <u>exact penalty functions</u> can also be derived directly from the general theory. If $x \in A_\nu$, for all ν larger than some $\bar{\nu}$, then from (2.15) it follows that $\bar{x} \in A$ and thus solves (5.1). This is the sufficiency theorem of Hahn and Mangasarian [27, Theorem 4.1].

On the other hand suppose that we are in the situation when the sequence of optimal solutions $\{x^*_\nu, \nu \in N\}$ admit x^* as a cluster point. If we assume that g_0 is locally Lipshitz--at least at x^*--then provided that the "slope" at x^* of $x \to p(\theta,x)$ on $R^n \setminus S$ becomes sufficiently steep, there will exist $\bar{\theta}$ such that for all $\theta_\nu \geqslant \bar{\theta}$, $x^* \in A_\nu$. By "slope" we mean here the following quantity:

$$\liminf_{V_\alpha \to \{x^*\}} \inf_{y \in V \cap (R^n \setminus S)} [p(\theta,y)/|y - x^*|]$$

where the $\{V_\alpha\}$ are nested collections of neighborhoods V_α of x^* such that $\cap V_\alpha = \{x^*\}$. For specific forms of the function p such as (5.4), more detailed conditions can be worked out; see e.g., [27, Theorem 4.4].

6. CONVERGENCE OF BIVARIATE FUNCTIONS

A number of algorithms for constrained optimization problems construct not only a sequence of approximate solutions but simultaneously build up approximates for the Lagrange multipliers. To study this type of convergence it is necessary to introduce a notion of convergence for bivariate functions that would have properties similar to e-convergence in the univariate case. Such a concept has been introduced recently by the authors [28], [29] and independently in the convex-concave case by Bergstrom and McLinden [30]. We shall only give here a sketchy description of e/h-convergence, all the implications having not yet been completely worked out.

Let $\{H_\nu, \nu \in N\}$ be a family of bivariate functions defined on $R^n \times R^m$ with values in $[-\infty, +\infty]$. A bivariate function H must be viewed as a representant of an equivalence class, D(H) is the subset of $R^n \times R^m$ on which the members of the class are defined without any ambiguity, see [31] for a detailed analysis. We say that the H_ν e/h-converge to a member H of an equivalence class of bivariate functions, if for all $(x,y) \in D(H)$, we have that

(6.1) for all $M \subset N$ and every sequence $\{x_\mu, \mu \in M \mid x_\mu \to x\}$, there exists $\{y_\mu, \mu \in M \mid y_\mu \to y\}$ such that

$$\liminf_\mu H_\mu(x_\mu, y_\mu) \geq H(x,y);$$

(6.2) for all $M \subset N$ and every sequence $\{y_\mu, \mu \in M \mid y_\mu \to y\}$, there exists $\{x_\mu, \mu \in M \mid x_\mu \to x\}$ such that

$$\limsup_\mu H_\mu(x_\mu, y_\mu) \leq H(x,y) \quad .$$

We refer to this type of convergence as e/h-convergence because the epigraph of $x \to H(x,y)$ is the limit of the epigraphs of $x \to H_\nu(x,y')$ with y' converging to y and the hypograph of $y \to H(x,y)$ is the limit of the hypographs of $y \to H_\nu(x',y)$ with x' converging to x. From this it follows that if H is the e/h-limit of a sequence of bivariate functions, it is necessarily lower semicontinuous with respect to x and upper semicontinuous with respect to y. For our purposes, the main consequence of the e/h-convergence of a family of bivariate functions is the implied convergence of the saddle points. More specifically: Suppose that for some $M \subset N$, the (x_μ, y_μ) are saddle points of the function H_μ, i.e., for all $y \in R^m$ and all $x \in R^n$, we have that

$$(6.3) \qquad H_\mu(x_\mu,y) \leq H_\mu(x_\mu,y_\mu) \leq H_\mu(x,y_\mu) \quad .$$

We assume that for all μ, $H_\mu(x_\mu,y_\mu)$ are finite. Moreover, suppose that the $\{H_\nu, \nu \in N\}$, e/h-converge to H, $(\bar{x},\bar{y}) = \lim_{\mu \in M}(x_\mu,y_\mu)$ and $(\bar{x},\bar{y}) \in D(H)$. Then $(\bar{x},\bar{y})$ is a saddle point of H with

$$(6.4) \qquad H(\bar{x},y) \leq H(\bar{x},\bar{y}) \leq H(x,\bar{y});$$

assuming again that $H(\bar{x},\bar{y})$ is finite.

To prove the assertion, we proceed by contradiction. Suppose that $(\bar{x},\bar{y})$ is not a saddle point. Then at least one of the two inequalities appearing in (6.4) must fail; without loss of generality, let us suppose that there exists x_ε such that

$$H(x_\varepsilon,\bar{y}) < H(\bar{x},\bar{y}) \quad .$$

Since $y_\mu \to \bar{y}$, by definition of e/h-convergence (6.2), there exists $\hat{x}_\mu \to x_\varepsilon$ such that

$$(6.5) \qquad \limsup H_\mu(\hat{x}_\mu,y_\mu) \leq H(x_\varepsilon,\bar{y}) \quad .$$

Recall that (x_μ,y_μ) is a saddle point which means that

$$H_\mu(x_\mu,y_\mu) \leq H_\mu(\hat{x}_\mu,y_\mu)$$

Taking liminf on both sides, we get

$$H(\bar{x},\bar{y}) \leq \liminf_{\mu \in M} H_\mu(x_\mu,y_\mu) \leq \liminf_{\mu \in M} H_\mu(\hat{x}_\mu,y_\mu) \quad ,$$

which combined with (6.5) yields

$$H(\bar{x},\bar{y}) \leqslant \liminf H_\mu(\hat{x}_\mu,y_\mu) \leqslant \limsup H_\mu(\hat{x}_\mu,y_\mu) \leqslant H(x_\varepsilon,\bar{y})$$

contradicting the working typothesis.

7. METHOD OF MULTIPLIERS

Our only purpose is to illustrate the potential use of the concept of e/h-convergence for bivariate functions to obtain convergence proofs for multiplier methods. We consider the problem

$$(7.1) \qquad \text{Minimize } g_0(x) \text{ subject to } g_i(x) = 0 \quad i=1,\ldots,m$$

where for $i=0,\ldots,m$, the functions g_i are continuous. As usual by $S = \{x \mid g_i(x)=0, i=1,\ldots,m\}$, we denote the feasibility region. The approximation to (7.1) are given by

$$(7.2) \qquad \text{Minimize } g_0(x) \text{ subject to } g_i(x) = \theta_i \quad i=1,\ldots,m$$

The idea being to have the θ_i tend to zero and the problems (7.2) would, in some sense, converge to (7.1). However, it is not quite in that form that we design the approximation scheme. To (7.2) we associate the bivariate function

$$(7.3) \qquad H_\nu(x,\theta) = g_0(x) + \tfrac{1}{2}\Sigma_{i=1}^{m}[(g_i(x)-\theta_i)^2\sigma_\nu - \theta_i^2\sigma_\nu^2]$$

As $\sigma_\nu \uparrow +\infty$, the family $H_\nu(x,\theta)$ e/h-converges to a member of H of an equivalence class of bivariate which on D(H) takes on the form

$$(7.4) \qquad H(x,\theta) = \begin{cases} g_0(x) & \text{if } x \in S \text{ and } \{\theta=0\} \\ +\infty & \text{if } x \notin S \text{ and } \{\theta=0\} \\ -\infty & \text{if } x \in S \text{ and } \{\theta\neq 0\} \end{cases}$$

To see this simply observe that if $(x,\theta) \in D(H)$ and a sequence $\{x_\mu, \mu \in M\}$ converges to x for some $M \subset N$, then simply setting $\theta_\mu \equiv \theta$, we see that (6.1) is satisfied, similarly if a sequence $\{\theta_\mu, \mu \in M\}$ converges to θ, then with $x_\mu \equiv x$ we obtain (6.2). Thus if the saddle points of the bivariate functions H_ν admit a cluster point in $D(H)$ it will be a saddle point if H and hence an optimal solution of (7.1).

Assuming that for $i=0,m$, the functions g_i are differentiables then if (x^ν, θ^ν) is a saddle point of H_ν satisfies the equations:

$$(7.5) \qquad \nabla g_0(x^\nu) + \Sigma_{i=1}^m \sigma_\nu (g_i(x^\nu) - \theta_i^\nu) \nabla g_i(x^\nu) = 0 \quad ,$$

$$(7.6) \qquad \theta_i^\nu = -g_i(x^\nu)/(\sigma_\nu - 1)$$

Substituting in (7.5) it yields

$$(7.7) \qquad \nabla g_0(x^\nu) + \sigma_\nu (1-\sigma_\nu^{-1})^{-1} \Sigma_{i=1}^m g_i(x^\nu) \ \nabla g_i(x^\nu) = 0$$

These conditions suggest a "multiplier method", where we solve (7.7), adjust θ^ν by means of (7.6) and then repeat. The method is just a variant of a penalty method and hence will be exact under some regularity conditions.

REFERENCES

[1] Burkholder, D., and R. Wijsman. 1963. Optimum properties and admissibility of sequential tests. Ann. Math. Statist. 34:1-17.

[2] Van Cutsem, B. 1971. Eléments Aléatoires à Valeurs Convexes Compactes. Thèse, Grenoble.

[3] Van Cutsem, B. 1972. Problems of convergence in stochastic linear programming. Pages 445-454 in: Techniques of Optimization, edited by A. Balakrishnan. New York: Academic Press.

[4] Lions, J-L., and G. Stampacchia. 1967. Variational inequalities. Commun. Pure Appl. Math. 20: 493-519.

[5] Mosco, U. 1967. Approximation of the solutions of some variational inequalities. Ann. Scuola Normale Sup. Pisa 21: 373-394; ibiol 765.

[6] Mosco, U. 1969. Convergence of convex sets and of solutions of variational inequalities. Adv. Math. 3: 510-585.

[7] Brezis, H. 1972. Opérateurs Maximaux Monotones et Semi-groupes de Contractions dans les Espaces d'Hilbert, Lecture Notes 5, Amsterdam: North-Holland.

[8] Joly, J-L. 1970. Une Famille de Topologies et de Convergences sur l'Ensemble des Fonctionelles Convexes. Thèse, Grenoble.

[9] Robert, R. 1974. Convergence de fonctionelles convexes. J. Math. Anal. Appl. 45:533-535.

[10] Attouch, H. 1976. E.D.P. Associées à des Familles de Sous-différentiels. Thèse, Paris VI.

[11] Attouch, H. 1979. Sur la T-Convergence, Séminaire Brézis-Lions, Collège de France.

[12] De Giorgi, E. 1975. Sulla convergenza di alcune successioni di integrali del tipo del l'area. Rend. Mat. Univ. Roma 8:277-294.

[13] Sbordone, C.C. 1975. Su alcune applicazioni di un tipo di convergenze variazionale. Ann. Scuola Norm. Pisa IV. 2:617-638.

[14] Zolezzi, T. 1978. Characterization of some variational perturbations of the abstract linear-quadratic problem. Siam J. Control & Optimiz. 16:106-121.

[15] Salinetti, G. 1978. Convergence for Measurable Multi-functions: An Application to Stochastic Optimization. Tech. Report, Univ. Roma.

[16] Robinson, S. 1979. Generalized equations and their solutions, Part 1: Basic theory. Mathematical Programming Study 10:128-141.

[17] Robinson, S. 1975. Stability theory for systems of linear inequalities, Part I: Linear systems. Siam J. Numer. Anal. 12:754-769.

[18] Spingarm, J. 1980. Fixed and variable constraints in sensitivity analysis. Siam J. Control & Optimiz. 18:297-310.

[19] Sargent, R. 1979. On the Paramentric Variation of Constraint Sets and Solutions of Minimization Problems. Tech. Report, Imperial College, London.

[20] Meyer, G. 1979. Asymptotic properties of sequences iteratively generated by point-to-set maps. Mathematical Programming Study 10:115-127.

[21] Tishadhigama S., E. Polak and R. Klessig. 1979. A comparative study of several general convergence conditions for algorithms modeled by point-to-set maps. Mathematical Programming Study 10: 172-190.

[22] Kuratowski, C. 1958. Topologie I, Warzawa: Panstwone Wydawnicto Naukowe.

[23] Fiacco, A. 1979. Barrier methods for nonlinear programming. In: Operations Research Support Methodology, edited by A. Holzman, M. Dekker, New York.

[24] Salinetti, G., and R. Wets. 1979. On the convergence of sequences of convex sets in finite dimensions. Siam Review 21:16-33.

[25] Fiacco, A., and G. McCormik. 1967. The sequential unconstrained minimization technique without parameters. Operations Research 15:820-827.

[26] Dolecki, S., and R. Wets. 1980. Convergence of Functions Equi-semicontinuity. Tech. Report, Univ. Kentucky.

[27] Han, S.-P., and O. Mangasarian. 1978. Exact penalty functions in nonlinear programming. Tech. Report, MRC, Univ. Wisconsin.

[28] Attouch, H., and R. Wets. 1980. A Variational Concept of Convergence for Bivariate Functions. Tech. Report, Univ. Kentucky.

[29] Attouch, H., and R. Wets. 1980. Variational Convergence of Convex-concave Bivariate Functions on Reflexive Banach Spaces. Tech. Report, MRC, Univ. Wisconsin.

[30] Bergstrom, R., and L. McLinden. Private Communication.

[31] Rockafellar, R. 1970. Convex Analysis. Princeton: Princeton Univeristy Press.

[32] Wets, R. 1980. Convergence of convex functions, variational inequalities, and convex optimization problems. Pages 375-403 in: Variational Inequalities and Complementarity Problems, edited by R. Cottle F. Gianessi and J.-L. Lions. Chichester: J. Wiley & Sons.